Testing for the impact of local spatial structure characteristics on house prices*

Liv Osland and Inge Thorsen†

June 14, 2006

Abstract

The starting point of this paper is a hedonic regression model where house prices are explained as a result of urban attraction and the accessibility to job opportunities in the region. The basic hypothesis is that house prices reflect that households in addition value accessibility to job opportunities in the neighborhood. We propose several measures of local labor market characteristics, and test for the impact on house prices. The alternative measures do not add considerably to the explanatory power. Still, some characteristics contribute significantly, and affect the size and interpretation of the relationship between local labor market conditions and house prices.

JEL-classification: R21, R31

Preliminary version, presented at the 46th Congress of the European Regional Science Association, Volos, Greece.

1 Introduction

In Osland and Thorsen (2005) spatial variation in housing prices were explained to result from an urban attraction and a labor market accessibility effect. Based on data from the southern parts of Rogaland County in the south west of Norway we found that such spatial characteristics added considerably to the explanatory power in an approach that also accounted for several house-specific attributes in a relatively macroscopical description of the geography. The empirical

*We would lik to thank Roger Bivand for helpful suggestions.
†Stord/Haugesund University College, Bjørnsonsgt. 45, 5528 Haugesund, Norway

results in Osland and Thorsen (2005) refer to a regional rather than an urban context, covering a connected labor and housing market area rather than just an urban area. The macroscopical perspective is reflected by the fact that the geography is subdivided into zones that extend over a relatively large area, and the fact that we consider interzonal rather than intrazonal variations in housing prices. At least in such a macroscopical perspective it can be argued that labour market accessibility and potential commuting distances are of vital importance for how readily saleable a house is, and what price that is achieved.

To capture the urban attraction and the labor market accessibility effects Osland and Thorsen (2005) introduced a labor market accessibility measure and the distance from the cbd in a hedonic model formulation. The one-dimensional measure of distance represents the position relative to the regional center, while the accessibility measure captures multicentric tendencies in the regional distribution of employment. In this paper we discuss the hypothesis that those two measures offer an adequate description of the geography for the purpose of explaining spatial variation in housing prices. We test for the possible impact of local characteristics which are not represented by the two globally defined measures of spatially structure.

The classical trade-off between commuting costs and housing prices represents a basic relationship in regional science and urban economics. The standard theoretical reference for the relationship is the "access-space-trade-off"of Alonso (1964), which gives rise to house prices falling with increased distance from the city center in a monocentric geography. This trade-off is represented by the labor market accessibility measure in Osland et al. (2005), while the distance from the cbd captures the urban attraction effect. The idea that the multicentric character of the labor market matters in explanation of housing prices is of course reflected in the literature, see for instance Dubin and Sung (1987), Richardson (1988), Heikkila et al. (1989), Waddell et al. (1993), and Adair et al. (2000). Most of those contributions emphasize the importance of including the distance to secondary employment centers, while for instance Adair et al. (2000) introduced a gravity based measure of transport accessibility in a study of the Belfast urban area.

In this paper we discuss and test whether relevant spatial labor market characteristics are adequately represented by a labor market accessibility measure defined from the spatial distribution of employment opportunities throughout the entire region. Does such a measure capture the impact of complex decision processes in modern households, or should model formulations also incorporate local labor market characteristics?

In general many authors account for spatial attributes that affect housing prices only in a small area. Heikkila et al. (1989) distinguish between macro-and microlocational effects, and implicitly introduce the impact related to the multipurpose nature of household spatial interaction. Households also value access to other activities than job opportunities. Li and Brown (1980) classify activities relative to three categories of attributes: aesthetic attributes, pollution sources and service facilities. Through fuzzy logic and very disaggregate data Theriault et al. (2003) account for information on how different categories of households perceive the accessibility to 17 different urban amenities.

Due to data restrictions we are not able to account for a wide range of possibly relevant local attributes and activities. We proceed through a zonal subdivision of the geography that corresponds to the most detailed spatial level for which official data are available. Still, this subdivision represents a relatively macroscopical description of the geography, and more spatially disaggregated data on local attributes would require a massive effort on data collection. In this paper we primarily focus on the impact of the location relative to labor market opportunities rather than a set of location-specific amenities. Considering our macroscopical perspective of the geography a high degree of residential interzonal homogeneity can be expected for many amenities, like for instance the view, the neighborhood quality, or the distance to nursery school. Many attributes of this kind 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. Similarly, we ignore the possible impact on housing prices of systematic variation in zonal socioeconomic characteristics. Labour market accessibility, on the other hand, is a location-specific characteristic with considerable interarea variation that is accounted for both through the globally defined accessibility measure and some measures reflecting the labor market situation within a zone and surrounding zones.

The lack of information on intrazonal location-specific attributes reduces the potential explanatory power of our estimation. To some degree the effect of the omitted variables might for instance be represented by location-specific dummy variables. This is not, however, a recommended procedure if focus is primarily on explaining and predicting spatial variation in housing prices. Our macroscopical approach means that we focus on general effects rather than on obtaining a highest possible explanatory power for our study area.

Based on the so-called hedonic method our ambition is to estimate the implicit price structure related to spatial structure characteristics, like the accessibility of job opportunities. Rosen (1974) offered a theoretical foundation for this method, interpreting the hedonic function in an equilibrium framework, enveloping the consumers' so-called "bid-functions"and suppliers' "offer-functions"(Quigley 1982).

