IDENTIFICATION IN A MODEL OF SORTING WITH SOCIAL EXTERNALITIES AND THE CAUSES OF URBAN SEGREGATION SUPPLEMENTARY APPENDIX

Maximilian Kasy

This supplementary appendix contains additional results, complementing those discussed in "Identification in a model of sorting with social externalities". Section 1 discusses the relationship between demand- and hedonic slopes and preferences. Section 2 presents a dynamic model of the local housing market. Section 3, finally, presents results which decompose linear IV coefficients as weighted averages of structurals slopes, with identified weights. This allows us to relate the observable data distribution to equilibrium comparative statics.

1. PRICE SLOPES AND PREFERENCES

This section discusses the relationship between household preferences and demand slopes as well as hedonic price gradients. We shall maintain the following assumptions: Households are characterized by the triple $(u(X, M, P), u^o, c)$, where u is their continuously differentiable indirect utility dependent on neighborhood characteristics, u^o is the utility of their best outside option and c is their type. Households locate in the given neighborhood iff $u(X, M, P) \geq u^o$. The outside utility u^o is exogenously determined, i.e., constant in (X, M, P). There is a continuum of households of total mass M^{tot} in the economy. The vector (u, u_X, u_M, u_P, u^o), evaluated at any (X, M, P), has a continuous joint distribution. D^c is the mass of households that want to locate in the given neighborhood,

$$D^c = M^{tot} \cdot \mathbb{P}(u \geq u^o, c).$$

Similarly $E = M^{tot} \cdot \mathbb{P}(u \geq u^o)$.

Since Rosen (1974), price slopes of the form P^*_X have often been used as estimates of household willingness to pay for X, which equals $-u_X/u_P$ in the notation used here. In the context of discrete choice models, it becomes evident that price slopes are not necessarily equal to willingness to pay for infra-marginal households. Proposition 1 and its corollary 1 show this in the present, nonparametric, setup. They represent the price gradient as an appropriately weighted average of willingness to pay of marginal households. If, however, there is a continuum of similar outside options, all households in the neighborhood are marginal and have identical willingness to pay, as illustrated by proposition 2.

Due to the possible presence of social externalities, $u_M \neq 0$, price slopes may also deviate from willingness to pay for X for marginal households. Price changes

Assistant professor, Department of Economics, Harvard University, and junior associate faculty, IHS Vienna. Address: Littauer Center 200, 1805 Cambridge Street, Cambridge, MA 02138. E-Mail: maximiliankasy@fas.harvard.edu.

P_X^* must compensate for the change in composition, M_X^*. This is made apparent by defining a notion of *counterfactual partial equilibrium*, as a point of reference. Counterfactual partial equilibrium gives the price P^+ that would prevail if M were determined exogenously and need not equal D. In proposition 1, the comparison of P_X^* and P_X^+ shows the bias in P_X^* relative to average marginal willingness to pay for X, P_X^+, due to externalities. This also suggests P_X^+ and P_M^+ as empirical objects of interest in their own right. In the main paper, partial sorting equilibrium was defined as the solution to equating total housing demand and supply as well as composition and type specific demand. Counterfactual equilibrium $(M^+(M,X),\ P^+(M,X))$ is defined as the solution to equating housing demand and supply, while the argument M to the demand functions is exogenously given and not necessarily equal to D:

Definition 1 (Counterfactual Partial Equilibrium) *A counterfactual partial equilibrium (M^+, P^+) given X and M solves the $C+1$ equations*

$$M^+ = D(X, M, P^+) \tag{1}$$

$$S(P^+, X) = \sum_c M^{+c}. \tag{2}$$

Let $(M^+(X,M), P^+(X,M))$ denote the function mapping (X,M) into the counterfactual partial equilibrium given (X,M).

Proposition 1 (Price gradients and marginal households' utility) *Under assumptions maintained in this section, the slope of total housing demand with respect to X is given by*

$$E_X = M^{tot} \cdot f^{u-u^o}(0) \cdot E[u_X | u = u^o],$$

where f^{u-u^o} denotes the density of $u - u^o$. Similarly for E_M, E_P, D_X^c, D_M^c, and D_P^c. Assume additionally that partial sorting equilibrium is unique or assume (M^*, P^*) is a differentiable selection from the set of partial equilibria, and assume $S_P = S_X = 0$. Then

$$P_X^+ = -\frac{E[u_X | u = u^o]}{E[u_P | u = u^o]},$$

$$P_M^+ = -\frac{E[u_M | u = u^o]}{E[u_P | u = u^o]},$$

and

$$P_X^* = P_X^+ + P_M^+ M_X^* = -\frac{E[u_X + u_M M_X^* | u = u^o]}{E[u_P | u = u^o]}.$$

Proposition 1 expresses the price gradients as ratios of average marginal utilities among marginal households. Since one can rewrite any ratio of averages as weighted average of ratios, the following corollary expresses the gradients as average willingness to pay for X, where the average is taken with respect to a reweighted distribution. The reweighting can be interpreted as a re-normalization of household utility to a constant marginal disutility of P, which implies a rescaling of the conditional density of marginal utilities among marginal households.

Corollary 1 (Price gradient as weighted average willingness to pay) *Under the assumptions of proposition 1, if $u_P < 0$ for all households,*

$$P_X^+ = \widetilde{E}\left[-\frac{u_X}{u_P}\middle| u = u^o\right],$$

where the expectation $\widetilde{E}$ is taken with respect to the density

$$f^{u_X, u_P | u - u^0}(u_X, u_P | 0) \cdot \frac{u_P}{E[u_P | u = u^0]}.$$

Similarly for P_M^+ and P_X^*.

If, relative to proposition 1, we assume additionally that there is a continuum of alternative location choices, as in hedonic models, tighter characterizations of equilibrium prices and sorting follow. All households in a neighborhood become marginal and have the same marginal willingness to pay.

Proposition 2 (Hedonic gradient given continuum of outside options) *In the setup considered in this section, assume additionally that u^o is bounded by the supremum of $u(X, M, P)$ over a set of outside options including an ϵ ball around X, and the corresponding equilibria $(M, P) \in (M^*(x), P^*(x))$:*

$$u^o \geq \sup_{\substack{x: ||x - X|| < \epsilon \\ (M,P) \in (M*(x), P^*(x))}} u(x, M, P)$$

Then

$$P_X^* = -\frac{u_X + u_M M_X^*}{u_P} \tag{3}$$

for all households choosing a neighborhood with given X.

This subsection concludes with a proposition characterizing local comparative statics of average reservation prices among households in the neighborhood in terms of average marginal willingness to pay of *all* households in the neighborhood. This proposition will be central in the characterization of prices in the dynamic model presented later. In this dynamic model, landowners will extract all surplus value generated by a match to a tenant, so that rents are equal to household reservation prices. Formally, define the reservation price of a household for living in the neighborhood, given X and M, as

$$P^{res} = P^{res}(X, M) := \sup\{P : u(X, M, P) \geq u^o\}.$$

Proposition 3 characterizes the dependence on X of average reservation prices, *conditional* on locating in the neighborhood, i.e., conditional on $P^{res} \geq P^*$. Changes in X can, in principle, influence average reservation prices in three ways: directly, through their effect on M, and through a reshuffling of residents. The last may matter if, under the new X, households with higher reservation prices crowd out the initial residents. The central message of proposition 3 is that this effect is not of first-order importance if housing supply is inelastic, so that the number of households in the neighborhood is constant, or if housing demand is elastic, so that all households in the neighborhood have reservation prices equal to P^*.

