Document de travail du LEM / Discussion paper LEM 2017- 15

Demand for housing choices in the North of France: a discrete approach

Véronique FLAMBARD
LEM UMR 9221 / veronique.flambard@univ-catholille.fr

http://lem.cnrs.fr/IMG/pdf/dp2017-15.pdf

http://lem.cnrs.fr/

Demand for housing choices in the North of France: a discrete approach

July 2017

Véronique Flambard

Université Catholique de Lille- FGES

LEM (UMR CNRS 9221)

60 boulevard Vauban

- CS 40 109, 59016 Lille Cedex France.

Tel: +33 3 59 31 50 66.

Fax:+ 33 3 20 13 40 40. Email: veronique.flambard@univ-catholille.fr

Abstract

Purpose:
This research analyses housing demand in northern France with respect to sociodemographic variables and the distance between the residence and the workplace.

Design:
Residential choices are estimated with a flexible heteroscedastic logit model, based on a discrete choice model.

Findings:
Econometric estimation confirms that residential choices are influenced by the age, the income and the size of the household, as well as by the rent to income ratio. An increase of any of these variables decreases the probability of choice of all the alternatives other than the most often chosen alternative (which is for this application house ownership in the suburb). Moreover the distance to work systematically influences the housing choice for single parent families and two-earner households. Additionally, preferences are found to significantly differ between local housing markets, specifically between Lille (a large agglomeration and capital city of the North area) and Dunkerque (an industrialized area). The geographical areas are defined based on Insee Employment Zones ("zones d'emplois").

Originality value:
This research represents one of the few applications of discrete choice models using French census data. Renting a dwelling in the social sector is explicitly modelled as a possible housing choice in this analysis, which is seldom done in the literature. The methodology and the issues raised are relevant for other European countries.

Key words: housing demand, location choice, job-housing balance, discrete choice model, heteroscedastic logit model

JEL classification : R21, C25

1. Introduction

There is no doubt that the types of housing units built and their location have profound implications for social mixity, urban sprawl and employability. However, housing demand is even more important in explaining these phenomena because households self-select based on tenure, type of building and location. For example, low income households will likely rent affordable apartments. Large families will tend to prefer houses to apartments. Households valuing job accessibilities will prefer high accessibility locations. Therefore having a better understanding of housing demand, in relation to existing housing supply is important for policy purposes and is the goal of this research. In addition, residential choice represents an important decision for a household and deserves attention for that reason. As a matter of fact, this choice will affect its well-being through housing cost, job opportunities, lifestyle, amenities or social networks. This choice is not an easy one because the household has to choose several dwelling dimensions: tenure (owning or not), type (house or apartment) and location (city centres or not) among others. Housing choices are likely to depend on the household's socio-demographic characteristics and on the housing characteristics including the housing cost.

Many studies concerning residential choices limit themselves to one aspect, which for the most part has been tenure choice. The challenge is to find the right number of dimensions of choices to analyse; having too many dimensions is intractable, whereas too few don't provide enough information. Parsimony is therefore desirable for that purpose. Here, twelve alternatives are considered. They are the alternatives that result from the combination of the three key choices of housing demand: tenure, type and location. Social housing is usually excluded in such models, leaving aside the choices of nearly one third of the households in northern France. This research fills that gap by explicitly including social renting in the analysis. It is also one of the only applications we are aware of for France, with the notable exception of De Palma et al. (2007). The objective of this work is to explain residential choices among a defined set of alternatives, based on households' socio-demographic characteristics and distance to their workplace. A survey will also be used to consider residential mobility projects in respect to the set of chosen alternatives.

Section 2 reviews the literature. Section 3 summarizes the econometric method. Section 4 presents the data. Section 5 outlines the main results and Section 6 concludes the analysis

2. Literature review for housing choices

Housing is a complex good with many attributes. Three important dimensions of dwelling choice are tenure (owning or renting in the private or social sector), dwelling type (detached house or not) and location (centre/suburb). In this research, the goal is to relate these residential choices to the socio-demographic characteristics of the household. For that purpose, firstly, the findings of the literature *by dimension* are explained and secondly the research addressing *multiple dimensions* are reviewed.

2.1. Explaining households' choices with respect to tenure, type or location

The choice of tenure has been the most extensively studied dimension of residential choice, both theoretically and empirically (for a review see Gobillon (2003)).

2.1.1. Tenure (owning, renting in the private or social sector)

Two different strands are present. In the first one, a household is willing to purchase if the utility of owning is greater than of renting. The utility may vary with the age, the number of children and the housing cost, according to the life cycle theory. Sinai and Souleles (2005) add the risk dimension. They confirm that the probability of homeownership increases with *rent* fluctuation risk and with longer time horizon for a given level of housing *price* fluctuation risk.

The second strand of the literature fully models the possible constraints on dwelling choice. Firstly, households may face credit market rationing. Secondly, the supply can be specialized: some goods may only be available in the countryside for example. Finally, neighbourhoods have different social or ethnic characteristics and can explain residential self-selection or/and residential discrimination and segregation. Haffner (2009) looked at the choice between the social and the market renting sector. In France, both sectors are about of equal sizes. The quality of dwellings is comparable, but social rented dwellings are usually apartments rather than houses, tend to be slightly bigger, are mainly concentrated in the zones of higher population density and most of all are much cheaper. The two sectors are however substitutes for each other in the intermediate price segment of the rental market. The second important choice is the dwelling type (apartment or house) presented in the next sub-section.

2.1.2. Type (house or apartment)

Most relocating households prefer detached or semi-detached homes with private gardens to apartments (see Senior et al. (2004)). With hedonic models, features associated with an apartment or a house can be estimated. Moreover, several empirical studies question the acceptability of more sustainable high density dwelling as opposed to less sustainable detached houses. As a matter of fact, households are often reluctant to live in apartments in big tower blocks. See Buys and Miller (2012) for insights into some desired features of dwelling design, and neighbourhood to enhance the acceptability of high-density living. The third key choice for a household is the location.

2.1.3. Location (city centre or suburb)

On one hand, urban theory (see Fujita (1989)) focuses on the fact that households may have to sacrifice space for more service or employment accessibility. On the other hand, Tiebout (1956) postulates that households vote with their feet by choosing the community with the public goods most closely aligned with their preferences. Hanushek et al. (2014) unifies these two propositions into a single model. Brueckner et al. (1999) go further by showing how preferences depend on the socio-demographic variables. They show that the relative location of different income groups also depends on the spatial pattern of amenities. When the centre has a strong amenity advantage over the suburbs, the rich are more likely to live in central locations. As a matter of fact, two opposing forces are identified by the model to explain this finding. Firstly, wealthy individuals are more attracted by low housing price in the suburb because they prefer larger homes. Secondly, people who are well-off tend to have a high opportunity cost of time and thus a high commuting cost per mile. Brueckner et al. (1999) assume that the ratio of commuting cost to housing consumption cost falls with income. Therefore the effect of higher consumption dominates and the rich tend to live in

the suburbs for similar levels of amenities in city centres and in the suburbs. They also assume that the marginal valuation of amenities rises sharply with income. If amenities fall rapidly with distance to the inner city, the attractiveness of the amenities of the centre combined with the higher marginal valuation of the amenities of the affluent people can be strong enough to dominate and to pull the rich toward the centre. This model can explain why high-income residents in U.S. urban areas tend living in the suburbs and that in many cases it is the reverse in Europe. In France, inner cities are usually wealthy districts but some pockets of poorer households also exist. Some contributions acknowledge that location choice for a two-earner households is more complex. Deding et al. (2009), for instance, show that the probability of moving increases with the commuting distances of *each* person (of the couple) and decreases with the distance between workplaces. So far in the literature review, residential choices have been analysed separately (for example the focus was on tenure ignoring the home's location). The next sub-section broadens the analysis to simultaneous choices.

