Measuring Urban Quality of Life Using Multivariate Geostatistical Models [1]

Alessandra Michelangeli
University of Milan-Bicocca, alessandra.michelangeli@unimib.it

Clarissa Ferrari, Marco Minozzo
University of Verona

Abstract: Urban quality of life (QOL) is usually measured through an index defined as the estimated value of a set of urban amenities. However there is an increasing awareness that omitted variables might seriously undermine the method's ability to accurately estimate QOL. Here we extend the hedonic approach using a multivariate geostatistical model to address the omitted variable bias by identifying the latent common factors responsible for the spatial distribution of the amenities. A new QOL index is then defined as a function of the latent factors whose implicit prices are estimated through hedonic regressions. Our methodology is shown on a data set of individual-level property transactions from the city of Vicenza. As a result we obtain the spatial distribution of QOL calculated according to the new index.

Keywords: Hedonic prices, Housing market, Monte Carlo EM algorithm, Spatial factor model.

1 Introduction

The well-being of people living in a city depends on the level of development of the city itself. The more the city is able to provide services and infrastructures, the better the living conditions are for its inhabitants. Obviously the range of factors that affect QOL is much wider and includes, among the others, climate conditions, environmental quality, the level of security for persons and things, and the socio-demographic environment. However, only partial statistical information is generally available, concerning a limited set of indicators that do not exhaust all the relevant factors. In this paper, we formulate a specific hypothesis to overcome the problem of shortage of statistical information. We suppose that the diverse factors affecting QOL can be subdivided into two groups; factors depending on the intervention of the public authority or of a private agent constitute the 'Agent-dependent factors group', whereas factors that do not depend on the intervention of some agent, as some characteristics of the landscape, constitute the 'Agent-independent factors group'. Then, assuming a model similar to that of Minozzo and Fruttini (2004), the

[1] We gratefully acknowledge funding from the Italian Ministry of Education, University and Research (MIUR) through PRIN 2008 project 2008MRFM2H, and Polo Scientifico Didattico "Studi sull'Impresa" (Vicenza) of the University of Verona.

two latent common factors behind these two groups are identified. Once the spatial distributions of these two latent factors are evaluated, we are able to assess their impact on QOL following the hedonic theoretical model of Rosen (1979)and Roback (1982). This approach usually defines a QOL index as a weighted sum of a set of urban amenities, where the weights are the hedonic prices of the amenities derived from the compensating differentials in the housing or in the labor markets, or in both.

We proceed as follows: in Section 2, we describe the different sources of available data. In Section 3 we briefly review the theoretical framework developed by Rosen (1979) and Roback (1982) from which the QOL is recovered. Then we define the multivariate geostatistical model used to identify the latent factors. Finally, we define a new QOL index as a function of the latent factors. The last section concludes the work.

2 Housing and Urban Data

The data come from different sources and are combined into a single data set. Housing market data come from the 'Osservatorio del Mercato Immobiliare' (OMI) managed by a public agency ('Agenzia del Territorio'), and refer to some 600 individual house transactions in Vicenza between 2004 and 2009. In addition to housing market values, the data set provides information also on structural characteristics of the properties. Housing prices are expressed in 2004 constant euros. On the other hand, the data on the amenities and the socio-demographic characteristics of the city are from the municipality and refer to environmental characteristics, educational services, commercial and administrative facilities, and public transports.All housing units and local amenities were geocoded by assigning to each of them a latitude and longitude coordinate by using a GIS-based geocoding application. Then, for each of the K housing units, that is, for each pointwise geographic location $\mathbf{x}_k$, $k = 1, \ldots, K$, of these units, we computed the Euclidean distance from the unit to the nearest representative of each of m categories of amenities. These distances, which we indicate with $y_1(\mathbf{x}_k), \ldots, y_m(\mathbf{x}_k)$, for $k = 1, \ldots, K$, constitute the key data for our geostatistical factor model.

3 Detection of Spatial Latent Factors for QOL

Our methodology is based on the model developed by Rosen (1979) and by Roback (1982) to assess urban QOL. The model depicts cities as interrelated bundles of wages, rents and amenities, with the specific combination of these elements differing across cities. Households and firms choose their location to, respectively, maximize utility and minimize production costs. Households are assumed as workers which are homogeneous in income and tastes. They maximize their utility function choosing the optimal bundle of composite good and residential location which allows access to local amenities. Firms combine capital and local labor using a production technology with constant returns to scale. The cost function depends both on input prices and the bundles of amenities. In equilibrium all households, regardless of their location, attain a common level of utility, and the unit production costs are equal to the unit production price. Equilibrium differentials for wages and housing prices can be used to compute implicit prices of amenities. Given the estimates of the implicit prices of amenities, the QOL index for any urban area is obtained by summing over all the average quantities of amenities using the implicit prices as weights. A serious drawback of this model is that the information about all urban attributes affecting QOL is unlikely to be available, and as Blomquist et al. (1988) note, even if all data were available, econometric specification problems such as collinearity would prevent the inclusion of all urban amenities. Furthermore, economic theory does not provide guidance for determining the optimal list of attributes. As a result, the empirical specification of the model may be plagued by omitted variable bias and measurement errors.

