WORKING PAPERS IN ECONOMICS

No. 06/05

LIV OSLAND, INGE THORSEN AND JENS PETTER GITLESEN

HOUSING PRICE GRADIENTS IN A GEOGRAPHY WITH ONE DOMINATING CENTER

Department of Economics
UNIVERSITY OF BERGEN

Housing price gradients in a geography with one dominating center[*]

Liv Osland[†], Inge Thorsen[‡] and Jens Petter Gitlesen[§]

March 17, 2005

---

[*] This paper has benefited from comments by Viggo Nordvik and Espen Bratberg.

[†] Stord/Haugesund University College, Bjørnsonsgt. 45, 5528 Haugesund, Norway (e-mail: liv.osland@hsh.no, fax: +47 52702765, phone: +4752702768)

[‡] Stord/Haugesund University College, Bjørnsonsgt. 45, 5528 Haugesund, Norway

[§] University of Stavanger, N-4036 Stavanger, Norway.

Abstract

We primarily focus on explaining housing prices and predicting housing price gradients in a Norwegian region with one dominating center (Stavanger). For such a geography spatial separation can be represented in a hedonic regression equation by a function of traveling distance from the city center. Several functions are tested, and some alternatives provide both a satisfying goodness-to-fit, consistent coefficient estimates, and intuitively reasonable predictions of housing price gradients. Still, not all commonly used functions are recommended. Spatial autocorrelation is removed when the hedonic function is properly specified.

Keywords: hedonic regression model, housing attributes, functional representation of spatial separation, spatial autocorrelation, housing price gradient, capitalization

1 Introduction

Housing price gradients represent important input to studies covering a wide range of regional and urban policy issues. One example is that investments in transportation infrastructure might cause reductions in traveling times to the cbd (central business district) and capitalize through property values. Our results are relevant in for instance studies related to constructing new roads (tunnels/bridges), changing speed limits or in analyzing investments inducing increased capacity and reduced queues on existing links.

Many empirical studies have aimed at finding rent gradients, land value gradients and/or housing price gradients. Ball (1973) and Bartik and Smith (1987) review several studies of the determinants of housing prices. In a few studies the variable indicating access to work came out with an insignificant sign, and occasionally a contra-intuitive sign was reported. Such results are explained either by multicollinearity problems, or by the fact that the study area in some cases involve a restricted urban area rather than a housing market area. Another reason for such results is that modern metropolitan areas tend to be multicentric. Dubin (1991) takes as his starting point the saying that ”the three most important determinants of housing prices are location, location and location”, and states that including distance from the cbd in a hedonic price function is the most common way of examining the effect of accessibility. Heikkila et al. (1989) state that most studies on land values have shown a significant and strongly negative land value gradient, when the hedonic price function is properly specified. Richardson (1988) also states that the main reason for insignificant or counter-intuitive results stems from a misspecified hedonic price function. This result is demonstrated in Waddell et al. (1993), who find that distance to the cbd is significant even when access to multiple employment centers and other nodes are accounted for. Adair et al. (2000) on the other hand claim that transport accessibility has limited explanatory power in modern segregated and segmented cities, and they recommend that studies focusing on the effect of spatial separation on housing prices are performed in homogenous markets.

In line with this tradition our main ambition is to study the functional form of the relevant gradient in a coherent labor and housing market area with one dominating center (Stavanger). Our study area is appropriate for this purpose, since it comes close to the geography underlying the traditional trade-off theory, while most empirical studies have considered more complex metropolitan areas. The breakthrough of this strand of analysis is represented by the “access-space-trade-off” model (Alonso 1964). Basically this model starts with a homogeneous landscape. There exists one central business district (cbd) where all employment and business activities are located. The households make two kinds of decisions: How much land do they want in what distance from the city centre? Households thus decide how much they are willing to pay for land by changing commuting time. This theoretical tradition of urban land use has been extended in many directions, see for instance Mills (1967), Muth (1969), Fujita (1986), and Brueckner (1987). The basic idea remains, however, represented by a steadily declining price of land and houses with increasing distance from the cbd.

Our empirical approach is based on the so-called hedonic method. The method is built on the idea that heterogeneous goods constitute of a number of attributes. Examples of attributes related to housing are the quality and size of the house. Achieved utility depends on the provisions of different attributes. When different combinations and amounts of attributes yield utility, and when there exist supply and demand for individual attributes, the hedonic approach introduces the implicit price structure of such attributes. These prices are defined as the increase in the total price of the corresponding good resulting from a marginal increase in the amount of the relevant attribute. The theoretical foundation of the hedonic method was established in Rosen (1974). Rosen showed that the hedonic price function is an equilibrium-function, enveloped by consumers' so-called "bid-functions" and suppliers' "offer-functions". Rosen's theory also offers the conditions for being able to interpret the implicit prices as the marginal willingness to pay for the respective attributes. For a review on hedonic analysis of housing markets, see Sheppard (1999).

A long tradition and a lot of literature in empirical hedonic housing market studies is devoted to the problem of defining a suitable delimitation of a housing market, both spatially and with respect to relevant types of dwellings, see for instance Straszheim (1974), Rothenberg et al. (1991), and Whitehead (1999). This problem is not addressed in this paper. The housing market in our prosperous study area is fairly homogenous. Spatially, the region is considered as one coherent housing and labor market area. Data on spatial labor market interaction clearly indicate that this is a reasonable assumption. We focus on one type of house only, that is privately owned single-family houses. In many cities the supply of this housing type is scarce relative to the situation in surrounding areas. This is not the case for Stavanger, where the proportion of privately owned single family houses is not markedly different from the average in Norway. On the other hand, the supply of other housing types, like terraced houses, is markedly underrepresented in more peripheral, rural areas of the region. Hence, such categories are less appropriate for the purpose of studying spatial variation in housing prices in a wide area.

In this paper we primarily search for general results on the relationship between housing prices and traveling time from the cbd. We are more concerned with predictability than in achieving a highest possible explanatory power, and we adopt a macroscopical perspective of the geography. Our macroscopical perspective is represented by a specific zonal subdivision of the geography, and the fact that we consider interzonal rather than intrazonal variations in housing prices. The zonal subdivision of the geography corresponds to the most detailed spatial level for which official Norwegian data are available. Data restrictions represent the main reason why we consider a relatively macroscopical description of the geography. Still, we strongly doubt that the additional insight and explanatory power resulting from a more disaggregated representation of the geography would be reasonably related to the enormous effort and resources required on data collection, if at all practically possible. For some spatially related attributes, like for instance the view, or the distance to nursery school, a relatively high degree of interzonal homogeneity can be expected. Many attributes are reasonably equally present in most of the (postal delivery) zones that we consider. We will of course account for the effect of some basic residence-specific attributes (internal living area, lot size, age of building etc.), but we ignore the impact of intrazonal location-specific amenities and services.

Section 2 provides documentation of our data, and the region is described and explained to be very appropriate for our purpose. As a benchmark for evaluating the impact of traveling time on housing prices, we start out by considering some non-spatial modeling alternatives. Those alternatives are described in Section 3, while the corresponding results are presented in Section 4. We also evaluate model performance for alternative delimitations of the geography. In Section 5 we evaluate traveling time and physical distance as alternative measures of spatial separation, once again for alternative delimitations of the geography. Section 6 focuses on different functional representation of the relationship between housing prices and traveling time. The evaluation is based both on explanatory power, and from the ability to predict reliable housing price gradients for a given set of values on other attributes than spatial separation from the cbd. Finally, in Section 7 we offer some concluding remarks.

2 The region and the data

2.1 The region

The study area in this paper is the southern parts of Rogaland, which is the southernmost county in Western Norway. This represents an integrated region with a connected road transportation network. There are 13 municipalities in the region, and each municipality is divided into postal delivery zones. All in all the region is divided into 98 (postal delivery) zones, as indicated in Figure 1. A strongly dominating part of journeys-to-work is made by cars. As an indicator of (commuting) distances by car, there is 79 km from the centre of Stavanger to Egersund in the south. According to our data the corresponding estimated shortest route traveling time is 69 minutes. This delimits a fairly large region in a hedonic context, motivated from the fact that our basic ambition is to study the relationship between labour and housing markets. The region comes close to what is defined as "an economic area" in Barkley et al. (1995), with a relatively self-contained labor market, and a relatively large central place (Stavanger) which influences on economic activity in a peripheral region. The high degree of intra-dependency within the region is very much due to physical, topographical, transportation barriers, that lengthen travel distances, and thereby deter economic relationships with other regions. The region is delimited by the North Sea in the west, fjords in the north and the east, while the southern delimitation is an administrative county border in a sparsely populated, mountainous area. Appendix A provides a list of municipalities and postal delivery zones, with corresponding figures of population and employment in 2001.

Figure 1: The division of the region into municipalities and zones

This prosperous region has experienced considerable population and employment growth over the last three decades. This is to a large extent due to the growth of petroleum based activities.

As mentioned in the introduction Stavanger is the dominating city centre of the region. Since 1970 population has increased by about 36% in Stavanger, while the corresponding population growth has been 91% in municipalities adjacent to Stavanger, and 49% in municipalities more peripherally located in the region. The corresponding population growth in Norway was about 17%. Considering such figures, it is important to know that land is relatively scarce in Stavanger, which is the fourth largest city in Norway. The suburbanization of both housing construction, shopping centers, and manufacturing firms has benefited the neighboring municipalities. The region has largely developed from employment growth in and close to the Stavanger city center. This contributes to make it appropriate for reaching reliable parameter estimates reflecting the "access-space-trade-off" rather than local characteristics of the central place system.