2.2. Modelling the residential choices

Researchers have relied on at least three methods to model multiple residential choices: gravity equation models, bid-rent gradient theory and discrete choice models. The *gravity model* offers an aggregate view of residential choices. Lowry (1964) assumes that households locate by taking into account the availability of retail and services and vice-versa with the gravity equation modelling the distance between the former and the latter. The *monocentric model* of Alonso (1964) considers that job opportunities are in the city centres and that a household arbitrages between housing's size, distance from the city centre and its consumption of non-housing goods and services. Extensions have been proposed to allow multiple centres or to account for other observed factors (for example distance to quality public goods or crime rates in the neighbourhood) or unobserved ones with the seminal work of Ellickson (1981). But the most used model is the *discrete choice model*. McFadden (1978) was one of the first to contribute to this approach by modelling the residential choices as *qualitative* choices between different alternatives. This type of model has some microeconomic foundations. As a matter of fact, the decision maker chooses the alternative that provides the highest utility among a mutually exclusive and exhaustive set of finite alternatives. The probability that a household chooses a particular alternative is expressed as a function of observed independent variables. Let us summarize some findings obtained with discrete choice models and specification issues.

2.2.1 Some findings with discrete choice models

Quigley (1985) explains that when a consumer chooses a dwelling unit, he also "selects a set of housing characteristics, neighbourhood and public service amenities and a journey to work" with a rent or purchase price. Families with different socio-economic and demographic variables have preferences over the set of services offered by the physical characteristics of the dwelling and the social and economic characteristics of a neighbourhood with all services available in close proximity and accessibility of that neighbourhood to the activities of the household. Quigley shows that in the Pittsburgh's metropolitan area households prefer single detached dwellings to duplexes or apartment dwellings. However accessibility to their workplace significantly affects their housing choice.

2.2.2 Specification issues with discrete choice models

Yates et al. (2006) have analysed the order sequence of three residential choices that will be addressed in this research: tenure, type of dwelling and location. They argue that the most intuitive decision making process is to choose tenure first, then dwelling type and finally location. But they find this is not always the case in practice and that the choice's sequence depends on households' preferences and market characteristics. The econometric specification should therefore be flexible and not impose the same decision making process for all households. *Heteroscedastic logit model,* for which the variance of the unobserved factors differs for alternatives or individuals (Bhat, 1995; Hensher, 1999), offers this advantage. Indeed, this category of model is general as it nests the conditional logit model. Furthermore, the heteroscedastic logit model is also the one that fits our data best as demonstrated in Section 5.1. Households' heterogeneity preferences will hence be captured.

3. Theoretical framework and econometric modelling

The microeconomic structure underlying the econometric model is the following: a consumer faces a finite number of mutually exclusive alternatives of which exactly one has to be chosen. The residential choices are treated as a categorically distributed dependent variable in the model. Even with the modest ambition of distinguishing tenure (private renting, social housing, owning), type (single-family or multiple-family housing) and location (centre, suburb) there are already 12 alternatives (3*2*2). The twelve categories are large enough to assume that all possibilities are available to every household. In other words, no constraints or rationing is introduced in the model. Indeed, at an aggregate level, any disequilibrium is likely to be much smaller, than at the scale of a neighbourhood or a town.

3.1 Model specification

Following McFadden (1980), utility is supposed to be a random function (equation (1)) to take into account the inability of households to discriminate perfectly between the possible choices and of the analyst to model it fully. In the present context, the econometric model is used to predict the *probability* P_{in} (equation (4)) that a household chooses a residential alternative i (one among the twelve) given a set of independent variables characterising both the households and the alternatives. The probability will be modelled by (3) and is derived from the household's utility summarized by (1).

More specifically, a sample of N consumers is assumed to choose among L discrete alternatives. Each alternative i brings an utility u_{in} to consumer n. The utility u_{in} takes the form of an index composed of a deterministic component v_{in} and an additive disturbance ϵ_{in} as follows:

$$u_{in} = v_{in} + \epsilon_{in} \qquad (1)$$

The deterministic part v_{in} is a function of the properties of alternative *i* (such as its price) as well as of the consumer n characteristics (such as its income). It can be written:

$$v_{in} = \beta X_{in} \qquad (2)$$

where X_{in} is a vector of characteristics of consumer *n* interacted with alternative *i*.

A consumer *n* maximises its utility when he prefers alternative *i* over alternative *j* if and only if $u_{in} > u_{jn}$. The probability to choose alternative I over j is therefore given by:

$$P_{in} = Prob(u_{in} > u_{jn} \forall\, i \neq j) = Prob(v_{in} + \epsilon_{in} > v_{jn} + \epsilon_{jn} \forall\, i \neq j)$$
$$P_{in} = Prob(\epsilon_{in} - \epsilon_{jn} > v_{jn} - v_{in} \forall\, i \neq j) \qquad (3)$$

Following Hensher et al. (1999), the heteroscedastic logit model is then derived from equation (3) by assuming that the ϵ_{in} are IID Gumbel distributed[1] and that the ϵ_{in} are independent across alternatives. The source of randomness in utility u_{in} comes from the heterogeneity of tastes among households as well as from measurement or perception error of the econometrician or the household. The variance of the error term ${\sigma_n}^2$ will be supposed to depend of a scale parameter λ_n which depends itself on household characteristics $Z_n\gamma$. If it were ignored it could cause the coefficient estimates to be inconsistent (see Yatchew and Griliches (1985)).

Following the methodology of Random Utility Maximisation models with the Gumbel distributed disturbance, the probability that alternative *i* is chosen by respondent *n* is given by:

$$P_{in} = \frac{\exp(\lambda_n v_{in})}{\sum_{i=1}^{I} \exp(\lambda_n v_{in})} = \frac{\exp(\lambda_n \beta X_{in})}{\sum_{i=1}^{I} \exp(\lambda_n \beta X_{in})} \qquad (4)$$

The error variance σ_n^2 is inversely proportional to the scale parameter λ_n:

$$\sqrt{\sigma_n^2} = \sigma_n = \frac{\pi}{\sqrt{6}} * \frac{1}{\lambda_n} \qquad (5)$$

with

$$\lambda_n = \exp(Z_n\gamma) \qquad (6)$$

This parametrised heteroscedastic model assumes households have different preferences modelled with the scale paremeter $\lambda_n \neq 1$ with $\gamma > 0$. However, when $\gamma = 0$, the parametrised heteroscedastic model reduces to the conditional logit with preferences assumed to be similar across households (since the scale parameter λ_n is then constant and equal to 1). A test for homoscedasticity is therefore a test for the error variance σ_n^2 being constant across consumers.

