ESSEC
BUSINESS SCHOOL

ROBUST LOCATION OF NEW HOUSING DEVELOPMENTS USING A CHOICE MODEL

RESEARCH CENTER
ESSEC WORKING PAPER 1521

2015

Juan Carlos Espinoza
Garcia Laurent Alfandari

L'esprit pionnier

Robust location of new housing developments using a choice model

Juan Carlos Espinoza Garcia Laurent Alfandari

October 26, 2015

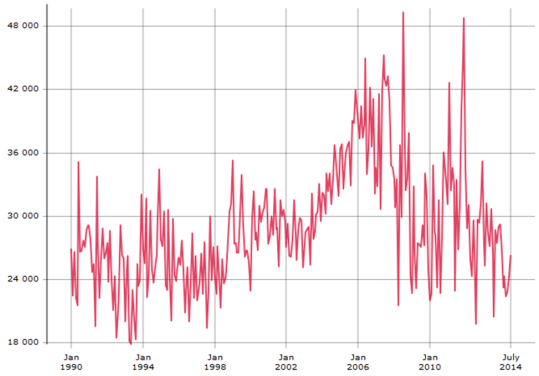
Abstract

We consider the issue of choosing a subset of locations to construct new housing developments maximizing the satisfaction of potential buyers. The allocation of demands to the selected locations is modeled by a choice model, based on the distance to the location, real-estate prices and incomes. We study two robust counterparts of the optimal location problem, where uncertainty lies on demand volumes for the first one, and on customer preferences for the second one. In both cases, the parameters subject to uncertainty appear both in the objective function and constraints. The second robust model combines a scenario-based approach with nominal, price-centric and distance-centric scenarios on customers preferences, and an uncertainty budget approach that limits the number of cities that can deviate from the nominal scenario. Computational experiments are conducted on instances of the Paris region to analyze the tractability of the problem and its robust counterparts, and derive insights for the new housing development issue.
Keywords Robust Optimization, Multinomial Logit Choice Models, Facility Location, Housing.

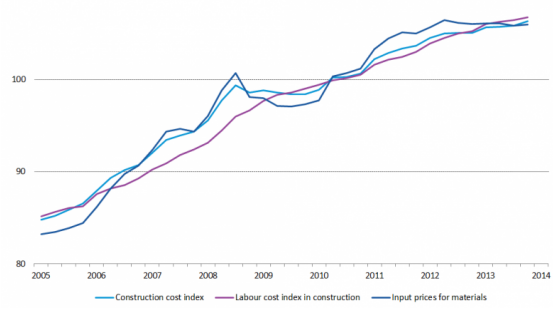
1 Introduction

Housing supply is a complex problem arising in multiple and diverse contexts. Driant (2011), Gilbert (2011), and Sandoval (2012) all provide examples of different problematic situations related to housing supply. In France, housing supply is an important issue with a shortfall of 800,000 to 1 million dwellings in 2012 Driant (2012). While housing policies mainly consist in building new housing, (Driant, 2011) addresses the existence of a mismatch between demand and offer.

At the same time, investments in new residential buildings have been increasing. To cite an example, in the U.S., the value of residential construction by the private sector alone has increased from $208 B in 1993 to $336 B in 2013 according to information from the United States Census Bureau (2014). In France, the National Institute of Statistic and Economic Studies (INSEE) and Eurostat, the statistical office of the European Union, paint a similar picture: a somewhat stable number of new dwellings and constantly increasing construction costs as shown in Figures 1a and 1b.



(a) New dwellings built in (continental) France 1990-2014



(b) Quarterly construction cost and its components gross 2005-2014

Figure 1: New buildings and construction costs statistics

Failing efforts at solving this problem see a proliferation of inadequate housing encouraged by the rising prices and the social policies which exacerbate the original issue, prompting the need for a better solution as to what type of and where to build new housing units (Driant, 2011). The issue of choosing locations for new housing developments in a network of possible sites can be seen as a more general facility location problem. The allocation of customers to the selected locations and the rules governing said allocation play a fundamental role in the solution of the problem.

Contrary to the supply chain sector where the decision-maker can assign supply sites to customers who are generally indifferent to the sourcing, in housing development the customers actually choose themselves their location to buy a flat, which naturally conducts to use a choice model based on consumers' utilities and willingness to buy (see the seminal paper of (Guadagni (1983)) for the Multinomial Logit(MNL) choice model).

Choice models have already been used in the context of housing supply, but not in an optimization framework to our knowledge. Bulent and Kenneth (2005) and Pagliara et al (2010) propose the use of scoring methods to address the decision of buyers in the housing context. Pagliara et al (2010) use a Stated Preference model to estimate the willingness to pay of a specific buyer for each alternative available, which is then used to estimate the probability of choosing a housing unit. Similarly, Bulent and Kenneth (2005) propose a discrete choice model to compute this probability. Additionally, de Palma et al (2007) develop a model for residential location choice where capacity constraints are considered.

When dealing with optimization and especially facility location, Haase and Müller (2014) have tested different linearizations of the multinomial logit model (MNL) for the problem of selecting a subset of r sites maximizing coverage, i.e. the part of the demand that does not go to competitors. They found the formulation of Aros-Vera et al (2013), which extends the simple formulation of Haase (2009), to achieve the best performance. Later, in Müller and Haase (2014) they extend this model to include customer segmentation in the retail industry. They used segmentation to reduce the predictive bias of the MNL. In Haase (2009), a linearization of the Independence of Irrelevant

Alternatives (IIA) property of choice models was also proposed for the case of binary variables associated with opening a site or not.

When positioning our paper compared to the work of Haase and Muller, beyond our specific new housing context we can distinguish two main features:

(i) Their work deals with maximizing the total demand covered, where part of the demand can flow to competitors. In our work all demands flow to the set of sites of the new housings developer as we consider well-identified customers engaged in a program, but to avoid drop-outs we maximize the average satisfaction of potential buyers.

(ii) We introduce a robustness approach where uncertainty lies on customers' demands and preferences. Due to (i), the robust model has the particularity that the demand parameters subject to uncertainty appear both in the constraints and the objective function.

Robustness approaches are natural for strategic facility location since decisions have a long term impact which is subject to uncertainty. The facility location literature typically considers uncertainty in demand quantities and service costs (Alumur et al, 2012; Baron et al, 2011; Gülpinar et al, 2013; Snyder, 2006). Snyder (2006) provides an extensive review of the literature on facility location under uncertainty where problems are classified according to the type of objective function stochastic location problems (mean-outcome models, probabilistic guaranties, where information on the underlying probability distribution of the uncertainty is required), and robust location problems.

Baron et al (2011) and Gülpinar et al (2013) deal with a robustness formulation of facility location problems. Baron et al (2011) model a location problem under box and ellipsoidal uncertainty, comparing the results in terms of number of facilities used, capacity assigned to them, and profit change w.r.t. the nominal solution. The authors also test different distribution functions for the demand and different conservativeness of the solution. Gülpinar et al (2013) address the issue of uncertainty in the distribution functions and asymmetric uncertainty sets. The authors explore the impact of the estimated moments of the distributions and that of the cost parameters. Alumur et al (2012) deal with the uncertainty in costs using a minimization of regret approach over a set of scenarios while the uncertainty in demands is handled using stochastic optimization. As mentioned before, this paper focuses on demand uncertainty.

As for robust optimization approaches in general, let us cite the minimax approach (Bertsimas and Sim, 2004), minimax regret, p-robustness (Gabrel et al, 2013) and the review of (Gabrel et al, 2014). See also the papers of Billionnet, Costa, and Poirion (2014) and Gabrel and Murat (2010) dealing with uncertainty on the right hand-sides of constraints (column-wise uncertainty instead of row-wise). Recently, Denoyel et al (2014) also studied a robust selection problem for healthcare using a choice model, but using a fractional formulation instead of linear formulation of the IIA property.

The paper is structured as follows. Section 2 presents the optimization problem and the underlying choice model for customers' willingness to buy in a given location. Section 3 describes the robust counterparts of the location problem, where uncertainty lies first on demand volumes, then on customers preferences. Section 4 reports numerical results on a study in the Paris region and provides an analysis of the results as well as managerial insights for new housing development.

