Housing Price Gradients in a Region with One Dominating Center

Authors Liv Osland, Inge Thorsen, and
Jens Petter Gitlesen

Abstract This paper primarily focuses on predicting housing price gradients in a Norwegian region with one dominating center. Spatial separation is represented by a function of the traveling distance from the city center in a traditional hedonic regression equation. Several functions are tested, and some alternatives provide a satisfying goodness-of-fit, consistent coefficient estimates, and intuitively reasonable predictions of housing price gradients. Still, not all commonly used functions are recommended. The findings also indicate that the strength of spatial autocorrelation is reduced when the hedonic function is properly specified.

The main ambition of this study is to estimate a housing price gradient for a region in the southern part of Western Norway. This region has a dominating city (Stavanger), and the study also tests for the appropriateness of the monocentric city model in this kind of area. The basic idea underlying this model is represented by a steadily declining unit price for houses with an increasing distance from the central business district (CBD). For a presentation of the modeling framework, comparative static results, and interesting extensions, see Anas, Arnott, and Small (1998).

Many empirical studies have aimed at finding rent gradients, land value gradients, and/or housing price gradients. In a few studies, the variable indicating the access to work came out with an insignificant sign, and occasionally a counter-intuitive sign was reported [see for instance Bartik and Smith (1987) for a review]. Such results can, for example, be explained by the fact that the area under study in some cases involves a restricted urban area rather than a housing market area. Another reason for such results is that modern metropolitan areas tend to be multicentric. Both Richardson (1988) and Heikkila, Gordon, Kim, Peiser, and Richardson (1989) state that the main reason for insignificant or counter-intuitive results stems from a misspecified hedonic price function. This is demonstrated in Waddell, Berry, and Hock (1993), who find that the impact of distance to the CBD is significant even when access to multiple employment centers and other nodes are accounted for. Adair, McGreal, Smyth, Cooper, and Ryley (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. According to McMillen (2004), however, the basic insights of the model also apply to complex polycentric cities, and he claims that the decline in the explanatory power of the model is a misunderstanding of the empirical evidence.

Most empirical studies on spatial variation in house prices consider complex metropolitan areas around large cities. Stavanger is the fourth largest city in Norway, but the population in the Stavanger municipality is only approximately 115,000. Still, the city is very dominating in the region, the structure is relatively monocentric, and the area is appropriate for studying housing price gradients. Alternative functional specifications of the relationship between the housing price and the distance from the CBD are considered here. Since housing price predictions are rather sensitive to the choice of functional form, it is important to find the form that best fits the data in the relevant type of area.

Housing price gradients represent an 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 and capitalization through property values. The findings, for instance, in studies related to constructing new roads (tunnels/bridges), changing speed limits or in analyzing investments inducing increased capacity and reduced queues on existing links. In some cases, the only available information on spatial characteristics might be the distance from the CBD. The prospects for reliable predictions then of course critically depend on the statistical robustness of the estimated relationship. This is the motivation for thoroughly examining statistical properties of alternative mathematical representations of the relationship between housing prices and the distance from the CBD. Therefore, the hedonic method used here includes several non-spatial attributes related to housing along with the distance from the CBD. For a review of hedonic analysis on housing markets, see Sheppard (1999).

Data on spatial labor market interaction clearly indicates that the study area can be considered as a coherent housing and labor market area, suitable for studies of the relationship between housing consumption and commuting costs. In addition, the specification of a housing market calls for delimitation with respect to the relevant types of dwellings. The housing market in the prosperous study area is fairly homogenous. The focus is on privately-owned, single-family houses. In many cities the supply of this housing type is relatively scarce. 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, such as terraced houses, is markedly underrepresented in more peripheral, rural areas of the region. Hence, such categories are less appropriate for studying spatial variations in housing prices in an area extending beyond the city center.

The zonal subdivision of the geography corresponds to the most detailed spatial level for which sufficient official Norwegian data is available. There is little additional insight and explanatory power that would result from a more disaggregated representation of the geography, which would require enormous effort and resources to collect such data, if at all practically possible. For some spatially-related attributes, such as the scenic view, or the distance to a nursery school, a relatively high degree of interzonal homogeneity can be expected. Many of these location-specific attributes are reasonably equally present in most of the (postal delivery) zones that are examined. No attempt is made to explicitly account for the possible impact of local anomalies and alternative characteristics of the spatial structure.

The next section provides documentation of our data. A description of the region is also given, and shown to be very appropriate for the study. Some general aspects related to the model formulation will be discussed in the third section. After this there will be a presentation of results, and finally, some concluding remarks are given.