3.2 Estimation method

The objective is to estimate the probability that alternative i is chosen by household n modelled by equation (4). In this equation, the utility v_{in} associated with alternative i for household n is explained by independent variables X_{in} characterising both the alternative and the household, that is to say: a constant, a ratio rent over income, the average distance to work for three categories of families as well as income, age and age squared and finally the size of the household. All the characteristics of the households are interacted with the alternatives[2]. The estimated probability also depends on the scale parameter of equation (6)

[1] This Gumbel distribution is used to model the distribution of a maximum (or minimum) of *n* independently and identically distributed variables as *n* approaches infinity following Hensher et al. (1999). It is a more natural choice than the normal distribution since households are assumed to choose housing *by maximising utility*. Besides it is a necessary and sufficient condition for choice probabilities to be consistent with random utility maximization model.

[2] It is important to multiply the characteristics of the household making the choice with the dwelling alternatives because only **relative** differences and not their absolute levels are relevant for a choice among alternatives. The explanatory variables which are characteristics of the household are themselves constant

which depends on Z_n. The variables Z_n which are assumed to influence λ_n are: the household income, the household size and the age of the head of the household.

In practice, the parameter vector $\theta = (\beta, \gamma)$ is estimated using the maximum likelihood method. The log-likelihood function is given by:

$$LL = \sum_{n=1}^{N} \sum_{i=1}^{I} y_{ni} * \ln(P_{ni}) \qquad (7)$$

where $y_{ni} = 1$ if alternative *i* is chosen by the household *n* and zero otherwise and where P_{ni} has been defined in (4).
With the estimated parameter vector $\hat{\theta}$, the estimated scale parameter of equation (6) can be deduced:

$$\widehat{\lambda_n} = \exp(Z_n \hat{\gamma})$$

and finally the estimated standard error of equation (5) can also be calculated:

$$\widehat{\sigma_n} = \frac{\pi}{\sqrt{6}} * \frac{1}{\hat{\lambda}_n} = \frac{\pi}{\sqrt{6}} * \frac{1}{\exp(Z_n \hat{\gamma})}$$

The model is fitted using a maximum likelihood approach, using the procedure clogith of STATA which is based on the econometric specification presented above. The data and the empirical issues are now presented.

4. Data

The application focuses on residential choices in northern France (in a region called Haut-de-France, near the Belgium frontier). Firstly, the database is described. Secondly, descriptive statistics are presented.

4.1 Construction of the database

The data are a random sample of one out of every twenty observations drawn from the French Population Census of 1999. This is the last exhaustive census with very fine spatial levels collected simultaneously in France[3]. As a matter of fact, in 2004, the format of the census was changed to that of a rolling census.

The census documents the housing characteristics and the socio-demographic features of the households. There is information on household age, educational attainment, size and family structure, occupation and employment location. The housing characteristics are quite detailed for the following factors: the location, whether the unit is owned or rented, the size, the type of structure and the age of the building. The census provides the location of jobs of each worker in a household. The Euclidian distance between the residence and the workplace will be calculated and included in the analysis.

---

across alternatives and cancel out unless they are multiplied by dwelling alternatives. Consequently the only way they can affect choice probability is by having a different impact on the various alternatives. For example, the household income is interacted with all discrete choice alternatives except owning a house in the suburb which serves as a reference alternative. An estimated negative coefficient for an alternative (say for "owning a house in city centre") in the regression shows that the household is less likely to choose this alternative than the reference one (which is "owing a house in the suburb").

[3] http://www.statistics.gov.hk/wsc/STS017-P5-S.pdf

The population of interest in this research is composed of households living and working in Nord-Pas-de-Calais, in an ordinary dwelling (for example motor-home or caravan are not included in the analysis), with one family per dwelling. This study focussed on towns of more than 5,000 inhabitants because market housing prices, which will be needed, are not published for smaller villages. There are a total of 22,699 households included in the database.

Nord-Pas-de-Calais is one of the **most** urbanised and densely populated regions in France, but the average income per household is lower than the French average. The proportion of renter is above the national average and the region has one of the highest proportion of social sector renters (the owner-occupation rate is therefore lower than the national average). For more information about Nord-Pas-de-Calais region relative to France, see INSEE data[4].

Alternatives for residential choices

Three key dimensions of choices are studied. Twelve alternatives result from the combination of tenure (owned housing, market rented housing, social rented housing), type (house, apartment) and location (centre, suburb).

[Table 1 approximately here]

House prices

The French Census does not provide house prices. Hedonic imputed housing prices would require access to detailed housing transactions for all cities in northern France, which we did not have unfortunately. Instead, published market values of 2000[5] are used. The method is the following. The published valuations are given per square metre for different categories of dwelling according to the tenure (owned or rented), the type of good (apartment by number of rooms, detached house, semi-detached house…), the period it was built, three levels of comfort and the location. These values and the coefficients to apply to recalculate house prices and rents based on characteristics of owner-occupied or rented dwellings are available for the 165 towns of more than 5,000 inhabitants of our database. To illustrate the method, let us take one example, the number of rooms. In the published market value, a scale is provided based on the number of main rooms. The published market value has to be increased by 32.5% for a studio, by 15% for a T1 (one main room plus kitchen and bathroom) and has to be decreased by 2.5% for a T3 (three main rooms plus kitchen and bathroom) and T5 (5 main rooms) and by 7.5% for a T4 (the reference is the two main rooms house called T2). To define the quality of the dwelling, the variables comfort in the census plus the presence of a digicode, of an elevator and a garage are used. As a proxy for amenities, the town inventory (Inventaire communal de l'INSEE de 1998) is used to characterize the equipment of each town in terms of residential services.

[4] See http://www.insee.fr/fr/themes/document.asp?reg_id=19&ref_id=12265#inter1

[5] Each year, market values are published by the edition Callon: http://www.editions-callon.com/valeurs-venales/

In summary based on the published market value scales provided for the main housing characteristics, this study imputes a market value per square metre for all owner-occupied dwellings[6]. Then this imputed value is multiplied by the surface area to obtain the housing price. For rented dwellings, rental values are calculated in a similar manner with the rental scales rather than the owner-occupied scale. For social housing units however, it is necessary to use a different method because the concept of market value does not apply to government subsidized housing units. The average rent per square metre for social rented units in 1999 (that is 4.30 euros per square metre) is used to calculate the rent of all social rented dwellings. The rent is then calculated by multiplying 4.30 by the dwelling's surface area provided in the Census. The average imputed rent for the private sector is 35% higher than the average imputed rent for the social sector. This corresponds to the expected difference of 30% to 40% according to housing surveys in France. Another major feature of the database has been the imputation of the household income.