2 Problem statement

We consider a set of demand sources indexed by $i \in \{1, \ldots, n\}$, with D_i the number of potential buyers of source i interested in new housing, and $V = \sum_i D_i$ the total number of potential buyers. Note that indices i can aggregate multiple smaller demand sources, in order to keep the optimization problem tractable. The set of possible locations for constructing new housing developments is indexed by $j \in \{1, \ldots, m\}$, with a variable cost c_j per housing unit built in location j (which is mainly a construction cost but can include other variable costs such as local taxes and management costs). The total cost of constructed housings should not exceed a given budget B, and the number

of selected locations should be no more than a given bound P, to avoid providing potential buyers a plethoric offer and to limit high fixed costs associated with opening multiple programs. Finally, the choice model that computes the volumes flowing from demand sources to selected locations, requires to know customer preference data, namely the utility u_{ij} of a potential buyer from demand source i for choosing location j. In a choice model where all locations would be open, the probability p_{ij} that i chooses location j is given by equation (1)

$$p_{ij} = \frac{e^{u_{ij}}}{\sum_{k \in J} e^{u_{ik}}} \tag{1}$$

where the utility u_{ij} aggregates criteria or attributes based on the characteristics of the housing unit, and the preferences of the potential buyer. Pagliara et al (2010) relate u_{ij} to the price of the housing unit, cost of traveling, location in the city, type of dwelling, among others. Bulent and Kenneth (2005) also consider the property tax of the location. As in the literature, we consider the utility as a linear combination of L locations characteristics and their relative value to the potential customers, so that

$$u_{ij} = \sum_{l=1}^{L} w_{ij}^{l} a_{ij}^{l} \tag{2}$$

where a_{ij}^{l} is the value of criterion l that a customer from demand source i perceives from choosing location j, and w_{ij}^{l} is a coefficient associated with criterion l, which can be a regression parameter or a weight. In this paper we mainly focus on price and distance attributes, as will be detailed later.

Since the subset of constructed locations is not known in advance, we introduce binary decision variables $y_j = 1$ if location j is selected for constructing new housing, 0 otherwise. The utilities are used to compute the distribution flows x_{ij}, i.e., the fraction of the demand of source i that flows from source i to location j:

$$x_{ij} = \frac{e^{u_{ij}} y_j}{\sum_{k \in J} e^{u_{ik}} y_k} \tag{3}$$

Distribution flow x_{ij}, as the ratio of two linear functions of variables y_j, is non linear. The New Housing Location Problem (NHLP) considered in this paper can be formulated as follows.

$$\max \quad \frac{1}{V} \sum_{i=1}^{n} \sum_{j=1}^{m} u_{ij} D_i x_{ij} \tag{4}$$

$$s.t. \quad \sum_{i=1}^{n} \sum_{j=1}^{m} c_j D_i x_{ij} \leq B \tag{5}$$

$$\sum_{j=1}^{m} y_j \leq P \tag{6}$$

$$\sum_{j=1}^{m} x_{ij} = 1 \qquad \forall i \tag{7}$$

$$x_{ij} = \frac{e^{u_{ij}} y_j}{\sum_{k \in J} e^{u_{ik}} y_k} \qquad \forall (i,j) \tag{8}$$

$$x_{ij} \geq 0 \qquad \forall (i,j) \tag{9}$$

$$y_j \in \{0,1\} \qquad \forall j \tag{10}$$

The objective function (4) maximizes the average utility of the potential buyers, in order to avoid drop-outs as much as possible. Constraints (5) and (6) are the budget constraints for the construction of the whole set of new developments. Constraints (7) assign the total demand of a source node to the selected locations, and distribution flows are modeled by (8). In the spirit of Haase, we use the Irrelevance of Independent Alternatives (IIA) property of choice models to linearize the above formulation as follows.

Proposition 2.1. *Non-Linear model (3)-(9) is equivalent to the following Linear Programming formulation NHLP:*

$$\max \quad \frac{1}{V}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ij} D_i x_{ij} \tag{11}$$

$$s.t. \quad \sum_{i=1}^{n}\sum_{j=1}^{m} c_j D_i x_{ij} \leq B \tag{12}$$

$$\sum_{j=1}^{m} y_j \leq P \tag{13}$$

$$\sum_{j=1}^{m} x_{ij} = 1 \qquad \forall i \tag{14}$$

$$x_{ij} \leq y_j \qquad \forall (i,j) \tag{15}$$

$$x_{ik} e^{u_{ij}-u_{ik}} + (2 - y_j - y_k) \geq x_{ij} \qquad \forall (i,j,k) | j \neq k \tag{16}$$

$$x_{ij} \geq 0 \qquad \forall (i,j) \tag{17}$$

$$y_j \in \{0,1\} \qquad \forall j \tag{18}$$

Proof. The linearization relies on the Independence of Irrelevant Alternatives (IIA) property of the MNL. IIA has been previously identified in the literature and simply states that equations (3) are equivalent to $x_{ij}/x_{ik} = e^{u_{ij}}/e^{u_{ik}}$ for all (j,k) such that $y_j = y_k = 1$. This can be formulated by linear constraints (15) and (16) similarly as in Haase (2009). Indeed, if $y_j = 0$ then $x_{ij} = 0$ by (15) so equations (3) are true for all (i,j) in that case. If $y_j = 1$, then for all k such that $y_k = 1$, by (16) we get $x_{ij}e^{u_{ik}} = x_{ik}e^{u_{ij}}$ so there exists $a > 0$ such that, given i, $x_{ik}/e^{u_{ik}} = a$ for all k. By summing over all k one obtains $a = \frac{\sum_k x_{ik}}{\sum_k e^{u_{ik}} y_k} = \frac{1}{\sum_k e^{u_{ik}} y_k}$ using (14). Since $x_{ij} = a\, e^{u_{ij}} y_j$ we get equation (8). □

Note that the NHLP model contains $m(n+1)$ variables and $O(nm^2)$ constraints of type (16). Section 4 will report numerical experiments to analyze the tractability of the above formulation, and the consistency of optimal locations with the attributes of the choice model, which will be described more precisely in that section. The next section presents a Robust Optimization (RO) approach for the New Housing Development Problem.

3 Robust Optimization formulations for the NHLP

We consider uncertainty on the demand data of Model NHLP, both volumes and customer preferences. It is quite common in the literature of Robust Optimization to consider variations in demands. Especially in new housing development, demands are more customers intents than real orders, as the locations are not chosen yet. Moreover, given the duration of the process from development planning to building/selling the units in it, said intentions are subject to change. Decisions that are robust to uncertainty on demands are thus more than needed for protecting oneself from possible variations. Before providing a robust formulation of NHLP, we briefly review the core of Robust Optimization in the sense of Bertsimas and Sim.

3.1 The Robust Optimization approach

The research by Bertsimas and Sim (2004) is the basis of this approach. This methodology has been previously used to address uncertainty in diverse contexts: Alvarez et al (2014) deal with yield uncertainty; Zokaee et al (2014) tackle a supply chain design with uncertainty in demands, costs, and supplies; Bohle et al (2010) solve a scheduling problem with uncertainty in productivity; and Bertsimas and Thiele (2006) address an inventory problem with uncertain demands.

In the remainder of this section we present Bertsimas and Sim's results as they apply to our problem. For a linear problem (P) of the form $\max \mathbf{c}'\mathbf{x} : \mathbf{Ax} \leq \mathbf{b}, \mathbf{x} \geq 0$, where wlog only A is subject to uncertainty, we note $\mathcal{A} = \{\tilde{A} \in R^{m\times n} | \tilde{a}_{ij} \in [a_{ij} - \hat{a}_{ij}, a_{ij} + \hat{a}_{ij}] \ \forall\, i, j,\ a_{ij}$ is the nominal value of the coefficient (i, j) of the constraint matrix, and $\hat{a}_{ij}$ is the maximum variation around this value. Contrary to Bertsimas and Sim (2004), in this brief review we restrict the notation to the case $x \geq 0$ as it is the case for NHLP.

Bertsimas and Sim consider a budget of uncertainty, Γ_i, controlling the maximum number of coefficients (i, j) where the variation $\hat{a}_{ij}$ can be realized. The Robust counterpart of (P) is the following program:

$$\begin{aligned} &\max \quad \mathbf{c}'\mathbf{x} \\ &s.t. \sum_j a_{ij}x_j + \max_{S_i\cup\{t_i\}||S_i|=\lfloor\Gamma_i\rfloor, t_i\in J\setminus S_i} \left\{ \sum_{j\in S_i} \hat{a}_{ij}x_j + (\Gamma_i - \lfloor\Gamma_i\rfloor)\hat{a}_{it_i}x_{t_i} \right\} \leq b_i \ \forall i \\ &\mathbf{x} \geq \mathbf{0} \end{aligned}$$

For a given vector $\mathbf{x}$, the protection against the uncertainty in $\tilde{a}_{ij}$ is given by the function $\beta_i(\mathbf{x}, \Gamma_i) = \max_{S_i\cup\{t_i\}||S_i|=\lfloor\Gamma_i\rfloor, t_i\in J\setminus S_i} \left\{ \sum_{j\in S_i} \hat{a}_{ij}x_j + (\Gamma_i - \lfloor\Gamma_i\rfloor)\hat{a}_{it_i}x_{t_i} \right\}$ and equals the objective function of the following linear program:

$$\begin{aligned} \beta_i(\mathbf{x}, \Gamma_i) = \max \quad & \sum_j \hat{a}_{ij}x_j z_{ij} \\ s.t. \quad & \sum_j z_{ij} \leq \Gamma_i \\ & 0 \leq z_{ij} \leq 1 \quad \forall j \end{aligned}$$

Bertsimas and Sim showed the optimal solution of the above problem consists of $\lfloor\Gamma_i\rfloor$ variables z_{ij} equal to 1 and one variable equal to $\Gamma_i - \lfloor\Gamma_i\rfloor$ which is equivalent to the selection of a subset $S_i \cup \{t_i\}||S_i| = \lfloor\Gamma_i\rfloor, t_i \in J\setminus S_i$. The maximum protection achieved by the function $\beta_i(\mathbf{x}, \Gamma_i)$ corresponds to the worst case scenario with an uncertainty budget Γ_i. This result is obtained through the dualization of $\beta_i(\mathbf{x}, \Gamma_i)$

$$\begin{aligned} \beta_i(\mathbf{x}, \Gamma_i) = \min \quad & q_i\Gamma_i + \sum_j r_{ij} \\ s.t. \quad & q_i + r_{ij} \geq \hat{a}_{ij}x_j \quad \forall j \\ & q_i, r_{ij} \geq 0 \quad \forall j \end{aligned}$$

The main result of Bertsimas and Sim is summarized in Theorem 1.

