Published in Revue économique, 2021, vol. 72, no. 6, pp. 1055-1077,
which should be cited to refer to this work
DOI : 10.3917/reco.726.1055

Titre de l'article

Une approche hybride de la définition des prix de l'immobilier : Une étude de cas en Suisse Romande

Prénoms et noms des auteurs

Maurin Baillif
Matthieu de Lapparent
Evanthia Kazagli

Résumé en français = 750 signes

L'estimation des prix des biens immobiliers est une question essentielle dans un marché hétérogène où des sous-marchés peuvent exister sans que les critères d'identification soient connus ou mesurables. Nous proposons un modèle de prix hédoniques à mélange discret en appliquant un algorithme d'Espérance-Maximisation simulé ainsi qu'un critère de précision afin d'évaluer les différences entre le modèle à classe latente et le modèle hédonique classique. La méthodologie est appliquée à des loyer de Suisse Romande, augmentées spatialement en utilisant uniquement des données en accès libre. Les résultats montrent que l'utilisation de l'approche proposée améliore la précision de la détermination des loyers.

Titre en anglais

A Hybrid Approach to Real Estate Price Definition: A Case Study in Western Switzerland

Résumé en anglais = 750 signes

Accurate estimation of property prices is a key issue within a heterogeneous market where submarkets may exist without the identification criteria being known or measurable. We propose a discrete mixture of market hedonic-pricing model implementing a Simulated Expectation-Maximization algorithm and an accuracy criterion to evaluate the differences between the latent class model and the classical hedonic pricing model. The methodology is applied to rent data from Western Switzerland, spatially augmented using only open source data. The results demonstrate that the use of the proposed approach improves the precision of rent determination.

Mots-clés en français, mots-clés en anglais

Immobilier, définition des prix, régression hédonique, classe latente, algorithme d'Espérance-Maximisation simulé (SEM).

Real estate, Price definition, Hedonic regression, Latent class, Simulated Expectation-Maximization algorithm (SEM).

Classification *JEL* = pas plus de cinq items

C51, C52, C53, D41, R31

Pour chaque auteur : appartenance professionnelle, suivie des adresses postales et électroniques

Maurin Baillif : School of Management and Engineering Vaud. HES-SO // University of Applied Sciences and Arts Western Switzerland. Avenue des sports 20 - CP 521 - 1401 Yverdon-les-Bains, Switzerland. maurin.baillif@heig-vd.ch

Matthieu de Lapparent : School of Management and Engineering Vaud. HES-SO // University of Applied Sciences and Arts Western Switzerland. Avenue des sports 20 - CP 521 - 1401 Yverdon-les-Bains, Switzerland. matthieu.delapparent @heig-vd.ch

Evanthia Kazagli : School of Management and Engineering Vaud. HES-SO // University of Applied Sciences and Arts Western Switzerland. Avenue des sports 20 - CP 521 - 1401 Yverdon-les-Bains, Switzerland. evanthia.kazagli@heig-vd.ch @heig-vd.ch

Remerciements (ou l'indication qu'il n'y a pas de remerciements ; nous l'ôterons)

Pas de remerciements.

1 INTRODUCTION

Price prediction is an important topic in the real estate sector. It supports the development and management of real estate investment portfolios, the guidance for urban development policies, the guidance to individuals on purchasing or selling, as well as the protection of interests of the concerned stakeholders, namely the landlords, the tenants, the banks, as well as the investment fund providers. As of today, the prices and rents in Switzerland have been established, mostly, based on financial models that do not take into account —or that take into account to a limited extent— the attributes of the housing, its surroundings and the evolution of the local and global economic conditions.

The hedonic method of valuation is a common and well-known approach to price estimation that accounts for these dimensions. It is based on the assumption that the price of a good can be represented by the sum of the prices of its component features (Rosen [1974]). These include both intrinsic features of the good, such as the size of the dwelling, as well as external factors affecting its price, such as the interest rates or the surroundings of the property. The price difference between two goods is explained by the differences in the number of their features and the marginal price (or hedonic price) of each of these features. The hedonic price is calculated by regressing the price of the good against its features.

The majority of studies on real estate price estimation either assume homogeneity of the hedonic prices across the area of study (e.g. Baranzini et al. [2010]; Bishop and Timmins [2011]), or sub-markets defined based on geographical criteria (e.g. Ruf [2017]; Laferrere [2005]). Indeed, a traditional approach is to define different sub-markets (geographical zones, type of dwellings, etc.) and to develop specific regression models. We are here limited with data: we consider only apartments and deal with regional differences by means of cantonal fixed effects, assuming a priori that we can pool data, i.e. other regression coefficients are the same all over cantons. We recognize that it is a stringent assumption and that the presented results should be considered specific to the proposed application.

Homogeneity here refers to the case in which a common feature to all dwellings, for instance the number of rooms, has the exact same effect on the final price of all properties across the study area. However, one would expect that different features may be relevant to the evaluation of different dwellings, and subsequently their contribution to the final price may vary. As an example, in the same geographical area, the price of some dwellings may be more sensitive to features such as the noise or the accessibility to public transport, while for other dwellings, the floorspace or the number of rooms may be more important. Such differences can be explained by the type and the characteristics of the property and modeled by means of finite mixture models.

The theoretical framework of hedonic pricing is flexible, i.e. there are no constraints on the functional form or the choice of the variables to be included in the regression equation. In this work, we exploit this flexibility to model the unobserved heterogeneity of real estate properties with respect to the hedonic prices using a latent class approach (Lazarsfeld [1950]). In particular, we assume that multiple groups (or classes) of dwellings —the prices of which vary by different features— exist in the study area. In order to tackle this problem, we propose a mixture hedonic price model.

The objective of this study is twofold. Firstly, it is to highlight the existence of latent classes of real estate goods and the fact that they are not geographically attached to a location, e.g. to a city. Secondly, it is to identify the determinant factors, as well as the signs and weights of the corresponding regression coefficients, in order to analyze the differences among the classes of properties.

The proposed model is applied to rental apartments announcement data from Western Switzerland, collected from the internet. The results show that the latent class approach significantly improves, in our case study, the accuracy of rent determination in comparison with the standard model. Each of the classes is spatially distributed over the area, and not concentrated in one geographical zone. Finally, we observe that each class evaluates differently the characteristics of the properties in the process of rent determination.

The remainder of the paper is structured as follows. Section 2 presents the literature review. Section 3 delineates the proposed methodology. Section 4 describes the case study and the definition of additional variables that complement the available dataset. Section 5 presents the application of the proposed methodology to the case study and discusses the results. Finally, Section 6 summarizes the conclusions of this work and identifies directions for future work.

2 LITERATURE REVIEW

Hedonic price analysis is often associated with Rosen [1974] and his seminal work on the theoretical structure to the hedonic pricing model[1]. He defined the short- and long-term market equilibrium conditions and suggested a two-step procedure for the estimation of the housing demand function. The first step consists in the regression of prices on observable attributes of the good, and the computation of implicit marginal prices in each point. This step is often referred to as the *market valuation of housing characteristics*. It is the most common approach to empirical analyses. The second step uses the estimated marginal prices from step one as endogenous variables for the estimation of the housing demand functions.

Since these studies that laid the foundations of the hedonic price method, many theoretical works have focus on hedonic prices (Palmquist [2006]; Hill [2013]). The fundamental assumption is that a housing good and its location are inseparable, they are sold together, and the consumer transforms them to get utility. Based on this observation, a large number of assumptions can be made concerning the variables to be used to describe the price. From the first papers using only intrinsic variables, such as the surface area or the number of rooms (Gross et al. [1990]), various extrinsic variables such as the noise (Baranzini et al. [2010]), the accessibility (Löchl and Axhausen [2010]; Boucq and Papon [2008]) or the view (Cavailhès et al. [2009]) were added as data collection evolved. In a second stream of works, authors propose segmentation and analyze the presence of sub-markets. Different segmentations have been proposed. For example, Ruf [2017] has grouped the dwellings according to their geographical area, Hoesli et al. [1997] in relation to the type of the accommodation (house vs. flat), Cavailhès [2005] in relation to the size of the city and Baranzini et al. [2008] in relation

[1] More generally, studies on the price definition of complex goods existed before. Waugh [1929] first analyzed the prices of vegetables as a function of characteristics related to their quality. Court [1939] provided the first formal contribution to hedonic price analysis in the context of automobile demand, later popularized by Griliches [1961].

to the type of the buyer (expatriate or native).

