The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1753-8270.htm

Insight into apartment attributes and location with factors and principal components

Alain Bonnafous and Marko Kryvobokov
Laboratory of Transport Economics, Lyon, France

Received 1 October 2010
Revised 23 January 2011
Accepted 3 February 2011

Abstract

Purpose – The purpose of this paper is to better understand the spatial structure of the Lyon urban area focusing on real estate. For this, two aims are formulated. The first aim is to identify and geographically analyse latent structure underlying apartment variables and location. The second aim is to decrease a number of explanatory variables in a hedonic model of real estate prices applying latent constructs.

Design/methodology/approach – For the first aim of a parsimonious representation among measured variables, exploratory factor analysis is applied. For the second aim of data reduction, principal component analysis (PCA) is used. The exploited regression methodologies are global and geographically weighted ordinary least squares.

Findings – Four factors are extracted, of which two represent apartment attributes and other two – location attributes. Principal components provide better insight into location attributes dividing the service employment centres into two geographical groups. The inclusion of principal components in hedonic price equation instead of initial location variables decreases goodness of fit, but does not gradually change non-location estimates and other parameters.

Originality/value – Differently from previous applications of factor analysis and PCA in the real estate domain, oblique rotation is applied, which allows the extracted factors or components to be correlated. The scores of factors and components are interpolated from points to raster maps creating a continuous geographical distribution. Hedonic models with and without principal components are compared in detail.

Keywords France, Urban areas, Real estate, Prices

Paper type Research paper

1. Introduction

A complex social nature of real estate price is a well-known phenomenon. In the academic world, the most popular way of its analysis is a hedonic regression modelling, where, in the cross-sectional version, the dependent variable is usually a price and the explanatory variables include real estate attributes and location attributes. The estimated parameters are interpreted as willingness to pay for different attributes (Rosen, 1974). This way was followed in the previous study of the Lyon urban area, where a hedonic model of apartment prices was created (Kryvobokov, 2010).

The study is part of project PLAINSUDD sponsored though French ANR (number ANR-08-VD-00). Provision of data on apartment prices and attributes by Perval and Pierre-Yves Péguy is acknowledged. The authors are grateful to Cristina Pronello for fruitful suggestions on factor analysis and to Nicolas Ovtracht and Valérie Thiebaut for calculation of travel times and the coordinates of apartments. The authors thank Chris Brunsdon, Stewart Fotheringham and Martin Charlton for the distribution of the GWR software. The paper benefited from the suggestions of two anonymous reviewers.

International Journal of Housing Markets and Analysis
Vol. 4 No. 2, 2011
pp. 155-171

1753-8270
DOI 10.1108/17538271111137930

Our current study is motivated by the necessity to better understand the spatial structure of the Lyon urban area for the practical needs of urban planning and real estate valuation. We can apply an analysis, which does not necessarily imply focusing on price as dependent variable. Such an analysis provides a better understanding of data themselves with insight into the hidden relationships between variables. The methods of this group include clustering, factor analysis, principal component analysis (PCA), artificial neuron networks and others. To higher or lesser degree, the results of these methods are related to pattern recognition and can be applied for identification of neighbourhoods or submarkets.

The first aim of our study is to identify and geographically analyse latent structure underlying apartment attributes and location. With this sort of analysis, we can also revert to a hedonic price modelling, where a numerous number of initial explanatory variables can be decreased (Des Rosiers *et al.*, 2000). This practical issue is the second aim of the paper.

Methodologically, the two aims can be achieved with two similar, but not identical techniques: while for the extraction of latent constructs, it is better to apply exploratory factor analysis (EFA), for data reduction PCA is appropriate (Fabrigar *et al.*, 1999). In both methods, we apply an oblique rotation, which permits correlation among factors or principal components. The exploited regression techniques are global and geographically weighted ordinary least squares (OLS).

The paper is structured as follows. The subsequent section is methodological. Section 3 describes the process of data preparation for factor analysis. Section 4 deals with the EFA itself and includes the interpretation and geographical demonstration of factor scores. Section 5 is about the application of PCA to location attributes and their inclusion into a regression model. The final section concludes.

2. Methodology

A relatively often used technique is a combination of factor analysis or PCA and cluster analysis. The extracted factors or principal components are used as data for clustering to determine submarkets and include them in a hedonic price equation. For this purpose, Dale-Johnson (1982) applied Q-factor analysis, whereas Maclennan and Tu (1996), Bourassa *et al.* (1999, 2003) exploited PCA. For example, Bourassa *et al.* (2003) found that the best results were obtained when cluster analysis was based on the two most important components.

The other application of PCA in hedonic modelling of real estate prices was proposed by Des Rosiers *et al.* (2000). The mentioned study as well as Des Rosiers and Thériault (2008) use PCA in the Quebec urban community for data reduction. In particular, to avoid severe multicollinearity in hedonic price model induced by 15 accessibility attributes of travel times and walking times to different objects, two principal components were obtained. Then these components were used in a regression model as substitutes for initial variables. The authors made a quite straightforward interpretation: the first component accounts for accessibility to regional services, while the second one refers to local accessibility. In the former study, it was also obtained four principal components on census attributes. After mapping of the principal components, Des Rosiers *et al.* (2000) conclude that PCA provides useful insights into housing market dynamics: it clearly highlights the marked concentration of low-income households dwelling as opposed to high-income households and also proves consistent with urban reality.

One more example of this kind of application of PCA is Öven and Pekdemir (2006). Among their five principal components, the first accounts for quality of location and the third for physical characteristics of building. In their regression model of office rent, Öven and Pekdemir (2006) incorporate a combination of specified principal components, dummy variables, and the variables found unrelated to any of the specified principal components.

With the first aim to identify latent constructs underlying our variables, in this study we apply the methodology of EFA. According to Fabrigar *et al.* (1999), it is an appropriate form of analysis "if the goal is to arrive at a parsimonious representation among measured variables". When the goal is data reduction, which is our second aim, PCA can be applied (Bonnafous, 1973; Fabrigar *et al.*, 1999). Though both methods represent the observed variables as linear combinations of factors or components and are closely related, they are not identical. PCA takes into account all variability in the variables, while factor analysis explains the variability, which exists due to common factors ("communality", which in this case can be much less than unity).

Applying EFA and PCA, the following peculiarities are important. The analysis is applicable to continuous data; therefore, dummy variables should be avoided (Kolenikov and Angeles, 2004). Normality is checked with skew and kurtosis taking into account the thresholds of 2 and 7, respectively (West *et al.*, 1995). Principal axes factoring is applied as the most widely used method in factor analysis (Warner, 2007). We select the number of factors and principal components using the criterion that the eigenvalues of the unreduced correlation matrix should be higher than one.