Theorem 1 (Bertsimas and Sim): The Robust counterpart of the generic problem (P) can be formulated as the following Linear Program:

$$\begin{aligned} \min \quad & \boldsymbol{c'x} \\ s.t. \quad & \sum_j a_{ij}x_j + q_i\Gamma_i + \sum_{j:(i,j)\in J} r_{ij} \leq b_i && \forall i \\ & q_i + r_{ij} \geq \hat{a}_{ij}x_j && \forall (i,j) \in J \\ & \boldsymbol{x, q, r} \geq \mathbf{0} \end{aligned}$$

where Γ_i is the uncertainty budget for constraint i, and J is the set of parameters subject to uncertainty. We now describe the robust counterpart of the NHLP when there is uncertainty on demand volumes first, then on customer utilities. Demands and utilities have in common that they both appear in the objective function and the budget constraint, which requires to define two categories of deviation variables in each robust model.

3.2 Uncertainty on demand volumes

We first consider uncertainty on demands such that $\tilde{D}_i \in [D_i - \hat{D}_i, D_i + \hat{D}_i]$. A robust approach with this respect seeks to maximize the worst-case average utility of potential buyers over all possible variations of the demand parameters, while ensuring feasibility of the solution in any case which, given a subset of selected locations, is only critical for the budget constraint (12).

Note that, since demand data D_i do not intervene in equations (16), the proportion of demand x_{ij} is not affected by fluctuations to D_i. Model NHLP can then be rewritten with a protection function $\beta(\mathbf{x}, \Gamma)$, where Γ is the uncertainty budget, to ensure feasibility, and a function $\delta(\mathbf{x}, \Gamma)$ to protect the objective as:

$$\max \quad \frac{1}{V}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ij}\bar{D}_i x_{ij} + \delta(x,\Gamma) \tag{19}$$

$$s.t. \quad \sum_{i=1}^{n}\sum_{j=1}^{m} c_j\bar{D}_i x_{ij} + \beta(x,\Gamma) \leq B \tag{20}$$

$$(14) - (18)$$

$$\text{where } \beta(x,\Gamma) = \max_{z} \sum_{i=1}^{n}\sum_{j=1}^{m} c_j\hat{D}_i x_{ij} z_i \qquad \delta(x,\Gamma) = \min_{z'} \frac{1}{V}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ij}\hat{D}_i x_{ij} z'_i$$

$$s.t. \sum_{i=1}^{n} |z_i| \leq \Gamma \qquad s.t. \sum_{i=1}^{n} |z'_i| \leq \Gamma$$

$$-1 \leq z_i \leq 1 \quad \forall i \qquad -1 \leq z'_i \leq 1 \quad \forall i$$

Note that, despite the fact that the same parameter D_i varies in the objective function and the budget constraint, one actually needs two sets of variables z_i and z'_i in the above model since the deviation vector z that achieves the worst added cost for the budget constraint is generally not the same that achieves the worst loss of utility in the objective function.

Proposition 3.1. *The robust counterpart of model NHLP (11)-(18) is the following model* ***R-NHLP***

$$\max \quad \frac{1}{V}\sum_{i=1}^{n}\sum_{j=1}^{m} u_{ij}\bar{D}_i x_{ij} - \Gamma q^{\delta} - \sum_{i=1}^{n} r_i^{\delta} \qquad \textbf{R-NHLP}$$

$$s.t. \sum_{i=1}^{n}\sum_{j=1}^{m} c_j\bar{D}_i x_{ij} + \Gamma q^{\beta} + \sum_{i=1}^{n} r_i^{\beta} \leq B$$

$$q^{\delta} + r_i^{\delta} \geq \frac{1}{V}\hat{D}_i \sum_{j=1}^{m} u_{ij} x_{ij} \quad \forall i$$

$$q^{\beta} + r_i^{\beta} \geq \hat{D}_i \sum_{j=1}^{m} c_j x_{ij} \quad \forall i$$

$$(14) - (18)$$

$$\boldsymbol{q}, \boldsymbol{r} \geq \mathbf{0}$$

Proof. In the sense of Bertsimas and Sim (2004), the protection function $\beta(x, \Gamma)$ for (20) is the optimal value of the problem:

$$\begin{array}{lll} \max\limits_{z} \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} c_j \hat{D}_i x_{ij} z_i & & \min\ \Gamma q^{\beta} + \sum\limits_{i=1}^{n} r_i^{\beta} \\ s.t. \sum\limits_{i=1}^{n} z_i \leq \Gamma & \iff & s.t.\ q^{\beta} + r_i^{\beta} \geq \hat{D}_i \sum\limits_{j=1}^{m} c_j x_{ij} \quad \forall i \\ z_i \leq 1 \quad \forall i & & q^{\beta} \geq 0 \\ z_i \geq 0 \quad \forall i & & r_i^{\beta} \geq 0 \quad \forall i \end{array}$$

where dual variable q^{β} is associated with the uncertainty budget constraint, and each dual variable r_i^{β} is associated with constraint $z_i \leq 1$ for $i = 1, \ldots, n$. Similarly, the protection function $\delta(x, \Gamma)$ for (19) is the optimal value of:

$$\begin{array}{lll} -\max\limits_{z'} \frac{1}{V} \sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} u_{ij} \hat{D}_i x_{ij} z'_i & & -\min\ \Gamma q^{\delta} + \sum\limits_{i=1}^{n} r_i^{\delta} \\ s.t. \sum\limits_{i=1}^{n} z'_i \leq \Gamma & \iff & s.t.\ q^{\delta} + r_i^{\delta} \geq \frac{1}{V} \hat{D}_i \sum\limits_{j=1}^{m} u_{ij} x_{ij} \quad \forall i \\ z'_i \leq 1 \quad \forall i & & q^{\delta} \geq 0 \\ z'_i \geq 0 \quad \forall i & & r_i^{\delta} \geq 0 \quad \forall i \end{array}$$

□

In some way, the above robust model is a simple extension of the robustness approach of Bertsimas and Sim (2004) where uncertainty impacts not only constraints but also the objective function; it will provide interesting insights though for new housing developments as shown in the numerical results section. The case when uncertainty lies on customer preferences or utilities is, to our mind, quite original on a purely research point of view, as we mix a scenario-based approach and a Γ-budget approach. This strategy leads to a more complex formulation where the distribution flows need to be modeled by specific variables and constraints for each scenario, given the particular structure of the choice model.

3.3 Uncertainty on customer utilities

The impact of uncertainty on customers' utilities is twofold: a direct impact on the objective function, which can be addressed as in the previous section, with the added consideration of updating the proportion of demand flowing from each demand source to the chosen locations. Contrary to the previous section where demand volumes are continuous within their variation range as in Bertsimas and Sim (2004), we choose to discretize the set of possible scenarios for utilities, for two main reasons. On one hand, considering continuous utilities leads to intractable non-linearity of the IIA constraints (16). On the other hand, since utilities are generally hard to calibrate in a multinomial logit choice model, we structure the uncertainty focusing on the two main criteria that drive customer preferences for new housing: distance (d) and price (p). Hence we consider for each demand source a base (nominal) scenario and two variants: one scenario that is more distance-centric than the base scenario, and one that is more price-centric.