Over the last few years, machine learning (ML) techniques have become extremely popular in a wide range of applications. Not surprisingly, they have also found application to the estimation of real estate prices. Several studies using ML techniques with "black box" effects, such as Random Forest, Gradient Boosting or Neural Network, can be found in the literature (see e.g. Oladunni and Sharma [2016]; Wezel et al. [2005]; Mayer et al. [2018]). These studies focus on the accuracy of the model and to a lesser extent on the economic interpretation of the coefficients. Unlike these "black box" algorithms, the Expectation-Maximization (EM) algorithm allows market segmentation and an economical interpretation of the results. Yet, very few studies in the real estate field use ML approaches based on the EM algorithm. For example, Howell and Peristiani [1987] applied it to the estimation of the demand and offer rent equation and Peng [2010] to the generalized repeat sales regression. Belasco et al. [2012] is, to our knowledge, one of the only studies using the EM algorithm to analyze the presence of latent classes and apply the hedonic method to three distinct submarkets. Our paper differs both on the method (EM vs. SEM) and on the data used (survey limited to one neighborhood vs. open data for an entire region).

This paper contributes to the current literature by proposing an implementation of the hedonic pricing model based on a simulated EM (SEM) algorithm that allows to analyze the differences among the latent classes. The model specification is based on open source online data and is extended to incorporate external variables, such as noise indicators and the accessibility to public transport. The SEM is preferred to the EM due to its fast convergence and the use of random draws at each iteration that prevent convergence to a local maximum (Celeux and Diebolt [1985]).

3 METHODOLOGY

This section delineates the proposed methodology, the backbone of which is the SEM algorithm. We first present the proposed mixture hedonic price model (MHPM). We then outline the SEM algorithm that is used for its estimation. Finally, an adaptation of the root-mean-square error criterion that accommodates the latent class approach is presented.

3.1 The mixture hedonic price model

Mixture models are probabilistic, semi-parametric models that account for the presence of unobserved sub-populations, i.e. latent classes, in the population of interest (Frühwirth-Schnatter [2006]). They are based on the assumption that each class is characterized by a different probability distribution. These are termed component distributions[2]. Given a population of real estate goods, each individual property i belongs to one and only one of the K underlying sub-populations. Here, we assume that the prices of real estate properties follow a log-normal distribution. Then, by definition, the log of the prices follow a normal distribution with the following probability density function :

$$f(\log(y_i)|X_i, \beta_k, \sigma_k^2) = \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{1}{2}\frac{(\log(y_i) - x_i^{'}\beta_k)^2}{\sigma_k^2}}, \tag{1}$$

where σ_k the standard deviation and X_i the features of a real estate property.

The logarithm of the price of a real estate good log(y_i) can be defined based on the typical hedonic price equation —the latter being specific to class k— as :

$$\log(y_i) = x_i^{'}\beta_k + \epsilon_{i,k}, \tag{2}$$

where $x_i^{'}$ are the features of a real estate property, β_k the regression parameters specific to class k, and $\epsilon_{i,k}$ the error term, with $\epsilon_{i,k} \sim \mathcal{N}(0,\, \sigma_k^2)$. For the time being, we assume that the same variables $x_i^{'}$ enter the hedonic price equation of all classes. As the class membership cannot be directly observed, the mixture distribution of each class k is express as :

$$P(\log(y_i)|X_i, \beta_1, \delta_1, \sigma_1^2, \ldots, \beta_k, \delta_k, \sigma_k^2) = \sum_{k=1}^{K} \delta_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{1}{2}\frac{(\log(y_i) - x_i^{'}\beta_k)^2}{\sigma_k^2}}, \tag{3}$$

[2] In this study, we further assume that all the component distributions belong to the same known parametric family.

where $\delta_1, \dots \delta_k$ are the mixing proportions, or else the weights of the components, for which $\delta_k \in [0, 1]$ and $\sum \delta_k = 1$ hold. Assuming that observations are i.i.d, the full information maximum likelihood function of the problem is defined as :

$$ll = \sum_{i=1}^{n} \log \left(\sum_{k=1}^{K} \delta_k \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{1}{2}\frac{(\log(y_i) - x_i^{'}\beta_k)^2}{\sigma_k^2}} \right). \quad (4)$$

In order to avoid dealing with complex non linearities, we build upon the SEM likelihood function. Equation 4 then becomes :

$$ll_{SEM} = \sum_{i=1}^{N} \sum_{k=1}^{K} d_{i,k}^{t} \left[\log(f(\log(y_i)|X_i, \beta_k^t, \sigma_k^{t^2})) + \log(\delta_k^t) \right], \quad (5)$$

where $d_{i,k}^t$ is a random assignment of each y_i, during iteration t, to one of the classes following some within-loop posterior distribution that depend on δ_k^{t-1} (see next subsection for details). The SEM algorithm, described in the next paragraph, is then used to estimate δ_k, β_k and σ_k.

3.2 Model estimation

The SEM is a stochastic extension to the EM algorithm [3,4] with the inclusion of an intermediate simulation step (S-step) between the E- and the M-steps. (e.g. Celeux and Diebolt [1985]; Celeux et al. [1995]; see also Grun and Leisch [2008] for practical implementation). In particular, an iteration of the SEM algorithm consists of the following three steps :

[3]The EM is a common approach to the maximum likelihood estimation of mixture models.

[4]The EM algorithm is an iterative process that was first proposed by Dempster et al. [1977] to obtain the maximum likelihood estimates of the parameters of a probabilistic model, when the latter depends on unobservable variables. The algorithm alternates between an expectation (E) step and a maximization (M) step. The E-step generates the function for the expectation of the log-likelihood that is evaluated using the current estimates of the parameters. Subsequently, the M-step calculates the parameters that maximize the expected log-likelihood function that was generated in the E step. The parameter estimates are then used to determine the distribution of the latent variables in the next E-step. The algorithm starts from an arbitrary initial vector of parameters at iteration 0.

E-step : For each observation, the objective is to estimate the expected outcome of class membership conditional to observed price. It is defined as a posterior probabilities, i.e. the probability that observation i belongs to class k given observed price $\log(y_i)$. Values of the parameters in the conditional posterior are estimated using $\hat{\delta}_k$, $\hat{\beta}_k$ and $\hat{\sigma}_k$ of the previous iteration, $t-1$.

$$\mathbb{E}(d_{i,k}|\log(y_i)) = \hat{\gamma}^t_{i,k} = \frac{\hat{\delta}_k^{(t-1)} \frac{1}{\sqrt{2\pi\hat{\sigma}_k^{(t-1)2}}} e^{-\frac{1}{2}\frac{\left(\log(y_i) - x_i'\hat{\beta}_k^{(t-1)}\right)^2}{\hat{\sigma}_k^{(t-1)2}}}}{\sum_{k=1}^{K} \hat{\delta}_k^{(t-1)} \frac{1}{\sqrt{2\pi\hat{\sigma}_k^{(t-1)2}}} e^{-\frac{1}{2}\frac{\left(\log(y_i) - x_i'\hat{\beta}_k^{(t-1)}\right)^2}{\hat{\sigma}_k^{(t-1)2}}}}, \quad (6)$$

where $\sum_{k=1}^{K} \hat{\gamma}^t_{i,k} = 1$.

Simulation-step : In order to reduce the risk of falling into a local maximum Celeux and Diebolt [1985] propose the introduction of a stochastic step between steps E and M that allows to better explore the surface of the log-likelihood function. Then, during iteration t, the individual class memberships $\hat{d}^t_{i,k}$ are pseudo-randomly drawn using values $\hat{\gamma}^t_{i,k}$ of step E, where $\hat{d}^t_{i,k} \sim \mathcal{M}(1, \hat{\gamma}^t_{i,k})$. Where $\mathcal{M}$ stands for the Multinomial distribution.