The rotation method usually exploited in PCA applications in the real estate domain (e.g. by Bourassa and colleagues or Des Rosiers and colleagues) is a varimax rotation, which involves an orthogonal transformation of variables into a new set of mutually independent components. As Fabrigar *et al.* (1999) noted, the methodological literature suggests little justification for using orthogonal rotation; it can be reasonable only if the oblique solution indicates that the factors are uncorrelated. In the current study, we apply an oblique rotation, which permits correlation among factors. We use the standard method of non-orthogonal rotation – direct oblimin and check correlation between factors or principal components.

To create a continuous representation of the geographical distribution of factors and principal components, we interpolate their scores to a raster. For this, a widely applied inverse distance weighted method is used.

The variables in our hedonic price model with initial attributes are selected on the base of the previous regression model of apartment prices in the Lyon urban area (Kryvobokov, 2010). The same data set is used in both studies, but the number of observations in regression equations is a bit different, mainly because in the current study, the data are prepared to the needs of PCA.

The hedonic price model with principal components is compared to the specification without them. In the comparison, we focus on regression performance, multicollinearity and spatial autocorrelation of residuals. In case when data are spatially dependent, which is practically always the case with real estate data, OLS estimates are inefficient and inconsistent (Dubin, 1998), while the estimate of the variance is biased (Anselin, 1988). To account for these issues, as an alternative to a global OLS model we use a geographically weighted regression (GWR) (Brunsdon *et al.*, 1996). The applied GWR model is OLS with the Gaussian error term, fixed kernel type and kernel bandwith determined by the Akaike information criterion; a separate equation is solved for each observation, and the overall results are compared with the global model.

3. Data preparation

Geographically, the area of study includes the cities of Lyon and Villeurbanne; the latter city is located to the north-east from Lyon (Figure 1). These adjacent cities with overall population of over 600,000 inhabitants have a common planning structure and transportation network including metro, tramway and trolleybus links and make up the core of the Lyon urban area, which is the second largest agglomeration by population in France.

The data on sale prices and apartment attributes were provided by Perval, which collects information about real estate transactions in France. Data on approximately 10,000 apartment sales selected randomly from all sales in the central part of the Lyon urban area in the period of 1997-2008 were obtained. With very few exceptions, the apartments are located in the urbanised area and mainly concentrated in Lyon and Villeurbanne.

We deleted observations with missing data and with prices lower than €20,000 and higher than €500,000 and with area of less than 20 square metres and more than 200 square metres. The 4,254 remained observations are used in the analysis. Exclusion of more observations with missing data could significantly lower sample size and the statistical power of results, while attributing mean scores for missing values reduces variation among observations and increases the potential for clumping and truncation (Vias and Kumaranayake, 2006). In our study, 25 per cent of observations have no data about the number of parking places and 60 per cent have no data about

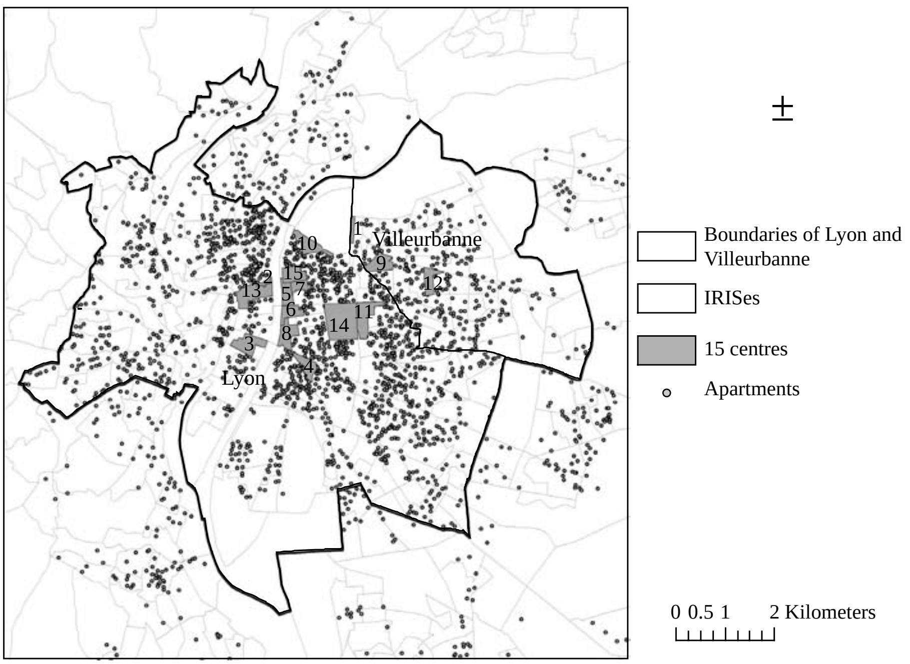

Figure 1.
Location of apartments

the quality of view. We choose to exclude these variables from factor analysis, because otherwise we would be enforced either to arbitrarily use mean scores or to considerably decrease our simple size.

Location of apartments is shown in Figure 1, where the boundaries of IRISes are shown with thin lines. les îlots regroupés pour l'information statistique (IRIS) is a French statistical unit used also as a transport analysis zone. The definitions of variables used in factor analysis and their descriptive statistics are presented in Tables I and II, of which the former includes apartment attributes and the latter describes location attributes.

As factor analysis is designed for continuous data, we treat our count variables (e.g. number of rooms) and categorical variables (e.g. apartment condition) not as dummies, but as continuous variables (Kolenikov and Angeles, 2004). Thus, years of transaction is treated as continuous variable *Year_Sale* equal to 1-12 corresponding to 1997-2008. Similarly, the variable addressing the condition of each apartment is represented as 1, 2, and 3, which correspond to bad condition (renovation is needed), medium condition (preventive maintenance is needed), and good condition, respectively, though we admit that this linear representation is rather artificial.

Building age is estimated as follows. There are seven construction periods available: before 1850; 1850-1913; 1914-1947; 1948-1969; 1970-1980; 1981-1991; and 1992 and later. We assume that a mean for the first construction period is 1800 and for the last period is 2000 and calculate the variable *Building_Age* as a difference between the year 2000 and the means of the earlier periods. For example, *Building_Age* for the second period is equal to $2000 - 1882 = 118$.