The geography we consider is not literally corresponding to the geography underlying the traditional trade-off theory, with a monocentric city in a featureless plain landscape. Still, it is probably hard to find geographies that come considerably closer to such a theoretical construction. The region has developed towards a central place system with centers at different levels, but Stavanger indisputably has a very dominating position, with a far higher rank than other centers in this system. This is not quite reflected in population figures for the municipalities in the region, since large parts of adjacent municipalities belong to the Stavanger residential area. This also applies for Sandnes, which is the second largest central place in this region, a lot larger than the third largest central place. The rapid population growth in Sandnes is explained primarily from the fact that it is located only 16 kilometers from Stavanger.

Land has become scarce in areas adjacent to the city center of Stavanger, and here has been a tendency that basic sector jobs have been decentralized to locations more distant from the cbd area. A large part of such activities is related to the administration of petroleum activities. The city center of Stavanger still has a dominating position concerning the supply of specific urban facilities, represented for instance by leisure and cultural services, and by shopping opportunities. The area has not developed into the characteristic multi-nodal structure observed in many metropolitan areas.

The fact that considerable commuting flows are observed from most parts of the region to Stavanger indicates that it makes good sense to consider the region as a suitable housing and labor market. Neither the labor market nor the housing market in any part of the region are significantly influenced by urban areas in other regions. The population growth throughout the region reflects the trade-off between housing prices and commuting costs. As mentioned above the municipalities adjacent to Stavanger has experienced a rapid population growth over the last decades. As a result of innovations in the transportation network, there is a tendency that this growth is spread to areas located in a longer physical distance from the Stavanger city area. As an example, Time is currently one of the municipalities in Norway that experiences the largest absolute increase in the number of dwellings. This is far from being matched by a corresponding increase in local employment opportunities. The traveling time from Bryne, the municipality center, to Stavanger is about 32 minutes, and Bryne is considered by many households to represent an attractive combination of traveling time and housing prices.

2.2 Data on housing prices

The housing market data consist of transactions of privately owned single-family houses in the period from 1997 through the first half of 2001. According to Statistics Norway, on average 52% of all sales in Norway are single-house dwellings. Our sample of 2788 observations represents approximately 50% of the total number of transactions of privately owned single-family houses in the region during the relevant period. In estimating the model we ignore transactions where information is missing for some variable(s). A lack of information on lot-size is the main reason why our sample is not even larger. To some degree the lack of information is positively related to the age of the houses. For other variables there is no indication in our data that our sample is not reasonably representative for the total number of transactions. We have in particular searched for possible spatial variation in the missing information. Despite some signs of varying inter-municipality practice in reporting data from transactions to official registers, we find no substantial tendency of systematic variation in available information across space.

The data on housing prices and housing attributes come from two different sources:

- Information on housing sales prices, postal codes, site in square meters, and type of building is collected from the national land register in Norway. This register is called the GAB-register. GAB is an abbreviation of ground parcel, address and building, and contains information on all ground parcels and buildings in Norway.

- Information on internal living space, garage, age of building, number of toilets and bathrooms as well as the time of sale is collected from Statistics Norway. These data come from a questionnaire that is sent to everyone who has bought a freeholder dwelling in Norway.

Each building is located on a postal code, and we use the postal codes to add information on distances and number of residents/employees.

Some information could be collected from both the GAB-register and Statistics Norway. In order to check which source is the most reliable, information was gathered from a third source, a local estate agent. Simple linear regressions were run between variables that existed in all three data-sets. If R^2 and the beta coefficient in the linear regression are 1, there is a perfect fit between the data-sets. This kind of analysis enabled us to discover serious problems with the living area variable found in the GAB-register, and the postal codes from Statistics Norway. By collecting data from three different sources, it was thus possible to verify that the postal codes in GAB were reliable, and that the variable for living area had to be collected from Statistics Norway.

In Appendix B we present descriptive housing market statistics. As expected both the average sales-price and the standard deviation are higher in Stavanger than in the other parts of the region. The dwellings with the smallest lot-size tend to be among the oldest in the sample, and the majority of those dwellings are located in Stavanger.

2.3 Data on population and employment

The division of the region into zones corresponds to the most detailed level of information which is available on individual residential and work location. The information is based on the Employer-Employee register, and provided for us by Statistics Norway. Access to data on zonal population is gained through the Central Population Register in Statistics Norway.

2.4 The matrix of physical distances and traveling times between the zones

The matrices of physical distances and traveling times were prepared for us by the Norwegian Mapping Authority. The calculations were based on the specification of the road network into separate links, with known distances and speed limits. Distances are given from a data base with an accuracy of ±2 meters for each link. In calculating traveling times it is accounted for the fact that actual speed depends on road category. Information of speed limits and road categories is converted into traveling times through instructions (adjustment factors for specific road categories) worked out by the Institute of Transport Economics. The centre of each (postal delivery) zone is found through detailed information on residential densities and the road network. Finally, both the matrix of distances and the matrix of traveling times is constructed from a shortest route algorithm. Interzonal distances are measured between zonal centers.

3 Non-spatial hedonic model formulations

In this study we do not attempt to account for accessibility to recreational facilities and shopping opportunities, and we ignore environmental conditions, location-specific amenities, and aesthetic attributes. This practice is partly explained from the fact that we consider interzonal rather than intrazonal variations in housing prices. Our approach is implicitly based on the assumption that such housing and location specific (micro-locational) attributes are not varying systematically across the zones. In other words we implicitly assume that the regional variation in such attributes can also be found within a zone, and that there is insignificant spatial variation in zonal average values.

We distinguish between two categories of attributes. One category is the physical attributes of the specific dwelling, the other is related to the location-specific attributes. In a general form the corresponding hedonic price equation can be expressed as follows:

$$P_{it} = f(z_{sit}, z_{lit}) \tag{1}$$

Here

P_{it} = the price of house i in year t

z_{sit} = value of dwelling-specific structural attribute s for house i in year t; $s = 1, ...S$, $i = 1, ...n$

z_{lit} = value of location-specific attribute l for house i in year t; $l = 1, ...L$, $i = 1, ...n$

All the hedonic regression results to be presented in this paper involve the same set of *dwelling-specific attributes*, z_{si}, which are defined in Table 1.

Table 1: List of non-spatial variables

Variable	**Operational definition**
REALPRICE	selling price deflated by the consumer price index, base year is 1998
AGE	age of building
LIVAREA	living area measured in square meters
LOTSIZE	lot-size measured in square meters
GARAGE	dummy variable indicating presence of garage
NUMBTOIL	number of toilets in the building
REBUILD	dummy variable indicating whether the building has been rebuilt/renovated

The ambition to capture the trade-off between housing market prices and labor market interaction calls for a large study area. As a working hypothesis we claim that unbiased estimates of this relationship cannot be based on a truncated specification of the market area. We prepare for a discussion on this by estimating the alternative model formulations for the following three subdivisions of the geography:

- The entire region
- The four most centrally located municipalities in the northern parts of the region (Stavanger, Sandnes, Randaberg and Sola, see the map in Figure 1)
- Stavanger

This corresponds to a natural subdivision of the geography, where Stavanger represents the cbd, while the four most centrally located municipalities represent an extended urban area in a relatively monocentric spatial structure. We have also experimented with the mathematical representation of the relationship between dependent and independent variables. As in most other empirical studies of the housing market we find that log-linear model formulations are superior to linear and semi-logarithmic model specifications. Only results based on log-linear specifications are presented.

4 Results based on non-spatial model formulations

Least squares estimation results based on the non-spatial model formulations are presented in Table 2. Model M1 refers to a specification where the estimation is based on data from the entire region, while the estimation of M2 is based on data from the four most centrally located municipalities, and M3 is based on data from Stavanger. To obtain a high number of observations, the data are aggregated through time. In order to account for increases in housing prices during the period, changing intercepts are introduced through dummy variables for each year. The dummy for 1998 is excluded in order to avoid perfect multicollinarity between these variables.

In general the coefficient estimates in Table 2 are supposedly biased, since important information on location is omitted from the model formulation underlying the results. This is especially evident for the variable LOTSIZE. The estimate of the coefficient corresponding to this variable has a counter-intuitive sign in the model specification based on data covering the entire region. We will return to a discussion of this result in Section 5.

Let β_A denote the coefficient attached to the variable AGE, while β_{AR} is attached to AGE $\cdot$ REBUILD. A significantly positive estimate of β_{AR} means that the negative impact of AGE upon housing price is reduced. It is intuitively reasonable, however, that $|\beta_A| > |\beta_{AR}|$, meaning that AGE has a negative influence on housing price, even if the house has been rebuilt.

White's general test (see for instance Greene 2003) is performed to test for heteroskedasticity. Since $\chi^2_{0,05} = 16,919$ it follows from Table 2 that the hypothesis of homoskedasticity is rejected in all model specifications. In order to make reliable inferences on the least square estimates when heteroskedasticity is present, the reported standard errors in all models will be estimated by the robust estimator of variance. In our data, however, this robust estimator of variance does not produce results that deviate much from estimates based on the ordinary least squares estimator.

According to Odland(1988) a test for spatial autocorrelation in the residuals should always be applied when using spatial data. Spatial heterogeneity implies violations of the assumption of a spherical error covariance matrix (Anselin 1988). This means that spatial heterogeneity represents an efficiency problem. If positive spatial dependence in the errors is present the estimated values of R^2, t- and F-values will be artificially high (Odland 1988). Inferences may hence be misleading and tests for heteroskedasticity may have reduced validity (Anselin and Rey 1991). According to Anselin (1988) and Odland (1988) the reason for spatial heterogeneity lies in the estimated model. This could be due to a poor functional representation of the relevant problem, and/or to the possibility that our list of independent variables fails to account for relevant spatial peculiarities or interdependencies in the data.