Let us now describe more precisely the characteristics of utilities u_{ij} for our housing problem, and which part of the utility that is impacted by uncertainty. We consider in expression (2) that the locations differentiate from one another through the level of the attributes, a_{ij}^l which depend on both i and j, but the coefficient or weight w_i^l only varies with the customer i and not with the location j. Furthermore, only parameters w_i^l are subjected to uncertainty, e.g. from the level of aggregation of the demand sources, changing of preferences with time, errors on the estimation

of the parameter. We discretize the set of possible values such as $\tilde{w}_i \in \{w_i^0, w_i^d, w_i^p\}$, where as mentioned before d and p respectively stand for distance and price, and 0 stands for the nominal scenario. This leads to the following three scenarios for a given demand source i:

- nominal (base) scenario : weight vector w_i^0 and utilities $u_{ij}^0 = \sum_{l=1}^{L} w_i^{l_0} a_{ij}^l$
- d-centric scenario : weight vector w_i^d and utilites $u_{ij}^d = \sum_{l=1}^{L} w_i^{l_d} a_{ij}^l$
- p-centric scenario : weight vector w_i^p and utilities $u_{ij}^p = \sum_{l=1}^{L} w_i^{l_p} a_{ij}^l$

Note that the d-centric and p-centric notion is relative to the base scenario. We define binary variables $z_i^d = 1$ if there is a shift from the nominal scenario to the d-centric scenario, and $z_i^p = 1$ if the shift is towards the p-centric scenario for demand source i. $z_i^d + z_i^p = 0$ means that the nominal scenario remains. The constraint $z_i^d + z_i^p \leq 1$ clearly holds, and the uncertain utility $\tilde{u}_{ij}$ can be written as:

$$\tilde{u}_{ij} = u_{ij}^0(1 - z_i^d - z_i^p) + u_{ij}^d z_i^d + u_{ij}^p z_i^p$$

The nominal scenario partly depends on the characteristics of the demand source, e.g. low income households are very sensitive to prices, whereas higher income would lead to a more distance-centric behaviour. Therefore, the nominal scenario being the most likely one, only a limited number of demand sources $\Gamma \in \mathbb{N}$ are allowed to deviate from this base scenario. This leads to add the uncertainty budget constraint $\sum_{i=1}^{n} z_i^d + z_i^p \leq \Gamma$. As uncertainty holds both in the budget constraint and the objective function, we need to duplicate the above variables z_i^d and z_i^p for feasibility into z'^d_i and z'^p_i for the objective function as in section 3.2.

We can rewrite model NHLP using variables z, z', and flow variables x_{ij}^s for the flow of demand in each possible scenario $s = 0, d, p$. A robust approach to Model NHLP seeking to protect the optimal value against the worst realization of the uncertainty and ensure feasibility of the solution can be written as (21)-(28):

$$\max_{\boldsymbol{x},\boldsymbol{y}} \left\{ \begin{array}{ll} \min\limits_{\boldsymbol{z}'} & \frac{1}{V}\left[\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} u_{ij}^{0} D_i x_{ij}^{0}(1-\sum\limits_{s\neq 0} z'^{s}_{i}) + \sum\limits_{s\neq 0}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} u_{ij}^{s} D_i x_{ij}^{s} z'^{s}_{i}\right] \\ s.t. & \sum\limits_{s\neq 0}\sum\limits_{i=1}^{n} z'^{s}_{i} \leq \Gamma \\ & \sum\limits_{s\neq 0} z'^{s}_{i} \leq 1 \qquad \forall i \\ & \boldsymbol{z}' \in \{\mathbf{0},\mathbf{1}\} \end{array} \right\} \tag{21}$$

$s.t.$

$$\left\{ \begin{array}{ll} \max\limits_{\boldsymbol{z}} & \left[\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} c_j D_i x_{ij}^{0}(1-\sum\limits_{s\neq 0} z_{i}^{s}) + \sum\limits_{s\neq 0}\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} c_j D_i x_{ij}^{s} z_{i}^{s}\right] \leq B \\ s.t. & \sum\limits_{s\neq 0}\sum\limits_{i=1}^{n} z_{i}^{s} \leq \Gamma \\ & \sum\limits_{s\neq 0} z_{i}^{s} \leq 1 \qquad \forall i \\ & \boldsymbol{z} \in \{\mathbf{0},\mathbf{1}\} \end{array} \right\} \tag{22}$$

$$\sum_{j=1}^{m} y_j \leq P \tag{23}$$

$$\sum_{j=1}^{m} x_{ij}^{s} = 1 \qquad \forall (i,s) \tag{24}$$

$$x_{ij}^{s} \leq y_j \qquad \forall (i,j,s) \tag{25}$$

$$x_{ij}^{s} \leq x_{ik}^{s} e^{u_{ij}^{s} - u_{ik}^{s}} + (2 - y_j - y_k) \qquad \forall (i,j,k,s) | j \neq k \tag{26}$$

$$\boldsymbol{x} \geq \mathbf{0} \tag{27}$$

$$\boldsymbol{y} \in \{\mathbf{0},\mathbf{1}\} \tag{28}$$

Unlike in Bertsimas and Sim (2004), we keep both positive and negative deviations from the estimated parameter since the direction of the objective function cannot be inferred directly from the direction of change in the parameter. Furthermore, the uncertainty subproblem to be dualized is no longer a classical knapsack problem due to this fact. (21)-(28) can be linearized in a similar matter as that proposed by Bertsimas and Sim using protection functions. (22), and analogously (21), can be rewritten with a deterministic and an uncertain component as

$$\sum_{i=1}^{n}\sum_{j=1}^{m} c_j D_i x_{ij}^{0} + \beta(x,\Gamma) \leq B, \qquad \text{where}$$

$$\beta(x,\Gamma) = \max \sum_{s\neq 0}\sum_{i=1}^{n}\sum_{j=1}^{m} c_j D_i (x_{ij}^{s} - x_{ij}^{0}) z_{i}^{s} \tag{29}$$

$$s.t. \sum_{s\neq 0}\sum_{i=1}^{n} z_{i}^{s} \leq \Gamma \tag{30}$$

$$\sum_{s\neq 0} z_{i}^{s} \leq 1 \qquad \forall i \tag{31}$$

$$\boldsymbol{z} \in \{\mathbf{0},\mathbf{1}\} \tag{32}$$

We now show that we can substitute the above $0-1$ subproblem with its linear relaxation. Doing so, we can replace the linear relaxation of the subproblem by its dual program in the model and find an equivalent formulation.

Lemma 3.1. *The linear relaxation of subproblem (29)-(32) has an optimal solution such that* $z_i^d, z_i^p \in \{0,1\}$ *for all* $i = 1, \dots, n$.

Proof. We transform the constraints into $\sum_{i=1}^{n} z_i^d + z_i^p + sl_0 = \Gamma$ and $z_i^d + z_i^p + sl_i = 1$ for $i = 1, \dots, n$, where sl_i, $i = 0, \dots, n$ are slack variables. In this standard form the LP has $3n+1$ variables and $n+1$ constraints. So we have at most $n+1$ non-zero basis variables. If $sl_0 > 0$, then there is exactly one non-zero variable for each of the n constraints $z_i^d + z_i^p + sl_i = 1$. In that case, one of the variables z_i^d, z_i^p or sl_i is one and the others are zero, and the solution is indeed binary. If $sl_0 = 0$, then there exists a single index i such that exactly two variables are non-zero. For this constraint i, since $\Gamma \in \mathbb{N}$, we necessarily have $sl_i = 0$ otherwise we would not have $sl_0 = 0$. Therefore, $z_i^d + z_i^p = 1$ and both variables are fractional. Since the solution is optimal, we necessarily have $\sum_j c_j(x_{ij}^d - x_{ij}^0) = \sum_j c_j(x_{ij}^p - x_{ij}^0)$, and setting $z_i^d = 1$ and $z_i^p = 0$ (or $z_i^d = 0$ and $z_i^p = 1$) does not deteriorate the objective value. □

We can then linearize (21)-(28) in the sense of Bertsimas and Sim (2004).

Proposition 3.2. *The robust counterpart of (21)-(28) is the following model* ***uR-NHLP***

$$
\begin{aligned}
\max \quad & \frac{1}{V}\sum_{i=1}^{n}\sum_{j=1}^{m} \bar{u}_{ij} D_i \bar{x}_{ij} - \Gamma q^{\delta} - \sum_{i=1}^{n} r_i^{\delta} && && \textbf{uR-NHLP}\\
s.t. & \sum_{i=1}^{n}\sum_{j=1}^{m} c_j D_i \bar{x}_{ij} + \Gamma q^{\beta} + \sum_{i=1}^{n} r_i^{\beta} \leq B && && (33)\\
& q^{\delta} + r_i^{\delta} \geq \frac{1}{V} D_i \sum_{j=1}^{m} \left(u_{ij}^0 x_{ij}^0 - u_{ij}^s x_{ij}^s \right) && \forall (i,s) | s \neq 0\\
& q^{\beta} + r_i^{\beta} \geq D_i \sum_{j=1}^{m} c_j (x_{ij}^s - x_{ij}^0) && \forall (i,s) | s \neq 0\\
& (23), (24) - (28)\\
& \boldsymbol{q}, \boldsymbol{r} \geq \mathbf{0}
\end{aligned}
$$

Proof. We illustrate the linearization of the equations regarding uncertainty (21)-(22) with the budget constraint (22) for simplicity. By Lemma 3.1 when dualizing the linear relaxation of subproblem (29)-(32) we indeed obtain the optimal value of its binary 0-1 counterpart. The linear relaxation of this subproblem is equivalent through dualization to :

$$
\begin{aligned}
\min \quad & \Gamma q^{\beta} + \sum_{i=1}^{n} r_i^{\beta}\\
s.t. \quad & q^{\beta} + r_i^{\beta} \geq D_i \sum_{j=1}^{m} c_j (x_{ij}^d - \bar{x}_{ij}) && \forall i\\
& q^{\beta} + r_i^{\beta} \geq D_i \sum_{j=1}^{m} c_j (x_{ij}^p - \bar{x}_{ij}) && \forall i\\
& q^{\beta}, \boldsymbol{r}^{\beta} \geq 0
\end{aligned}
$$

which can be substituted in (22). The same reasoning holds for (21). □

We end this subsection by studying a variant of the problem where demand sources have *asymmetric uncertainty budgets*. Indeed, following the Γ-robustness framework developed by Bertsimas and

Sim, we previously defined the different elements of the scenario set as deviations from a base (most likely) scenario. One can consider however in the housing context that it is more likely to deviate to the price-centric scenario than the distance-centric scenario, or the reverse. In this case, we need to differentiate uncertainty budgets, i.e., set a distinct Γ_s for each scenario s. This provides subproblem (34)-(37).

$$\beta(x,\Gamma) = \max \sum_{s\neq 0}\sum_{i=1}^{n}\sum_{j=1}^{m} c_j D_i (x_{ij}^s - x_{ij}^0) z_i^s \tag{34}$$

$$s.t. \sum_{i=1}^{n} z_i^s \leq \Gamma_s \qquad s \neq 0 \tag{35}$$

$$\sum_{s\neq 0} z_i^s \leq 1 \qquad \forall i \tag{36}$$

$$\boldsymbol{z} \in \{\mathbf{0},\mathbf{1}\} \tag{37}$$

We show that, as in the general case, replacing the above $0-1$ subproblem by its linear relaxation does not change the problem, so the dualization process remains valid also in that case.