M-step : The augmented log-likelihood is maximized using the values of the previous step and Equation (5). The maximization step involves the computation of the following for each class: the beta coefficients of the regressions :

$$\hat{\beta}^t_k = \left[\sum_{i=1}^{N} \hat{d}^t_{i,k} X_i X_i'\right]^{-1} \left[\sum_{i=1}^{N} \hat{d}^t_{i,k} \log(y_i)\right], \quad (7)$$

the standard deviations of regressions :

$$\hat{\sigma}^{2,t}_k = \frac{\sum_{i=1}^{N} \hat{d}^t_{i,k} (\log(y_i) - X_i'\hat{\beta}^t_k)^2}{\sum_{i=1}^{N} \hat{d}^t_{i,k}}, \quad (8)$$

and the weighting of each class :

$$\hat{\delta}_k = \frac{\sum_{i=1}^{N} \hat{d}_{i,k}^t}{n} \tag{9}$$

the EM and SEM algorithms are asymptotically equivalent to the Full Information Maximum Lilkelihood (FIML) for parameter estimation (Celeux and Diebolt [1985]; Dempster et al. [1977]). However, when calculating t-stats, the EM algorithm requires the calculation of the Hessian at the convergence point, whereas the SEM algorithm allows the direct calculation of draws. It is the latter approach that is applied in this paper.

3.3 Model accuracy criterion

The root-mean-square error criterion (RMSE) measures the accuracy of a model using both the estimated and the observed values. The classic formula of the RMSE is adapted for the latent class model as follows:

$$RMSE_{SEM} = \sqrt{\frac{\sum_{i=1}^{N} \sum_{k=1}^{K} d_{i,k}(y_{pred,i,k} - y_i)^2}{N}}, \tag{10}$$

where $d_{i,k}$ is the membership of an observation i to a class k, N the overall sample size, y_i the observed value of an observation, $y_{pred,i,k}$ the estimated value for an observation in class k.

This criterion allows to compare the RSME of a standard hedonic regression and hedonic regression using the SEM algorithm.

4 DATA DESCRIPTION

This section presents the real estate announcements dataset for apartments in Western Switzerland that is used to demonstrate the methodology described in section 3. To complement this dataset, we define additional variables pertaining to noise levels and accessibility indicators.

4.1 Data sources

The dataset was constructed by combining several sources of open data. In particular, data on properties and their key attributes are extracted from announcements on the internet. The raw data contains 5682 observations for which information about the exact address, the geographic coordinates, the floor space, the number of rooms, the canton and the rent of each apartment was collected. Objects with surface above 500 square meters, as well as 0.5% of the least and the most expensive dwellings were removed from the sample[5].

The monthly rent per square meter is chosen as the dependent variable. In the remainder of this document, we refer to the rents as prices. Figure 1 shows the location and the price over floorspace of the collected real estate properties in the different cantons of Western Switzerland. The data is well distributed over the entire territory, with a higher concentration around the Lake Geneva region. The price over floorspace appears to be higher in cities (e.g. in Lausanne and Geneva). These observations are according to the expectation.

In order to enrich the online data, additional variables were created from various open-data sources. An important missing variable in the announcements is the age of building. This variable was extracted from the StatBL[6] survey made by FSO. The data is geo-referenced on a 100 by 100 meters grid. That is, the data provides the average age of the buildings that occupy a given cell. As it is not possible to define the exact age of each specific building, we use this data as a measure of the age of the buildings in the neighborhood.

Employment and population data were taken from the Federal Statistical Office (FSO). More precisely, the employment data comes from the STATENT-2015[7] survey, while the population data from the STATPOP-2016[8] survey. The

[5] A prototype of the model was created and tested using a sample of such data from the 1st and 2nd of November, 2017.

[6] https://www.bfs.admin.ch/bfs/fr/home/services/geostat/geodonnees-statistique-federale/batiments-logements-menages-personnes/batiments-logements-des-2010.html

[7] https://www.bfs.admin.ch/bfs/fr/home/services/geostat/geodonnees-statistique-federale/etablissements-emplois/statistique-structurel-entreprises-statent-depuis-2011.html

[8] https://www.bfs.admin.ch/bfs/fr/home/services/geostat/geodonnees-statistique-federale/batiments-logements-menages-personnes/population-menages-depuis-2010.html

data is aggregated at the hectare level.

For the definition of the noise level and accessibility indicators we use additional sources. In particular, information regarding noise exposure on a georeferenced 10 by 10 metes grid is provided by the SonBase[9] dataset of the Federal Office for Environment (FOEN). The data pertaining to road noise is from 2010, while the data pertaining to railways noise is from 2011. The definition of the noise indicators is presented in Section 4.2.1. The necessary information for the definition of the accessibility indicators, presented in Section 4.2.2, comes from the 2016 "Public transport stops"[10] database of the Federal Statistical Office (FSO).

[Insert Figure 1]

Remarks : One implication of the announcement data is that the reported rent on the announcement may not correspond the rent that the tenant actually pays[11]. As a consequence, the endogenous variable is subject to measurement error. However, this should only concern a very small proportion of the data.

The years of data collection of the various open data differ. We assume that this will have only a small impact on the results as (i) the orders of magnitude are representative at this aggregate level and (ii) no major changes, demographic or economic, have taken place over the last few years.

We have chosen to limit the number of standard (intrinsic) variables used in the analysis, while diversifying the information about the properties in question, by including characteristics of environment, i.e. the noise and accessibility indicators. Up to now, it has not been possible to extract information about some important housing features, such as the availability of a parking spot, the view or the vacancy rate.

[9]https://www.bafu.admin.ch/bafu/fr/home/themes/bruit/etat/banque-de-donnees-sig-sonbase.html

[10]https://www.bav.admin.ch/bav/fr/home/themes-a-z/geoinformation/geodonnees-de-base/arrets-des-transports-publics.html

[11]In Switzerland, a tenant can contest the rent after concluding the rental agreement.

4.2 Definition of additional variables

In what follows, we define the noise and accessibility to public transport indicators.

4.2.1 *Noise indicators*

Road Noise : Road noise is an important element in price estimation. Its definition is not straightforward though. Intuitively, the higher the level of noise, the greater the noise disturbance, and therefore the less attractive the property is. The actual noise impact depends on various factors, such as the time of day, the intensity, the duration, etc. A critical and complex dimension that comes into play is the perception of the individual regarding the noise disturbance. Drawing from the work of Schultz [1978] on noise disturbance, we use a day-night index proposed by Baranzini et al. [2010, 2008] for the hedonic housing model. The exponential form of this index allows to distinguish between the contribution of day- and night-time noise, by accounting for their effective duration and by increasing the night-time disturbance. It is defined as follows :

$$N_{i,dn} = 10 \log \left[10^{N_{i,d}/10}(15/24) + 10^{(N_{i,n}+10)/10}(9/24) \right], \tag{11}$$

where $N_{i,d}$ and $N_{i,n}$ are the noise levels by day and night.

Train Noise : Unlike road noise, train noise cannot be assessed solely by the noise index, as the proximity to the train station comes into play and complicates the identification of the noise disturbance. On the one hand, the noise results in a disturbance that reduces the attractiveness of the property. On the other hand, it implies proximity to a train station, i.e. higher accessibility to public transport and, in consequence, it increases the attractiveness of the property.

Therefore, in order to capture this phenomenon, we propose a non-linear relationship among the price, the noise and the distance from the station, representing the attractiveness of a nearby train station. Inuitively, the definition of the attractiveness postulates that, ceteris paribus, the price should increase when the distance decreases, but it should decrease when the noise increases. It is defined in the following way :

$$T_i = \frac{1}{N^2_{i,dn\ Train} + D^2_{i,j} + 1}, \tag{12}$$

where $N_{i,dn\ Train}$ the day-night train index and $D_{i,j}$ the distance to the nearest train station. The attractiveness of all properties is affected by road noise. This is not the case for train noise. The indicator associated with the latter is only applied when the train noise is positive, and is set to zero otherwise. Figure 2 illustrates the relationship among the price, the noise index and the distance from the station.