Income

As a matter of fact, household income is not available in the French Census. To compensate, income I for household *i* is estimated by using the French Housing Survey (Enquête Nationale Logement of INSEE 2002) based on the following household characteristics C: social category, last diploma earned, age of the household head, age squared, dummy variables for male, unemployed and civil servant.

$$I_i = \alpha + \delta C_i$$

The estimation results are presented in Table 5 in the appendix. Only data from Nord-Pas-de-Calais is used since there are large income differences between regions in France. In the Housing Survey, not only is the salary available but also all the other income sources. These estimated coefficients of the income equation are used, in a second step, to calculate the average estimated income for each household of the Census sample.

Other imputed variables

Finally, binary variables for the commuting distances to work are constructed by type of households. The distances between cities in the region Nord-Pas-de-Calais are calculated using the centroid provided by the software mapinfo. Distances between the residence and the workplace for each household are then deduced. For single families, binary variable "dsingfam" is defined such that it is zero for non-single families and it is equal to the average residence-workplace distance in the sample per housing choice otherwise. For this type of model, the dataset is organized so that for each household there is one row per alternative. The size of the dataset is therefore equal to the number of households that is multiplied by the 12 alternatives. In a similar fashion, binary variables denoted "d1worker" and "d2workers" are constructed for respectively the distance to the workplace for single earner households and two-earner households.

[6] The published market values are calculated based on all available housing transactions. The database from the census is obtained with a random draw of one out of twenty observations. Both are therefore representative of the housing market. A substitution principle can be used to calculate housing prices of the census based on published housing values.

4.2 Descriptive statistics

The socio-demographic variables used in the estimation of the model are described in Table 2. The average yearly income is higher in Lille, the capital city of northern France, by approximately 23,200 euros than the region as a whole whereas Dunkerque average income is below 22,000 euros. In the three studied zones, the family size is approximately equal to 3 and the housing cost represents approximately 32% of the household income.

[Table 2 approximately here]

Descriptive statistics (Table 3) shows that household's characteristics differ by alternative. Owners are older than renters and spend more on housing. Households in the social sector are the lowest income and larger families. The relative cost of housing is the highest for owned properties, and the lowest for social rented dwellings. Living in houses tends to be more expensive than in flats. Finally for a given type of dwelling and tenure, the user cost is higher in the centre than in the suburb.

[Table 3 approximately here]

The average distance to the workplace[7] also varies with the alternative chosen and with the type of family. Single-families tend to live closer to their job than couples with one or two workers (regardless of the alternatives), reflecting a strong preference for accessibility to their workplace, probably because they have a larger domestic burden. One-worker and two-worker families average distance to work depend on the chosen alternatives. It may suggest they have different preferences.

To complete the exploratory statistical analysis, a hierarchical clustering is used to identify similarities between households with respect to their residential choices. The active variables are the household's characteristics already described: estimated income, number of persons in the household and the age and age squared of the head of the household. The passive variables are the twelve studied housing choices. Tenure is the more discriminant factor, separating owners and tenants in two broad categories (mainly because of two variables: the income and the age of the head of the household). Owners have on average higher income and are older. The second most discriminant variable is the location (households tend to be wealthier in the suburb than in the centre). Finally the less discriminant variable is the type of housing (in houses, families tend to be of larger size and older than in apartments).

5. Econometric estimation and main findings

After the descriptive analysis, econometric regressions are used to estimate the model three times: at the regional scale of Nord-Pas-de-Calais and at a sub-regional scale for Lille and Dunkerque. Table 2 shows differences between the three samples. In Lille, households have a higher income on average and are younger. In Dunkerque, the rent to income ratio is higher. It will be interesting to see if preferences vary between these three areas or not.

[7] It would be very interesting as well to compare commuting time, but unfortunately the information is not available in the census. There is no reliable way to estimate it for the North of France (while there is for Ile-de-France).

5.1. Interpretation of the Estimation Results with Respect to the Econometric Specification

[Table 4 approximately here]

The estimated coefficients presented in Table 4 represent the taste weight β in the utility function (equation (2)) and indicates whether a factor is likely to increase or decrease the probability of choosing the reference alternative. The most frequent chosen alternative (an owned house in suburb, abbreviated "osh") is used as the reference alternative. This is done simply by omitting all household characteristics interacted with this alternative. A Lagrange multiplier test for heteroscedasticity[8] confirms the rejection of the homoscedastic model. The variance of the error term depends significantly on household income, age and size (see the last three lines of Table 4). The variance of the error term (which captures heterogeneity of preferences and/or measurement error as defined before in equation (5)) varies significantly with these three characteristics. This confirms that if heteroscedasticy had not been taken into account, some variables would have been wrongly ignored (inc_rpcf; age_rsch; age2_rsch; size_och; size_rpch) while others would not have the correct direction (size_rpsh; size_rsch and size_rssh).

The estimated model is consistent with the random utility model since all coefficients of the parametrised variance are negative, resulting in λ_n (defined previously in equation (6)) which are less than 1 for all households as it should be. For the Nord-Pas-de-Calais region as a whole and Lille, heteroscedasticity mainly comes from family size while it principally comes from the age for Dunkerque.

5.2 Determinants of the Residential Choices for Nord-Pas-de-Calais region

The interpretation of the coefficients of the estimated heteroscedastic logit model for the whole region Nord-Pas-de-Calais, in column two of Table 4, shows the following, everything else equal. An increase in income reduces the probability of choosing the eleven alternatives compared to the reference alternative (which is “osh”, owning a house in the suburb). Indeed, the characteristics of the household making the choice, such as income, are interacted with all the dwelling alternatives except the reference alternative “osh”. Therefore a negative coefficient for a dwelling alternative (say for “och”, owning a flat in city centre) indicates the household is less likely to choose “och” than the reference alternative “osh”.

All the coefficients of size by alternative (size-och to size-rssf) are also significantly different from zero at 5% and negative. Therefore households of larger sizes are more likely to choose the reference alternative “osh” than the other alternatives. The same thing can be said about the age: older households are more likely to choose the baseline alternative osh. The only exception is for owning a house in the centre for which the coefficient is not significantly different from zero meaning the age does not influence this particular choice. The effect of the age is nonlinear (since the coefficients of the age squared are significant at 5% for most of the alternatives and positive). The interpretation is the following: households

[8] Usual tests for parameter restrictions (likelihood ratio, Wald and Lagrange multiplier tests) can be employed to test the null hypothesis of homoscedasticity.

are more likely to become homeowners when they get older but the probability increases first quickly and progressively more slowly with the age.

