The Demand for Density: Evidence from New York City

Benjamin K. Couillard

November 2022

Preliminary and incompletee.
Some of the figures in the data appendix have not been recently updated. See the data section in the body of the paper for accurate numbers.

1 Introduction

Cities provide large productivity externalities (Rosenthal and Strange, 2004) but are afflicted by regulation-induced land scarcity (Gyourko and Molloy, 2015), which is why in recent years researchers have attempted to quantify the implications of zoning liberalization so that more workers can live in the most productive regions (Hsieh and Moretti, 2019; Duranton and Puga, 2019). Other research has focused on the role of endogenous urban amenities, like neighbourhood population characteristics, in location choice (Diamond, 2016; Couture et al., 2019; Bayer et al., 2007; Almagro and Domınguez-Iino, 2019). Although the existence of zoning regulations implies that households prefer lower-density neighbourhoods, and research has found that households prefer more regulated neighbourhoods (Chiumenti et al., 2022; Kulka, 2019; Song, 2021), to my knowledge no research studying the counterfactual effects of zoning reform has incorporated this countervailing force. Furthermore, senior urban economists have recently written that "[we] know virtually nothing about urban costs and how they vary with city population," (Combes et al., 2018). The contribution of this paper is to estimate the demand for population density itself using publicly-available data, and I find suggestive evidence that a general urban cost does exist: for white and hispanic households density is a homothetic bad valued at 1.5% of income for a 10% increase in density, although statistically significant only for those with the highest income. If these results can be replicated with better data and greater precision, then the effects of zoning reform in the literature would be overstated since households have disutility over dense environments.

To estimate the demand for density, I consider population density as a location characteristic that affects utility in a neighbourhood choice model. However, high-utility locations will mechanically have larger populations, so there is a severe endogeneity problem. Additionally, the urban housing supply crisis has resulted in price endogeneity that is more severe than usual: households may make location decisions on the basis of financial returns, a dynamic consideration that may bias estimates from a static model. To address the latter concern, I adapt the dynamic neighbourhood choice model of Bayer et al. (2016), which purges any dynamic price bias by including the possibility of wealth changes resulting from price changes in the continuation value of a choice-specific value function. Furthermore, the marginal utility of wealth is identified from the effect of mostly observable financial moving costs on the probability of not moving, which can then be used to (1) partial out the effect of static price on utility and (2) obtain the willingness to pay for other characteristics (namely density). Finally, I address the endogeneity of density by using instrumental variables in the second-step regressions.

For identifying variation in population density, I leverage variation in the co-location of wide streets and various zoning categories, which permits more relaxed zoning along several policy dimensions. The instrument is the share of residential land that meets multiple conditions simultaneously and qualifies for more relaxed zoning, which I refer to as the "microgeographic interaction" of the conditions. Because zoning and the street grid certainly affect utility directly, it is important to include the share of residential land that meets these conditions individually, and their "multiplicative interaction," as controls. The multiplicative interaction is interpreted as the expectation of the instrument if the distributions of the conditions (zoning and street width) are independent, and so the identifying variation is the residual variation in the co-location of wide streets and zoning categories after conditioning on the expectation. A failure of the exclusion restriction would require households to have utility over the co-location of wide streets and zoning categories below the neighbourhood level.

The BMMT framework requires observing households as they make repeated location decisions over time, which I accomplish using New York City's Pluto data. This data is a panel of properties spanning two decades and counting, crucially including the name of the owner. Using string-matching techniques and sales microdata, I construct a panel of households, following movers across properties. To this I join imputed household race and income, using a naive Bayesian classifier of races based on names, and the distribution of income for households of a given race who purchase in a given tract-year respectively. I validate the data construction and imputation against mover flows observed in ACS microdata (Ruggles et al., 2022).

Finally, I propose a definition of general equilibrium in the dynamic BMMT context and a solution method. Prices are set by a shift along a local housing supply curve estimated by Baum-Snow and Han (2019), and are solved in the present (without uncertainty) and in the future (with uncertainty) with a fixed point algorithm. The procedure also leverages the simple function relating present and future value functions (a reduced-form regression model from the estimation phase) and a Taylor approximation to the probability that a household does not move.

Secion 2 describes the data. Section 3 explains the model and my contribution to it. Section 4 presents first-step results from the BMMT procedure. Section 5 discusses the instrument, and section 6 presents IV results. Section 7 concludes.

2 Data

2.1 Overview

To estimate a BMMT-style dynamic model generally requires proprietary data in order to construct a panel of fine household location choices, along with several important household-level characteristics. I build such a panel using New York City's Pluto dataset, a panel of properties that crucially includes the names of owners, along with sales microdata. More detail on data construction can be found in the appendix.

I first remove obvious non-owner occupants (*e.g.*, Smith Reality Corp.) and consolidate plausibly single-household names within a property (*e.g.*, John and Jane Smith, Hillary Rodham and Hillary Clinton). Then I change the dates of owner household changes to align with dates of sale. I compare the names of households that sold a property with those that bought one at a similar time, setting a cutoff of four years. I also drop households that owned two properties for an overly long spell of time.

In the end, I arrive at a panel of 290,000 households that contribute 2.91 million household-years. Additionally, there are 22,000 households that move, contributing 364,000 household-years. Movers are 7.7% of households but contribute a relatively higher 12.5% of household-years because non-movers are observed for one "spell" at a property, while movers are observed for (almost always) two spells that are slightly shorter on average. However, there are still important household-level variables that are not in the data.

I classify households into four broad racial categories (White, Black, Asian, Hispanic) based on their names using a naive Bayesian classifier. This method uses data on the racial distributions of first and last names, from the Census Bureau, mortgage applications (Tzioumis, 2018), and voter registration files for six Southern states (Rosenman et al., 2022). Rather than take the most likely race based on a name, I draw from the output probability distribution over races for each household, and I choose the prior such that these draws replicate the racial distribution of homeowners in the sample region and period as measured in the ACS. Although this prior is not exactly "prior" to observing the data, it does result in better benchmarking results, discussed below.

Then, I draw income from distributions defined by purchase-year, purchase-tract, and race in public HMDA data. When there are zero buyers in a relevant HMDA cell (this may occur due to measurement error), I define the distribution as centred on the prediction from a regression model with race-tract fixed effects and separate year fixed effects, and with spread defined by the race-tract residuals.

2.2 Benchmarking Mover Flows to ACS

I benchmark normalized flows of movers against the 2005-2020 ACS (Ruggles et al., 2022) in order to test the accuracy of my hard-won data. Specifically, I use ACS respondents who own their homes, live in single unit structure, hae a mortgage, and reported moving from a different house in the last year. The county (and hence, borough) of residence one year ago is provided, as is the PUMA of current residence. I also disaggregate by race, year, and income (above or below media), so that each observation is a borough-PUMA-year-race-income flow. Since the ACS provides population estimates, but the administrative Pluto data is essentially a sample (due to many dropped observations for data quality purposes), I normalize the borough-PUMA-race-year flows by taking them as a share of the sum of all flows in the data source. For the purpose of merging Pluto to the ACS I need to choose a year for the Pluto flow when the year of sale and purchase are not the same; the year of sale outperforms the year of purchase and a random year between the two.

The scatterplot of these flows, along with the 45 degree line, best fit line, and local regression line, is provided as Figure 1. All three lines are close together, and the local regression line only departs from the other two at the extremes of the distribution, when the variance of the estimator increases since individual observations become more influential. Since the ACS is less granular than the Pluto flows, very small flows tend to either be missed by the ACS sampling process or included in the ACS but weighted to be larger than they should; the local regression shows that these two cases are balanced such that the conditional expectation lies close to the 45 degree line.

Next, I examine the different components of variation by including fixed effects for each combination of four our of the five sets of categories that define the cells, as well as a baseline that includes no fixed effects. For example, in the second column of table 1, I include destination-year-race-income fixed effects, so that all variation occurs within a destination-year-race-income, *i.e.*, over origins, indicated in the table footer. In three out of the five fixed-effects models in 1, the coefficient is close to one, indicating that the variation under inspection in the Pluto flows is representative of the same variation in the population measured in the ACS. The under-performing sources of variation are time and income, which, although positively correlated with the ACS flows, have little explanatory power.

To conclude, this validation exercise shows that the household mobility and race data derived from Pluto is representative of population analogues. A decomposition of the different sources of variation shows that all five sources (origin, destination, race, and year) of variation in Pluto are positively correlated with the same sources in the ACS. However, between-year and between-income variation is substantially less correlated with the population variation.

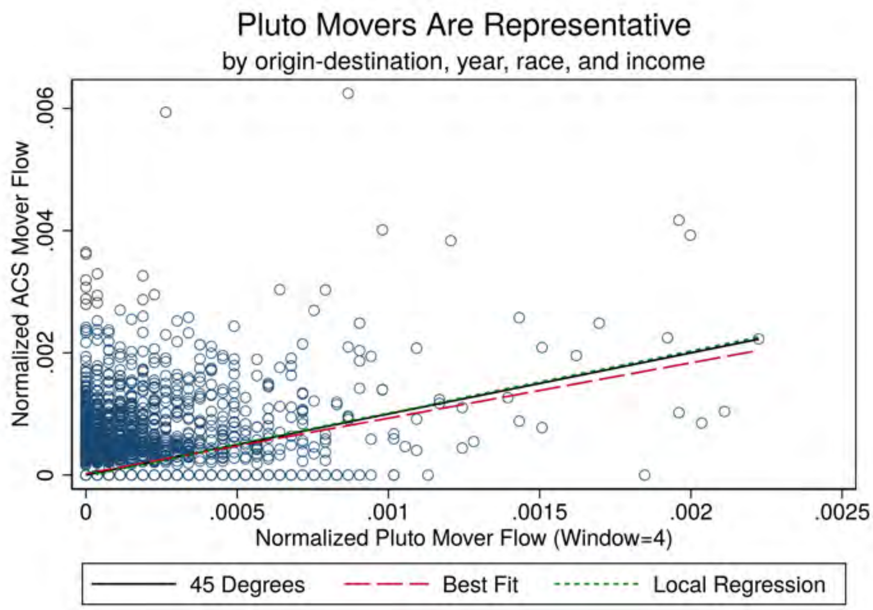

Figure 1: Scatterplot of normalized origin-destination-race-year mover flows in Pluto versus ACS

	ACS	ACS	ACS	ACS	ACS	ACS
Pluto Flow	0.906	1.093	1.023	0.225	0.807	0.171
	(0.0664)	(0.0727)	(0.0688)	(0.119)	(0.0817)	(0.134)
variation	all	origin	destination	year	race	income
r2_within		0.176	0.156	0.00437	0.0990	0.00134
r2	0.129	0.435	0.211	0.345	0.416	0.635

Standard errors in parentheses

Table 1: Regression of ACS flows on Pluto flows, with four-way fixed effects. Remaining variation is indicated in the table footer. Robust standard errors.

3 Model and First Stage Estimation

This is a dynamic discrete choice model, where agents choose one neighbourhood to live in (or the outside option) while taking into account expected future amenity values since moving is costly. Houses within a neighbourhood are taken to be identical. Households are discretized into types τ defined by immutable characteristics (in this case race and income) and endogenous housing wealth. As in the static case, households receive idiosyncratic preference shocks for each alternative, such that the conditional choice probabilities (CCPs) of choosing an alternative are a logistic function of value functions. Households take into account that choosing a different location from the one they enter the period with incurs a moving cost, after paying that moving cost they transition to a lower wealth level, and changes in the price of the chosen alternative also cause transitions to a different wealth level. The moving cost depends directly on household characteristics, interpreted as psychological moving costs, and on the interaction of household characteristics and the price of the sold house, interpreted as financial moving costs.

3.1 Value Functions

Naturally, households are modeled as choosing the neighbourhood j that maximizes the sum of present and expected future utility in every period, motivating the following Bellman equation:

$$V(s_{it}, \epsilon_{it}) = \max_j \Big\{ \{ v_{jt}^{\tau(\bar{s}_{it})}(\bar{s}_{it}) - Z^{\tau(\bar{s}_{it})\prime}\gamma_p + \phi \mathbb{1}(g_i = G_j) + \epsilon_{ijt} \}_{j<J+1}, \\ \{ v_{jt}^{\tau(s_{it})}(s_{it}) + \phi \mathbb{1}(g_i = G_j) + \epsilon_{ijt} \}_{j=J+1} \Big\}$$

The solution to this problem is the index of the optimal location j, which is a function of state variables $d_{it} = d^*_{it}(s_{it}, \epsilon_{it})$. Most but not all of individual i's characteristics can be discretized into a type τ, which I will write for now as a mapping of the state $\tau(s_{it})$. The set of state variables is $s_{it} = (X_t, \xi_t, Z_{it}, h_{it}, g_i)$ and their lags, where X is observable location characteristics, ξ is the unobservable location characteristic, Z is observable household characteristics, g is the unobservable household characteristic, and h_{it} is the index of the neighbourhood that household i resides in at time t. The household's type τ is determined by observable characteristics Z but not g,[1] since it is not observable, nor h, since a household of type τ may in principle decide to live anywhere.

We denote the index of deciding to stay in the same neighbourhood as $d_{it} = J + 1$. When this does not occur, psychological (non-financial) moving are incurred, which are assumed to be (1) invariant to the origin or destination neighbourhood, (2) linear in observed household characteristics, and (3) eval-

[1]Because of this relationship I choose to write $Z^{\tau(s_{it})} = Z_{it}$.

uated at the post-move characteristics.[2] We denote post-move characteristics, which are identical except for a decrease in wealth due to moving costs, with a bar: $Z \to \bar{Z}$, $s \to \bar{s}$, $\tau \to \bar{\tau}$.[3] Also, choosing some $j \neq J+1$ causes a different value function v_{jt}^{τ} to be operative, described below. There is unobserved heterogeneity in taste for broad location, and when a household lives in a neighbourhood that matches with these preference it receives an additional ϕ utility. As usual, there are idiosyncratic shocks to taste for neighbourhoods ϵ_{ijt}, which are assumed to be additively separable from utility and conditionally independent from observable states (Rust, 1987).

Under the assumptions specified above, we can write the choice-specific value function as the sum of flow utility and expected future utility,

$$v_{jt}^{\tau(s_{it})}(s_{it}) = u_{jt}^{\tau(s_{it})} + \beta \mathbb{E}\Bigg[\ln \bigg(\exp\big(v_{J+1,t+1}^{\tau(s_{i,t+1})}(s_{i,t+1}) + \phi \mathbb{1}(g_i = G_{J+1})\big) \\ + \sum_{k=0}^{J} \exp\big(v_{k,t+1}^{\tau(\bar{s}_{i,t+1})}(\bar{s}_{i,t+1}) - Z^{\tau(\bar{s}_{i,t+1})\prime}\gamma_p + \phi \mathbb{1}(g_i = G_k)\big)\bigg) \Big| s_{it}, d_{it} = j \Bigg].$$

The financial costs of moving are implicit, operating through the state argument to post-move type $\tau(\bar{s}_{i,t+1})$.

3.2 Conditional Choice Probabilities

Since the idiosyncratic tastes follow an extreme value distribution, the conditional choice probabilities are a logistic function of the value functions. This functional form means that the value functions identified by maximum likelihood estimation are unique only up to an additive constant (since adding a constant to the value of every alternative yields the same probabilities). Formally, we identify $\tilde{v}_{jt}^{\tau} = v_{jt}^{\tau} - m_t^{\tau}$.

It is useful to think of the household's optimization problem in stages. First, it chooses whether to move or stay. Conditional on moving, it chooses between the outside option or choosing among the inside option. Conditional on moving and not choosing the outside option, it chooses an inside option.

Considering the move/stay decision first, the necessary and sufficient condition for staying is

$$v_{J+1,t}^{\tau(s_{it})}(s_{it}) + \phi \mathbb{1}(g_i = G_{J+1}) + \epsilon_{i,J+1,t} > \max_{k \neq J+1} \{ v_{kt}^{\tau(\bar{s}_{it})}(\bar{s}_{it}) + \phi \mathbb{1}(g_i = G_k) + \epsilon_{ikt} \} - Z^{\tau(\bar{s}_{it})\prime}\gamma_p$$

$$\iff \tilde{v}_{J+1,t}^{\tau(s_{it})}(s_{it}) + \phi \mathbb{1}(g_i = G_{J+1}) + \epsilon_{i,J+1,t} > \max_{k \neq J+1} \{ \tilde{v}_{kt}^{\tau(\bar{s}_{it})}(\bar{s}_{it}) + \phi \mathbb{1}(g_i = G_k) + \epsilon_{ikt} \} - Z^{\tau(\bar{s}_{it})\prime}\gamma_p - (m_t^{\tau(s_{it})} - m_t^{\tau(\bar{s}_{it})}).$$

Hence, the disutility incurred from financial moving costs is given by this difference, $m_t^{\tau(s_{it})} - m_t^{\tau(\bar{s}_{it})}$. We take the monetary value of these costs to be $0.06 p_{h_{it}}$

[2]Although, (4) this last distinction does not actually matter since wealth is the only endogenous or time-varying characteristic and we exclude it from the set of characteristics that determine the psychological moving costs.