[Insert Figure 2]

4.2.2 *Accessibility indicator*

A simple way to define accessibility is the Euclidean distance (Vickerman [1974]). The gravity model improves the Euclidean model by taking into account topographic effects (Hansen [1959]). In this case, the effort to go from i to j is measured by a *generalized cost* and structured so that accessibility decreases as distance increases. The choice of the functional form is open to interpretation. Here, we choose the negative exponential, which is the most common functional form and the most representative of the actual behavior. The accessibility indicator A_i can then be express as :

$$A_i = \sum N_j \exp^{-\lambda CG_{i,j}}, \tag{13}$$

where N_j is the number of opportunities at location j, λ a sensitivity parameter fixed to 1 and $CG_{i,j}$ the generalized cost between i and j. It is a variant of the usual Hansen indicator in spatial geography (CG appears in log form), yet it matches a consumer surplus formulation in microeconomics The generalized cost is defined using the notion of the "effort kilometer" [12] [13] (kme), accounting for mountain trails. The kme transforms the notion of the elevation gain into kilometers, the positive difference in elevation being more penalized than the negative.

[12]The use of this concept is a working assumption whose values may vary.

[13]https://randochevalblog.wordpress.com/2017/01/27/notion-de-kilometre-effort-km-e/

This allows to transform a distance between two points with an altitude difference to what it would be on a flat surface. The standard kme is defined as follows :

$$kme_{i,j} = \frac{d_{i,j}}{1'000} + \frac{\sum Z^{+}_{i,j}}{100} + \frac{\sum Z^{-}_{i,j}}{300}, \tag{14}$$

where $d_{i,j}$ the distance between the dwelling and a bus stop, subway or train station. $Z^{+}_{i,j}$ the positive altitude difference and $Z^{-}_{i,j}$ the negative altitude difference. All variables are expressed in meters. A distinction is made between a positive and a negative gradient in the sense that, at the same distance, the going up requires greater effort than the going down. Thus, a 800 distance to a bus stop with a positive difference in altitude of 5 meters will have a kme of 0.85. In this context, the underlying assumption is that an individual assesses accessibility solely on the basis of uphill distance. Hence, the accessibility indicator can be revised as follows :

$$A_i = \sum N_j \exp^{-\lambda(\frac{d_{i,j}}{1'000} + \frac{\Delta Z^{+}_{i,j}}{100})}. \tag{15}$$

4.3 Descriptive statistics

Subsequently, the variables used in the analysis are described in Table 1. Table 2 summarizes the descriptive statistics of the data. For the Western Switzerland, the "average property" is an apartment of $83.5\ m^2$, with 3.5 rooms and located in a district where the average age of the buildings is 54 years. The road noise index is 51.5 db and the train noise indicator is 0.74, indicating that most of the dwellings are far from the train stations with few train-related nuisances. Within a 10 km radius, the average population is $128'500$ and the employment opportunities are $64'000$. The 10 km radius was chosen because, all other variables being the same, it is the one that maximizes R^2. Intuitively, one could imagine the average property to be located on the periphery of a large city.

Finally, Vaud is the most represented canton with 36%, while Jura, with 4% of the observations, is the least represented. Taking into consideration the population of each canton, the distribution by cantons appears to be representative of the Western Switzerland area, with the exception of the canton of Geneva that is under-represented and would have been expected to account for around 20% of the observations.

[Insert Table 1]

[Insert Table 2]

5 APPLICATION

This section applies the methodology described in Section 3 to the dataset presented in Section 4. The proposed model, denoted by MHPM, is compared with the standard hedonic price model (SHPM) that is used as a benchmark. All models are estimated based on the same specification of the hedonic price equation, using on the variables presented in Table 1. The MHPM can accommodate a varying number of classes depending on the needs of the specific application. We applied the methodology on different numbers of classes and chose a 3-class model as it is the one that performed best based on the BIC criterion.

We begin by presenting a comparison of the results between the SHPM and MHPM, followed by an analysis of the classes of the MHPM.

5.1 SHPM vs MHPM

Table 3 presents the results of the SHPM and the MHPM. We will first study and discuss the results of the SHPM and then, interpret the potential differences of the MHPM.

SHPM : We start with the standard hedonic pricing model that is used as a reference point to compare the results of the MHPM. All the signs are in line with our expectations — bearing in mind that the dependent variable is the log of the price over floorspace. The variables that increase the price over floorspace are the number of rooms, the population, the employment number and the accessibility to public transport. Whereas the ones that lower it are the floorspace, the age of the buildings, the road noise index and the train noise indicator.

The supply and demand relationship is the key element when defining prices. Real estate is no exception to this rule and has the particularity of being dependent on internal and external factors. We can separate the variables affecting the price into 3 categories: those internal to the property, those external to the property on a small scale (e.g. street, neighbourhood) and those on a larger scale (e.g. city, canton, region).

Concerning the internal categories, generally, the price over floorspace decreases when the total surface increases. For the same floorspace, dwellings with more (bed)rooms are in high demand on the real estate market, whether by families or students in shared flats close to universities. Finally, a new property, or one located in a newer neighborhood, is usually rented for more than the same one in an older neighborhood.

For external factors in the neighbourhood, quiet dwellings and high accessibility are more sought-after. Then, when road noise increases price decreases and when accessibility increases, when accessibility increases, the attractiveness increases and so does the price. Finally, based on its definition, the negative sign of the train noise indicator indicates that, when the relation noise / distance to train station increases, price increases. This means that, overall, being close to a train station is not an attractive feature.

Large-scale external factors are the simplest to interpret. An increase in population implies an increase in demand, while an increase in employment implies an increase in attractiveness. In both cases, the consequence is an increase in the prices of the region.

The Cantons deserve to be treated separately. With Fribourg as a reference, the positive signs for the cantons of Vaud and Geneva indicate that, for an identical good, the price is higher in these cantons. Moreover, if we look at the values, we see that, for the same features, an apartment in Geneva has a higher price over floorspace than in Vaud. Similarly, negative coefficients indicate that the same dwelling will have a lower price over floorspace than in the reference canton. These interpretations are representative of the current real estate market in the study area.

MHPM : We now analyze the MHPM and compare the results with the

SHPM. It is worth noting that the signs of coefficients are, in most cases, aligned.

For class 1, the only significant sign that is different is the train noise indicator. According to its definition (Section 4.2), this indicator takes into account the noise of the station as well as the distance between the station and the apartment. Therefore, a positive sign indicates that the attractiveness of being close to the station compensates for the inconvenience of noise. This attractiveness also translates into an increase in the coefficient linked to the accessibility variable.

For class 2, the only significant sign that is different compared with the SHPM is the one of the employment. The negative sign indicates that, within a close radius, employment is not a factor that increases the price over floorspace as, for the people in these dwellings, it is not a problem to work further away. This can be explained by the higher mobility of the population in Switzerland, where people travel an average of 30 km to go to work [14].

The only significant result of class 3 that differs from the SHPM is the population. This negative sign can be explained by the high number of new dwellings built in some agglomerations in recent years. For these categories of housing the supply may be higher than the demand, which pushes the prices over flooorspace of those annoucments downwards.

[Insert Table 3]

Model performance : The individual class performances of the MHPM as well as the comparison with the SHPM are summarized in Table 4. First, we notice that class 3 of MHPM has, as the same time, a R^2 much lower than the other classes and a high $RMSE$. Taking into account the small size of this class and its characteristics, this result indicates that class 3 gathers the outliers of the dataset. Classes 1 and 2 perform better than the SHPM, concerning both the R^2 and the $RMSE$, suggesting that for the concerned sub-populations these classes better describe the prices in comparison with the SHPM. The overall R^2 and $RMSE$ of the MHPM model are better than the SHPM. These results indicate that, overall, the MHPM estimates real estate prices better than the SHPM and confirms the value of the latent class approach.