Moran's I is reported as a descriptive measure of whether there exists spatial autocorrelation in the residuals[1]. In the computation a binary row standardized weight matrix is used to define

---

[1] Moran's I values are computed in the program R by using R-packages developed by Roger Bivand, Norwegian School of Economics and Business Administration.

Table 2: Results based on non-spatial modeling specifications

	M1	M2	M3
Constant	11.83	11,77	11,26
	(0,179)	(0,183)	(0,136)
LOTSIZE	-0,0154	0,0644	0,0952
	(0,0136)	(0,0129)	(0,0159)
AGE	-0,045	-0,0397	-0,0461
	(0,0075)	(0,0076)	(0,0080)
AGE·REBUILD	0,0171	0,0130	0,0154
	(0,0040)	(0,0036)	(0,0043)
GARAGE	0,081	0,052	0,0435
	(0,0149)	(0,0134)	(0,0168)
LIVAREA	0,4330	0,3774	0,4665
	(0,0418)	(0,0459)	(0,0341)
NUMBTOIL	0,2235	0,1701	0,1226
	(0,0267)	(0,0268)	(0,0243)
YEARDUM97	-0,1181	-0,1250	-0,1523
	(0,0178)	(0,0177)	(0,2410)
YEARDUM99	0,1565	0,1454	0,1305
	(0,0188)	(0,0176)	(0,0232)
YEARDUM00	0,2835	0,2787	0,2582
	(0,0178)	(0,0173)	(0,0218)
YEARDUM01	0,3120	0,3127	0,2834
	(0,0185)	(0,0185)	(0,0234)
n	2788	2051	1188
R^2	0,5221	0,5944	0,6606
R^2-adj.	0,5204	0,5924	0,6578
L	-580	7	57
APE	312233	277302	273133
White test statistic	278	135	110
Moran's I	0,3590	0,1554	0,0718
Standard Normal Deviate (z_I)	184	71	26

Note: Results based on observations from the period 1997-2001, robust standard errors in parentheses. M1 refers to a specification where the estimation is based on data from the entire region, while the estimation of M2 is based on data from the four most centrally located municipalities, and M3 is based on data from Stavanger.

the relation between observations. Zones which have common borders in the geography are neighbors. All houses within a zone are also neighbors. A house is not a neighbor to itself.

The values of Moran's I range approximately from -1 to 1. Positive values indicate positive autocorrelation. A value of zero (or more precisely, $-\frac{1}{n-1}$ (Florax et al. 2002)), may indicate no spatial autocorrelation. In Table 2 we have also reported values of the standard normal deviate (z_I) that is constructed from values of the mean and the variance of the Moran statistic (see for example Anselin (1988) for details on the estimation of such values). Corresponding to this transformed variate the null hypothesis is the absence of spatial autocorrelation in the residuals. The alternative hypothesis is spatial autocorrelation of some unspecified kind (Anselin 1988). This means that the null is rejected at the 5% significance level if $z_I > 1,645$. It follows from Table 2 that the null is rejected in all the non-spatial model specifications. It also follows from the table that (positive) spatial autocorrelation is strongest when estimation is based on data from the entire region. According to the values of Moran's I a natural hypothesis is that the systematic pattern of spatial interdependencies across zones is positively related to how large part of the region that is considered.

In addition to R^2 we have reported values of two measures to evaluate the goodness of fit abilities of various model alternatives. L is the log-likelihood value. The Average Prediction Error (APE) is explicitly based on a comparison between the observed and the predicted housing prices; APE $= \frac{\sum_i (|\hat{P}_i - P_i|)}{n}$. Here $\hat{P}_i$ is the predicted price of house i, while n is the observed number of houses. APE has an obvious interpretation, but this measure is not appropriate for statistical testing and to discriminate between alternative model specifications.

According to Table 2 the non-spatial independent variables explain more of the variation in housing prices the smaller part of the region that is considered. This justifies the hypothesis that information on spatial characteristics like distances is particularly important in studies based on a wide delimitation of the study area. If the study area is restricted for instance to a specific urban area distances probably contribute less to the explanation of housing price variation, and non-spatial housing attributes explain more of the variation in housing prices. The corresponding hypothesis is that short distances within a sub-area have only marginal impact on individual spatial labour market behavior and housing prices. We will return to this discussion in forthcoming sections, where spatial characteristics are explicitly taken into account.

Notice from Table 2 that log-likelihood values apparently are obscure in the logarithmic model specifications. Positive log-likelihood values might in general result in rare cases for density functions with very small variance, allowing for density values exceeding 1,0. Such cases typically are met in problems where dependent variables are defined for a relatively small range of high values. In our study the logarithm of housing prices defines a function that is very flat for the relevant range of values, with correspondingly low variance.

5 Results based on hedonic model formulations including traveling time from the labor market center in the region

In order to estimate the housing-price gradient, it is necessary to identify the center of the geography. Since we focus on the trade-off between housing prices and commuting costs it is reasonable to find the labor market center. According to Plaut and Plaut (1998) much of the empirical literature in the field assumes that the location of the center is known in advance. In our study the location of the labor market center is found endogenously. For this purpose we introduced a gravity based accessibility measure. Assume that distance (traveling time from origin j to destination k, d_{jk}) appears through a negative exponential function in the definition of the accessibility measure, and let σ be the weight attached to distance; $\sigma < 0$. The Hansen type of accessibility measure (Hansen 1959), S_{ij} is then defined as follows:

$$S_j = \sum_{k=1}^{w} D_k \exp(\sigma d_{jk}) \tag{2}$$

Here, D_k represents the number of jobs (employment opportunities) in destination (zone) k. The measure S_{ij} is based on the principle that the accessibility of a destination is a decreasing function of relative distance to other potential destinations, where each destination is weighted by its size, or in other words the number of opportunities available at the specific location. Hence, it can be interpreted as an opportunity density function.

It is not a priori obvious that the zones in the city center of Stavanger represent the labor market center of the region. In addition there is in particular a high labour demand originating from an area hosting for instance large industrial firms and administrative units related to petroleum activities. This industrial area is located in between the Sandnes and Stavanger. Another category of zones with high labor market accessibility are those located in an intermediate position between the center of Stavanger and the mentioned industrial area. Still, our results reveal a clear tendency that model performance is best when centrally located zones in Stavanger is chosen as the labor market center. To be more specific zone 10 is found to represent a natural labor market center in the region.

As another starting point for the results to be presented in this section we found that model performance is significantly better when traveling time rather than physical distance is used as the measure of spatial separation.

Studies of journeys-to-work report significant distance deterrence effects, see for instance Thorsen and Gitlesen (1998). One general problem in empirical spatial interaction studies is to choose an appropriate specification of the distance deterrence function. Some studies are based on a power function ($d_{ij}^{\beta_p}$), while others are based on an exponential deterrence function ($e^{\beta_e d_{ij}}$). In both specifications the distance deterrence parameter is of course negative; $\beta_p < 0$ and $\beta_e < 0$. The choice of the distance deterrence function has been considered to be essentially a pragmatic one in the literature, see for instance Nijkamp and Reggiani (1992). For some problems this might be correct, but some studies have concluded that the appropriateness of the functional form should be critically examined, see for instance a study of US migration flows in Fik and Mulligan (1998).

Combined with the three alternative delimitations of the geography (the entire region, the four most centrally located municipalities, and Stavanger), we consider two alternative specifications of the spatial separation function:

M8-M10: a negative exponential function, three alternative delimitations of the geography

M11-M13: a power function specification, three alternative delimitations of the geography

The models which are linear in all parameters are estimated by ordinary least squares, whereas the models which are nonlinear in at least one parameter is estimated by using nonlinear least squares and maximum likelihood estimation. The general rule is that the result from the general least squares estimations are used as starting values for the maximum likelihood estimation. In all the nonlinear models the two methods give identical results.

In the previous section we found that especially the coefficient related to lot size varied considerably with respect to the subdivision of the geography. In the case where estimation is based on data from the entire region we even found that this coefficient has a counter-intuitive, negative, sign. This is probably a result of the fact that non-spatial model specifications ignore the positive covariation between lot size and the distance from the cbd. Such counter-intuitive results do not appear when relevant measures of spatial separation are included.

As another attempt to capture area-specific variation in coefficient estimates we have introduced a dummy variable representing the most rural areas in the geography. This dummy variable takes the value 1 if the municipality is located in the most rural area, else 0. In those rural, spatially adjacent, municipalities the ratio of inhabitants to open land is considerably lower than in the rest of the geography, see Table 3. As a comparison to the figures in this table, Eigersund has 34 inhabitants per km^2, which is the fifth lowest in the region. Stavanger has the highest population density, with 1631 inhabitants per km^2.

Table 3: The most rural areas in the geography

	Inhabitants per km^2	Average distance from labour market center (minutes)
Gjesdal	16	25
Bjerkreim	4	46
Lund	9	87
Sokndal	13	88
Regional average	233	19

We have introduced in our models the product of the dummy variable representing the rural areas and the variable LOTSIZE. This new variable, RURLOT, applies only when the estimation is based on data from the entire region. By comparing model specifications M8b and M11b to the corresponding versions with no such stratification of the geography, it follows from Table 4 that the introduction of RURLOT increases model performance significantly. Testing the joint significance of the two variables LOTSIZE and RURLOT by a Wald test, indicates significant differences in the elasticity of LOTSIZE in the rural and non-rural areas. No other variable had significantly different slopes when stratifying our data this way.

The number of observations in this study is relatively large, hence multicollinearity is not expected to be a problem. This is also confirmed by estimated values of the variance inflation factor (see for instance Greene (2003)), which is not reported here. The parameter estimates are significant, and relatively stable in regressions referring to different time periods, different sets of variables, and different delimitations of the geography.