In summary, for the area of Nord-Pas-de-Calais as a whole, the household size and age play a role in residential choices, which is consistent with the life-cycle theory. An older and larger family will more likely prefer a house (which is generally larger than an apartment). The probability of being an owner increases with income which is consistent with the tenure choice theory emphasizing credit constraints faced by some households.
Furthermore, all coefficients rentn-och to rent-rssf are significantly different from zero and negative. In other words, an increase in the price or in the rent of each alternative reduces its probability of being chosen[9]. It is something that is obviously expected.

Finally, both the composition of the household (single family, family with one or two workers) and the distance to work of the current job may influence the residential choice. The estimated coefficients for dsingfam, d1worker and d2worker capture these average effects. The results show that a longer distance to work reduces the probability to choose the reference alternative (”osh”) for single families and for two-income households compared to single persons (the coefficients of dsingfam and d2worker are negative), while it increases it for one-worker households (the coefficient of d1worker is positive). This suggests different trade-offs for the reference alternative, according to family groups. Single-families (and to a lesser extent, two-worker households), as shown previously by the descriptive statistics, live closer to their workplace in comparison to the other types of families. The multivariate analysis confirms it still holds everything else equal. Single and two-worker families are more likely to choose a location closer to their workplace than single persons. The effect is not significant for one-worker households.

5.3 Geographical differences in the Nord-Pas-de-Calais region

The model is also fitted on two other areas by restricting the population to a given employment zone: Lille on one hand and Dunkerque on the other hand. Results are presented respectively in column three and four of Table 4. The results are different than for Nord-Pas-de-Calais whole area both for the estimated function of the variance of the error terms (the coefficients of the parametrised variance toward the bottom of the Table differ), and for the significance of the variables in the regression (coefficients from cst-och to d2worker). This indicates that preferences can vary at a sub-regional scale (possibly because of differences in population characteristics and housing alternatives).

[9] The explanatory variables which are characteristics of the household are themselves constant across the alternatives and are interacted with all choices except one that serves as a reference choice or alternative. See footnote 1. In contrast, housing price or rent is not a household characteristic but a *housing* one. It differs across all alternatives. This is why housing price or rent is interacted with all choices, including the reference alternative. When the coefficient of rent interacted with an alternative (rent-och to rent-rssf) is negative, it shows that the household is less likely to choose that particular alternative. It is also true of the reference alternative since the coefficient rent-osh is negative. Indeed, the probability to choose one alternative (say “och”) is not defined relatively to the reference alternative “osh” for the *housing* characteristics as it is the case for the *household* characteristics.

Lille

Lille is the capital city of the Nord-Pas-de-Calais region. In the Lille employment zone, housing choices seem to globally respond to the same factors at the regional level with two exceptions. The age of the head of the household does not have a direct effect on housing choices (it plays a role only through the estimated scale parameter λ_n). The age coefficients for the Lille area (age-och to age2-rssf, in Table 4, column 3) are not significantly different from zero for most of the alternatives. In the whole Nord-Pas-de-Calais region however older households are more likely to choose the reference alternative. One reason for this could be that the population is not the same in the two areas and they have different preferences. For instance, the population is one year younger in Lille than in Nord-Pas-de-Calais (as shown in Table 2).

However, the distance to workplace does not have any impact on choosing the reference alternative for one-worker households in Lille (d1worker is not significant in the 3rd column of Table 4). However, the distance to work matters for single families and two-earner households (the coefficients of dsingfam and d2worker are negative). A plausible explanation is that accessibility tends to be high everywhere in the Lille area and that all alternatives are satisfactory enough when at least one member of the household stays at home. In sum, residential housing choices do not differ very much from those at the regional scale. Similarities and differences between the whole region and Dunkerque are now analysed.

Dunkerque

The Dunkerque employment zone is a highly industrialized area (especially in terms of metallurgical, chemical and pharmaceutical industries). Pollution and industrial risks are known to be high with a dozen major industrial risk sites and the largest nuclear power plant in Western Europe.

Preferences for city centre living in Dunkerque seem to be higher than elsewhere. Indeed, a higher income increases the probability of owning a house in the centre rather than in the suburbs (the coefficient of inc_och is positive in the last column of Table 4). It is likely that the centre of these cities have stronger amenities over their suburbs (marked by industrialization). If so, it would be consistent with Brueckner et al. (1999) who emphasises the relative level of amenities in the centre to explain choice location for European Cities. Grislain-Letrémy et al. (2013) show that proximity to “Seveso sites” (firms with major industrial risks) reduces housing prices and therefore exhibit negative externalities. City centres become in this context even more attractive than the suburbs suffering from negative externalities.

In addition, households are willing to spend more to rent a house in the private sector both in the centre and in the suburb (the coefficients of rentn_rpch and rentn_rpsh are positive rather than negative). This result could be the consequence of a scarcity of houses relative to flats in this area. According to the database a lower proportion of households rent a house in the private sector in the Dunkerque employment zone than in Nord-Pas-de-Calais (3.56% versus 4.87% in the centre of cities and 3.62% versus 6.02% in the suburbs). Besides, the share of social housing in Dunkerque (39% of all housing alternatives) is higher than the regional average (approximately 30%). The supply of this particular housing segment

(generally composed of affordable apartments) is well represented and reduces the attractiveness of apartments rented in the private sector.

The willingness to pay more for houses in Dunkerque than in the whole region of Nord-Pas-de-Calais could also be an indicator of stronger preferences for houses over apartments in this area or just come from the fact that the households have different characteristics. Indeed, in the Dunkerque area the yearly average income of the households located in the centre is higher than in the suburb (22,572 euros in the centre relatively to 20,619 euros in the suburb). It is different than what is observed at the regional scale and for Lille (22,544 euros against 21,139 euros for the whole region and 21,579 euros compared to 24,428 euros for Lille).

Finally, the Dunkerque employment zone also differs from the whole northern area for the distance to work for single worker households (d1worker has a positive and not a negative coefficient). For Nord-Pas de Calais, households are less likely to choose a residence when the distance to work increases. It is also the case in the Dunkerque area except for one worker household: this type of household tends to live farther away from their work when other parameters are controlled for (perhaps to enjoy a better dwelling or environment). In conclusion, residential choices appear to be quite specific in Dunkerque compared to the whole region of Nord-Pas-de-Calais

5.4 Stated choices- Results validation

With the French Census, housing choices have been analysed based on observed housing tenure, type and location (which reveal their preferences). To complete the analysis, a survey from Market Audit with household's choices and stated preferences for residential mobility is employed. The survey was administered to 1,313 respondents. The sample is representative of the households of the Nord-Pas-de-Calais region in September 2006.

Households were asked whether they were generally satisfied with their accommodation: 93% were satisfied (58% fully satisfied and 35% quite satisfied) and 7% were not (4% not really satisfied and 3% not satisfied at all). The housing choices analysed with the census should therefore provide information consistent with the household's stated preferences of the survey.