Proposition 3 (Comparative statics of average reservation prices) *Assume partial sorting equilibrium is unique or assume (M^*, P^*) is a differentiable selection from the set of partial equilibria. Under the assumptions maintained in this section*

$$\frac{\partial}{\partial X} E\left[P^{res} | P^{res} \geq P^*\right] = E\left[P_X^{res} + P_M^{res} M_X^* | P^{res} \geq P^*\right] - \left(\frac{\partial}{\partial X} \log \mathbb{P}\left(P^{res} \geq P^*\right)\right) \cdot \left(E\left[P^{res} | P^{res} \geq P^*\right] - P^*\right), \tag{4}$$

where $P_X^{res} = -u_X/u_P$ and $P_M^{res} = -u_M/u_P$. In particular, if housing supply is price inelastic and constant in X, i.e., $S_P = S_X = 0$, or if all households in the neighborhood are marginal, i.e., $E\left[P^{res} | P^{res} \geq P^*\right] = P^*$, then

$$\frac{\partial}{\partial X} E\left[P^{res} | P^{res} \geq P^*\right] = E\left[P_X^{res} + P_M^{res} M_X^* | P^{res} \geq P^*\right]. \tag{5}$$

2. A DYNAMIC EXTENSION OF THE STATIC MODEL WITH SEARCH FRICTIONS

The model discussed in the main paper is static. We can think of it as describing an economy with negligible search frictions in which equilibrium is instantaneously achieved. Alternatively, it could be considered as describing the long run steady state of an economy with frictions. However, explicitly considering dynamics and frictions reveals additional sources of identification.

A well established literature in labor economics discusses the dynamics and comparative statics of unemployment and wages in models with search frictions. Its central presumption is that finding a job or an employee takes time and unemployment is due to this search time. Pissarides (2000) provides an extensive overview of this literature. Wheaton (1990) applies the insights of this literature to the housing market. The focus of either of these is the relationship between vacancies (unemployment) and prices. Wheaton (1990) in particular models housing vacancies as corresponding to the search time of households who decided to move due to lifecylce events (shocks), found another place and now attempt to sell their old home. The present section extends the basic sorting model of the main paper using similar techniques as these papers.

Relative to the static model, the main extensions in the dynamic model are as follows. There is an explicit, continuous time dimension, and exogenous location characteristics X can change over time. Households that would like to move to a different neighborhood are subject to search frictions. If they decide to search for a new home, offers arrive at Poisson rate λ. Similarly, owners of vacant units have to search for tenants and find them at rate μ. Households are maximizing expected discounted utility, and make their search decisions in a forward looking way. Due to search frictions, composition M changes continuously over time and only reacts with delay to shocks in X. Finally, once a match is formed between homeowner and household, they are in a situation of bilateral monopoly: By breaking the match they both would have to search again, and thereby incur a loss of utility. Therefore, they have to negotiate over the division of the surplus, and rents are match-specific.

The purpose of this extension is twofold. First, the delayed adjustment of composition M to changes in exogenous characteristics X generates independent variation between X and M, contrary to the static case where M is essentially a function of X. This allows, under certain conditions, to separately identify household willingness to pay for X and for M. Second, the dynamic structure provides a connection between multiplicity of equilibria in the static sense, and multiplicity of equilibria in a dynamic sense. A test for the latter will be constructed below.

This section presents a search model of the *rental* market for housing. Considering homeownership would add the additional complication of housing being an asset in addition to being a consumption good. Under complete financial markets, the results derived for the rental market of housing immediately extend to the more general case however, as will be discussed briefly at the end of this section.

For simplicity of notation, household and time superscripts are mostly dropped. As before, we consider one fixed neighborhood.

Assumption 1 (The local economy, dynamic setup)

- *There are $\mathscr{C}$ types of households, $c = 1, \ldots, \mathscr{C}$.*
- *Households can be in one of four states: Living in the neighborhood and not searching, living in the neighborhood and searching for a place outside, living outside and searching for a place in the neighborhood, living outside and not searching for a place in the neighborhood.*
- *Housing units can be in one of two states, vacant or occupied by one house-*

hold.

- *A neighborhood, at each point of time $t \in \mathbb{R}$, is characterized by*
  1. *the number of households of each type living in the neighborhood, $M = (M^1, \dots, M^{\mathscr{C}})$*
  2. *an exogenous vector X of* all *other location characteristics and factors influencing demand or supply.*
- *The time paths of X and M are piecewise differentiable.*
- *There is a match specific rental price P for each match between a unit and a household.*
- *Households living outside searching for a place in the neighborhood, or living in the neighborhood and searching for a place outside, find a match at rate λ. Vacancies are matched to a household at rate μ. These rates can vary over time but are constant across households and units.*
- *Vacant units and searching households are matched uniformly at random.*

Assumption 2 (The household probprop)

- *Households are characterized by their type c, their flow utility $u(X, M, P)$ of living in the given neighborhood, their discount rate r and the value of their outside option V^o. Except for type, all of these may depend on time t. V^o does not depend on X, M.*
- *Households have the choice between searching or not. They do so to maximize their expected discounted utility.*
- *There are no costs of search.*
- *There is a continuum of households of total mass M^{tot} in the economy.*

Denote the value of living in the given neighborhood by $V = \max(V^s, V^{ns})$, where V^s and V^{ns} are the values of searching and not searching, respectively. Denote the time derivative of V by $\dot{V}$. The value functions are to be understood as conditional expectations, given the information set at time t, as are their time derivatives. Assumptions 1 and 2 imply

$$rV^s = u(X, M, P) + \lambda(V^o - V) + \dot{V} \tag{6}$$

and

$$rV^{ns} = u(X, M, P) + \dot{V}. \tag{7}$$

A household living in the neighborhood wants to search for a place outside if and only if $V^o > V$, and V satisfies

$$(r + \lambda)V = u(X, M, P) + \lambda \max(V^o, V) + \dot{V}. \tag{8}$$

Let us now turn to the landowners.

Assumption 3 (The landowner's probprop)

- *Landowners are risk neutral, maximize their discounted stream of incomes and are otherwise indifferent about the residents of their units. Their discount rate is denoted by r.*
- *Owners of vacant units can and do search for renters among the pool of households that search for a home in the given neighborhood.*

Denote the value of an occupied unit by $W = \max(W^s, W^{ns})$ where W^s and W^{ns} are the values of the unit when the renting household is searching and not searching, respectively. Denote the value of a vacant unit by W^v. Under assumptions 1 and 3, the value of an occupied unit where the renter is not searching for a new place is characterized by

$$rW^{ns} = P + \dot{W}. \tag{9}$$

The value of an occupied unit with a searching renter is

$$rW^s = P + \lambda(W^v - W) + \dot{W}. \tag{10}$$

The value of a vacant unit satisfies, finally,

$$rW^v = \mu(W^{new} - W^v) + \dot{W}^v. \tag{11}$$

Note that the value of a match to the landowner is household specific, and therefore W^{new}, the expected value of a match with a new renter, is in general different from the value of the current match, W. These values describe the expected discounted revenue for a given unit.

Once a potential renter and a landowner holding a vacant unit meet, they have to negotiate a rental contract.

Assumption 4 (Rent determination)

- *The contract specifies rental payments. Contracts can be continuously renegotiated.*
- *Each of the contract parties can unilaterally decide at any time to end the contract and initiate search of the renting household, where this decision is reversible. The renter can not be evicted before she has found a new place, but can be committed to search.*
- *Rents are determined by Nash bargaining over the division of the surplus relative to the outside option of searching (not searching)*[1]*, that is, current rents maximize $(V^{ns} - V^s)^\beta (W^{ns} - W^s)^{(1-\beta)}$, where $\beta \in [0, 1]$ is the relative bargaining power of tenants.*

[1]Note that the potential outside option of searching for a different home in the same neighborhood is always strictly dominated. It leaves the household indifferent and the landowner strictly worse off, since she foregos rents while searching for a new tenant.