In Table 4 we also report the p-values of the Ramsey reset test (see for instance Davidson and MacKinnon (1993)). This is usually referred to as an omitted variable test. The null is that the model is correctly specified. At the 5% level of significance we find that the null is rejected for model specifications M4, M4b, M6 and M7. This suggests that these models are incorrectly specified. Note also that M4 and M7 do not include the variable RURLOT.

In comparing the parameter estimates in Table 4 to their corresponding counterparts in models M1-M3 in Table 2, the most noticeable difference appears for the parameter estimate related to LOTSIZE. In the non-spatial model specifications this parameter estimate varied considerably with respect to the alternative subdivisions of the geography. When the entire region is considered the relevant parameter is even estimated to be negative. Lot is a completely immobile attribute, while all other attributes can be modified independent of location. Since LOTSIZE tends to be largest in areas distant from the labour market center, with relatively low housing prices, this variable will be systematically correlated with omitted variables in the non-spatial model specifications. Hence, the relevant parameter estimates in Table 2 to some degree captures the effect of a falling housing price gradient. The estimation results also indicate that the impact of this effect is positively related to how large part of the region that is considered. This is reasonable, since the negative correlation between LOTSIZE and housing prices probably is not particularly strong when the study area is restricted to a rather small part of the region (for instance the regional center, Stavanger).

Table 4: Results based on model specifications where spatial separation is measured by traveling time from the labour market center of the geography

	M4	M4b	M5	M6	M7	M7b	M8	M9
Constant	10,9235	10,8766	10,6641	10,51	12,15	12,09	11,87	11,66
	(0,0887)	(0,0873)	(0,1006)	(0,1358)	(0,0811)	(0,0896)	(0,1020)	(0,1364)
LOTSIZE	0,1126	0,1181	0,1310	0,1371	0,1240	0,1319	0,1351	0,1338
	(0,0110)	(0,0106)	(0,0113)	(0,0156)	(0,0088)	(0,0100)	(0,0111)	(0,0154)
RURLOT	-	-0,02914	-	-	-	-0,0319	-	-
	(-)	(0,0033)	(-)	(-)	(-)	(0,0033)	(-)	(-)
AGE	-0,0799	-0,0763	-0,0698	-0,0728	-0,0931	-0,0897	-0,0768	-0,0761
	(0,0066)	(0,0065)	(0,0072)	(0,0087)	(0,0048)	(0,0065)	(0,0073)	(0,0084)
AGE·REBUILD	0,0106	0,01053	0,0113	0,01537	0,0109	0,0107	0,0116	0,0158
	(0,0028)	(0,0030)	(0,0032)	(0,0041)	(0,0028)	(0,0030)	(0,0032)	(0,0041)
GARAGE	0,0683	0,0639	0,0414	0,02748	0,0749	0,0706	0,0479	0,0339
	(0,0113)	(0,0110)	(0,0119)	(0,0161)	(0,0104)	(0,0111)	(0,0120)	(0,0160)
LIVAREA	0,3583	0,3558	0,3865	0,4190	0,3581	0,3553	0,3743	0,4107
	(0,0179)	(0,0179)	(0,0218)	(0,0323)	(0,0162)	(0,0179)	(0,0219)	(0,0327)
NUMBTOIL	0,1515	0,1491	0,1351	0,1236	0,15933	0,1557	0,1421	0,1292
	(0,0151)	(0,0147)	(0,0167)	(0,0227)	(0,0142)	(0,0148)	(0,0167)	(0,0228)
β_e (exponential)	-0,0256	-0,0218	-0,0254	-0,0303	-	-	-	-
	(0,0013)	(0,0015)	(0,0018)	(0,0047)	(-)	(-)	(-)	(-)
β_p (power)	-	-	-	-	-0,2363	-0,2198	-0,1794	-0,1500
	(-)	(-)	(-)	(-)	(0,0054)	(0,0060)	(0,0082)	(0,0127)
YEARDUM97	-0,1352	-0,1343	-0,1302	-0,1551	-0,1342	-0,1332	-0,1307	-0,1557
	(0,0139)	(0,0136)	(0,0158)	(0,0231)	(0,0133)	(0,0136)	(0,0158)	(0,0231)
YEARDUM99	0,1305	0,1261	0,1383	0,1466	0,1378	0,1329	0,1421	0,1437
	(0,0140)	(0,0137)	(0,01561)	(0,0217)	(0,0139)	(0,0138)	(0,0156)	(0,0218)
YEARDUM00	0,2700	0,2678	0,2710	0,2627	0,2714	0,2690	0,2718	0,2624
	(0,0139)	(0,0136)	(0,0154)	(0,0205)	(0,0135)	(0,0135)	(0,0154)	(0,0204)
YEARDUM01	0,3079	0,3019	0,3062	0,3027	0,3128	0,3065	0,3107	0,3008
	(0,0140)	(0,0137)	(0,0163)	(0,0224)	(0,0144)	(0,0137)	(0,0164)	(0,0223)
n	2788	2788	2051	1188	2788	2788	2051	1188
R^2	0,7231	0,7350	0,6751	0,6910	0,7175	0,7319	0,6734	0,6949
R^2-adj.	0,7220	0,7339	0,6734	0,6881	0,7163	0,7308	0,6717	0,6920
L	204,4133	266	257	112	176	249	252	120
APE	221409	217571	234292	255226	227961	221501	235668	254110
White test statistic	253	261	192	114	249	263	146	114
Moran's I	0,0271	0,0066	0,0061	0,0102	0,0431	0,0077	0,0037	0,0064
Standard normal deviate (z_I)	11,49	3,44	2,50	3,73	16,09	4,46	2,07	2,83
Ramsey reset test (p-value)	0,0000	0,0001	0,3976	0,0056	0,0294	0,1523	0,68	0,47

Note: Results based on observations from the period 1997-2001, robust standard errors in parentheses. Models M8-M10 are based on a negative exponential spatial separation function, M11-M13 are based on a power function specification. The number of observations (n) indicates the relevant delimitation of the geography.

We will not enter into a detailed discussion of specific parameter estimates in Table 4. As a general comment there is a tendency that the introduction of distance results in more precise parameter estimates, and that the estimates are less dependent on what subdivision of the geography they refer to. Hence, it is important to account for an appropriate measure of spatial separation to reach a satisfying identification of how partial variation in the independent variables affects housing prices.

The introduction of distance from the labour market center in general improves the goodness-of-fit considerably. This especially applies for the case where the estimation is based on data from the entire region, with R^2 (adjusted) increasing from 0,5120 (model M1) to around 0,73 (models M4b and M7b)). In the case where the study area is restricted to Stavanger the corresponding increase only range from 0,6578 (M2) to around 0,67 (M6 and M9).

With reference to the trade-off theory it is natural that the contribution of distance in explaining housing prices increases with the spatial extension of the study area within a labour marked region. This pattern is reflected in the values of all the reported indicators of model performance in Table 4. Notice, however, that the introduction of distance significantly improves model performance for all the reported subdivisions of the geography. In the case based on data from the entire region (comparing M7b to M1) the value of the likelihood ratio test statistic is 1658, which of course by far exceeds the critical value of a chi square distribution with two degrees of freedom. When only Stavanger is considered (comparing M9 to M2) the value of this test statistic is 126.

By comparing Table 2 and Table 4 it follows that the values of Moran's I are considerably reduced when a distance from the cbd is introduced in the model specifications. By comparing M4 to M4b and M7 to M7b it also follows that spatial autocorrelation is reduced when the variable RURLOT is introduced. This clearly indicates that at least a large part of the spatial autocorrelation in the non-spatial modeling alternatives was due to the fact that important information was omitted from the model specifications. RURLOT and, in particular, distance from the cbd represent characteristics of spatial structure that influence housing prices. Notice from Table 4, however, that the hypothesis of no spatial autocorrelation has to be rejected in all the models M4-M9. Hence, the presence of autocorrelation is not removed, and the estimates of for instance R^2 might still be artificially high.

6 Housing price gradients and alternative functional representation of the relationship between housing prices and traveling time

6.1 Comparing the exponential and the power function specification

The two gradients in Figure 2 apply for a given set of values on all independent variables except the traveling time from the cbd. We consider a fictional house that has not been rebuilt, and with a garage. The fictional house is not located in the most rural areas, and the price in the figure refers to year 2000. For the remaining independent variables we use observed average values. Given this fixed set of values on attributes we predict housing prices in varying distances from the cbd. The dashed curve refer to a path where spatial separation is represented by $e^{\hat{\beta}_e d_{ij}}$, while the solid curve reflects the power function specification ($d_{ij}^{\hat{\beta}_p}$). The parameter values are based on models M4b and M7b, which means that $\hat{\beta}_e = -0,0218$ and $\hat{\beta}_p = -0,2198$.

In general it follows from the results in Table 4 that the two alternative functions of traveling time only result in marginal differences in explanatory power. For all the measures of explanatory power a tendency can be found, however, that the exponential function performs best when the

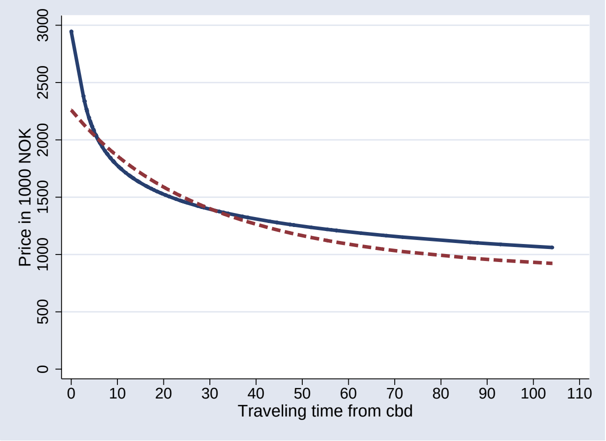


Figure 2: Predicted housing price gradients for a fictional standard house. Estimated parameters are based on data from the entire region. The dashed curve reflects an exponential relationship between housing price and traveling time, the solid curve is based on the power function.

estimation is based on data from the entire region, while the power function performs best in the case where data is restricted to Stavanger. In terms of a log-likelihood ratio test such differences are statistically significant. Still, our results do not mean that we can conclude that the relative performance of the exponential function is positively related to how large part of the region data refer to. Our results are based on just one observation of a region, with only three subdivisions of this geography into specific areas. At the same time, however, our results clearly indicate that the prediction of a housing price gradient is very sensitive both to the choice of how traveling time is represented and to how large part of the region data refer to.