[3]Under this notation, we have $\bar{\tau} = \tau(\bar{s}_{it})$, and $\bar{Z}^{\tau} = Z^{\bar{\tau}} = Z^{\tau(\bar{s}_{it})}$.

due to realtor costs and parameterize its effect on utility as $0.06 p_{h_{it}} Z^{\tau(\bar{s}_{it})}\gamma_f$. Then, we may write the probability of a household of observable type τ, unobserved type g, residing in neighbourhood j, and at the beginning of time t deciding to stay in j as

$$\mathbb{P}^{\tau,\text{stay}}_{jtg} = \frac{\exp(\tilde{v}^{\tau}_{jt} + \phi\mathbb{1}(G_{J+1} = g))}{\exp(\tilde{v}^{\tau}_{jt} + \phi\mathbb{1}(G_{J+1} = g)) + \sum_{k=0}^{J} \exp\left(\tilde{v}^{r(\tau,p_{jt})}_{kt} + \phi\mathbb{1}(G_k = g) - 0.06 p_{jt} Z^{r(\tau,p_{jt})\prime}\gamma_f - Z^{r(\tau,p_{jt})\prime}\gamma_p\right)}.$$

Here, I have omitted all references to the household index i because it is not necessary: h_{it} is implicit since we are dealing with a stay probability, the unobserved type g_i is assumed, and conditional on the observable type τ and local price p_{jt} the post-move type is determined and represented with the mapping $r(\tau, p_{jt})$.

The probability of choosing the outside option conditional on moving is

$$\mathbb{P}^{\bar{\tau}}_{0tg} = \frac{\exp(\tilde{v}^{\bar{\tau}}_{0t})}{\exp(\tilde{v}^{\bar{\tau}}_{0t}) + \sum_{k=0}^{J} \exp\left(\tilde{v}^{\bar{\tau}}_{kt} + \phi\mathbb{1}(G_k = g)\right)}.$$

Here, I use type $\bar{\tau}$ to reflect that households choosing the outside option have made a decision to move and transitioned to a lower wealth level. This expression combines all households who have dropped to the wealth level in $\bar{\tau}$, regardless of the precise combination of house prices and pre-move wealth that got them there.

Finally, the probability of choosing an inside option ℓ conditional on moving and not to the outside option is

$$\mathbb{P}^{\bar{\tau}}_{\ell tg} = \frac{\exp(\tilde{v}^{\bar{\tau}}_{\ell tg} + \phi\mathbb{1}(G_\ell = g))}{\exp(\tilde{v}^{\bar{\tau}}_{\ell tg} + \phi\mathbb{1}(G_\ell = g)) + \sum_{k=0}^{J} \exp\left(\tilde{v}^{\bar{\tau}}_{kt} + \phi\mathbb{1}(G_k = g)\right)}.$$

3.3 Maximum Likelihood Estimation

Since I will rely on a draw from a conditional distribution for income data on homebuyers (tract-year-race), I will only have a guess of income for households that bought a home in NYC during 2003-19. So, in qualitative terms, I observe the same sample of households as BMMT: buyers from when they buy until they sell or the end of the sample, and sellers that buy again in the region (with the caveat that my race and income measures are much noisier).

Abstracting away from unobserved heterogeneity for a moment, any household i can be mapped onto the relevant set of CCPs (conditional on coverage described immediately above), by looking at their observable characteristics τ and their location j at time t, and the type they would transition to conditional on a move $\bar{\tau}$. The unconditional probability of choosing the outside option is the product $\mathbb{P}^{\text{stay}}\mathbb{P}_0$, and the unconditional probability of moving and choosing an inside option ℓ is the product $\mathbb{P}^{\text{stay}}\mathbb{P}_0\mathbb{P}_\ell$. In year t and assuming $g_i = g$, household i's contribution to the likelihood is then

$$
\begin{aligned}
L^i_{tg}(\tilde{v},\gamma_f,\gamma_p,\phi) = {} & \left(\mathbb{P}^{\tau(s_{it}),\text{stay}}_{h_{it}tg}(\tilde{v},\gamma_f,\gamma_p,\phi)\right)^{\mathbb{1}(d_{it}=J+1)}\left(1-\mathbb{P}^{\tau(s_{it}),\text{stay}}_{h_{it}tg}(\tilde{v},\gamma_f,\gamma_p,\phi)\right)^{\mathbb{1}(d_{it}\neq J+1)} \\
& \times\left(\mathbb{P}^{\tau(\bar{s}_{it})}_{0tg}(\tilde{v},\phi)\right)^{\mathbb{1}(d_{it}=0)}\left(1-\mathbb{P}^{\tau(\bar{s}_{it})}_{0tg}(\tilde{v},\phi)\right)^{\mathbb{1}(0<d_{it}<J+1)} \\
& \times\prod_{\ell=1}^{J}\left(\mathbb{P}^{\tau(\bar{s}_{it})}_{\ell tg}(\tilde{v},\phi)\right)^{\mathbb{1}(d_{it}=\ell)}.
\end{aligned}
$$

In the case with unobservable heterogeneity, we are not just estimating normalized value functions $\tilde{v}$ and moving cost parameters $\gamma_f,\ \gamma_p$, but also the utility from living in the correct broad area ϕ and the distribution of preferences over broad areas π. A household's contribution to the likelihood is then an expectation over potential unobserved types. Denote the first and last year that i is observed in the data with T^i_0 and T^i_1 respectively. The complete log-likelihood is then

$$
\mathcal{L}(\tilde{v},\gamma_f,\gamma_p,\phi,\pi)=\sum_i \ln\left(\sum_g \pi_g \prod_{t=T^i_0}^{T^i_1} L^i_{tg}(\tilde{v},\gamma_f,\gamma_p,\phi)\right).
$$

The formal definition of the estimator is

$$
(\hat{\tilde{v}},\hat{\gamma_f},\hat{\gamma_p},\hat{\phi},\hat{\pi}) = \arg\max_{\tilde{v},\gamma_f,\gamma_p,\phi,\pi}\mathcal{L}(\tilde{v},\gamma_f,\gamma_p,\phi,\pi) \text{ s.t. } \bar{\mathbb{P}}^{\tau}_{jt}=\sum_g \pi_g \mathbb{P}^{\tau}_{jtg}(\tilde{v},\phi)
$$

where the left hand side of the constraint is the empirical, frequency-based choice probability (which needs to be smoothed to deal with sparseness in the type-space). For a guess of $(\gamma_f,\gamma_p,\phi,\pi)$, the optimal $\tilde{v}$ can be found by using the constraint as a contraction in the style of Berry (1994):

$$
\tilde{v}^{\tau,b+1}_t=\tilde{v}^{\tau,b}_t+\ln(\bar{\mathbb{P}}^{\tau}_t)-\ln\Big(\sum_g \pi_g\mathbb{P}^{\tau}_{tg}(\tilde{v}^{\tau,b}_t,\phi)\Big).
$$

Since the total number of $\tilde{v}$ is equal to the product of the number of neighbourhoods, time periods, and types, this contraction greatly speeds estimation.

3.4 Forecasting and Flow Utility

I will describe the way that household expectations are modeled in the original BMMT. My contribution follows later.

From $\hat{\gamma}$, we have $m^{\tau}_t - m^{\bar{\tau}}_t$. When utility from zero wealth is normalized to zero (for a race-income group), we obtain m^{τ}_t and hence $v^{\tau}_t=\hat{\tilde{v}}+m^{\tau}_t$. Rather than predict time-varying amenities directly, we use present characteristics to predict future value functions:

$$
v^{\tau}_{jt}=\rho^{\tau}_{0j}+\sum_{\ell=1}^{L}\rho^{\tau}_{1\ell}v^{\tau}_{j,t-\ell}+\sum_{\ell=1}^{L}X'_{j,t-\ell}\rho^{\tau}_{2\ell}+\rho^{\tau}_{3j}t+\omega^{\tau}_{jt}.
$$

Notably, this model is run at the type level and includes lagged value functions, lagged amenities, a time trend, and neighbourhood fixed effects.

Assume for now that a price forecast is available (discussed in the next section).[4] This price forecast $\hat{p}_{j,t+1}$ also implies forecasted moving costs $0.06\hat{p}_{j,t+1}$ and transition probabilities for household wealth $\hat{\tau}_{j,t+1}|\tau_t$. This allows the simulation of the expectation below, with flow utility as the residual:

$$u_{jt}^{\tau} = v_{jt}^{\tau} - \beta\mathbb{E}\Bigg[\ln\Big(\exp\big(v_{J+1,t+1}^{\tau_{t+1}}\big) + \sum_{k=0}^{J}\exp\big(v_{k,t+1}^{\bar{\tau}_{t+1}} - Z^{\bar{\tau}_{t+1}\prime}\gamma_p\big)\Big)\big|s_{it}, d_{it} = j\Bigg].$$

Here, the complete set of right hand side variables in the preceding two regressions is represented by s, since the inclusion of lags and fixed effects proxies for the unobservable ξ. Furthermore, no individual information *per se* is needed - all the relevant information is contained in the type τ.

Finally, utility may be studied with regressions of the following form, where ξ is treated as a residual and c indexes counties (boroughs):

$$u_{jt}^{\tau} = \alpha_0^{\tau} + \alpha_c^{\tau} + \alpha_t^{\tau} + X_{jt}'\alpha_x^{\tau} + \xi_{jt}^{\tau}.$$

Although increases in prices into the future are good for owners, high prices in the present still are not. When dealing with flow utility in a dynamic model, we can represent this with user costs C_{jt}, and take them to be 5% of prices. Then, if the benefit of a dollar in the present is the same as the benefit of a dollar in the value function, we can use the marginal utility of wealth γ_f to partial out the user costs:

$$u_{jt}^{\tau} + \hat{\gamma}_f^{\tau}C_{jt} = \alpha_0^{\tau} + \alpha_c^{\tau} + \alpha_t^{\tau} + A_{jt}'\alpha_x^{\tau} + \xi_{jt}^{\tau},$$

where $A = \{X/p\}$ are all non-price characteristics (amenities). A correctly-estimated price coefficient eases identification of other parameters, but other endogenous variables on the right hand side still need specific identification strategies.

3.5 Dynamic Model Motivation

Although population density is highly auto-correlated, implying that a static model would do well at measuring households' MWTP for it, if we take seriously the idea that households gain utility from wealth then expected price changes would enter flow utility as an omitted variable under the static model. Consider the following definition and lemma, proven in the appendix.

[4] BMMT used a similar regression model to the one used to forecast value functions:

$$p_{jt} = \phi_{0j} + \sum_{\ell=1}^{L} X_{j,t-\ell}'\phi_{2\ell} + \phi_{3j}t + \nu_{jt}.$$

Assume throughout that there is no unobserved heterogeneity. To simplify notation, denote the continuation value for any given choice-specific value function with

$$c_{jt}^{\tau} = \ln\Big(\exp(v_{j,t+1}^{\tau_{t+1}}) + \sum_{k=0}^{J}\exp(v_{k,t+1}^{\bar{\tau}_{t+1}} - Z^{\bar{\tau}_{t+1}\prime}\gamma_p)\Big).$$

Lemma 1 *Under the assumptions of the BMMT model with no unobserved heterogeneity, the continuation value may be expressed as*

$$c_{jt}^{\tau} = m_{t+1}^{\tau_{t+1}} + \tilde{v}_{j,t+1}^{\tau_{t+1}} - \ln(\mathbb{P}_{j,t+1}^{stay,\tau_{t+1}}).$$

With this result, it can be shown that a correlation between amenities and price changes will bias results unless the correct model is used.

Proposition 1 *Let the static model be a model where the value function is a sum of unchanging discounted flow utility over an infinite horizon, because the neighbourhood is assumed to be unchanging and the household is assumed to never move. If BMMT is the true model but the static model is estimated instead, then flow utility will be incorrectly estimated as*

$$\bar{u}_{jt}^{\tau} = \beta^{\tau}X_{jt}^{\tau} + \xi_{jt}^{\tau} + \delta Z^{\tau\prime}\gamma_f^{\tau}\mathbb{E}[p_{j,t+1} - p_{jt}|s_t] + \delta\mathbb{E}[\tilde{v}_{j,t+1}^{\tau_{t+1}} - \tilde{v}_{jt}^{\tau} - \ln(\mathbb{P}_{j,t+1}^{stay,\tau_{t+1}})|s_t],$$

where the expected change in price $\delta Z^{\tau\prime}\gamma_f^{\tau}\mathbb{E}[p_{j,t+1} - p_{jt}|s_t]$ enters as an omitted variable.

The proof is in the appendix.

There are three conceptually distinct problems with this measure of utility. Even if amenities are close to constant, if they are correlated with changes in prices then estimates will be biased. Variation in price growth may occur even when amenities are constant if the population is growing and there is variation in housing supply elasticity.

Next, $\mathbb{E}[\tilde{v}_{jt}^{\tau_{t+1}} - \tilde{v}_{jt}^{\tau}|s_t]$ represents that as wealth changes preferences also change and that forward-looking households take this into account. Finally, $\mathbb{E}[\ln \mathbb{P}_{jt}^{\text{stay},\tau_{t+1}}|s_t]$ represents a bias related to option value and the "churn" of moving and location choices. If, despite wealth effects and changing preferences, the household believes it will always stay in the same location (perhaps due to prohibitive moving costs), this bias is zero, whereas if households tend to move a lot then this bias will be non-zero.

3.6 Forecasting Prices

One area in which the baseline model could be improved is the formation of price forecasts; I propose to model the distribution of future prices as the result of future demand shifts along a local housing supply function, using estimates from Baum-Snow and Han (2019). Housing demand can be expressed using an accounting identity as the combination of the present population distribution and future CCPs. The present population distribution is known, and the distribution of future CCPs depends on the distribution of future value functions, which is also known. The difficulty lies in that (1) the distribution of types within a neighbourhood at the beginning of the next period depends on future prices, and (2) future CCPs depend on non-linearly on future prices and (3) errors in value functions and prices.

The evolution of the type-neighbourhood distribution over time is complicated but not complex. For now, assume away all components of type that are not wealth, and so take τ to be an integer representing wealth. Write the type that a household of type τ transitions to after selling its house for price p as $\bar{\tau} = r(\tau, p)$. Assume that this function is invertible with respect to τ so that we can write the type that transitions to τ as $\hat{\tau} = r^{-1}(\tau, p)$.[5] Finally for notational convenience let $\tilde{p}_{jt} = p_{jt} - p_{j,t-1}$. Then, the following accounting identity must hold:

$$N_{jt}^{\bar{\tau}} = N_{j,t-1}^{\bar{\tau}-\tilde{p}_{jt}} \mathbb{P}_{jt}^{\text{stay},\bar{\tau}} + \sum_{k \neq j} N_{k,t-1}^{\tau-\tilde{p}_{kt}} (1 - \mathbb{P}_{kt}^{\text{stay},\tau})(1 - \mathbb{P}_{0t}^{\tau})\mathbb{P}_{jt}^{\tau}.$$

The present population consists of stayers and movers. The movers must choose to move, not choose the outside option, and choose j. The movers must also be the type that would transition to $\bar{\tau}$ after paying moving costs. We also need to incorporate that if we are considering a wealth level of *e.g.* \$100,000, and the price has increased by *e.g.* \$10,000, then we are concerned with the number of people who had $\tau - \tilde{p}_{kt} = \$90{,}000$ in wealth the year prior.

I model future prices as a shift along a housing supply curve, where the shift is given by the change in aggregate housing demand and the housing supply elasticity is taken from Baum-Snow and Han (2019), plus an idiosyncratic error term:[6]

$$p_{j,t+1} = \left(\frac{N_{j,t+1} - N_{jt}}{N_{jt}} \right) \frac{p_{jt}}{\eta_j} + p_{jt} + \epsilon_{j,t+1}^{p}.$$

Provided that a CCP exists for every type of household that exists (including renters, and making strong assumptions about the rental market), we can use the accounting identity to write the future aggregate location choices as an aggregation of past location choices and future CCPs, which both depend on future prices – location choices through the indexes, and CCPs through moving costs. The distribution of future prices that households use to form expectations is centred on the vector that minimizes prediction errors, has the same spread as those prediction errors, and satisfies the accounting identity below:

$$\hat{N}_{j,t+1}^{\bar{\tau}} = \mathbb{E}\Big[N_{jt}^{\bar{\tau}-\tilde{p}_{j,t+1}} \mathbb{P}_{j,t+1}^{\text{stay},\bar{\tau}}(p_{j,t+1}) + \sum_{k \neq j} N_{kt}^{\tau-\tilde{p}_{k,t+1}} (1 - \mathbb{P}_{k,t+1}^{\text{stay},\tau}(p_{k,t+1}))(1 - \mathbb{P}_{0,t+1}^{\tau})\mathbb{P}_{j,t+1}^{\tau} \Big| s_t \Big].$$

[5] In any quantitative implementation, these will be correspondences and not functions and hence not invertible, but abstract away from this for now.

[6] In the future it would be desirable to incorporate a city-wide shock to housing prices; this shock may be autocorrelated to reflect that house prices took several years to bottom out in the housing-financial crisis, and grew steadily before and after.

Aggregate housing demand is then given by $\hat{N}_{j,t+1} = \sum_\tau \hat{N}^\tau_{j,t+1}$.

Under specific conditions, the definition of elasticity is a contraction mapping, such that iterating it will lead to fixed points for expected prices and location decisions.

Proposition 2 *Let*

$$\tilde{v}^\tau_{jt} = \tilde{v}^{\bar{\tau}}_{jt} \ \forall \ \tau, \bar{\tau},$$

(preferences for locations do not vary over wealth), let

$$\mathbb{E}\Big[(1 - \mathbb{P}^{stay,\tau}_{k,t+1}(p_{k,t+1}))(1 - \mathbb{P}^\tau_{0,t+1})\mathbb{P}^\tau_{j,t+1}\Big| s_t, \hat{p}_{j,t+1}\Big] \\ = \mathbb{E}\Big[(1 - \mathbb{P}^{stay,\tau}_{k,t+1}(p_{k,t+1}))(1 - \mathbb{P}^\tau_{0,t+1})\mathbb{P}^\tau_{j,t+1}\Big| s_t, \Big],$$

(price forecasts do not affect expected movers), let $\mathbb{P}^{stay,\tau}_{jt} > 1/2$ *for all* j, t, τ, *let*

$$\hat{N}_{j,t+1}(x) = \sum_{\bar{\tau}} \Bigg(\mathbb{E}\Big[N^{\bar{\tau}}_{jt} \mathbb{P}^{stay,\bar{\tau}+\tilde{p}_{j,t+1}}_{j,t+1}(p_{j,t+1}) \\ + \sum_{k \neq j} N^\tau_{kt}(1 - \mathbb{P}^{stay,\tau+\tilde{p}_{k,t+1}}_{k,t+1}(p_{k,t+1}))(1 - \mathbb{P}^\tau_{0,t+1})\mathbb{P}^\tau_{j,t+1}\Big| s_t, \hat{p}_{j,t+1} = x\Big]\Bigg),$$

and let

$$T(x) = \left(\frac{\hat{N}_{j,t+1}(x) - N_{jt}}{N_{jt}} \right) \frac{p_{jt}}{\eta_j} + p_{jt}.$$

Then, there exists an $\underline{x} \in \mathbb{R}$ *such that* T *is a contraction mapping with the Euclidean norm on the metric space* $\{x \in \mathbb{R} : x > \underline{x}\}$.

