

Discrete choice models with capacity constraints: an empirical analysis of the housing market of the greater Paris region

André de Palma* Nathalie Picard† Paul Waddell‡

April 5, 2006

Abstract

Discrete choice models are based on the idea that each user can choose both freely and independently from other users in a given set of alternatives. But this is not the case in several situations. In particular, limitations and interactions can occur when the number of available products of one type is smaller than the total demand for this type. As a consequence, some individuals can be denied their preferred choice. We develop a methodology to address those constraints and we apply it to residential location choice, where our empirical data suggest that availability constraints may bias actual choices. The analysis provides some theoretical developments and elaborates an iterative procedure for estimating demand in the presence of capacity constraints. The empirical application relies on the location choice model developed and estimated in [5] for Ile de France (Paris region) and generalizes it to integrate capacity constraints.

Keywords: Residential location, constrained Logit, capacity constraints, sampling, price endogeneity, Ile de France.

JEL: C35, R21, R31

*University of Cergy-Pontoise, THEMA and Ecole Nationale des Ponts et Chaussée. Member of the Institut universitaire de France.

†University of Cergy-Pontoise, THEMA.

‡University of Washington, Evans School of Public Affairs, corresponding author.

1 Introduction

The choices that individual households make in the housing market produce aggregate outcomes that shape urban traffic conditions, patterns of poverty concentration and ethnic segregation, the quality of public schools, access to economic opportunities, and the decline and revitalization of neighborhoods – all of which are important, long-term, and related policy concerns. Efforts to model individual household residential location choices using discrete choice frameworks date at least to the pathbreaking work by Lerman [8], McFadden [9], Quigley [10], Weisbrod [11], and Williams [12], among others. Over the past 25 years an extensive literature has examined household location choices in the housing market, significantly advancing the behavioral underpinnings and the methods used, and including efforts to represent the residential location choice as a dynamic process: see, e.g., Anas [1].

One important issue that has not yet received sufficient attention in the literature, and that is the central focus of the empirical analysis presented in this paper, is the role of availability constraints in households' choices. In estimating location choice models by observing agents' choices among a set of alternatives, an implicit assumption is made that the alternatives are all available, as they would commonly be if the choice were made among standard (durable or non durable) commodities which are not in short supply. However, in the housing market and in many other market situations, such as air travel, limited availability is not at all uncommon. For example, a particular neighborhood may be highly desired, while few vacancies may be available to those searching in this area. Seats may be sold out on the particular flight itinerary a traveler wishes to choose. A standard assumption in economics is that prices adjust and clear the market, and therefore putting prices on the right hand side of the model is sufficient to address this concern. However, casual observation supported by data suggest that this assumption may be too strong in many market conditions, including the housing market. Various forms of friction make the housing market less than perfectly efficient. High transactions costs, attachments to social networks, non-trivial search costs, asymmetric information, low turnover in some locations, and sellers' willingness to withhold a property from the market rather than suffer a loss, among other factors, suggest that prices may not fully clear the market.

Whether prices actually do fully clear the housing market should be an empirical question rather than a strong assumption. If the assumption that prices clear the market is not valid, then it follows that coefficients estimated for discrete choice models in markets that experience some level of availability constraints will be biased, confounding the effect of the constraints with the agents' preferences. An important policy implication of this methodological concern is that if these constraint effects are not corrected for in estimation of a choice model, predicted shifts in demand in response to an exogenous change (such as the change in accessibility due to major transportation investments) would also be biased, leading to potentially misleading conclusions regarding the relative costs and benefits of alternative policy choices. A related concern is that housing prices are jointly determined with location choices in constrained housing markets due to the role of prices in clearing the market, even if they do not completely clear it. It is therefore necessary to address this source of endogeneity bias in the estimation process (see dePalma [5] or Guevara and Ben-Akiva [7]).

In this paper, we are interested in the allocation process when supply is smaller than demand in some alternatives. Such an allocation has been described as a search mechanism by dePalma and Lefevre [4], later on applied in the regulated housing market in Holland by dePalma and Rouwendal [6]. Here we adopt a completely different approach to address the problem of disequilibrium between supply and demand. We assume that the final allocation mechanism should obey some simple rules (or axioms). From these assumptions, we are able to compute the ex post allocation. That is, we only describe the ex ante choice (i.e. the choice which ignores the capacity constraints) and the ex post allocation (i.e. the individual choice once the competition for scarce housing resources has taken place). Note that in this approach, we do not specify the complex dynamics which characterizes the adjustment process from ex ante to ex post allocation. Below we describe the organization of the paper.

In this paper, we develop models of residential location and housing price for the Ile-de-France metro-

politan region centered on Paris, testing for endogeneity between prices and location choices. We develop and apply an empirical estimation procedure that accounts for the effects of constraints on the availability of some alternatives. Following a brief description of the study area and the data used in the empirical work, we develop the model specifications and estimation algorithm. We then present the estimation results and an analysis of the sensitivity of the algorithm. The paper concludes with an assessment of the contribution of this research and proposed extensions of it.

2 The study context

The Ile-de-France is one of the most important economic regions in Europe. It houses 11 million people (2 million in Paris) and 5.1 million jobs within a space of 12,000 sq. km. Careful preservation of the historic central city of Paris and its many landmarks has had the effect of making Paris a major tourist mecca for the globe, largely by restricting supply of new housing and office space in the historic core, and protecting historic buildings from alteration or replacement. As a consequence, real estate prices are high in the center, and lower-income housing is concentrated in the several suburbs and new towns, mainly (but not only) in the eastern portion of the region. This is an important motivation for the focus of this paper, since the desirability of central Paris, coupled with strong supply constraints, provides a prime example of the kind of constrained market we wish to analyze. Most of the growth in population and employment are outside the core of the city, fueling rapid suburbanization and growth in travel. The regional express roadway network contains 4,500 lane-km, and despite the traditional rush-hour traffic jams, the average duration of a car trip is still only 19 minutes (EGT, 2001). With a ubiquitous metro system in central Paris and half of households in the city not owning a car, transit ridership is quite high. Half of commuting trips in the region in 2001 were by private cars, with 36% using public transportation and 14% using a bicycle or walking, though the transit mode share has declined 6% over the past twenty years as the region has decentralized. An efficient radial public transport system (RER) and freeways connect Paris to the suburbs.

Table 1 shows important differences in average housing prices by district: prices are higher in Paris, intermediate in the close suburbs and decline in the more distant suburbs, which we refer to as the inner ring and outer ring, respectively. In addition, prices are higher in the western part of the area than in the eastern part, consistent with the spatial patterns of social stratification. Note that the heterogeneity of prices is probably underestimated when using such average prices computed at the commune level, because they do not correct for the characteristics of the dwellings, which are typically large houses in the West suburbs and small flats inside Paris. This may explain why the largest average price is in district 92, and is nearly twice the largest average price inside Paris.

Table (1): Prices by district (in year 1998).

Sub-region	District	Average	Standard Deviation	Minimum	Maximum
Paris	75	294,500	165,241	83,939	694,375
Close Suburbs (Inner Ring)	92 (West)	247,556	205,038	66,966	1,198,950
	93 (North)	115,709	49,055	47,876	259,163
	94 (South)	144,098	74,603	53,356	373,499
Far away suburbs (Outer Ring)	78 (West)	135,122	65,714	38,112	373,815
	91 (South)	114,826	46,740	24,719	332,338
	95 (North)	104,375	41,670	25,154	241,692
	77 (East)	91,539	37,220	18,028	253,827

Source: Author's computations from notaries' database.

The geographic units used in this analysis follow the administrative boundaries used in France. The

smallest administrative unit, and the one we use as the basis for the residential location choice model is the commune, of which there are 1300 in the Ile-de-France. These roughly correspond to small cities, in terms of having local administrative control of land use decisions. Note that local administrative control operates at two different levels within Paris: each of the 20 Arrondissements has his own mayor, who has less administrative control than the mayor of Paris. However, we will still refer to Paris Arrondissements as communes for simplicity. The 1300 communes are grouped into 8 districts (departements). Table 2 presents the origin and destination rings and districts for the moves observed in 1998. Most of the moves are within the same district, although households who move to Paris come predominantly from outside Ile-de-France, following a classical migration staging pattern observed in many other cities. The most common destination for immigrants from outside Ile-de-France region is Paris, followed by district 92 in the inner ring, which shares many attributes with Paris.

Table (2): The distribution of moves between different rings in 1998 (origin by destination).

		Origin district				
Current District		Outside	Paris	Inner Ring	Outer Ring	Total
Paris	Frequency	77.579	67.027	18.192	18.023	180.821
	Percent	42.9%	37.1%	10.1%	10.0%	30.68%
Inner Ring	Frequency	61.135	22.633	103.205	20.168	207.141
	Percent	29.5%	10.9%	49.8%	9.7%	35.15%
Outer Ring	Frequency	49.936	9.299	23.967	118.191	201.393
	Percent	24.8%	4.6%	11.9%	58.7%	34.17%
Ile-de-France	Frequency	188.650	98.959	145.364	156.382	589.355
	Percent	32.01%	16.79%	24.66%	26.53%	100.00%

Source: Census, 1999.

According to the 1999 Census, 8.06% of the dwellings located in Ile-de-France were vacant in March 1999. While this represents a reasonable estimate of the vacancy rate over a very short period of time, for our analysis, we wish to approximate the availability of housing over a longer period of one year (one year supply).¹ Unfortunately, we have no information on how long each dwelling was vacant before a household moved in. However, during the period between the time a household moved out and another moved in, the dwelling was vacant and could have been chosen by another household. For this reason, some fraction, denoted by λ , of the dwellings occupied in March 1999 by a household which moved in 1998 should be included in the one year supply. If we assume that dwellings are vacant, on average, during half a year ($\lambda = 0.5$), then the one-year supply is made of 409,491 dwellings vacant in March 1999, and half the 589,355 dwellings in which a household moved in 1998, which represents a total supply of 704,168.5 dwellings in the region. We consider it as an upper bound for the supply, leading to a lower bound for the constraints, while we will also explore smaller values for λ . More interestingly, the time a dwelling remains vacant is shorter in Paris than in most of the other districts, which means that λ varies across the region. We therefore computed the λ fraction for each district under a stationarity assumption.² We assume that the ratio of the (cumulated) stock of vacant dwelling to the (cumulated) supply stock is equal to the ratio of the one-year flow of movers to the supply defined above, which can be considered as a one-year supply flow.³ Table 3 presents the vacancy rate and λ fraction computed by district and shows

¹We observe the households which moved during one year (1998), without information on the exact date they moved. In addition, we know which district they come from, but not which commune they come from. Moreover, we have no information on the number of households who left Ile de France in 1998.

²The λ fraction computed at a lower geographical level would be too sensitive to the randomness of the number of movers during a given year.

³Let M and V denote the numbers of movers and vacant units, respectively. The annual supply is $\lambda M + V$, so that the annual supply over demand ratio is $(\lambda M + V) / M$. Let T and S denote the total housing stock and the secondary residence units, respectively. The aggregate supply over demand ratio is $(T - S) / (T - S - V)$. Equating these two ratios, we find an expression for λ : $\lambda = \frac{T-S}{T-S-V} - \frac{V}{M}$.

that, although Paris displays the highest vacancy rate, it also displays the second lowest λ fraction.