In Section 2 we present the region and our data, while the modeling framework is introduced in Section 3. Alternative measures of local spatial structure are proposed in Section 4. Section 5 offers results based on the proposed measures, and results based on semi-parametric approaches are evaluated in Section 6. Finally, some concluding remarks are given in Section 7.

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. 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. As an indicator of (commuting) distances, there is 79 km from the center of Stavanger to Egersund in the south. Stavanger is the dominating city in the region, with about 115000 inhabitants. The region is described in more detail in Osland et al. (2005), and is very appropriate for studies of the relationship between spatial labor market interaction and the housing market. The suitability is due to the fact that it is a fairly large integrated and autonomous region; the landscape is fairly homogeneous and the topographical barriers protect from disturbances in other regions, rather than causing spatial submarkets and disconnections in the intraregional transportation network. The region is more or less like an island with one dominating city, with a tendency of a steadily increasing rural profile as the distance increases from this city center.

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

2.2 The data

The housing market data consist of transactions of privately owned single-family houses in the period from 1997 through the first half of 2001. Our sample of 2788 property transactions represents approximately 50% of the total number of transactions of privately owned single-family houses in the region during the relevant period. The transactions data on the freeholder dwellings have been provided for us from two sources: the national land register in Norway and Statistics Norway. For more details on those data, and descriptive housing market statistics for separate parts of the region, see Osland et al. (2005).

The division of the region into zones corresponds to the most detailed level of information which is officially available on residential and work location of each individual worker within the region. The information is based on the Employer-Employee register, and provided for us by Statistics Norway. Our analysis also requires data on total population in the (postal delivery) zones. We gained access to this information through the Central Population Register in Statistics Norway. 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 massive effort and resources required on data collection.

The matrices of Euclidean distances and traveling times were prepared for us by the Norwegian Mapping Authority, who have at their disposal all the required information on the road network and the spatial residential pattern.

The calculations were based on the specification of the road network into separate links, with known distances and speed limits, and it is accounted for the fact that actual speed depends on road category. Information of speed limits and road categories is converted into travelling times through instructions (adjustment factors for specific road categories) worked out by the Institute of Transport Economics. The center 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.

3 The modeling framework

As indicated in the introduction our approach is implicitly based on the assumption that location specific (microlocational) amenities are not varying systematically across the zones. In other words we implicitly assume that the regional variation in such amenities can also be found within a zone, and that there is insignificant spatial variation in zonal average values. Based on this assumption we focus on how systematic variations in centrality and labour market accessibility influence average zonal housing prices. We distinguish between two categories of attributes. One category is the physical attributes of the specific dwelling, the other is related spatial structure characteristics and the accessibility to labor market opportunities. In a general form the hedonic price equation can be written 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; $l = 1, ...L$, $i = 1, ...n$

Table 1 offers a list of non-spatial dwelling-specific attributes incorporated in our modeling framework.

Table 1: List of non-spatial dwelling-specific variables

Variable	**Operational definition**
REALPRICE	selling price of property
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

In addition to the dwelling-specific attributes we introduce the variable RURLOT into our regression model specifications. This variable is based on a stratification of the geography into rural and urban areas. The rural areas include four municipalities, see Osland et al. (2005) for details and criteria. RURLOT is defined to be the product of the dummy variable representing rural areas and the variable LOTSIZE, defined in Table 1. Osland et al. (2005) found that this spatial characteristic variable increased the explanatory power of the model significantly.

For a separate discussion of non-spatial modeling alternatives, see Osland et al. (2005). Based on the same data set that is considered in this paper, Osland et al. (2005) primarily focused on model formulations incorporating the distance from the cbd. They further considered model performance for different spatial delimitations of the housing market, and they experimented with different mathematical representations of the relationship between dependent and independent variables, as well as different measures of spatial separation (physical distance and traveling time). It followed from their evaluation that the use of more complex and flexible functional specifications of traveling time contributes significantly to the explanatory power compared to a one-parameter approach. In addition the more flexible forms are found to represent a more reliable basis for predicting housing price gradients. Based on explanatory power in combination with pragmatic, theoretical, econometric, and interpretational arguments, Osland et al. (2005) recommended a power function specification supplemented by a quadratic term. According to this approach traveling time appears in the regression equation through the following expression:

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

According to the idea of a trade-off between housing prices and commuting costs, Osland and Thorsen (2005) introduced a gravity based measure of labor market accessibility that captures the fact that job opportunities are not solely concentrated to the cbd. In this Hansen type (Hansen 1959) of accessibility measure distance appears through a negative exponential function. Let $\sigma_e < 0$ be the weight attached to distance, and γ_e the parameter attached to the number of job opportunities, D_k. The accessibility measure, S_j, is then defined as follows:

$$S_j = \sum_{k=1}^{w} D_k^{\gamma_e} \exp(\sigma_e d_{jk}) \tag{3}$$

Here, D_k represents the number of jobs (employment opportunities) in destination (zone) k.

The measure S_j 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, introduced to account for the possibility that the relevant kind of spatial pull originates from several destination opportunities. The basic hypothesis underlying the introduction of the measure is that workers prefer a location with favorable job opportunities within a reasonable distance from their residential site. Hence, labor market accessibility influences the number of households bidding for a house that is for sale, explaining spatial variation in housing prices. The Appendix offers estimates of the relative labor market accessibility of all the zones in our study, defined by $\frac{S_j}{\frac{1}{98}\sum_{j=1}^{98} S_j}$.