Apartments in buildings constructed before 1850, which include some observations of the eighteenth century, compose the smallest group with only 3 percent of the sample. They are concentrated in the historical arrondissements of Lyon on the banks of the Rhône and the Saône and especially between the two rivers. This old housing stock is to a considerable degree deteriorated. Each of the second and the third groups contains 7 percent of the sample. These apartments are a bit more geographically dispersed; nevertheless, most of them are located in the historical arrondissements of Lyon as well as in the east of Lyon and the south-west of Villeurbanne adjacent to Lyon.

The post-Second World War period was characterised by intensive residential development in all available parts of Lyon and Villeurbanne. The largest apartment group, consisted of more than one-third of observations, was constructed after 1992.

Variable	Description	Mean	Minimum	Maximum	SD	Skew	Kurtosis
Price	Transaction price (€)	122,215.96	20,000.00	500,000.00	70,212.55	1.44	2.90
Year_Sale	Count for year of transaction	6.86	1	12	2.85	−0.10	−0.87
Area	Apartment area (square metre)	69.11	20	196	25.63	0.84	1.61
Rooms	Number of rooms	3.07	1	8	1.18	0.26	−0.16
Floor	Floor	2.84	0	18	2.25	1.36	3.82
Building_Age	Building age, years	33.65	0	200	42.93	2.12	4.93
Condition	Apartment condition	2.78	1	3	0.48	−2.12	3.77
Cellars	Number of cellars	0.70	0	2	0.49	−0.45	−0.85

Table I. Definition of apartment variables and descriptive statistics

Variable	Description	Mean	Minimum	Maximum	SD	Skew	Kurtosis
%LowIncome	Percentage of low-income households	29.40	10.24	52.12	5.78	− 0.09	− 0.05
%MidIncome	Percentage of middle-income households	58.01	42.70	66.20	3.30	− 0.15	0.09
%HighIncome	Percentage of high-income households	12.58	4.34	28.77	2.92	0.51	0.68
TT_1	Travel time to Stalingrad	11.34	1.21	24.28	4.86	0.42	− 0.26
TT_2	Travel time to Louis Pradel	11.25	2.12	29.32	5.37	0.60	− 0.04
TT_3	Travel time to CBD	11.07	0.45	31.28	4.99	0.87	0.72
TT_4	Travel time to Victor Bach	9.67	0.45	28.20	4.99	0.49	0.01
TT_5	Travel time to Molière	10.48	0.45	29.27	5.29	0.67	− 0.07
TT_6	Travel time to Jussieu	10.52	0.45	30.26	5.22	0.70	− 0.05
TT_7	Travel time to Saxe-Bossuet	10.12	0.45	28.40	5.35	0.61	− 0.24
TT_8	Travel time to Mutualité-Liberté	10.13	0.45	30.14	5.14	0.75	0.25
TT_9	Travel time to Charles Hernu	11.23	0.45	26.34	5.39	0.33	− 0.67
TT_10	Travel time to Les Belges	11.05	0.45	27.42	5.36	0.47	− 0.46
TT_11	Travel time to Villette Gare	10.74	0.45	29.20	5.36	0.35	− 0.82
TT_12	Travel time to Gratte Ciel est	11.71	0.45	25.33	5.69	0.17	− 0.79
TT_13	Travel time to Terreaux-Bat d'Argent	11.04	0.45	30.18	5.23	0.78	0.29
TT_14	Travel time to Part-Dieu	10.69	0.45	29.41	5.26	0.44	− 0.73
TT_15	Travel time to Marechal Lyautey	10.45	0.45	28.38	5.32	0.62	− 0.19

Table II.
Definition of location variables and descriptive statistics

These groups of newer apartments are located outside the historical part of Lyon. Many communes outside Lyon and Villeurbanne, especially in the east, south and north, were converted into residential areas with multi-storey buildings. Historically, Western suburbs of Lyon with many villas were richer, while eastern were poorer. This tendency originally conditioned by geography (hills in the west provided better air quality and view) has been continued during the 1960s and 1970s, when newly constructed high-density residential districts to the east and south-east from Lyon and Villeurbanne were populated by poorer inhabitants with high share of immigrants. Moreover, the post-Second World War era of increased car ownership was characterised by the movement of richer population from the central part of Lyon to the suburbs, mainly in the Western direction. This process, which is similar to that in American urban areas, leaded to socioeconomic mix in the city centre of Lyon.

The gravity of the problem of social segregation in France is confirmed by recent urban riots as well as by the increasing number of French urban studies devoted to this subject. Korsu and Wenglenski (2010) expose in detail the problem of residential segregation in the Paris region describing the social reasons of long-term unemployment. Their study confirms the hypothesis that unemployment of poor residents is partially caused by poor accessibility of low-skilled jobs. Pan Ké Shon (2010) focuses on ethnic segregation in France. Mignot *et al.* (2009) who among other issues analyse the differences between the Eastern and Western suburbs of Lyon state that social segregation there is a reality. They also mention the process of "ghettoization" discussed in French context by Bresson *et al.* (2004).

Our location variables in Table II include percentages of households in three income groups and travel times to urban centres in minutes. Data about the ethnic composition of population are not available. All location attributes are calculated per IRIS. The middle-income group includes households in the middle 60 percent of the income range, and the lowest and highest 20 percent margins compose the other two groups. We also take into consideration 15 service employment centres, which were identified with residual analysis in Kryvobokov, 2010. Location of service employment centres is shown in Figure 1. Travel times for the travel by car between IRISes in the morning peak were obtained from the MOSART transportation model for the Lyon urban area. When an apartment is located in one of the centres (i.e. the same IRIS is both the origin and the destination), the travel time is equal to the minimum among all the other cases, thus null values are avoided.

The highest skew (for *Building_Age* and *Condition*) is only a bit higher than 2, whereas kurtosis for all variables is lower than 5. The two variables with the highest skew are specific and will not be changed: *Building_Age* better represents a building than, e.g. a counter for construction period, while *Condition* is an important apartment attribute though measured with a linear scale.

Many available apartment attributes are not included in factor analysis because of their strong non-normality. It refers for example to the number of bathrooms. Table III describes the dummy variables, which are not used in factor analysis, but are applied in regression modelling. These are the dummies for the number of bathrooms, parking places, quality of view and the existence of a garden and a terrace. As many observations contain no data about the number of parking places and quality of view, the specific dummy variables are created.

4. Factor analysis

It was impossible to include all the variables from Tables I and II into the analysis. In particular, *Area* and *Rooms* could not be presented simultaneously, and the former variable was chosen. The variable *Floor* demonstrated so low communality, that it was excluded. *Year_Sale* was skipped as well. Of income groups, the two marginal ones were included.