The proof can be found in the appendix. Two of these assumptions are meaningful / restrictive. First, the assumption that preferences for locations do not vary over wealth was made to avoid a mathematical dead end, but it is possible that the data may be such that the function is not a contraction if this tenet is not enforced. However, I do believe that endogenous preferences are a bit odd, and so they might not be missed. Second, this proposition is only useful if prices are in the space where it is a contraction; if they are not sufficiently large then it won't work. But the assumption is perhaps better understood as a condition on stay probabilities, which increase with prices due to moving costs. We need the stay probabilities to be large so that their derivatives are small. Considering that they stay probabilities measured in the ACS are about 95%, and the stay probabilities in the estimation panel a bit higher yet, I am optimistic that the assumption will hold. As I explain in the proof, the other assumptions are probably innocuous.

One would use this contraction by: making a guess of $\hat{p}_{t+1}$, obtaining the residuals $\hat{\epsilon}^p_{t+1}$, drawing from these residuals and the value function residuals $\hat{\epsilon}^v_{t+1}$ to simulate $\hat{N}^{\bar{\tau}}_{t+1}$, aggregating to $\hat{N}_{j,t+1}$, and evaluating and updating $\hat{p}_{t+1}$.

While this procedure would appear to be very computationally costly, I have found that the stay probability can be approximated with a Taylor expansion that is somewhat separable from price:

$$\mathbb{P}^{\text{stay},\tau}_{j,t+1}(p_{j,t+1}) \approx \sum_{n=0}^{N-1} (-1)^n \exp(-Z^{\tau\prime}\gamma_p - Z^{\tau\prime}\gamma_f 0.06 p_{j,t+1})^n \left[\frac{1}{D} \sum_{d=1}^{D} \left(\frac{\sum_k \exp(\hat{v}^{\bar{\tau}}_{kt} + \epsilon^{\bar{\tau}v}_{dkt})}{\exp(\hat{v}^{\tau}_{jt} + \epsilon^{\tau v}_{djt})} \right)^n \right]$$

Here, d indexes draws of value function residuals and n indexes terms in the Taylor expansion. The implication is that I can precalculate the simulation results that depend on value function residual draws, paying some large fixed cost in order to be assured of faster fixed point iterations. These simulation results would be indexed by j, t, τ, n and multiplied with a price guess and residual term also indexed by j, t, τ, n before being aggregated up to j, t, τ.

A numerical test shows that this Taylor approximation performs quite well. I calculated the Taylor approximation for a 10% sample of my data, using same-period prices (guaranteed to be the right magnitude), until the each stay probability stopped changing. I compare the Taylor approximation (X-axis) against properly simulated stay probabilities (Y-axis) in Figure 2. The fit is extremely strong, with some poorly-fitted outliers on the low end - but these observations account for a tiny minority, with the vast majority lying almost exactly on top of the 45 degree line close to one. Qualitatively, this result is unsurprising, since the Taylor expansion is around the point $\exp(-Z^{\tau\prime}\gamma_p - Z^{\tau\prime}\gamma_f 0.06 p_{j,t+1}) = 0$, which corresponds to the asymptotic case where moving costs are infinite and the stay probability is one.

Since the BH elasticities are 10-year elasticities, but I wish to use them for 1-year differences, I will use the housing supply elasticity corresponding to new units and floorspace in existing buildings. This margin is more likely to respond within a year, compared to new structures.

3.7 Estimating Renter CCPs

Although renters are not the main focus of the paper, modeling their behaviour is important for several reasons. First, the price forecast method depends on forecasting total housing demand, of owners and renters. Second, the utility of owners depend on the number and characteristics of all neighbours, so if renters' behaviour changes in response to some counterfactual then owner behaviour will also be impacted in a way that will be missed if renters are not explicitly modelled. Finally, it will be interesting and informative to analyze renter utility and contrast the results with the main owner results.

Although I do not observe renter mobility, renter value functions and CCPs can be recovered when reasonable assumptions are made about their moving costs. Discretize renters by $\tau =$ (race, income), ignoring wealth. The number of renters of a given type τ in a given location j at a given time t can be decomposed into the lagged vector of that types location distribution and current CCPs:

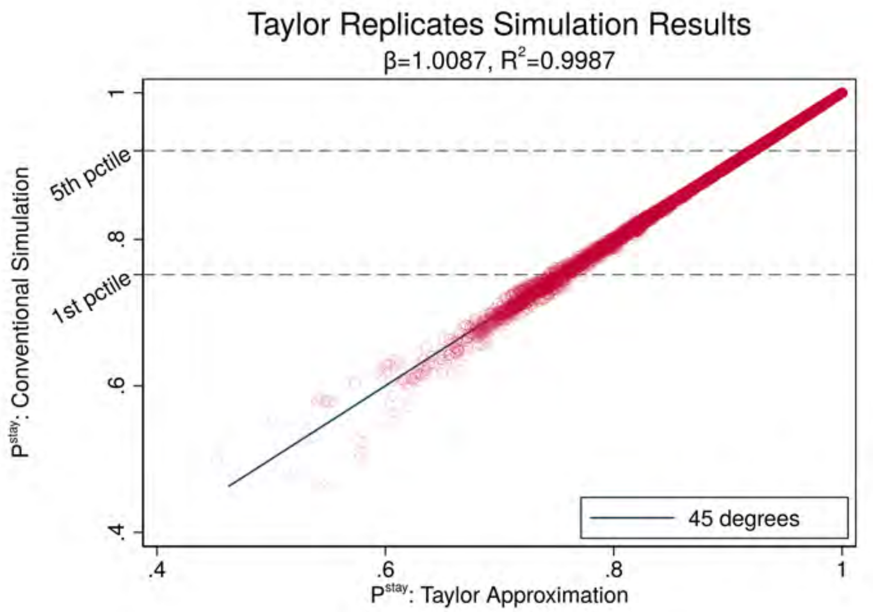

Figure 2: Taylor Approximation Matches Conventional Simulation

$$N_{jt}^{\tau} = N_{j,t-1}^{\tau}\mathbb{P}_{jt}^{\text{stay},\tau} + \sum_{k \neq j} \Big(N_{k,t-1}^{\tau}(1 - \mathbb{P}_{kt}^{\text{stay},\tau})(1 - \mathbb{P}_{0t}^{\tau})\mathbb{P}_{kt}^{\tau} \Big).$$

Assume that financial moving costs are zero (since there are no realtor fees paid) and that the psychological moving costs are the same as owners of the same type. Then, given a type and time period, there are J locations and J unknown value functions, and so the value functions can be recovered. Furthermore, the limited information on mobility in ACS microdata can be used as an over-identification check, or as a target to calibrate parameters governing a conversion of owner psychological moving costs to renter psychological moving costs.

3.8 Solving Counterfactual Equilibria

Once parameters have been estimated, it will be useful to study how the equilibrium would change if the facts were different. To that end, I define an equilibrium and propose how to solve for endogenous variables, given parameters and exogenous variables.

Let utility be an additive function of an exogenous component (a stand-in for exogenous observable time-varying amenities) and endogenous population density and user cost of housing:

$$u_{jt}^{\tau} = u_{jt0}^{\tau} + \beta^{\tau} \ln\left(\frac{N_{jt}}{a_j}\right) - \gamma_f^{\tau} 0.05 p_{jt}$$

Then, an equilibrium is a sequence of location decisions N_{jt}^{τ} and prices p_{jt} such that

- agents choose the location that maximizes their (individual) value function
  - the (choice-specific) value function is the sum of present utility and the expected discounted continuation value
  - the CCPs are a logistic function of value functions
  - location decisions are an "aggregative function" of past location decisions and CCPs
- prices are given by shifts along a housing supply function
  - the shifts are given by changes in location decisions
- future value functions are given by the expected future value functions plus idiosyncratic errors, and future prices are given by the expected future prices plus idiosyncratic errors
  - expected future CCPs are given by integrating the logistic function of value functions over idiosyncratic value function and price errors
  - expected future location decisions are given by the expectation of the "aggregative function" of present location decisions and future CCPs
  - future prices are given by shifts along a housing supply function, where the shifts are given by expected changes in location decisions, plus an idiosyncratic error

I propose to solve for N_{jt}^{τ} and p_{jt} by numerically solving the following equation, which must hold for all j, t, τ:

$$v_{jt}^{\tau} = u_{jt0}^{\tau} + \beta^{\tau} \ln\left(\frac{N_{jt}(v_{jt}, p_{jt})}{a_j}\right) - \gamma_f^{\tau} 0.05 p_{jt}(N_{jt}(v_{jt}, p_{jt}))$$
$$+ \delta \mathbb{E}\left[\ln\left(\exp(v_{j,t+1}^{\tau_{t+1}}) + \sum_{k=0}^{J} \exp(v_{k,t+1}^{\bar{\tau}_{t+1}} - Z^{\bar{\tau}_{t+1}\prime} \gamma_p)\right) \Big| s_t, v_t \in s_t \right].$$

Start with a guess of v_{jt}^{τ}, and use observed exogenous amenities u_{jt0}^{τ}. Solve for the N_{jt}^{τ} and p_{jt} that jointly satisfy the aggregative function and the definition of elasticity (using $N_{j,t-1}^{\tau}, p_{j,t-1}$). This is the same procedure as the one described for future prices and location decisions, but without uncertainty over value functions since in the present they are observed to be what I have guessed (I am undecided about how uncertainty in prices should be handled in the present). I will repeat the procedure as described in the previous section to solve for forecasted prices and location decisions. Now, we have all information necessary to construct the expectation of the continuation value as in estimation. Then, we have an output v_{jt}^{τ}; compare it with the input and choose a new output. When this equation holds, all equilibrium conditions are met. In this way we can solve for the equilibrium moving forward in time. Rather than using observed u_{jt0}^{τ}, one could estimate an AR regression and take draws in order to extrapolate the equilibrium path into the future.

This function has not (yet) been shown to be a contraction mapping, and its calculation is computationally demanding: two fixed point algorithms and a large number of residual draws for each iteration. However, the first fixed point algorithm for present prices will require many fewer residual draws (possibly none), and the second for future prices can be sped up by using the Taylor approximation discussed in the previous section. Finally, by writing the continuation value in its alternate form, we can see that it actually will not require additional simulation:

$$\mathbb{E}[c_{jt}^{\tau}|s_t] = \mathbb{E}[m_{t+1}^{\tau_{t+1}} + \tilde{v}_{j,t+1}^{\tau_{t+1}} - \ln(\mathbb{P}_{j,t+1}^{\text{stay},\tau_{t+1}})|s_t].$$

The first term can be calculated directly from the expected future price and current-period data:

$$\mathbb{E}[m_{t+1}^{\tau_{t+1}}|s_t] = \gamma_f^{\tau}\left(\mathbb{E}[p_{j,t+1}|s_t] - p_{jt}\right) + m_t^{\tau}.$$

The second term can be calculated using the fitted values from the value function regressions along with the price forecast and the marginal distribution of price residuals. That is, once the price forecast is calculated, the price residual distribution (which will be based on relatively small data and cheap to compute) will tell us the distribution over types, which will be combined with the fitted $\tilde{v}_{j,t+1}^{\tau}$. Finally, the stay probability will have been directly simulated already; I will need only to save those results and take the expectation of the log.

4 Maximum Likelihood Estimation

4.1 Preliminaries

At this stage, I present results for a straight BMMT replication using NYC data, having not yet had the opportunity to enact my proposed extensions related to renters, price forecasts, or counterfactual simulations. Furthermore, I have not yet had the opportunity to estimate the model with unobserved heterogeneity. Rather, I use the analytical solution for v_{jt}^{τ} followed by estimation of $\hat{\gamma}_p, \hat{\gamma}_f$ using move/stay binomial logit.

I construct wealth in the household panel as housing equity less debts and moving costs. Housing equity is calculated by imputing price for non-sale years using a small-area repeat sales index and the original purchase price. Housing debt is calculated assuming that every household takes on 30-year fixed rate mortgage with the average national interest rate in the purchase year, a 20%

	γ_p	γ_f
CONSTANT	8.2971	0.023937
constant se	0.0066873	0.00019027
constant x black	0.30515	0.0091142
constant x black se	0.021012	0.00072267
constant x asian	0.18648	-0.0002506
constant x asian se	0.013722	0.00037443
constant x hispanic	0.3829	0.0015649
constant x hispanic se	0.015025	0.0004673
INCOME	6.5218e-05	-2.1166e-05
se	4.3777e-05	1.0445e-06
income x black	0.0011289	3.4898e-06
income x black se	0.0001383	3.9046e-06
income x asian	0.0012767	-6.4865e-06
income x asian se	9.6111e-05	2.3551e-06
income x hispanic	-0.00024261	2.0823e-06
income x hispanic se	0.00010266	2.734e-06

Table 2: MLE results

down payment, and constant yearly payments thereafter. Moving costs are 6% of a sale price.

For estimation, I discretize real wealth and income (2012 USD) into twelve quantiles calculated using the entire sample. I use four race groups: white, black, asian, and hispanic. Due to sparsity in the type space, I smooth the number of choosers of a neighbourhood and potential choosers (numerator and denominator of empirical choice probabilities) over income and wealth and within race-year groups. I use a Gaussian kernel and a bandwidth of two income or wealth groups – the smallest integer bandwidth that does not result in smoothed choice probabilities of zero.

4.2 Results

I let moving cost parameters vary by race and income and do not include a time trend. Parameter estimates and provisional standard errors are depicted in a table.[7] To ease interpretation, I also plot the parameter estimates for $\hat{\gamma}_p$ (the psychological cost moving), $\hat{\gamma}_f$ (the marginal utility of wealth), and $\hat{\gamma}_p/\hat{\gamma}_f$ (the monetary value of the psychological moving costs, or the WTP to avoid a move even if financial moving costs were zero).

The first result of note is that the constant and income interactions for whites are similar to those estimated by BMMT: a constant PMC of $8.3 \approx 9.5$,[8], a

[7] I say that the standard errors are provisional because they are calculated using the Hessian matrix output by Matlab's fminunc function, and the Matlab documentation says that this is not good practice – but it is better than nothing for now.

[8] With the caveat that $\exp(-8.3)/\exp(-9.5) \approx 3.3$ – not so close.

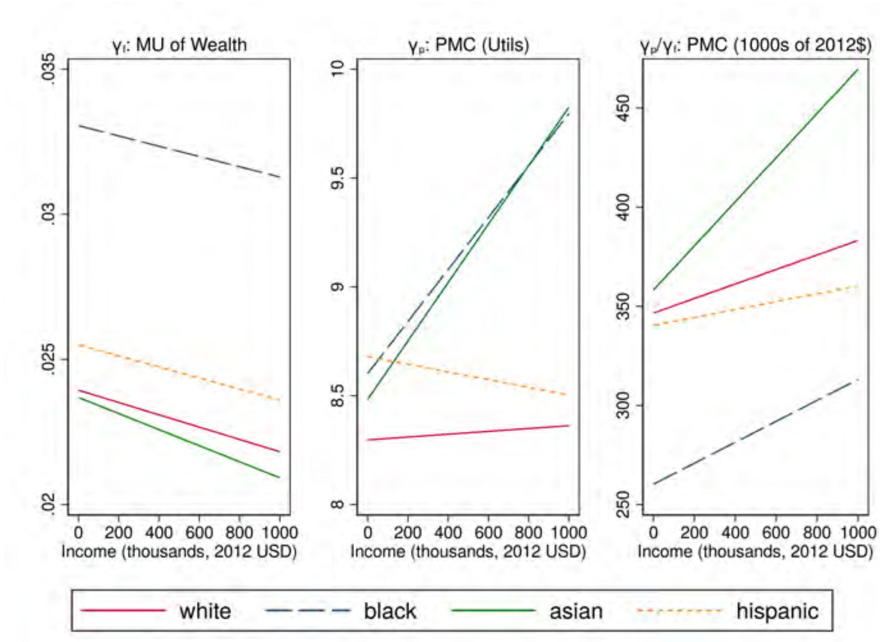

Figure 3: MLE results

constant FMC of 0.024, which when converted to 2000 USD is $0.032 \approx 0.035$.

I also find that the marginal utility of wealth is declining with income for all racial groups, although at a slower rate for whites than BMMT, but nonetheless a good and sensible result. My result that PMC increases with income for most groups is in contrast to BMMT and not so sensible. Taken together, the monetary value of psychological moving costs is large and increasing with income. Although that this increases with income is worrisome, the magnitude is not. As explained by Kennan and Walker (2011) and BMMT, large psychological moving costs reflect the high cost of moving to a random place at a random time. The moves that are observed in the data are those with very large unobserved idiosyncratic shocks, such as a new job.

4.3 Forecasts and Flow Utility

I run value function and price forecasts almost exactly as specified by BMMT, although I exclude my regressor of interest – log population density – from these regressions, since it is extremely endogenous. I calculate flow utility by using 1000 draws to estimate the expected continuation value.