In this paper we take as our starting point a model formulation (the "Basic model" (BM)) that incorporates both travelling time from the cbd, through Equation (2), and the gravity based labor market accessibility measure S_j:

$$\begin{aligned}\log P_{it} \quad = \quad & \beta_0 + \beta_1 \log \mathrm{LOTSIZE}_i + \beta_2(\mathrm{RUR} \log \mathrm{LOT})_i + \beta_3 \log \mathrm{AGE}_i + \beta_4(REBUILD \log \mathrm{AGE})_i + \\ & + \beta_5 \mathrm{GARAGE}_i + \beta_6 \log LIVAREA_i + \beta_7 \log \mathrm{NUMBTOIL}_i + \beta \log \mathrm{TIMECBD}_i + \\ & + \beta_q(\log \mathrm{TIMECBD}_i)^2 + \beta_8 \log \mathrm{ACCESSIBILITY}_i + \sum_{t=97}^{01} \beta_t \mathrm{YEARDUM}t_i + \epsilon_{it} \qquad (4)\end{aligned}$$

4 Alternative local spatial structure characteristics

In this paper we test for the possibility that labour market accessibility should be defined at two separate spatial levels of aggregation. The motivation for this test springs out from the hypothesis that residential location choices can be considered as the result of a hierarchical, two-step, decision process. As a first step of such a decision process the households determine what parts (municipalities) of the region that is relevant in their search for a house. Residential location preferences can for example in general be due to environmental conditions, location-specific amenities, public services, friendships and family relations, or simply preferences imprinted from childhood experiences. In this first, macroscopical, step of the decision process, labour market considerations are important, since households, ceteris paribus, prefer a location with favourable job opportunities within a reasonable distance from their residential site. In our model formulations the job opportunity density is represented by the accessibility measure S_i, and the underlying hypothesis is that labour market accessibility influences the number of households bidding for a house that is for sale, explaining spatial variation in housing prices.

The second step of the decision process concerns the choice of a residential site within the relevant search area. Any location within this area represents an acceptable combination of job search and realisation of residential site preferences. This does of course not mean, however, that all location alternatives within the area are evaluated to be equivalent. The evaluations are influenced by a multitude of attributes, and individual households do not put the same weight on different attributes. Due to data restrictions we have of course no chance to capture this heterogeneity in preferences and location pattern of attributes. Our ambition is to test for the possible impact from spatial structure and labour market characteristics that are not captured through the accessibility measure S_j. Such labour market characteristics might systematically affect individual evaluations, and the willingness-to-pay for a house that is for sale.

The interpretation of housing demand in terms of a two-step search procedure is analoguous to the hypothesis of hierarchical destination evaluation in spatial interaction analysis. Fotheringham (1983) contributed to this hypothesis by introducing the competing destinations model to improve the ability of the gravity modelling tradition to capture spatial structure effects. The competing destinations model includes a measure of accessibility into the structural model equation, to capture how alternative destinations appear in clusters according to their position relative to a specific origin. This model formulation can be interpreted from a two-stage hierarchical decision process, see Fotheringham (1988) and Pellegrini and Fotheringham (1999). The introduction of choice probabilities and choice restrictions are motivated by the idea that the capacity of humans to process large amounts of information is limited. Decision makers are assumed to conduct a hierarchical processing strategy rather than a simultaneous evaluation of all alternatives. First, they select the set of alternatives that are relevant destination choices. Second, a specification destination is selected from this set of alternatives. Thorsen and Gitlesen (2000) offer an economic interpretation of the competing destinations model as a framework for studying job-search problems. We will not enter into a similar thorough analysis of the search procedure in the housing market, in this paper we just formulate the hypothesis that the combination of regional accessibility measure and local measures of excess labour demand can be explained from a hierarchical decision process.

Since our study area has a very dominating center, labor market accessibility covariates strongly with distance from the cbd. Still, we found in the previous subsection that labor market accessibility adds significantly to explain housing prices also in the case where distance from the cbd is accounted for. A natural hypothesis is that the accessibility measure captures some polycentric characteristics and/or local anomalies of the geography. Some zones in our study area can best be described as bedroom communities for the Stavanger cbd, but neither local nor basic production sectors are in general entirely concentrated to the cbd of a region. For a theoretical discussion of the spatial distribution of local sector employment, see Gjestland et al. (2006). In this section we examine the possibility that such characteristics are better represented by alternative local measures of the spatial structure rather than by a simple aggregate measure of regional labor market accessibility.

In other words the challenge is to identify general spatial structure characteristics that reflect complex systematic multipurpose decisions in the households. Some labor market considerations are not, however, captured by a simple measure of regional labor market accessibility and/or a one-dimensional function of distance from the cbd. Two-worker households might, for instance, prefer residential locations with favorable job opportunities in the close neighborhood. This facilitates the logistics of running the household, and potentially reduces transport costs, for instance by reducing the need for disposing two cars. Hence, it can be argued that the model should incorporate spatial structure measures identifying local clusters of favorable job opportunities.

Residential location decisions are of course not determined by labor market considerations alone. In an empirical study based on official data rather than for instance a questionnaire, we cannot account for the impact of individual interdependencies and preferences related to childhood experience or to the presence of specific amenities. As mentioned in the introduction evidence can be found in the housing market literature that households value access to other activities than those related to the job situation. Both proximity to schools and shopping centers might for instance explain why a location is attractive for residential purposes, resulting in high housing prices.