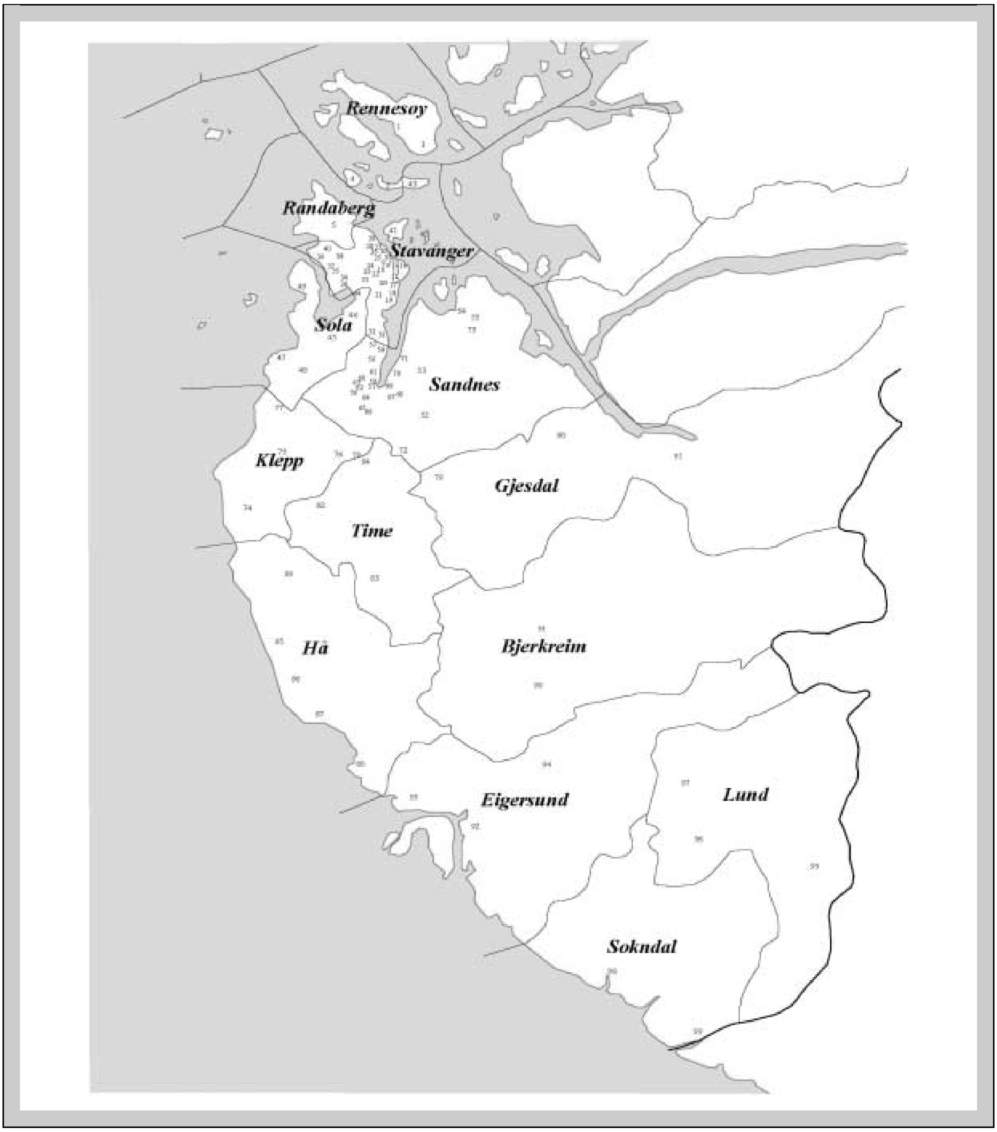
The Region and the Data

The study area is the southern part of Rogaland County in Western Norway. This represents an integrated region with a connected road transportation network. There are 13 municipalities, and each municipality is divided into postal delivery zones. Altogether, the region is divided into 98 zones, as indicated in Exhibit 1. As an indicator of (commuting) distances, there are 79 km between the center of Stavanger and the center of Eigersund in the south. The region is delimited by the North Sea in the west, and fjords in the north and the east, while the southern delimitation is an administrative county border in a sparsely populated, mountainous area. Hence, the demarcation of the region is mainly determined by natural boundaries. For further information on the region, see for instance Osland, Thorsen, and Gitlesen (2005).

The ambition to capture the trade-off between housing market prices and labor market interaction calls for a large geographic area of study. As a working hypothesis, unbiased estimates of this relationship cannot be based on a truncated specification of the market area. This study estimates the alternative model formulations for different subdivisions of the geography. The results are based both on the entire region and are based on data only from Stavanger (the CBD). In addition, Osland et al. (2005) present results for an extended urban area, represented by the four most centrally located municipalities in the northern part of the region.

Estimation results presented in this paper are based on housing market data for transactions of privately-owned, single-family houses in the period from 1997 through the first half of 2001. The sample consists of 2,788 transactions of privately owned, single-family houses in the region during the relevant period. The transaction data on the freeholder dwellings have been provided by two sources: the national land register in Norway and Statistics Norway. All the

Exhibit 1 | Division of the Region into Municipalities and Zones



hedonic regression results to be presented involve the same set of dwelling-specific attributes, which are defined in Exhibit 2.

In order to obtain a high number of observations, the data is drawn from several years of house sales. In order to account for increases in housing prices during the period, changing intercepts are introduced through dummy variables for each year (*YEARDUM*). The dummy for 1998 is excluded in order to avoid perfect

Exhibit 2 | List of Dwelling-Specific 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 the presence of a garage.
TOILETS	Number of toilets in the building.

multicollinearity. Finally, the product of the dummy variable representing the rural areas and the variable *LOTSIZE* is introduced. This product defines the variable *RURLOT* in forthcoming sections. The dummy variable takes the value 1 if the municipality is located in the most rural area, otherwise it is 0. Based on information concerning the ratio of inhabitants to open land, four municipalities in the southern part of the region are defined to represent the most rural areas in the region [see Osland et al. (2005) for details].

The information on the spatial distribution of jobs is based on the Employer-Employee register, and was provided by Statistics Norway. The matrices of Euclidean distances and traveling times were prepared by the Norwegian Mapping Authority, who have at their disposal all the required information on the road network and the spatial residential pattern. For further details on that data, descriptive housing market statistics for separate parts of the region, and on the region in general, see Osland et al. (2005).

General Aspects Related to the Model Formulation and Reported Statistics

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

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

Where:

P_{it} = The price of house i in year t;
z_{sit} = The value of dwelling-specific structural attribute s for house i in year t; $s = 1,...S$, $i = 1,...n$; and
z_{lit} = The value of location-specific attribute l for house i in year t; $l = 1,...L$, $i = 1,...n$.

A macroscopical perspective of the geography is adopted. This means that no attempt was made to explicitly account for the accessibility to recreational facilities and shopping opportunities, or the possible impact of variations in environmental conditions, location-specific amenities, and aesthetic attributes. The approach is implicitly based on the assumption that such (micro-locational) attributes are not varying systematically across the zones. In other words, it is implicitly assumed 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. The only location-specific attribute that is explicitly accounted for in this paper is the distance from the CBD.

Most housing studies in the literature are based on data from large metropolitan areas. In such studies spatial variation in property values in general also reflects local tax rates, the quality of local public services (such as local public education), the crime rate, and socioeconomic characteristics of neighbors. Tax rates are uniformly distributed over the area considered here, the crime rate is low, and the society is relatively egalitarian, at least in an international perspective. Still, some zonal variation is expected both in the provision of public services, the crime rate, and in the neighborhood composition; however, no attempt was made to explicitly account for such variation. The explanatory power reported in the subsequent section indicates that the potential for further improvements is limited, compared to the time resources required to collect sufficient information. Another relevant argument is that some of those variables are implicitly accounted for as endogenous variables in the reduced form. The socioeconomic composition of the population to some degree depends systematically on the distance from the CBD, which in turn influences for instance the quality of public services.

The study examined alternative mathematical representations of the general relationship between the dependent and the non-spatial independent variables. As in most other empirical studies of the housing market, the log-linear model formulations were found to be superior to linear model specifications. Hence, only results based on log-linear specifications of the relationship between housing prices and non-spatial attributes are reported. Given a power function specification of traveling time to the CBD (*TIMECBD*), this means that the hedonic regression model formulation is given by:

$$\begin{aligned}\log P_{it} = {} & \beta_0 + \beta_1 \log LOTSIZE_i + \beta_2 (RUR \log LOT)_i \\ & + \beta_3 \log AGE_i + \beta_4 (REBUILD \log AGE)_i + \beta_5 GARAGE_i \\ & + \beta_6 \log LIVAREA_i + \beta_7 \log TOILETS_i + \beta \log TIMECBD_i \\ & + \sum_{T=97}^{01} \beta_i YEARDUMt_i + \varepsilon_{it}, \end{aligned} \qquad (2)$$

where log(·) denotes the natural logarithm, and ε_{it} is the error of disturbance for a specific observation.

According to the reported values of the White test statistic, the hypothesis of homoscedasticity is rejected in all models $\chi^2_{0.05} = 16.92$. Still, the robust estimator of variance does not produce results that deviate much from estimates based on the ordinary least squares estimator.

In the computation of Moran's I values, a binary row standardized weight matrix is used to define the relationship between observations.[1] Zones that have common borders in the geography are neighbors. All houses within a zone are also neighbors. A house is not a neighbor to itself. Since geocoded data is not used, the resulting weight matrix represents a relatively crude measure of proximity. A more disaggregated study and modeling of spatial dependence among individual properties is beyond the scope of this paper and the data. A similar approach regarding the construction of the spatial weight matrix is found in Kim, Phipps, and Anselin (2003). As pointed out by Anselin (1988), the theoretical requirements are less stringent when the weight matrix is only used in a hypothesis test rather than in analyzing the structure of spatial dependence. When used for testing, the weight matrix should be related to the alternative hypothesis stating the spatial dependence of some unspecified kind. Values of the standard normal deviate (z_I) are also reported, which is constructed from values of the mean and the variance of the Moran's I statistic [see, for example, Anselin (1988) for details on the estimation of such values]. The null hypothesis of no spatial autocorrelation is rejected at the 5% significance level if $z_I > 1.645$.