Table (3): Vacancy rate and λ fraction.

Sub-region	District	Vacancy rate	λ fraction
Paris	75	0.10325	0.36773
Close Suburbs (Inner Ring)	92 (West)	0.08447	0.41411
	93 (North)	0.08461	0.28478
	94 (South)	0.07479	0.38813
Far away suburbs (Outer Ring)	78 (West)	0.06188	0.49909
	91 (South)	0.06596	0.42586
	95 (North)	0.06286	0.45274
	77 (East)	0.06557	0.44376

Source: Author’s computations from census data.

Most of the movers (71%) are male headed. The “poor households” (that is, the 33% households in the region with the lowest per capita income, defined as household income divided by the square root of the number of persons in the household) are unevenly distributed in the region: only 26% of households living in district 78, to the West of Paris, are poor, whereas this fraction goes up to 41% in district 93, to the East. These same two districts contain the highest (38% in district 78) and the lowest (21% in district 93) concentrations of rich households. Single-person households are highly concentrated in Paris city (52% of households in Paris are single). Between 25 to 30% percent of households of all the districts have two members. The larger families represent a larger share in rural districts in outer ring. About 25% percent of households have no working member, and out of these, 28% percent live in Paris city. Nearly 50% of the families in outer ring have two or more workers. Foreign households are most concentrated in district 93 (19%), and are less represented in the outer ring (9%). About 25% of households have a young head. They represent a larger share in Paris center and district 92 (31% and 27%) and their share is uniform in other districts (23%).

3 Model specification

In this section, we develop the specification for the model, beginning with the household residential location choice, and elaborating the basic model to address constraints. For clarity of exposition, we begin with a stronger assumption of homogeneous agents (no individual characteristics are considered), and then relax this to accommodate heterogeneous agents. We follow with the description of the iterative solution algorithm, first when the demand is known (parameters are given for the residential location choice model), and then when the demand is unknown. In this latter case, we need to estimate the parameters of the location choice model to check if the capacity constraint in a commune is violated or not, and to separate this constraint effect from the effect of preferences. The housing price model is a hedonic regression model that incorporates measures of aggregate demand and aggregate supply in each commune. We refer to this variation of the hedonic model as a *semi-hedonic* model, since it allows estimation of the degree to which prices adjust to help clear imbalances in the market, by incorporation both aggregate demand and supply. We turn now to the specification of the demand model.

3.1 Household residential location choice: basic model and notation

We refer to households that make a location choice within a given year as movers. These represent the set of agents producing demand for housing. Supply of housing is considered to be the set of housing units that are available for locating households (see the previous Section and footnote 3). The set of alternatives in the housing market are represented by $J = 1300$ communes, denoted by \mathcal{J} , with $Card(\mathcal{J}) = J$. The

demand for alternative j is denoted by D_j and the supply (or capacity) of alternative j is denoted by S_j . Individual decision makers are indexed by i , $i = 1 \dots N$. The utility of individual i selecting alternative j is:

$$U_j^i = V_j^i + \varepsilon_j^i, \quad i = 1 \dots N, \quad j \in \mathcal{J}, \quad (1)$$

where V_j^i represents the systematic component of the utility and where ε_j^i are i.i.d. random variables, with a double exponential distribution with zero mean variance equal to $\mu^2 \pi^2 / 6$. The probability \mathbb{P}_j^i that individual i prefers alternative (commune) j is given by the multinomial logit formula (See [2] for details):

$$\mathbb{P}_j^i = \frac{\exp(V_j^i / \mu)}{\sum_{j' \in \mathcal{J}} \exp(V_{j'}^i / \mu)}, \quad i = 1 \dots N, \quad j \in \mathcal{J}. \quad (2)$$

In the homogeneous case where we treat agents as though they were identical, systematic utilities and choice probabilities do not vary between individuals, so they are denoted by V_j and $\mathbb{P}_j = \frac{\exp(V_j)}{\sum_{j' \in \mathcal{J}} \exp(V_{j'})}$,

respectively. The expected demand, D_j , for alternative j is:

$$D_j = \begin{cases} \sum_{i=1}^N \mathbb{P}_j^i & \text{in the heterogeneous case} \\ N \mathbb{P}_j & \text{in the homogeneous case} \end{cases}, \quad j \in \mathcal{J}. \quad (3)$$

3.2 Introducing capacity constraints

The analysis of the residential choices in Ile-de-France (based on preliminary demand estimates) shows that many alternatives (about half, depending on the specification) have greater demand than supply. We consider below the situation where demand D_j strictly exceeds supply S_j for at least one alternative j . In that case, we say that the system is *constrained*. We denote $D_j - S_j$ as the excess demand, which is positive for at least one alternative when the system is constrained. When the system is constrained, the choices based only on preferences, or *ex ante* choices, differ from the actual allocation consistent with preferences and with capacity constraints, or *ex post* allocation, because the constraints are binding and some households are forced to take an alternative which was not their first preference. Note that even when only one alternative is constrained, ex post demands is a priori different from ex ante demand, for all alternatives. We refer to the demand after constraints are imposed as the *ex post demand*. It is in this sense that we will refer to *equilibrium*, or *market clearing*. The time frame for the analysis is considered short-term, with supply assumed to be constant.

In order to guarantee that there exists at least one feasible allocation, we also assume that constraints are not globally too severe. That is, we impose a feasibility condition that aggregate supply is sufficient to accommodate aggregate demand, resulting in a vacancy rate that is strictly positive:

$$N = \sum_{j \in \mathcal{J}} D_j < \sum_{j \in \mathcal{J}} S_j \quad (4)$$

When the system is constrained, the (ex post) probability that individual i is allocated to alternative j is denoted by $\tilde{\mathbb{P}}_j$ in the homogeneous case (or by $\tilde{\mathbb{P}}_j^i$ in the heterogeneous case). The ex ante demand D_j defined in (3) corresponds to ex ante choices, whereas the *observed demand* (or *ex post demand*), denoted by \tilde{D}_j , corresponds to the ex post allocation:

$$\tilde{D}_j = \begin{cases} \sum_{i=1}^N \tilde{\mathbb{P}}_j^i & \text{in the heterogeneous case} \\ N \tilde{\mathbb{P}}_j & \text{in the homogeneous case} \end{cases}. \quad (5)$$

Note that, when the system is not constrained, then $\tilde{\mathbb{P}}_j^i = \mathbb{P}_j^i$, $j \in \mathcal{J}$, for all alternatives and all individuals, and the observed demand is equal to the ex ante demand for all alternatives. When the system is

constrained, \tilde{D}_j is bounded by S_j ($j \in \mathcal{J}$) and the constraint is binding for at least one alternative. For a given alternative j , two situations may arise:

1. Either alternative j is *unconstrained ex ante*: $D_j < S_j$. In this case, alternative j is *unconstrained ex post* if $\tilde{D}_j < S_j$, while alternative j is *constrained ex post* if $\tilde{D}_j = S_j$.
2. Or alternative j is *constrained ex ante*: $D_j \geq S_j$. In this case, it can be shown, under simple regularity conditions, that $\tilde{D}_j = S_j$, i.e. alternative j is also *constrained ex post*.

We define the following sets of unconstrained alternatives: $\bar{\mathcal{C}} = \{j \in \mathcal{J} \mid \tilde{D}_j < S_j\}$ denotes the alternatives unconstrained ex post, while $\bar{\mathcal{C}}(0) = \{j \in \mathcal{J} \mid D_j < S_j\}$ denotes the alternatives unconstrained ex ante.

Similarly, we define the following sets of constrained alternatives: $\mathcal{C} = \{j \in \mathcal{J} \mid \tilde{D}_j = S_j\}$ denotes the alternatives constrained ex post, while $\mathcal{C}(0) = \{j \in \mathcal{J} \mid D_j \geq S_j\}$ denotes the set of alternatives constrained ex ante.

These four sets determine two partitions of \mathcal{J} (since $\mathcal{C} \cap \bar{\mathcal{C}} = \mathcal{C}(0) \cap \bar{\mathcal{C}}(0) = \emptyset$ and $\mathcal{C} \cup \bar{\mathcal{C}} = \mathcal{C}(0) \cup \bar{\mathcal{C}}(0) = \mathcal{J}$). When the system is constrained, the actual allocation is the result of a complex mechanism which depends on the assumptions (priority rules) introduced below. Such assumptions characterize ex post allocation.

Constraints have two consequences: first, if alternative j is constrained ex ante ($j \in \mathcal{C}(0)$), a fraction of the individuals who would select alternative j without capacity constraints (D_j) must instead choose a less desirable alternative k . Second, due to the first consequence, the excess demand generated in alternatives constrained ex ante is reallocated to alternatives which were not constrained ex ante. Therefore, the observed demand is larger than (or equal to) the ex ante demand in all alternatives unconstrained ex ante. Some of the alternatives which are not constrained ex ante may be constrained ex post, due to the reallocation of excess demand.

We now introduce a first assumption: *free allocation*, which means that if an individual prefers an alternative j , which is unconstrained ex post, he can be sure to be allocated to it.

Assumption 1 (Free allocation) *Let $j \in \bar{\mathcal{C}}$. Then*

$$\mathbb{P}(\{i \text{ allocated to } j \mid i \text{ prefers } j\}) = 1, \forall i = 1 \dots N.$$

However, note that an individual may be denied access to his preferred choice j if it is constrained ex post, and be forced to choose another alternative (that may be constrained or unconstrained ex ante). Assumption 1 implies that the IIA property (specific to the MNL model) is valid for the unconstrained alternative, that is, for two alternatives unconstrained ex post, the ratio of (ex post) allocation probabilities is equal to the ratio of (ex ante) choice probabilities. Assumption 1 implies the IIA property in the following sense:

Lemma 1 (Alternative unconstrained ex post) *If Assumption 1 holds and if alternative j and k are unconstrained ex post, then the allocation probabilities satisfy the IIA property in the sense that:*

$$\frac{\hat{\mathbb{P}}_j^i}{\hat{\mathbb{P}}_k^i} = \frac{\mathbb{P}_j^i}{\mathbb{P}_k^i}, \forall i = 1 \dots N, \forall j, k \in \bar{\mathcal{C}}. \quad (6)$$

Moreover:

$$\hat{\mathbb{P}}_j^i \geq \mathbb{P}_j^i \text{ and } \tilde{D}_j \geq D_j, \forall i = 1 \dots N, \forall j \in \bar{\mathcal{C}}. \quad (7)$$

The equality (6) states that the allocation probabilities satisfy the IIA property for unconstrained alternatives. The inequalities (7) state that the probability that individual i is allocated to j is larger than the probability that he prefers j and the observed demand addressed to j is larger than the ex ante demand if j is unconstrained. An immediate consequence of Assumption 1 and Lemma 1 is that $\mathcal{C}(0) \subset \mathcal{C}$

and $\bar{\mathcal{C}} \subset \bar{\mathcal{C}}(0)$. That is: if alternative j is constrained ex ante, then it is also constrained ex post. If alternative j is unconstrained ex post, then it is also unconstrained ex ante.