In comparing model formulations with different functional representations of traveling time, the estimated marginal impact of most attributes does not differ considerably, and neither do the values of important test statistics. Some parameter values do not even seem to be particularly sensitive with respect to the choice between spatial and non-spatial model formulations. Such results should be interpreted with care. We deal with nonlinear relationships and mathematical transformations of dependent and independent variables. Predictions might differ substantially even with small differences in parameter estimates and in the measures of explanatory power.

It follows from Figure 2 that predictions on changes in housing prices in some cases are rather sensitive to the choice of model, despite the fact that the models do not differ substantially with respect to explanatory power. One important application of the results is to discuss the impact on housing prices of changes in the road transportation infrastructure. Assume for instance that a new road connection reduces traveling distance to the cbd from 30 minutes to 15 minutes for a specific zone. According to our estimates the predicted increase in housing prices in this zone is 100000 NOK higher when the model is based on an exponential representation of traveling time than it is in the case with a power function. This difference represents almost 30% of the predicted change in housing prices. As can be seen from Figure 2 predictions are even far more sensitive to the choice of the relevant functional specification in cases where the changes in the transportation infrastructure network affect areas close to the cbd.

It also follows from Figure 2 that the two alternative model specifications do not predict considerable differences in the level of housing prices for locations beyond about 5 minutes from the cbd. The situation is different in a case where the parameter estimates are based on observations from only Stavanger. The housing price gradients illustrated in Figure 3 are based on the same fictional standard house that was considered in Figure 2. Notice first that a predicted housing price in a long distance from the cbd is now a lot more sensitive to the choice of a functional representation of traveling time than in the case where estimation is based on observations from the entire region. This illustrates the trivial but important point that identification of such a relationship is positively related to the deviation in observed values of the independent variable. Hence, a predicted housing price gradient is in general more reliable if it is based on observations covering a wide range of values on traveling time from the cbd. Our study area is very appropriate for this purpose, since the local housing market only marginally interferes with housing markets in adjacent areas outside the region. The region can in this respect be considered as an isolated island with one dominating central place.

Another point to notice is that the sensitivity of the housing prices with respect to variations in short distances from the cbd changes considerably for both model specifications if only data from Stavanger are used in the estimation procedure. It is reasonable that the two alternative model specifications predict very similar housing prices for distances within 15 minutes from the cbd in a case where the estimation is primarily based on data for such values of distance. At the same time those results also contribute to an evaluation of the two alternative housing price gradients in Figure 2, with data based on observations from the entire region. To be specific, the results clearly indicate that the power function approach gives a considerably biased slope of the housing price gradient for low values of traveling time from the cbd. This conclusion also corresponds to a combination of intuition and knowledge of the local geography. Hence, gradients based on the power function seem to predict too radical changes in housing prices for variations in distances close to the cbd. With data based on the entire geography the exponential function results in a more reasonable housing price gradient for this rather monocentric geography.

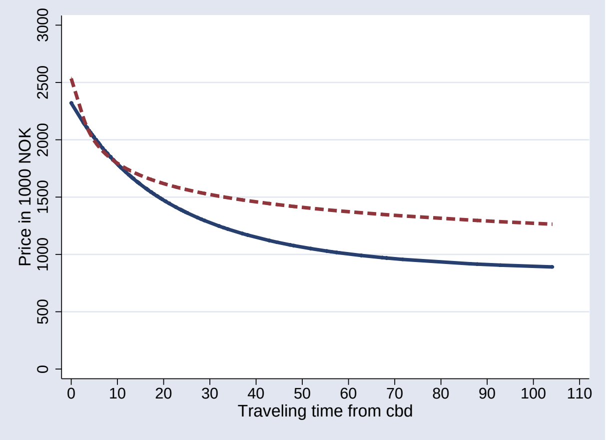


Figure 3: Predicted housing price gradients for a fictional standard house. Estimated parameters are based on data from Stavanger. The dashed curve reflects an exponential relationship between housing price and traveling time, the solid curve is based on the power function.

Housing price gradients can be stated in terms of physical distance rather than traveling time. In such cases the graph will not in general be autonomous to changes in road transportation infrastructure. An innovation in the road network might for instance result in reduced traveling time even if the physical distance is unaffected. This can be due to higher speed limits and/or improved road standard. In such a case a housing price gradient expressed in terms of physical distance will shift to the right, and the slope will be reduced. With traveling time represented on the horizontal axis predictions follow directly from movements along the curve.

6.2 More complex specifications of the relationship between housing prices and traveling time

It follows from the discussion in the previous subsection that an empirically based evaluation of distribution and capitalization effects of investments in road transportation infrastructure is rather sensitive to the functional specification of spatial separation in the model. There are of course many specifications alternative to the exponential and the power function. One alternative is a logistic function, where traveling time enters through the following expression:

$$f(d_{ij}) = \frac{\beta_0}{1 + e^{\beta_1 + \beta_2 \cdot d_{ij}}}$$

Another alternative is the conventional Box-Cox transformation. This transformation is expressed through the specification of the variable Z_{ij}, defined by:

$$Z_{ij} = \frac{d_{ij}^{\lambda} - 1}{\lambda}$$

In Table 5 we let β_z represent the marginal impact of changes in Z_{ij} on housing prices. In principle a procedure where λ is estimated allows for a more flexible way to take the effect of traveling time into account. It is not a priori obvious that spatial separation should be represented by either d_{ij} or $\ln d_{ij}$ in the relevant relationship. As in the previous subsection we assume a log-linear relationship between the dependent and the remaining independent variables. In the estimation intrazonal distances of zero are substituted by arbitrarily small numbers.

To summarize, we consider the following alternative model specifications:

M10: traveling time is represented through a logistic function

M11: traveling time enters through a Box-Cox transformation

Results are presented in Table 5. All results in this table are based on data from the entire model. Compare first M10 to M4b. It then follows that the model formulation based on a logistic function adds significantly to the explanatory power, since the value of the likelihood ratio test statistic is approximately 9, which exceeds the critical value of a chi square distribution with two degrees of freedom. It also follows from Table 5 that the Box-Cox specification adds even more to explanatory power, with only one parameter more than the exponential distance deterrence function in M4b.

This means that the explanatory power is significantly improved as a result of the conventional Box-Cox transformation. One objection against this approach concerns the interpretation of the transformation parameter λ. In a case with heteroskedasticity it is well known (see for instance Kmenta (1983)) that the estimate of this parameter will almost certainly be less than 1, it is biased towards zero. The Box-Cox transformation adjusts both for heteroskedasticity and for an incorrect functional form. Low values of λ reduce heteroskedasticity in residuals. Hence, we cannot be sure that our parameter estimate defines an appropriate functional form of spatial separation in the model. This also means that the resulting housing price gradient is biased.

Rather than a Box-Cox transformation Kmenta (1983) recommends a spline-function approach to test for nonlinearity. Such an approach has been used for instance by Dubin and Sung (1987), for the estimation of housing rent gradients in non-monocentric cities. We assume a function that is piecewise log-linear, and introduce two knots, defining three segments of the housing price gradient. At the outset a component-plus-residual plot (Ezekiel 1924, Larsen and McCleary 1972) was used to detect nonlinearity. The locations of the knots were then determined through a search procedure, identifying the values that were maximizing the explanatory power of the regression model. We let the spline function with two knots (d_{ij}^1 and d_{ij}^2) be represented by the following specification of the relevant function:

$$g_0(d_{ij}) = d_{ij}^{\beta_0 + \delta_{ij}^{(1)}\beta_1 + \delta_{ij}^{(2)}\beta_2}$$

Here, $\delta_{ij}^{(1)}$ and $\delta_{ij}^{(2)}$ are Kronecker deltas, defined by:

$$\delta_{ij}^{(k)} = \begin{cases} 1 & \text{if } \ d_{ij} > d_{ij}^k \\ 0 & \text{otherwise} \end{cases} \qquad k = 1, 2$$

With such a parametric specification β_1 and β_2 can be interpreted as discontinuous corrections in the effects of variations in traveling time by moving from one segment to the next, and the specification refers to model M12 in Table 5:

M12: traveling time is represented by a piecewise log-linear spline function with two knots

Since this spline function enters into a log-linear relationship, the elasticity of housing price with respect to distance is constant within each of the three segments of the gradient. According to the results in Table 5 there is a significant discontinuous change in housing prices at a distance corresponding to 20 minutes of traveling time from the cbd. The relevant elasticity increases from -0,1649 to -0,3527 when traveling time exceeds 20 minutes. A natural hypothesis is that this reflects a discontinuous change in commuting behavior at such distances. The second knot that is reported in Table 5 appears for a traveling time of 55 minutes from the cbd. The segment represented by traveling times exceeding 55 minutes has an elasticity of -0,1539. The parameter related to this knot is not, however, significant at the 5% level.

According to the log-likelihood ratio test this spline function model specification fits data significantly better than the Box-Cox modeling approach. The value of the likelihood ratio test statistic is 7.3, which exceeds the critical value at the 5 % significant level.