It is in general important to account for the interdependency between spatial interaction behavior and location decisions. Our distinction between local and regional accessibility is motivated from the relationship between housing prices and spatial labor market interaction. A similar distinction was applied in Handy (1993), in a study focusing on spatial differences in average shopping distances and shopping frequencies for a given residential location pattern. Accessibility was in general defined relative to commercial, non-industrial, activities, and local accessibility was defined with respect to "convenience" establishments, such as local supermarkets, drugstores etc. Handy (1993) defined gravity based accessibility measures, and parameters were estimated through data from a travel survey for shopping trips. The estimation results were applied in an analysis of how characteristics of the spatial structure affect automobile travel and gasoline consumption; communities with low local but high regional accessibility tend to induce the most amount of automobile travel, motivating a policy providing high levels of local accessibility.

As mentioned in the introduction our subdivision of the region into zones corresponds to a rather macroscopic, spatially aggregate, description of the geography. This especially applies for the most peripheral parts of the region. The focus on regional measures of spatial structure is implicitly based on the assumption of a relatively high degree of interzonal homogeneity. Most relevant activities can be performed within each zone, and we ignore potentially relevant microscopic locational aspects. Our primary ambition has been to capture effects related to general spatial and labor market characteristics rather than housing price variations explained by local, intrazonal, conditions. In this subsection we will discuss the possibility that local variation in spatial structure characteristics influences housing prices also in a dataset corresponding to a relatively macroscopic description of the geography.

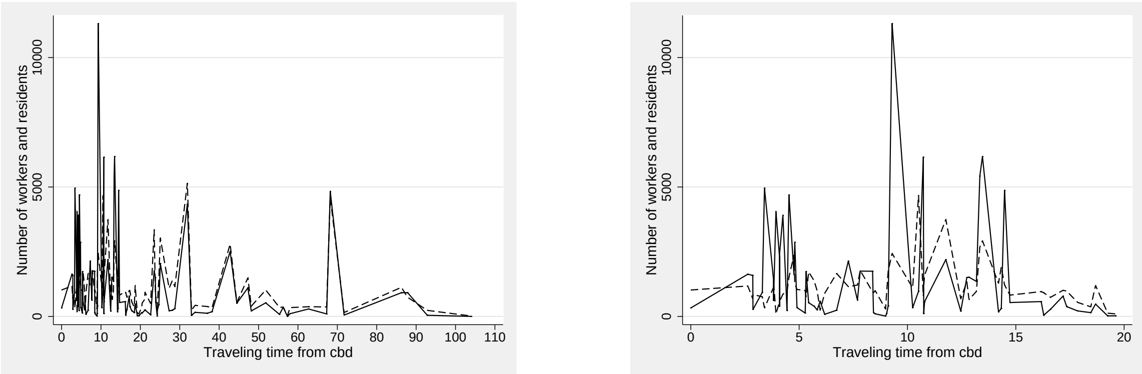
This aggregate subdivision of the geography into rather wide-spreading zones affects the prospects for defining appropriate local accessibility measures. Accessibility can be measured by the cumulative opportunities of the relevant activities, that is the number of activities reached within a given travel time (see Handy and Niemeier (1997)). Analogously Yinger (1979) suggests to specify rings of employment around the cbd, to capture the fact that not all jobs are located within the cbd. The fact that our data refers to a rather aggregate subdivision of the geography limits the possibility to specify continuously defined measures of the cumulative opportunities of activities with respect to travel time. Despite this lack of spatially very disaggregate data we attempt to capture the effect of relevant local characteristics of the labor market also through rough specifications of this kind of accessibility measures. Besides pragmatic arguments based on available data the use of simple specifications has the advantage of being more straightforward to interpret than complex measures focusing primarily on explanatory power for the specific study area. In addition to the technical specification of the measure a variable representing the activity level has to be defined. The chosen specifications of variables, or activities, are all motivated by labor market considerations. Below, we will be more specific on the formulation local accessibility measures.

Subcenters

Despite the fact that both employment and population are strongly concentrated to Stavanger and adjacent municipalities, some other regional subcenters can be identified. Guiliano and Small (1991) focus on how subcenters typically develop as a conflict between agglomeration forces and congestion effects, and they discuss empirical criteria for identifying subcenters. The criteria are applied to identify subcenters in the large and complex Los Angeles region. Both McDonald (1987) and Guiliano and Small (1991) argue that employment, not population, is the key to understand the formation of centers, and that a subcenter is a zone whose measure of employment concentration is higher than all adjacent zones. From such arguments Guiliano and Small (1991) propose criteria based on a specific density cutoff of employees per acre and a minimum total employment. A subcenter is identified if those criteria are met, and if all immediately adjacent zones have density below this cutoff. We do not apply such explicit criteria. It is of course more straightforward to identify subcenters in our study area than in the Los Angeles region, but the line of arguing is relevant also in the kind of regions that we consider.

The left part of Figure 2 illustrates how employment and population are distributed across our study area, with travel time from the peak of the Stavanger cbd represented on the horizontal axis. The figure indicates that two marked subcenters can be identified outside the most central parts of the region. Those are the centers of the municipalities Time and Eigersund, respectively, and they are represented by two marked peaks in employment densities, in a traveling time by car of about 32 and 68 minutes from the regional center. Notice also from the left part of

Figure 2 that the spatial distribution of workers (population) has a marked peak in those two subcenters, where the number of jobs is approximately balanced to the number of workers. The right part of the figure illustrates that jobs are spatially considerably less balanced to workers in the central part of the region, where the subdivision of the geography into zones is more disaggregate. Based on information of commuting flows, Statistics Norway categorizes the two zones as subregional centers in the geography.



a) The entire region

b) The central area of the region

Figure 2: The spatial distribution of jobs and workers. The solid lines represent the number of jobs, while the dashed line represents the number of workers residing at alternative locations.