In addition to R^2, values of two measures to evaluate the goodness of fit abilities of various model alternatives are reported. L is the log-likelihood value that is maximized in the estimation procedure. The Average Prediction Error (APE) is explicitly based on a comparison between the observed and the predicted housing prices; $APE = \Sigma_i(|\hat{P}_i - P_i|)/n$. Here, $\hat{P}_i$ is the predicted price of house i, while n is the observed number of houses. The dependent variable is the logarithm of the real housing price. Used for prediction purposes, the logarithm of price has to be transformed. According to Wooldridge (2003), there are several ways this can be done, but none of them are unbiased. This study uses the following transformation, which is consistent and relies on normality of the errors (Wooldridge, 2003):

$$\hat{P} = \exp(\ln\hat{P})\exp\frac{\hat{\sigma}^2}{2}, \tag{3}$$

where $\hat{\sigma}^2$ is an unbiased estimator of the residual variance. The dependent variable in the models has also been transformed by using an estimator that is robust to non-normal errors (Wooldridge, 2003). This yielded the same result for $\hat{P}$ as the method described above. APE has an obvious interpretation, but this measure is not appropriate for statistical testing or to discriminate between alternative model specifications.

In the forthcoming sections, positive log-likelihood values are reported. In general, in rare cases this might result in density functions with very small variances, allowing for density values exceeding 1.0. Such cases are typically found in problems where dependent variables are defined for a relatively small range of high values. In this study, the logarithm of housing prices defines a function that is very flat for the relevant range of values, with correspondingly low variance.

As an alternative to the approach chosen in this paper, the impact of variations in distance on housing prices can be represented by semi-parametric estimators. Semi-parametric approaches are less likely to produce edge effects, and in particular they are useful in studies with highly non-linear relationships (see McMillen and Thorsnes, 2000), as well as in studies focusing on local spatial variation in the dependent variables (Clapp, 2003). This paper does not focus on potential local anomalies. Given the relatively ideal study area, problems related to edge effects are probably of minor importance. In addition, the monocentric model implies a shape for the housing price function (see McMillen, 2004) that is adequately represented by a two-parameter specification. Since a region with one strongly dominating center is considered here, a parametric approach probably is appropriate. Further examination of this hypothesis is left for future research. In general terms, both McMillen and Thorsnes (2000) and Clapp (2003) reported that the substantial results of the parametric approach and different data-driven semi-parametric approaches are very similar.

Results

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

Results Based on Non-spatial Model Formulations

Least squares estimation results based on non-spatial model formulations are presented in the first two columns of Exhibit 3. M1 refers to a model specification where the estimation is based on data from the entire region, while M2 is based on data from Stavanger municipality. Since important information on location is omitted from those model formulations, the parameter estimates are biased. This is especially evident for the variable *LOTSIZE*. Other parameter estimates are not substantially different from estimates based on more reliable model specifications. This applies for instance to the variables representing age and renovation information. Let β_A denote the coefficient attached to the variable *AGE*, while β_{AR} is attached to *AGE·REBUILD*. A significantly positive estimate of β_{AR} means that the negative impact of *AGE* upon the housing price is reduced. It is intuitively reasonable, however, that $|\beta_A| > |\beta_{AR}|$, meaning that *AGE* has a negative influence on housing prices, even if a house has been rebuilt.

It follows from Exhibit 3 that the null hypothesis of no spatial autocorrelation clearly is rejected in all the non-spatial model specifications. The values of Moran's I indicate that the systematic pattern of spatial interdependencies across zones is positively related to how large a part of the region that is considered. On the other hand, the non-spatial independent variables tend to explain more of the variation in housing prices the smaller the part of the region that is considered. This justifies the hypothesis that information on general spatial characteristics such as distances contributes less to the explanation of housing price variations if the study is restricted, for instance, to a specific urban area.

Exhibit 4 plots the mean of the residuals for each of the 98 zones. The plot illustrates how the residuals more or less systematically tend to fall off with the traveling time from the CBD, and clearly indicates that this measure of spatial separation should be explicitly incorporated into an appropriate model formulation.

Results Based on Hedonic Model Formulations Where Traveling Time from the Labor Market Center is Represented by a One-Parameter Function

The ambition of estimating a gradient reflecting the trade-off between housing prices and commuting costs calls for the identification of the labor market center of the region. It is likely, but not obvious, that the zones in the city center of Stavanger represent the labor market center. In addition, there is a particularly high labor demand originating from an area hosting large industrial firms and administrative units related to petroleum activities. This industrial area is located in between Sandnes and Stavanger. Through experiments using a gravity-based accessibility measure, however, zone 10 within the Stavanger CBD (see the map

Exhibit 3 | Estimation Results Based on Alternative Model Specifications

	M1	M2	M3	M4	M5	M6	M7	M8
Constant	11.608	11.260	10.921	10.874	10.514	12.083	12.038	11.924
	(0.116)	(0.136)	(0.089)	(0.087)	(0.136)	(0.090)	(0.088)	(0.089)
LOTSIZE	−0.022	0.095	0.112	0.118	0.137	0.132	0.122	0.126
	(0.013)	(0.016)	(0.011)	(0.011)	(0.016)	(0.010)	(0.010)	(0.010)
RURLOT	—	—	—	−0.029	—	−0.032	−0.033	−0.030
	(−)	(−)	(−)	(0.003)	(−)	(0.003)	(0.003)	(0.003)
AGE	−0.043	−0.046	−0.080	−0.076	−0.073	−0.090	−0.083	−0.083
	(0.007)	(0.008)	(0.007)	(0.007)	(0.009)	(0.007)	(0.007)	(0.007)
AGE·REBUILD	0.017	0.015	0.011	0.011	0.015	0.011	0.011	0.011
	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)
GARAGE	0.080	0.044	0.069	0.065	0.028	0.071	0.069	0.068
	(0.015)	(0.017)	(0.011)	(0.011)	(0.016)	(0.011)	(0.011)	(0.011)
LIVAREA	0.489	0.467	0.359	0.357	0.419	0.356	0.353	0.358
	(0.022)	(0.034)	(0.018)	(0.018)	(0.032)	(0.018)	(0.018)	(0.018)
TOILETS	0.197	0.123	0.151	0.149	0.124	0.155	0.152	0.152
	(0.020)	(0.024)	(0.015)	(0.015)	(0.023)	(0.015)	(0.015)	(0.015)
β_e (exponential)	—	—	−0.026	−0.022	−0.030	—	—	—
	(−)	(−)	(0.002)	(0.002)	(0.005)	(−)	(−)	(−)

Exhibit 3 | (continued)

Estimation Results Based on Alternative Model Specifications

	M1	M2	M3	M4	M5	M6	M7	M8
β_p (power)	—	—	—	—	—	−0.220	—	−0.068
	(−)	(−)	(−)	(−)	(−)	(0.006)	(−)	(0.021)
β_0 (spline. power)	—	—	—	—	—	—	−0.164	–
	(−)	(−)	(−)	(−)	(−)	(−)	(0.009)	(−)
β_1 (spline. power)	—	—	—	—	—	—	−0.191	—
	(−)	(−)	(−)	(−)	(−)	(−)	(0.026)	(−)
β_2 (spline. power)	(−)	(−)	(−)	(−)	(−)	(−)	0.209	—
	(−)	(−)	(−)	(−)	(−)	(−)	(0.128)	(−)
β_q (quadratic)	—	—	—	—	—	—	—	−0.030
	(−)	(−)	(−)	(−)	(−)	(−)	(−)	(0.004)
YEARDUM97	−0.120	−0.152	−0.135	−0.134	−0.155	−0.133	−0.135	−0.133
	(0.018)	(0.241)	(0.014)	(0.014)	(0.023)	(0.014)	(0.014)	(0.014)
YEARDUM99	0.151	0.131	0.130	0.126	0.147	0.133	0.129	0.129
	(0.019)	(0.023)	(0.014)	(0.014)	(0.022)	(0.014)	(0.014)	(0.014)
YEARDUM00	0.281	0.258	0.270	0.268	0.263	0.269	0.268	0.269
	(0.018)	(0.022)	(0.014)	(0.014)	(0.021)	(0.014)	(0.014)	(0.014)
YEARDUM01	0.308	0.283	0.307	0.301	0.303	0.305	0.303	0.303
	(0.018)	(0.023)	(0.014)	(0.014)	(0.022)	(0.014)	(0.014)	(0.014)