5 Empirical Strategy

With flow utility in hand, I turn towards second stage estimation: instrumental variable regression of utility on population density. In particular, I am interested in the general effect of population density, to test the idea that NIMBYs genuinely dislike density *per se*, while saying little about what the exact mechanisms are. That means that we want a source of variation in density that simply adds people to an area, rather than radically altering the built form, demographics, or economic characteristics alone.

Since this is demand estimation in a broad sense, a natural place to look for exogenous variation in quantity (population) is the supply side. Although supply side instruments are probably correlated with housing supply elasticity η_j, and hence correlated with expectations of future density and price appreciation, the construction of flow utility using value function and price forecasts accounts for precisely this mechanism. And, a natural place to look for a supply shifter in a dense, supply-constrained urban environment is the zoning code.

The strategy that I will describe leverages the microgeographic interactions of three separate aspects of New York City's Zoning Resolution to generate identifying variation in log population density. These three characteristics are the hierarchical density level, the binary "contextual" status, and proximity to a wide street. When these three characteristics align, the zoning code allows greater density. The instrument itself is the share of land that meets the three requirements, while controlling for the share of land that meets each characteristic individually and the multiplicative interaction of these shares.

This instrument has some desirable qualities when it comes to estimating the general effect of density. It works through very marginal changes in the number of people in an area, by marginally making buildings larger to an extent that is most likely imperceptible to a layperson. The built form of the land that it operates on comprises structures of about 4-5 storeys and up, meaning that it does not directly densify single-family-zoned areas – which would be highly perceptible and possibly correlated with other characteristics – but instead adds people to a wide range of neighbourhoods across the city in buildings from low-rises and up. Since these kinds of structures are never too distant from single-family neighbourhoods, the instrument does add people to these neighbourhoods too. Since NYC is generally diverse and many people live in larger buildings, increasing the number of people living in larger buildings will probably not disproportionately change demographics or income levels.

5.1 History of Zoning in New York City

The year 1961 was a watershed in zoning and land use in New York City, seeing the adoption of the Zoning Resolution,[9] the ur-text that is still operative today (expanded to over 3,000 pages). These rules established residence districts numbered one to ten (referred to as *e.g.* R6) with specific regulations pertaining

[9]Primary source: City Planning Commission (2022), City Planning Commission (1961), secondary source: Department of City Planning (2018).

to building bulk and height and urban density. Similar numbered districts were established for commercial and manufacturing usages (*e.g.* C4, M2). These rules were generally focused on controlling density through the floor area ratio (FAR).

To the authors of the 1961 resolution, the predominant public amenities that needed to be protected from encroaching density were sunlight and open space, and so higher maximum FARs were granted to development projects on the larger lots through the use of a zoning concept called the "sky exposure plane". Such a plane starts at a setback from the street and continues at an upward angle until it reaches the rear of the lot; buildings may generally not cross this plane. A tall building could be considered a "tower" and be exempt from the sky exposure plane and subject to tower regulations instead if is located in an R9 or R10 district and satisfies a maximum lot coverage requirement. The urban form that results from the 1961 resolution is referred to as "tower-in-the-park" or Height Factor Zoning (HFZ).

However, by 1987, this model had come under criticism from residents and urbanists who resented these projects' departures from the pre-1961 built form of the surrounding properties. In a major update to the zoning resolution, planners introduced the Quality Housing Program[10] (QHP) as an optional alternative to HFZ and a total of eleven "contextual districts" where participation in the QHP would be mandatory. These new rules allowed projects that were shorter but bulkier, and better aligned with the "character" of the surrounding neighbourhood. Contextual districts augmented the basic set of zoning districts, denoted with an extra character at the end of the code (*e.g.*, R6A). The contrasts between QHP, sky exposure plane, and tower regulation is shown in Figure 4, borrowed from NYC's Zoning Handbook Department of City Planning (2018) (New York City Department of City Planning © 2018).

In both contextual and non-contextual districts, zoning is slightly relaxed when the lot abuts or is near a wide street, and these clauses kick in at different categorical density levels. A "wide street" is defined as a street that is wider than 75ft, and "near a wide street" generally means within 100ft of one. The width of the street matters for R9 and R10 contextual districts, and for R6-R10 in non-contextual districts (*e.g.*, R8X) even if the developer chooses to be governed by QHP regulations as in contextual districts. A guide for when street width matters and when it does not is given in Table 3; the instrument is the share of land that is in a category where street width matters and is near a wide street.

I have placed some modern and historical tables from city government in an appendix that show how street width affects different zoning regulation parameters. In contextual districts where developers must follow QHP regulations, street width operates through base height (height at which building width must narrow), building height, and number of stories (mostly redundant since height is also limited). In non-contextual districts where a developer chooses to follow QHP regulations, street width operates through the maximum lot coverage ra-

[10] Primary source: City Planning Commission (1987), secondary source: Kober (2018).

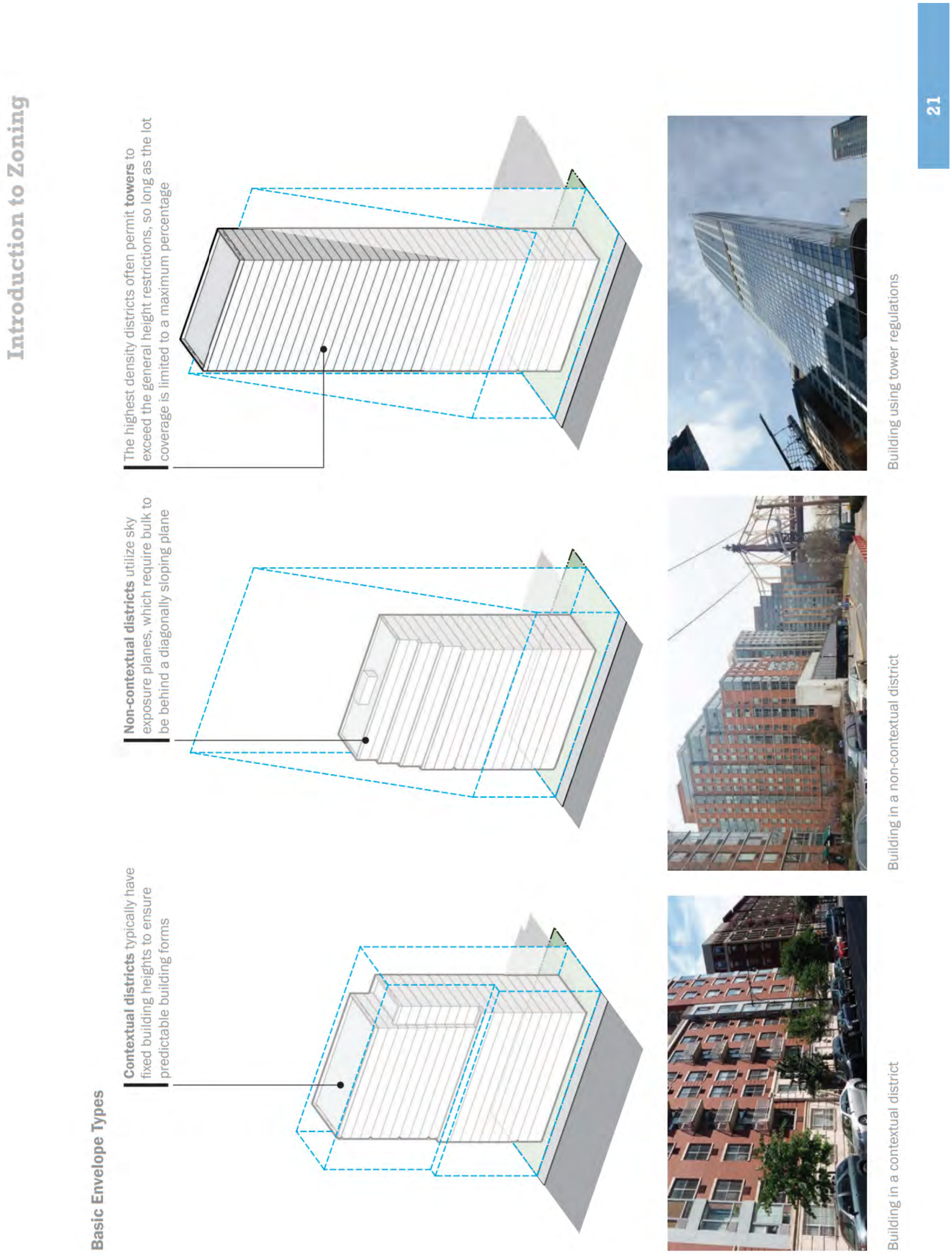

Figure 4: Zoning regulation types visualized. New York City Department of City Planning © 2018.

X => street width matters
N/A => regulations not applicable

	R1-5	R6-8	R9-10
Contextual, QHP			X
Non-contextual, QHP		X	X
Non-contexual, HFZ, SEP		X	X
Non-contextual, HFZ, tower	N/A	N/A	X

Table 3: Guide for when street width matters and when it does not.

tio, FAR, base height, building height, and number of stories. In non-contextual districts where a developer chooses to follow HFZ regulations, and specifically sky exposure plane regulations, street width operates through the setback at which the sky exposure plane begins and the slope of the sky exposure plane.

In non-contextual districts where a building chooses to follow HFZ regulations, and specifically tower regulations, street width operates by compelling developers to follow tower-on-a-base regulations over regular tower regulations. Tower-on-a-base regulations *require* that there be a base of 60-85ft height that is *not* set back from the street (sidewalks count as part of the street), and that the tower component above the base cover at *least* 30% of the lot. Compared to this, regular tower regulations do not include a wider base, must be set back from the street, and do not have a minimum lot coverage. The tower in Figure 4 follows regular tower regulations, not tower-on-a-base regulations.

5.2 Discussion of Identifying Variation

Supply constraints tend to be erected in response to demand, or take the form of natural barriers that may enter utility, and so they cannot generally be taken to induce exogenous variation (Davidoff et al., 2016). In the worst case, we can take this to mean that zoning levels are set to accommodate or foil demand, the local street grid provides better transportation or excessive noise, and contextuality is used as down-zoning by another name or provides pleasing architecture: any of these effects would mean that identification has failed. These stories are a mix of reverse causality (X causing Z) and exclusion restriction (Z causing Y, not through X) concerns, which I address by using control variables based on these three characteristics. As long as certain flexible assumptions (which I discuss immediately below) about how these characteristics enter utility are satisfied, the share of land satisfying all three conditions satisfies the exclusion restriction. I also make an argument for why a reverse-causal explanation is unlikely.

If preferences for these individual attributes (and/or their unobserved correlates) are additively separable at the neighbourhood level, then we can achieve causal identification by controlling the individual attributes and taking the area satisfying all three conditions to be exogenous. If preferences for these attributes (and/or their unobserved correlates) are not additively separable but do depend only on the *expected* share of land satisfying any combination of categories at the tract level, then we can achieve causal identification by controlling the individual attributes and their multiplicative interactions. It is only if preferences for these attributes (and/or their unobserved correlates) depend on the actual share of land satisfying multiple conditions that identification fails. Hence, the identifying assumption for this IV strategy is that preferences depend (at most) on shares of individual land use categories or expected shares of combination land use categories.

Since I condition on neighbourhood aggregates, the reverse causality story demands that population pressure against legal limits be resolved by an allocation by planners of the three characteristics together in space, while holding the levels of each constant. However, the evidence for this possibility is not strong. First, the width of streets is outside of city government's control.[11] Second, historical evidence[12] indicates that different types of contextual districts were designed for neighbourhoods with different street widths, suggesting that they were not intended to be systematically used with either wide or narrow streets. All told, it seems like reverse causality would be a strained interpretation of a positive first stage relationship, when we also know that exogenous variation in the instrument would cause there to be more legal space for people to live in.

Other important controls will include transit access, streetscape beauty (proxied by trees), and local industrial composition (probably the first principal component(s) of industry shares of establishments). Of these, only transit access is currently present in the models that I will present. Other dynamic amenities that I anticipate including and taking to be exogenous include crime, air quality, and possibly racial/ethnic composition (I would like to eventually instrument racial/ethnic composition).

5.3 First Stage Results

To reiterate, the instrument is the share of land in zoning categories where street width matters *and* near a wide street. We can refer to the confluence of these conditions in the same literal physical space as the *microgeographic interaction* of these conditions. I use the share of land meeting these conditions individually and their *multiplicative interactions* (in the sense of an interaction term in a regression) as controls, which addresses specific kinds of exclusion restriction violations discussed in the subsection prior. These multiplicative interactions can be interpreted as the expectation of the microgeographic interactions. To see why, please refer to Figure 5.

If 50% of the land in a neighbourhood meets Condition A, and 50% of the land meets Condition B, then if these conditions are independent we would

[11] The definition of a wide street has been 75ft since at least the original 1961 Zoning Resolution (City Planning Commission, 1961). Sidewalks are typically included in street width, and the streets themselves are owned by the New York Department of Transportation (Department of City Planning, 2018).

[12] "The districts with a B suffix are intended primarily for narrow streets, with height and setback regulations reflecting the limited light and air available... The A and X districts were contemplated primarily for wide streets with better access to light and air," (p. 9, City Planning Commission (1987)).

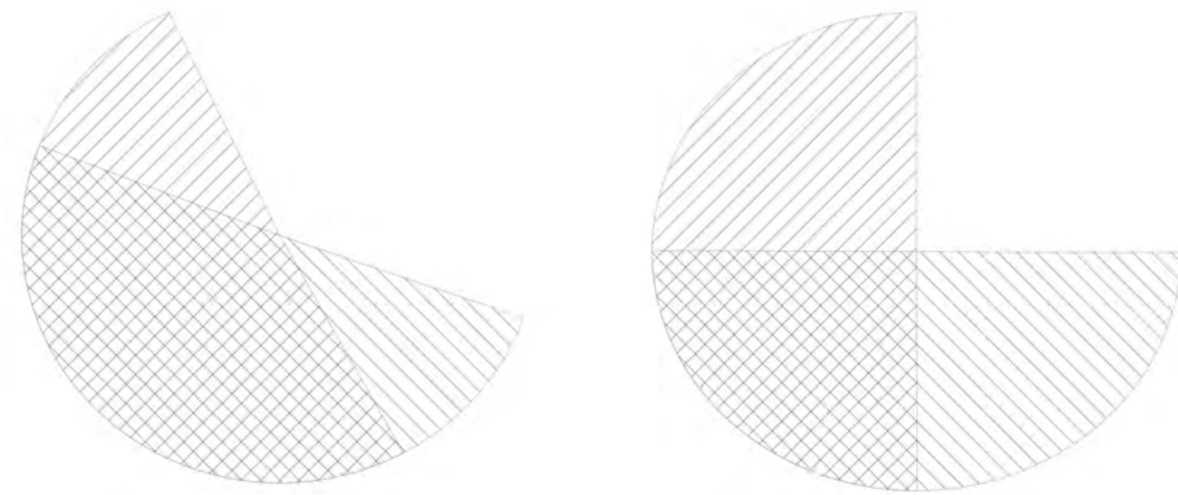

Figure 5: Microgeographic and Multiplicative Interactions

expect 25% – the product, the multiplicative interaction – of the land to meet both, depicted at right in Figure 5. But draws of random variables are not always equal to their expectation, and so it may be the case that the share of land meeting both conditions is actually 40%, depicted at left in Figure 5. The 50% of the land meeting condition A, the 50% of the land meeting condition B, and the 25% of the land expected to meet both enter as controls, and the power of the instrument in predicting density then comes from idiosyncratic, sub-neighbourhood-level variation in the co-location of land meeting these specific conditions.

Specifically, the instrument is the share of residential land that is zoned R9-10 and contextual and near a wide street, or zoned R9-10 and non-contextual and near a wide street, or zoned R6-8 and non-contextual and near a wide street (also depicted in Table 3). So we can divide all land into three density categories (1-5, 6-8, 9-10), two contextuality categories (yes/no), and two street width categories (yes/no). I let R1-5, non-contextual, and not near a wide street be the base categories, so there are four individual share variables as controls. I then take the set of fully-saturated interactions (except for share R6-8 interacted with share 9-10, obviously). This generates four single-variable terms, five two-variable interaction terms, and two three-variable interaction terms. To this I add borough fixed effects, year fixed effects, and fixed effects for quintiles of distance to the nearest subway station (to approximate an arbitrary functional form).

This first stage specification tends to be very strong when run at a small-geography level but lose power after aggregating to feasible neighbourhood sizes. So, I use the Max-P algorithm Duque et al. (2012) to aggregate Census Block

Groups up to units of at least 10,000 housing units that maximize the internal homogeneity of zoning and street variables. Put differently, I aggregate so that variation in zoning and street width within units is minimized, and variation between units is maximized. This generates 120 neighbourhoods and first stage results that meets conventional benchmarks.[13]

Finally, although the instrument is nominally constructed at the neighbourhood-year level because I use yearly data on zoning, relevant changes in zoning are so infrequent that it is more accurate to say that the variation is only cross-sectional. Therefore, I cluster standard errors at the neighbourhood level.

First stage results for specifications where the instrument is disaggregated (into different zoning categories where street width matters) are provided in Table 4. Each instrument is positive and significant at at least the 10% threshold, and usually significant at more stringent thresholds, for several definitions of population density. The first definition is the most literal definition: people per area. The second definition excludes parks, while the third excludes parks and streets themselves.

Since each instrument is significant alone, it makes sense to combine them in order to increase the power of the first stage. These results are shown in Table 5. Here, I have also included the Kleibergen-Paap weak identification statistic, which tends to be about 10-13, meeting the conventional benchmark.

I do not know why the coefficients are so large in magnitude. At the tract level, they were smaller but about as significant as displayed here, and I am convinced that it is a valid instrument at the tract level. But in honesty, I cannot rule out the case that aggregation with the Max-P algorithm introduces econometric problems, causing overly-large point estimates.