[14]https://www.bfs.admin.ch/bfs/fr/home/statistiques/mobilite-transports/transport-personnes/pendularite.html

[Insert Table 4]

The F-test statistic comparing ratio of variances of residuals (H0= equal variances vs. H1= variance of MHPM < variance of SHPM) is equal to 0.58 and leads to rejection of the H0 assumption (p-value $< 2.2e-16$). The variances of the residuals of the models are statistically different from each other and the MHPM one is strictly lower than the SHPM one. Figure 3 presents the scatter plots of the logarithm of the real price (X-axis) versus the logarithm of the estimated prices (Y-axis), either for the SHPM or the MHPM. For the SHPM, this corresponds to the value estimated using the coefficients obtained by the model. Finally, for the MHPM, this corresponds to the estimated values for each of the classes of the model, individually weighted by the posterior class membership probabilities (Equation (6)). These two diagrams support the results presented in the previous sections. In line with Belasco et al. [2012], the MHPM is able to better capture heterogeneity in data.

[Insert Figure 3]

5.2 MHPM - Study of the classes

In this section, we use the model posterior probabilities, described in Equation (6), to assign each observation to one class, in order to be able to characterize them. We analyze the distribution of the observations allocated in the classes of the MHPM as well as the statistics on intrinsic variables (Table 5), completed by the distribution of the prices over floorspace per class compared to the overall distribution (Figure 4). Note that we are referring to the real observed values of the goods and not the estimated values.

Class 1 is the biggest class, with a weight of 49$\%$ of the dataset. The size of this class and the location of the dwellings associated with it in the center of the price distribution, indicate that it represents what we would term as "standard" and most common apartments. According to the model, on average, a "standard" property in Western Switzerland has a price over floorspace of 21.56 CHF/m^2, for a floorspace of 81.37 m^2, and a mean number of rooms of 3.38.

With a weight of 42% and prices also located in the center of the price distribution, Class 2 has an average price over floorspace of 22.95 CHF/m^2, for a mean floorspace of 85.38 m^2, and a mean number of rooms of 3.47. On average, Class 2 is characterized by larger and more expensive dwellings that could be preferred by higher income households.

Finally, class 3 with a proportion of 9% completes the first two classes and is located throughout the entire distribution. These observations have on average higher prices ovre floorspace (23.36CHF/m^2) and higher average floorspace (86.48 m^2) and an average number of rooms similar to the other classes (3.38). The weight of this class, the mean values of these intrinsic variables as well as their high standard deviations suggests that it is a collection of outliers.

The logarithms of the mean prices over floorspace presented in Table 5 are finally quite close. In order to verify that the means and the standard deviations of the MHPM classes are statistically different, we perform a Welch's t-test and a F-test of variance comparison. For classes 1 and 2, the statistical value of the t-test is equal to 6.46 with a negligible p-value (e.g. < 0.001). These results indicate that the mean price over floorspace of class 1 is significantly different from that of class 2 at the 95% level of confidence. With a F-test equal to 2.22 and a negligible p-value the variances of those two classes are also significantly different at the 95% level of confidence. With a t-test of 0.73 (and a p-value of 0.46), the means of classes 2 and 3 are not significantly different at the 95% confidence level. But the variances with an F-test of 1.76 (with a negligible p-value) are considered to be different. Finally for class 1 and 3, the means (t-test of 3.36 and a negligible p-value) and the variances (F-test 3.90 and a negligible p-value) are condidered significantly different at the 95% level of confidence.

Finally, the study of the location of the MHPM classes presented in Appendix A confirms that the observations are, for each class, well distributed in space and not geographically attached to a location.

[Insert Table 5]

[Insert Figure 4]

6 CONCLUSION

This paper applies discrete mixture of market hedonic-pricing model to a real estate market case study in order to investigate the presence of unobserved heterogeneity in the preferences for real estate goods and compare the proposed approach that is estimated using a SEM algorithm to the standard hedonic pricing model. The accuracy of statistical models, including the latent classes, is evaluated based on the RMSE criterion.

The data is extracted from internet announcements of appartments in Western Switzerland and completed by combining several open data sources and variables created from the raw data. The first constructed variable concerns the road noise and transforms the average day and night noise by weighting their durations and increasing the disturbance caused by night noise. The second concerns the train noise and proposes an indicator combining noise and distance to the nearest station in order to assess the attractiveness of train stations. Finally, the last one concerns the accessibility to public transport, integrating both the distance and the difference in altitude from the dwelling to public transport stops, using a generalized cost, and more specifically the notion of "effort kilometer" used for mountain trails. Despite the effort to augment the data, there are still many missing characteristics. This may limit to a certain extent the explanatory power of the study, but not its methodological interest.

The results confirm our hypothesis that latent classes exist in the price definition of dwellings. Different characteristics may be relevant to the price of dwellings in different classes. The results demonstrate that the SEM algorithm with 3 classes is overall, for this case study , more accurate than the standard hedonic pricing model. The largest class accounts for 49% of the observations and is assumed to represent standard dwellings. The second class accounts for 42% of the observations and is assumed to be preferred by higher income households, while the last class with 9% of the observations captures the outliers. Finally, each latent class behaves independently with different estimated coefficients. The signs of the coefficients are in most cases the same as those of the hedonic pricing model, and when this is not the case, it helps to understand the price behavior of the latent classes.

One remark is that the data used here has been extracted from rental announcements. Even though they are useful for identifying trends, they do not represent the real estate market in Western Switzerland per se. Future work will focus in applying this methodology to real transactions data and to study the evolution of the market over time. Finally, we are interested in developing a methodology that is capable of predicting the price of new goods by estimating its probability of belonging to a class, using supervised clustering.

REFERENCE

BARANZINI A., SCHAERER C., RAMIREZ J.V. and THALMANN P. [2008], « Do Foreigners Pay Higher Rents for the Same Quality of Housing in Geneva and Zurich?», *Swiss J Economics Statistics* 144, 703–730.

BARANZINI A., SCHAERER C. and THALMANN P. [2010], **«** Using measured instead of perceived noise in hedonic models **»,** *Transportation Research Part D: Transport and Environment*, 15[8]:473 – 482.

BELASCO E., FARMER M. and LIPSCOMB C. [2012], **«** Using a finite mixture model of heterogeneous households to delineate housing submarkets », *Journal of Real Estate Research*, 34.

BISHOP K. C. and TIMMINS C. [2011], « Hedonic prices and implicit markets: Estimating marginal willingness to pay for differentiated products without instrumental variables », *Working Paper 17611, National Bureau of Economic Research.*

BOUCQ E. and PAPON F. [2008], « Assessment of the real estate benefits due to accessibility gains brought by a transport project: The impacts of a light rail infrastructure improvement in the hauts-de-seine department », *European Transport Trasporti Europei*, 40:51–68.

CAVAILHÈS J. [2005], « Le prix des attributs du logement », *Economie et statistique*, n°381-382.

CAVAILHÈS J., BROSSARD T., FOLTÊTE J.-C., HILAL M., JOLY D., TOURNEUX F., TRITZ C. and WAVRESKY P. [2009], « Gis-based hedonic pricing of landscape », *Environmental and Resource Economics*, 44:571–590.

CELEUX G., CHAUVEAU D. and DIEBOLT J. [1995], « On Stochastic Versions of the EM Algorithm », *Research Report RR-2514, INRIA.*

CELEUX G. and DIEBOLT J. [1985], « The sem algorithm : a probabilistic teacher algorithm derived from the em algorithm for the mixture », C*omputational Statistics Quarterly*, 2, 73-82.

COURT A. [1939], « Hedonic price indexes with automotive examples. The dynamics of automobile demand », *The Dynamics of Automobile Demand*, pp. 99–117. General Motors Corporation, New York.

DEMPSTER A. P., LAIRD N. M. and RUBIN, D. B. [1977], « Maximum likelihood from incomplete data via the EM algorithm », *Journal of the Royal Statistical Society: Series B*, 39:1–38.