Lemma 3.2. *The linear relaxation of subproblem (34)-(37) has an optimal solution such that* $z_i^d, z_i^p \in \{0,1\}$ *for all* $i = 1,\ldots,n$.

Proof. We transform the constraints using slack variables into $\sum_{i=1}^n z_i^d + sl_d = \Gamma_d$, $\sum_{i=1}^n z_i^p + sl_p = \Gamma_p$ and $z_i^d + z_i^p + sl_i = 1$ for $i = 1,\ldots,n$. In this standard form the LP has $3n+2$ variables and $n+2$ constraints. Therefore there are at most $n+2$ non-zero basis variables. There are three cases w.r.t variables (sl_d, sl_p): i) $sl_d, sl_p > 0$, ii) $sl_d, sl_p = 0$, or iii) $sl_d \oplus sl_p = 0$. If i), there is exactly one non-zero variable for each of the n constraints $z_i^d + z_i^p + sl_i = 1$, and the solution is binary as in the proof of lemma 3.1.

If ii), there exists a single pair of indices (i,i') such that exactly two variables are non-zero for each of the n constraints $z_i^d + z_i^p + sl_i = 1$, since $\Gamma_d, \Gamma_p \in \mathbb{N}$. For this pair (i,i'), either $sl_i + sl_{i'} = 0$ or $sl_i + sl_{i'} = 1$. In the first case, $z_i^d + z_i^p = 1$ and both variables are fractional. Since the solution is optimal, we necessarily have $\sum_j c_j(x_{ij}^d - \bar{x}_{ij}) = \sum_j c_j(x_{ij}^p - \bar{x}_{ij})$, and setting $z_i^d = 1$ and $z_i^p = 0$ (or $z_i^d = 0$ and $z_i^p = 1$) does not deteriorate the objective value. The same reasoning follows for i'. If $sl_i + sl_{i'} = 1$, either $z_i^d z_{i'}^d > 0$ and $z_i^p + z_{i'}^p = 0$ or vice-versa. Since the solution is optimal, we necessarily have $\sum_j c_j(x_{ij}^d - \bar{x}_{ij}) = \sum_j c_j(x_{i'j}^p - \bar{x}_{i'j})$, and setting $(z_i^d, z_{i'}^d, sl_i, sl_{i'}) = (1,0,0,1)$, or $(z_i^d, z_{i'}^d, sl_i, sl_{i'}) = (0,1,1,0)$, does not deteriorate the objective value.

If iii), $sl_d > 0$ and $sl_p = 0$ implies that there exists exactly one constraint $z_i^d + z_i^p + sl_i = 1$ where $z_i^d + sl_i = 1$ since $\Gamma_p \in \mathbb{N}$ and $z_i^d, sl_i > 0$. It necessarily follows that $\sum_j c_j(x_{ij}^d - \bar{x}_{ij}) = 0$, since the solution is optimal. Then setting $z_i^d = 1$ and $sl_i = 0$ (or $z_i^d = 0$ and $sl_i = 1$) does not deteriorate the objective value. The same is true for the case where $sl_p > 0$ and $sl_d = 0$ □

We now present numerical experiments on a dataset of French cities, as well as some managerial insights for housing development.

4 Numerical Experiments

4.1 Data

Model NHLP was tested on instances built from French data of the Paris region. 30 neighborhoods were selected according to their location and population characteristics. With information gathered from the National Institute of Statistic and Economic Studies (INSEE), we built a number of instances from a combination of the following parameters:

- Number of demand sources $n \in \{30, 40, 50, 60\}$[1]
- Number of possible locations for new housings $m \in \{15, 20\}$
- Maximum number of locations allowed for development $P \in \{2, 4, 6\}$
- Customers' utilities computed from one of two behavioral traits: {distance-centric, price-centric}

The utility combines two criteria a^d and a^p with weight vector w:

- *Distance ratio* : a^d_{ij} is the ratio of the distance between i and j and the average distance from i to all sites (a small ratio is preferred)

$$a^d_{ij} = -\frac{d_{ij}}{\overline{d}_{i\cdot}}$$

  where $\overline{d}_{i\cdot} = \sum_j d_{ij}/m$.

- *Relative deviation of Real-estate price vs Purchasing Power*: a^p_{ij} is the deviation in % between the real-estate price per meter-square at location j, denoted by p_j , and the price that demand node i would be willing to pay given its income (Purchasing Power, noted PP_i), where a small deviation is preferred:

$$a^p_{ij} = \frac{|p_j - PP_i|}{PP_i}$$

Let us give more details on how purchasing powers (PP) were estimated in our study. We denote by v_i the city of demand source i. The INSEE French institute provides income data for two different population segments in each city: non-taxed (lower-income) and taxed (higher-income) households. We set $PP_i = p_{v_i}$ if i corresponds to the lower-income population segment of city v_i, and $PP_i = (I_i/\bar{I}_i)p_{v_i}$ if i corresponds to the higher-income population of city v_i, where I_i is the average income of segment i and $\bar{I}_i$ is the average income of city v_i. Indeed, we make the assumption that the average PP of the lower-income segment is the average real-estate price of the city this population lives in, but the higher-income segment has a higher PP, with a multiplicative factor which is the ratio between the segment income and the city income.

As no survey was available with data to calibrate the parameters of the utility function, we scale the two factors a^d_{ij} and a^p_{ij} using z-scores ((value-mean)/standard-deviation), which are noted $\hat{a}^d_{ij}$ and $\hat{a}^p_{ij}$. We obtain comparable measures that can be aggregated to compute utilities u_{ij} using weights which sum over one. Thus we express the utility of a potential buyer i choosing housing in j as:

$$u_{ij} = w^d_i \hat{a}^d_{ij} + w^p_i \hat{a}^p_{ij}$$

where $w^d_i + w^p_i = 1$. We call "price-centric" the case when $(w^d_i, w^p_i) = (1/3; 2/3)$ and "distance-centric" the reverse case $(w^d_i, w^p_i) = (2/3; 1/3)$. We estimated demand volumes using INSEE data on households. The real-estate prices were obtained from MeilleursAgents.com. Table 1 reports data about cities. We selected a set of cities with high diversity in population size and income to get richer instances.

[1]Each neighborhood was partitioned into two demand sources according to information available on income level.

Table 1: Characteristics of the demand sources

City	Population	Price €/m^2	Taxed households Demand	Taxed households Income (€)	Non-Taxed households Demand	Non-Taxed households Income (€)
Boulogne-Billancourt	114205	7740	382	57966	146	10280
Argenteuil	103125	2811	189	31193	183	8886
Courbevoie	87469	5847	284	44904	103	9746
Cergy	56988	2534	120	31404	97	9006
Asnières-sur-Seine	82327	4795	216	42371	122	9459
Sarcelles	58614	2139	68	28385	106	8996
Nanterre	89185	4547	175	34116	151	8704
Franconville	33097	2768	86	33564	45	10112
Colombes	85398	4417	190	38949	132	9280
Garges-les-Gonesse	40012	2150	41	25742	79	8607
Rueil-Malmaison	79426	5561	240	54771	79	10993
Pontoise	29548	2594	67	33905	50	9176
Levallois-Perret	64253	7117	216	51419	81	9995
Ermont	27446	2992	67	36381	39	10244
Neuilly-sur-Seine	61754	9350	214	114418	66	13649
Bezons	27987	3085	54	30760	45	9332
Issy-les-Moulineaux	64355	6249	207	45857	73	9697
Sannois	26090	2923	62	36254	36	9695
Clichy	58916	4925	147	31640	113	8750
Taverny	26144	2765	64	36175	32	10258
		mean	155	42009	89	9743
		st. dev.	92	19186	43	1119

4.2 Numerical results for the non-robust model NHLP

In this section we present computational results for model NHLP. We analyze the consistency of solutions with the defining assumptions of the choice model, and explore its tractability. Table 2 shows numerical results on instances considering $m = 15$ potential sites for new developments, $n \in \{30, 60\}$ and $P \in 2, 4, 6$. v^* is the optimal value, IGap is the ratio $\bar{v}/v^*$, where $\bar{v}$ is the value of the linear relaxation, Ts is CPU time in seconds, and #Sites is the number of sites actually selected in the optimal solution. The complexity of the problem naturally increases with each parameter, however, the number P of sites to select seems to have a particularly high impact. Additionally to the parameters described in 4.1, we tested the model against four budgetary levels. For a given cost vector c, we computed its median (2Q), mean ($\bar{c}$), 3rd Quartile (3Q), and mean plus standard deviation ($\bar{c}+\sigma_c$) as average budget per housing unit, so that $B/V \in \{c^{2Q}, \bar{c}, c^{3Q}, \bar{c}+\sigma_c\}$. We consistently find larger integrality gaps on instances with smaller budgets ($(c^{2Q}, \bar{c})$). The computation times however are generally shorter for these instances.