Variable	Description	Mean	SD
Bath1	Dummy for one bathroom	0.93	0.26
Bath2	Dummy for two bathrooms	0.07	0.25
Bath3	Dummy for three bathrooms	<0.01	0.04
ParkUn	Dummy for cases with no data about parking places	0.25	0.43
Park0	Dummy for no parking place	0.16	0.37
Park1	Dummy for one parking place	0.50	0.50
Park2	Dummies for two parking places	0.08	0.27
Park3	Dummies for three parking places	<0.01	0.06
ViewNo	Dummy for cases with no data about view	0.60	0.49
ViewGood	Dummy for view increasing value	0.38	0.49
ViewBad	Dummy for view decreasing value	0.02	0.13
Garden	Dummy for existence of garden	0.05	0.22
Terrace	Dummy for existence of terrace	0.09	0.29

Table III. Definition of dummy variables and descriptive statistics

Of 15 variables of travel times to service employment centres, it was possible to include eight. Among the centres included are Bellecour-Sala usually referred to as the CBD and the two other commonly recognisable centres of Louis Pradel and Part-Dieu. The communalities of the attributes presented in Table IV range from 0.98 to 0.11 with a mean of 0.72. It would be good to compare these figures with findings from similar studies, but it is problematic to find in the factor analysis literature any reported communalities referred for real estate. Instead, we can consider some empirical results of factor analysis conducted in psychological domain, where a good example is the above-mentioned study of Fabrigar *et al.* (1999), who analysed the data sets from Breckler (1984) and Crites *et al.* (1994). Our communalities are in general in line with their figures, but we should note that the number of observations in the mentioned sources is considerably less than in our case and our communalities for *Condition* and *Cellars* are rather low. Nevertheless, we will keep both variables because they, especially the former one, represent important apartment attributes.

There are four factors with the eigenvalues of the unreduced correlation matrix higher than one. Their eigenvalues are 7.50, 1.87, 1.60 and 1.22. The fifth factor has the eigenvalue of 0.83. The scree plot of eigenvalues (Figure 2) supports our choice: starting from the fifth factor, the slope becomes gentler. The factor correlation matrix (Table V) shows that factors 1 and 3 are negatively correlated with the coefficient of −0.54. Thus, the decision to apply a non-orthogonal rotation is correct.

EFA is a common factor model, where each measured variable is a linear function of one or more common factors (that influence more than one measured variables) and one unique factor (that influences only one measured variable). Factor loadings for structure matrix and pattern matrix are presented in Table IV. The first matrix represents the variance in a measured variable explained by a factor in both a unique and common contributions basis. The pattern matrix represents only unique contributions.

We will focus on loadings higher 0.30 and lower −0.30, which are in italics in Table IV. It is clearly seen that factors 1 and 3 are location factors, whereas factors

		Factors							
		Structure matrix				Pattern matrix			
Variable	Communality	1	2	3	4	1	2	3	4
Price	0.71	−0.19	*0.82*	−0.07	−0.07	−0.15	*0.82*	−0.01	−0.12
Area	0.72	<0.01	*0.84*	−0.14	−0.08	0.10	*0.86*	0.06	0.05
Building_Age	0.51	−0.23	−0.03	0.19	*0.68*	−0.11	−0.04	0.12	*0.68*
Condition	0.17	0.04	0.07	−0.04	*−0.40*	<0.01	0.09	−0.02	*−0.40*
Cellars	0.11	0.03	0.16	−0.11	0.28	0.02	0.14	−0.07	0.27
%LowIncome	0.94	*−0.52*	−0.13	*0.97*	0.04	0.01	0.04	*0.98*	0.04
%HighIncome	0.90	*0.55*	0.12	*−0.95*	−0.04	0.04	−0.04	*−0.94*	−0.03
TT_1	0.50	*0.70*	−0.09	*−0.40*	0.02	*0.68*	−0.06	−0.04	0.08
TT_2	0.91	*0.94*	−0.11	*−0.46*	−0.24	*0.94*	−0.04	0.04	−0.16
TT_3	0.78	*0.88*	−0.03	*−0.50*	−0.18	*0.86*	0.03	−0.03	−0.10
TT_4	0.82	*0.89*	−0.02	*−0.54*	0.04	*0.87*	0.06	−0.07	0.12
TT_5	0.98	*0.99*	−0.10	*−0.55*	−0.15	> *0.99*	−0.03	0.04	−0.06
TT_8	0.97	*0.98*	−0.03	*−0.54*	−0.04	*0.98*	0.03	−0.01	0.04
TT_13	0.93	*0.95*	−0.09	*−0.49*	−0.24	*0.95*	−0.02	0.02	−0.17
TT_14	0.82	*0.88*	−0.03	*−0.53*	0.12	*0.86*	0.01	−0.06	0.19

Table IV. Communalities and factor loadings

Figure 2.
The scree plot of eigenvalues

Factor	1	2	3	4
1	1.00	−0.06	−0.54	−0.08
2	–	1.00	−0.17	0.04
3	–	–	1.00	<0.01

Table V.
Correlation between factors

2 and 4 are the factors of apartment attributes. Significant difference between the structure and the pattern matrices exists only for the two location factors. The unique contribution of factor 1 has negligible correlation with income variables, thus high correlation with these variables in the structure matrix demonstrates segregation at the expense of common contributions. The unique contribution of factor 3 has low correlations with travel times (not higher than −0.07), but at the expense of common contributions the correlations are much higher (up to −0.55) in the structure matrix.

For each factor, we interpolate its score to a raster. The inverse distance weighted method is used with 12 neighbours, power 2 and output cell size of 10 metres. The raster maps are shown in Figures 3-6, where factor scores are grouped in nine classes.

Factor 1 is highly positively correlated with travel times to centres and thus represents locations farther from centres, where high-income households live as opposed to low-income population. In Figure 3, it is represented as a central core of low score and belts of increasing score. Note also that in the north, the third belt crosses the administrative boundary of Lyon. This district named Caluire-et-Cuire is urbanised and has the metro and trolleybus links with the central part of Lyon.