FRÜHWIRTH-SCHNATTER S. [2006], « Finite mixture and Markov switching models », *Springer Series in Statistics*, Springer-Verlag New York

GRILICHES Z. [1961], « Hedonic Price Indexes for Automobiles: An Econometric of Quality Change ». *The Price Statistics of the Federal Goverment, NBER Chapters*, pages 173–196. National Bureau of Economic Research, Inc.

GROSS D. J., SIRMANS C. F. and BENJAMIN J. D. [1990], « An empirical evaluation of the probabilistic bid-rent model : The case of homogenous households », *Regional Science and Urban Economics*, 20[1]:103–110.

GRUN B. and LEISCH F. [2008], « Flexmix version 2: Finite mixtures with concomitant variables and varying and constant parameters », *Journal of Statistical Software, Articles*, 28[4]:1–35.

HANSEN W. G. [1959], « How accessibility shapes land use », *Journal of the American Institute of Planners*, 25[2]:73–76.

HILL R. [2013], « Hedonic price indexes for residential housing: A survey, evaluation and taxonomy », *Journal of Economic Surveys*, 27.

HOESLI M., GIACCOTTO C. and FAVARGER P. [1997], « Three new real estate price indices for Geneva, Switzerland », *The Journal of Real Estate Finance and Economics*, 15[1]:93–109.

HOWELL J. F. and PERISTIANI S. [1987], « The estimation of a hedonic asking and offer rent equation model: An em algorithm approach », *Empirical Economics*, 12[4]:203–220.

LAFERRERE A. [2005], « Hedonic housing price indexes: the French experience », *International Settlements, B*., Real estate indicators and financial stability, volume 21 of BIS Papers chapters, pages 271–287. Bank for International Settlements.

LAZARSFELD P. F. [1950], « The logical and mathematical foundation of latent structure analysis », *Studies in Social Psychology in World War II Vol. IV : Measurement and Prediction*, pages 362–412.

LÖCHL M. and AXHAUSEN K. W. [2010], « Modelling hedonic residential rents for land use and transport simulation while considering spatial effects », *Journal of Transport and Land Use*, 3[2].

MAYER M., BOURASSA S., HOESLI M., and SCOGNAMIGLIO D. [2018], « Estimation and updating methods for hedonic valuation », *SSRN Electronic Journal*.

OLADUNNI T. and SHARMA S. [2016], « Hedonic housing theory : A machine learning investigation », *15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Anaheim, CA, USA, 2016, pp. 522-527.

PALMQUIST R. B. [2006], « Property value models », in K. G. MÄLER and J. R. Vincent, editors, *Handbook of Environmental Economics, volume 2, chapter 16*, pages 763–819. Elsevier, 1 edition.

PENG L. [2010], « Repeat sales regression on heterogeneous properties », *Journal of Real Estate Finance and Economics*, 45:1–24.

ROSEN S. [1974], « Hedonic prices and implicit markets: Product differentiation in pure competition », *Journal of Political Economy*, 82[1]:34–55.

RUF D. [2017], « Agglomeration Effects and Liquidity Gradients in Local Rental Housing Markets », *Working Papers on Finance 1702*, University of St. Gallen, School of Finance.

SCHULTZ T. J. [1978], « Synthesis of social surveys on noise annoyance », *The Journal of the Acoustical Society of America*, 64 2:377–405.

VICKERMAN R. W. [1974], « Accessibility, attraction, and potential: A review of some concepts and their use in determining mobility », *Environment and Planning A: Economy and Space*, 6[6]:675–691.

WAUGH FREDERICK V. [1929], « Quality as a Determinant of Vegetable Prices: a Statistical Study of Quality Factors Influencing Vegetable Prices in the Boston Wholesale Market », New York Chichester, *West Sussex: Columbia University Press*.

WEZEL M., KAGIE M. and POTHARST R. [2005], « Boosting the accuracy of hedonic pricing models », *Erasmus University Rotterdam, Econometric Institute, Econometric Institute Report*.

A MHPM : LOCATION OF REAL ESTATE PROPERTY

Figure 5 to 7 illustrate the spatial distribution of classes 1, 2 & 3. We can see that the observations are well distributed in space and not geographically attached to a location and that prices over floorspace are fairly well distributed and go from less than 16 to more than 28 for all classes. Then, there is not one class that gather observations around one location (canton or city) or absorbs all the rents with the highest (or lowest) prices over floorspace.

We also note that dwellings in class 2, high prices are particularly concentrated around the Lake Geneva area (Figure 6). This is not the case for class 1 (Figure 5), and in particular for dwellings located in the Canton of Geneva, whose prices over floor space appear to be lower than in the other Cantons. The prices of goods located in the countryside seem to be well distributed for each of these two classes. Finnaly, dwellings of class 3 (Figure 7), identify as outliers, are, despite the small number of observations, well spatially distributed.

[Insert Figure 5]

[Insert Figure 6]

[Insert Figure 7]

B ELASTICITIES

Residuals from the models presented in Table 3 are likely correlated with price. Insofar as we have not dealt with the endogeneity bias of the models, elasticities interpretation of floorspace and number of rooms are presented in appendix, knowing this issue. As the variable explained is the log of the price over floorspace and explanatory variables either linear or logarithmic its implies two claculating methods of elasticity. Elasticity of a Log-Linear specification is expressed as :

$$\epsilon_{\text{Log-Lin}} = \beta_{\text{Log-Lin}} \cdot X \quad (16)$$

This implies that an increase of one unit of x increases y by $\beta_{\text{Log-Lin}} \cdot 100\%$. While the elasticity of a Log-Log specification is expressed as :

$$\epsilon_{\text{Log-Log}} = \beta_{\text{Log-Log}} \quad (17)$$

This implies that a 1% increase in x increases y by $\beta_{\text{Log-Log}} \cdot 1\%$.

Table 6 presents the elasticities of the dependent variables with respect to floorspace and number of rooms.

[Insert Table 6]

Putting aside the outlier class, we notice that the elasticities of the number of rooms are almost similar. The differences are more pronounced for the floorspace : a modification in the MHPM class 1 implies a greater change in price over floorspace compared to the MHPM class 2 or the SHPM. Therefore, for a dwellings of the MHPM class 1, a $10m^2$ increase in floorspace, while keeping the same number of rooms, results in an average drop in the price over floorspace of $2.38\ CHF/m^2$. Still for MHPM class 1, the creation of an additionnal room, without changing the floorspace, increases the price over floorspace by an average of $1.554\ CHF/m^2$. Finally, both the creation of an additionnal room and a $10m^2$ increase in floorspace results in an average drop of $0.826\ CHF/m^2$.

The same reasoning applied to other classes will give different results. Note that for normalized variables (e.g., population, employment and age of buildings) or variables transforming raw data (e.g., road noise index, train noise indicator and accessibility indicator) elasticities need to be calculated with caution.

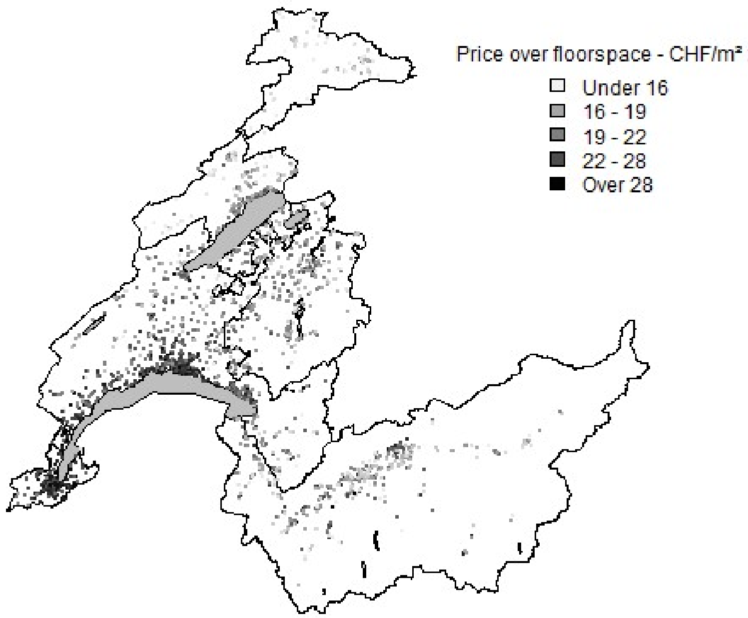

Figure 1: Location and price over floorspace of the collected real estate properties.