The presence of the two subcenters is represented by dummy variables in the model formulation:

$$\text{SUBi} = \begin{cases} 1 & \text{if the house is located in subcenter } i; i = 1, 2 \\ 0 & \text{otherwise} \end{cases}$$

In addition, a natural hypothesis is that housing prices vary systematically with distance from those subcenters, even in a model formulation where regional labor market accessibility is accounted for. Is here a similar attraction effect that was identified for the Stavanger cbd area? Is it possible that households, like firms, are attracted to centers through some kind of agglomeration effects, for instance related to the probability of having matching neighbors? Such hypothesis and questions motivate our modeling alternative LM1:

LM1: The basic model (BM) extended by two dummy variables (SUB1 and SUB2) representing the presence of the two subcenters, and corresponding variables (SUB1DIST and SUB2DIST) representing traveling times within a specific cutoff value of 20 minutes from the subcenters SUB1 (Bryne) and SUB2 (Egersund).

The choice of a cutoff value of 20 minutes is a result of experiments with several alternative values. The cutoff value represents the distance where the influence of the subcenter on housing prices is no longer noticeable. We have also experimented by specifying Sandnes as a subcenter in the model. Our results indicate, however, that the center of Sandnes is an integrated part of the Stavanger urban area, and that this subcenter is adequately represented by the spatially defined variables in the basic model.

Another hypothesis is that housing prices is systematically higher nearby the administrative center in a municipality than elsewhere. This hypothesis can be motivated from the possibility that households on average find it attractive and convenient to reside close to services offered by local authorities. A modeling alternative corresponding to this hypothesis is:

LM2: The basic model (BM) extended by a dummy variable (ADMCENTER) representing the administrative center of a municipality.

The dummy variable is defined by

$$\text{ADMCENTER} = \begin{cases} 1 & \text{if zone is the administrative center of its municipality} \\ 0 & \text{otherwise} \end{cases}$$

Cumulative opportunities of employment

As mentioned above households might prefer residential locations with favorable job opportunities in the close neighborhood, since short journeys-to-work facilitates the logistics of running a household. One hypothesis is that this effect can be represented by a simple cumulative opportunities measure of accessibility, for instance defined by the number of job opportunities reached within a travel time by car of 5 minutes. Ideally, the measure should reflect the probability of receiving relevant job offers, capturing both the labor market turnover (vacancies) and the diversity of job opportunities. The number of jobs within an area represents, of course, only a rough proxy variable of the relevant labor market situation in alternative areas, but we doubt that the payoff in form of more significant results is reasonably related to the considerable amount of data collection required to study the matters in more detail. Hence, the modeling alternative M10 is implicitly based on the assumption that the number of local jobs adequately represents the relevant labor market situation.

Another data-driven simplifying assumption is related to our aggregate subdivision of the geography into rather wide-spreading zones. This complicates a confident specification of employment rings corresponding to a specific traveling times from alternative locations. As an alternative measure of the local labor market situation we have instead used the intrazonal employment:

LM3: The basic model (BM) extended by a variable (JOBS) representing the number of jobs within the zone.

Population density

Though our subdivision of the geography into zones results in a relatively high degree of interzonal homogeneity, some characteristics can be expected to vary systematically across zones, and this might influence the attractiveness and housing prices of a zone. The presence of higher level schools, centers for physical training and shopping might for instance contribute positively to the attractiveness of an area for residential location. Such facilities are of course not equally spread across the zones. A reasonable hypothesis is that their presence is positively related to the population density. The population density might in principle be represented by the number of workers residing within rings of a specific traveling time from a location. Once again, however, the use of such a simple measure is complicated by our aggregate subdivision of the geography into zones. Instead, we test the hypothesis that the population within a zone affects the housing prices. We assume that population is represented by the number of workers residing within a zone.

LM4: The basic model (BM) extended by a variable (POPULATION) measuring the number of workers residing in a zone.

The number of jobs per worker

We have argued that the attractiveness of a location for residential purposes depends on the probability of receiving relevant job offers locally. Due to distance deterrence effects in the job-search procedure and to costs related to the journey-to-work it further can be argued that this probability depends positively on the number of jobs per inhabitant within a zone. This hypothesis is examined through the following model formulation

LM5: The basic model (BM) extended by a variable measuring the number of jobs per worker (BALANCE) residing within a zone.

Relative local labor market accessibility

As pointed out by Guiliano and Small (1991) local subcenters can also be identified through gravity based measures of accessibility. Analogously, we characterize the labor market position of a zone through a measure of relative accessibility. Let

$$g(i,j) = \begin{cases} 1 & \text{if zone } i \text{ and zone } j \text{ have a common boundary} \\ 0 & \text{otherwise} \end{cases}$$

$$n(i) = \text{ \# zones with a boundary common to zone } i$$

and

$$Z(j) = [j : g(i,j) = 1]$$

where $Z(j)$ = the set of zones with a boundary common to zone i

The relative accessibility of a zone is then defined by:

$$\text{RELACC}_i = \frac{S_i}{\frac{1}{n(i)} \sum_{j \in Z(j)} S_j} \tag{5}$$

where S_i is the labor market accessibility of a zone, as defined by Equation (3). A high value of this measure means that the corresponding zone has a high local labor market accessibility. Model LM5 is introduced through the ambition of testing whether this contributes positively to explain variation in housing prices:

LM6: The basic model (BM) extended by the variable RELACC_i, reflecting local variations in labor market accessibility.