The spatial distribution of factor 3 is different irrespective of its correlation with factor 1. Factor 3 is highly positively correlated with low-income households and highly negatively correlated with high-income households. This is similar to the finding of Des Rosiers *et al.* (2000) in respect to the Quebec urban community. For the common contributions of our factor 3, it is also important to be closer to urban centres. Figure 3 clearly shows that the area with its highest score in the central part of Lyon is located between Victor Bach and Mutualité-Liberté (shown as centres 4 and 8, respectively) and overlaps with Guillotière – a problematic low-income area located remarkably close to

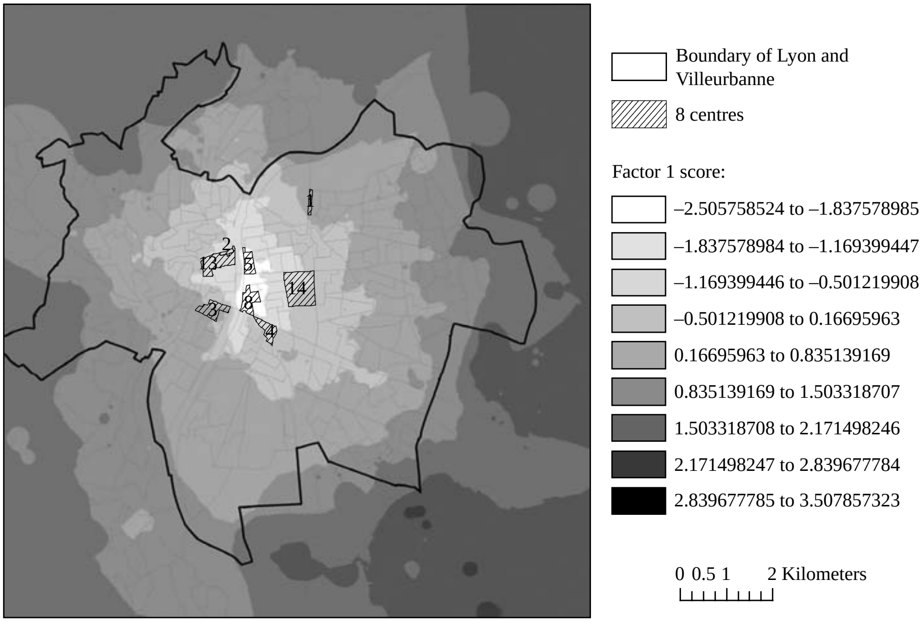

Figure 3.
Raster map of factor 1

the CBD, populated by immigrants and being the object of the specific attention of the