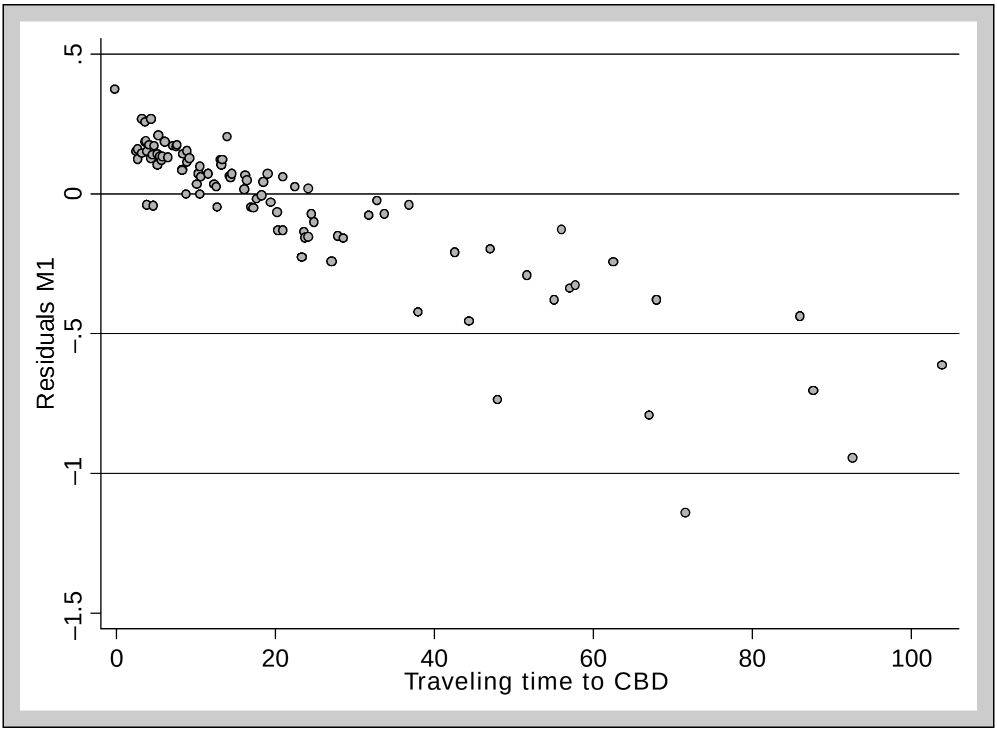
Exhibit 3 | (continued)

Estimation Results Based on Alternative Model Specifications

	M1	M2	M3	M4	M5	M6	M7	M8
R^2	0.522	0.661	0.724	0.736	0.691	0.732	0.739	0.738
R^2-adj.	0.520	0.658	0.723	0.734	0.688	0.731	0.738	0.737
L	−556	57	207	269	112	251	287	282
White test statistic	278	110	198	217	114	264	288	265
Moran's I	0.365	0.072	0.027	0.020	0.019	0.025	0.011	0.019
Standard normal deviate (z_I)	186	26.39	13.82	9.85	6.62	13.77	6.54	7.13

Note: Results are based on observations from the period 1997–2001, with robust standard errors in parentheses. Models M1and M2 are based on a non-spatial model specification. Traveling time from the CBD is represented by a negative exponential function in models M3–M5, while model M6 is based on a power function specification. Traveling time is represented by a piecewise log-linear spline function with two knots in model M7, while the power function is supplemented by a quadratic term in model M8. The number of observations (n) indicates the relevant delimitation of the geography; n = 2,788 for M1, M3, M4, M6–M8 and n = 1,188 for M2 and M5. The APE is as follows: M1 = 310,291, M2 = 274,597, M3 = 221,946, M4 = 218,508, M5 = 255,226, M6 = 221,402, M7 = 216,954, and M8 = 216,941.

Exhibit 4 | Plot Based on the Estimation of M1 and Represents the Mean of the Residuals for Each of the 98 Zones



in Exhibit 1) is found to represent a natural labor market center in the region [see Osland et al. (2005) for details on the experiments]. As a general result, the model performance is significantly better when traveling time rather than physical distance is used as the measure of spatial separation.

The model specifications M3–M5 are based on a negative exponential function, while M6 is based on a power function specification. The variable *RURLOT* is accounted for in M4, M6, M7, and M8. The model specifications M2 and M5 are estimated by data restricted to the Stavanger municipality.

In the non-spatial model specifications, the estimated parameter values related to *LOTSIZE* varied considerably with respect to the alternative subdivisions of the geography. In the case where estimation is based on data from the entire region, this coefficient is found to have a counter-intuitive, negative sign. Such counter-intuitive results do not appear when relevant measures of spatial separation are included. Since *LOTSIZE* tends to be largest in areas that are distant from the labor market center, this variable is systematically correlated with omitted variables in the non-spatial model specifications. Hence, the relevant parameter estimates in Exhibit 3 to some degree capture 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 a part of the region is considered. This is reasonable, since the negative correlation between *LOTSIZE* and housing prices is weaker when data are restricted to a small part of the region.

It follows from an evaluation of models M4 and M6 that the introduction of the variable *RURLOT* increases the model performance significantly, and it leads to larger predicted differences in housing prices between central and peripheral areas of the region. Testing for the joint significance of the two variables *LOTSIZE* and *RURLOT* by a Wald test indicates significant differences in the elasticity of *LOTSIZE* in rural and non-rural areas. No other variable had significantly different slopes as a result of this stratification of the data.

There will not be a detailed discussion of the remaining parameter estimates in Exhibit 3. As a general comment, there is a tendency that the introduction of a variable representing spatial separation (traveling time) 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 traveling time from the labor 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 around 0.51 (M1) to around 0.72 (M4 and M6, excluding *RURLOT*). In the case where the study area is restricted to Stavanger, the corresponding increase only ranges from around 0.66 (M2) to around 0.69 (M5). This pattern is also reflected through the other indicators of model performance reported in Exhibit 3. With reference to the trade-off theory, it is natural that the contribution of traveling time in explaining housing prices increases with the spatial extension of the study area within a labor market region.

It follows from Exhibit 3 that the values of Moran's I are considerably reduced when traveling time from the CBD is introduced into the model specifications. By comparing M4 to M3, it also follows that spatial autocorrelation is reduced when the variable *RURLOT* is introduced. This means 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. This result complies with the findings in for instance McMillen (2003); tests indicating spatial autocorrelation may reflect the impact of omitted variables that are correlated in space, and/or the impact of a functional form misspecification. *RURLOT* and traveling time from the CBD represent characteristics of spatial structure that influence housing prices. Notice from Exhibit 3, however, that the hypothesis of no spatial autocorrelation has to be rejected in all model formulations (M1–M6). Hence, the presence of autocorrelation in residuals is not removed.