6 Results

I present the results of IV regressions run at the neighbourhood-year-type level. I run separate models for each race. Within a race, I let utility from density vary by income, but constrain the coefficients on controls to be equal. Within a race-income type, there are multiple wealth types, so I use the distribution of households over types as weights to ensure that more frequent race-income-wealth types have more influence on the average parameter estimates. Since population density varies over neighbourhood-year, but the models are run at the neighbourhood-year-type level, I cluster standard errors by neighbourhood-year.

Before discussing the main results, it is informative to compare the results of the dynamic model with those of a static model. An important part of the motivation for using a dynamic model is that changes in prices could be

[13]When choosing a minimum number of housing units per geographical units, there are several factors that need to be considered. A threshold lower than 10,000 might deliver a stronger first stage because it involves less aggregation. On the other hand, it would also mean that measures constructed at the neighbourhood level would be noisier, principally structural utility and also the repeat-sales index. Given that 10,000 was used by BMMT for their algorithm, and that it delivers a KP statistic of just about 10, I decided to let it be.

	Log(Pop.Dens.)	Log(Pop.Dens.),ver.2	Log(Pop.Dens.),ver.3
	b/se/t	b/se/t	b/se/t
R6-8,non-cont,wide	2.430	2.767	2.905
	1.290	1.095	1.089
	(1.884)$^{+}$	(2.527)*	(2.668)**
R9-10,non-cont,wide	5.837	5.448	5.194
	1.331	1.261	1.208
	(4.386)***	(4.320)***	(4.298)***
R9-10,cont,wide	6.317	6.122	5.714
	1.478	1.422	1.311
	(4.274)***	(4.305)***	(4.357)***
Observations	1320	1316	1320
r2	0.735	0.785	0.788

+=10% *=5% **=1% ***=0.1%. 120 neighbourhoods x 11 years.
Three distinct instruments.
Controls: fully saturated land use and street interactions,
borough, year, subway distance quintile FE.
SEs clustered by neighbourhood.

Table 4: Disaggregated First Stage Results

	Log(Pop.Dens.)	Log(Pop.Dens.),ver.2	Log(Pop.Dens.),ver.3
	b/se/t	b/se/t	b/se/t
Z	3.921	4.022	3.968
	1.250	1.161	1.107
	(3.136)**	(3.464)***	(3.585)***
Observations	1320	1316	1320
weakID	9.836	12.00	12.85
r2	0.727	0.779	0.784

+=10% *=5% **=1% ***=0.1%. 120 neighbourhoods x 11 years.
Single combined instrument.
Controls: fully saturated land use and street interactions,
borough, year, subway distance quintile FE.
SEs clustered by neighbourhood.

Table 5: Disaggregated First Stage Results

	(1) static, white	(2) dynamic, white	(3) static, black	(4) dynamic, black
ln(pop.dens.)	-0.491*	-0.240	0.446	1.681***
	(0.245)	(0.384)	(0.367)	(0.504)
88kUSD × ln(pop.dens.)	-0.0285	-0.0596*	-0.129***	-0.0870**
	(0.0233)	(0.0242)	(0.0297)	(0.0289)
126kUSD × ln(pop.dens.)	-0.0804	-0.203***	-0.235***	-0.208***
	(0.0511)	(0.0409)	(0.0624)	(0.0555)
321kUSD × ln(pop.dens.)	-0.200**	-0.731***	-0.304***	-1.071***
	(0.0649)	(0.0752)	(0.0740)	(0.146)
pval_88kUSD	0.0324	0.433	0.390	0.00156
pval_126kUSD	0.0162	0.244	0.568	0.00331
pval_321kUSD	0.00360	0.0105	0.702	0.229
N	15144	190080	13740	189120

Standard errors in parentheses
120 neighb. x 11 years x 12 inc. x 12 wealth.; 4th, 8th, 12th income types shown
weighted by race-income-wealth counts.
Controls: same as 1st stage + income FE, income × race shares and log med. neighb. inc.
SEs clustered by neighbourhood-year.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6: Static vs Dynamic

correlated with amenities (in this case, population density) even if the amenities tend to be stable over time, which biases parameter estimates if households gain utility from increasing wealth.

I present static and dynamic results for whites and blacks in Table 6. I display the baseline coefficient on log population density, as well as the interactions with income groups 4, 8, and 12 (there are 12 total, with the base level being the lowest). I include the average real income for an income group in the label as well. I display p-values for tests of the significance of the sum of the baseline coefficient and the interaction in the footer. Finally, I partial out the user cost (5% of the average price, as calculated using a repeat-sales index) using the marginal utility of wealth for each group. These parameters are organically estimated by the dynamic model; partialling out user cost for the static model using these estimates is analogous to assuming that prices are correctly instrumented in the static case.

Table 6 shows that the static model understates the income gradient of the density parameter for both groups, which makes sense when we consider the nature of the bias in the static model. As income increases, the observed location choices *vis-à-vis* density do not change much (conditional on covariates), which the static model interprets as similar taste for density across income. But, as

	(1) dynamic, white	(2) dynamic, black	(3) dynamic, hispanic	(4) dynamic, asian
ln(pop.dens.)	-0.240	1.681***	-0.154	0.674
	(0.384)	(0.504)	(0.446)	(0.472)
88kUSD × ln(pop.dens.)	-0.0596*	-0.0870**	-0.0707***	-0.0946*
	(0.0242)	(0.0289)	(0.0186)	(0.0421)
126kUSD × ln(pop.dens.)	-0.203***	-0.208***	-0.285***	-0.216**
	(0.0409)	(0.0555)	(0.0455)	(0.0664)
321kUSD × ln(pop.dens.)	-0.731***	-1.071***	-0.916***	-0.564***
	(0.0752)	(0.146)	(0.0957)	(0.0899)
pval_88kUSD	0.433	0.00156	0.613	0.215
pval_126kUSD	0.244	0.00331	0.319	0.325
pval_321kUSD	0.0105	0.229	0.0153	0.812
N	190080	189120	189480	189240

Standard errors in parentheses
120 neighb. x 11 years x 12 inc. x 12 wealth.; 4th, 8th, 12th income types shown
weighted by race-income-wealth counts.
Controls: same as 1st stage + income FE, income × race shares and log med. neighb. inc.
SEs clustered by neighbourhood-year.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Main Results

density increases, so does the expected price change, and the dynamic model recognizes that the larger wealth increase compensates the rich for their greater distaste for density. On the other hand the poor like density more and want wealth more, but they are deterred by higher user costs, such that location choices over density do not vary much with income. While the total coefficient – sum of baseline and interaction – is significant for whites of all income levels in the static model, the larger standard error for the baseline dynamic coefficient means that we can only reject the null hypothesis of no effect for the richest whites.

Now that we have a rational story for how theoretical biases of the static model manifest empirically, we can move on to the main results of the dynamic model, presented in Table 7 (columns 1 and 3 here are the same as columns 2 and 4 in Table 6). Black households also have a steep gradient, but starting from a strong taste for density, while asian households have no gradient to speak of starting from a baseline coefficient that is just short of positive 5% significance.

Main results are also presented graphically in Figure 6. At the top left is the income gradient of the density parameter, relative to the baseline, reiterating the result that all racial groups have a negative income gradient. The gradient is roughly linear, possibly flattening out around 200k.

At top right are the sum of the gradient coefficients and the baseline coefficients, for which I have not yet calculated standard errors. We can see that differences within race and across income are smaller than differences between races. At bottom left I have taken the ratio of 10% of these coefficients to the marginal utility of wealth to obtain the willingness to pay for a 10% increase in density, in thousands of 2012 USD. Ranging in the tens of thousands, it is substantial.

At bottom right, and in greater size in Figure 7, we can see the willingness to pay for a 10% increase in as a share of income. The flatness of the lines for white and hispanic households at about 1.5% of income indicates that density is a homothetic bad for these groups. On the other hand, the line for black households declines quickly from 10% toward 0, indicating that density is an inferior good. Since the line for asian households is based on estimates that are universally non-significant, I draw no conclusion.

Although the heterogeneity between races is extreme, I do think it makes sense when considering the generality of population density. I purposely chose this broad measure because it carries with it multiple effects, and this instrument because (in theory) it is not dominated by any one of these effects while still removing the effect of any correlation with ξ^{τ}_{jt}. These different effects can be disentangled by adding additional controls. For example, it is possible that black households have a taste for density due to the consumption externalities it entails. Although location sorting models generally may conflate the preferences of groups that are discriminated against with those of the groups doing the discriminating, the use of IV should protect against this bias.

7 Conclusion

In this paper, I estimate the demand for population density as a non-market neighbourhood characteristic, with heterogeneity by race and income, using publicly-available data form New York City. To obtain consistent estimates I leverage variation in population density resulting from the microgeography of the combination of zoning and the street grid, and I neutralize the effects of prices (in levels and changes) using the dynamic demand framework of Bayer et al. (2016). I find that density is a homothetic bad for white and hispanic households valued at about -1.5% of income for 10% of density, and an inferior good for black households with a willingness to pay that declines from 10% to less than 1%. However, while economically significant, the results are only statistically significant for the highest-income whites hispanics and low-to-mid-income blacks, and so the evidence is suggestive at most. Further research is necessary before substantive conclusions can be drawn.

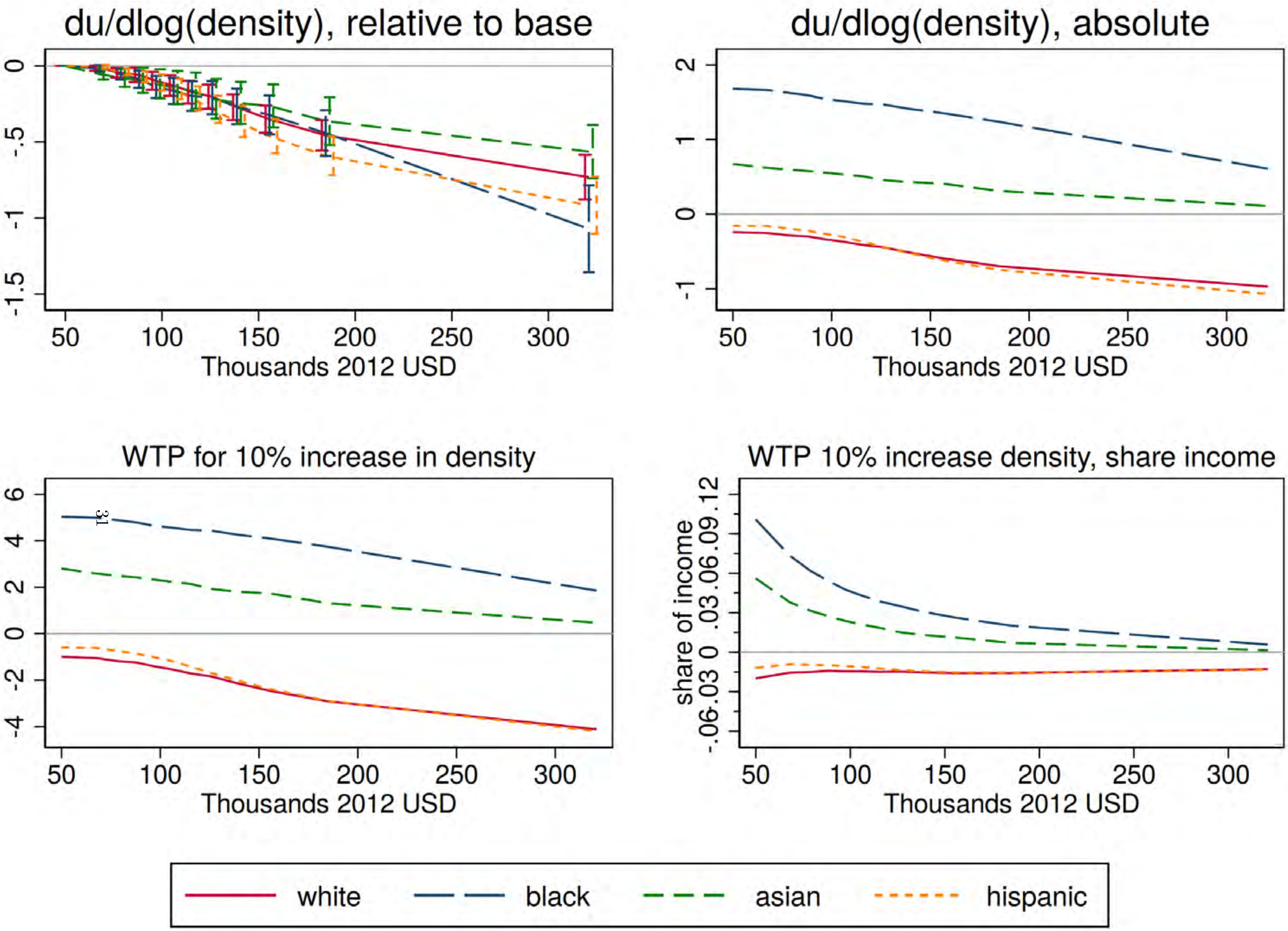

Figure 6: Main Results

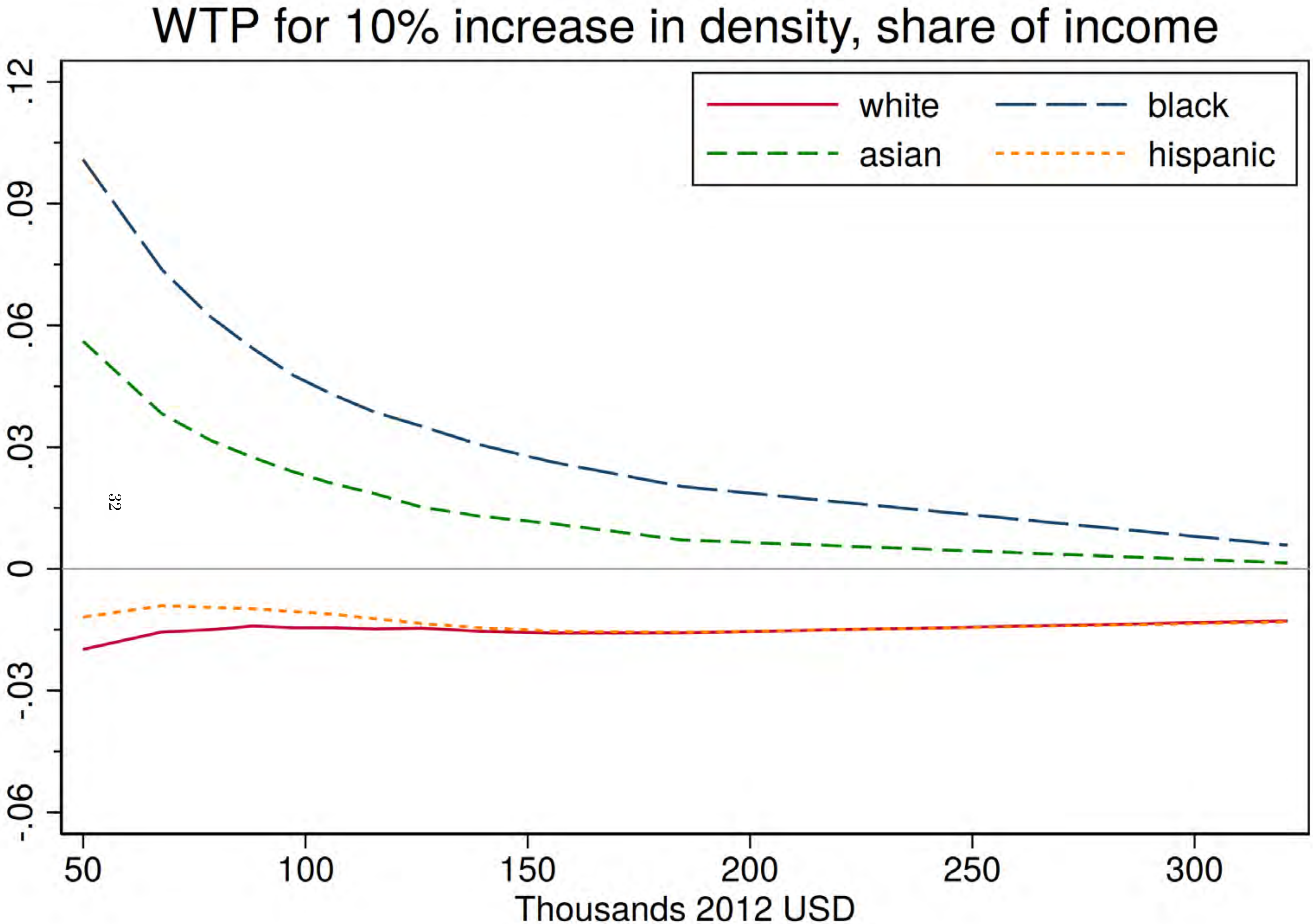

Figure 7: Willingness to Pay for a Doubling of Density as a Share of Income

References

Milena Almagro and Tomás Domınguez-Iino. Location sorting and endogenous amenities: Evidence from amsterdam. *NYU, mimeograph*, 2019.

Nathaniel Baum-Snow and Lu Han. The microgeography of housing supply. *Work in progress, University of Toronto*, 2019.

Patrick Bayer, Fernando Ferreira, and Robert McMillan. A unified framework for measuring preferences for schools and neighborhoods. *Journal of political economy*, 115(4):588–638, 2007.

Patrick Bayer, Robert McMillan, Alvin Murphy, and Christopher Timmins. A dynamic model of demand for houses and neighborhoods. *Econometrica*, 84 (3):893–942, 2016.

Nicholas Chiumenti, Amrita Kulka, and Aradhya Sood. How to increase housing affordability: Understanding local deterrents to building multifamily housing. 2022.