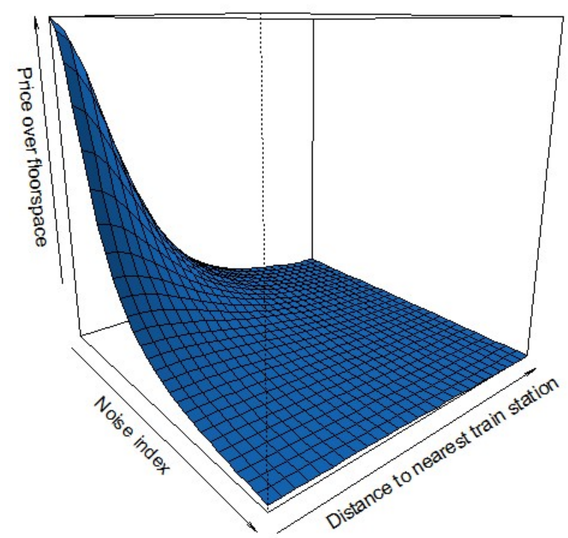

Figure 2: Representation of the train noise indicator.

Table 1: Variable description

Variable	**Description**
Floorspace	Log of the surface, in m^2
Number of rooms	Number of rooms in the real estate property
Population number	Log of the population in a given radius (10 km) - normalized value
Employment number	Log of number of workplaces (full-time equivalent) in a given radius (10 km) - normalized value
Age of buildings	Log of the average age of buildings within a hectar - normalized value
Road noise index	Log day-night road noise index
Train noise indicator	Defined using both the Log day-night noise index and the distance to the nearest train station
Accessibility indicator	Defined using both the distance and the difference in altitude from the dwelling to public transport stops
Canton	Canton where the object is located:VD, VS, GE, JU,FR (reference : FR)

Table 2: Descriptive statistics

Variable	Mean	Standard deviation	Median	Minimum	Maximum
Floorspace	83.51	34.17	80	5	382
Number of rooms	3.42	1.2	3.5	1	9
Population	128453.97	124302.11	80373	1282	480627
Employment number	64303.74	77441.41	33287.64	258.41	282579.53
Age of buildings	54.11	26.32	53	3	100
Road noise index	51.57	8.79	50.91	19.08	81.77
Train noise indicator	0.74	0.17	0.67	0.4	1
Accessibility indicator	15.54	12.69	12.25	0.04	63.23
Canton GE	0.08	0.27	0	0	1
Canton JU	0.04	0.19	0	0	1
Canton NE	0.13	0.34	0	0	1
Canton VD	0.36	0.48	0	0	1
Canton VS	0.20	0.4	0	0	1
Canton FR	0.19	0.39	0	0	1

Table 3: Results of the SHPM and the MHPM - 3 classes

Variable	**SHPM**			**MHPM**								
				Class 1			Class 2			Class 3		
	Est.	t-stat		Est.	t-stat		Est.	t-stat		Est.	t-stat	
Intercept	4.537	57.790	***	4.588	69.964	***	4.486	56.257	***	4.844	10.129	***
Log Floorspace	-0.431	-34.467	***	-0.498	-46.215	***	-0.487	-39.609	***	-0.250	-3.617	***
Number of rooms	0.049	10.433	***	0.047	11.744	***	0.059	13.050	***	0.064	2.384	*
Log Population	0.543	4.398	***	0.377	3.708	***	1.889	14.065	***	-1.506	-2.125	*
Log Employment number	0.112	0.934		0.037	0.381		-1.124	-8.734	***	2.011	2.814	**
Log age of buildings	-0.066	-5.015	***	-0.056	-5.582	***	-0.060	-4.435	***	-0.186	-1.807	.
Log road noise index	-0.041	-2.458	*	0.024	1.776	.	0.005	0.276		-0.235	-2.149	*
Train noise indicator	-0.116	-6.398	***	0.063	4.190	***	-0.142	-7.855	***	-0.509	-4.033	***
Accessibility indicator	0.001	3.852	***	0.004	13.392	***	0.000	0.760		-0.001	-0.416	
Canton (ref. is fribourg)												
Geneva	0.253	16.370	***	0.105	8.341	***	0.447	28.562	***	0.381	3.794	***
Jura	-0.179	-10.103	***	-0.160	-11.369	***	-0.140	-7.896	***	-0.278	-1.858	.
Neuchatel	-0.148	-13.456	***	-0.033	-3.747	***	-0.188	-16.759	***	-0.158	-1.846	.
Vaud	0.138	17.222	***	0.021	3.276	**	0.267	33.302	***	0.305	4.957	***
Valais	-0.082	-9.078	***	-0.048	-6.685	***	-0.103	-11.714	***	-0.112	-1.572	

Levels of significance : *** 0.1%; ** 1%; * 5%; . 10%

Table 4: Performance of the SHPM and the MHPM - 3 classes

	SHPM	**MHPM**		
		Class 1	Class 2	Class 3
R^2	0.607	0.793	0.865	0.288
			0.722	
RMSE	0.205	0.116	0.130	0.422
			0.172	

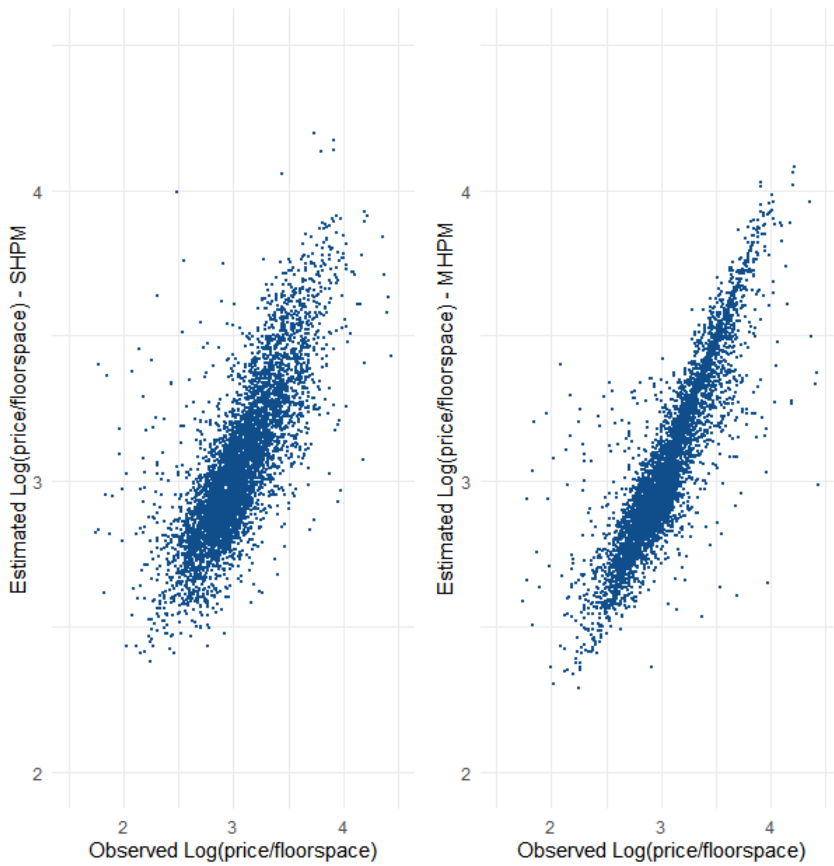

Figure 3: Comparision of models : log(price/floorspace) vs. $\mathbb{E}$(log(price/floorspace))