Studies of journeys-to-work report significant distance deterrence effects (e.g., 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 ($e^{\beta_e d_{ij}}$); $\beta_p < 0$ and $\beta_e < 0$. The choice of the distance deterrence function has been considered to be essentially a pragmatic one in the literature (e.g., Nijkamp and Reggiani, 1992). Some studies have concluded, however, that the appropriateness of the functional form should be critically examined (see for instance a study of U.S. migration flows in Fik and Mulligan, 1998).

The two gradients in Exhibit 5 apply for a standard set of values on all independent variables except the traveling time from the CBD. The standard house has not been rebuilt, it has a garage, it is not located in the most rural areas, and the price in the exhibit refers to the year 2000. The observed average values are used for the remaining independent variables. Given this fixed set of attribute values, house prices will vary with distances from the CBD. The dashed curves refer to a path where spatial separation is represented by $e^{\hat{\beta}_e d_{ij}}$, while the solid curves reflect the power function specification ($d_{ij}^{\hat{\beta}_p}$). The estimated parameter values are based on M4 and M6, which means that $\hat{\beta}_e = -0.022$ and $\hat{\beta}_P = -\ 0.220$.

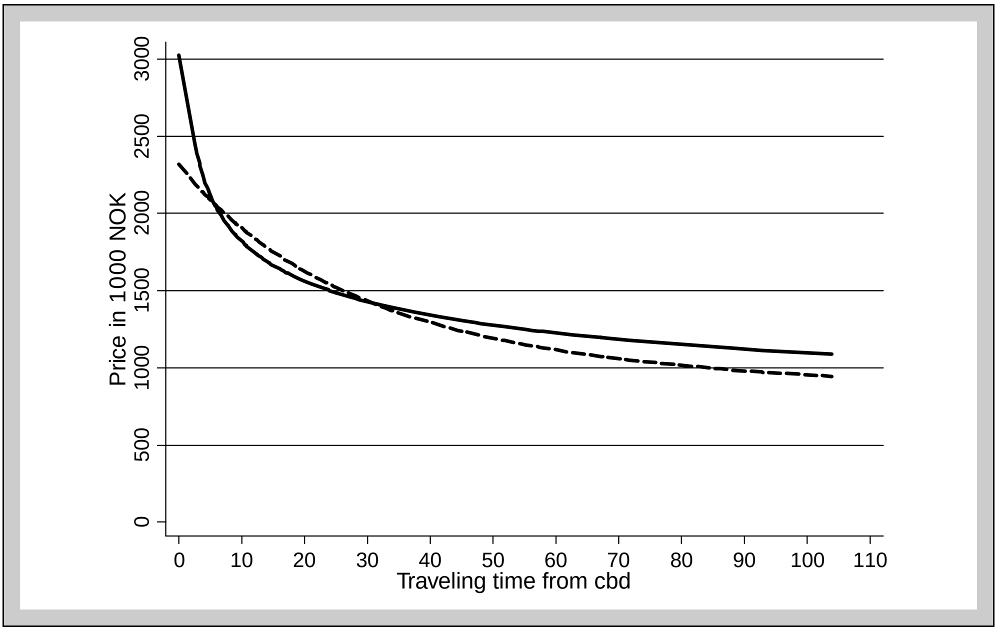
According to the results, the two alternative functions of traveling time only result in marginal differences in explanatory power. Without reporting the details, the estimation results reveal a tendency that the exponential function performs best when the estimation is based on data from the entire region, while the power function performs best in the case where data is restricted to Stavanger. The estimated marginal impact of most attributes does not differ considerably between the modeling alternatives, and neither do the values of important test statistics. Such results should be interpreted with care as the estimation deals with nonlinear relationships and mathematical transformations of dependent and independent variables. Predictions might differ substantially even with small differences in parameter estimates and measures of explanatory power. The predicted housing price gradient is very sensitive both with respect to the representation of traveling time and with respect to how large a part of the region the data refers to.

Assume for instance that a new road connection reduces the traveling distance to the CBD from 30 minutes to 15 minutes for a specific zone. The predicted increase in housing prices in this zone is 100,000 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. The sensitivity with respect to functional specification is even higher in cases where the changes in transportation infrastructure network affect areas closer to the CBD.

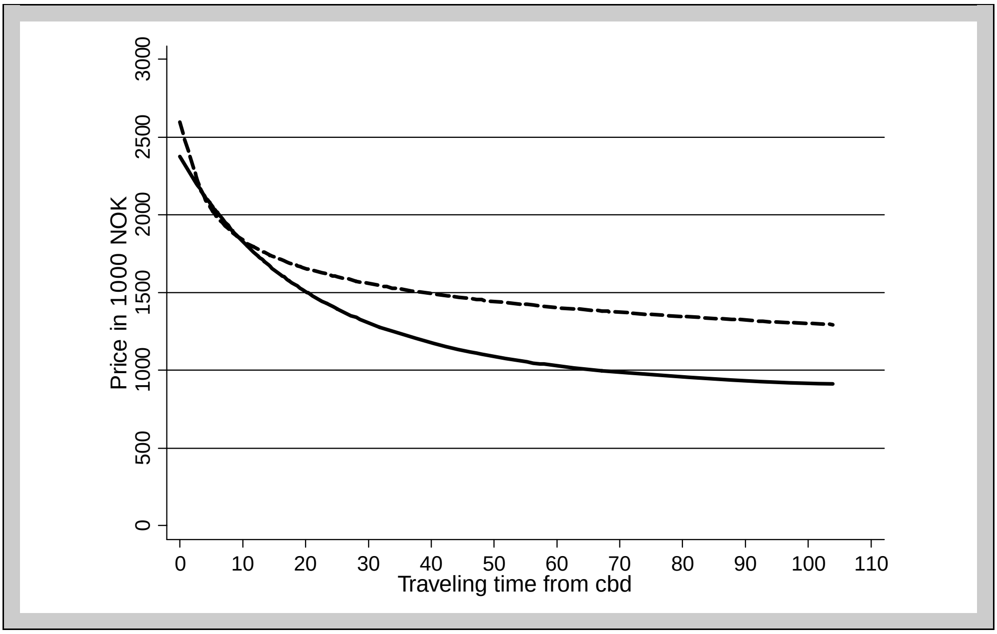
In Panel A of Exhibit 5, the parameter estimates are based on data restricted to the Stavanger municipality. Notice that a predicted housing price a long distance from the CBD is now much more sensitive to the choice of a functional representation of traveling time than in the case where the estimation is based on observations from the entire region. This illustrates the trivial but important point

Exhibit 5 | Predicted Housing Price Gradients for a Standard House

Panel A: Based on Data from the Entire Region



Panel B: Based on Data from Stavanger



The dashed curve reflects an exponential relationship between the house price and traveling time, and the solid curve is based on the power function.

that a reliable identification of such a relationship is positively related to the deviation in observed values of the independent variable. 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.

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 the results clearly indicate that the exponential function results in a more reasonable housing price gradient in cases where the estimation is based on data from the entire region; gradients based on the power function seem to predict too radical changes in housing prices for variations in distances close to the CBD. This conclusion also corresponds to a combination of intuition and knowledge of the local geography.