We have $(V^{ns} - V^s)^\beta (W^{ns} - W^s)^{(1-\beta)} = \lambda (V - V^o)^\beta (W - W^v)^{(1-\beta)}$, and the first order condition for the solution to Nash bargaining is

$$(V - V^o) = \beta \left[-u_P (W - W^v) + (V - V^o) \right]. \tag{12}$$

If there exists a price P such that both $V > V^o$ and $W > W^v$, then these conditions must hold under the bargaining solution, no matter what β is. Search happens if and only if there is no such P, and the search decision is always consensual. This is a feature of any privately efficient contract. For households living outside the neighborhood, the decision to search in this given neighborhood is independent of β. Our last assumption pins down bargaining power β:

Assumption 5 (Bargaining power) *All bargaining power lies with the landowners, i.e.,* $\beta = 0$.

This assumption allows for a clean characterization of price dynamics, since all changes in household utility will be compensated by price changes. If we were to drop assumption 5, only a fraction $1-\beta$ of utility changes would be compensated by price changes.

Readers familiar with the literature on search models of unemployment will notice a central feature of these models missing in the assumptions just stated. Neither λ and μ, nor housing supply, are explicitly modeled. Common practice in the literature, for instance in the models reviewed in Pissarides (2000), is to assume a matching technology where the rates λ and μ are a function of the ratio of searching workers (households) to vacancies, and there is free entry of firms (landlords). This is crucial in the context of search models of the labor market that attempt to explain unemployment and vacancy rates. It also has important implications for the speed of adjustment following shocks. As neither vacancies nor variation in the speed of adjustment are of central concern in the present context, exposition is simplified by not explicitly modeling intertemporal variation in λ and μ.

2.1. *Implications of the model*

The rest of this section develops some central properties of the model described by assumptions 1 through 5. First, the dynamics of composition are shown to follow a differential equation of the form $\dot{M} = \lambda \cdot (D - M)$, where D is a dynamic generalization of the demand function. If we specialize this to the two type case and consider discrete intervals of time, then changes in composition m follow the difference equation $\Delta m := m^1 - m^0 = \kappa \cdot (d - m^0)$, where κ is a rate derived from λ.

Next, the reaction of prices to shocks in X is studied. In the short run, because of search frictions, M is not affected by such shocks. Under assumptions 4 and 5, rents immediately adjust so that all surplus of the match is appropriated by landlords, and changes in rents correspond to household willingness to pay for changes in X. In the long run, M does change however. Rental changes occurring with delay correspond to household willingness to pay for this change in M.

Finally, the relationship between this dynamic model and the static model studied so far is clarified. First, the long run comparative statics of M, as a function of X, are the same as those of an appropriately defined corresponding static model. Second, if search frictions are low and/or discount rates high, then the dynamic model is approximated by the static model in a sense made precise below.

The dynamics of composition

Under assumptions 1 and 2 we have, at each point in time, a set of households of each type c that want to move out of the neighborhood, because for them $V^o > V$, and a corresponding outflow. Similarly, at each point in time there is a set of households of each type c that want to move into the neighborhood, because for them $V \geq V^o$, and a corresponding inflow. The net flow will equal λ times the difference between the number of households that *want* to live in the neighborhood, i.e., for which $V \geq V^o$, and those that *do* live in the neighborhood, M. This motivates the following definition of demand D in the dynamic model.

Definition 2 (Demand in the dynamic model) *Denote by D^c the mass of households of type c for which $V \geq V^o$:*

$$D^c = M^{tot} \cdot \mathbb{P}(V \geq V^o, c)$$

Proposition 4 (Dynamics of composition) *Make assumptions 1, 2, 3 and 4. Then*

$$\dot{M} = \lambda \cdot (D - M), \tag{13}$$

where $\dot{M}$ is the expected time derivative of M.

The following result specializes to a two-type model and describes changes of composition m over discrete time intervals.

Proposition 5 (Dynamics of composition in the two-type model) *Make the assumptions of proposition 4, as well as the assumption that there are only two types of households and that $d = d(m, X)$. Then the change in m from time 0 to time 1 is given by*

$$\Delta m := m^1 - m^0 = \kappa \cdot (d(m, X) - m^0) \tag{14}$$

for some m, X at a time in $[0, 1]$, where

$$\kappa = 1 - \exp\left(- \int_0^1 \lambda \cdot \frac{D^1 + D^2}{M^1 + M^2} ds \right) > 0. \tag{15}$$

The determination of prices

Under assumption 5, the landowner appropriates all surplus from the match, and equation 12 implies that the participation constraint for the renter is binding at all times, i.e., $V = V^o$. By equation 7, the household specific rental price is then determined by

$$u(X, M, P) = rV^o - \dot{V}^o, \tag{16}$$

where X, M, and V^o are given to the household and landowner. This implies in turn that changes in rents must directly reflect willingness to pay for changes in X and M, for any household that lives in the neighborhood. This is reflected in the following propositions 6 and 7. Proposition 6 additionally uses the fact that composition M is constant in the short run.

Proposition 6 (Short run comparative statics of prices) *Make assumptions 1, 2, 3, 4, specifying the dynamic model, and 5 on bargaining power. Assume that $X = x$ before time 0, $X = x + \xi$ for a jump ξ after time 0, and (u, V^o) is continuously differentiable with respect to time for all households. Then $\frac{\partial}{\partial \xi} \lim_{t \to 0^+} E[P] = E\left[-\frac{u_X}{u_P}\right]$ where the expectation is taken over the set of households living in the neighborhood at time $t = 0$.*

We recover short-run comparative statics of prices in response to changes in X and M which look similar to the ones in the static model in the absence of social externalities and with inelastic housing supply. In the static model, the neighborhood rental gradient P_X^+ equals the average marginal willingness to pay of marginal households, according to corollary 1, whereas here the match specific rent gradient P_X equals the marginal willingness to pay of any given household. In the static case, marginal households had to be kept indifferent by changes in X and P^* for demand to be constant. In the present case, all households have to be kept indifferent by changes in X and P, since by the assumption on bargaining power all households are marginal, in the sense that their utility is equal to their reservation utility.

As households re-sort across neighborhoods, however, prices adjust further for two reasons. First, holding outside options as well as X and M constant, some households will want to move in which have a willingness to pay for the given bundle (X, M) which is higher than the willingness to pay of the current residents. Second, composition M will adjust over time, and influence the households' valuation of the neighborhood. If housing supply is constant or all households are marginal, the first reason can be ignored to first order, however. This follows from proposition 3. As a consequence, long run effects of changes in X on rents P reflect the sum of the willingness to pay for X and of the willingness to pay for the change in M induced by X.

Proposition 7 (Long run comparative statics of prices) *Make assumptions*

1, 2, 3, 4, specifying the dynamic model, and 5 on bargaining power. Assume that housing supply is constant or all households are marginal.
Assume that $X = x$ before time 0, $X = x + \xi$ for a jump ξ after time 0, and (u, V^o) is constant for all households. Denote $M^{lr} = lim_{t\to\infty} M$, where it is assumed that this limit exists.
Then $\frac{\partial}{\partial \xi} \lim_{t\to\infty} E[P] = E\left[-\frac{u_X + u_M M_\xi^{lr}}{u_P}\right]$, where the expectation is taken over the set of households living in the neighborhood at time $t = 0$.