We have also tested a model specification where the distance function entering into the hedonic regression model is assumed to be piecewise linear:

M13: traveling time is represented by a piecewise linear spline function with three knots

This model specification does of course not imply that the housing price gradient is piecewise linear, since the distance function appears in an equation where the dependent variable appears through its natural logarithm. Based on the component-plus-residual plot we found it natural to specify three knots (d_{ij}^1, d_{ij}^2, and d_{ij}^3) at the outset. The distance deterrence relationship is then represented by the following expression in the model:

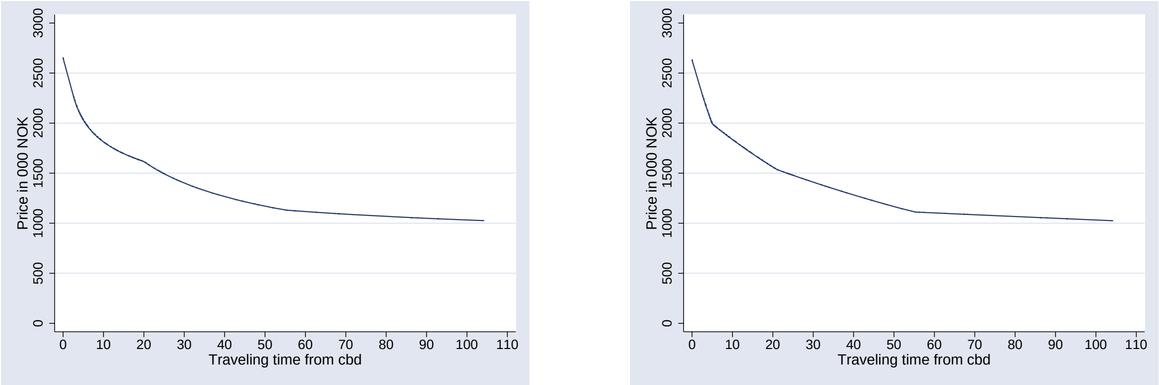
$$g_1(d_{ij}) = (\beta_0 + \delta_{ij}^{(1)}\beta_1 + \delta_{ij}^{(2)}\beta_2 + \delta_{ij}^{(3)}\beta_3)d_{ij}$$

Once again:

$$\delta_{ij}^{(k)} = \begin{cases} 1 & \text{if } \ d_{ij} > d_{ij}^k \\ 0 & \text{otherwise} \end{cases} \qquad k = 1, 2, 3$$

This estimation resulted in the following three knots: $\delta_{ij}^{(1)} = 5$, $\delta_{ij}^{(2)} = 21$, and $\delta_{ij}^{(3)} = 55$.

The two versions of the spline function are illustrated in Figure 4. In both cases we predict only marginal reductions in housing prices when traveling time increases beyond 55 minutes. For all practical purposes the two curves are more or less totally overlapping. According to log-likelihood values reported in Table 5 the approach where traveling time appears through a piecewise linear function results in a significantly higher explanatory power than all the alternative model specifications.



M12 M13

Figure 4: Predicted housing price gradients for a fictional standard house. The left part of the figure represents model M12, the right part is based on model M13.

Objections can be raised against the data-driven spline function approach. One such objection is that very close neighbors on either side of a knot is implicitly assigned distinctly different distance responsiveness in commuting demand. Another kind of arbitrariness concerns the number of knots. In general explanatory power is positively related to the number of knots. At the limit a specification can be chosen where workers with the same number of minutes traveling time from the cbd are assigned a specific distance responsiveness. Such an approach probably fits well to the observations, but it does definitely not represent a satisfying general hypothesis of commuting behavior. Hence, it also at best represents a questionable basis for predicting effects of changes in for example road transportation infrastructure.

In the previous subsection we concluded that a model specification based on an exponential representation of traveling time is superior to an approach based on a power function. One reason for this was that the power function results in a housing price gradient where housing prices are unreasonably sensitive to variations in short distances from the cbd. This tentative conclusion was supported by the results following from the spline function approach considered above. Hence, both intuition and numerical results indicate that the assumption of a globally constant elasticity of housing prices with respect to distance does not provide a satisfying explanation of our observations. As an alternative to the spline function approach this can be adjusted for by introducing a quadratic term in the regression equation. Traveling time then appears in the regression equation through the following expression:

$$h(d_{ij}) = d_{ij}^{\beta} \cdot ((d_{ij})^2)^{\beta_q}$$

This corresponds to model specification M14 in Table 5:

M14: the power function is supplemented by a quadratic term

With such a specification the elasticity of housing price with respect to distance is:

$$El_{d_{ij}} P_i = El_{d_{ij}} h(d_{ij}) = \beta + 2\beta_q \ln d_{ij}$$

Notice first from Table 5 that both parameters related to distance are significantly negative; $\hat{\beta} = -0,0689$ and $\hat{\beta}_q = -0,0295$. This means that housing prices become increasingly more elastic with respect to distance for movements downwards along the housing price gradient. The point elasticity is $-0,0689$ when the traveling time is 1 minute from the cbd, while it is for instance $-0,2048$ and $-0,2997$ for locations with respectively 10 and 50 minutes of traveling time to the cbd.

Compared to the pure power function specification (M7b) the quadratic term adds significantly to the explanatory power. Model M14 also fits data significantly better than the approach based on an exponential distance deterrence function, M4b; the value of the relevant likelihood ratio test statistic is about 14. Evaluated from the measures of model performance reported in Table 5 it is not possible to distinguish significantly between the quadratic approach and the Box-Cox approach.

We have seen that the alternative, somewhat more complex, functions contribute significantly to the explanatory power compared to a simple exponential function. For most practical purposes, however, this difference in explanatory power probably does not matter very much. From a practical point of view criteria related to predictability of housing price gradients represent a more important basis for the choice of a model specification. Such criteria are difficult to formulate in an empirical context. Physically more or less identical houses probably can be found at different locations, but prices are also influenced by variation in local aspects that we have not incorporated into the model. Hence, the choice of a model specification has to be based on more tentative considerations, and/or sensitivity analysis.

Figure 5 offers predictions of housing price gradients for alternative functional specifications of a the relationship between housing prices and traveling time. The curves once again refer to the fictional, average, house that has been considered above. The logistic function, the Box-Cox approach, and the model specification incorporating a quadratic term generate housing price gradients in an intermediate position to the exponential and the power function. The predicted profiles are rather similar for those three alternatives.

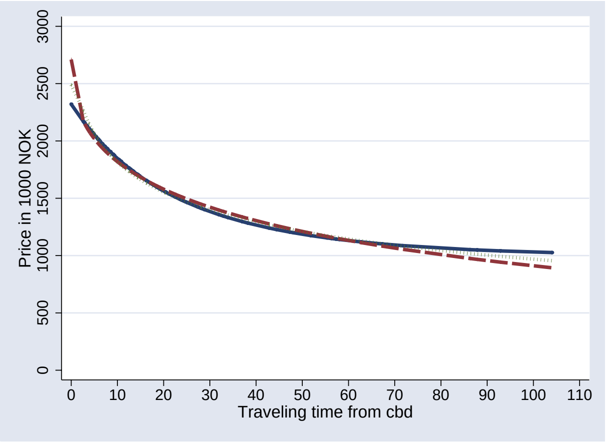


Figure 5: Predicted housing price gradients for a fictional standard house. The dotted curve is based on model M10, the dashed curve is based on model M11, while the solid curve is based on model M14.

As mentioned in Section 4 spatial autocorrelation can be explained by omitted variables or a poor functional representation of the relevant problem. In Section 5 we have seen that spatial autocorrelation is considerably reduced when traveling time from the cbd is explicitly accounted for in the model formulation. By comparing the results in Table 4 and Table 5 it follows that spatial autocorrelation is additionally reduced when distance appears in a more appropriate function. In fact, the null hypothesis of no spatial autocorrelation cannot be rejected in models M11, M12, M13, and M14.

We will not enter into a thorough discussion of what is the preferred specification for predicting housing price changes of investments in the road infrastructure network. In general, however, it is an argument in such a discussion to keep things simple. In addition, theoretically based objections can be raised against the alternatives that perform best according to explanatory power (the spline function approach). Given such considerations the results of our experiments indicate that not much is gained by introducing more complex functional specifications of the relationship between housing prices and traveling time from the cbd. Such a conclusion should be interpreted with care, however, since increased flexibility pays off considerably in some situations. Assume for instance that a housing price gradient for some reason is defined from a starting point in a lower rank central place than the dominating cbd in the region. Figure 6 illustrates three alternative housing price gradients based on the traveling time from the center of Sandnes (see Figure 1). All gradients refer to the same standard fictional house that has been considered above. The dashed curve represents the predicted housing price gradient in a case where traveling time is represented by a simple negative exponential function in the model to be estimated, while the dotted curve is based on a power function specification of the relevant relationship. Notice that the slope of those gradients is reduced compared to the corresponding gradients originating from the cbd in the region (see Figure 2). Still, both gradients predict housing prices to fall monotonically as traveling time from Sandnes increases. As indicated by the gradient based on the more flexible model M14 this is not a reasonable conclusion. The more flexible function captures the fact that Sandnes is not the spot in the region that is associated with the highest willingness-to-pay for housing. Without entering into further details on the local geography and housing market, there is no doubt that the corresponding non-monotonic housing price gradient in Figure 6 is a lot more reasonable than the monotonic counterparts. Notice also that the gradient based on an exponential function fits better to the more reasonable path than the gradient based on a power function.