Proposition (1) states that the individual ratio $\tilde{\mathbb{P}}_j^i/\mathbb{P}_j^i$ of the actual allocation probability to the choice probability is the *same accross all unconstrained alternatives*, since equality (6) can be rewritten as:

$$\frac{\tilde{\mathbb{P}}_j^i}{\mathbb{P}_j^i} = \frac{\tilde{\mathbb{P}}_k^i}{\mathbb{P}_k^i} \stackrel{def}{=} \Omega^i, \quad \forall i = 1 \dots N, \quad \forall j, k \in \bar{\mathcal{C}},$$

where the common value of this ratio is denoted by Ω^i . We interpret Ω^i as the *individual allocation ratio* for individual i . The computation of Ω^i cannot be performed before the allocation rules are defined for the ex post constrained alternatives.

We now introduce a second assumption concerning the allocation in alternatives which are constrained ex post. We assume that if an individual i has a stronger preference (ex ante) for the alternative j constrained ex post than another individual i' , in the sense that the choice probability for i is larger than the choice probability for i' , individual i will also have proportionally more chances than individual i' to be allocated ex post to alternative j in the following sense:

Assumption 2 (No priority rule) *If $j \in \mathcal{C}$, the individual allocation ratio of alternative j , constrained ex post is the same for all individuals:*

$$\frac{\tilde{\mathbb{P}}_j^i}{\mathbb{P}_j^i} = \frac{\tilde{\mathbb{P}}_j^{i'}}{\mathbb{P}_j^{i'}} \stackrel{def}{=} \Phi_j, \quad \forall i, i' = 1 \dots N, \quad \forall j \in \mathcal{C}.$$

Assumption (2), which states that the individual allocation ratio for alternative j , denoted by Φ_j , is the same for each individual ($i = 1 \dots N$). Assumption 2 can therefore be interpreted as a fairness criterion.

This assumption suffices to determine the queue discipline for the alternatives which are constrained ex post. Indeed, it is straightforward to show that the common value (accross individuals) of the allocation ratio $\tilde{\mathbb{P}}_j^i/\mathbb{P}_j^i$ is equal to the relative supply (measured by S_j/D_j):

Lemma 2 (Alternative constrained ex post) *Consider an individual i . If Assumption 2 holds and alternative j is constrained ex post, the common value of the individual allocation ratio is equal to the relative supply, i.e.:*

$$\Phi_j = \frac{S_j}{D_j} < 1, \quad \forall i = 1 \dots N, \quad \forall j \in \mathcal{C}.$$

Assumptions 1 and 2 hold, throughout the rest of the paper. They are essentially sufficient to solve for an equilibrium. It is necessary, also, to eliminate extreme preferences (which are hardly observed in our data), which corresponds to a kind of “no preference reversal” assumption.⁴

We provide below the solution for the homogeneous case and sketch the solution for the heterogenous case.

3.3 Equilibrium solution with capacity constraints

The solution requires the computation of two unknown pieces of information: (1) which alternatives are constrained ex post and (2) the value of Ω^i . We assume for the moment that we know which alternatives are constrained ex post and will return to the computation of this list subsequently.

⁴We need to eliminate extreme preferences. With that respect, one supplementary assumption is the “no preference reversal” : no individual in the population has a strong preference for alternatives, which are constrained ex post but not ex ante. More specifically, we assume that $\sum_{j \in \mathcal{C} \cap \bar{\mathcal{C}}(0)} \mathbb{P}_j^i$ is not too large.

Recall that, for an unconstrained alternative j , the individual allocation ratio, $\tilde{\mathbb{P}}_j^i/\mathbb{P}_j^i = \Omega^i$, is independent of the alternative. For any alternative j unconstrained ex post, the individual allocation ratio $\tilde{\mathbb{P}}_j^i/\mathbb{P}_j^i$ is larger than 1 since some individuals are reallocated from their preferred constrained alternative to alternative j , whereas, according to Assumption 1, all individuals who prefer j are allocated to j . The value of the individual allocation ratio in unconstrained alternatives is given by:

Lemma 3 (Individual allocation ratio) *Consider an individual i . The individual allocation ratio is the same for each alternative unconstrained ex post and is given by :*

$$\frac{\tilde{\mathbb{P}}_j^i}{\mathbb{P}_j^i} = \Omega^i = \frac{1 - \sum_{j \in \mathcal{C}} \frac{S_j}{D_j} \mathbb{P}_j^i}{\sum_{k \in \bar{\mathcal{C}}} \mathbb{P}_k^i} > 1, \forall j \in \bar{\mathcal{C}}.$$

Lemma 3 shows that the individual allocation ratio $\Omega^i = \tilde{\mathbb{P}}_j^i/\mathbb{P}_j^i$, $\forall j \in \bar{\mathcal{C}}$ is uniquely determined once the set \mathcal{C} of alternatives constrained ex post is known. Note that $\Omega^i = \frac{1 - \sum_{j \in \mathcal{C}} \frac{S_j}{D_j} \mathbb{P}_j^i}{1 - \sum_{j \in \mathcal{C}} \mathbb{P}_j^i}$, which is larger than 1 since $\frac{S_j}{D_j} < 1, \forall j \in \mathcal{C}(0)$ and $\sum_{j \in \mathcal{C} \cap \bar{\mathcal{C}}(0)} \mathbb{P}_j^i$ is not too large (see footnote 4), so that the negative difference $\sum_{j \in \mathcal{C} \cap \bar{\mathcal{C}}(0)} \left(\mathbb{P}_j^i - \frac{S_j}{D_j} \mathbb{P}_j^i \right)$ cannot offset the positive difference $\sum_{j \in \mathcal{C}(0)} \left(\mathbb{P}_j^i - \frac{S_j}{D_j} \mathbb{P}_j^i \right)$.

The ex ante demand is given by (3), while the ex post demand is defined by: $\tilde{D}_j = \sum_{i=1}^N \tilde{\mathbb{P}}_j^i = \sum_{i=1}^N \Omega^i \mathbb{P}_j^i$ (see equation (5)). The aggregate allocation ratio, $\bar{\Omega}_j$, is defined as a weighted average of individual allocation ratios Ω^i :

$$\bar{\Omega}_j \stackrel{def}{=} \frac{\sum_{i=1}^N \Omega^i \mathbb{P}_j^i}{\sum_{i=1}^N \mathbb{P}_j^i} > 1, \forall j \in \mathcal{J}. \quad (8)$$

Note that $\bar{\Omega}_j > 1$ if $j \in \bar{\mathcal{C}}$ (see Lemma 3). In this case, we have (using (3)):

$$\tilde{D}_j = \bar{\Omega}_j D_j \text{ and } \tilde{\mathbb{P}}_j^i = \Omega^i \mathbb{P}_j^i, \forall j \in \bar{\mathcal{C}}.$$

Collecting the previous results, we have:

Lemma 4 (Aggregate allocation ratio) *Consider an alternative unconstrained ex post $j \in \bar{\mathcal{C}}$. The alternative-specific aggregate allocation ratio is :*

$$\bar{\Omega}_j = \frac{\sum_{i=1}^N \Omega^i \mathbb{P}_j^i}{\sum_{i=1}^N \mathbb{P}_j^i} = \frac{\tilde{D}_j}{D_j}, \quad j \in \bar{\mathcal{C}},$$

where Ω^i are given by Lemma 3.

Recall that the value of the individual allocation ratio Ω^i can be computed once the set \mathcal{C} is known (see Lemma 3). Conversely, assume that the ratios Ω^i (and therefore $\bar{\Omega}_j$) are known. It can be shown that an alternative j is constrained ex post if and only if $\bar{\Omega}_j D_j = \sum_{i=1}^N \Omega^i \mathbb{P}_j^i$ is larger than (or equal to) S_j . Therefore, we have:

$$j \in \bar{\mathcal{C}} \Leftrightarrow \bar{\Omega}_j \geq \frac{S_j}{D_j}. \quad (9)$$

Note that, if alternative j is constrained ex ante, then $S_j/D_j < 1$, and the condition (9) is always satisfied (ex ante constraint implies ex post constraint). If alternative j is constrained ex post but not ex ante, then $S_j/D_j \geq 1$, but $\bar{\Omega}_j$ is even larger. In both cases ($j \in \mathcal{C}(0)$ and $j \in \mathcal{C} \cap \bar{\mathcal{C}}(0)$), we have:

$$\tilde{D}_j = S_j \text{ and } \tilde{\mathbb{P}}_j^i = \frac{S_j}{D_j} \mathbb{P}_j^i, \quad j \in \mathcal{C}.$$

The proofs of existence and uniqueness of a global solution (Ω^i, \mathcal{C}) remain an open issue (although numerical simulation suggest that existence and uniqueness are guaranteed).

Below, we provide a method for computing the allocation equilibrium, either when the demand is known and the parameters are given, or when it is unknown and the parameters need to be estimated in tandem with the iterative procedure.

3.4 Computational method

3.4.1 Allocation Probabilities

The allocation probabilities are given by different expressions according to whether the alternatives are constrained or not ex post. The allocation probabilities can still be written as a multinomial logit model, but with a additional term, or “correction factor”, $\ln(\pi_j^i)$, which expresses the allocation ratio.

Theorem 1 (Allocation probabilities) *If Assumptions 1 and 2 hold, the allocation probabilities are given by the adjusted MNL formula:*

$$\tilde{\mathbb{P}}_j^i = \frac{\exp(\tilde{V}_j^i/\mu)}{\sum_{k \in \mathcal{J}} \exp(\tilde{V}_k^i/\mu)}, \quad \text{with} \quad (10a)$$

$$\tilde{V}_j^i = V_j^i + \mu \ln(\pi_j^i), \quad \text{with} \quad (10b)$$

$$\pi_j^i = \begin{cases} \frac{S_j}{D_j} & \text{if } j \in \mathcal{C} \\ \Omega^i & \text{if } j \in \bar{\mathcal{C}} \end{cases}. \quad (10c)$$

Proof. *The denominator in (10a) is*

$$\begin{aligned} \sum_{k \in \mathcal{J}} \exp(\tilde{V}_k^i/\mu) &= \sum_{k \in \mathcal{C}} \frac{S_k}{D_k} \exp(V_k^i/\mu) + \Omega^i \times \sum_{k \in \bar{\mathcal{C}}} \exp(V_k^i/\mu) \\ &= \sum_{j \in \mathcal{J}} \exp(V_j^i/\mu) \times \left[\sum_{k \in \mathcal{C}} \frac{S_k}{D_k} \frac{\exp(V_k^i/\mu)}{\sum_{j \in \mathcal{J}} \exp(V_j^i/\mu)} + \Omega^i \sum_{k \in \bar{\mathcal{C}}} \frac{\exp(V_k^i/\mu)}{\sum_{j \in \mathcal{J}} \exp(V_j^i/\mu)} \right] \\ &= \sum_{j \in \mathcal{J}} \exp(V_j^i/\mu) \times \left[\sum_{k \in \mathcal{C}} \frac{S_k}{D_k} \mathbb{P}_k^i + \Omega^i \sum_{k \in \bar{\mathcal{C}}} \mathbb{P}_k^i \right] \\ &= \sum_{j \in \mathcal{J}} \exp(V_j^i/\mu) \times \sum_{j \in \mathcal{J}} \tilde{\mathbb{P}}_j^i = \sum_{j \in \mathcal{J}} \exp(V_j^i/\mu). \end{aligned}$$

Therefore:

$$\frac{\exp(\tilde{V}_j^i/\mu)}{\sum_{k \in \mathcal{J}} \exp(\tilde{V}_k^i/\mu)} = \pi_j^i \times \frac{\exp(V_j^i/\mu)}{\sum_{j \in \mathcal{J}} \exp(V_j^i/\mu)} = \pi_j^i \mathbb{P}_j^i = \tilde{\mathbb{P}}_j^i.$$

■

The reader should still keep in mind that the difficult part of this approach (and of any approach dealing with constraints) is the determination of the set of alternatives which are constrained ex post.