This measure of local labor market accessibility can also be defined relative to specific areas of the geography. The most central parts of the region represent one such area, including the Stavanger urban area as well as lower rank central places and suburban communities in the municipalities surrounding Stavanger (Stavanger, Sola, Randaberg, and Sandnes, see the map in Figure 1, and relevant data in Appendix A). Those four most centrally located municipalities are denoted as the urban area. The rural area represents four municipalities (Sokndal, Lund, Bjerkreim, and Gjesdal) in the hinterland in the southern parts of the region, where the ratio of inhabitants to open land is considerably lower than in other municipalities. The remaining zones are neither located in the most urban nor in the most rural parts of the region, and define a semi-urbanized area. The three subareas represents a natural subdivision of the region into clusters of adjacent and reasonably similar zones. Each area is identified through a dummy variable, like for instance:

$$\text{URBAN}(i) = \begin{cases} 1 & \text{if zone } i \text{ belongs to the most urban parts of the region} \\ 0 & \text{otherwise} \end{cases}$$

The variables RURAL(i) and SEMI(i) are similarly defined, and the corresponding areas are defined by $U(i) = [i : \text{URBAN}(i) = 1]$, $R(i) = [i : \text{RURAL}(i) = 1]$, and $SU(i) = [i : \text{SEMI}(i) = 1]$. Let

$$n(\text{U}) = \# \text{ the number of zones within the urban area}$$

The relative local labor market accessibility of zone i within the urban area is defined by:

$$\text{RELACC(U)}_i = \frac{\text{RELACC}_i}{\frac{1}{n(\text{U})} \sum_{j \in U(j)} \text{RELACC}_j} \cdot \text{URBAN}(i) \tag{6}$$

RELACC(R)_i and RELACC(SU)_i are similarly defined as the relative local labor market accessibility of zone i within the rural and the semi-urbanized zones, respectively. This specification complies to the idea that a local measure of labor market accessibility should refer to the location within a subarea rather than the entire region. The basic hypothesis is that the residential preferences of households might be in favor of a particular kind of area, like an urban area, a rural area, or a semi-urban area, and that high accessibility to job opportunities on average is considered as an attractive location attribute within this area. This suggests that the parameter estimates corresponding to the area-specific accessibility measures are positive. The corresponding model formulation is represented by

LM7: The basic model (BM) extended by variables reflecting local variations in labor market accessibility within specific subareas of the region.

The alternative accessibility measures are introduced log-linearly in the corresponding hedonic regression models. Referring to model LM7 as an example this means that the hedonic regression formulation is given by:

$$\begin{aligned}\log P_{it} \quad = \quad & \beta_0 + \beta_1 \log \text{LOTSIZE}_i + \beta_2(\text{RUR} \log \text{LOT})_i + \beta_3 \log \text{AGE}_i + \beta_4(REBUILD \log \text{AGE})_i + \\ & + \beta_5 \text{GARAGE}_i + \beta_6 \log LIVAREA_i + \beta_7 \log \text{NUMBTOIL}_i + \beta \log \text{TIMECBD}_i + \\ & + \beta_q(\log \text{TIMECBD}_i)^2 + \beta_8 \log \text{ACCESSIBILITY}_i + \beta_9 \log \text{RELACC(U)}_i + \\ & + \beta_1 0 \log \text{RELACC(R)}_i + \beta_1 1 \log \text{RELACC(SU)}_i + \sum_{t=97}^{01} \beta_t \text{YEARDUM}t_i + \epsilon_{it} \qquad (7)\end{aligned}$$

where $\log(\cdot)$ denotes the natural logarithm, and ϵ_{ij} is the error of disturbance for a specific observation.

Finally, we have tested model formulations combining several of the proposed local measures of spatial structure:

LM8: The basic model (BM) extended by several characteristics of local spatial structure.

The results are presented in Table 2. Contrary to for instance Adair et al. (2000) and Handy and Niemeier (1997) all parameters are estimated simultaneously rather than through a stepwise procedure, where values of the accessibility measure are estimated from commuting flow data before they enter into the hedonic housing model.

5 Results

In this section we present estimation results based on the alternative model formulations that were proposed in the preceding section. We also search for possible local characteristics through a data-mining semi-parametric approach.

5.1 An empirical evaluation of the alternative model formulations

The analysis to follow is based on the use of pooled cross section data. This explains the introduction of the time-dummies in our models. The advantage of this procedure is that it enables an increase in sample size, and greater variations in the independent variables.

Results from the experiments with measures of the local spatial structure are presented in Table 2. Consider first the results based on LM1. Compared to the basic model (BM) all the measures of explanatory power are improved, but the changes are not very convincing. Still, the value of the likelihood ratio test statistic is $(2 \cdot (314,21 - 296,79) \approx)$ 34,8, which exceeds the critical value of a chi square distribution with 4 degrees of freedom at the 5 percent significance level. According to the results corresponding to LM1 in Table 2, the presence of subcenters has ambiguous effects on house prices. The partial impact of a location in Bryne is estimated to be positive, but the effect is not significant at the 5% level. The estimated partial effect of a location in Egersund is, on the other hand, significantly negative. In interpreting this result, remind that effects of job concentrations are accounted for through the labor market accessibility measure. It also follows that the position of Egersund as a center in the southern parts of the region is reflected in the parameter estimate corresponding to the variable SUB2DIST.