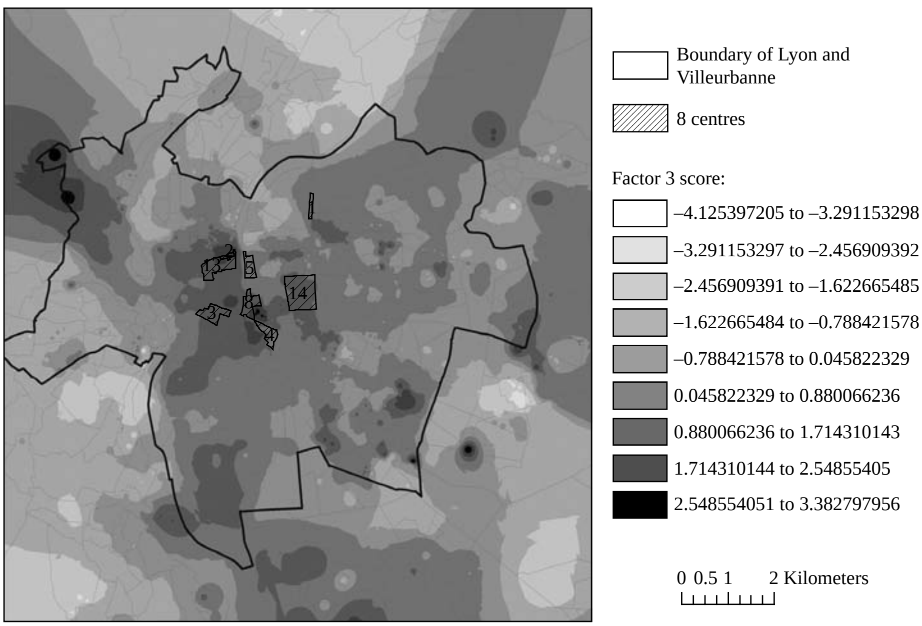

Figure 4.
Raster map of factor 3

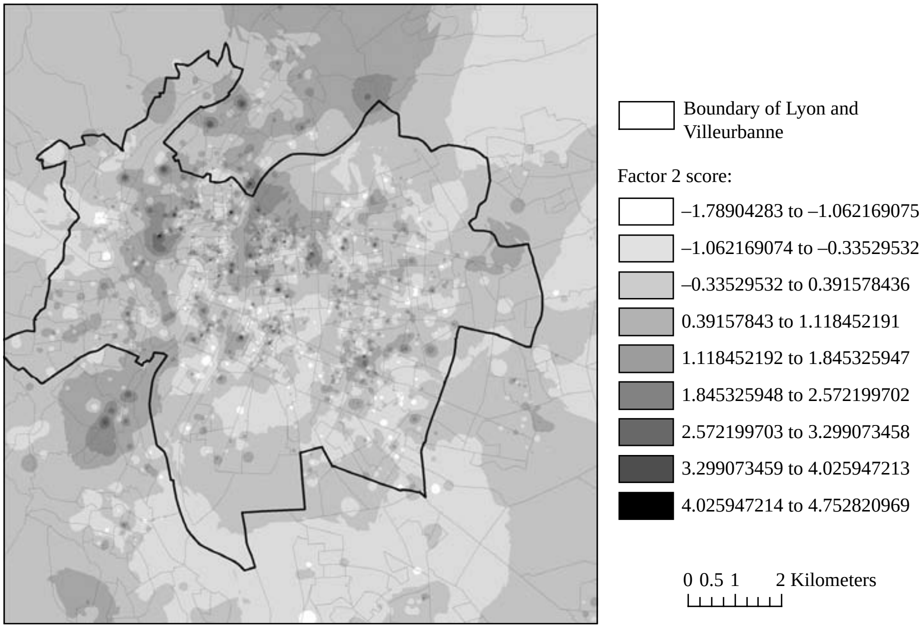

Figure 5.
Raster map of factor 2

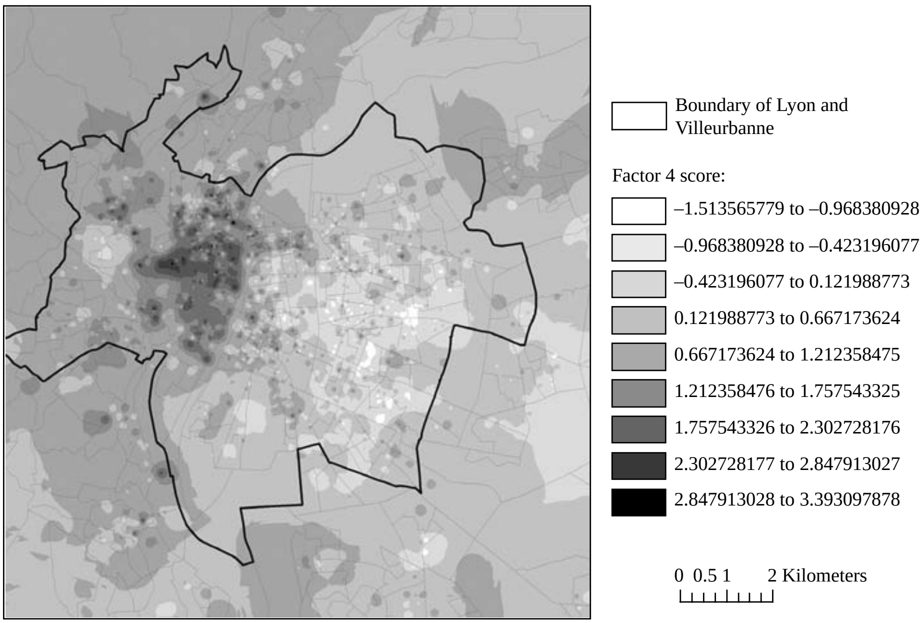

Figure 6.
Raster map of factor 4

police. The phenomenon of this area cannot be explained by the low level of employment accessibility as in Korsu and Wenglenski (2010).

Factors 2 and 4 account for internal apartment attributes. The former describes big and expensive apartments, the highest concentrations of which is seen in the most picturesque locations in the west of Croix-Rousse as well as near Cité International and Park de la Tête d'Or, which is regarded as the best urban park in France (Figure 5). Factor 4 refers to older apartments (whose attribute is cellars, though correlation with this variable is a bit lower than 0.30) in bad condition, the maximum is observed in the area near St Paul station with many buildings of the eighteenth and nineteenth centuries, while there are areas of low factor scores in the east of Lyon and south of Villeurbanne as well as to the east from their boundary (Figure 6).

5. PCA of location attributes and hedonic regression

We can execute one more exercise with location attributes by analysing travel times to all 15 service employment centers. As our second aim is data reduction and hedonic modelling, we can do this with PCA. With direct oblimin rotation for 15 travel time variables and two income variables for low- and high-income groups, we obtain three principal components whose eigenvalues are higher than unity. The communalities of variables are very high, ranged from 0.82 to 0.99 with a mean of 0.94. Correlation between the first and the second components is 0.54, between the first and the third is -0.49, while between the second and the third it is -0.32.

The configuration of a raster map of the third principal component is very similar to that of factor 3 shown in Figure 4. Figures 7 and 8 showing the two first principal components also show the boundary between Lyon and Villeurbanne. The first and the second components can tell more about the urban structure than factor 1 told. While the

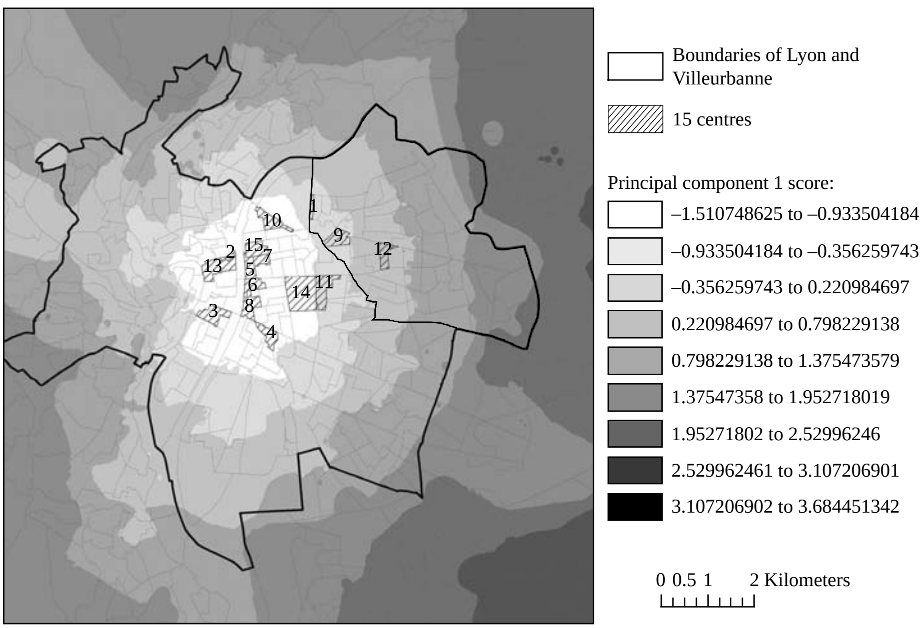

Figure 7.
Raster map of the first principal component for location

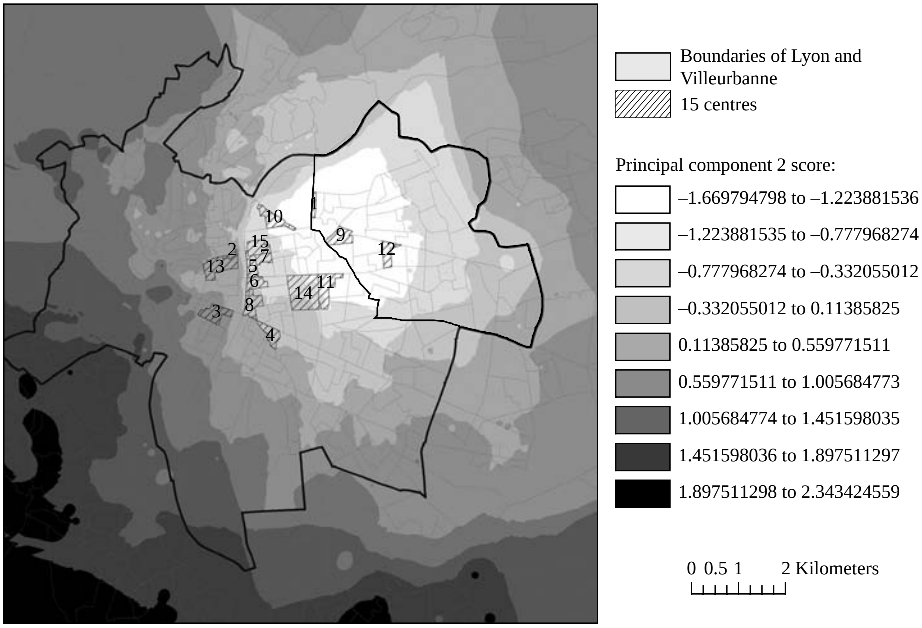

Figure 8.
Raster map of the second principal component for location

spatial distribution of the first component resembles those of factor 1, its core covers bigger areas including 11 of 12 service employment centres located in Lyon (Figure 7). The geographical distribution of the second component is shifted to Villeurbanne, and the three centres of Villeurbanne are located on the diagonal of its central core (Figure 8). Thus, the latent variables highlight the fact that though Lyon and Villeurbanne have many things in common, the influence of their centres is different and not yet amalgamated spatially.