We first describe the iterative procedure to find the allocation probabilities and the ex post constrained alternatives, when the demand is known, as would be the case once the model is estimated and is being used to make predictions. This simplifies the initial exposition. Next we treat the case when demand is unknown and has to be estimated.

3.4.2 Solution algorithm with known demand

The formulas are given by Lemma 3 for the individual allocation ratio Ω^i , by Lemma 4 for the aggregate allocation ratio $\bar{\Omega}_j$ and by condition (9) for the set \mathcal{C} of alternatives j constrained ex post. The following algorithm allows computation of Ω^i , $\bar{\Omega}_j$ and \mathcal{C} with less than J iterations.

Algorithm K: known demand

1. Check that aggregate demand is smaller than aggregate supply (see Condition (4)).
2. Iteration $l = 0$ (initialization): set $\Omega^i = 1$; compute the set $\mathcal{C}(0) = \{j \in \mathcal{J} \mid S_j \leq D_j\}$ of alternatives constrained at iteration zero (that is, constrained ex ante). The set of alternatives unconstrained at iteration zero (that is, unconstrained ex ante) is given by: $\bar{\mathcal{C}}(0) = \mathcal{J} \setminus \mathcal{C}(0)$.
3. Compute the individual allocation ratios $\Omega^i(0) = \frac{1 - \sum_{j \in \mathcal{C}(0)} \frac{S_j \mathbb{P}_j^i}{D_j}}{1 - \sum_{j \in \mathcal{C}(0)} \frac{\mathbb{P}_j^i}{D_j}}$, using Lemma 3.
4. Compute the alternative-specific allocation ratios using Equation (8):

$$\bar{\Omega}_j(0) = \frac{\sum_{i=1}^N \Omega^i(0) \mathbb{P}_j^i}{\sum_{i=1}^N \mathbb{P}_j^i}.$$

5. Update iteration: $l \rightarrow l + 1$. Update the constrained choice set using Condition (9) and Definition (8):

$$\mathcal{C}(l+1) = \left\{ j \in \mathcal{J} \mid S_j \leq \bar{\Omega}_j(l) D_j = \sum_{i=1}^N \Omega^i(l) \mathbb{P}_j^i \right\}.$$

6. Update: $\Omega^i(l+1) = \frac{1 - \sum_{j \in \mathcal{C}(l+1)} \frac{S_j \mathbb{P}_j^i}{D_j}}{\sum_{k \in \bar{\mathcal{C}}(l+1)} \frac{\mathbb{P}_k^i}{D_k}}$

7. Update $\bar{\Omega}_j(l+1) = \frac{\sum_{i=1}^N \Omega^i(l+1) \mathbb{P}_j^i}{\sum_{i=1}^N \mathbb{P}_j^i}$

8. Stop at iteration $l + 1$ if $\mathcal{C}(l + 1) = \mathcal{C}(l)$ (and therefore $\Omega^i(l + 1) = \Omega^i(l)$ for all $i = 1 \dots N$), else go to 5. By definition, convergence is attained at $l = \infty$.

9. Compute the correction factor

$$\ln(\pi_j^i) = \begin{cases} \ln(S_j/D_j) & \text{if } j \in \mathcal{C} \\ \ln(\Omega^i(\infty)) & \text{if } j \in \bar{\mathcal{C}} \end{cases}.$$

using Theorem 1.

In the homogenous case, it can be shown that this algorithm converges. Simulation experiments suggest that it also converge in the heterogenous case.

3.4.3 Solution algorithm with unknown demand

When the demand is unknown, we need to jointly estimate the parameters of the model and to determine the allocated demand, given a set of parameters. The second set of iterations (module 2) determines the alternatives constrained ex post, the allocated demand and the correction factors given the demand parameters. It corresponds to algorithm K. The first set of iterations (routine 1) generates a set of estimated parameters in the adjusted MNL formula.

The approach in module 2 is to iterate between an estimation module, which incorporates the correcting factor $\ln(\pi_j^i)$, computed during the previous iteration and module 1, which in turn relies on the previous iteration estimated demand. The estimation module and module 1 are embedded in a double loop. The outer loop (estimation) is indexed by l' and the inner loop (module 1) is indexed by l .

We require additional notations: $\hat{D}_j(l')$, $\hat{\mathbb{P}}_j^i(l')$ and $\ln[\pi_j^i(l')]$ represent the estimation of the demand, the choice probabilities and the correction factor at iteration l' . Moreover, $\mathcal{C}(l', l)$ denotes the set of alternatives constrained ex post at iteration (l', l) , where l' corresponds to the demand estimation and l corresponds to the constraints. We define $\Omega^i(l', l)$ and $\bar{\Omega}_j(l', l)$ in a similar way. Below, we sketch the structure of the solution algorithm when unknown demand is estimated.

Algorithm U: unknown demand

1. Check that the aggregate demand is smaller than the aggregate supply (Condition (4)).
2. Iteration $l' = 0$ (initialization of demand of the estimation module): estimate a MNL model assuming initially no constraints (i.e. $\Omega^i = \bar{\Omega}_j = \pi_j^i = 1$). This leads to the estimated choice probabilities $\hat{\mathbb{P}}_j^i(0)$ and to the estimated demand $\hat{D}_j(0) = \sum_{i=1}^N \hat{\mathbb{P}}_j^i(0)$. The correcting factors $\ln[\pi_j^i(0)]$ are initialized to 0 and play no role in Equation (10a).

- (a) Iteration $l = 0$ (initialization of the constraints): compute the set

$$\mathcal{C}(0, 0) = \left\{ j \in \mathcal{J} \mid S_j \leq \hat{D}_j(0) \right\}$$

of alternatives constrained ex ante for the initial estimated demand, using Condition (9).

- (b) Compute the individual allocation ratio, using Lemma 3:

$$\Omega^i(0, 0) = \left(1 - \sum_{j \in \mathcal{C}(0,0)} \frac{S_j}{\hat{D}_j(0)} \hat{\mathbb{P}}_j^i(0) \right) / \left(1 - \sum_{j \in \mathcal{C}(0,0)} \hat{\mathbb{P}}_j^i(0) \right).$$

- (c) Compute the alternative-specific allocation ratio using Equation (8):

$$\bar{\Omega}_j(0, 0) = \left(\sum_{i=1}^N \Omega^i(0, 0) \hat{\mathbb{P}}_j^i(0) \right) / \left(\sum_{i=1}^N \hat{\mathbb{P}}_j^i(0) \right).$$

- (d) Update iteration for algorithm K: $l \rightarrow l + 1$. Update the set of alternatives constrained using Condition (9):

$$\mathcal{C}(0, l + 1) = \left\{ j \in \mathcal{J} \mid S_j \leq \bar{\Omega}_j(0, l) \hat{D}_j(0) = \sum_{i=1}^N \Omega^i(0, l) \hat{\mathbb{P}}_j^i(0) \right\}$$

- (e) Update $\Omega^i(0, l + 1) = \left(1 - \sum_{j \in \mathcal{C}(0, l+1)} \frac{S_j}{\hat{D}_j(0)} \hat{\mathbb{P}}_j^i(0) \right) / \left(1 - \sum_{j \in \mathcal{C}(0, l+1)} \hat{\mathbb{P}}_j^i(0) \right)$