Table 2: Results for **NHLP** with m=15,20 potential locations

				v^*		IGap		Ts		#Sites	
n	P	d/p	B	15	20	15	20	15	20	15	20
30	2	d	nc^{2Q}	0,582	0,498	1,328	1,298	1,3	3,4	1	2
30	2	d	n$\bar{c}$	0,582	0,517	1,350	1,513	1,5	7,7	1	2
30	2	d	nc^{3Q}	0,653	0,670	1,305	1,312	1,4	6,7	2	2
30	2	d	n$(\bar{c}+\sigma_c)$	0,653	0,672	1,305	1,308	1,5	4,3	2	2
30	2	p	nc^{2Q}	0,504	0,075	1,488	6,642	2,1	10,1	1	1
30	2	p	n$\bar{c}$	0,504	0,420	1,528	1,776	2,4	9,0	1	2
30	2	p	nc^{3Q}	0,760	0,735	1,296	1,370	1,2	5,2	2	2
30	2	p	n$(\bar{c}+\sigma_c)$	0,760	0,791	1,296	1,282	1,1	3,1	2	2
30	4	d	nc^{2Q}	0,618	0,498	1,602	1,753	13,2	92,1	4	2
30	4	d	n$\bar{c}$	0,618	0,604	1,621	1,691	24,6	169,4	4	4
30	4	d	nc^{3Q}	0,690	0,730	1,594	1,564	28,9	99,7	4	4
30	4	d	n$(\bar{c}+\sigma_c)$	0,692	0,736	1,590	1,557	30,7	90,1	4	4
30	4	p	nc^{2Q}	0,548	0,312	1,596	2,020	10,8	85,5	4	3
30	4	p	n$\bar{c}$	0,550	0,545	1,635	1,625	12,0	93,1	4	3
30	4	p	nc^{3Q}	0,769	0,810	1,583	1,541	26,8	92,2	4	4
30	4	p	n$(\bar{c}+\sigma_c)$	0,794	0,844	1,538	1,501	10,1	64,8	4	4
30	6	d	nc^{2Q}	0,618	0,524	1,832	1,871	39,1	126,2	4	5
30	6	d	n$\bar{c}$	0,623	0,647	1,843	1,820	42,1	271,0	5	6
30	6	d	nc^{3Q}	0,703	0,756	1,783	1,739	76,2	332,9	6	6
30	6	d	n$(\bar{c}+\sigma_c)$	0,703	0,756	1,783	1,739	72,6	310,8	6	6
30	6	p	nc^{2Q}	0,548	0,343	1,735	2,041	30,2	93,7	4	5
30	6	p	n$\bar{c}$	0,552	0,545	1,773	1,774	30,3	197,2	5	3
30	6	p	nc^{3Q}	0,775	0,820	1,689	1,652	59,5	381,8	5	6
30	6	p	n$(\bar{c}+\sigma_c)$	0,794	0,844	1,648	1,614	39,4	229,6	4	4
60	2	d	nc^{2Q}	0,569	0,462	1,361	1,342	5,6	14,3	1	2
60	2	d	n$\bar{c}$	0,569	0,490	1,386	1,571	6,8	25,5	1	2
60	2	d	nc^{3Q}	0,658	0,655	1,329	1,345	5,4	19,3	2	2
60	2	d	n$(\bar{c}+\sigma_c)$	0,667	0,672	1,311	1,311	4,8	12,7	2	2
60	2	p	nc^{2Q}	0,516	0,060	1,487	8,153	6,9	34,4	1	1
60	2	p	n$\bar{c}$	0,516	0,448	1,534	1,673	7,3	32,7	1	2
60	2	p	nc^{3Q}	0,764	0,726	1,296	1,377	3,4	20,1	2	2
60	2	p	N$(\bar{c}+\sigma_c)$	0,764	0,787	1,297	1,280	3,0	13,3	2	2
60	4	d	Nc^{2Q}	0,586	0,462	1,637	1,786	37,3	224,8	4	2
60	4	d	N$\bar{c}$	0,588	0,591	1,656	1,666	88,1	354,5	4	4
60	4	d	Nc^{3Q}	0,685	0,691	1,568	1,593	93,2	429,3	4	4
60	4	d	N$(\bar{c}+\sigma_c)$	0,685	0,691	1,568	1,593	89,4	418,9	4	4
60	4	p	Nc^{2Q}	0,575	0,306	1,541	2,000	31,8	337,4	4	3
60	4	p	n$\bar{c}$	0,583	0,510	1,565	1,728	38,4	505,4	4	4
60	4	p	nc^{3Q}	0,780	0,805	1,531	1,514	79,6	319,5	4	4
60	4	p	n$(\bar{c}+\sigma_c)$	0,780	0,805	1,531	1,517	43,2	301,6	4	4
60	6	d	nc^{2Q}	0,589	0,468	1,824	1,986	133,8	516,9	6	5
60	6	d	n$\bar{c}$	0,602	0,613	1,816	1,840	187,9	1081,4	5	6
60	6	d	nc^{3Q}	0,695	0,723	1,729	1,725	205,8	1688,1	6	6
60	6	d	n$(\bar{c}+\sigma_c)$	0,695	0,723	1,729	1,725	176,4	1537,8	6	6
60	6	p	nc^{2Q}	0,575	0,338	1,648	1,996	85,2	287,7	4	5
60	6	p	n$\bar{c}$	0,583	0,552	1,671	1,731	79,2	828,9	4	6
60	6	p	nc^{3Q}	0,782	0,805	1,638	1,632	134,0	1266,6	5	4
60	6	p	n$(\bar{c}+\sigma_c)$	0,782	0,805	1,638	1,634	120,7	1138,8	5	4

$(c^{2Q}, \bar{c}, c^{3Q}, \bar{c}+\sigma_c)$:(median, mean, 3^{rd} Quartile, mean + standard deviation) values of vector c.



Figure 2: Demographics of an instance with 15 cities (The size of the circles is proportional to corresponding D_i/PP_i). Paris is the empty space on the bottom-right corner.

Figure 2 presents the main demographic characteristics of the area under study (Figure 2). The upper level of the figure shows a higher concentration of the demand in the south of the region. Furthermore, it points to a prevalence of high income potential buyers. On the lower level of the figure, the left side shows the relative Purchasing Power (PP) for the high income-segments. A comparison with the right-hand side shows that the demand sources with the highest PP from the high income-segment also present the highest PP from the low income-segment. Figure 3 shows the result of the optimization model with demand sources which are also potential locations (□), and locations selected for new developments (■ with some specific color). Additionally, the figure draws, in form of a pie chart, the fraction of demand (x_{ij}) flowing from each demand source to the selected locations ($y_j = 1$). As could have been expected, the choice of new development locations was influenced by the concentration of demand in the southern region. The consistency of model NHLP can be appreciated by comparing the top and bottom halves of Figure 3. When the utility is mainly driven by the distance criterion (top half of figure 3), we can see similar distributions of demands when comparing the two segments (low-income and high-income) of a given city. This is logical as these two segments mainly differ by their purchasing power but the price is not the main choice criterion for the d-centric case, so income has a minor impact contrary to the p-centric case (bottom half of the figure).