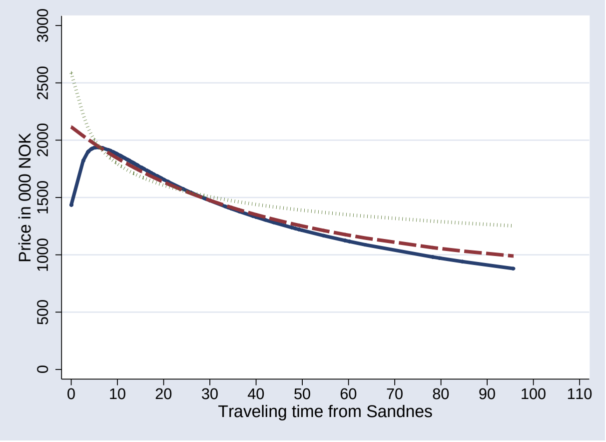


Figure 6: Predicted housing price gradients based on traveling time from Sandnes. The dashed curve represent an exponential spatial separation function, while the dotted curve is based on a power function specification. The non-monotonic curve is based on model M14 (quadratic).

6.3 An approach based on gradients that may vary with direction from the cbd

Several authors have pinpointed that housing price gradients may vary with direction from the cbd, see for instance Dubin and Sung (1987) or Plaut and Plaut (1998). One approach to test for this possibility is to introduce dummy variables for houses located south and north of Stavanger. The dummy variable representing the north was multiplied by traveling time from the cbd. The north-dummy and the interaction variable were included in a regression based on model M11.

Table 5: Results based on alternative specifications of spatial separation and spatial structure

	M14	M15	M16	M17	M18
Constant	11,0262	11,9378	12,0408	12,0342	11,9275
	(0,0878)	(0,0878)	(0,0881)	(0,0939)	(0,0810)
LOTSIZE	0,1203	0,1172	0,1224	0,1225	0,1262
	(0,0103)	(0,0098)	(0,0102)	(0,0101)	(0,0085)
RURLOT	-0,0303	-0,0303	-0,0326	-0,0311	-0,0300
	(0,0032)	(0,0032)	(0,0031)	(0,0032)	(0,0026)
AGE	- 0,0784	- 0,0805	- 0,0826	-0,0824	- 0,0828
	(0,0064)	(0,0063)	(0,0065)	(0,0066)	(0,0048)
AGE·REBUILD	0,0105	0,0109	0,017	0,0107	0,0106
	(0,0029)	(0,0027)	(0,0029)	(0,0029)	(0,0027)
GARAGE	0,0661	0,0654	0,0677	0,0679	0,0670
	(0,0109)	(0,0110)	(0,0110)	(0,0110)	(0,0101)
LIVAREA	0,3593	0,3538	0,3517	0,3514	0,3573
	(0,0178)	(0,0178)	(0,0178)	(0,0178)	(0,0157)
NUMBTOIL	0,1495	0,1546	0,1522	0,1518	0,1522
	(0,0147)	(0,0147)	(0,0146)	(0,0147)	(0,0137)
β_0 (logistic)	3820,666	-	-	-	-
	(8,5879)	(-)	(-)	(-)	(-)
β_1 (logistic)	8,4126	-	-	-	-
	(0,0399)	(-)	(-)	(-)	(-)
β_2 (logistic)	0,0313	-	-	-	-
	(0,0034)	(-)	(-)	(-)	(-)
β_z (Box-Cox)	-	- 0,0641	-	-	
	(-)	(0,0017)	(-)	(-)	(-)
λ (Box-Cox)	-	0,4313	-	-	-
	(-)	(0,0251)	(-)	(-)	
β_0 (spline, power)	-	-	-0,1649	-	
	(-)	(-)	(0,0086)	(-)	
β_1 (spline, power)	-	-	-0,1878	-	
	(-)	(-)	(0,0263)	(-)	(-)
β_2 (spline, power)	-	-	0,1988	-	-
	(-)	(-)	(0,1278)	(-)	(-)
β_0 (spline, linear)	-	-	-	-0,0557	-
	(-)	(-)	(-)	(0,0060)	(-)
β_1 (spline, linear)	-	-	-	0,0395	-
	(-)	(-)	(-)	(0,0064)	(-)
β_2 (spline, linear)	-	-	-	0,0067	-
	(-)	(-)	(-)	(0,0016)	(-)
β_3 (spline, linear)	-	-	-	0,0078	-
	(-)	(-)	(-)	(0,0025)	(-)
β (quadratic)	-	-	-	-	- 0,0689
	(-)	(-)	(-)	(-)	(0,0202)
β_q (quadratic)	-	-	-	-	- 0,0295
	(-)	(-)	(-)	(-)	(0,0038)
YEARDUM97	-0,1336	-0,1346	-0,1350	-0,1348	- 0,1333
	(0,0135)	(0,0135)	(0,0135)	(0,0135)	(0,0128)
YEARDUM99	0,1275	0,1288	0,1289	0,1288	0,1295
	(0,0137)	(0,0137)	(0,0137)	(0,0137)	(0,0134)
YEARDUM00	0,2684	0,2672	0,2684	0,2686	0,2686
	(0,0135)	(0,0136)	(0,0135)	(0,0135)	(0,0130)
YEARDUM01	0,3031	0,3024	0,3045	0,3044	0,3041
	(0,0136)	(0,0137)	(0,0137)	(0,0136)	(0,0139)
n	2788	2788	2788	2788	2788
R^2	0,7368	0,7378	0,7385	0,7396	0,7376
R^2-adj.	0,7356	0,7367	0,7372	0,7382	0,7364
L	275,4067	280,78	284,4316	290,4447	279,68
APE	217020,33	216901,08	216921,28	216222,25	216736,08
White test statistic	252	255	288	306	264
Moran's I	0,0036	0,0019	0,0012	0,0010	0,0015
Standard normal deviate (z_I)	2,3151	1,4027	1,1896	1,1255	1,3068
Ramsey reset test (p-value)	0,4241	0,7860	0,9280	0,8853	0,8274

Note: Results based on observations from the period 1997-2001, robust standard errors in parentheses.

The number of observations is 118 in the north, and 2670 in the south; the region only ranges over 24 minutes in the northern direction. The two new coefficients got robust t-values of about 2,3. A Wald test was performed, testing the null hypothesis of equality between the general coefficient representing traveling time, and the coefficient representing traveling times toward the north. This null hypothesis could not be rejected. Hence, the price distance gradient is not considered to be statistically different in the north and in the south of the cbd. Once again, this is an indication that the area we consider is fairly homogenous.

7 Concluding remarks

In this paper we have studied the impact of spatial and non-spatial variables on housing prices in the southernmost region in Western Norway. We have primarily focused on the impact of spatial separation on housing prices, and on the form of housing price gradients corresponding to a specific set of other attributes. The region has one dominating center (Stavanger), and the diffusion of new residential areas has to a large degree been determined by employment growth in, and close to, this center. This is one reason why the region is very appropriate for the purpose of identifying reliable housing price gradients, reflecting to a relatively large degree housing market equilibrium forces rather than characteristics of more complex multicentric and multinodal systems.

We started out by considering non-spatial modeling alternatives, primarily representing a benchmark for evaluating more satisfying model specifications. The introduction of spatial separation measures results in more precise parameter estimates and improved explanatory power, especially in a case where the estimation is based on data from the entire region. We have evaluated alternative functional specifications of spatial separation in the hedonic model formulation. Our main findings are:

- Our results indicate that predicted housing price gradients are in general not reliable if estimation is based on observations covering only a part of the relevant labor and housing market originating from the cbd. This represents one possible reason why some studies report intuitively strange gradients.

- According to our results a power function specification of traveling time is not an appropriate approach to predict housing price gradients. A log-linear regression model results in biased gradients, and tends to over-predict housing prices in locations close to the cbd. The exponential function performs better, and results in more reliable gradients in the case where estimation is based on observations from the entire labor/housing market. Predicted housing prices might differ substantially even if values of estimated parameters and explanatory power are similar.

- The use of more complex and flexible functional specifications of traveling time contributes significantly to the explanatory power compared to a one-parameter approach. For most practical purposes the difference in explanatory power does not matter very much. The more flexible approaches lead to housing price gradients in an intermediate position between the one-parameter exponential and power function approaches, and they probably represent a more reliable basis for predictions.

- Our results do not distinguish clearly between the alternative flexible function approaches that we have considered. Results on explanatory power should be considered in combination with pragmatic, theoretical and interpretational arguments. Based on such a consideration we especially find the approach incorporating a quadratic term appealing.

In this kind of empirical research it is important to consider potential econometric problems. In particular one such potential problem is related to spatial autocorrelation. We find severe spatial autocorrelation in models where no measure of spatial separation is accounted for. A large part of this autocorrelation is removed when traveling time is introduced through a one-parameter function. Autocorrelation is further reduced in the model formulations based on more flexible functions. In the more flexible approaches we find that the hypothesis of no spatial autocorrelation cannot be rejected. Our results also indicate that increased functional flexibility pays off in terms of more reliable predictions of housing price gradients if the geography is more multicentric and/or multinodal than the one we consider, with less obvious identification of a regional center.

All in all we achieve encouraging results, with satisfying goodness-to-fit, reliable coefficient estimates, and intuitively reasonable predictions of housing price gradients. Our results represent important input in an evaluation of for instance residential construction programs, urban renewal, and/or investments in transportation infrastructure. In addition our results contribute to a discussion of how forces relating to the housing market can be incorporated into a general spatial equilibrium framework constructed for a region with one dominating center.