- (f) Update $\bar{\Omega}_j(0, l+1) = \left(\sum_{i=1}^N \Omega^i(0, l+1) \hat{\mathbb{P}}_j^i(0) \right) / \left(\sum_{i=1}^N \hat{\mathbb{P}}_j^i(0) \right)$.
- (g) Stop at iteration $l+1$ if $\mathcal{C}(0, l+1) = \mathcal{C}(0, l)$ (and therefore $\Omega^i(0, l+1) = \Omega^i(0, l)$, $i = 1 \dots N$), else go to step d.
- (h) When convergence is attained ($l = \infty$), we have $\mathcal{C}(0) \stackrel{def}{=} \left\{ j \in \mathcal{J} \mid S_j \leq \bar{\Omega}_j(0, \infty) \hat{D}_j(0) \right\}$, $\Omega^i(0) \stackrel{def}{=} \Omega^i(0, \infty)$, and $\bar{\Omega}_j(0) \stackrel{def}{=} \bar{\Omega}_j(0, \infty)$.
3. Update iteration of the estimation module: $l' \rightarrow l'+1$: update $\pi_j^i(l'+1) = \begin{cases} \frac{S_j}{\hat{D}_j(l')} & \text{if } j \in \mathcal{C}(l') \\ \Omega^i(l') & \text{if } j \in \bar{\mathcal{C}}(l') \end{cases}$.
4. Stop if $|\pi_j^i(l'+1) - \pi_j^i(l')| < 10^{-4}$, $\forall i = 1 \dots N$, $\forall j \in \mathcal{J}$ ($l' = \infty$). The solution is: $D_j = \hat{D}_j(l') = \hat{D}_j(\infty)$, $j \in \mathcal{J}$, $\mathcal{C} = \mathcal{C}(l') = \mathcal{C}(\infty)$, $\Omega^i = \Omega^i(l') = \Omega^i(\infty)$ and $\bar{\Omega}_j = \bar{\Omega}_j(l') = \bar{\Omega}_j(\infty)$; else go to step 5.
5. Update the MNL estimates with updated $\pi_j^i(l'+1)$ among the explanatory variables in \tilde{V}_j^i . This gives $\tilde{\mathbb{P}}_j^i(l'+1) = \frac{\exp(\tilde{V}_j^i(l'+1))}{\sum_{k \in \mathcal{J}} \exp(\tilde{V}_k^i(l'+1))}$, $\hat{\mathbb{P}}_j^i(l'+1) = \frac{\tilde{\mathbb{P}}_j^i(l'+1)}{\pi_j^i(l'+1)}$ and $\hat{D}_j(l'+1) = \sum_{i=1}^N \hat{\mathbb{P}}_j^i(l'+1)$
- (a) Iteration $l = 0$ (initialization): Compute the set $\mathcal{C}(l'+1, 0) = \left\{ j \in \mathcal{J} \mid S_j \leq \hat{D}_j(l'+1) \right\}$ of alternatives constrained ex ante for demand at step $(l'+1)$.
- (b) Compute the individual allocation ratio
- $$\Omega^i(l'+1, 0) = \left(1 - \sum_{j \in \mathcal{C}(l'+1, 0)} \frac{S_j}{\hat{D}_j(l'+1)} \hat{\mathbb{P}}_j^i(l'+1) \right) / \left(1 - \sum_{j \in \mathcal{C}(l'+1, 0)} \hat{\mathbb{P}}_j^i(l'+1) \right).$$
- (c) Compute the alternative-specific allocation ratio $\bar{\Omega}_j(l'+1, 0) = \frac{\sum_{i=1}^N \Omega^i(l'+1, 0) \hat{\mathbb{P}}_j^i(l'+1)}{\sum_{i=1}^N \hat{\mathbb{P}}_j^i(l'+1)}$.
- (d) Update iteration of Algorithm K: $l \rightarrow l+1$. Update
- $$\mathcal{C}(l'+1, l+1) = \left\{ j \in \mathcal{J} \mid S_j \leq \bar{\Omega}_j(l'+1, l) \hat{D}_j(l'+1) = \sum_{i=1}^N \Omega^i(l'+1, l) \hat{\mathbb{P}}_j^i(l'+1) \right\}.$$
- (e) Update $\Omega^i(l'+1, l+1) = \left(1 - \sum_{j \in \mathcal{C}(l'+1, l+1)} \frac{S_j}{\hat{D}_j(l'+1)} \hat{\mathbb{P}}_j^i(l'+1) \right) / \left(1 - \sum_{j \in \mathcal{C}(l'+1, l+1)} \hat{\mathbb{P}}_j^i(l'+1) \right)$.
- (f) Update $\bar{\Omega}_j(l'+1, l+1) = \left(\sum_{i=1}^N \Omega^i(l'+1, l+1) \hat{\mathbb{P}}_j^i(l'+1) \right) / \left(\sum_{i=1}^N \hat{\mathbb{P}}_j^i(l'+1) \right)$.
- (g) Stop Algorithm K at iteration $l+1$ if $\mathcal{C}(l'+1, l+1) = \mathcal{C}(l'+1, l)$ (and therefore $\Omega^i(l'+1, l+1) = \Omega^i(l'+1, l)$ and $\bar{\Omega}_j(l'+1, l+1) = \bar{\Omega}_j(l'+1, l)$), else go to step d.
- (h) When convergence of Algorithm K is attained ($l = \infty$), we have
- $$\mathcal{C}(l'+1) = \left\{ j \in \mathcal{J} \mid S_j \leq \bar{\Omega}_j(l'+1, \infty) \hat{D}_j(l'+1) \right\}, \Omega^i(l'+1) = \Omega^i(l'+1, \infty), \text{ and}$$
- $$\bar{\Omega}_j(l'+1) = \bar{\Omega}_j(l'+1, \infty).$$
6. Go to step 3.

In the next section, we use Algorithm U to estimate the residential location demand model when the system is constrained. In the simulation results reported (and not reported), Algorithm U always converged to a unique solution.

4 Empirical results

The data for household location model come from the 1999 census, for which we were able to access household-level data for the entire population, allowing us to test the sensitivity of the estimation to using a range of sampling rates from the population, a range of numbers of alternatives and various specifications for demand equation and supply computation (value of the λ fraction). The analysis focuses on “recent movers”: households who settled or moved to the region recently, specifically, during 1998. Among the 4,510,369 households living in the study area in March 1999, 589,355 moved into or within the region of interest during year 1998. The housing price data used in the model come from the “base de données des Notaires” and contains the average price of single-family dwellings sold in the commune in 1999. Note that the single-family dwellings are usually larger and are on larger lots in the suburbs, so the differences between Paris and the suburbs are smaller than differences in price per square meter. The attributes of the alternatives have been computed from different sources, mainly drawing on data from the IAURIF metropolitan planning agency. See [5] for details on the data and on the estimation of the price equation.

4.1 Housing price

We have estimated hedonic and semi-hedonic regression models to predict housing prices, and to test for the endogeneity of prices in the residential location choice model. In the semi-hedonic regressions, we put measures of housing supply (considered static in the short-run period we are modeling), and demand, which is computed from the residential location choice model, on the right hand side of the hedonic regression. This specification, linking the hedonic regression and the residential location choice model via aggregate demand, allows us to empirically test the degree to which price adjustments based on the varying relationship of demand and supply serve to clear the market. As is standard in the hedonic literature, we specify the model as semi-log, with the natural log of housing prices as the dependent variable.

The estimated coefficients for housing price model are presented in Table 4. The R^2 for the model is 53%, which is rather high considering that we are using only the average sales price in each commune and therefore have no attributes of individual houses entered in the model.⁵ The only effects are commune characteristics and the aggregate market conditions of demand and supply. We obtain the expected signs for demand and supply but they are not exactly opposed. From a purely structural equation (results not reported here, available on request) with only supply and demand, we cannot reject the hypothesis that coefficients are exactly opposed, which means that the price only depends on the supply/demand ratio, and not separately on supply and demand. This result does not hold *ceteris paribus*, that is in the semi-hedonic regression. Note that the R^2 of the purely structural price equation is very low (about 5%), suggesting that prices are not very efficient at clearing the market. This is also supported by the fact that, when controlling for some commune characteristics, supply and demand are not so significant in the price equation.

A decrease in average travel time⁶ significantly increases the price: 10 minutes less imply a 2.8% increase in housing price. The price is very sensitive to socio-economic structure of the commune: a 10% increase in the proportion of one-member households is associated with a 50% increase of the price. Such an increase for the proportion of two-members households corresponds to a 19% increase of the price. Similarly, the fraction of households with no or only one working member has a positive effect on the price. Surprisingly, the fraction of foreign households has a positive effect on price. We should note however, that the data do not distinguish the nationality of the foreigners, and make no difference between OECD

⁵Note that the R^2 for separate price equations for houses (single-family dwellings) and flats led to larger R^2 (72% and 75%, respectively).

⁶Weighted average travel time by public transport, from the commune considered to any destination. The weights correspond to the number of trips observed from this commune during a given day.

and third world countries. In addition, we are controlling for the income of the commune, which shows negative and highly significant effects of the proportion of low and intermediate income families on the price.

Table (4): Housing price estimation results

Variable	Coefficient	Standard error	t-statistic	p-value
Intercept	11.02668	0.12800	86.14	<.0001
Log(Supply)	-0.04791	0.02466	-1.94	0.0522
Log(Demand)	0.09918	0.02244	4.42	<.0001
Average travel time from j to work (minutes)	-0.00280	0.000851	-3.28	0.0011
% households with 1 member	5.09136	0.37884	13.44	<.0001
% households with 2 members	1.87960	0.34135	5.51	<.0001
% households with no working member	1.25241	0.30954	4.05	<.0001
% households with 1 working member	0.82300	0.33762	2.44	0.0149
% poor households	-6.63187	0.50316	-13.18	<.0001
% households with medium income	-4.54311	0.33102	-13.72	<.0001
% households with a foreign head	1.58406	0.36279	4.37	<.0001

4.2 Location choice without capacity constraints

Given the large size of the databases, there is a trade-off between the number of households, the number of alternatives and the number of explanatory variables which can be used for estimating the model. Table 5 contains the results of the residential location choice model estimation, estimated on a 0.5% sample of households (that is 2,946 households) with all the alternatives, and assuming that the system is not constrained (iteration $l' = 0$ in algorithm U). With a pseudo- R^2 (McFadden’s likelihood ratio index) of 30.6%, this model has a rather good explanatory power at the individual level. Later on, we will explore the validity of this model at the aggregate level (see Section 4.4.4). We find a very significant effect of the “same district as before move” variable, confirming (as expected) a strong preference of households to move in the same district or neighbourhood in which they lived before the move. Testing the effect of the distance from last residence would be interesting, but it was not possible to perform this test with our available data, which did not contain information on previous residence location more detailed than the district. The “Paris” coefficient is not significant, suggesting that the relative attractiveness of Paris is captured in local amenities and other explanatory variables.

As expected, housing price has a negative effect on location preference for a commune. We tested for the endogeneity of prices in the location choice model introducing the residual of the price equation (results not reported here). Adding the residuals of the price equation as an explanatory variable, the estimated coefficients changes trivially and the coefficient of this new variable is not at all significant. This result confirms that housing price is *not endogenous* (in the econometric sense) with regard to the location choice model. In other words, the variables used in these two models fully explain the correlation between prices and location choice. The price effect increases with the age of the household head and decreases with as the household income increases. That is, older heads of households are more sensitive to price and richer households are less sensitive to it, as would be expected. Since price is entered using three variables to capture average effects as well as interactions with age and income, the combined effects are complex. We note that the average price effect as well as the age and income interactions, all have expected signs. However, for a small subset of the population, namely very young and very rich households, the net price effect from the interaction of these three coefficients would be predicted by this model to show a slight positive preference for higher prices in communes where the neighbouring households are in the same socio-economic category and which have more amenities. The relative sensitivity to price is as expected, though the potential for a small positive preference for higher prices for this specific subpopulation and sample of locations is likely to be due to some amenities that are not accounted for in the model, rather

than an actual preference for a higher price, *ceteris paribus*. In that case, the price effect would become negative for all households if amenities were totally controlled for (see, e.g., Guevara and Ben-Akiva [7]).

In this specification, increase of the average travel time by public transit decreases equally the preference of households headed by a man or by a woman. On the other hand, location is not sensitive, *ceteris paribus*, to the average travel time by private car, or to the distance to highway. These latter results are probably explained by the large correlation between “TC” (public transit), “VP” (private car) and “Distance to highway” variables. In this model, the demand is not sensitive either to the number of subway (metro) or railway stations. This may also be explained by a large correlation with other explanatory variables such as the “Paris” dummy.