Table 5: Statistics of the SHPM and the MHPM - 3 classes

	SHPM	**MHPM**		
		Class 1	Class 2	Class 3
Number of observations	5682	2789	2380	513
Weights	1	0.49	0.42	0.09
Mean Price over floorspace	22.30	21.56	22.95	23.36
Sd Price over floorspace	8.02	6.00	8.94	11.86
Mean Floorspace	83.51	81.37	85.38	86.48
Sd Floorspace	34.17	30.48	35.31	45.33
Mean Number of rooms	3.42	3.38	3.47	3.38
Sd Number of rooms	1.20	1.13	1.23	1.41

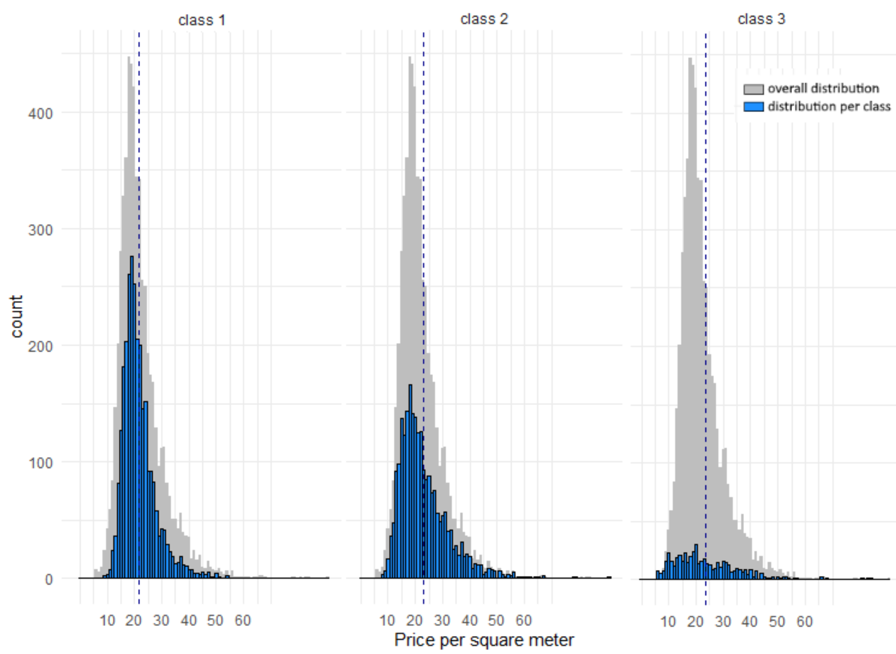

Figure 4: MHPM - Distribution of the real prices over floorspace per class compared to the overall distribution.

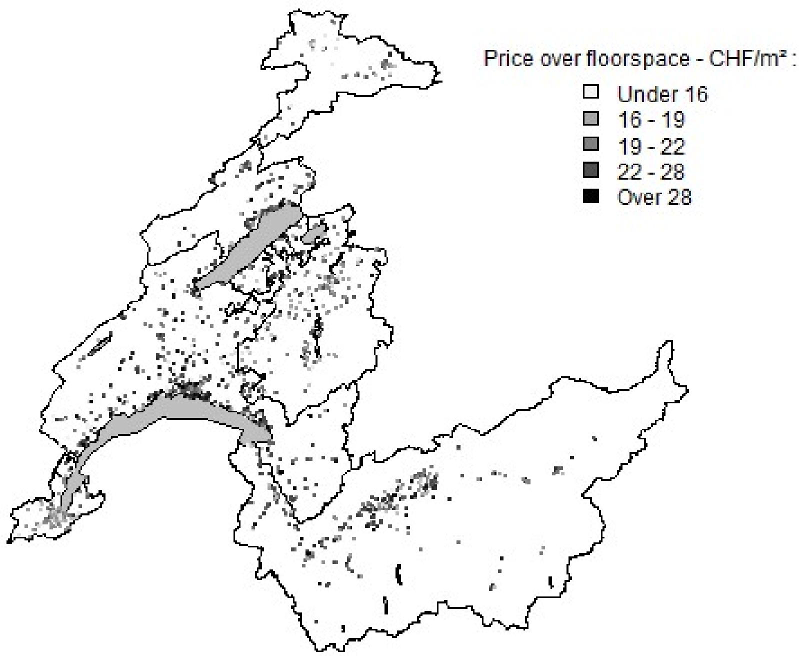

Figure 5: MHPM - Location and price of the real estate property for class 1.

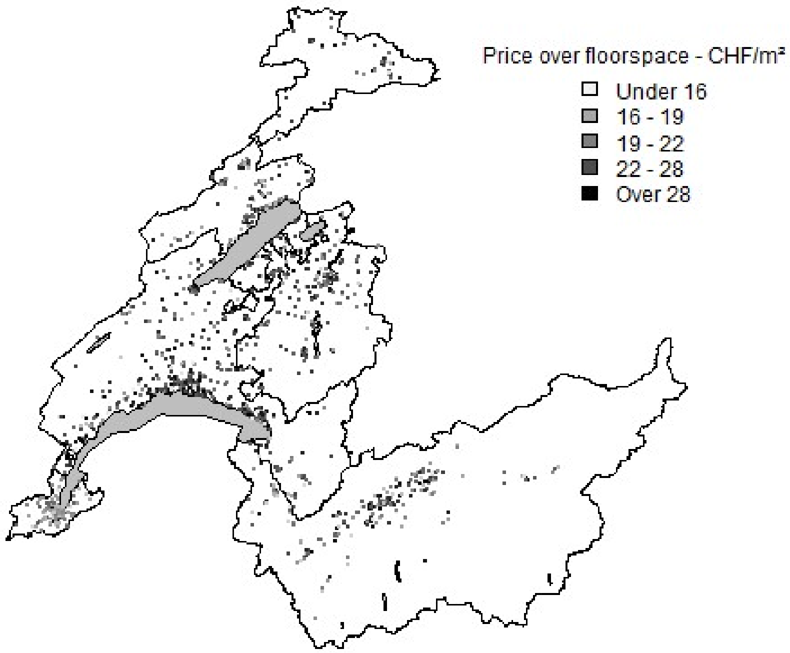

Figure 6: MHPM - Location and price of the real estate property for class 2.

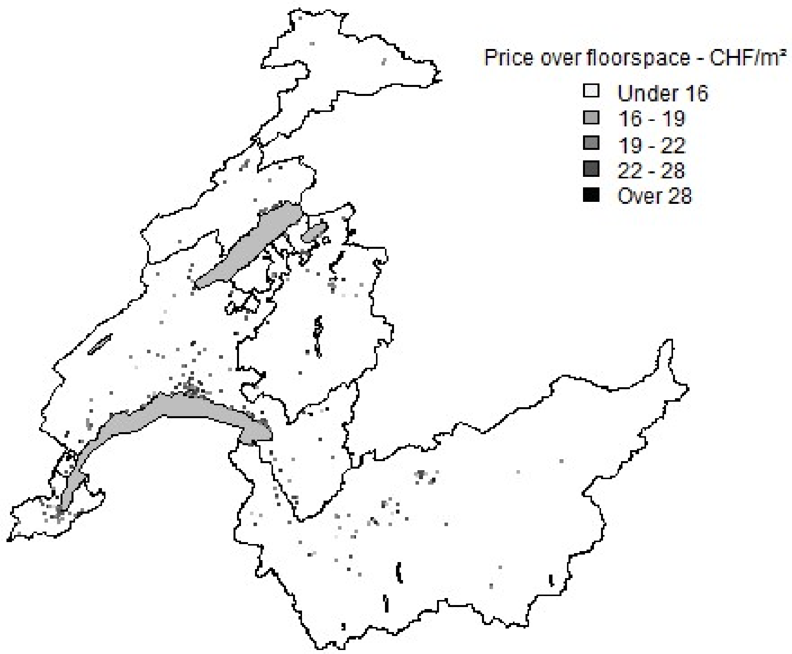

Figure 7: MHPM - Location and price of the real property for class 3.

Table 6: Elasticities with respect to floorspace and number of rooms

Variable	**SHPM**	**MHPM**		
		Class 1	Class 2	Class 3
Floorspace (Log)	-0.155	-0.174	-0.178	-0.095
Number of rooms (Lin)	1.085	1.002	1.365	1.506