The estimated marginal impact of changes in the variables SUB1 and SUB2 has the intuitively expected sign; housing prices are estimated to be significantly negatively related to variations in traveling time within the cutoff value of 20 minutes traveling time from the two subcenters. Notice from Table 2 that the estimated effect of variations in distance is considerably larger for Egersund (SUB2DIST) than for Bryne (SUB1DIST). This is a reasonable result. Bryne is surrounded by smaller centers of a lower rank, while Egersund is a center for a more rural area in a considerably longer distance from the central parts of the region. The Egersund area to a larger degree represents a separate housing submarket, while the housing market in the Bryne area is more influenced by the situation in the cbd of the region. The coefficient related to SUB1DIST reflects a very marginal effect of variations in distance on housing prices. The estimate implies that the price of a standard house falls by about ... NOK from the center of Bryne to a location 20 minutes from this center. For Egersund the estimate implies a corresponding reduction of about ... NOK.

The introduction of the variables representing the subcenters into the model formulation does only lead to marginal changes in most of the remaining parameter estimates. The parameters that are relatively most sensitive to the model extension are β and β_q and the parameter attached to the accessibility measure. This is not surprising, considering the fact that the two additional variables capture effects of spatial structure characteristics. If such characteristics are not accounted for in the model formulation an estimation bias will result, especially for parameters reflecting effects of spatial separation and spatial structure. Notice in particular that the effect of the quadratic term in the function representing distance from the cbd becomes redundant in the case where relevant local structure attributes are taken explicitly into account. In such a case the flexibility of this function is not required ... capture local anomalies in .

According to our estimation results the dummy variable ADMCENTER in general has no significant influence on housing prices in the region; the corresponding parameter estimate is 0,0134, with a standard error of 0,0124, and the adjusted R^2 resulting from this model specification is 0,7396. We have not, however, included the results based on this general formulation of the model in Table 2. Based on an inspection of residuals we found a tendency that the basic model underpredicts house prices in the 5 most centrally located administrative centers outside Stavanger (the centers of Sola, Randaberg, Sandnes, Bryne, and Gjesdal). As a result of this data-mining procedure we reached a model specification performing better than a general representation of the variable ADMCENTER. This is the model specification underlying LM2 in Table 2. According to the table the variable ADMCENTER contributes significantly to explain spatial variation in house prices, and all the goodness-of-fit indices come out with marginally more satisfying values. Hence, the conclusion is that a dummy variable representing a subset of accessible administrative centers contributes to reveal systematic spatial variation in house prices.

The results based on the variable JOBS, representing the number of jobs within a zone, are neither very encouraging. This variable is not found to influence housing prices significantly, and it does not lead to a significantly improved goodness-of-fit (see the results based on LM3 in Table 2). Such results do not mean, however, that we can jump to the conclusion that the local supply of jobs does not influence housing prices. We can of course not ignore the possibility that the results are due to our specification of the local labor supply, through the variable JOBS. We have carried through experiments with adjustments in this specification, both by restricting the variable to groups of municipalities, and by including adjacent zones in the measure. None of those experiments offered more encouraging results than LM3. Still, we can of course not rule out the possibility that a more disaggregate subdivision of the geography and/or a more detailed description of job categories could lead to the conclusion that the local labor market situation affect housing prices significantly.

It follows from Table 2 that the results based on LM5 give no support for the hypotheses that

Table 2: Results based on alternative specifications of local spatial structure characteristics