Table (5): Residential location choice, all alternatives, no constraints (0.5% sample)

Variable	Coefficient	Std error	t-stat.
Same district as before move	2.585	0.047	54.80
Paris	0.047	0.128	0.364
Log(Price)	-7.108	1.445	-4.920
Log(Price)* (Age-20)/10	-0.220	0.0473	-4.649
Log(Price)* Log(Income)	0.746	0.146	5.110
Number Railway stations	0.0053	0.0151	0.352
Number Subway stations	-0.0011	0.0070	-0.156
Average travel time from j, commuting (TC) [hour]	-0.635	0.157	-4.050
TC*(Dummy female) [hour]	-0.066	0.231	-0.287
Average travel time from j, by private car (VP) [hour]	0.219	0.265	0.827
Distance to highway [10^3 km]	-3.238	3.441	-0.941
% households with 1 member * 1 member in h	3.769	0.452	8.336
% households with 2 members * 2 members in h	-0.591	1.678	-0.352
% households with 3+ members * 3+ member in h	2.523	0.455	5.549
% hh with no working member * no working member in h	6.489	1.283	5.059
% hh with 1 working member * 1 working member in h	0.855	0.789	1.085
% hh with 2+ working member * 2+ working member in h	-0.890	0.556	-1.601
% poor households * h poor	1.459	0.664	2.197
% households with a foreign head * foreign head in h	5.125	0.911	5.628
% households with a foreign head * French head in h	-3.575	0.556	-6.432
Employment [1,000]	0.263	1.292	0.204
Density (Population/Surface) [1000 persons/km]	-6.625	4.360	-1.52
Log(Population)	1.011	0.029	35.00
% change in population, 1990 to 1999	0.185	0.097	1.898

The estimated coefficients corresponding to the socio-economic structure of the commune show a general preference of the households to live with neighbors of the same social category. This preference is very strong for singles, for large families, for households without workers, or with a foreign head. Households of French origin tend to avoid locations in which there are higher concentrations of foreign households. The coefficient for the total number of jobs is insignificant in this specification. Households prefer less dense communes, and the communes that have absorbed more population during the 1990-99 period attract still more households. Note that the very large and significant coefficient of “Log(Population)” should not be interpreted as a preference for more populated communes but rather as a proxy for a series of commune dummies allowing a better fit for aggregate demand.

4.3 Sensitivity analysis

In order to determine how robust these results are with respect to the sampling procedures we are using, we present below the results of sensitivity analyses to test for the effects of different sampling rates for households, and different sizes of sampled alternatives. In order to make the results comparable and to keep tractable dataset size, we use as a benchmark 0.5% of households with 100 alternatives and increase sequentially the number of alternatives (Table 6) and the number of households (Table 7).

4.3.1 Sampling alternatives

Table 6 presents the results when 100, 200, 500 and all alternatives are considered for each household. When only a fraction of alternatives is considered, they are selected randomly using an importance sampling strategy (at the commune level), with replacement. More specifically, the probability of inclusion of a commune in the choice set is proportional to the number of dwellings in the commune; each alternative is selected independently from the other alternatives considered for the same household and for the other households, and repetitions are allowed. This is equivalent to a uniform sampling of dwellings.⁷ The last column in Table 6 corresponds to the coefficients reported in Table 5. In order to save space, the significance of the coefficients is indicated using signs (*, † and ‡).

Table (6): Impact of the number of alternatives on residential location choice (0.5% sample)

Variable \ Household Sampling Rate	100	200	500	1300
Same district as before move	2.578‡	2.583‡	2.582‡	2.585‡
Paris	0.079	0.246*	0.229*	0.047
Log(Price)	-7.847‡	-8.062‡	-7.790‡	-7.108‡
Log(Price)* (Age-20)/10	-0.233‡	-0.235‡	-0.232‡	-0.220‡
Log(Price)* Log(Income)	0.808‡	0.822‡	0.799‡	0.746‡
Number Railway stations	-0.0014	-0.0032	-0.0080	0.0053
Number Subway stations	0.011*	0.011*	-0.0013	-0.0011
Average travel time from j, commuting (TC) [hour]	-0.251*	-0.177	-0.493‡	-0.635‡
TC*(Dummy female) [hour]	-0.108	-0.137	-0.102	-0.066
Average travel time from j, by private car (VP) [hour]	-0.101	0.012	0.263	0.219
Distance to highway [10 ³ km]	-4.814*	-5.351*	-2.922	-3.238
% households with 1 member * 1 member in h	3.246‡	3.325‡	3.660‡	3.769‡
% households with 2 members* 2 members in h	0.137	0.257	-0.082	-0.591
% households with 3+ members* 3+ member in h	2.962‡	2.887‡	2.627‡	2.523‡
% hh with no working member * no working member in h	6.384‡	6.793‡	6.940‡	6.489‡
% hh with 1 working member * 1 working member in h	-0.118	0.016	0.537	0.855
% hh with 2+ working member * 2+ working member in h	-0.112	-0.234	-0.758*	-0.890*
% poor households * h poor	1.148*	1.337†	1.412†	1.459†
% households with a foreign head * foreign head in h	5.140‡	4.628‡	4.771‡	5.125‡
% households with a foreign head * French head in h	-3.849‡	-4.327‡	-4.026‡	-3.575‡
Employment [1,000]	1.951*	3.289‡	3.913‡	0.263
Density (Population/Surface) [1000 persons/km]	0.661	2.167	3.079	-6.625*
Log(Population)	0.211‡	0.377‡	0.728‡	1.011‡
% change in population, 1990 to 1999	0.201†	0.232‡	0.185*	0.185†
Pseudo R ²	14.4%	15.7%	21.3%	30.6%

*: significant at the 10% level; †: significant at the 5% level; ‡: significant at the 1% level

⁷In this case, repetitions of the same dwelling in the same individual choice set should be excluded, but the probability of repetition in a set of more than 5 million dwellings is negligible.

The first noteworthy result is that the variations among the coefficients as a function of the number of alternatives are very limited for the coefficients which are significant. This is consistent with the theory that importance sampling leads to unbiased coefficients when a correcting term (size variable) is introduced in the model.⁸ On the other hand, the insignificant coefficients tend to be very unstable (e.g. number of stations, commuting travel time for women and travel time or distance to highway). This conforms with expectations and probably results from the large correlation between these variables. However, the impact of commuting travel time becomes significant in larger samples. The coefficient of the “log(Population)” significantly varies as a function of the number of alternatives.

The second notable result is that the explanatory power dramatically increases with the number of alternatives (from 14.4% to 30.6%). This is consistent with the fact that the coefficient of the size term “Log(Population)” becomes larger and more significant. It is possible that these two results are due to the range in size of communes from extremely sparsely populated to quite large, which in smaller samples of alternatives would tend to under-represent the smaller communes. In larger samples of alternatives smaller communes would be more likely to be represented in the choice set, and could be distinguished by "Log(Population)", increasing the range of the alternatives, and improving the overall capacity to distinguish location preferences within this range.

4.3.2 Sampling households

We now turn to the sensitivity of the model to the sampling ratio for households (rather than sampling of alternatives). Table 7 presents the results with 100 alternatives for each household, when the sampling rate for households varies between 0.5% and 5% of the population of moving households. The first column is the same as in Table 6. Computing time is significantly affected by household sampling rate, especially when dataset size is above RAM capacity (2 Gb), and we were not able to increase sample size beyond this level.

The coefficients which were significant in the 0.5% sample generally remain significant and stable. Some coefficients (e.g. Paris dummy, or density) which were not significant in the 0.5% sample tend to become more significant and stable at higher household sampling rates. This is consistent with the fact that the variance of a coefficient is inversely proportional to sample size. Other coefficients (e.g. travel times) remain insignificant, suggesting that the corresponding variable has really no impact, *ceteris paribus*.

Table (7): Impact of sampling rate on residential location choice (100 alternatives)

⁸In the case of importance sampling, the coefficient of this correcting term should be equal to 1 (see [3]). However, the size variable can also enter preferences. Its coefficient can therefore be estimated freely, since it merges preferences, fitting considerations (see above) and correcting factor.

Variable \ Household Sampling Rate	0.5%	1%	2%	3%	4%	5%
Same district as before move	2.578‡	2.601‡	2.599‡	2.581‡	2.588‡	2.577‡
Paris	0.079	-0.126*	-0.055	-0.073*	-0.082*	-0.082‡
Log(Price)	-7.847‡	-8.055‡	-7.422‡	-6.669‡	-7.029‡	-6.849‡
Log(Price)* (Age-20)/10	-0.233‡	-0.162‡	-0.146‡	-0.121‡	-0.128‡	-0.126‡
Log(Price)* Log(Income)	0.808‡	0.812‡	0.748‡	0.673‡	0.709‡	0.689‡
Number Railway stations	-0.0014	-0.016*	-0.011*	-0.012‡	-0.017‡	-0.013‡
Number Subway stations	0.011*	0.018‡	0.018‡	0.016‡	0.016‡	0.015‡
Average travel time from j, commuting (TC) [hour]	-0.251*	-0.109	0.003	0.046	0.046	-0.001
TC*(Dummy female) [hour]	-0.108	-0.173	-0.204*	-0.246‡	-0.260‡	-0.204
Average travel time from j, by private car (VP) [hour]	-0.101	0.042	-0.029	0.050	0.083	0.146*
Distance to highway [10 ³ km]	-4.814*	-4.674*	-4.201‡	-2.825‡	-4.813‡	-3.437‡
% households with 1 member * 1 member in h	3.246‡	3.456‡	3.181‡	3.124‡	3.038‡	3.456‡
% households with 2 members* 2 members in h	0.137	-1.046	-1.806‡	-1.512‡	-1.954‡	-1.046
% households with 3+ members* 3+ member in h	2.962‡	3.253‡	3.234‡	3.372‡	3.273‡	3.253‡
% hh with no working member * no work. memb. in h	6.384‡	7.633‡	7.805‡	7.391‡	8.366‡	7.633‡
% hh with 1 working member * 1 work. memb. in h	-0.118	0.807*	0.620*	0.843‡	0.973‡	0.807*
% hh with 2+ working member * 2+ work. memb. in h	-0.112	-0.328	-0.364*	-0.360*	-0.200	-0.328
% poor households * h poor	1.148*	0.916‡	1.246‡	1.360‡	1.484‡	0.916‡
% households with a foreign head * foreign head in h	5.140‡	3.598‡	3.190‡	3.647‡	3.646‡	3.598‡
% households with a foreign head * French head in h	-3.849‡	-3.947‡	-3.709‡	-3.721‡	-3.873‡	-3.947‡
Employment [1,000]	1.951*	1.405*	1.397‡	1.604‡	1.750‡	1.947‡
Density (Population/Surface) [1000 persons/km]	0.661	4.568*	4.528‡	3.781‡	3.560‡	4.729‡
Log(Population)	0.211‡	0.228‡	0.228‡	0.244‡	0.242‡	0.228‡
% change in population, 1990 to 1999	0.201‡	0.255‡	0.251‡	0.298‡	0.291‡	0.255‡
Pseudo R ²	14.4%	14.6%	14.4%	14.3%	14.4%	14.3%

*: significant at the 10% level; †: significant at the 5% level; ‡: significant at the 1% level

In the largest samples, both coefficients associated with the numbers of stations become significant. The number of subway (metro) stations in a commune increases the probability of location but the number of railway stations decreases it, after accounting for transit accessibility and other effects. These results may reflect the relative effects of positive and negative externalities associated with subway stations and railway stations. Metro stations are more likely than railway stations to be located within clusters of shopping and service employment or adjacent to major cultural attractions, and railway stations are larger and may be more likely to have negative localized externalities on the immediate neighbourhood, such as traffic, noise, and possibly petty crime.

4.4 Location choice with capacity constraints

4.4.1 Estimation of a more parsimonious model with and without constraints

In order to estimate the effect of the availability constraints based on the yearly demand, it is necessary to infer an annual supply of dwellings from the current supply observed in March 1998. This annual supply depends on the λ ratio defined in Section 2.