We overcome these problems by complementing this hedonic approach with the use of a hierarchical geostatistical factor model which will allow to arrive at a new QOL index. This model, which allows to deal with non-Gaussian data, can be seen as an extension to the multivariate context, of the classical geostatistical linear model of coregionalization (or of the spatial generalized linear model (Wang and Wall, 2003)). For a given set of distances $y_1(\mathbf{x}_k), \ldots, y_m(\mathbf{x}_k)$, for $k = 1, \ldots, K$, we assume that these are the realization, at the set of spatial locations $\mathbf{x}_1, \ldots, \mathbf{x}_K$, of a set of m random functions $Y_1, \ldots, Y_m$, for which we assume that, for $\mathbf{x} \in \mathbf{R}^2$,

$$Y_i(\mathbf{x})|Z_i(\mathbf{x}) \sim f_i(y; M_i(\mathbf{x})),$$

and

$$Z_i(\mathbf{x}) = \sum_{p=1}^{P} a_{ip} F_p(\mathbf{x}) + \xi_i(\mathbf{x}),$$

where $f_i(y; M_i(\mathbf{x}))$ is a Gamma density and $M_i(\mathbf{x}) = \mathrm{E}[Y_i(\mathbf{x})|Z_i(\mathbf{x})]$ is the conditional mean of the data given the latent part of the model, which is linked to $Z_i(\mathbf{x})$ by the link function $h_i(M_i(\mathbf{x})) = \beta_i + Z_i(\mathbf{x})$. As we said, $Y_i(\mathbf{x}_k)$ represents the minimum distance between the housing unit located at point $\mathbf{x}_k$ and the ith amenity. The latent part of the model resembles the classical linear factor model. Here, $F_p(\mathbf{x})$, for $p = 1, \ldots, P$, and $\xi_i(\mathbf{x})$, for $i = 1, \ldots, m$ are the common and the unique factors of the model, that we assume Gaussian, and for which we assume a spatial autocorrelation structure depending on a spatial autocorrelation function $\rho(\mathbf{h})$, for $\mathbf{h} \in \mathbf{R}^2$, such that $\rho(0) = 1$, and $\rho(\mathbf{h}) \to 0$, as $|\mathbf{h}| \to \infty$. In particular, we assume that $\mathrm{Cov}[F_p(\mathbf{x}), F_p(\mathbf{x} + \mathbf{h})] = \rho(\mathbf{h})$, and $\mathrm{Cov}[\xi_i(\mathbf{x}), \xi_i(\mathbf{x} + \mathbf{h})] = \psi_i \rho(\mathbf{h})$, for $\psi_i > 0$ (for more details see Minozzo and Fruttini, 2004, Minozzo and Ferrari, 2010). According to our proposal, we assume here that $P = 2$, that is, that there are two common latent factors responsible for the spatial distribution of the two groups of urban amenities ('Agent-dependent' and 'Agent-independent'). For this model, the intercept parameters $\beta_1, \ldots, \beta_m$, the coefficients a_{i1}, a_{i2}, for $i = 1, \ldots, m$, and the variances $\psi_1, \ldots, \psi_m$, can be estimated through Monte Carlo EM algorithms, whereas predictions of the common latent factors $F_1(\mathbf{x}_k)$ and $F_2(\mathbf{x}_k)$, at the spatial locations $\mathbf{x}_k$, for $k = 1, \ldots, K$, can be obtained by MCMC methods. Once the predicted spatial distributions of the two common latent factors have been obtained over the city area, and in particular at the set of K property locations, we can proceed in using this predictions to obtain the spatial distribution of our QOL index, defined as the weighted sum of the two common latent factors, using the implicit prices as weights. These implicit prices can be estimated by simple hedonic regressions or adopting more sophisticated techniques.

Preliminary assessments confirm a monocentric structure of the city QOL is higher in the city centre where there is a greater quantity and variety of amenities and decreases non monotonically as the distance from the center increases.

4 Concluding remarks

This paper extends the hedonic approach to measure urban QOL by using a hierarchical geostatistical factor model. The proposed methodology improves our ability to assess the spatial distribution of QOL and to identify its main causes. The standard approach developed by Rosen (1979) and Roback (1982) gives a synthetic measure of the QOL that people can on average enjoy in the considered areal unit, for example a city or a neighbourhood. Our approach allows to determine the QOL level at each spatial point obtaining the entire spatial distribution of QOL index instead of its average value.

References

Roback J. (1982) Wages, rents, and the quality of life, *Journal of Political Economy* University of Chicago Press, 90, 1257–1278.

Rosen S. (1979) Wage-based indexes of urban quality of life, in *Current Issues in Urban Economics*, Mieszkowsi, P. & Stratzheim, M. (Eds.), John Hopkins Press, Baltimore, 74–104.

Minozzo M., Ferrari C. (2010) A hierarchical geostatistical factor model for multivariate Poisson count data, *Working Paper Series, Department of Economics, University of Verona*.

Minozzo M., Fruttini D. (2004) Loglinear spatial factor analysis: an application to diabetes mellitus complications, *Environmetrics*, 15, 423–434.

Wang F., Wall M. M. (2003) Generalized common spatial factor model, *Biostatistics*, 4, 569–582.