The main reason of desired mobility for the *renters* is the desire to *become an owner* (34%) of a house (9 out of 10 respondents). Households who desire to become home owners in the next two years are most often 25 to 34 years old and have a budget of 136,000 euros. Changing dwelling type (house versus apartment) is less often the stated cause of desired mobility (7%). Moving closer to workplace is not even mentioned. The average commuting time for all households is 18 minutes in the Nord-Pas-de-Calais and the main transportation mode is the car (in 81% of the cases). Indeed, the Nord-Pas-de-Calais is a region with good road and highway network which can explain why moving closer to work is not a priority.

For the *owners*, residential mobility is desired to get more space (22%). Owners claim they want to remain home owners and for most of them plan to buy a new larger house costing less than 192,000 euros. Other causes of stated project of mobility are the desire to have a garden (9%), to change the dwelling type from house to flat or vice-versa (7%) and to move

closer to work (5%). Owners tend to live further away from their workplace which could explain why for some of them proximity to workplace becomes an explanatory factor.

All the households who declare they would like to move are asked their three main criteria of choices. The main criteria are the environment (48%), the price (38%) and the size (37%). Proximity of the workplace is the 6th criterion after the comfort and the presence of a garden. The type of dwelling (house versus apartment) is the 8th criterion just after proximity to services (schools, nursery schools…). This confirms one finding of the analysis: proximity to workplace matters but is not at the top of the priorities of the household. In northern France where highways can be used free of charge, households with a range of location choices prefer the environment, space and houses over apartments even if it is at an increased distance from work. It is interesting to observe that tenure is not mentioned as a criterion of choice while it is the main stated reason for residential mobility (one third of renters said they would like eventually to move to become an owner). This survey confirms that the preferred alternative is being a home owner of a house. The environment and the space available are among the top three priorities of households explaining why the preferred alternative is osh (owning a house in the suburb). Being close to the workplace matters but is not among the top priorities for most of the households. The census analysis has shown that this close proximity is however important for single parents.

6. Conclusion

The main purpose of this paper has been to estimate a reliable model of residential choices for northern France. Almost no studies have been conducted using discrete choice models in France before and no attempt to include social housing exist. A limit of this approach is the need for parsimony in the number of considered alternatives. Twelve categories have been chosen resulting from the different combinations of three tenure statuses, two dwelling types and two locations (city centre and suburb). Three variables representing the socioeconomic background of the household (age, income, size) and two attributes of housing alternatives (housing cost to income ratio and the distance between the residence and the workplace) are the chosen explanatory variables. The econometric specification best fitting the probability to choose a housing alternative in northern France is the heteroscedastic logit model, which does not impose the homogeneity of preferences across households. The estimations are performed on the region Nord-Pas-de-Calais as a whole, as well as on two sub-areas which are Lille and Dunkerque. The geographical areas are defined based on Insee Employment Zones (“zones d’emplois”).

The findings of this research reveal that the distribution of housing choices varies with the geographical zone, the characteristics of the household and to a lesser extent the housing cost to income ratio. For the whole region, an increase of income, age, size of the household or housing cost to income ratio increases the probability of being an owner of a house in the suburb. The residential choices in the sub-area of Lille do not differ markedly from the whole Nord-Pas-de-Calais region but in the sub-area of Dunkerque, they do. In Dunkerque, households are more likely to prefer living in the city centre (as opposed to living in the suburb), probably because amenities in the centre make up for a less spacious dwelling and also because dangerous industrial activities are obviously not located in inner cities. The coefficients of income, size and age of the households are not significant in the regression

for Dunkerque. This result could be an indicator of a strong heterogeneity of preferences between households or of well-diversified housing supply within each housing alternative in this area.

This research was motivated by the desire to understand residential choices of households in order to make recommendations for real estate and urban policy makers. From this research, it can be firstly deduced that the supply of housing should be diversified to accommodate different household preferences and needs, stages in the life cycle (represented by the size of families and the age variable) and capacity constraints (the income variable). A lack of diversification would reinforce segregation by income level, type of families or age structure and would result in less social mixity. The analysis shows that, in northern France, households with similar characteristics tend to prefer the same type of real estate properties. Secondly, an important conclusion for sustainable development is the importance of improving city centre amenities relative to those of the suburb or of increasing the services associated with high density dwelling because clearly the most desired alternative remains a house in the suburb. This preference for a detached home, space and nice environment at an affordable price is clear from the actual choices of households in the census data and from their stated preferences in the survey. The case of the housing market in the Dunkerque area, which is characterised by a strong industrial landscape with port and heavy-duty industrial activities, sheds additional light on the circumstances surrounding urban sprawl in this context. In this area, amenities provided by the city centre become strong enough to offset the attraction of a house in the suburb. Moreover, the regression results show that workplace proximity is a desired feature but only for single families with dependent children and for two-worker households relative to single persons. In their case, domestic constraints likely explain why they are less willing to travel to access the preferred alternative (which is to own a house in the suburb). Proximity to the workplace is confirmed by the survey as not being a priority of households when other factors are considered. This can be explained in northern France by the existence of free highways in a context where 80% of the household go to work by car. Policy makers must be aware of the effects of better road accessibility on urban sprawl. Furthermore, the effectiveness of jobs-housing balance policies (which consists of building homes in close proximity to jobs) largely depends on the willingness of a significant number of households to opt for better access to work in exchange for less spacious and less green homes. Typically a large number of households need to be willing to live in an apartment close to the city centre in order for these policies to be successful. It does not seem to be the case to a larger extent in northern France.

Acknowledgments: My gratitude goes out to the editors and to two anonymous referees for their detailed reports which helped to improve the article. Also, I would like to thank Julien Salingue for helping with the database work.

References

Alonso, W. (1964), *The form of Cities in Developing Countries*, Papers in Regional Science, 13(1), 165-173.

Bhat, C. R. (1995), "A heteroscedastic extreme value model of intercity travel mode choice", *Transportation Research B*, Vol. 29 No 6, pp. 471-483.

Brueckner, J., Thisse, J.F. and Zenou, Y. (1999), "Why is central Paris rich and down-town Detroit poor? An amenity based theory", *European Economic Review*, Vol. 43 No 1, pp. 91-107.

Buys, L. and Miller, E. (2012), "Residential satisfaction in inner urban higher-density Brisbane, Australia: role of dwelling design, neighbourhood and neighbours", *Journal of Environmental Planning and Management*, Vol. 55 No 3, pp. 319-338.

Callon (ed.), *La côte annuelle des valeurs vénales, immobilières et foncières*, au 1er janvier 2000, vente et location, éditions Callon, 2000.

Deding, M., Filges, T. and Van Ommeren, J. (2009), "Spatial Mobility and Commuting: the Case of Two-earner Households", *Journal of Regional Science*, Vol. 49, pp. 113–147.

De Palma, A., Picard, N. and 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*, Vol. 62 No 2, pp. 204-230.

Ellickson, B. (1981), "An Alternative Test of the Hedonic Theory of Housing Markets", *Journal of Urban Economics*, Vol. 9 No 1, pp. 56-79.