The results can also be interpreted in terms of the monocentric city theory. Assume the version of the monocentric model where the distance from the CBD does not appear in the utility function, but only enters through monetary costs in the budget constraint. The slope of the bid-rent gradient for housing is then given by $t(x)/H(x)$, where $t(x)$ is the marginal round-trip commuting cost at location x, and $H(x)$ is the quantity of housing at this location (e.g., McMillen, 2004; and Coulson, 1991). Deterministic asset pricing theory indicates that housing value is related to rents through the formula $1/(r - g)$, where r is the interest rate and g is the growth rate of rents. Hence, the slope of the housing price gradient is given by $t(x)/(r - g)H(x)$. Are these empirically-based findings broadly consistent with this theoretical result? This question will be addressed briefly through a back-of-the-envelope calculation.

In Håndbok-140 (1995), the Directorate of Public Roads in Norway recommends cost benefit analyses that are based on an estimate of operating costs for private cars of 0.86 NOK per km. In addition, the Directorate recommends that time costs related to commuting are set equal to 65 NOK ($\approx$ \$10) per hour. The data on distances and traveling time allow for a calculation of the average speed for workers in alternative zones. This information is used to find estimates of the average speed for 5-minute intervals of traveling time from the CBD. This average is found to be about 30 km/hr for workers living within 5 minutes from the CBD, increasing to about 60 km/hr for workers living 20 minutes from the CBD, and more or less stable around this level hereafter. As a next step this information is used to find time costs per kilometer for workers residing in different locations. Through this procedure the total round-trip marginal costs of commuting (sum of operating costs and time costs) is found to be approximately 6 NOK (\$0.92) for workers living within 5 minutes of the CBD, reducing steadily to a level of 3.9 NOK for workers residing more than 20 minutes from the CBD. A car is the dominating modal choice of commuting in the area. Hence, there is no discussion of how commuting by buses, for instance, might affect the slope of the relevant gradient.

The data on the quantity of housing refers to privately owned, single-family houses. Due to systematic spatial variation in the density of various housing types, the data does not offer reliable information on how the amount of housing develops with increasing distances from the CBD. Therefore, lot size is used as a proxy for how the quantity of housing is substituted against commuting costs. The average lot size for residences is found to increase from about 400 m^2 for residences within 5 minutes of traveling time from the CBD, and then more or less steadily increases to about 800 m^2 for residences beyond a traveling time of 40 minutes from the CBD. In the formula representing the slope of the housing price gradient, lot size is scaled to a level corresponding to the average living area of houses.

Consider next a standard house (see the definition above), and assume a total of 200 working days a year. Assume also that the growth rate of rents was equal to the inflation rate of about 3%. The interest rate was on average about 8%. Based on all this information and the set of simplifying assumptions, an extra minute of traveling time from a location of 5 minutes from the CBD results in an increase in housing prices of about 33,000 NOK. The corresponding estimate is about 24,200 NOK at a location 10 minutes from the CBD, about 18,000 NOK 15 minutes from the CBD, and about 15,000 NOK 20 minutes from the CBD. This procedure further suggests that the housing price gradient continues to fall by about 10,800 NOK per minute of traveling time when the traveling time from the CBD extends beyond 40 minutes.

These estimates are theoretically based on the monocentric city model. Both the theoretically-based considerations and the empirically-based estimates suggest a convex housing price gradient. By inspection of the housing price gradient resulting from model M4, an extra minute of traveling time from a location of 5 minutes from the CBD is found to result in an increase in housing prices of about 39,600 NOK. The corresponding estimates are 32,100 NOK, 27,100 NOK, 22,300 NOK, and 11,600 NOK for locations at10, 15, 20, and 40 minutes from the CBD, respectively. The main difference is that the empirically estimated gradient flattens out at distances beyond 40 minutes from the CBD, while the theoretically-based gradient continues to fall. It is not unreasonable that the monocentric model fails to explain housing price variation in peripheral areas of a region. Those back-of-the envelope calculations indicate that the estimation results are at least not strongly opposed to the monocentric model. In other words, this means that the reduction of housing prices as the distance increases from the CBD approximately corresponds to the increase in commuting costs, as suggested by the theory.

More Flexible Specifications of the Relationship between Housing Prices and Traveling Time

Consider the impact of introducing additional parameters in the relationship between housing prices and traveling time. One alternative is a logistic function, another is the conventional Box-Cox transformation. These two approaches result in log-likelihood values of about 275 and 280, respectively. This means that they add a relatively large contribution to the explanatory power, compared to the approaches with only one parameter [see Osland et al. (2005) for details].

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. 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 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. Let the spline function with two knots (d_{ij}^1 and d_{ij}^2) be represented by the following specification of the relevant function:

$$g_0(d_{if}) = d_{ij}^{\beta_0+\delta_{ij}^{(1)}\beta_1+\delta_{ij}^{(2)}\beta_2}, \tag{4}$$

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

$$\delta_{ij}^{k} = \begin{cases} 1 \text{ if } d_{ij} > d_{ij}^{k} & k = 1,2 \\ 0 \text{ otherwise} & \end{cases}. \tag{5}$$

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 following specification refers to model M7 in Exhibit 3:

M7: Traveling time is represented by a piecewise log-linear spline function with two knots.

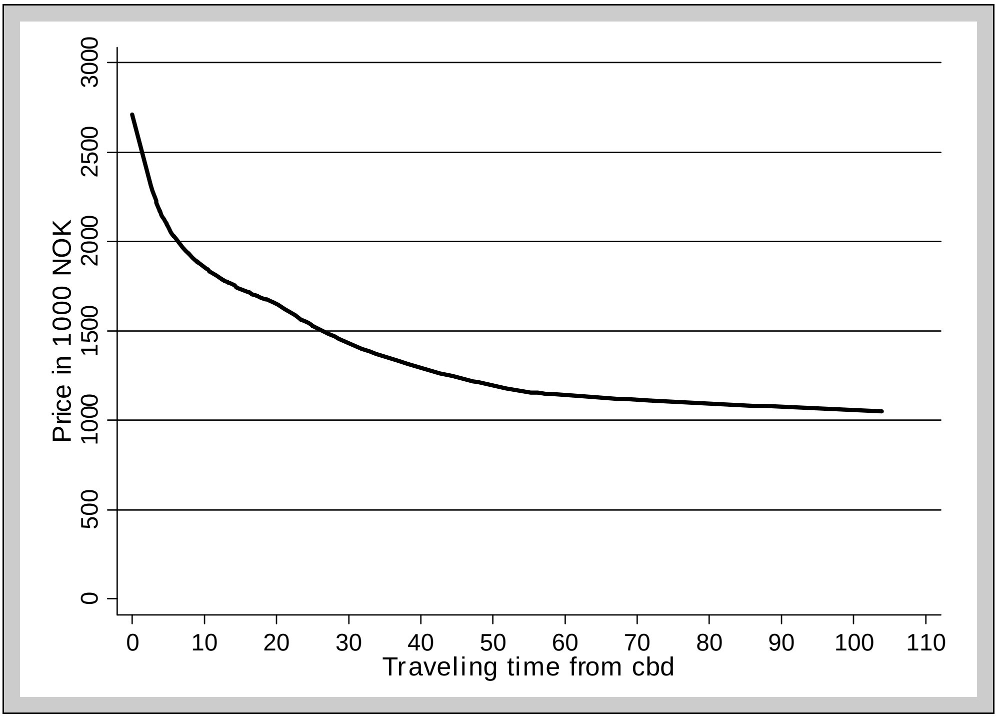
Since this spline function enters into a log-linear relationship, the elasticity of housing prices with respect to distance is constant within each of the three segments of the gradient. According to the results in Exhibit 3, there is a significant discontinuous change in housing prices at a distance corresponding to 20 minutes of traveling time from the CBD. The relevant elasticity increases from -0.165 to -0.353 when traveling time from the CBD 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 Exhibit 3 appears for a traveling time of 55 minutes from the CBD. The segment represented by traveling times exceeding 55 minutes has an elasticity of -0.154. According to the maximum likelihood estimation, this spline function approach results in more satisfying log-likelihood values than both the logistic approach and the Box-Cox transformation.