Demand in the dynamic and the static model
D, as given by definition 2, equals the number of households for which $V \geq V^o$. In the static model, under the assumption of household utility maximization, D was equal to the number of households for which $u \geq u^o$. How do these notions of demand relate to each other? To connect the dynamic model to our discussion of the static model, the following definition is useful. It derives a static model from the given dynamic model. Equilibrium prices in the static model correspond to cut-off prices, below which landlords do not accept a tenant in the dynamic model in steady state. The utility of households' outside option, u^o, is implicitly given by V^o. Corresponding static demand, finally, is equal to the mass of households for which flow utility u is bigger than outside utility u^o. As shown in proposition 8, the static model defined in this way describes the long run comparative statics of composition in the dynamic model. Proposition 9 implies that it also approximates the short run behavior of the dynamic model in the case of low search frictions or high discount rates.

Definition 3 (The corresponding static model) *Under assumptions 1, 2, 3, 4 and 5, the corresponding static model is defined as follows: Let $u^o := rV^o - \dot{V}^o$. Let $P^{res} = \max\{P : u(X, M, P) \geq u^o\}$ be the reservation price for each household. Let P^* be the "cut-off" price below which landowners in the dynamic model would not accept a tenant in steady state. As will be shown, this cut-off is given by*

$$P^* = \frac{r\mu}{r + \mu} E^s[P^{res} | P^{res} > P^*], \tag{17}$$

where the expectation E^s is taken over the set of households searching for a place in the neighborhood. This equation implicitly defines the corresponding static housing supply.
The corresponding static demand of type c is equal to

$$\widetilde{D}^c = M^{tot} \cdot \mathbb{P}(u(X, M, P^*) > u^o | C = c) = M^{tot} \cdot \mathbb{P}(P^{res} > P^* | C = c) \tag{18}$$

for all c. Let $\widetilde{E} = \sum_c \widetilde{D}^c$. Denote the equilibrium (set) of this corresponding static model by (M^*, P^*).

Proposition 8 (Long run comparative statics of composition) *Make assumptions 1, 2, 3, 4, specifying the dynamic model, and 5 on bargaining power. Assume that λ is uniformly bounded away from zero for $t > 0$, and that X and (u, u^o) is constant over time for all households.*
If $M^{lr} := \lim_{t\to\infty} M$ exists, then $M^{lr} \in M^*$, i.e., composition converges to an equilibrium composition of the corresponding static model.

Proposition 9 (Low-friction limit of the dynamic model) *Make assumptions 1, 2, 3 and 4. Define u^o as $u^o = (r+\lambda)V^o - \dot{V}^o - \lambda \max(V, V^o)$. Assume u and u^o are continuous in time and bounded. Then, for V, V^o, u and u^o evaluated at time t^0,*

$$\lim \frac{V - V^o}{\int_{t^0}^{\infty} e^{-\int_{t^0}^{t}(r+\lambda)ds} dt} = u - u^o$$

as $r + \lambda \to \infty$ uniformly in a neighborhood of t^0, if $r + \lambda$ remains bounded away from 0 uniformly on $[t^0, \infty)$.

Proposition 9 says that, if discount rates are large or search frictions low, then relative values are approximately equal to relative flow utilities. Similarly, if u and u^o are constant over time, relative values equal relative flow utilities. If either of these is the case uniformly across households, then dynamic demand D is approximately equal to demand in the corresponding static model $\widetilde{D}$. It is in this sense that the static model can be regarded as an approximation to the dynamic model in the cases of either "myopic" behavior (high discount rates), low search frictions (high λ) or steady state (constant u).

Home ownership

So far we have been discussing the market for housing rentals. What about home ownership? Under an assumption of perfect financial markets a no-arbitrage condition between either renting and holding financial assets or home-ownership must hold. In particular, we could extend the above model assuming that at each point in time a landowner can decide to sell her unit to the tenant or to another potential landowner, if the latter agrees. The price for such a (potential) sale into ownership is $\mathbf{P}$. Agreement on such a sale requires that each party is indifferent between holding financial assets and home ownership. Such indifference implies the asset equation

$$r\mathbf{P} = P + \dot{\mathbf{P}}, \tag{19}$$

where r is now a market rate of return. Tenant households could at the same time be landowners, for instance for units previously occupied. The interest rate r implicitly already incorporates a risk premium and a compensation for depreciation.

The focus of the present paper is identification of the determinants of the consumption value u. This consumption value conceptually maps more closely to rental prices P rather than home values $\mathbf{P}$, since decisions about homeownership reflect both consumption and investment considerations. This explains our focus on rental prices.

3. DECOMPOSITION REPRESENTATIONS OF LINEAR IV COEFFICIENTS

In this section, a series of representations of linear IV coefficients in terms of weighted average slopes is developed. These results resemble closely the LATE representations introduced by Imbens and Angrist (1994). The distinguishing feature of the results presented here is that all weights are defined in terms of observable and identifiable quantities, as opposed to first stage structural slopes (in the binary case, compliance versus noncompliance). This allows to describe the distribution of any observable covariates for the population over which structural slopes are averaged to obtain the linear IV coefficients. In the terminology of Imbens and Angrist (1994), we don't know who the compliers are but we do know how they look like. Results similar in spirit were used by Kling (2001).

The first set of results is stated for a random coefficient, cross-sectional setup. These results suggest to plot densities of covariates with respect to a reweighted distribution, and to plot conditional IV coefficients in the case of linear IV with controls. Then, these results are generalized to the fully non-parametric panel difference case, which is the setup relevant for the present paper.

Proposition 10 (Crossectional IV, random coefficient case) *Assume that*

$$Y^i = \alpha^i + \beta^i X^i \tag{20}$$

and assume $Cov(Z, \alpha) = 0$. Then

$$\beta^{IV} = \frac{Cov(Y, Z)}{Cov(X, Z)} = E\left[\beta^i \cdot \omega\right]$$

for a weighting function

$$\omega = \frac{X(Z - E[Z])}{E[X(Z - E[Z])]}.$$

Proposition 11 (Crossectional OLS with controls, random coefficient case) *Assume that*

$$Y^i = X^{1,i}\beta^{1,i} + X^{2,i}\beta^{2,i} + \epsilon \tag{21}$$

for a scalar X^1 and a vector X^2. Assume $X^1 \perp (\beta, \epsilon)|X^2$, and $E[X^{2,i}\beta^{2,i} + \epsilon|X^2]$ is linear in X^2. Then the coefficient on X^1 in OLS regression of Y on X is in *expectation equal to*

$$\beta^{1,OLS} = E\left[E[\beta^{1,i}|X^2]\frac{E[X^1 e|X^2]}{E[X^1 e]}\right],$$

where e is the residual from OLS regression of X^1 *on* X^2*.*

Proposition 12 (Crossectional IV with controls, random coefficient case) *Assume that*

$$Y^i = X^{1,i}\beta^{1,i} + X^{2,i}\beta^{2,i} + \epsilon \tag{22}$$

for a scalar X^1 *and a vector* X^2*. Assume* $Z \perp (\beta^2, \epsilon)|X^2$ *for a scalar instrument* Z*, and* $E[X^{2,i}\beta^{2,i} + \epsilon|X^2]$ *is linear in* X^2*. Denote by* e *the residual of OLS regression of* Z *on* X^2*.*

Then the coefficient on X^1 *in IV regression of* Y *on* X*, instrumented by* (Z, X^2)*, is in expectation equal to*

$$\beta^{1,IV} = \frac{E[Ye]}{E[X^1 e]} = E\left[\frac{E[\beta^{1,i}X^1 e|X^2]}{E[X^1 e]}\right] = E\left[\beta^{1,i}\cdot\omega\right]$$

for a weighting function

$$\omega = \frac{X^1 e}{E[X^1 e]}.$$

These propositions give a LATE representation of IV coefficients. In the setup of proposition 12, the following two exercises seem instructive:

Suggestion 1: Plot the distribution of covariates (in particular of components of X^2), reweighted by ω. In the terminology of Imbens and Angrist (1994), this gives the distribution of covariates for the set of compliers.