4.3 Numerical results for uncertainty on demands

In this section we present the impact of demand uncertainty on NHLP using the **R-NHLP** robust model. The main findings in Table 3 highlight the usefulness of a robust model when dealing with uncertainty. For each original (nominal) instance with demand vector $\bar{D}$, we generated 25 "deviated" instances noted $t = 1, \ldots, 25$ with demand vector $\tilde{D}_t$ inside uncertainty range $[\bar{D} - \hat{D}, \bar{D} + \hat{D}]$ for each instance. Let us note by y_R the robust solution and y_N the nominal solution, i.e., the optimal solution for NHLP with nominal demand $\bar{D}$. For each original instance we computed over the 25 deviated instances, for y_R and y_N: the average loss of satisfaction in the objective

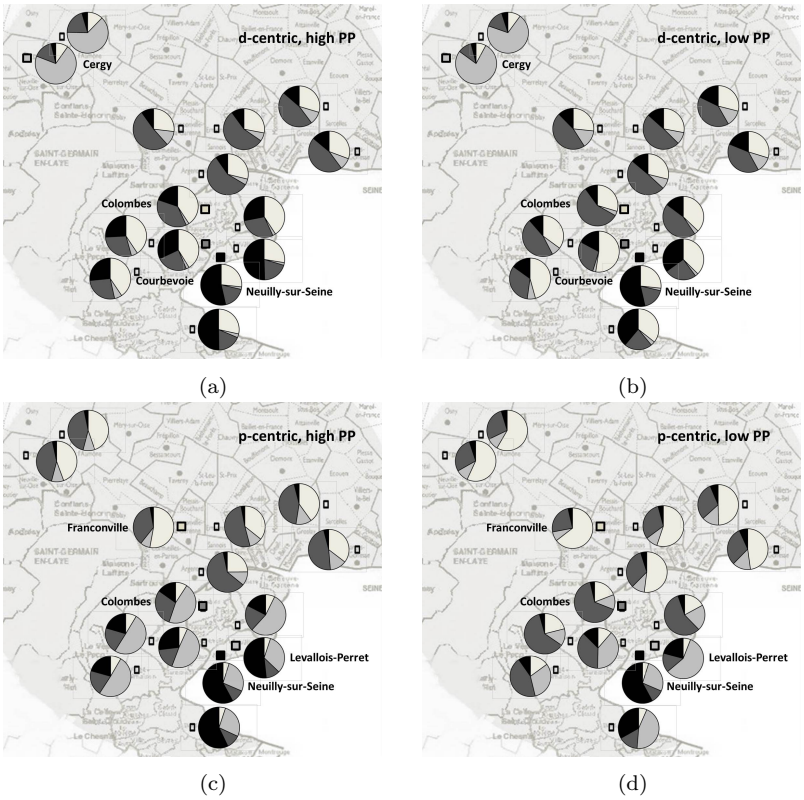


Figure 3: Distribution of the demand for an instance with 30 demand sources. Selected locations for development are assigned a colored square and shown with their name on the side. Demand flows from each demand source are shown in the pie charts. For example, the black fraction in a pie flows to Neuilly-sur-Seine.

function, the number of infeasibilities, and the number of times solutions achieved optimality. We call $\bar{\Delta}_R$ the average loss of customer satisfaction (utility) of the robust solution y_R over runs $t = 1, \ldots, 25$, when compared to the value v_t^* of the optimal solution of deviated instance t:

$$\bar{\Delta}_R = \frac{1}{25} \sum_{t=1}^{25} \frac{v_t(y_R) - v_t^*}{v_t^*}$$

where $v_t(y)$ is the value of solution y on deviated instance t. Similarly, we compute the average loss of satisfaction $\bar{\Delta}_N$ achieved by the nominal solution y_N:

$$\bar{\Delta}_N = \frac{1}{25} \sum_{t=1}^{25} \frac{v_t(y_N) - v_t^*}{v_t^*}$$

We consider both the budget constraint (12) and the flow definition constraints (16) to be truly binding. In this way, if a solution y_R or y_N is infeasible, then the satisfaction derived from it is 0. Table 3 shows the effect of uncertainty on customer satisfaction with average losses $\bar{\Delta}_R, \bar{\Delta}_N$.

Generally, this loss is greater in the price-centric case. This can be explained by the interaction between the demand and the budget constraint which reduces feasibility. The initial setting of parameters Γ and $\hat{D}$ also has an impact. Not surprisingly, enlarging Γ (10 vs 5) or the maximum demand deviation $\hat{D}$ (+30% $\bar{D}$ vs +15 % $\bar{D}$) generates higher losses of satisfaction, mainly when the resources (B) are scarce as reported in the table. When comparing y_R, which always remains feasible, and y_N, it evidences a greater need for a robust formulation for tighter budget constraints (scenarios c^{2Q} and $\bar{c}$), as shown by the larger number of unfeasible repetitions (#inf) occurring with y_N. Finally, the number of times y_R and y_N achieved optimality tells us that the advantages of using the robust model **R-NHLP** diminish when the budget constraint is less tight (scenarios c^{3Q} and $\bar{c} + \sigma_c$).

Table 3: Results for a **R-NHLP** with (m,P)=(15,4)

							# inf		# best		Time (s)	
n	Γ	U	B	$\hat{D}$	$\bar{\Delta}_R$	$\bar{\Delta}_N$	R	N	R	N	R	N
30	5	d	nc^{2Q}	15%D	-0.13	-0.05	0	1	0	24	19	23
30	5	d	nc^{2Q}	30%D	-0.12	-0.38	0	8	0	14	13	23
30	5	d	n$\bar{c}$	15%D	-0.09	0.00	0	0	0	23	18	15
30	5	d	n$\bar{c}$	30%D	-0.14	-0.06	0	1	0	20	16	15
30	5	d	nc^{3Q}	15%D	0.00	0.00	0	0	20	20	32	32
30	5	d	nc^{3Q}	30%D	-0.03	0.00	0	0	0	14	36	32
30	5	d	n$(\bar{c}+\sigma_c)$	15%D	0.00	0.00	0	0	2	23	29	30
30	5	d	n$(\bar{c}+\sigma_c)$	30%D	0.00	0.00	0	0	4	20	32	30
30	5	p	nc^{2Q}	15%D	-0.16	-0.31	0	7	0	8	27	10
30	5	p	nc^{2Q}	30%D	-0.26	-0.38	0	9	0	6	14	10
30	5	p	n$\bar{c}$	15%D	-0.14	-0.05	0	1	0	23	14	12
30	5	p	n$\bar{c}$	30%D	-0.22	-0.35	0	7	0	13	14	12
30	5	p	nc^{3Q}	15%D	-0.05	-0.01	0	0	0	10	28	26
30	5	p	nc^{3Q}	30%D	-0.07	-0.01	0	0	0	9	15	26
30	5	p	n$(\bar{c}+\sigma_c)$	15%D	0.00	0.00	0	0	11	14	11	10
30	5	p	n$(\bar{c}+\sigma_c)$	30%D	0.00	0.00	0	0	11	14	24	10
30	10	d	nc^{2Q}	15%D	-0,14	-0,08	0	2	0	21	15	23
30	10	d	nc^{2Q}	30%D	-0,20	-0,21	0	5	0	9	9	23
30	10	d	n$\bar{c}$	15%D	-0,13	-0,04	0	1	0	16	28	15
30	10	d	n$\bar{c}$	30%D	-0,21	-0,09	0	2	0	9	13	15
30	10	d	nc^{3Q}	15%D	-0,04	0,00	0	0	0	14	36	31
30	10	d	nc^{3Q}	30%D	-0,06	-0,01	0	0	0	12	38	31
30	10	d	n$(\bar{c}+\sigma_c)$	15%D	0,00	0,00	0	0	10	8	34	29
30	10	d	n$(\bar{c}+\sigma_c)$	30%D	-0,01	-0,01	0	0	10	8	34	29
30	10	p	nc^{2Q}	15%D	-0,21	-0,13	0	3	0	9	12	10
30	10	p	nc^{2Q}	30%D	-0,41	-0,22	0	5	0	7	13	10
30	10	p	n$\bar{c}$	15%D	-0,19	-0,09	0	2	0	16	23	11
30	10	p	n$\bar{c}$	30%D	-0,34	-0,22	0	4	0	3	7	11
30	10	p	nc^{3Q}	15%D	-0,07	-0,06	0	1	0	6	27	26
30	10	p	nc^{3Q}	30%D	-0,16	-0,10	0	2	0	3	14	26
30	10	p	n$(\bar{c}+\sigma_c)$	15%D	0,00	0,00	0	0	13	13	22	10
30	10	p	n$(\bar{c}+\sigma_c)$	30%D	-0,08	0,00	0	0	0	11	31	10
60	5	d	nc^{2Q}	15%D	-0.08	-0.43	0	6	0	11	49	49
60	5	d	nc^{2Q}	30%D	-0.14	-0.43	0	6	0	8	55	49
60	5	d	n$\bar{c}$	15%D	-0.05	-0.42	0	7	0	3	101	41
60	5	d	n$\bar{c}$	30%D	-0.16	-0.42	0	7	0	1	47	41
60	5	d	nc^{3Q}	15%D	0.00	0.00	0	0	25	25	99	92
60	5	d	nc^{3Q}	30%D	-0.01	0.00	0	0	0	24	108	92
60	5	d	n$(\bar{c}+\sigma_c)$	15%D	0.00	0.00	0	0	25	25	94	85
60	5	d	n$(\bar{c}+\sigma_c)$	30%D	0.00	0.00	0	0	24	24	93	85
60	5	p	nc^{2Q}	15%D	-0.15	-0.38	0	6	0	8	39	30
60	5	p	nc^{2Q}	30%D	-0.23	-0.38	0	6	0	4	48	30
60	5	p	n$\bar{c}$	15%D	-0.15	-0.40	0	6	0	10	55	37
60	5	p	n$\bar{c}$	30%D	-0.23	-0.43	0	6	0	7	40	37
60	5	p	nc^{3Q}	15%D	-0.02	0.00	0	0	0	25	88	82
60	5	p	nc^{3Q}	30%D	-0.06	-0.04	0	1	0	24	45	82
60	5	p	n$(\bar{c}+\sigma_c)$	15%D	0.00	0.00	0	0	25	25	73	75
60	5	p	n$(\bar{c}+\sigma_c)$	30%D	0.00	0.00	0	0	25	25	74	75