Fujita, M. (1989), *Urban Economic Theory, Land Use and City Size*, Cambridge University Press, 366 pages.

Gobillon, L. (2003), « Une Synthèse de la Littérature sur les Choix de Logement des Ménages »; working paper (available at http://laurent.gobillon.free.fr/Papers%20in%20French.html).

Goodman, A. C. (1988), "An Econometric model of Housing Price, Permanent Income, Tenure Choice ,and Housing Demand", *Journal of Urban Economics*, Vol. 23, pp. 327-353.

Grislain-Letrémy, C. and Katossky, A. (2013), « Les risques industriels et le prix des logements », *Economie et Statistiques*, Vol. 460-461, pp. 79-106.

Haffner M., J., Hoekstra, Oxley, M. and Heijden H. van der (2009), "Substitutability between Social and Market Renting in Four European Countries", *International Journal of Housing Policy*, Vol. 9 No 3, pp. 241-258.

Hanushek, E. A., Leung, C. K. Y. and **Yilmaz,** K**. (**2014), "Borrowing Constraints, College Aid, and Intergenerational Mobility", ***Journal of Human Capital***,** Vol. 8 No 1, pp. 1-41.

Hensher, D.A. (1999), "HEV choice Models as a search Engine for the Specification of Nested Logit Tree Structures", *Marketing Letters*, Vol. 10 No 4, pp. 339-349.

Hensher, D., Louviere, J. and Swait, J. (1999), "Combining sources of preference data", *Journal of Econometrics*, Vol. 89, pp. 197-221.

Lowry I.S., A. (1977), *Model of Metropolis, The Rand Corporation, California*, 150 pages

Manski C. (1977), “The structure of random utility models”, *Theory and Decision*, Vol. 8, pp. 229-254.

McFadden, D. (1980), “Econometric Models for Probabilistic Choice Among Products”, *The Journal of Business*, Vol. 53 No 3, pp. S13-S29.

McFadden, D. (1978), *Modelling the choice of residential location*. In A. Karlquist et al., editor, Spatial interaction theory and residential location, pp. 75-96, Amsterdam, North-Holland.

Quigley, J. (1985), “Consumer choice of dwelling, neighbourhood and public services”, *Regional Science and Urban Economics*, Vol. 15, pp. 41-63.

Senior M., Webster, C. and Blank, N. (2004), “Residential preferences versus sustainable cities: Quantitative and qualitative evidence from a survey of relocating owner-occupiers”, *Town Planning Review*, Vol. 75 No 3, pp. 337–357.

Sinai, T. and Souleles, N.S. (2005), “Owner-Occupied Housing as a Hedge Against Rent Risk”, *Quaterly Journal of Economics*, Vol. 120 No 2, pp. 763-789.

Tiebout, C. M. (1956), “A Pure Theory of Local Expenditures”, *The Journal of Political Economy*, Vol. 64 No 5, pp. 416-424.

Yatchew, A. and Griliches, Z. (1985), “Specification error in probit models”, *Review of Economics and Statistics*, Vol. 67, pp. 134-139.

Yates, J. and Mackay, D.F. (2006), “Discrete choice modelling of urban housing markets: a critical review and an application”, *Urban Studies* Vol. 43 No 3, pp. 559-581.

Appendix

As explained in 4.1 we had to impute an income to every household in the database, according to their socioeconomic and education characteristics.
The results are presented in Table 5.

Table 1: Abbreviations used in the econometric modelling

A. Possible alternatives in the discrete housing choice model

Choice	*Tenure*	*Location*	*Type*	*Mnemonic*
1	Own	Centre	House	och
2	Own	Centre	Flat	ocf
3	Own	Suburb	House	osh
4	Own	Suburb	Flat	osf
5	Rent in private sector	Centre	House	rpch
6	Rent in private sector	Centre	Flat	rpcf
7	Rent in private sector	Suburb	House	rpsh
8	Rent in private sector	Suburb	Flat	rpsf
9	Rent in social sector	Centre	House	rsch
10	Rent in social sector	Centre	Flat	rscf
11	Rent in social sector	Suburb	House	rssh
12	Rent in social sector	Suburb	Flat	rssf

B. Explanatory variables

Variables	*Abbreviation used for this variable*
Intercept	Cst
Estimated income	inc
Age of the head of the household	age
Age squared of head of the household	age2
Family size	Size
Ratio rent over income (multiplied by 1,000)	Rentn
Average distance to the workplace for single-family	Dsingfam
Average distance to work for one worker-family	d1worker
Average distance to work for two workers-family	d2worker

Table 2: Summary statistics per geographical area

	Nord-Pas-de-Calais (whole area)	Lille (first sub-area)	Dunkerque (second sub-area)
Yearly average income (in euro)	21,913	23,181	21,459
Family size	3.0	2.9	3.0
Age	39	38	39
Ratio rent/income (in %)	31.5	31.5	31.9
Number of observations (households) in our sample	22,699	5,779	1,770

Table 3: Summary statistics per housing alternative for the Nord-Pas-de-Calais
(Mnemonic is used to denote alternatives, see notations in Table 1)
A. Distribution of the alternatives in the sample of Nord-Pas-de-Calais

Alternatives:	Mnemonic for this alternative	Proportion in the sample (in %)
1	och	17.31
2	ocf	2.14
3 (reference alternative)	osh	28.49
4	osf	1.05
5	rpch	4.87
6	rpcf	7.79
7	rpsh	6.02
8	rpsf	2.67
9	rsch	3.24
10	rscf	9.54
11	rssh	7.36
12	rssf	9.51

B. Characteristics of the households living in each alternative in Nord-Pas-de-Calais

	Yearly average income of household living in this type of housing (in euros)	Average family size living in this type of housing (in number of persons)	Age of household head living in this type of housing on average (in number of years)	Distance from work for single families living in this type of housing (in km)	Distance from work for one worker families living in this type of housing (in km)	Distance from work for two-workers families living in this type of housing (in km)	Monthly average user cost for families living in this type of housing (in euros)	Number of observations in the sample
och	26,384	3.2	43	5	7	9	692	3,930
ocf	22,863	2.3	42	3	6	9	463	486
osh	26,698	3.1	42	6	8	9	690	6,467
osf	24,028	2.3	41	5	6	9	452	238
rpch	19,891	3.1	36	4	8	7	561	1,106
rpcf	17,232	2.2	31	4	6	11	349	1,769
rpsh	20,552	3.2	37	5	8	9	549	1,367
rpsf	17,777	2.3	33	6	11	9	348	606
rsch	18,014	3.6	39	3	7	9	333	636
rscf	16,123	2.8	35	3	6	8	296	2,166
rssh	18,401	3.6	38	6	9	9	328	1,670
rssf	15,740	2.8	34	4	7	8	293	2,158

Table 4: Determinants of the probability of choice for the Nord-Pas-de-Calais region and two sub-areas that are Lille and Dunkerque