A model specification where the distance function entering into the hedonic regression model is assumed to be piecewise linear with three knots was also tested. For all practical purposes, this results in a housing price gradient that is more or less totally overlapping the gradient resulting from model specification M7, which is illustrated in Exhibit 6.

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 are 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. Increasing the number of knots does not, however, offer a satisfying general hypothesis of commuting behavior. Hence, such an approach represents a questionable basis for predicting the effects of changes in, for example, road transportation infrastructure.

As discussed above, the power function results in a 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. Hence, both intuition and numerical results indicate that the

Exhibit 6 | Predicted Housing Price Gradients for a Fictional Standard House, Based on a Piece-wise Log-Linear Spline Function with Two Knots.



assumption of a globally constant elasticity of housing prices with respect to distance is unreasonable. As an alternative to the spline function approach, this can be adjusted for by introducing a quadratic term in the regression equation:

M8: The power function is supplemented by a quadratic term; $h(d_{ij}) = d_{ij}^{\beta} \cdot ((d_{ij})^2)^{\beta_q}$.

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

$$El_{d_{ij}}P_i = El_{d_{ij}}h(d_{ij}) = \beta + 2\beta_q \ln d_{ij}. \tag{6}$$

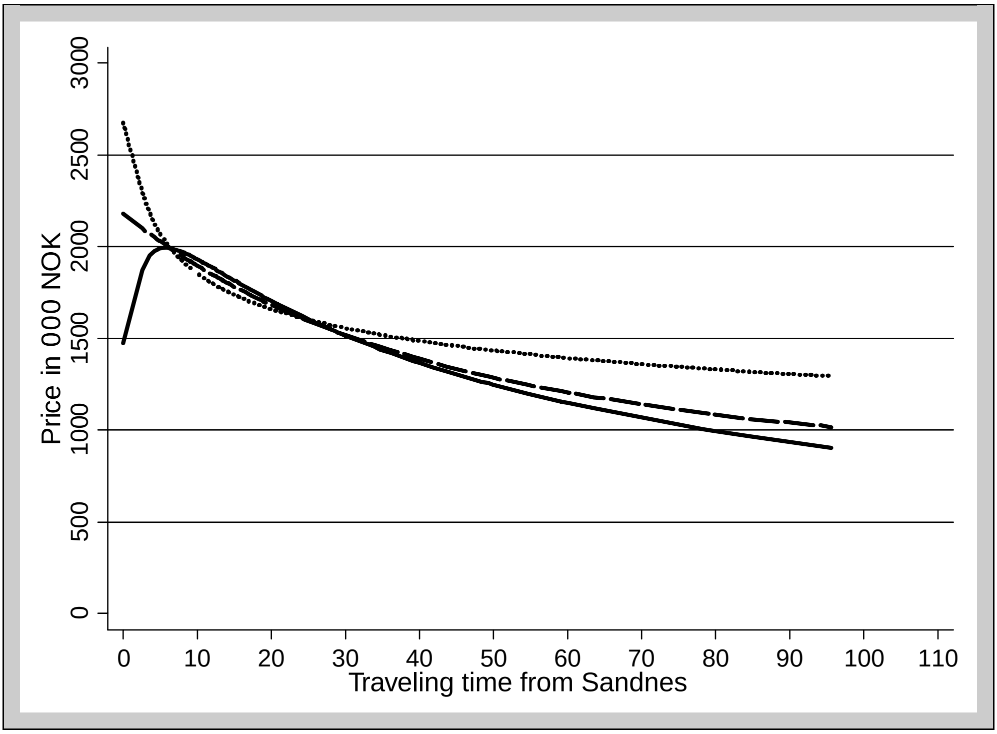
From Exhibit 3, notice that both parameters related to distance are significantly negative; $\hat{\beta} = -0.069$ and $\hat{\beta}_q = -0.030$. This means that house prices become increasingly more elastic with respect to distance for movements downward along the gradient. The point elasticity is −0.069 when the traveling time is 1 minute from the CBD, while it is, for instance, −0.205 and −0.300 for locations with, respectively, 10 and 50 minutes of traveling time from the CBD.

Since the number of observations in this study is relatively large, multicollinearity is not expected to be a problem. This is also confirmed by estimated values of the variance inflation factor (VIF, see for instance Greene, 2003). The mean VIF-value is 1.5 in M6 and 4.22 in M8. 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.

M8 is a hierarchical extension of M6, which is based on a power function specification of the spatial separation. In comparing those two model formulations, it follows from the values of L in Exhibit 3 that the value of the likelihood ratio test statistic is approximately 31. This exceeds by far the critical value of a chi-square distribution with one degree of freedom, at any level of significance. Hence, the quadratic term significantly adds to the explanatory power, and M8 fits the data significantly better than the approaches based on one-parameter representations of the traveling time. The two models, M7 and M8, can be compared by means of a non-nested test. Based on a Davidson-MacKinnon test (Wooldridge, 2003), M7 is found to be statistically superior to M8. As mentioned above, however, here are theoretical arguments disfavoring the spline function approach.

It follows from Exhibit 3 that spatial autocorrelation is additionally reduced when the traveling time appears in a more flexible function compared to a one-parameter function, using data covering the whole area. The null hypothesis of no spatial autocorrelation, however, still has to be rejected in M7 and M8. In order to find whether the problem is spatial autocorrelation or spatial heterogeneity, further testing is needed. Lagrange Multiplier tests were performed, which explicitly

Exhibit 7 | Predicted Housing Price Gradients Based on the Traveling Time from Sandnes



The dashed curve represents an exponential spatial separation function, while the dotted curve is based on a power function specification. The non-monotonic curve is based on M8.

specify alternative hypotheses regarding the two mentioned problems (Florax, Voortman, and Brouwer, 2002). The tests asymptotically follow a χ^2 distribution with one degree of freedom. The results of the test show that the problem is spatial heterogeneity, which leads to spatial dependence in the error. This problem is assumed to originate from erroneously omitted spatial variables, functional misspecifications, or parameters that are not stable across the geographical area. As a result, the ordinary least squares estimator is unbiased but inefficient. In order to study whether this causes incorrect inferences, spatial autoregressive error models (Anselin, 1988) have been estimated. These models typically account for spatial heterogeneity by specifying a spatial autoregressive process in the error. Correcting for spatial heterogeneity only gives minor changes in the estimated coefficients and standard errors, however, and the results are not reported in this paper. Models estimated by ordinary least squares are robust for these kinds of misspecifications.

The values of the goodness-of-fit indices resulting from the experiments do not indicate that the more complex functional representations of spatial separation contribute with a substantially better model formulation. 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. Exhibit 7 illustrates three alternative housing price gradients based on the traveling time from the center of Sandnes (see Exhibit 1). All gradients refer to the standard house. The dashed curve represents the predicted housing price gradient in a case where the 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. The slope of those gradients is reduced compared to the corresponding gradients originating from the CBD in the region (see Exhibit 5). Still, both gradients predict housing prices to fall monotonically as the traveling time from Sandnes increases. As indicated by the gradient based on the more flexible model M8, this is not a reasonable conclusion. A 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 Exhibit 7 is much more reasonable than the monotonic counterparts.