For the second aim of the study, we create two hedonic models of apartment prices: with initial variables and with principal components substituting the initial variables. We compare regression performance, multicollinearity and spatial autocorrelation of residuals of the alternative models applying global and GWR techniques.

The dependent variable is the logarithm of *Price*. The logarithmic transformations of the independent variables *Area* and *Building_Age* are used as well as of *%LowIncome*, *%HighIncome* and travel times when these initial variables are included. Beside the dummy variables from Table III, we include also the dummies for transaction year (*Year98-Year08*), number of rooms (*Rooms2-Rooms8*), floor (*Floor1*, *Floor2_4*, *Floor5_8* and *Floor9more*), medium and bad condition (*CondMed* and *CondBad*) and number of cellars (*Cellar1* and *Cellars2*). The following attributes are used as default values in the hedonic modelling: transaction year 1997, one room, ground floor, good condition, no cellar, *Bath1*, *Park0* and *ViewGood*.

Table VI exhibits the comparison of the examined hedonic models. Only variables significant at the 5 percent level at least in one of the global estimations are presented, *t*-values are in parenthesis. In the models with principal components, the second component (Figure 8) is insignificant. The first component refers to higher income households, who live farther from service employment centres. Its negative sign indicates

Table VI.
Hedonic regression models

Variable	With initial variables Global	GWR	With principal components Global	GWR
Constant	8.99 (39.00)	9.47	6.95 (94.37)	7.13
First principal component	–	–	−0.16 (−31.28)	−0.17
Third principal component	–	–	−0.07 (−16.98)	−0.05
TT_3	−0.08 (−7.86)	−0.06	–	–
TT_2	−0.24 (−24.07)	−0.20	–	–
TT_14	−0.02 (−2.24)	−0.01	–	–
%LowIncome	−0.33 (−8.33)	−0.40	–	–
Year99	0.10 (3.74)	0.05	0.10 (3.73)	0.05
Year00	0.18 (6.91)	0.15	0.18 (6.79)	0.15
Year01	0.26 (10.14)	0.24	0.26 (9.66)	0.23
Year02	0.32 (12.27)	0.29	0.31 (11.87)	0.29
Year03	0.46 (18.23)	0.45	0.46 (17.80)	0.45
Year04	0.63 (24.45)	0.62	0.62 (23.85)	0.61
Year05	0.80 (30.75)	0.77	0.79 (30.08)	0.77
Year06	0.91 (34.68)	0.89	0.91 (34.01)	0.88
Year07	1.00 (37.70)	0.98	1.00 (37.03)	0.97
Year08	0.98 (33.68)	0.97	0.98 (32.98)	0.96
Area	0.96 (49.12)	0.93	0.99 (50.66)	0.94
Building_Age	−0.06 (−22.61)	−0.06	−0.05 (−19.12)	−0.05
Rooms4	−0.05 (−2.34)	−0.01	−0.07 (−3.11)	−0.01
Rooms5	−0.03 (−1.21)	0.02	−0.07 (−2.42)	0.03
Floor1	0.06 (4.09)	0.06	0.06 (4.32)	0.06
Floor2_4	0.09 (7.27)	0.10	0.10 (7.76)	0.11
Floor5_8	0.10 (6.86)	0.12	0.10 (6.79)	0.12
Floor9more	0.07 (2.62)	0.12	0.07 (2.46)	0.12
Bath2	0.06 (3.92)	0.04	0.07 (4.13)	0.04
CondMed	−0.10 (−10.30)	−0.10	−0.11 (−10.39)	−0.10
CondBad	−0.20 (−9.13)	−0.21	−0.20 (−8.82)	−0.21
Cellar1	0.02 (2.33)	0.06	< −0.01 (−0.01)	0.05
Cellars2	0.08 (2.72)	−0.01	0.06 (2.16)	−0.01
ParkUn	0.06 (5.31)	0.04	0.08 (6.51)	0.03
Park1	0.14 (12.58)	0.11	0.13 (11.54)	0.10
Park2	0.20 (11.84)	0.15	0.19 (11.10)	0.15
ViewNo	−0.03 (−4.30)	−0.03	−0.03 (−4.20)	−0.03
ViewBad	−0.09 (−3.35)	−0.08	−0.09 (−3.20)	−0.08
Garden	0.06 (2.99)	0.05	0.07 (3.52)	0.05
Terrace	0.05 (3.54)	0.05	0.05 (3.38)	0.05
Adjusted R^2	0.878	0.879	0.842	0.845

that irrespective the fact that in the central part of Lyon, higher percentage of low-income households live, apartment prices are more expensive there. At the same time, the third component, which accounts for low-income households located in some districts closer to the centres, also negatively influences apartment prices.

We should note that due to high multicollinearity we are able to include in the model without components only a few variables of travel times. The best specification contains the travel times to the three commonly recognised centres. To investigate multicollinearity, we can compare the maximums of variance inflationary factors (VIF). In both global models, VIF is lower than the threshold of 10; it is equal to 8.25 for the model with variables and 8.19 for the model with principal components. *Moran's I* measuring spatial autocorrelation of residuals is calculated with the row-standardised weight matrix of inverse squared distances. *Moran's I* is at the same level of 0.28 for both models.

For the GWR estimates, median values are shown in Table VI. The goodness of fit in comparison with global OLS is increased only marginally. For most location attributes and structure variables, GWR provides a bit lower estimates than global models. All location variables are found to be spatially non-stationary. The non-location attributes have similar coefficients and significance in the two global models, but when the models are spatially weighted, their estimates become practically indiscernible. Though in the models with principal components the intercept is reduced, their performance is by 3.6 percent and 3.4 percent lower in comparison to the models with initial location variables.

6. Conclusion

The first aim of the study to identify latent structure of apartment attributes and location is achieved with the geographical analysis of factors. EFA with oblique rotation is found to be applicable for extraction of latent variables providing an insight into apartment attributes and the complexity of urban structure. The results are intuitively easy to interpret. Factor analysis did not find a strong interaction between apartment attributes and location attributes: separate factors were obtained for the two groups. Of the two factors of apartment attributes, one accounts for big and expensive apartments and the other represents older apartments in bad condition. One of the two location factors demonstrates that the existing city boundary in the north seems to be outdated, while the other highlights the existence of a problematic low-income area in the central part of Lyon. The latter phenomenon cannot be explained by the low level of job accessibility. Taking into account the social reality of French cities nowadays, this finding is a visible alarm signal for policy-makers.