City Planning Commission. Zoning maps and resolution, 1961. `https://www1.nyc.gov/assets/planning/download/pdf/about/city-planning-history/zoning_maps_and_resolution_1961.pdf`.

City Planning Commission. Zoning resolution: Articles i-xiv and appendices, N 870385(A) ZRY 1987. `https://www1.nyc.gov/assets/planning/download/pdf/about/cpc/870385a.pdf`.

City Planning Commission. Zoning resolution: Articles i-xiv and appendices, 2022. `https://zr.planning.nyc.gov`.

Pierre-Philippe Combes, Gilles Duranton, and Laurent Gobillon. The Costs of Agglomeration: House and Land Prices in French Cities. *The Review of Economic Studies*, 86(4):1556–1589, 10 2018. ISSN 0034-6527. doi: 10.1093/restud/rdy063. URL `https://doi.org/10.1093/restud/rdy063`.

Victor Couture, Cecile Gaubert, Jessie Handbury, and Erik Hurst. Income growth and the distributional effects of urban spatial sorting, 2019. National Bureau of Economic Research Working Paper No. 26142. Available at `https://www.nber.org/papers/w26142`.

Fred J Damerau. A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3):171–176, 1964.

Thomas Davidoff et al. Supply constraints are not valid instrumental variables for home prices because they are correlated with many demand factors. *Critical Finance Review*, 5(2):177–206, 2016.

Department of City Planning. Zoning handbook, 2018. `https://www1.nyc.gov/assets/planning/download/pdf/about/publications/zoning-handbook/zoning-handbook.pdf`.

Rebecca Diamond. The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review*, 106(3):479–524, 2016.

Juan C Duque, Luc Anselin, and Sergio J Rey. The max-p-regions problem. *Journal of Regional Science*, 52(3):397–419, 2012.

Gilles Duranton and Diego Puga. Urban growth and its aggregate implications. Technical report, National Bureau of Economic Research, 2019.

D. Gale and L. S. Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962. ISSN 00029890, 19300972. URL `http://www.jstor.org/stable/2312726`.

Joseph Gyourko and Raven Molloy. Chapter 19 - regulation and housing supply. In Gilles Duranton, J. Vernon Henderson, and William C. Strange, editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 1289–1337. Elsevier, 2015. doi: https://doi.org/10.1016/B978-0-444-59531-7.00019-3. URL `https://www.sciencedirect.com/science/article/pii/B9780444595317000193`.

Chang-Tai Hsieh and Enrico Moretti. Housing constraints and spatial misallocation. *American Economic Journal: Macroeconomics*, 11(2):1–39, 2019.

John Kennan and James R. Walker. The effect of expected income on individual migration decisions. *Econometrica*, 79(1):211–251, 2011.

Eric Kober. Housing the next million new yorkers near transit. Technical report, New York University Wagner Rudin Center for Transportation Policy and Management, July 2018. `https://wagner.nyu.edu/files/faculty/publications/Housing_July27-18_0.pdf`.

Amrita Kulka. Sorting into neighborhoods: The role of minimum lot sizes. *job market paper, unpublished*, 2019.

Vladimir I Levenshtein et al. Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady*, volume 10, pages 707–710. Soviet Union, 1966.

Evan Rosenman, Santiago Olivella, and Kosuke Imai. Race and ethnicity data for first, middle, and last names, 2022. URL `https://doi.org/10.7910/DVN/SGKW0K`.

Stuart S Rosenthal and William C Strange. Evidence on the nature and sources of agglomeration economies. In *Handbook of regional and urban economics*, volume 4, pages 2119–2171. Elsevier, 2004.

Steven Ruggles, Sarah Flood, Ronald Goeken, Megan Schouweiler, and Matthew Sobek. *Integrated Public Use Microdata Series (IPUMS): Version 12.0,* [Machine-readable database]. Minneapolis: University of Minnesota, 2022.

John Rust. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society*, pages 999–1033, 1987.

Jaehee Song. The effects of residential zoning in us housing markets. *Available at SSRN 3996483*, 2021.

Konstantinos Tzioumis. Data for: Demographic aspects of first names, 2018. URL `https://doi.org/10.7910/DVN/TYJKEZ`.

A Appendix 2: Model

A.1 Motivation

Lemma 1 *Under the assumptions of the BMMT model with no unobserved heterogeneity, the continuation value may be expressed as*

$$c_{jt}^{\tau} = m_{t+1}^{\tau_{t+1}} + \tilde{v}_{j,t+1}^{\tau_{t+1}} - \ln(\mathbb{P}_{j,t+1}^{stay,\tau_{t+1}}).$$

Proof. Denote with c_{jt}^{τ} the continuation value for the households DP problem, which is a random variable.

$$c_{jt}^{\tau} = \ln\big(\exp(v_{j,t+1}^{\tau_{t+1}}) + \sum_{k=0}^{J}\exp(v_{k,t+1}^{\bar{\tau}_{t+1}} - Z^{\bar{\tau}_{t+1}\prime}\gamma_p)\big)$$

Recall that value functions are the sum of the demeaned part and the normalizing constant, $v_{jt}^{\tau} = \tilde{v}_{jt}^{\tau} + m_t^{\tau}$. The continuation value may be rewritten as

$$\begin{aligned}
c_{jt}^{\tau} =\ & \ln\big(\exp(\tilde{v}_{j,t+1}^{\tau_{t+1}} + m_{t+1}^{\tau_{t+1}}) + \sum_{k=0}^{J}\exp(\tilde{v}_{k,t+1}^{\bar{\tau}_{t+1}} + m_t^{\bar{\tau}_{t+1}} - Z^{\bar{\tau}_{t+1}\prime}\gamma_p)\big) \\
=\ & \ln\Big(\exp(m_{t+1}^{\tau_{t+1}})\big(\exp(\tilde{v}_{j,t+1}^{\tau_{t+1}}) + \sum_{k=0}^{J}\exp(\tilde{v}_{k,t+1}^{\bar{\tau}_{t+1}} + m_{t+1}^{\bar{\tau}_{t+1}} - m_{t+1}^{\tau_{t+1}} - Z^{\bar{\tau}_{t+1}\prime}\gamma_p)\big)\Big) \\
=\ & m_{t+1}^{\tau_{t+1}} + \ln\Big(\exp(\tilde{v}_{j,t+1}^{\tau_{t+1}}) + \sum_{k=0}^{J}\exp(\tilde{v}_{k,t+1}^{\bar{\tau}_{t+1}} - Z^{\bar{\tau}_{t+1}\prime}\gamma_p - 0.06p_{j,t+1}Z^{\bar{\tau}_{t+1}\prime}\gamma_f)\Big) \\
=\ & m_{t+1}^{\tau_{t+1}} + \ln\left(\Big[\frac{\exp(\tilde{v}_{j,t+1}^{\tau_{t+1}})}{\exp(\tilde{v}_{j,t+1}^{\tau_{t+1}})}\Big]\Big[\exp(\tilde{v}_{j,t+1}^{\tau_{t+1}}) + \sum_{k=0}^{J}\exp(\tilde{v}_{k,t+1}^{\bar{\tau}_{t+1}} - Z^{\bar{\tau}_{t+1}\prime}\gamma_p - 0.06p_{j,t+1}Z^{\bar{\tau}_{t+1}\prime}\gamma_f)\Big]\right) \\
=\ & m_{t+1}^{\tau_{t+1}} + \ln\Big(\exp(\tilde{v}_{j,t+1}^{\tau_{t+1}})(\mathbb{P}_{j,t+1}^{\text{stay},\tau_{t+1}})^{-1}\Big) \\
=\ & m_{t+1}^{\tau_{t+1}} + \tilde{v}_{j,t+1}^{\tau_{t+1}} - \ln(\mathbb{P}_{j,t+1}^{\text{stay},\tau_{t+1}}),
\end{aligned}$$

where we have used the normalization and the definition of $\mathbb{P}_{j,t+1}^{\text{stay},\tau_{t+1}}$. □

Proposition 1 *Let the static model be a model where the value function is a sum of unchanging discounted flow utility over an infinite horizon, because the neighbourhood is assumed to be unchanging and the household is assumed to never move. If BMMT is the true model but the static model is estimated instead, then flow utility will be incorrectly estimated as*

$$\bar{u}_{jt}^{\tau} = \beta^{\tau}X_{jt}^{\tau} + \xi_{jt}^{\tau} + \delta Z^{\tau\prime}\gamma_f^{\tau}\mathbb{E}[p_{j,t+1} - p_{jt}|s_t] + \delta\mathbb{E}[\tilde{v}_{j,t+1}^{\tau_{t+1}} - \tilde{v}_{jt}^{\tau} - \ln(\mathbb{P}_{j,t+1}^{stay,\tau_{t+1}})|s_t],$$

where the expected change in price $\delta Z^{\tau\prime}\gamma_f^{\tau}\mathbb{E}[p_{j,t+1} - p_{jt}|s_t]$ enters as an omitted variable.

Proof. The static model observes the same $\bar{v}^{\tau}_{jt} = v^{\tau}_{jt} = \tilde{v}^{\tau}_{jt} + m^{\tau}_{t}$, but decomposes it into $\bar{v}^{\tau}_{jt} = \bar{u}^{\tau}_{jt} + \delta\bar{v}^{\tau}_{jt}$, falsely assuming that neighbourhood characteristics are constant over time and that the household will never move. The difference in utility is given by

$$\begin{aligned}
\bar{u}^{\tau}_{jt} - u^{\tau}_{jt} &= \cancel{\bar{v}^{\tau}_{jt} - v^{\tau}_{jt}} - \delta\bar{v}^{\tau}_{jt} + \delta\mathbb{E}[c^{\tau}_{jt}|s_t] \\
&= -\delta(\tilde{v}^{\tau}_{jt} + m^{\tau}_{t}) + \delta\mathbb{E}[m^{\tau_{t+1}}_{t+1} + \tilde{v}^{\tau_{t+1}}_{j,t+1} - \ln(\mathbb{P}^{\text{stay},\tau_{t+1}}_{j,t+1})|s_t] \\
&= \delta\mathbb{E}[m^{\tau_{t+1}}_{t+1} - m^{\tau}_{t}|s_t] + \delta\mathbb{E}[\tilde{v}^{\tau_{t+1}}_{j,t+1} - \tilde{v}^{\tau}_{jt} - \ln(\mathbb{P}^{\text{stay},\tau_{t+1}}_{j,t+1})|s_t] \\
&= \delta Z^{\tau\prime}\gamma^{\tau}_{f}\mathbb{E}[p_{j,t+1} - p_{jt}|s_t] + \delta\mathbb{E}[\tilde{v}^{\tau_{t+1}}_{j,t+1} - \tilde{v}^{\tau}_{jt} - \ln(\mathbb{P}^{\text{stay},\tau_{t+1}}_{j,t+1})|s_t],
\end{aligned}$$

where the expected discounted utility from changing wealth is equal to the expected discounted utility from an increase in the price of the house. Then, the false static utility is given by

$$\bar{u}^{\tau}_{jt} = \beta^{\tau}X^{\tau}_{jt} + \xi^{\tau}_{jt} + \delta Z^{\tau\prime}\gamma^{\tau}_{f}\mathbb{E}[p_{j,t+1} - p_{jt}|s_t] + \delta\mathbb{E}[\tilde{v}^{\tau_{t+1}}_{j,t+1} - \tilde{v}^{\tau}_{jt} - \ln(\mathbb{P}^{\text{stay},\tau_{t+1}}_{j,t+1})|s_t],$$

where the expected change in price $\delta Z^{\tau\prime}\gamma^{\tau}_{f}\mathbb{E}[p_{j,t+1} - p_{jt}|s_t]$ enters as an omitted variable.

A.2 Forecasting Prices

In this section I will show that the definition of elasticity can be used to obtain a price forecast fixed point, some preliminaries are required before the formal statement and proof.

When future prices are given by shifts along a housing supply curve plus an idiosyncratic error,

$$p_{j,t+1} = \left(\frac{N_{j,t+1} - N_{jt}}{N_{jt}}\right)\frac{p_{jt}}{\eta_j} + p_{jt} + \epsilon^{p}_{j,t+1},$$

then the expected future price is given by

$$\mathbb{E}[p_{j,t+1}|s_t] \equiv \hat{p}_{j,t+1} = \left(\frac{\overbrace{\hat{N}_{j,t+1}}^{\equiv\mathbb{E}[N_{j,t+1}|s_t]} - N_{jt}}{N_{jt}}\right)\frac{p_{jt}}{\eta_j} + p_{jt},$$

since the error is zero in expectation and the present prices and locations are known. The shifts are themselves affected by realizations of future prices, shown by the following accounting identity:

$$\begin{aligned}
\hat{N}_{j,t+1} = \sum_{\bar{\tau}}\Bigg(\mathbb{E}\Big[&N^{\bar{\tau}-\tilde{p}_{j,t+1}}_{jt}\mathbb{P}^{\text{stay},\bar{\tau}}_{j,t+1}(p_{j,t+1}) \\
&+ \sum_{k\neq j}N^{\tau-\tilde{p}_{k,t+1}}_{kt}(1 - \mathbb{P}^{\text{stay},\tau}_{k,t+1}(p_{k,t+1}))(1 - \mathbb{P}^{\tau}_{0,t+1})\mathbb{P}^{\tau}_{j,t+1}\Big|s_t\Big]\Bigg).
\end{aligned}$$

(Type is taken to represent just integer-valued wealth for now.) The stay CCP depends on prices through moving costs. Additionally, changes in prices $\tilde{p}_{j,t+1}$ determine which group of households maps to which CCP. We can re-index this summation so that the change in wealth appears in the CCP index:

$$\hat{N}_{j,t+1} = \sum_{\bar{\tau}} \Bigg(\mathbb{E}\Big[N_{jt}^{\bar{\tau}} \mathbb{P}_{j,t+1}^{\text{stay},\bar{\tau}+\tilde{p}_{j,t+1}}(p_{j,t+1}) + \sum_{k \neq j} N_{kt}^{\tau}(1 - \mathbb{P}_{k,t+1}^{\text{stay},\tau+\tilde{p}_{k,t+1}}(p_{k,t+1}))(1 - \mathbb{P}_{0,t+1}^{\tau})\mathbb{P}_{j,t+1}^{\tau} \Big| s_t \Big] \Bigg).$$

Originally, the $\bar{\tau}$ summation index represented types in the next period after price changes, which necessitated subtracting price changes in the N index. But we can also let $\bar{\tau}$ represent types in the current period, and we keep track of the changing mapping between populations and CCPs in the CCP index. In an empirical implementation, types need to be defined such that there is not a one-to-one mapping, but for now I abstract away from this concern.

Now, we may write the expected future housing demand as a function of the expected future price:

$$\hat{N}_{j,t+1}(x) = \sum_{\bar{\tau}} \Bigg(\mathbb{E}\Big[N_{jt}^{\bar{\tau}} \mathbb{P}_{j,t+1}^{\text{stay},\bar{\tau}+\tilde{p}_{j,t+1}}(p_{j,t+1}) + \sum_{k \neq j} N_{kt}^{\tau}(1 - \mathbb{P}_{k,t+1}^{\text{stay},\tau+\tilde{p}_{k,t+1}}(p_{k,t+1}))(1 - \mathbb{P}_{0,t+1}^{\tau})\mathbb{P}_{j,t+1}^{\tau} \Big| s_t, \hat{p}_{j,t+1} = x \Big] \Bigg).$$

With formal statements about the mutual dependence of $\hat{p}_{j,t+1}$ and $\hat{N}_{j,t+1}$, we may define the function

$$T(x) = \left(\frac{\hat{N}_{j,t+1}(x) - N_{jt}}{N_{jt}} \right) \frac{p_{jt}}{\eta_j} + p_{jt}$$

I will show that this function is a contraction mapping, in the case where relative preferences for locations do not vary over wealth, and given some further technical assumptions (and a strong assumption that is a stand-in for a future argument). Then, iterating this function will eventually lead to a price and population forecast that satisfies this definition and the accounting identity for the future population distribution.

In the argument to follow, τ will only denote immutable characteristics (race, income) and not endogenous wealth, since currently I only have a proof for the case where people of different wealth levels have the same preferences. This means that arithmetic in superscripts will be removed. Households will continue to gain utility from wealth, but their preferences over locations and moving costs do not change.

Proposition 2 *Let*

$$\tilde{v}_{jt}^{\tau} = \tilde{v}_{jt}^{\bar{\tau}} \ \forall \ \tau, \bar{\tau},$$

(preferences for locations do not vary over wealth), let

$$\mathbb{E}\Big[(1-\mathbb{P}^{stay,\tau}_{k,t+1}(p_{k,t+1}))(1-\mathbb{P}^{\tau}_{0,t+1})\mathbb{P}^{\tau}_{j,t+1}\Big|s_t,\hat{p}_{j,t+1}\Big] = \mathbb{E}\Big[(1-\mathbb{P}^{stay,\tau}_{k,t+1}(p_{k,t+1}))(1-\mathbb{P}^{\tau}_{0,t+1})\mathbb{P}^{\tau}_{j,t+1}\Big|s_t,\Big],$$

(price forecasts do not affect expected movers), and let $\mathbb{P}^{stay,\tau}_{jt} > 1/2$ for all j,t,τ. Then, there exists an $\underline{x} \in \mathbb{R}$ such that T is a contraction mapping with the Euclidean norm on the metric space $\{x \in \mathbb{R} : x > \underline{x}\}$.