As a preliminary sensitivity analysis, we consider two cases: $\lambda = 0.5$ and $\lambda = 0.45$. The results obtained when estimating a more *parsimonious* model (with only 5 explanatory variables) on a 2% sample of households with all alternatives are reported in Table 8.⁹ The “initial” estimates corresponds

⁹The dataset size would prevent from estimating the full model on the same sample since four times more variables roughly means four times more RAM necessary.

to the model estimated assuming that the system is not constrained (iteration $l' = 0$ in Algorithm U), whereas the “final” estimates correspond to (iteration $l' = \infty$ in Algorithm U).

Since random sampling introduces heterogeneity in the individual allocation ratios Ω^i , the homogeneous model can only be estimated using the universal set of alternatives (1,300 municipalities). However, the dataset with universal choice set and 100% of the movers would contain 766 million lines and its size would be over 2 terabytes with all the variables used in the previous section! In order to illustrate the method on a more tractable data set, we used a representative random sample of 2% households, and we selected a small number of explanatory variables (Log(Price); Average travel time from j, commuting (TC); % households with a foreign head; Log(Population); and Same district as before move in the heterogeneous case). This selection is based on results obtained in the previous sections and on other results not reported here in order to save space. With this parsimonious model, the explanatory power is still reasonable in the heterogeneous case, and sample size is tractable (15,321,800 lines, 11,786 movers and slightly more than 1 Gb). Although convergence is attained after a small number of iterations (less than 5 iterations for constraints at each estimation loop, less than 10 demand loops), about 2 hours are necessary for estimating the homogeneous version of the model, and computing time goes over 15 hours for the heterogeneous version model.¹⁰ In the future, working on a more powerful computer and improving the efficiency of the algorithm will allow increasing significantly both the sampling rate and the number of explanatory variables. A trade-off between the number of households, the number of alternatives by individual and the number of explanatory variables will remain, however.

The overall fit of this parsimonious model is quite good given the low number of explanatory variables: in the 2% sample with all alternatives, the Pseudo- R^2 is nearly 30% in the heterogeneous case and 23% in the homogeneous case. The unexpected positive sign of the price coefficient reflects an omission bias: some alternative attributes (e.g. % poor households) are correlated both with price and with demand. Because of this omission bias (which is corrected by the introduction of additional explanatory variables, as in the full model), we will not interpret the value of the price coefficient. The extreme sensitivity of the price coefficient to omission bias has already been noticed by Guevara and Ben-Akiva [7]. In addition, results not reported here show that the price coefficient value and significance level are very different in a 1% sample of households, and even more different in the full model. The most significant variable remains the dummy “same district as before move” and the corresponding estimated coefficient is stable across sample sizes. Strangely enough, in this parsimonious specification, the price coefficient is unchanged when the factor correcting for constraints is introduced.

We could anticipate identification problems because both price and correcting factor $\log(\pi_j^i)$ reflect the supply/demand ratio in some sense: $\pi_j^i = S_j/D_j$ in the alternatives constrained ex post and equilibrium on the dwelling market suggests that price depends on the supply/demand ratio. However, the correlation between $\log(\text{Price})$ and supply/demand ratio is only -4.85% (significant at the 10% level but not at the 5% level) for initial demand estimates and -1.12% (not significant at all), suggesting that prices are only marginally influenced by the frictions in the housing market. In other words, prices do not actually clear the market, a conclusion we can only generalize at this point to the Ile-de-France housing market. Note that the price coefficient is not significant in the 1% sample and only marginally significant in the 2% sample. However, the price coefficient becomes more significant in larger samples and negative when enough covariates are added, as in the full model. In the parsimonious model, the coefficients of the accessibility variable (average travel time from the commune) and of the size variable ($\log(\text{Population})$) are stable across sample sizes and between initial and final estimates. By contrast, the fraction of foreign households is more sensitive both to the sample size and to the introduction of constraints.

Table (8): Comparison of results in the homogeneous and heterogeneous cases (2% sample)

¹⁰Using SAS software on an IBM laptop with 2 Gb RAM.

Variable		Homogeneous case Coefficient (t-stat)	Heterogeneous case Coefficient (t-stat)
Same district as before move	initial, $\lambda = 0.45$	does not apply	2.5765 (105.71)
	final, $\lambda = 0.45$	does not apply	2.5890 (106.05)
	initial, $\lambda = 0.5$	does not apply	2.5765 (105.71)
	final, $\lambda = 0.5$	does not apply	2.5800 (105.93)
Log(Price)	initial, $\lambda = 0.45$	0.0280 (1.67)	0.0219 (1.30)
	final, $\lambda = 0.45$	0.0321 (1.92)	0.0268 (1.59)
	initial, $\lambda = 0.5$	0.002501 (0.11)	0.0219 (1.30)
	final, $\lambda = 0.5$	0.000250 (0.01)	0.0260 (1.54)
Average travel time from j	initial, $\lambda = 0.45$	-0.2443 (-2.27)	-0.2876 (-2.64)
	final, $\lambda = 0.45$	-0.1392 (-1.28)	-0.1999 (-1.82)
	initial, $\lambda = 0.5$	-0.3170 (-2.07)	-0.2876 (-2.64)
	final, $\lambda = 0.5$	-0.3302 (-2.14)	-0.2567 (-2.34)
% households with a foreign head	initial, $\lambda = 0.45$	-0.4483 (-2.28)	-0.5612 (-2.77)
	final, $\lambda = 0.45$	-0.8548 (-4.21)	-0.9530 (-4.57)
	initial, $\lambda = 0.5$	-0.9955 (-3.46)	-0.5612 (-2.77)
	final, $\lambda = 0.5$	-1.0966 (-3.71)	-0.8345 (-4.05)
Log(Population)	initial, $\lambda = 0.45$	1.0846 (105.27)	1.0722 (100.23)
	final, $\lambda = 0.45$	1.0590 (98.82)	1.0495 (94.76)
	initial, $\lambda = 0.5$	1.0807 (105.27)	1.0722 (100.23)
	final, $\lambda = 0.5$	1.0733 (68.33)	1.0496 (94.23)
Log(π_j^i)	final, $\lambda = 0.45$	0.5157 (8.27)	0.3888 (7.88)
	final, $\lambda = 0.5$	0.7848 (7.22)	0.7848 (7.22)
Pseudo-R ²	initial, $\lambda = 0.45$	22.90%	29.68%
	final, $\lambda = 0.45$	22.94%	29.72%
	initial, $\lambda = 0.5$	23.15%	29.84%
	final, $\lambda = 0.5$	23.19%	29.92%

We are also interested in the sensitivity of the results to the sampling ratio and compare the results obtained on a 1% and on a 2% representative samples of households. Because of constraints on the data sets size, we only explore the parsimonious models (see Table 8). The “Same district as before move” dummy is used in the heterogeneous case, but not in the homogeneous one.

4.4.2 Constraints and allocation ratio in the parsimonious model

We turn now to an assessment in the homogeneous case (parsimonious model) of the effect on the fraction of alternatives constrained and on the allocation ratio of the sampling rate of households and of λ . Table 9 presents the results for 1% and 2% samples of households, using $\lambda = 0.5$ and $\lambda = 0.45$.

Table (9) Fraction alternatives constrained and allocation ratio, homogeneous case

	$\lambda = 0.5, 1\%$	$\lambda = 0.5, 2\%$	$\lambda = 0.45, 2\%$
% alt. constrained:			
Ex ante, initial demand estimates	36.2%	40.2%	47.31%
Ex post, initial demand estimates	39.3%	43.5%	52.69%
Ex ante, final demand estimates	46.5%	51.5%	61.77%
Ex post, final demand estimates	51.2%	56.2%	70.85%
Average [min;max] S_j/D_j ratio:			
Initial demand estimates	1.18 [0.06; 5.48]	1.14 [0.06; 4.97]	1.09 [0.06; 4.94]
Final demand estimates	1.08 [0.01; 5.26]	1.03 [0.03; 4.64]	0.91 [0.00; 4.53]
Allocation ratio Ω :			
Ex ante, initial demand estimates	1.019	1.021	1.037
Ex post, final demand estimates	1.050	1.039	1.127

For all values of λ , we have the following result: the fraction of constrained alternatives is always larger once algorithm K has converged. This is true when the initial demand is considered and when the final demand is considered. For the final demand, the fraction of constrained alternatives is (almost always) larger than for the initial demand at any stage of algorithm K. In other words, convergence of the inner or outer loop increases the fraction of constrained alternatives. The reason behind this intuition is as follows: at the initial demand loop, the observed demand under-estimates the ex ante demand in constrained alternatives and over-estimates ex ante demand in unconstrained alternatives. These biases are corrected at the final demand loop. As expected, ceteris paribus, when λ decreases, the level of constraints increases since a lower value of λ means a smaller number of dwellings available on average during one year. Note that a very small variation of λ induces an important variation of the fraction of alternatives constrained.

The average ratio of supply to demand, which is inversely proportional to the degree of constraint, decreases from the initial demand to the final demand (since this increases the amount of constraints). The same reasoning applies to Ω , which is an increasing function of the level of constraints. According to the final estimates, when $\lambda = 0.5$, 4% to 5% of the households are not located in their preferred commune. Similar results are obtained in the heterogeneous case with the parsimonious model. The results for the heterogeneous case (not reported here) are qualitatively similar to the ones of the homogeneous case in terms of the fraction of alternatives constrained and the supply over demand ratio. Note that the allocation ratio in the heterogeneous case is specific either to the household or to the alternative. When $\lambda = 0.5$, the average individual-specific allocation ratio is approximately equal to the homogeneous allocation ratio, but this is no more true when $\lambda = 0.45$. In that case, the fraction of households not located in their preferred commune increases to 13% when $\lambda = 0.45$.

4.4.3 Extent of the constraints and allocation ratio in the full model (heterogeneous case)

Here we examine in the heterogeneous case the effect on the fraction of alternatives constrained and on the allocation ratio of the sampling rate of households and of alternatives. Table 10 presents the results in the case when demand is estimated using the full model (with 24 explanatory variables) and a ratio λ specific to the district, in place of the invariant λ used previously. Note that the model could not be estimated on the 1300 alternatives/0.5% sample because it ran out of memory after the estimation of the initial demand. The results are significantly different from those obtained in the previous section, and are highly sensitive to the number of alternatives used for the estimation. The fraction of alternatives constrained ex ante with initial demand estimates is also sensitive to the sampling ratio, but the fraction of alternatives constrained ex post with final demand is less sensitive to the sampling ratio. However, the constraints are not very severe in most of these alternatives, and the average allocation ratio remains close to 1.13 ex ante and 1.30 ex post in all sample sizes with 100 alternatives, which means that 30% of households are not located in their preferred commune. This suggests that constraints are more severe

than in the previous section. This is consistent with the fact that the λ ratio specific to the district is on average less than 0.45, indicating that the constraints are a priori more severe than in the previous section.