The limitation of EFA is its inability to work with many highly correlated variables. To include into analysis, the travel times to all the service employment centres, we applied PCA with non-orthogonal rotation. With more variables included, a more complex latent structure was delineated with separation between the centres of Lyon and those of Villeurbanne in spite of their long common boundary, common planning structure and transportation network and their common role of the central core of the Lyon urban area.

For the second aim to decrease the number of multiple explanatory variables in a hedonic price model, we used principal components instead of initial location variables. This does not gradually change either the estimates of non-location variables, especially when geographically weighted methodology is applied, or the indicators of multicollinearity and spatial autocorrelation of residuals. Two principal components with significant regression coefficients illustrate a contradiction between the traditionally positive perception of the central part of Lyon and its rather complicated social reality of today.

The principal component, which is geographically grouped around the centres of Villeurbanne, is insignificant in the regression model. This finding merits attention from both theoretical and practical viewpoints and the interesting question is to which extent it can be explained by the historical memory and socioeconomic perception of these two cities as separate objects and to which extent by the physical differences in distances or travel times.

The hedonic model with principal components accounts for all location attributes, while the model without components includes only three variables of travel times to centres and two variables of income groups. Nevertheless, the model with principal components does not demonstrate superiority. On the contrary, its goodness of fit is reduced. To investigate further the potential of PCA, a future study should focus on the clusters of principal components as proxies of apartment submarkets.

References

Anselin, L. (1988), *Special Econometrics: Method and Models*, Kluwer Academic, Dordrecht.

Bonnafous, A. (1973), *La logique de l'investigation économétrique*, Dunod, Paris.

Bourassa, S.C., Hamelink, F., Hoesli, M. and MacGregor, B.D. (1999), "Defining housing submarkets", *Journal of Housing Economics*, Vol. 8 No. 2, pp. 160-83.

Bourassa, S.C., Hoesli, M. and Peng, V.S. (2003), "Do housing submarkets really matter?", *Journal of Housing Economics*, Vol. 12 No. 1, pp. 12-28.

Breckler, S.J. (1984), "Empirical validation of affect, behaviour, and cognition as distinct component of attitude", *Journal of Personality and Social Psychology*, Vol. 47 No. 6, pp. 1191-205.

Bresson, G., Madre, J.-L. and Pirotte, A. (2004), "Is urban sprawl stimulated by economic growth?", paper presented at the 10th World Conference on Transport Research (WCTR), Istanbul, 4-8 July.

Brunsdon, C.F., Fotheringham, A.S. and Charlton, M.E. (1996), "Geographically weighted regression: a method for exploring spatial nonstationarity", *Geographical Analysis*, Vol. 28 No. 4, pp. 281-98.

Crites, S.L. Jr, Fabrigar, L.R. and Petty, R.E. (1994), "Measuring the affective and cognitive properties of attitudes: conceptual and methodological issues", *Personality and Social Psychology Bulletin*, Vol. 20 No. 6, pp. 619-34.

Dale-Johnson, D. (1982), "An alternative approach to housing market segmentation using hedonic price data", *Journal of Urban Economics*, Vol. 11 No. 3, pp. 311-32.

Des Rosiers, F. and Thériault, M. (2008), "Mass appraisal, hedonic price modelling and urban externalities: understanding property value shaping process", in Kauko, T. and d'Amato, M. (Eds), *Mass Appraisal Methods: An International Perspective for Property Valuers*, Blackwell, Oxford, pp. 111-47.

Des Rosiers, F., Thériault, M. and Villeneuve, P.-Y. (2000), "Sorting out access and neighbourhood factors in hedonic price modelling", *Journal of Property Investment & Finance*, Vol. 18 No. 3, pp. 291-315.

Dubin, R.A. (1998), "Spatial autocorrelation: a primer", *Journal of Housing Economics*, Vol. 7 No. 4, pp. 304-27.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. and Strahan, E.J. (1999), "Evaluating the use of exploratory factor analysis in psychological research", *Psychological Methods*, Vol. 4 No. 3, pp. 272-99.

Kolenikov, S. and Angeles, G. (2004), "The use of discrete data in PCA: theory, simulations, and applications to socioeconomic indices", Working Paper No. WP-04-85, MEASURE/Evaluation Project, 20 October, Caroline Population Center, University of North Caroline, Chapel Hill, NC.

Korsu, E. and Wenglenski, S. (2010), "Job accessibility, residential segregation and risk of long-term unemployment in the Paris region", *Urban Studies*, Vol. 47 No. 11, pp. 2279-324.

Kryvobokov, M. (2010), "Is it worth identifying service employment (sub)centres when modelling apartment prices?", *Journal of Property Research*, Vol. 27 No. 4, pp. 337-56.

Maclennan, D. and Tu, Y. (1996), "Economic perspectives on the structure of local housing systems", *Housing Studies*, Vol. 11 No. 3, pp. 387-406.

Mignot, D., Aguiléra, A., Bloy, D., Caubel, D. and Madre, J.-L. (2009), "Formes urbaines, mobilités et ségrégation. Une comparaison Lille, Lyon et Marseille (France)", *Recherche Transports Sécurité*, Vol. 102, pp. 47-59.

Öven, V.A. and Pekdemir, D. (2006), "Office rent determinants utilising factor analysis – a case study for Istanbul", *Journal of Real Estate Finance & Economics*, Vol. 33 No. 1, pp. 51-73.

Pan Ké Shon, J.-L. (2010), "The ambivalent nature of ethnic segregation in France's disadvantaged neighbourhoods", *Urban Studies*, Vol. 47 No. 8, pp. 1603-23.

Rosen, S. (1974), "Hedonic prices and implicit markets: product differentiation in pure competition", *Journal of Political Economy*, Vol. 82 No. 1, pp. 34-55.

Vias, S. and Kumaranayake, L. (2006), "Constructing socio-economic status indices: how to use principal components analysis", *Health Policy and Planning*, Vol. 21 No. 6, pp. 459-68.

Warner, R.M. (2007), *Applied Statistics: From Bivariate through Multivariate Techniques*, Sage, Thousand Oaks, CA.

West, S.G., Finch, J.F. and Curran, P.J. (1995), "Structural equation models with nonnormal variables: problems and remedies", in Hoyle, R.H. (Ed.), *Structural Equation Modeling: Concepts, Issues and Applications*, Sage, Newbury Park, CA, pp. 56-75.

Corresponding author
Marko Kryvobokov can be contacted at: mark.kryvobokov@let.ish-lyon.cnrs.fr