Suggestion 2: Calculate conditional IV given (components of) X^2: Let $\widehat{E}$ denote some flexible ("nonparametric") estimator of the conditional expectation. For components of X^2, plot (nonparametric) regressions of

$$\widehat{\beta}^{IV}(X^2) := \frac{\widehat{E}\left[Ye|X^2\right]}{\widehat{E}\left[X^1 e|X^2\right]}$$

on these components. The estimator $\widehat{\beta}^{IV}(X^2)$ converges to a conditional weighted average of the structural slope β^2,

$$E\left[\beta^1\frac{X^1 e}{E[X^1 e|X^2]}\middle| X^2\right].$$

In, practice, however, such estimates of $\beta^{IV}(X^2)$ might be poorly behaved. If the denominator, $\widehat{E}\left[X^1 e|X^2\right]$, has positive mass around 0, then $\widehat{\beta}^{IV}(X^2)$ might not have a finite expectation. In that case, it can still be insightful to plot the "conditional reduced form" estimator $\widehat{E}\left[Ye|X^2\right]$.

The following propositions extend the previous results to the panel-difference case.

Proposition 13 (Panel difference IV, random coefficient case) *Assume that*

$$Y^{it} = \alpha^{it} + \beta^{it} X^{it} \tag{23}$$

for $t \in \{0,1\}$*, and assume* $\Delta Z \perp (\Delta\alpha + \Delta\beta \cdot X^{i,1})$.[2] *Then*

$$\beta^{IV,\Delta} := \frac{Cov(\Delta Y, \Delta Z)}{Cov(\Delta X, \Delta Z)} = E\left[\beta^{i,0} \cdot \omega\right]$$

for a weighting function

$$\omega = \frac{\Delta X(\Delta Z - E[\Delta Z])}{E[\Delta X(\Delta Z - E[\Delta Z])]}.$$

Proposition 14 (Panel difference IV, nonparametric case) *Assume that*

$$Y^{it} = g(X^{it}, \epsilon^{it}) \tag{24}$$

for $t \in [0,1]$*, and assume*

$$\Delta Z \perp \int_0^1 g_\epsilon(X^{it}, \epsilon^{it}) \cdot \epsilon_t \, dt.$$

Then

$$\beta^{IV,\Delta} := \frac{Cov(\Delta Y, \Delta Z)}{Cov(\Delta X, \Delta Z)} = E\left[g_X \cdot \omega\right]$$

for a weighting function

$$\omega = \frac{X_t(\Delta Z - E[\Delta Z])}{E[X_t(\Delta Z - E[\Delta Z])]}.$$

All expectations here are taken over the product distribution of the crosssectional distribution over the i *and the uniform distribution over the time interval* $[0,1]$.

Proposition 15 (Panel difference IV, nonparametric case, if exclusion is violated) *Assume that*

$$Y^{it} = g(X^{1,it}, X^{2,it}, \epsilon^{it}) \tag{25}$$

[2] That is, ΔZ is uncorrelated with the counterfactual change in Y which would have occurred if X had stayed constant.

for $t \in [0,1]$*, and assume*

$$\Delta Z \perp \int_0^1 g_\epsilon(X^{it}, \epsilon^{it}) \cdot \epsilon_t \, dt.$$

Then

$$\beta^{Y1,IV,\Delta} := \frac{Cov(\Delta Y, \Delta Z)}{Cov(\Delta X^1, \Delta Z)} = E\left[g_{X^1} \cdot \omega^1\right] + R$$

for weighting functions ($j = 1, 2$*)*

$$\omega^j = \frac{X_t^j(\Delta Z - E[\Delta Z])}{E[X_t^j(\Delta Z - E[\Delta Z])]}$$

and an error term

$$R = E\left[g_{X^2} \cdot \frac{X_t^2}{X_t^1} \cdot \omega\right] = E[g_{X^2}\omega^2] \cdot \beta^{21,IV,\Delta} = E[g_{X^2}\omega^2] \cdot \frac{Cov(\Delta X^2, \Delta Z)}{Cov(\Delta X^1, \Delta Z)}.$$

All expectations here are taken over the product distribution of the crosssectional distribution over the i *and the uniform distribution over the time interval* $[0,1]$*.*

Suggestion 3: Bound the error term by making a-priori assumptions giving bounds on $E[g_{X^2}\omega^2]$. Estimate $\beta^{21,IV,\Delta} = Cov(\Delta X^2, \Delta Z)/Cov(\Delta X^1, \Delta Z)$ directly from the data.

This appendix concludes with a characterization of cross-sectional linear IV in a triangular system, where the weights in this proposition are now expressed in terms of first stage structural slopes.

Proposition 16 (Cross-sectional linear IV in nonparametric triangular systems) *Consider the triangular system* $Y = g(X, \epsilon)$*,* $X = h(Z, \eta)$*,* $Z \perp (\epsilon, \eta)$*, where all variables are continuously distributed and* g, h *are continuously differentiable. Then*

$$\beta^{IV} = \frac{Cov(Y, Z)}{Cov(X, Z)} = E\left[g_x(X, \epsilon)\omega(Z, \eta)\right]$$

for a weighting function ω *which is given, up to normalization, by*

$$\omega(z, \eta) = const. \cdot \frac{h_z(z, \eta)}{f(z)} \cdot (E[Z|Z > z] - E[Z|Z \leq z])\, \mathbb{P}(Z > z)\mathbb{P}(Z \leq z).$$

The constant is such that $E[\omega] = 1$*.*

3.1. *Application*

The formal results just stated show that linear instrumental variables estimates describe the local average treatment effect (LATE) for the subpopulations for which the instruments do affect the treatment. What are the characteristics of these subpopulations of neighborhoods for our instruments? The decomposition results of this section can be used to shed some light on this question. In particular, the IV coefficient controlling for covariates can be decomposed as a weighted average of structural slopes over the sampling population, where the weights ω are identifiable and are given by

$$\omega = \frac{\Delta m \cdot e}{E[\Delta m \cdot e]}.$$

In this expression, Δm is the regressor of interest and e is the residual of a regression of the instrument on the controls.

We shall now apply this to the data analyzed in the main paper. Figures 1 through 3 show the unweighted density of the initial Hispanic share across neighborhoods for the sample used, as well as this density reweighted by ω, for weights ω corresponding to the various instruments. They furthermore show plots of estimates of the "conditional first stage" and "conditional reduced form," $E[\omega|m]$ and $E[\Delta Y \cdot e|m]$, where ΔY corresponds to the change of various outcomes of interest. The plots of the reweighted densities are particularly instructive. They show that the specification using the subgroup instrument estimates a LATE for neighborhoods with a medium to high initial Hispanic share, using the spatial instrument yields a LATE for neighborhoods with lower Hispanic shares (although still upweighting higher shares relative to the population), and the dynamic instrument estimates a LATE for neighborhoods somewhere in between. The conditional expectation estimates for higher values of m should be interpreted with caution, as they are quite imprecisely estimated due to limited support of Hispanic share in the right tail.