Conclusion

This paper primarily contributes to the literature by demonstrating the importance of incorporating spatial separation through an appropriately specified function and an appropriate delimitation of the geography. In addition, it contributes by estimating implicit prices of dwelling-specific attributes in a region where such studies have not been carried through so far. In fact, most empirical housing market studies refer to metropolitan areas rather than a wider regional context, and relatively few European studies are reported in the literature. The findings also reveal that information on the character of sub-areas (through the variable *RURLOT*) contributes to explain spatial variation in housing prices.

The southern part of Western Norway is very appropriate for the purpose of identifying reliable housing price gradients, since it can be considered to be more or less like an isolated island with one dominating central place in a comprehensive regional labor market. This contributes to make it appropriate for reaching reliable parameter estimates reflecting the ‘‘access-space-trade-off’’ rather than edge effects and/or local characteristics of the central place system. The region is not literally corresponding to the geography underlying the traditional trade-off theory (Alonso, 1964), 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 of different ranks, but Stavanger indisputably has a very dominating position. The characteristic multi-nodal structure observed in many metropolitan areas is less dominant in the study area, and the diffusion of new residential areas has to a large degree been determined by employment growth in, and close to, the Stavanger city center.

The experiments demonstrate that unbiased estimates of the housing price gradients cannot be based on a truncated specification of the markets. If estimation is based on observations covering only a part of the relevant labor and housing market, the results indicate that the predicted housing price gradients are not reliable, and that coefficient estimates associated with dwelling-specific variables are biased. As mentioned in the introduction, this might be one reason why many studies report counter-intuitive results on systematic spatial variations in housing prices.

As in Coulson (1991), the study area provides a suitable laboratory for testing the theorems of the monocentric city model. The reduction in housing prices as the traveling time from the CBD increases approximately corresponds to the increase in commuting costs, as suggested by the theoretical modeling framework. In other words, the tentative calculations indicate that the predicted housing price gradient fits reasonably well to the corresponding theoretically-based gradient.

Another non-obvious insight is that an exponential function specification of the traveling time results in more reliable housing price gradients than a power function specification. A log-linear regression model tends to over-predict housing prices in locations close to the CBD. Compared to the one-parameter approaches, however, model performance is improved if appropriate flexible functional specifications of traveling time are introduced. In evaluating alternative flexible functional forms, results on explanatory power should be considered in combination with pragmatic, theoretical, and interpretational arguments. Therefore, the approach incorporating a quadratic term is especially appealing. In addition, the specification of flexible functional forms reduces problems related to spatial autocorrelation. The 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 considered, with less obvious identification of a regional center.

All in all, there are encouraging results, with satisfying goodness-to-fit, reliable coefficient estimates, and intuitively reasonable predictions of housing price gradients. The results represent important input in an evaluation of, for instance, residential construction programs, urban renewal, and/or investments in transportation infrastructure. In addition, the findings 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.

Endnote

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

References

Adair, A., S. McGreal, A. Smyth, J. Cooper, and T. Ryley. House Prices and Accessibility: The Testing of Relationships within the Belfast Urban Area. *Housing studies*, 2000, 15, 699–716.

Alonso, W. *Location and Land Use. Toward a General Theory of Land*. Cambridge, MA: Harvard University Press, 1964.

Anas, A., R. Arnott, and K.A. Small. Urban Spatial Structure. *Journal of Economic Literature*, 1998, 36, 1426–64.

Anselin, L. *Spatial Econometrics: Methods and Models*. London, UK: Kluwer Academic Publishers, 1988.

Bartik, T.J. and V. K. Smith. *Urban Amenities and Public Policy*. Handbook of Regional and Urban Economics, Volume 2, B.V., Netherlands: Elsevier, 1987.

Clapp, J.M. A Semiparametric Method for Valuing Residential Locations: Application to Automated Valuation. *Journal of Real Estate Finance and Economics*, 2003, 27, 303–20.

Coulson, N.E. Really Useful Tests of the Monocentric Model. *Land Economics*, 1991, 67, 299–307.

Dubin, R.A. and C.H. Sung. Spatial Variation in the Price of Housing Rent Gradients in Non-Monocentric Cities. *Urban Studies*, 1987, 24, 193–204.

Fik, T.J. and G.F. Mulligan. Functional Form and Spatial Interaction Models. *Environment and Planning*, 1998, 30, 1497–1507.

Florax, R.J.G.M., R.L. Voortman, and J. Brouwer. Spatial Dimensions of Precision Agriculture: A Spatial Econometric Analysis of Millet Yield on Sahelian Coversands. *Agricultural Economics*, 2002, 27, 425–43.

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

Håndbok-140, Del 1, Prinsipper og Metodegrunnlag, Statens vegvesen, 1995.

Heikkila, E., P. Gordon, J.I. Kim, R.B. Peiser, and H.W. Richardson. What Happened to the CBD-Distance Gradient?: Land Values in a Polycentric City. *Environment and Planning A*, 1989, 21, 221–32.

Kim, C.W., T.T. Phipps, and L. Anselin. Measuring the Benefits of Air Quality Improvement: A Spatial Hedonic Approach. *Journal of Environmental Economics and Management*, 2003, 45, 24–39.

Kmenta, J. *Elements of Econometrics*. New York City, NY: Macmillan, 1986.

McMillen, D. Spatial Autocorrelation or Model Misspecification. *International Regional Science Review*, 2003, 26, 208–17.

——. Testing for Monocentricity. Prepared for *A Companion to Urban Economics*, R. Arnott and D. McMillen (Eds.), 2004.

McMillen, D.P. and P. Thorsnes. The Reaction of Housing Prices to Information on Superfund Sites: A Semiparametric Analysis of the Tacoma, Washington Market. *Advances in Econometrics*, 2000, 14, 201–28.

Nijkamp, P. and A. Reggiani. *Interaction, Evolution and Chaos in Space.* Springer-Verlag, 1992.

Osland, L., I. Thorsen, and J.P. Gitlesen. Housing Price Gradients in a Geography with One Dominating Center. Working Papers in Economics, Department of Economics, University of Bergen, no. 06/05, 2005.

Richardson, H.W. Monocentric vs. Policentric Models: The Future of Urban Economics in Regional Science. *The Annals of Regional Science*, 1988, 22, 1–12.

Sheppard, S. Hedonic Analysis of Housing Markets. *Handbook of Regional and Urban Economics*. E.S. Mills and P. Cheshire (Eds.), 1999.

Thorsen, I. and J.P. Gitlesen. Empirical Evaluation of Alternative Model Specifications to Predict Commuting Flows. *Journal of Regional Science*, 1998, 38, 273–92.

Waddell, P., B.J.L. Berry, and I. Hock. Residential Property Values in a Multinodal Urban Area: New Evidence of the Implicit Price of Location. *Journal of Real Estate Finance and Economics*, 1993, 7, 117–41.

Wooldridge, J.M. *Introductory Econometrics. A Modern Approach*. Thomson, South Western, 2003.

This paper has benefited from comments by Espen Bratberg, Viggo Nordvik, and two anonymous referees. We would also like to thank Roger Bivand for useful suggestions.

Liv Osland, Stord/Haugesund University College, Haugesund, Norway or liv.osland@hsh.no.

Inge Thorsen, Stord/Haugesund University College, Haugesund, Norway or inge.thorsen@hsh.no.

Jens Petter Gitlesen, University of Stavanger, Stavanger, Norway or jens.p.gitlesen@uis.no.