4.4 Numerical results for uncertainty on utilities

In this section we focus on the robust model **uR-NHLP** where uncertainty lies on customer utilities. We classify the demand sources into two groups:

- Lower-than-average income: demand sources i with $I_i < \bar{I}$, where $\bar{I}$ is the average income over all cities, are considered more sensitive to price and better described by $w^0 = \{1/3, 2/3\}$.
- Higher-than-average income: demand sources with $I_i > \bar{I}$ are considered more sensitive to distance than price, and assigned a nominal weight vector $w^0 = \{2/3, 1/3\}$

Table 4: Results for a u**R-NHLP** with (n,m,P)=(30,15,4)

				# inf		# best		Time (s)	
Γ	B	$\bar{\Delta}_R$	$\bar{\Delta}_N$	R	N	R	N	R	N
5	nc^{2Q}	0.000	0.000	0	0	24	24	68	15
5	$n\bar{c}$	-0.001	-0.001	0	0	21	21	67	33
10	nc^{2Q}	-0.053	-0.001	0	0	0	21	58	15
10	$n\bar{c}$	-0.015	-0.003	0	0	1	17	74	33
15	nc^{2Q}	-0.056	-0.120	0	3	0	22	70	15
15	$n\bar{c}$	-0.004	-0.004	0	0	19	19	87	33
20	nc^{2Q}	-0.043	-0.400	0	10	0	14	151	14
20	$n\bar{c}$	0.000	0.000	0	0	20	20	50	28
25	nc^{2Q}	-0.035	-0.680	0	17	0	8	120	14
25	$n\bar{c}$	0.000	0.000	0	0	23	23	46	28

In Table 4 we see again that the robust solution y_R is more interesting than the nominal solution y_N when the uncertainty budget Γ is quite large and the budget constraint is in the tightest scenario $B = nc^{2Q}$. The nominal solution starts to be infeasible for some of the 25 deviated instances from $\Gamma = 15$, i.e. $0.5n$, and is very often infeasible for greater Γ (# inf = 10 out of 25 for $\Gamma = 20$, and inf = 17 out of 25 for $\Gamma = 25$. Nevertheless, for smaller values of Γ, the nominal solution behaves very well with no infeasible solution and a higher proportion of best solutions. We can conclude that the price of robustness, i.e. the price to pay for systematic feasibility, seems to be higher for uncertainty on utilities.

5 Conclusion

We have presented a robust approach for selecting new housing programs, where uncertainty lies first on demand volumes, then on customer utilities. We showed that when uncertainty lies on utilities with a discrete set of deviation scenarios from the nominal data, the robustness subproblem with binary variables can be replaced by its linear relaxation with no consequence on the original model, which enables to use the classical dualization method. Since consumer behaviors are modeled by a choice model, we used the Irrelevance of Independent Alternatives (IIA) property to describe the flows from demand sources to selected sites in our models, which are reformulated in a linear way. Numerical experiments on data of the Paris region show from which level of the uncertainty budget and which degree of tightness of the budget constraint, the robust model starts to be more accurate than just using the optimal solution of the nominal problem, that becomes more and more infeasible when these parameters increase. To our knowledge, this paper is the first contribution of optimization to selecting new housing programs with a choice model, beyond the robustness aspect.

References

Alumur, S. A., Nickel, S., & Saldanha-da-Gama, F. (2012). Hub location under uncertainty. *Transportation Research Part B: Methodological*, 46(4), 529-543.

Alvarez, P. P., & Vera, J. R. (2014). Application of robust optimization to the sawmill planning problem. *Annals of Operations Research*, 219(1), 457-475.

Aros-Vera, F., Marianov, V., & Mitchell, J. E. (2013). p-Hub approach for the optimal park-and-ride facility location problem. *European Journal of Operational Research*, 226(2), 277-285.

Baron, O., Milner, J., & Naseraldin, H. (2011). Facility location: a robust optimization approach. *Production and Operations Management*, 20(5), 772-785.

Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations research*, 52(1), 35-53.

Bertsimas, D., & Thiele, A. (2006). A robust optimization approach to inventory theory. *Operations Research*, 54(1), 150-168.

Billionnet, A., Costa, M. C., & Poirion, P. L. (2014). 2-Stage Robust MILP with continuous recourse variables. Discrete Applied Mathematics, 170, 21-32.

Bohle, C., Maturana, S., & Vera, J. (2010). A robust optimization approach to wine grape harvesting scheduling. *European Journal of Operational Research*, 200(1), 245-252.

Bulent, U., & Kenneth, B. H. (2005). Impact of local public services and taxes on dwelling choice within a single taxing jurisdiction: a discrete choice model. *Journal of Real Estate Research*, 27(4), 427-443.

Denoyel V., Alfandari L., & Thiele A. (2014), Optimizing healthcare network design under Reference Pricing and parameter uncertainty. Optimization on-line.

Driant, Jean-Claude translated by Christina Mitrakos. Why Build New Housing? Metropolitics, 28 January 2011. URL : http://www.metropolitiques.eu/Why-Build-New-Housing.html

Driant, Jean-Claude and translated by Oliver Waine. Why isnt there enough housing in France? Metropolitics, 11 January 2012. URL : http://www.metropoliti ques.eu/Why-isn-t-there-enough-housing-in.html

Eurostat. Construction cost index overview, January 2015. URL : http://ec.europa.eu/eurostat/statistics-explained/ index.php/ Construction _cost _index _overview

Gabrel, V., & Murat, C. (2010). Robustness and duality in linear programming. Journal of the Operational Research Society, 61(8), 1288-1296.

Gabrel, V., Murat, C., & Wu, L. (2013). New models for the robust shortest path problem: complexity, resolution and generalization. *Annals of Operations Research*, 207(1), 97-120.

Gabrel, V., Murat, C., & Thiele, A. (2014). Recent advances in robust optimization: An overview. *European Journal of Operational Research*, 235(3), 471-483.

Gilbert, A. (2011). Ten Myths Undermining Latin American Housing Policy. Revista de Ingeniera, (35), 79-87.

Guadagni P.M., Little J.D.C., (1983) A logit model of branc choice calibrated on scanner data, *Marketing Science* 2(1), 203-238.

Gülpinar, N., Pachamanova, D., & Çanakoglu, E. (2013). Robust strategies for facility location under uncertainty. *European Journal of Operational Research*, 225(1), 21-35.

Haase, K. (2009). Discrete location planning. *Institute of Transport and Logistics Studies Working Paper*, (ITLS-WP-09-07).

Haase, K., & Müller, S. (2014). A comparison of linear reformulations for multinomial logit choice probabilities in facility location models. *European Journal of Operational Research*, 232(3), 689-691.

Müller, S., & Haase, K. (2014). Customer segmentation in retail facility location planning. *Business research* 7(2), 235-261.

Pagliara, F., Preston, J., & Simmonds, D. (2010). *Residential location choice*. Dordrecht: Springer.

de Palma, A., Picard, N., & Waddell, P. (2007). Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater Paris region. *Journal of Urban Economics*, 62(2), 204-230.

Sandoval Alarcn, Francisco. Familias abandonan 2 millones de viviendas por no poder pagarlas. Animal Politico. 8 February 2012. URL : http://www.animal politico.com/2012/02/familias-abandonan-2-millones-de-viviendas-por-no-poder-pagarlas/#ixzz2sYtFCloh

Snyder, L. V. (2006). Facility location under uncertainty: a review. *IIE Transactions*, 38(7), 547-564.

United States Census Bureau. Value of Private Residential Construction Put in Place excluding rental, vacant, and seasonal residential improvements - Not Seasonally Adjusted. URL : https://www.census.gov/construction/c30/pdf/residentialnsa.pdf.

Zokaee, S., Jabbarzadeh, A., Fahimnia, B., & Sadjadi, S. J. (2014). Robust supply chain network design: an optimization model with real world application. *Annals of Operations Research*, 1-30.

PARIS

ESSEC Business School
3, avenue Bernard-Hirsch
CS 50105 Cergy
95021 Cergy-Pontoise Cedex
France
Tél. +33 (0)1 34 43 30 00
www.essec.fr

ESSEC Executive Education
CNIT BP 230
92053 Paris-La Défense
France
Tél. +33 (0)1 46 92 49 00
www.executive-education.essec.fr

ESSEC Asia Pacific
2 One-North Gateway
Singapore 138502
Tél. +65 6884 9780
www.essec.edu/asia

SINGAPOUR

Contact :

Centre de Recherche
+33 (0)1 34 43 30 91
research.center@essec.fr

affilié à la
CCI PARIS ILE-DE-FRANCE