Appendix A

Table 6: Zonal data

Zone	Working population	Jobs	Observations
Rennesøy			
1	725	552	16
2	98	24	4
3	354	145	5
4	127	23	4
Randaberg			
5	3748	2195	89
Stavanger			
6	328	4961	12
7	95	4058	1
8	769	1736	11
9	688	1586	36
10	1021	328	47
11	1177	1630	41
12	863	3905	23
13	1125	1398	21
14	555	2339	34
15	1274	2864	41
16	1382	396	26
17	1518	4695	8
18	1151	2141	29
19	1750	407	47
20	1637	392	16
21	1777	1751	102
22	2367	1627	40
23	1340	627	45
24	959	226	33
25	846	271	16
26	1042	341	27
27	1001	132	23
28	997	254	46
29	1662	239	42
30	945	1746	29
31	1212	630	28
32	2436	11309	10
33	1719	529	44
34	760	930	24
35	240	583	4
36	999	101	35
37	919	147	28
38	284	14	14
39	1106	338	16
40	1169	110	22
41	4674	968	135
42	237	37	13
43	92	11	1
Sola			
44	893	83	34
45	2925	6178	70
46	945	115	34
47	497	63	22
48	514	131	11
49	2681	5423	74
Sandnes			
50	1215	4870	22
51	1338	1506	43
52	1090	218	16
53	371	147	8
54	1383	240	57
55	1150	302	40
56	543	214	4
57	788	6151	25
58	1592	570	55
59	651	1515	10
60	678	207	19
61	1280	175	10
62	1911	307	53
63	966	1355	23
64	824	537	21
65	737	276	6
66	1010	787	22
67	979	380	21
68	914	49	10
69	960	574	25
70	1198	477	23
71	942	253	13
72	668	240	24
73	21	3	3
Klepp			
74	429	158	5
75	3034	2043	72
76	1047	1502	16
77	340	208	2
78	1457	457	10
Gjesdal			
79	3354	1760	129
80	336	184	16
81	362	353	1
Time			
82	5148	4343	93
83	383	123	5
84	1457	457	27
Hå			
85	1493	1106	35
86	1021	525	12
87	348	81	6
88	376	289	10
89	2795	2511	62
Bjerkreim			
90	395	213	8
91	540	511	8
Eigersund			
92	4612	4830	148
93	367	97	7
94	342	106	1
Lund			
95	742	920	10
96	235	45	2
97	152	53	1
Sokndal			
98	1125	916	21
99	17	1	3

Appendix B

Table 7: Descriptive housing market statistics

	OBSERVATIONS	MEAN	STD. DEV.	MINIMUM	MAXIMUM
STAVANGER					
REALPRICE	1188	1751240	662891	160000	4747500
Price pr square meter	1192	10902	3335	1778	34451
LIVAREA	1188	167	61	43	500
LOTSIZE	1188	512	334	40	5243
GARAGE	1188	0,68	0,47	0	1
NUMBTOIL	1188	2	0,84	1	6
AGE	1188	41	37	0	187
REBUILD	1188	0,42	0,50	0	1
TIMECBD	1188	6,58	3,2	0	25,8
RANDABERG, SOLA, AND SANDNES					
REALPRICE	863	1582665	518372	353425	4401000
Price pr square meter	863	9386	3040	1767	24265
LIVAREA	863	178	60	47	387
LOTSIZE	863	679	346	80	4070
GARAGE	863	0,79	0,41	0	1
NUMBTOIL	863	2	0,81	1	6
AGE	863	22	18	0	116
REBUILD	863	0,33	0.47	0	1
TIMECBD	863	16,1	6,2	6,2	57,3
MUNICIPALITIES IN THE REST OF THE REGION					
REALPRICE	737	1100264	419500	204600	3892950
Price pr square meter	737	7631	2696	897	22827
LIVAREA	737	151	54	37	361
LOTSIZE	737	728	326	37	2487
GARAGE	737	0,74	0,44	0	1
NUMBTOIL	737	2	0,72	0	6
AGE	737	26	26	0	163
REBUILD	737	0,31	0,46	0	1
TIMECBD	737	48	20	21	220
THE ENTIRE REGION					
REALPRICE	2788	1526976	622323	160000	4747500
Price pr square meter	2788	9567	3358	896	33451
LIVAREA	2788	166	60	37	500
LOTSIZE	2788	621	349	37	5243
GARAGE	2788	0,73	0,44	0	1
NUMBTOIL	2788	2	0,81	0	6
AGE	2788	31	30	0	220
REBUILD	2788	0,36	0,48	0	1
TIMECBD	2788	19	18	0	104

Note: Prices are measured in NOK, they have been adjusted for inflation, and 1998 represents the base year. LIVAREA and LOTSIZE are measured in square metres, and is measured TIMECBD in minutes.

References

[1] Adair A, S McGreal, A Smyth, J Cooper, and T Ryley, 2000, “House prices and accessibility: the testing of Relationships within the Belfast urban area”, *Housing studies*, **15** 699-716.

[2] Alonso W, 1964, *Location and land use. Toward a general theory of land*, Harvard University Press, Cambridge, Massachusetts.

[3] Anselin L, 1988, *Spatial econometrics: methods and models.* Kluwer Academic Publishers, London.

[4] Anselin L and Rey S, 1991, “Properties of tests for spatial dependence in linear regression models”, *Geographical Analysis* **23**, 113-131.

[5] Ball M J, 1973, ”Recent empirical work on the determinants of relative house prices”. *Urban Studies* **10** 213-233.

[6] Ball M J and R M Kirwan, 1977, “Accessibility and supply constraints in the urban housing market”, *Urban studies* **14** 11-32.

[7] Barkley, D L, Henry M S, Bao S, Brooks K R, 1995, “How functional are economic areas? Tests for intra-regional spatial association using spatial data analysis,” *Papers in Regional Science* **74** 297-316.

[8] Bartik T J and V K Smith, 1987, *Urban Amenities and public policy*, Handbook of Regional and Urban Economics, vol 2, Elsevier science publisher B.V, Netherlands.

[9] Brueckner J, 1987, *The structure of urban equilibria: a unified treatment of the Muth-Mills model*, Handbook of Regional and Urban Economics, vol 2, Elsevier science publisher B.V, Netherlands.

[10] Davidson R and J G MacKinnon, 1993, *Estimation and Inference in Econometrics*, Oxford University Press. New York.

[11] Dubin, R A and C H Sung, 1987, ”Spatial variation in the price of housing rent gradients in non-monocentric cities”, *Urban Studies* **24**, 193-204.

[12] Dubin R A, 1992: “Spatial autocorrelation and neighbourhood quality”, *Regional Science and Urban Economics* **22** 433-452.

[13] Ezekiel, M, 1924, ”A method of handling curvilinear correlation for any number of variables” *Journal of the American Statistical Association* **19**, 431-453.

[14] Fik, T J and G F Mulligan, 1998, “Functional form and spatial interaction models”, *Environment and Planning*, volume 30, 1497-1507.

[15] Florax R J G M, Voortman R L and Brouwer J, 2002, “Spatial dimensions of precision agriculture: a spatial econometric analysis of millet yield on Sahelian coversands”, *Agricultural Economics*, **27**, 425-443.

[16] Greene W H, 2003, *Econometric Analysis*, Fifth edition, Prentice Hall.

[17] Hansen W G, 1959, “How accessibility shapes land use”, *Journal of the American Institute of Planners* **25** 73-76.

[18] Heikkela E, P Gordon, J I Kim, R B Peiser, H W Richardson, 1989, “What happened to the CBD-distance gradient?: land values in a polycentric city”, *Environment and Planning A* **21** 221-232.

[19] Kmenta J, 1986, *Elements of Econometrics*, Macmillan, New York.

[20] Laakso S, 1997, “Urban housing prices and the demand for housing characteristics. A study on housing prices and the willingness to pay for housing characteristics and local public goods in the Helsinki Metropolitan Area”, The research Institute of the Finnish Economy, Helsinki.

[21] Larsen, W A and S J McCleary, 1972 ”The use of partial residual plots in regression analysis”, *Technometrics* **14**, 781-790.

[22] Mills E, 1967, ”An aggregated model of resource allocation in a metropolitan area”, *American Economic Review* **57** 197-210.

[23] Muth R F, 1969, *Cities and housing*, University of Chicago Press.

[24] Nijkamp, P. and A. Reggiani. 1992. *Interaction, evolution and chaos in space,* Springer-Verlag

[25] Odland, J, 1988, *Spatial autocorrelation*, Sage Publicationd Inc.

[26] Plaut P O and S E Plaut, 1998, “Endogenous identification of multiple housing price centers in metropolitan areas”, *Journal of Housing Economics* **7** 193-217.

[27] Rosen S, 1974, “Hedonic prices and implicit markets: product differentiation in pure competition”, *Journal of Political Economy* **82** 34-55.

[28] Rothenberg J, G C Galster, R V Butler, and J Pitkin, (1991), *The maze of urban markets. Theory, evidence and policy*, University of Chicago Press.

[29] Sheppard S, 1999, “Hedonic analysis of housing markets”, *Handbook of Regional and Urban Economics*, ed. by E S Mills and P Chesire.

[30] Strazheim M R, 1974, ”Hedonic estimation of housing market prices. A further comment” *The Review of Economics and Statistics* **56** 404-406.

[31] Thorsen I, Gitlesen J P, 1998, “Empirical evaluation of alternative model specifications to predict commuting flows”, *Journal of Regional Science*, **38** 273-292

[32] Waddell P, Berry B J L, and I Hock, 1993, “Residential Property Values in a Multinodal Urban Area: New Evidence of the Implicit Price of Location” *Journal of Real Estate Finance and Economics* **7** 117-141.

[33] Whitehead C M E, 1999, “Urban housing markets: theory and policy”, *Handbook of Regional and Urban Economics*, ed. by E S Mills and P Chesire.

Department of Economics
University of Bergen
Fosswinckels gate 6
N-5007 Bergen, Norway
Phone: +47 55 58 92 00
Telefax: +47 55 58 92 10
http://www.svf.uib.no/econ