The graphs of the conditional reduced form of price responses, $E[\Delta P \cdot e|m]$ for the spatial and dynamic instrument, when compared to the conditional first stage, $E[\omega|m]$, are somewhat worrisome. They suggest significant variation of the conditional IV coefficient given m over the range of m. This does not imply invalidity of the instrument, but it cautions to be careful when extrapolating the willingness-to-pay results to different populations.

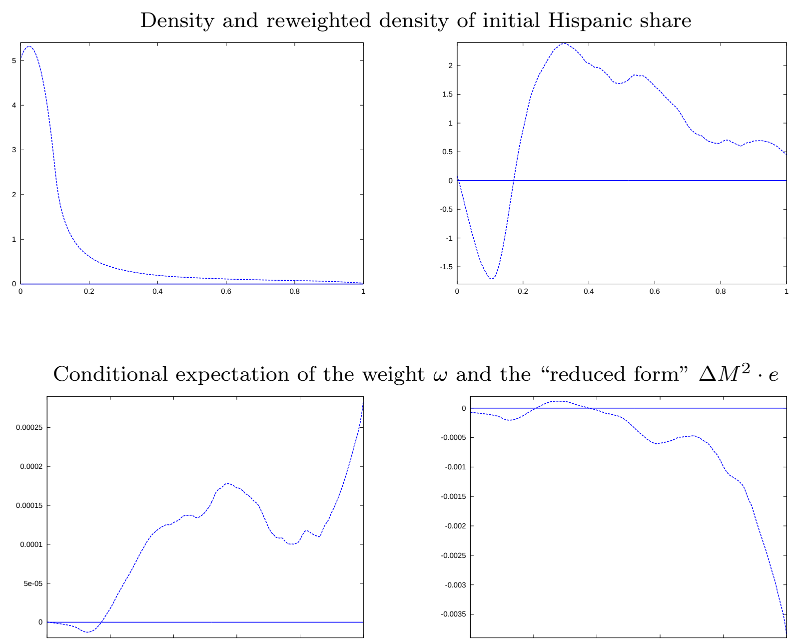
APPENDIX A: PROOFS

Section 1

Proof of proposition 1 :
Plugging 1 into 2 and differentiating w.r.t. X gives

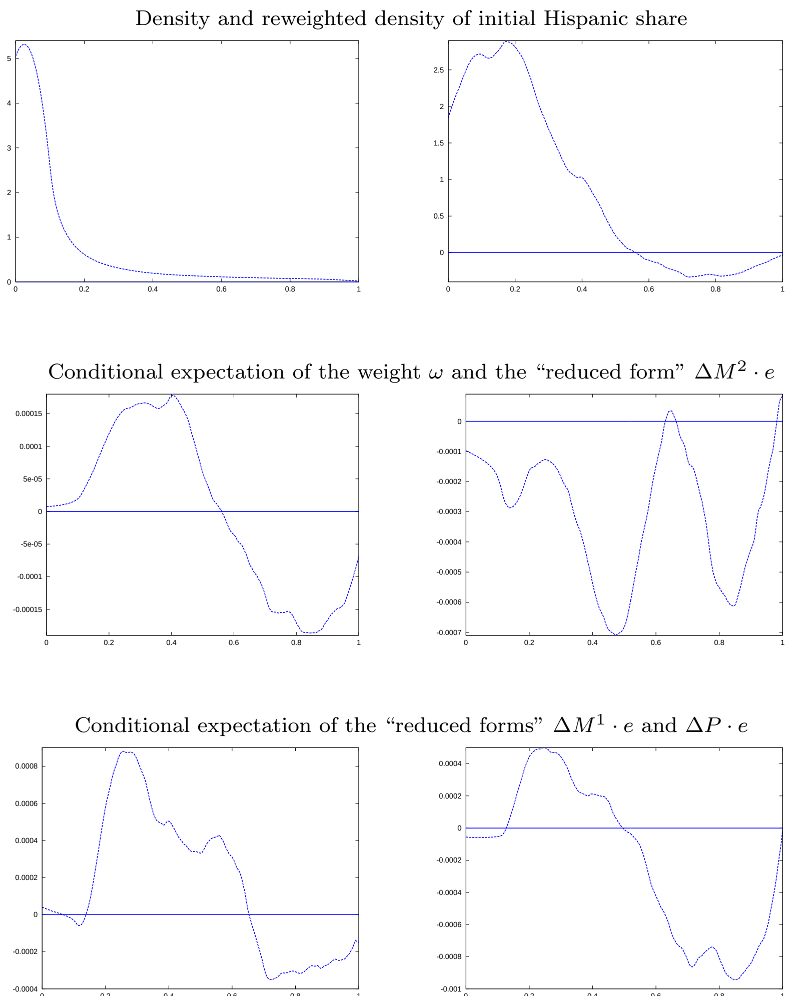
$$S_X + S_P P_X^+ = \sum_c M_X^{+c} = \sum_c (D_X^c + D_P^c P_X^+) = E_X + E_P P_X^+.$$

Figure 1.— Decomposition of the subgroup instrumental variable estimate



Notes: These graphs decompose the IV estimate based the subgroup instrument shown in table **??**, according to proposition 12. The top row shows a kernel estimate of the density of initial Hispanic share across neighborhoods in the sample, as well as this density reweighted to give the distribution among the population for which the LATE is estimated. The bottom row shows kernel estimates of the conditional expectation of the weight ω, as well as the "conditional reduced form", $\Delta M^2 \cdot e$.

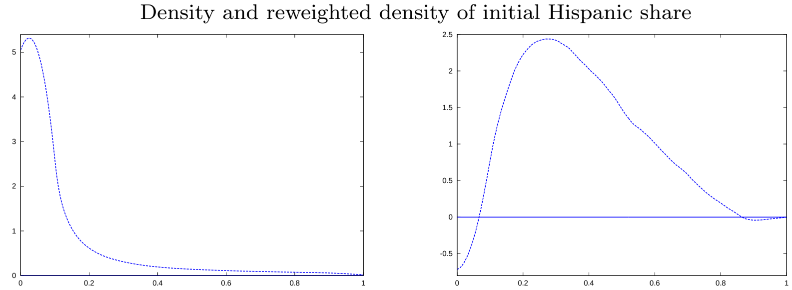
FIGURE 2.— DECOMPOSITION OF THE SPATIAL INSTRUMENTAL VARIABLE ESTIMATE



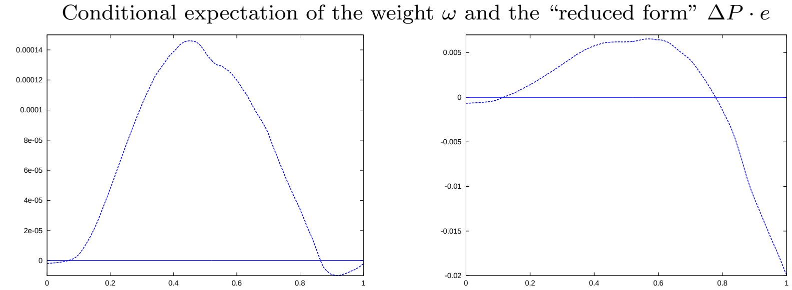
Notes: These graphs replicate those of figure 1 for the spatial instrument, and display conditional reduced forms for the additional outcome variables M^1 and P.

FIGURE 3.— DECOMPOSITION OF THE DYNAMIC INSTRUMENTAL VARIABLE ESTIMATE

Density and reweighted density of initial Hispanic share



Conditional expectation of the weight ω and the "reduced form" $\Delta P \cdot e$



Notes: These graphs replicate those of figure 1 for the dynamic instrument, where the conditional reduced form is for the outcome variable P.