Proof. T is a contraction mapping if for two potential price forecasts x, y and some $\lambda < 1$ we have

$$|T(x) - T(y)| \leq \lambda |x - y|.$$

This statement is equivalent to

$$|\hat{N}_{j,t+1}(x) - \hat{N}_{j,t+1}(y)| \leq \frac{\lambda \eta_j N_{jt}}{p_{jt}} |x - y|,$$

where

$$\begin{aligned}
\hat{N}_{j,t+1}(x) - \hat{N}_{j,t+1}(y) = \sum_{\tau} \Bigg(& \mathbb{E}\Big[N^{\tau}_{jt}\mathbb{P}^{\text{stay},\tau}_{j,t+1}(p_{j,t+1})\Big|s_t, \hat{p}_{j,t+1} = x\Big] \\
- \ & \mathbb{E}\Big[N^{\tau}_{jt}\mathbb{P}^{\text{stay},\tau}_{j,t+1}(p_{j,t+1})\Big|s_t, \hat{p}_{j,t+1} = y\Big] \\
+ \ & \sum_{k \neq j} \mathbb{E}\Big[N^{\tau}_{kt}(1-\mathbb{P}^{\text{stay},\tau}_{k,t+1}(p_{k,t+1}))(1-\mathbb{P}^{\tau}_{0,t+1})\mathbb{P}^{\tau}_{j,t+1}\big|s_t, \hat{p}_{j,t+1} = x\Big] \\
- \ & \sum_{k \neq j} \mathbb{E}\Big[N^{\tau}_{kt}(1-\mathbb{P}^{\text{stay},\tau}_{k,t+1}(p_{k,t+1}))(1-\mathbb{P}^{\tau}_{0,t+1})\mathbb{P}^{\tau}_{j,t+1}\big|s_t, \hat{p}_{j,t+1} = y\Big]\Bigg).
\end{aligned}$$

If not for the differing forecasts that movers are conditioned on, all mover terms (lines 3 and 4) would cancel, leaving only the different stayer terms (lines 1 and 2). In the assumptions of the proposition, I have included that these mover terms do not depend on price forecasts x and y, so they would cancel. This assumption is probably innocuous (should be examined in future work), for reasons explained below.

Movers do depend on price forecasts, but the effect is ambiguous. On one hand, a higher price in j means less people will move to j. On the other hand, a higher price in j means fewer movers from j, less housing demand in k, a lower price in k, and more movers from k, some of whom will choose j. My strong intuition is that the first effect dominates, because demand curves slope downward. If this is the case, then lines 3 and 4 are negative. Ignoring these lines is the same as adding a positive number, so the "true" value of the expression is less than the "simplified" value. To show that the "simplified" expression meets the requirements of the proof then implies that the "true" expression does also. We will then continue with the result that

$$\hat{N}_{j,t+1}(x)-\hat{N}_{j,t+1}(y) = \sum_{\tau} N_{jt}^{\tau}\left(\mathbb{E}\Big[\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})\Big|s_t,\hat{p}_{j,t+1}=x\Big]-\mathbb{E}\Big[\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})\Big|s_t,\hat{p}_{j,t+1}=y\Big]\right).$$

Then, (the necessary conditions should be satisfied such that) the derivative of the expectation is equal to the expectation of the derivative:

$$\begin{aligned}
&\frac{\partial}{\partial x}\mathbb{E}\Big[\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})\Big|s_t,\hat{p}_{j,t+1}=x\Big]\\
&=\int\int\frac{\partial}{\partial x}\left(\frac{\exp(\hat{\hat{v}}_{j,t+1}^{\tau}+\epsilon_{j,t+1}^{\tau v}+Z^{\tau\prime}\gamma_p+Z^{\tau\prime}\gamma_f 0.06(x+\epsilon_{j,t+1}^{p}))}{\exp(\hat{\hat{v}}_{j,t+1}^{\tau}+\epsilon_{j,t+1}^{\tau v}+Z^{\tau\prime}\gamma_p+Z^{\tau\prime}\gamma_f 0.06(x+\epsilon_{j,t+1}^{p}))+\sum_k\exp(\hat{\hat{v}}_{k,t+1}^{\tau}+\epsilon_{k,t+1}^{\tau v})}\right)dF(\epsilon_{j,t+1}^{\tau v})dF(\epsilon_{j,t+1}^{p})\\
&=Z^{\tau\prime}\gamma_f 0.06\mathbb{E}\Big[\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})(1-\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1}))\Big|s_t,\hat{p}_{j,t+1}=x\Big]>0.
\end{aligned}$$

This is positive and by the same argument

$$\begin{aligned}
\frac{\partial}{\partial x}\left(\frac{\partial}{\partial x}\mathbb{E}\Big[\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})\Big|s_t,\hat{p}_{j,t+1}=x\Big]\right)&=(Z^{\tau\prime}\gamma_f 0.06)^2\mathbb{E}\Big[\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})(1-\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1}))(1-2\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1}))\Big|s_t,\hat{p}_{j,t+1}=x\Big]\\
&>0 \iff \mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})>1/2,
\end{aligned}$$

which is assumed (and almost universally true in the data). Hence the expectation of the stay CCP is increasing and concave in the price forecast. Then, it must be that the difference in the stay CCPs for two price forecasts x, y is less than the linear approximation of the stay CCP where the derivative is evaluated at the lower point, which we will take to be y. That is,

$$\hat{N}_{j,t+1}(x)-\hat{N}_{j,t+1}(y)\le(x-y)\sum_{\tau}N_{jt}^{\tau}Z^{\tau\prime}\gamma_f 0.06\mathbb{E}\Big[\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})(1-\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1}))\Big|s_t,\hat{p}_{j,t+1}=y\Big].$$

Finally, T is a contraction as long as

$$(x-y)\sum_{\tau}N_{jt}^{\tau}Z^{\tau\prime}\gamma_f 0.06\mathbb{E}\Big[\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})(1-\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1}))\Big|s_t,\hat{p}_{j,t+1}=y\Big]\le\frac{\lambda\eta_j N_{jt}}{p_{jt}}(x-y),$$

which is the same as

$$\sum_{\tau}N_{jt}^{\tau}(m^{\tau}-m^{\bar{\tau}(\tau,p_{jt})})\mathbb{E}\Big[\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})(1-\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1}))\Big|s_t,\hat{p}_{j,t+1}=y\Big]\le\lambda\eta_j N_{jt}.$$

If we take the limit

$$\lim_{y\to\infty}\mathbb{E}\Big[\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1})(1-\mathbb{P}_{j,t+1}^{\text{stay},\tau}(p_{j,t+1}))\Big|s_t,\hat{p}_{j,t+1}=y\Big]=0,$$

which intuitively is because as a choice attribute goes to infinity then the CCP goes to one (or zero), and so further changes in any attribute have smaller impacts on the CCP. Because the left hand side approaches zero, there must be some $\underline{y}$ such that the expression holds for all $y>\underline{y}$. □

B Appendix 2: Zoning Tables

Contextual districts R9-10 depend on street width.

R6-R10 Contextual Districts

Use			R6A	R6B	R7A	R7B	R7D	R7X	R8A	R8B	R8X	R9A	R9D	R9X	R10A	R10X
Single-Family	Detached	Use Group 1	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Single & Two-Family / Multi-Family	All	Use Group 2	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Community Facility		Use Groups 3, 4	•	•	•	•	•	•	•	•	•	•	•	•	•	•
Bulk																
Lot Area (min)	All		1,700 sf													
Lot Width (min)	All		18 ft													
Rear Yard (min)	All		30 ft													
Lot Coverage (max)	Corner Lot		100%													
	Other Lot	Narrow Street / Wide Street	65%	60%	65%			70%								
Residential FAR	Basic	Narrow Street / Wide Street	3.00	2.00	4.00	3.00	4.20	5.00	6.02	4.00	6.02	7.52	9.00		10.00	
	MIH / VIH	Narrow Street / Wide Street	3.60	2.20	4.60	n/a	5.60	6.00/ 5.00	7.20	n/a	7.20	8.50	10.00	9.70	12.00	
Community Facility FAR			3.00	2.00	4.00	3.00	4.20	5.00	6.50	4.00	6.00	7.50	9.00		10.00	
Base Height (min-max) Outside MN Core	Basic (w/QGF)	Narrow Street	40-60 ft (65ft)	30-40 ft (45 ft)	40-65 ft (75 ft)	40-65 ft	60-85 ft	60-85 ft (95 ft)	60-85 ft (95 ft)	55-65 ft	60-85 ft (95 ft)	60-95 ft	60-85 ft or 15-25ft (if facing elevated rail)	60-120 ft (125 ft)	60-125 ft	60-85 ft
		Wide Street										60-105 ft		105-120ft (125 ft)	125-155ft	
	MIH/VIH (w/QGF)	Narrow Street	40-65 ft	30-40 ft (45 ft)	40-75 ft	n/a	60-95 ft	60-105 ft/ 60-85 ft (95 ft)	60-105 ft	n/a	60-105 ft	60-125 ft		60-145 ft	60-155 ft	60-85 ft
		Wide Street												105-145ft	125-155 ft	
Building Height (max) Outside MN Core	Standard (w/QGF)	Narrow Street	70 ft (75 ft)	50 ft (55 ft)	80 ft (85 ft)	75 ft	100 ft (105 ft)	120 ft (125 ft)	120 ft (125 ft)	75 ft	150 ft (155 ft)	135 ft	n/a	180 ft (185 ft)	185 ft	n/a
		Wide Street										145 ft		170 ft (175 ft)	230 ft (215 ft)	
	MIH (w/QGF)	Narrow Street	80 ft (85 ft)	50 ft (55 ft)	90 ft (95 ft)	n/a	110 ft (115 ft)	140 ft (145 ft)/ 120 ft (125 ft)	140 ft (145 ft)	n/a	170 ft (175 ft)	160 ft (165 ft)	n/a	190 ft (195 ft)	230 ft (215 ft)	n/a
		Wide Street										170 ft (175 ft)		200 ft (205 ft)	230 ft (235 ft)	
Number of Stories (max)	Standard (w/QGF)	Narrow Street	n/a (7)	n/a (5)	n/a (8)	n/a	n/a (10)	n/a (12)	n/a (12)	n/a	n/a (15)	n/a	n/a	n/a (16)	n/a	n/a
		Wide Street												n/a (17)	n/a (21)	
	MIH / VIH (w/QGF)	Narrow Street	8	n/a (5)	9	n/a	11	14 / n/a (12)	14	n/a	17	16	n/a	19	21	n/a
		Wide Street										17		20	23	
Tower Lot Coverage (min-max)			n/a										33-40%	n/a		33-40%
Dwelling Unit Factor	All		680													
Parking																
General (min % of d.u.)	for Group Parking Facilities		50%						40%	50%	40%					
Reduced and Waived Requirements (min % of d.u.)	IRHU-outside Transit Zone		25%		15%	25%	15%		12%	15%	12%					
	AIRS-outside Transit Zone		10%													
	IRHU/AIRS-inside Transit Zone		0%													
	Small Lot	10,000sf or less	50%		30%	50%	30%		0%	n/a	0%					
		10,000-15,000 sf	n/a						20%	n/a	20%					
	waived if small # spaces		5		15	5	15									

Non-contextual districts R6-10 following QHP regulations depend on street width.

R6-R10 Non-Contextual Districts (Quality Housing)

Use			R6 QH	R7 QH R7-1	R7 QH R7-2	R8 QH	R9 QH	R10 QH
Single-Family	Detached	Use Group 1	•	•		•	•	•
Single & Two-Family / Multi-Family	All	Use Group 2	•	•		•	•	•
Community Facility		Use Groups 3, 4	•	•		•	•	•
Bulk								
Lot Area (min)	All		1,700 sf					
Lot Width (min)	All		18 ft					
Rear Yard (min)	All		30 ft					
Lot Coverage (max)	Corner Lot		100%					
	Other Lot	Narrow Street	60%	65%		70%		
		Wide Street	65%					
Residential FAR	Standard	Narrow Street	2.20	3.44		6.02	7.52	10.00
		Wide Street	3.60	4.00		7.20		
	MIH	Narrow/Wide Street	3.60	4.60		7.20	8.00	12.00
Community Facility FAR			4.80	4.80	6.50	6.50	10.00	10.00
Base Height (min-max) Outside MN Core	Standard	Narrow Street	30-45 ft	40-65 ft		60-85 ft	60-95 ft	60/125 ft
		Wide Street	40-65 ft	40-75 ft		60-95 ft	60-105 ft	125-155 ft
	MIH / VIH	Narrow Street	30-45 ft	40-65 ft		60-105 ft	60-125 ft	60-155 ft
		Wide Street	40-65 ft	40-75 ft				125-155 ft
Building Height (max) Outside MN Core	Standard (w/QGF)	Narrow Street	55 ft	75 ft		115 ft	135 ft	185 ft
		Wide Street	70ft (75 ft)	80 ft (85 ft)		130 ft (135 ft)	145 ft	210 ft
	MIH (w/QGF)	Narrow Street	80 ft (85 ft)	100 ft (105 ft)		215 ft	160 ft (165 ft)	210 ft (215 ft)
		Wide Street					170 ft (175 ft)	230 ft (235 ft)
	VIH (w/QGF)	Narrow Street	55 ft	75 ft		140 ft (145 ft)	160 ft (165 ft)	210 ft (215 ft)
		Wide Street	80 ft (85 ft)	100 ft (105 ft)			170 ft (175 ft)	230 ft (235 ft)
Number of Stories (max)	Standard (w/QGF)	Narrow Street	n/a	n/a		n/a	n/a	21
		Wide Street	n/a (7)	n/a (8)		n/a (13)		
	MIH (w/QGF)	Narrow Street	8	13		14	16	21
		Wide Street					17	23
	VIH (w/QGF)	Narrow Street	n/a	n/a		21	16	21
		Wide Street	n/a (8)	8			17	23
Tower Lot Coverage (min-max)			n/a					
Dwelling Unit Factor	All		680					
Parking								
General (min % of d.u.)	for Group Parking Facilities		50%			40%		
Reduced and Waived Requirements (min% of d.u.)	IRHU-outside Transit Zone		25%	15%		12%		
	AIRS-outside Transit Zone		10%					
	IRHU/AIRS-inside Transit Zone		0%					
	Small Lot Area	10,000sf or less	50%		0%			
		10,000-15,000sf	n/a		30%	20%		
	waived if small # spaces		5	5(15)	15			

R6-R10 Non-Contextual Districts (Height Factor and Tower)

Use			R6 HF	R7 HF R7-1	R7 HF R7-2	R8 HF	R9 HF	R9 TOB	R10 ST	R10 TOB
Single-Family	Detached	Use Group 1	•	•		•	•		•	
Single & Two-Family / Multi-Family	All	Use Group 2	•	•		•	•		•	
Community Facility		Use Groups 3, 4	•	•		•	•		•	
Bulk										
Lot Area (min)	All		1,700 sf							
Lot Width (min)	All		18 ft							
Rear Yard (min)	All		30 ft							
Residential FAR	Standard		0.78-	0.87-3.44		0.95-	0.99-	7.52	10	
	MIH		2.43			6.02	7.52	8	10	12
Community Facility FAR			4.8	4.8	6.5	6.5	10			
Sky Exposure Plane	begins at:		60 ft			85 ft				
Tower Lot Coverage (min-max)			n/a				n/a- 40%	30- 40%	n/a- 40%	30- 40%
Dwelling Unit Factor	All		680							
Parking										
General (min % of d.u.)	for Group Parking Facilities		70%	60%	50%	40%				
Reduced and Waived Requirements (min % of d.u.)	IRHU (outside Transit Zone)		25%	15%		12%				
	AIRS (outside Transit Zone)		10%							
	Small Lot Area (sf)	10,000 or less	50%	30%	0%					
		10,000-15,000	n/a		30%	20%				
	If required small # spaces		5		15					

All R6-R10 Districts

Streetscape

Street Tree (min)	
All Contextual Districts / All Non-Contextual Districts	Shall be provided within a planting strip for every 25 ft of Street Frontage
Planting (min)	
All Contextual Districts / All Non-Contextual (QH Option) Districts	Area between street line and building street wall shall be planted at ground level or in raised planting beds

Streetscape

Street Wall Location Provision			
Contextual Districts			
R6B,R7B,R8B		Lot < 50 ft	no closer nor further
		Lot > 50 ft	no closer than adjacent street wall
R6A,R7A,R7D, R7X,R9D		All	
Non-Contextual Districts (QH Option)			
R6, R7	Narrow St.	Lot < 50 ft	no closer nor further
		Lot > 50 ft	no closer than adjacent street wall
	Wide St.	All	
R8,R9, R10	All within 50 ft of W. St.		70% of within 8ft
	N. St. beyond 50 ft W. St.		70% of within 15ft
Non-Contextual Districts (HF and Tower Options)			
n/a			

Non-contextual districts R6-10 following sky exposure plane regulatioins depend on street width. Smaller setback and higher slope mean bigger building.

#Initial Setback Distance# (in feet)		Maximum Height of a Front Wall or other portion of a #Building or Other Structure# within the #Initial Setback Distance#	#Sky Exposure Plane#				
			Height above #Street Line# (in feet)	Slope over #Zoning Lot# (expressed as a ratio of vertical distance to horizontal distance)			
				On #Narrow Street#		On #Wide Street#	
On #Narrow Street#	On #Wide Street#			Vertical Distance	Horizontal Distance	Vertical Distance	Horizontal Distance
R6 or R7 Districts							
20	15	60 feet or six #stories#, whichever is less	60	2.7	to 1	5.6	to 1
R8 R9 or R10 Districts							
20	15	85 feet or nine #stories#, whichever is less	85	2.7	to 1	5.6	to 1

In 1961, non-contextual districts (contextual districts didn't yet exist) depended on street width.