Zone	Homoscedastic logit Nord-Pas-de-Calais	Heteroscedastic logit Nord-Pas-de-Calais	Heteroscedastic logit Lille	Heteroscedastic logit Dunkerque
cst-och	2.482*	10.170***	30.022***	-23.782**
cst-ocf	19.107***	63.104***	78.997***	13.622
cst-osf	16.799***	70.367***	91.463***	-8.892
cst-rpch	5.326***	45.416***	73.627***	-5.854
cst-rpcf	14.066***	56.153***	60.607***	32.935***
cst-rpsh	5.017***	44.354***	33.466***	3.036
cst-rpsh	11.414***	55.996***	65.623***	36.706**
cst-rsch	0.994	52.685***	88.179***	4.950
cst-rscf	11.465***	47.225***	68.988***	34.442
cst-rssh	0.586	38.561***	55.320***	5.113
cst-rssf	10.112***	44.405***	55.111***	29.594***
inc-och	-0.053***	-0.171***	-0.483***	0.296**
inc-ocf	-0.286***	-0.896***	-1.316***	-0.206
inc-osf	-0.176***	-0.868***	-1.373***	-0.402
inc-rpch	-0.140***	-0.743***	-1.732***	0.16
inc-rpcf	0.005	-0.253***	-0.296***	-0.282*
inc-rpsh	-0.095***	-0.608***	-0.379***	0.117
inc-rpsf	-0.101***	-0.522***	-0.816***	-0.572*
inc-rsch	-0.396***	-1.420***	-2.683***	-0.919**
inc-rscf	-0.140***	-0.748***	-1.481***	-0.477**
inc-rssh	-0.155***	-0.995***	-1.398***	-0.227
inc-rssf	-0.127***	-0.751***	-1.175***	-0.550***
age-och	-0.068	-0.139	-0.367	0.633
age-ocf	-0.311***	-0.449*	-0.05	0.198
age-osf	-0.457***	-0.805*	-0.592	1.070
age-rpch	-0.198***	-0.918***	-0.661	-0.128
age-rpcf	-0.414***	-1.146***	-0.365	-0.768**
age-rpsh	-0.212***	-1.020***	-0.395	-0.527
age-rpsf	-0.271**	-0.906***	-0.447	-1.025*
age-rsch	0.125	-0.715***	-0.423	0.941
age-rscf	-0.366***	-1.042***	-0.858**	-0.896**
age-rssh	-0.065	-0.659***	-0.879	-0.237
age-rssf	-0.331***	-0.938***	-0.781**	-0.663*
age2-och	0.001	0.001	0.002	-0.008
age2-ocf	0.003**	0.002	-0.006	-0.004
age2-osf	0.004**	0.006	0.001	-0.013
age2-rpch	0.002**	0.006**	0.000	-0.002
age2-rpcf	0.0003***	0.006***	-0.009	0.006
age2-rpsh	0.002***	0.008***	-0.004	0.002

age2-rpsf	0.002	0.004	-0.005	0.009
age2-rsch	-0.001	0.006*	-0.001	-0.013
age2-rscf	0.003***	0.008***	0.004	0.008*
age2-rssh	0.001	0.005**	0.005	0.002
age2-rssf	0.003***	0.007***	0.003	0.005
size-och	0.055	-1.388***	-4.173***	0.230
size–ocf	-2.650***	-14.160***	-16.469***	-4.859**
size-osf	-3.045***	-15.976***	-16.099***	-4.520*
size-rpch	0.099	-5.156***	-8.768***	-0.764
size-rpcf	-2.122***	-11.091***	-14.170***	-3.848***
size-rpsh	0.209***	-4.101***	-7.616***	-0.067
size-rpsf	-2.072***	-12.698***	-13.447***	-2.543*
size-rsch	0.706***	-3.455***	-6.805***	-0.207
size-rscf	-0.812***	-4.936***	-5.758***	-1.094*
size-rssh	0.508***	-1.772***	-4.321***	0.265
size-rssf	-0.823***	-4.733***	-4.164***	-0.730*
rentn-och	-23.279***	-53.913***	-45.004***	6.815*
rentn–ocf	-44.116***	-96.034***	-95.083***	-14.133
rentn-osh	-19.999***	-48.173***	-31.560***	-4.591
rentn-osf	-35.997***	-93.183***	-105.500***	-14.173
rentn-rpch	-27.122***	-62.436***	-69.847***	13.844***
rentn-rpcf	-32.233***	-69.848***	-54.405***	-7.015
rentn-rpsh	-28.067***	-66.772***	-32.645***	11.450**
rentn-rpsf	-37.971***	-77.404***	-70.144***	-14.445
rentn-rsch	-64.975***	-147.336***	-180.891***	-57.235*
rentn-rscf	-45.320***	-98.090***	-110.767***	-22.774**
rentn-rssh	-48.052***	-122.331***	-97.380***	0.890
rentn-rssf	-42.723***	-96.543***	-86.205***	-16.624**
dsingfam	-0.088	-0.462**	-0.995**	-0.708*
d1worker	0.087*	0.267***	-0.117	0.469**
d2worker	-0.318***	-1.082***	-2.495***	-0.515*
Inclusive values				
iv for own	2.859***			
iv for private rental	2.559***			
iv for social rental	3.405***			
Parametrised variance				
Household income*100		-0.002***	-0.002***	-0.001*
Family size		-0.348***	-0.317***	-0.081
Age		-0.011***	-0.019***	-0.015**

Log-likelihood	-40,482.37	-40,297.229	-10,186.313	-3,374.595
Number of households	22,699	22,699	5,779	1,770
LL/number of households	-1.783	-1.775	-1.763	-1.907

Table 5: Coefficients of the estimated income

Variables	Coefficients
Intercept	7.14**
Socio-occupational categories	(ref: not in the labour force)
Farmer	0.38*
Self-employed businessman or manager	0.75**
Higher intellectual occupation	1.31**
Intermediate white-collar occupation	1.08**
Office workers	0.74**
Manual worker	0.88**
Retired (farmer)	-0.20*
Retired (white collar)	0.71**
Retired (blue collar)	0.30**
Unemployed (farmer, self-employed, etc)	0.74
Unemployed (higher intellectual occupation)	0.76*
Unemployed (intermediate white collar)	0.47*
Unemployed (office workers)	0.27*
Unemployed (manual workers)	0.20*
Education	(ref: middle school certificate or vocational high school diploma)
No diploma	-0.27**
Former primary-school certificate (CEP)	-0.17**
Professionnal or technological baccalaureate	0.09*
General baccalaureate	0.11*
University degree (undergraduate)	0.12*
University degree (graduate)	0.28**
Other variables	
Age of the father	0.04**
Age of the father squared	-0.00**
Male	0.35**
Long time employed	-0.22*
Civil servant	0.09*
Legend: * $p<0.05$; ** $p<0.01$ Number of observations: 2996 Adjusted R^2=0.41	