Table (10) Fraction alternatives constrained and allocation ratio, heterogeneous case

Nb.alt. sampling rate	500 0.5%	200 0.5%	100 0.5%	100 1%	100 2%	100 3%	100 4%	100 5%
% alt. constr.:								
ex ante, initial	27.0%	29.5%	40.3%	38.7%	43.8%	38.9%	38.2%	40.9%
ex post, initial	43.7%	50.8%	57.7%	59.2%	63.0%	58.5%	59.2%	60.4%
ex ante, final	34.8%	37.7%	45.8%	44.2%	48.5%	41.8%	43.0%	46.2%
ex post, final	62.5%	69.1%	67.9%	68.8%	75.5%	69.2%	71.7%	73.8%
Average Ω^i :								
ex ante, initial	1.167	1.179	1.138	1.130	1.136	1.134	1.129	1.131
ex post, final	1.478	1.444	1.314	1.302	1.311	1.303	1.303	1.321
[min;max] Ω^i :								
ex ante, initial	[1.07; 1.35]	[1.06; 1.55]	[1.02; 1.74]	[1.02; 1.86]	[1.01; 2.06]	[1.01; 2.54]	[1.01; 1.95]	[1.01; 1.95]
ex post, final	[1.00; 3.70]	[0.91; 5.35]	[0.87; 5.25]	[0.83; 7.38]	[0.61; 8.02]	[0.23; 13.4]	[0.44; 9.15]	[0.42; 12.3]
Average Ω_j :								
ex ante, initial	1.194	1.199	1.159	1.157	1.169	1.157	1.153	1.156
ex post, final	1.717	1.636	1.441	1.442	1.472	1.431	1.445	1.482
[min;max] Ω_j :								
ex ante, initial	[1.13; 1.25]	[1.08; 1.41]	[1.03; 1.55]	[1.08; 1.43]	[1.09; 1.39]	[1.09; 1.34]	[1.08; 1.33]	[1.10; 1.31]
ex post, final	[1.23; 2.43]	[−; 3.06]	[−; 4.49]	[1.08; 3.45]	[1.11; 2.35]	[1.14; 2.56]	[1.14; 2.36]	[1.16; 3.05]

4.4.4 Comparison of constrained and unconstrained aggregate demand

We now turn to the aggregate results and compare the demand predicted by the different models in each commune to actual (observed) demand, that is the number of movers observed locating in each commune in 1998. At a first glance, the correlations between aggregate demand (D_j) estimated in the different specifications and samples is very high. However, the precision of aggregate demand deteriorates in small municipalities when sample size becomes too small (see Section 4.4.4).

Table 11 reports the correlation coefficient between the actual demand (number of households which moved in to each commune in 1998) and the demand predicted by the different models. The “Full model” was estimated using a random sampling of 8 alternatives for each household, with 24 explanatory variables, as reported in Table 5. The “Parsimonious model” was estimated using the same random sampling of 8 alternatives, but with only 5 explanatory variables, as reported in Table 6. The estimates with constraints are those of the homogeneous case with $\lambda = 0.5$. Table 11 shows that the precision of the aggregate estimated remains satisfactory when the sampling ratio decreases. It is interesting to notice that precision is not significantly altered in the parsimonious model. It seems that the precision becomes poorer (correlation less than 90% in the 1% sample, less than 93% in the 2% sample) when the “size” (measured as the number of observed moves into the commune in 1998) of the alternatives decreases. The correlation with actual demand remains quite high in the 1148 municipalities with less than 1000 movers in 1998, but it diminishes in the 1054 municipalities with less than 500 movers in 1998 and becomes even lower in smaller alternatives. However, precision significantly improves in the 1% and 2% samples of households when a universal set of alternatives is used for the demand predictions. Note that with a universal set of alternatives the precision is the same in the 1% and 2% samples of households, whereas, with a random sampling of 8 alternatives, precision was far poorer in the 1% sample compared to the 2% sample of households (in the small municipalities). Integrating the constraints significantly improves the precision of the demand estimates, especially for small alternatives.

Table (11): Correlation between actual number of movers and demand predicted by different models

Model / sampling rate / Nb alt.	Size of Commune: Number of Movers in 1998					
	All alt.	< 1000 1148	< 500 1054	< 200 948	< 100 841	< 50 725
Full / 100% / 8	0.9979	0.9941	0.9893	0.9764	0.9599	0.9205
Parsimonious / 100% / 8	0.9968	0.9877	0.9765	0.9626	0.9444	0.8979
Full / 20% / 8	0.9979	0.9919	0.9855	0.9672	0.9329	0.8726
Parsimonious / 20% / 8	0.9967	0.9864	0.9726	0.9535	0.9206	0.8595
Full / 2% / 8	0.9968	0.9735	0.9398	0.8701	0.7441	0.6258
Parsimonious / 2% / 8	0.9958	0.9702	0.9361	0.8878	0.7775	0.6702
Full model, 1% / 8	0.9958	0.9554	0.8952	0.7659	0.6180	0.5124
Parsimonious / 1% / 8	0.9950	0.9531	0.8986	0.7935	0.6563	0.5529
Parsimonious / 2% / all	0.9953	0.9803	0.9635	0.9427	0.9218	0.8582
<i>Parsimonious / 2% / all, with constraints</i>	<i>0.9955</i>	<i>0.9834</i>	<i>0.9722</i>	<i>0.9625</i>	<i>0.9426</i>	<i>0.9005</i>
Parsimonious / 1% / all	0.9953	0.9799	0.9624	0.9397	0.9176	0.8582
<i>Parsimonious / 2% / all, with constraints</i>	<i>0.9997</i>	<i>0.9828</i>	<i>0.9712</i>	<i>0.9610</i>	<i>0.9413</i>	<i>0.9005</i>

5 Conclusion and extensions

Constraints on the availability of alternatives are clearly present in some markets, such as the housing market in the Ile-de-France. We have demonstrated that housing prices do tend to capitalize the imbalances between demand and supply in the short run period of one year, but do not fully accomplish the task of resolving short-term market disequilibrium. The specification linking residential location choices and housing prices through the aggregation of individual demands provides a tractable means to address this economic endogeneity issue, and our results demonstrate that the market clearing role for prices is significant. However, even after accounting for the price effect, a large fraction of alternatives still had excess demand, imposing availability constraints on consumers that ultimately must make suboptimal choices from unconstrained alternatives. This is a significant empirical finding, since virtually all previous literature in discrete choice modeling in market contexts implicitly or explicitly assume that prices clear the market.

We have developed and demonstrated a viable mechanism to address the influence of constrained alternatives on choice outcomes, by using relatively weak assumptions consistent with the property of independence of irrelevant alternatives inherent in the multinomial logit model. Basically, we assume that (1) if an individual prefers an alternative j , which is unconstrained ex post, he can be sure to be allocated to it, and (2) the individual allocation ratio is the same for all individuals. We then develop a discrete choice model specification and computational algorithm that corrects for capacity constraints in a way that is consistent with the IIA assumption. It can be shown that this algorithm converges to a unique equilibrium in the case of homogeneous agents. Though we have not proven the existence and uniqueness of an equilibrium solution in the heterogeneous case, simulation results suggest that convergence and uniqueness occur. We leave the theoretical proofs for future research.

Our sensitivity testing confirms that most parameter estimates in the unconstrained model are fairly robust to sampling of households, down to quite small proportions such as 1 to 2%, where precision decreases. Due to the size of the data and the use of a full enumeration of 1300 alternatives in the heterogeneous case, we were limited to using small sampling rates for households for testing a very parsimonious model that clearly reflects omitted variables bias, but allowed the initial testing of the algorithm to proceed. The results of this testing confirm that the algorithm converges after a small number of iterations, and that parameters can be adjusted for the bias resulting from capacity constraints.

Application of this constrained discrete choice algorithm shows that the share of alternatives initially estimated to be constrained significantly increases once the parameter estimates are adjusted to correct

for the effect of the constraints. In most of our sensitivity tests, the percentage of alternatives constrained increased by 30% or more from the initial ex ante percentage constrained. Finally, our analysis of demand predictions confirms that the constrained discrete choice algorithm significantly improves predictions of demand, or market shares, particularly for those alternatives with relatively small market share. This may have important implications for research in which the alternatives of interest reflect small market shares, and the system contains some capacity constraints – even if the alternatives of interest are not themselves constrained ex ante.

6 Acknowledgements

This research has been funded in part by PREDIT, the French Department of Transportation (contract awarded to IAURIF, University of Cergy-Pontoise and adpC). We would like to thank Dany Nguyen (IAURIF), for helpful comments and for providing us with the data for the Ile-de-France region. We benefited from the comments of the seminar participants of the second Khumo conference (in particular from Esko Niskanen, Eduardo Marcucci, Robin Lindsey and Stef Proost), of the Dijon conference organized by Alban Thomas, and from the Fiesta conference participants at the Ecole Nationales des Ponts et Chaussées (in particular Fabien Laurent, Dominique Swartz and Joel Maurice). We would like also to thank the participants of the seminar organized in the honour of Ken Small, and in particular Chen Hsiao and Ken Train for extremely helpful comments and suggestions. Finally, the sampling methods were discussed with Daniel McFadden and Moshe Ben-Akiva, who provide us with useful advice. The third author wishes to acknowledge National Science Foundation Grants EIA-0121326 and IIS-0534094, and Environmental Protection Agency Grant R831837.

References

- [1] A. Anas and J. Cho. A dynamic, policy oriented model of the regulated housing market; the Swedish prototype. *Regional Science Urban Economics*, 18:210–231, 1988.
- [2] S. Anderson, A. de Palma, and J. Thisse. *Discrete Choice Theory of Product Differentiation*. MIT Press, 1993.
- [3] M. Ben-Akiva and S. R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press, 1985.
- [4] A. de Palma and C. Lefèvre. A probabilistic search model. *Journal of Mathematical Sociology*, 8:43–60, 1981.
- [5] A. de Palma, K. Motamedi, N. Picard, and P. Waddell. A model of residential location choice with endogenous housing prices and traffic for the Paris region. *European transport*, page Under press, 2006.
- [6] A. de Palma and J. Rouwendal. Availability constraints in the housing market. *Journal of housing economics*, 5:105–132, 1996.
- [7] C. A. Guevara and M. Ben-Akiva. Endogeneity in residential location choice models. Unpublished manuscript, 2006.
- [8] S. Lerman. Location, housing, automobile ownership, and the mode to work: A joint choice model. *Transportation Research Board Record*, 610:6–11, 1977.
- [9] D. McFadden. Modelling the choice of the residential location. In A. K. et Al., editor, *Spatial Interaction Theory and Residential Location*, pages 75–96. North Holland, Amsterdam, 1978.

- [10] J. Quigley. Housing demand in the short-run: An analysis of polytomous choice. *Explorations in Economic Research*, 3:76–102, 1976.
- [11] G. Weisbrod, S. Lerman, and M. Ben-Akiva. Tradeoffs in residential location decisions: Transportation versus other factors. *Transport Policy and Decision Making*, 1:13–26, 1980.
- [12] R. C. J. Williams. A logit model of demand for neighborhood. In D. Segal, editor, *The Economics of Neighborhood*, pages 17–42. Academic Press, New York, 1979.