Summary of Bulk Regulations in Residence Districts (Continued)

(for Residential Buildings)

	Height and Setback Requirements													Minimum Dimensions of Courts				Minimum Distance between Windows and Walls or Lot Lines	
	Standard Regulations							Alternate Regulations						Minimum width of outer court		Minimum area and minimum dimension of inner court			
	Initial setback distance (in feet)		Maximum height of front wall or building within setback distance		Height above street line or front yard line (in feet)	Sky exposure plane: Slope over zoning lot (vertical distance to horizontal distance)		Depth of optional front open area (in feet)		Sky exposure plane: Height above street line or front yard line (in feet)	Sky exposure plane: Slope over zoning lot (vertical distance to horizontal distance)		Minimum Spacing between Any Two Buildings on Same Zoning Lot (in feet)						
District	Narrow street	Wide street	(in feet)	(in stories)		Narrow street	Wide street	Narrow street	Wide street		Narrow street	Wide street		If less than 30 feet wide	If more than 30 feet wide	Area (in square feet)	Dimension (in feet)	Between window and any wall or rear or side lot line (in feet)	Between window on inner court and court wall (in feet)
R1	None	None	Street level	Street level	25[6]	1 to 1	1 to 1	..	..	..	..	..	30 or formula[9]	2 x depth	1 x depth	1,200	30	30[11]	⅓ height of wall
R2	None	None	Street level	Street level	25[6]	1 to 1	1 to 1	..	..	..	..	..	30 or formula[9]	2 x depth	1 x depth	1,200	30	30[11]	⅓ height of wall
R3	None	None	Street level	Street level	25[6]	1 to 1	1 to 1	..	..	..	..	..	30 or formula[9]	2 x depth	1 x depth	1,200	30	30[11]	⅓ height of wall
R4	None	None	Street level	Street level	25[6]	1 to 1	1 to 1	..	..	..	..	..	30 or formula[9]	2 x depth	1 x depth	1,200	30	30[11]	⅓ height of wall
R5	None	None	Street level	Street level	35[6]	1 to 1	1 to 1	..	..	..	..	..	30 or formula[9]	2 x depth	1 x depth	1,200	30	30[11]	⅓ height of wall
R6	20	15	60	6	60[7]	2.7 to 1	5.6 to 1	15	10	60[7]	3.7 to 1	7.6 to 1	30 or formula[9]	2 x depth	1 x depth	1,200	30	30[11]	⅓ height of wall
R7	20	15	60	6	60[7]	2.7 to 1	5.6 to 1	15	10	60[7]	3.7 to 1	7.6 to 1	30 or formula[9]	2 x depth	1 x depth	1,200	30	30[11]	⅓ height of wall
R8	20	15	85	9	85[7]	2.7 to 1	5.6 to 1	15	10	85[7]	3.7 to 1	7.6 to 1	30 or formula[9] or [10]	2 x depth	1 x depth	1,200	30	30[11]	⅓ height of wall
R9	20	15	85	9	85[7]	2.7 to 1	5.6 to 1	15	10	85[7]	3.7 to 1	7.6 to 1	30 or formula[9] or [10]	2 x depth	1 x depth	1,200	30	30[11]	⅓ height of wall
R10	20	15	85	9	85[7]	2.7 to 1	5.6 to 1	15	10	85[7]	3.7 to 1	7.6 to 1	30 or formula[9] or [10]	2 x depth	1 x depth	1,200	30	30[11]	⅓ height of wall

C Data Construction

Although this appendix faithfully describes the methodology, some of the figures in this appendix have not been updated for several months. For the most accurate numbers, refer to those in the body of the paper.

C.1 Cleaning Names

The main property data I use is Primary Land Use Tax Lot Output (Pluto), available from NYC's open data platform. Pluto excludes some condo properties (building classes R3 and R6), so I supplement this with a distinct property tax roll dataset that contains much of the same information. The crucial field in both of these datasets is the name of the owner.

I drop properties that disappear and then reappear in the data after large gaps, are zoned as a park, have zero or more than three residential units, or have a non-private owner type. I remove from the owner string non-ASCII characters, as well as common substrings that could lead to a false match (e.g. "llc", "trustee", "corp", "jr", "junior", etc). When removing these common substrings, I am careful to only remove substrings that are enclosed by a space on either side or are at the beginning/end of the string and immediately followed/preceded by a space. I then split (by spaces) both the owner and the address into tokens, compare each token, and set the owner to missing if there is any match. This screens out corporations that are named for a single property that it owns. For additional certainty, I remove from the owner string common street name tokens (e.g., "street", "road") and their abbreviations ("st", "rd"). Now, the owner strings are ready for the first algorithm: detecting matches with other owner strings associated with the property in other years.

C.2 Matching Names Within Properties

For this task, we want to be more lenient with matches than if we were comparing names across the whole city. Names that have one token in common are more likely to be members of the same family, or a woman who married and changed her name, or a person with a typo in one year, than two strangers who share a first or last name. For every combination of two unique names that ever owned a given property, I compute two measures of similarity.

C.2.1 Levenshtein Token Sort Ratio

The first is the "token set ratio" variant of the Levenshtein difference (Levenshtein et al., 1966) implement by python library TheFuzz,[14] maintained by SeatGeek. The Levenshtein difference is a popular edit distance algorithm, essentially the minimum number of insertions, deletions, or substitutions necessary to transform one string into another. For example, $Lev(\text{seinfeld}, \text{senfeld}) = 1$ because we may delete one character from the left token or insert one in the right,

[14] `https://github.com/seatgeek/thefuzz`

while $Lev(\text{jerry}, \text{gary}) = 3$, since we must substitute "j" for "g" and "e" for "a" and insert or delete one "r". The difference is frequently calculated as a similarity ratio $sim(s_0, s_1) = 1 - (Lev(s_0, s_1)/min(length(s_0), length(s_1))$, *i.e.* one is identical, zero means no common characters, and it is normalized to the length of the shorter string. The token set ratio function in TheFuzz additionally does not penalize variation in the order of tokens ($sim(\text{jerry seinfeld}, \text{seinfeld jerry}) = 1$) or repetition of tokens ($sim(\text{jerry seinfeld}, \text{jerry jerry seinfeld}) = 1$) for multi-token strings. Considering that there is no consistent formatting of owner names in Pluto, this is a strength.

C.2.2 Custom Levenshtein-Damerau

The other similarity measure is a custom function based on the Levenshtein-Damerau (Damerau, 1964) distance. This distance is identical to the Levenshtein distance, except it counts transpositions of two characters as one operation. For example, while $Lev(\text{seinfeld}, \text{sienfeld}) = 2$, $LD(\text{seinfeld}, \text{sienfeld}) = 1$. For a given combination of names, I also generate combinations where one space from each is omitted, in order to successfully match cases like "lorenzo demedici" and "catherine de medici".

I take the L-D distance for each combination of tokens in the two strings (ignoring tokens with length less than two) and then compare it with a benchmark that is harsher for shorter strings and more lenient for longer strings, to permit mangled spellings of long uncommon names but exclude false matches between short, usually East Asian names (for example, Lin and Ling are more likely to be distinct names than a typo). Setting a benchmark similarity ratio is equivalent to a linear inequality constraint with a zero intercept (*e.g.*, taking only $sim(s_0, s_1) \geq 0.9$ implies $Lev(s_0, s_1) \leq 0.1 \times \ell$ where $\ell = min(length(s_0), length(s_1)))$, while this other benchmark is a linear inequality with a negative intercept: taking only $LD(s_0, s_1) \leq -(3/2) + (1/2)\ell$, where ℓ is defined as before. By construction, matches of tokens of length two or less are impossible, there is no tolerance for any error until lengths of five or more, and the particular slope was chosen after carefully inspecting the data.

Because names were truncated to 21 characters in earlier years, if an original unaltered name string is 21 characters long and the final token is at least five character long then I compare it with tokens in the other string that are truncated to the same length. For example, if we were comparing "johnathan alex bhattacharya", truncated to "johnathan alex bhatta", with "jane bhattacharya", then the DL distance between "bhatta" and "bhattacharya" would be six and we would fail to match. However, accounting for possible truncation, we instead compare "bhatta" with "bhatta" and obtain a match.

In the end, I take two names to be from the same household if at least one token matches in the manner described immediately above or if the Levenshtein token sort ratio is at least 0.75.

C.3 Aligning Owner Changes With Sales

Next, I reconcile the detected owner changes with sales microdata, which includes precise dates of sales. This is important because it will increase precision when detecting matching the seller of one property to the buyer of another property in a window of time around the sale. The combination of errors in detecting owner changes, errors in when owner changes resulting from sales appear in Pluto, and the existence of non-arms-length sales between members of the same household (that cannot always be detected with unreasonably low prices) complicate this endeavour.

When all sales (for at least $1,000) of a property are separated by at least three years, I deem these properties to be non-overlapping (in the sense that I can construct three year windows around each sale that do not overlap). When a property is non-overlapping, I deem apparent owner changes outside of these three year windows to be transfers within a household. This may occur if for example there is a married couple with different surnames. If there is a window around a sale with no apparent owner change, then I deem this to be a non-arms-length sale between members of the same household. Then, the number of sales and owner changes are the same and I assign the dates of the owner changes to be the same as the dates of the sales.

If a property is overlapping but there is the same number of owner changes and sales within the union of all the windows, then I assign the dates of the owner changes to be the same as the dates of the sales.

The situation is more complicated when the property is overlapping but there are different numbers of sales and owner changes within the union of the windows. I generate all possible combinations of sequences with sales or owner changes removed (such that there is the same number of each) and take the one that minimizes an objective function. That objective function is:

$$Obj = \sum_j Q_j^2 \text{ where } Q_j = \begin{cases} c_j - s_j & \text{if } c_j - s_j < 1 \\ c_j - s_j - \frac{1}{2} & \text{if } c_j - s_j = 1 \\ c_j - s_j - 1 & \text{if } c_j - s_j > 2 \end{cases}$$

and where c_j is the year of the jth owner change and s_j is the year of the jth sale. This objective function is informed by the distribution of year differences when a property has exactly one owner change and one sale. Most frequently, Pluto accurately reports the owner as of December 31st of each year, such that if there was a sale during the year then the seller appears as the owner the year prior and the buyer appears as the owner for the year of the sale, so I apply a penalty of 0^2 to that case. Second most frequently, the sale is reported the year after, so I apply a penalty of $\frac{1}{2}^2$ to that case. These two cases comprise 95.6% of all properties with exactly one owner change and one sale. The tails relative to these two central points are substantively the same on either side, and so a difference of 2 is treated the same as a difference of -1, etc. The lack of mass relative to these two central points motivates squaring the difference. I only execute this algorithm for properties where the number of owner changes is between one fewer than and two more than the number of sales. Properties where the difference is larger than this are only about 1% of the total to this point, and might inject too much noise into the data.

C.4 Identifying Movers

The foregoing work to match members of the same household has generated a set of names associated with each household. I check sellers in every year against buyers in every year, having not yet taken a stand on what time window to use. I execute the same fuzzy matching algorithms (TheFuzz's token set ratio, my custom Damerau-Levenshtein token matching algorithm) on all combinations of names. This time, since we are matching names of all participants in the NYC housing market, we want to be much more discriminating than in the case where we are matching names of people who have ever owned the same property. The basic eligibility requirements are now: (1) if one token matches then the similarity ratio must be at least 0.94 and each name must be no more than two tokens long, (2) if two tokens match then the similarity ratio must be at least 0.9 and one of the two names must be two tokens long, (3) if three tokens match then the similarity ratio must be at least 0.9, and (4) any two names that have four tokens matching is a match.

However, there are many non-unique matches in this preliminary set of matches. That is, a seller may be linked to multiple buyers, and a buyer may be linked to multiple sellers. Furthermore, any number of these multiple buyers and sellers may be linked to multiple other sellers and buyers, creating a complex graph structure. I resolve this problem by drawing on matching market algorithms from microeconomic theory.

We can think of buyers and sellers as opposite sides of a matching market (*e.g.*, the marriage market). We can define preferences by specifying a function of the characteristics of the match. In this case, I use

$$u_{sb} = 100(sim(n_s, n_b) - 0.9) + 5 \times \mathbb{1}(tm(n_s, n_b) \geq 2) - 5\left|y_s - y_b + \frac{1}{2}\right| - \mathbb{1}(t_s \neq t_b) + \epsilon_{sb}.$$

That is, I take the sum of the percentage that the similarity ratio exceeds 90% by, five if at least two tokens match, -5 times the absolute value of the difference in years (with a small intential bias informed by the data), and -1 if the number of tokens in each name does not match exactly. The tradeoff between similarity and token matching is informed by careful inspection of the data. The bias in the difference of the years pushes the optimal point to between buying in the same year and one year prior, which I chose because usually people buy and then sell but also a difference of zero is the most frequent in the data. The penalty for the number of tokens is intended only to break ties so that “Adam Smith” is matched with “Adam Smith” over “Adam Quincy Smith”.

There is a random error $\epsilon_{sb} \sim U[0, 1]$ to ensure that “preferences” are strict, since this is required by matching models. Due to the scale of preferences and the error, the error only serves to break ties. Any combination of names that does not meet the preliminary eligibility requirements specified above is taken to be strictly dominated by being alone for each seller and buyer.

Then, I execute a deferred-acceptance algorithm (Gale and Shapley, 1962) twice, with sellers "proposing" and buyers "accepting", and then the reverse. I find that each algorithm generates the same result (which we do not expect in general from deferred-acceptance algorithms), which is probably related to the symmetry of preferences ($u_{sb} = u_{bs}$). This is a pleasant surprise, in addition to the expected stability (no match that is better for both the seller and buyer).

In the end, I am able to identify 53,000 sellers as buyers of other properties. Of these, 17,500 transactions occurred within one year of each other, while in 32,000 cases the the purchase occurred within four years of each other. In order to understand what this says about the size of the population of movers in NYC, we should compare this to the population of 388,000 uncensored sales, of which 53,000, 32,000, and 17,500 are 13.7%, 8.2%, and 4.5% respectively. In order to understand the implications for estimation, we should compare this to the population of all 427,000 sales (which includes sales where we do not observe one party because the sale occurred close to the endpoints of the sample period), of which 53,000, 32,000, and 17,500 are 12.4%, 7.5%, and 4.1% respectively. No comparable figure is directly presented in the original BMMT.[15]

The histogram of the difference in years is given in Figure 8, where we can see that after a difference of four years the change in the size of the bars becomes small. I present a cross-tabulation of the window with the quality of the match in Table 8. I define a match to be "perfect" if the similarity ratio is equal to one and the number of matching tokens is equal to the minimum number of tokens across the two names. In particular, I present the percent perfect matches for each window value. This definition of the window is marginal rather than cumulative, that is $window = 1$ corresponds to a difference of -1 or 1 but not 0. From this table, we see that the the largest dropoff in the quality of matches occurs when considering a window of 1 versus 0, at about 7%. When widening the window to two, there is a decline of 2.5%, with a further decline 4% when choosing the widest possible window. Alone, this evidence provides conflicting results about how best to identify movers.

[15] By closely reading we can make a guess. In the 2016 published version, footnote 14: "We can locate the seller's previous purchase in the data in 30% of cases". In the 2011 working paper, page 7: "[Merging to HMDA] allows us merge in information about sellers for approximately 35-40 percent of the sample". This implies that they were able to locate the seller's *next* purchase in 5-10% of cases. The 2011 working paper had 800,000 transactions with buyer information due to HMDA; if we take this as "the sample" then 5-10% is 40,000-80,000. However, we also need to take into account that BMMT only looked in a window of one year around a sale to match a buyer's name, suggesting that the quality of these 40,000-80,000 is high. The likely reasons why my figures are lower despite a more lenient matching procedure are (1) more homes being owned by landlords due to greater implausibility of homeownership for a family in 2003-2019 New York City versus 1994-2004 Bay Area, and (2) the limited geographical extent of New York City compared to the Bay Area excluding from the data moves of young homeowners from the city to the suburbs for childrearing reasons.

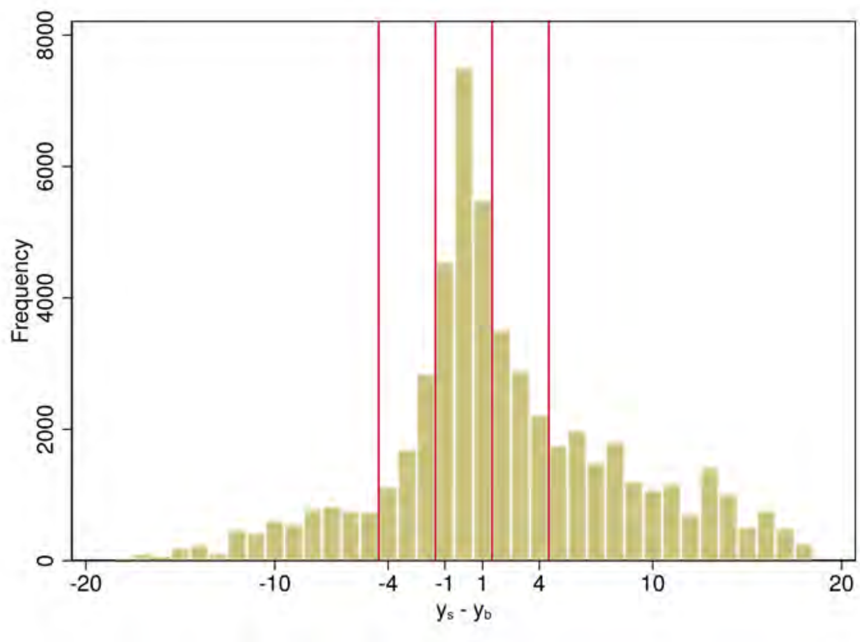

Figure 8: Histogram of difference between years of purchase and sale for detected movers

	window=0		window=1		window=2		window=4		window=10		window=18	
	num	pct	num	pct	num	pct	num	pct	num	pct	num	pct
perfect=0	1271	16.96	2417	24.15	1680	26.48	2223	28.17	4142	30.72	2372	30.43
perfect=1	6223	83.04	7590	75.85	4664	73.52	5669	71.83	9340	69.28	5423	69.57
Total	7494	100	10007	100	6344	100	7892	100	13482	100	7795	100

Table 8: Tabulation of mover quality by window (marginal, not cumulative)