Inelastic supply $S_P = 0$ and constancy $S_X = 0$ imply

$$P_X^+ = -\frac{E_X}{E_P}.$$

Analogously,

$$P_M^+ = -\frac{E_M}{E_P}$$

and

$$P_X^* = -\frac{E_X + E_M M_X^*}{E_P}.$$

By the assumption of household utility maximization and iterated expectations, we can write $E = M^{tot} \cdot E[\mathbb{P}(u \geq u^o | u_X)]$. Denote $f^{u-u^o|u_X}$ the conditional density of $u - u^o$ given u_X. We get

$$\frac{1}{M^{tot}} E_X = E\left[\frac{\partial}{\partial X}\mathbb{P}(u - u^o \geq 0 | u_X)\right] = E\left[u_X f^{u-u^o|u_X}(0|u_X)\right]$$

$$= \int u_X \frac{f^{u-u^o,u_X}(0, u_X)}{f(u_X)} f(u_X) du_X = f^{u-u^o}(0) E[u_X | u = u^o].$$

Similarly for E_M and E_P and for D. □

Proof of Corollary 1:
Immediate from proposition 1, once we check that this density integrates to one and is non-negative. □

Proof of proposition 2:
$u(X, M^*(X), P^*(X)) \geq u^o$ implies the first order condition

$$u_X + u_m M_X^* + u_P P_X^* = 0$$

for all the households choosing the given neighborhood. □

Proof of proposition 3:
For simplicity of notation, the superscript *res* will be omitted from reservation prices in this proof. Furthermore, assume for the moment that there are no social externalities, i.e., $u_M = 0$. The general case is completely analogous. By iterated expectations we can write

$$E[P | P \geq P^*] = E[E[P \cdot \mathbf{1}(P \geq P^*) | P_X]] / \mathbb{P}(P \geq P^*).$$

In integral notation, the conditional expectation is given by

$$E[P \cdot \mathbf{1}(P \geq P^*) | P_X] = \int_{P^*}^{\infty} P f(P | P_X) dP.$$

Differentiating this conditional expectation gives

$$\frac{\partial}{\partial X} E[P \cdot \mathbf{1}(P \geq P^*) | P_X] = P_X \cdot \mathbb{P}(P \geq P^* | P_X) + P^* \cdot (P_X - P_X^*) \cdot f^{P-P^*|P_X}(0 | P_X).$$

The second term is due to the change in the boundaries of integration. Hence

$$\frac{\partial}{\partial X} E[P \cdot \mathbf{1}(P \geq P^*)] = E[P_X \mathbf{1}(P \geq P^*)] + P^* \cdot E[P_X - P_X^* | P = P^*] \cdot f^{P-P^*}(0).$$

Similarly

$$\frac{\partial}{\partial X}\mathbb{P}(P \geq P^*) = E[P_X - P_X^* | P = P^*] \cdot f^{P-P^*}(0).$$

Collecting terms then gives

$$\begin{aligned}
\frac{\partial}{\partial X} E[P|P \geq P^*] &= \left(E[P_X \mathbf{1}(P \geq P^*)] + P^* \frac{\partial}{\partial X} \mathbb{P}(P \geq P^*) \right) / \mathbb{P}(P \geq P^*) \\
&- E[P|P \geq P^*] \frac{\partial}{\partial X} \mathbb{P}(P \geq P^*) / \mathbb{P}(P \geq P^*) \\
&= E[P_X | P \geq P^*] \\
&- \left(\frac{\partial}{\partial X} \log \mathbb{P}\left(P \geq P^*\right) \right) \cdot \left[E\left[P^{res} | P \geq P^* \right] - P^* \right].
\end{aligned}$$

Finally, inelastic housing supply implies that, in equilibrium, the number of households must be constant, i.e., $\mathbb{P}\left(P \geq P^*\right)$ does not depend on X. □

Section 2

Proof of Proposition 4:
We can divide households of type c into four classes, depending on whether or not they live in the neighborhood (indexed by 1 and o) and depending on whether the want to stay (s) or to move (m) into or out of the neighborhood. Denote these classes by $D^{c,1,s}, \ldots, D^{c,o,m}$. By definition $M^c = D^{c,1,s} + D^{c,1,m}$ and $D^c = D^{c,1,s} + D^{c,o,m}$. A fraction λ of those that want to move will be successful per time unit, giving

$$\begin{aligned}
\dot{M}^c &= \lambda \left(D^{c,o,m} - D^{c,1,m} \right) = \lambda \left((D^{c,1,s} + D^{c,o,m}) - (D^{c,1,s} + D^{c,1,m}) \right) \\
&= \lambda \left(D^c - M^c \right).
\end{aligned}$$

□

Proof of Proposition 5:
Recalling the definitions $m = M^1/(M^1 + M^2)$ and $d = D^1/(D^1 + D^2)$, and using the result of the previous proposition,

$$\dot{m} = \frac{\partial m}{\partial M} \dot{M} = \lambda \cdot \frac{\partial m}{\partial M} \cdot (D - M) = \check{\lambda} \cdot (d - m),$$

where

$$\check{\lambda} := \lambda \cdot \frac{\frac{\partial m}{\partial M} \cdot (D - M)}{d - m} = \lambda \cdot \frac{D^1 + D^2}{M^1 + M^2}.$$

The second equality in this expression follows from

$$\frac{\frac{\partial m}{\partial M} \cdot (D - M)}{d - m} = \frac{\frac{1}{(M^1+M^2)^2} \left(M^2, -M^1 \right) \cdot \left(D^1 - M^1, D^2 - M^2 \right)'}{\frac{D^1}{D^1+D^2} - \frac{M^1}{M^1-M^2}} = \frac{D^1 + D^2}{M^1 + M^2}.$$

By assumption **??**, the price and scale elasticities of both types are identical and hence $d = d(X, m)$. Therefore $\dot{m} = \check{\lambda} \cdot (d(X, m) - m)$.
Taking the time path of d and $\check{\lambda}$ as given, the solution to this differential equation can be written as

$$m^t = m^0 e^{-\int_0^t \check{\lambda} ds} + \int_0^t \check{\lambda} d e^{-\int_s^t \check{\lambda} du} ds.$$

This gives m^t as a weighted average of initial m^0 and d in the time interval from 0 to t. Letting $\kappa = 1 - e^{-\int_0^1 \check{\lambda} ds}$ and (m, X) some appropriate intermediate values in the time interval $[0, 1]$ the claim follows. □

Proof of Proposition 6:
From equation 16 it is immediate that, for any given household, $P_{X,m} = \frac{(u_X, u_M)}{-u_P}$. By assumption, due to search frictions, M has a smooth time path and in particular $\frac{\partial}{\partial \xi} \lim_{t \to 0^+} M = 0$. □

Proof of Proposition 7: For any given household, it can be shown as in proposition 6 that $P_{X,m} = -\frac{u_X + u_M M^{lr}_{\xi}}{u_P}$. To prove the claim we have to show, that resorting of households according to willingness to pay has no first order effect on the average reservation price within the neighborhood. But this follows immediately from proposition 3. □

Proof of Proposition 8:
First, $M \in M^*$ are the only constant solutions of the differential equation 13: Any stable solution must imply $M = D$. By constancy of X, M and u, u^o, $\dot{V} = 0$ and $V = u/r$, $V^o = u^o/r$. Hence $V > V^o$ if and only if $u > u^o$, and D is equal to demand $\widetilde{D}$ in the corresponding static model. A landlord accepts a tenant if and only if W for this tenant is greater than W^v, i.e., if

$$P = rW \geq rW^v = \frac{r\mu}{r+\mu} E\left[P^{new}\right].$$

By random matching $E\left[P^{new}\right] = E^s[P^{res}|P^{res} > P^*]$, and hence D equals to demand $\widetilde{D}$ of the corresponding static model.
The claim follows, since any limit of a converging path must satisfy $\dot{M} = 0$. □

Proof of Proposition 9:
Let w.l.o.g. $t^0 = 0$. If we denote $V^{\max} = \max(V^o, V)$ and impose a transversality condition, we can solve equation 8 for V and get

$$V = \int_0^\infty e^{-\int_0^t (r+\lambda) ds} \left[u(X, M, P) + \lambda V^{\max}\right] dt. \tag{26}$$

This is again to be understood as a conditional expectation given the information set at time 0. A similar equation holds for V^o.
Equation 26 implies

$$V - V^o = \int_0^\infty e^{-\int_0^t (r+\lambda) ds} \left[u - u^o\right] dt$$

and hence

$$\frac{V - V^o}{\int_0^\infty e^{-\int_0^t (r+\lambda) ds} dt} = \int_0^\infty w^t \left[u - u^o\right] dt, \tag{27}$$

where we denote

$$w^t := \frac{\phi^t}{\int_0^\infty \phi^t dt}$$

for $\phi^t = e^{-\int_0^t (r+\lambda) ds}$ The weights w^t integrate to one. Let $\epsilon > 0$ be such that $r + \lambda > C^1$ and $\left|(u^{1,t} - u^{2,t}) - (u^{1,0} - u^{2,0})\right| < \delta$ on the interval $[0, \epsilon]$, and assume $r + \lambda > C^2$ on $[0, \infty)$. We get

$$\left|(u^{1,0} - u^{2,0}) - \int_0^\infty w^t \left[u - u^o\right] dt\right| < C^3 \delta + (1 - C^3) \sup_t (u - u^o) \tag{28}$$

for $C^3 = \int_0^\epsilon w^t dt$. Some algebraic manipulation yields

$$C^3 = \frac{1}{1 + \frac{\phi^\epsilon}{\int_0^\epsilon \phi^t dt} \frac{\int_\epsilon^\infty \phi^t dt}{\phi^\epsilon}}.$$

By $r + \lambda > C^1$ on $[0, \epsilon]$

$$\int_0^\epsilon \phi^t dt > \phi^\epsilon \int_0^\epsilon e^{C^1 [\epsilon - t]} dt = \frac{\phi^\epsilon}{C^1} \left[e^{C^1 \epsilon} - 1\right].$$

By $r+\lambda > C^2$ on $[\epsilon, \infty)$

$$\int_\epsilon^\infty \phi^t dt < \phi^\epsilon \int_\epsilon^\infty e^{-C^2[t-\epsilon]} dt = \frac{\phi^\epsilon}{C^2}.$$

Hence, as $C^1\epsilon \to \infty$

$$C^3 > \frac{1}{1+\frac{C^1C^2}{e^{C^1\epsilon}-1}} \to 1.$$

The claim now follows from equation 28. □

Section 3

Proof of proposition 10:
Since $Cov(\alpha^i, Z) = 0$, we have $Cov(Y, Z) = Cov(\beta^i X, Z) = E[\beta^i X(Z - E[Z])]$ □

Proof of proposition 11:
By the Frisch-Waugh theorem, $\beta^{1,OLS} = \frac{E[Ye]}{E[X^1e]}$, where e is the residual from OLS regression of X^1 on X^2. By linearity of $E[X^{2,i}\beta^{2,i} + \epsilon|X^2]$ and independence of $\beta^{2,i}, \epsilon^i$ and $(X^{1,i}, X^{2,i})$, $E[Ye] = E[\beta^{1,i}X^{1,i}e]$. By independence of $\beta^{1,i}$ and $(X^{1,i}, X^{2,i})$, $E[\beta^{1,i}X^{1,i}e|X^{2,i}] = E[\beta^{1,i}|X^2]E[X^1e|X^2]$. The claim then follows from iterated expectations. □

Proof of proposition 12:
$\beta^{1,IV} = \frac{E[Ye]}{E[X^1e]}$ follows again from the Frisch-Waugh theorem, applied to the two-stage least-squares representation of $\beta^{1,IV}$, and $E[Ye] = E[\beta^{1,i}X^{1,i}e]$ from linearity of $E[X^{2,i}\beta^{2,i} + \epsilon|X^2]$ and conditional independence $Z \perp (\beta^2, \epsilon)|X^2$. □

Proof of proposition 13:
Immediate from proposition 10, with differences replacing levels. □

Proof of proposition 14:
Under appropriate smoothness assumptions, we can write

$$\Delta Y = \int_0^1 (g_X(X^{it}, \epsilon^{it}) \cdot X_t + g_\epsilon(X^{it}, \epsilon^{it}) \cdot \epsilon_t)\, dt.$$

By exogeneity of the instrument, we then get

$$Cov(\Delta Y, \Delta Z) = E\left[\int_0^1 g_X(X^{it}, \epsilon^{it}) \cdot X_t\, dt(\Delta Z - E[\Delta Z])\right] == E\left[g_X \cdot \omega\right].$$

□

Proof of proposition 15:
This is an immediate extension of proposition 14. □

Proof of proposition 16:
First, consider the covariance of X and Z. Denote $\mu(Z) := E[X|Z] = E[h(Z, \eta)|Z]$. Then

$$\begin{aligned}
Cov(X, Z) = E[\mu(Z)(Z - E[Z])] &= \int_{-\infty}^{\infty}\int_{-\infty}^{z} \mu_z(\widetilde{z})(z - E[Z])f(z)d\widetilde{z}dz = \\
&= \int_{-\infty}^{\infty} \mu_z(\widetilde{z}) \int_{\widetilde{z}}^{\infty} (z - E[Z])f(z)dzd\widetilde{z} \\
&= \int_{-\infty}^{\infty} \mu_z(\widetilde{z}) \left(E[Z|Z > z] - E[Z|Z \le z]\right) \mathbb{P}(Z > z)\mathbb{P}(Z \le z)d\widetilde{z}dz = \\
&= E\left[h_z(Z, \eta)\widetilde{\omega}(Z)\right],
\end{aligned}$$

where

$$\widetilde{\omega}(z) := \frac{1}{f(z)} \left(E[Z|Z > z] - E[Z|Z \leq z]\right) \mathbb{P}(Z > z)\mathbb{P}(Z \leq z).$$

Similarly,

$$Cov(Y, Z) = E\left[g_x(X, \epsilon) h_z(Z, \eta) \widetilde{\omega}(Z)\right].$$

The assertion follows from $\beta^{IV} = \frac{Cov(Y,Z)}{Cov(X,Z)}$. $\square$

REFERENCES

Imbens, G. and J. Angrist (1994): "Identification and estimation of local average treatment effects," *Econometrica*, 62, 467–475.

Kling, J. (2001): "Interpreting instrumental variables estimates of the returns to schooling," *Journal of Business and Economic Statistics*, 19, 358–364.

Pissarides, C. (2000): *Equilibrium unemployment theory*, The MIT press.

Rosen, S. (1974): "Hedonic prices and implicit markets: product differentiation in pure competition," *The Journal of Political Economy*, 82, 34–55.

Wheaton, W. (1990): "Vacancy, search, and prices in a housing market matching model," *The Journal of Political Economy*, 98, 1270–1292.