	BM	LM1	LM2	LM3	LM4	LM5	LM6	LM7	LM8
Constant	11,1835	11,1318	11,2885	11,2320	11,22391	11,1874	11,1874	11,2272	11,3366
	(0,1687)	(0,1819)	(0,1711)	(0,1722)	(0,1681)	(0,1687)	(0,1695)	(0,1685)	(0,2058)
LOTSIZE	0,1308	0,1302	0,1294	0,1320	0,1301	0,1326	0,1303	0,1336	0,1332
	(0,0099)	(0,0100)	(0,0100)	(0,0099)	(0,0099)	(0,0100)	(0,0100)	(0,0102)	(0,0103)
RURLOT	-0,0271	-0,0304	-0,0303	-0,0273	-0,0274	-0,0271	-0,0270	-0,0972	-0,1004
	(0,0031)	(0,0031)	(0,0035)	(0,0031)	(0,0031)	(0,0031)	(0,0031)	(0,0298)	(0,0301)
AGE	-0,0849	-0,0839	-0,0856	-0,0854	-0,0848	-0,0853	-0,0849	-0,0848	-0,0842
	(0,0066)	(0,0065)	(0,0066)	(0,0067)	(0,0066)	(0,0067)	(0,0066)	(0,0066)	(0,0065)
AGE·REBUILD	0,0104	0,0104	0,0106	0,0104	0,0104	0,0104	0,0105	0,0107	0,0106
	(0,0029)	(0,0029)	(0,0029)	(0,0029)	(0,0029)	(0,0029)	(0,0029)	(0,0029)	(0,0029)
GARAGE	0,0645	0,0644	0,0636	0,0646	0,0634	0,0653	0,0645	0,0638	0,0629
	(0,0108)	(0,0108)	(0,0108)	(0,0108)	(0,0109)	(0,0109)	(0,0108)	(0,0108)	(0,0108)
LIVAREA	0,3552	0,3554	0,3564	0,3562	0,3564	0,3560	0,3551	0,3536	0,3550
	(0,0177)	(0,0176)	(0,0175)	(0,0177)	(0,0177)	(0,0177)	(0,0177)	(0,0177)	(0,0175)
NUMBTOIL	0,1475	0,1473	0,1482	0,1474	0,1474	0,1474	0,1476	0,1456	0,1451
	(0,0146)	(0,0145)	(0,0146)	(0,0146)	(0,0146)	(0,0145)	(0,0146)	(0,0145)	(0,0145)
β (quadratic)	-0,1095	-0,1352	-0,1181	-0,1059	-0,1074	-0,1087	-0,1158	-0,1381	-0,1506
	(0,0218)	(0,0268)	(0,0217)	(0,0220)	(0,0219)	(0,0218)	(0,0250)	(0,0280)	(0,0280)
β_q (quadratic)	-0,0104	-0,0017	-0,0102	-0,0134	-0,0108	0,0111	-0,0081	-0,0011	0,0000
	(0,0053)	(0,0077)	(0,0053)	(0,0056)	(0,0056)	(0,0053)	(0,0069)	(0,0082)	(0,0083)
ACCESSIBILITY	0,0776	0,0844	0,0684	0,0688	0,0631	0,0754	0,0825	0,0839	0,0659
	(0,0159)	(0,0181)	(0,0160)	(0,0173)	(0,0170)	(0,0160)	(0,0179)	(0,0182)	(0,0202)
SUB1	-	0,0386	-	-	-	-	-	-	0,0550
	(-)	(0,0233)	(-)	(-)	(-)	(-)	(-)	(-)	(0,0299)
SUB1DIST	-	-0,0140	-	-	-	-	-	-	-0,0055
	(-)	(0,0057)	(-)	(-)	(-)	(-)	(-)	(-)	(0,0066)
SUB2	-	-0,0645	-	-	-	-	-	-	-0,0213
	(-)	(0,0329)	(-)	(-)	(-)	(-)	(-)	(-)	(0,0515)
SUB2DIST	-	-0,1351	-	-	-	-	-	-	-0,1349
	(-)	(0,0452)	(-)	(-)	(-)	(-)	(-)	(-)	(0,0455)
ADMCENTER	-	-	0,0359	-	-	-	-	-	0,0100
	(-)	(-)	(0,0130)	(-)	(-)	(-)	(-)	(-)	(0,0173)
JOBS	-	-	-	0,0041	-	-	-	-	-
	(-)	(-)	(-)	(0,0036)	(-)	(-)	(-)	(-)	(-)
POPULATION	-	-	-	-	0,0126	-	-	-	0,0067
	(-)	(-)	(-)	(-)	(0,0067)	(-)	(-)	(-)	(0,0080)
BALANCE	-	-	-	-	-	0,0027	-	-	-
	(-)	(-)	(-)	(-)	(-)	(0,0033)	(-)	(-)	(-)
RELACC	-	-	-	-	-	-	-0,0441	-	-
	(-)	(-)	(-)	(-)	(-)	(-)	(0,0913)	(-)	(-)
RELACC(U)	-	-	-	-	-	-	-	-0,0947	-0,0682
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(0,0916)	(0,1567)
RELACC(SU)	-	-	-	-	-	-	-	-0,1232	-0,1100
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(0,0958)	(0,1592)
RELACC(R)	-	-	-	-	-	-	-	0,3273	0,3505
	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(0,1949)	(0,2195)
YEARDUM97	-0,1362	-0,1366	-0,1357	-0,1360	-0,1364	-0,1361	-0,1363	-0,1372	0,1375
	(0,0135)	(0,0135)	(0,0135)	(0,0135)	(0,0135)	(0,0134)	(0,0135)	(0,0134)	(0,0135)
YEARDUM99	0,1297	0,1326	0,1294	0,1296	0,1283	0,1300	0,1296	0,1299	0,1318
	(0,0136)	(0,0134)	(0,0136)	(0,0136)	(0,0137)	(0,0136)	(0,0136)	(0,0136)	(0,0134)
YEARDUM00	0,2700	0,2717	0,2701	0,2703	0,2699	0,2700	0,2698	0,2694	0,2713
	(0,0135)	(0,0134)	(0,0135)	(0,0135)	(0,0135)	(0,0135)	(0,0135)	(0,0134)	(0,0134)
YEARDUM01	0,3030	0,3033	0,3032	0,3045	0,3025	0,3035	0,3028	0,3029	0,3032
	(0,0136)	(0,0136)	(0,0135)	(0,0136)	(0,0136)	(0,0136)	(0,0135)	(0,0136)	(0,0136)
n	2788	2788	2788	2788	2788	2788	2788	2788	2788
R^2	0,7409	0,7441	0,7415	0,7410	0,7412	0,7410	0,7409	0,7419	0,7453
R^2-adj.	0,7396	0,7424	0,7401	0,7396	0,7398	0,7396	0,7395	0,7403	0,7431
L	296,79	314,21	300,26	297,39	295,76	297,29	296,91	301,95	320,48
APE	215144	214320,	214999	215235	215044	215203	215178	214744	213875
SRMSE	0,2046	0,2038	0,2045	0,2046	0,2044	0,2047	0,2046	0,2045	0,2036
White test statistic	281,47	324,22	287,47	292,10	298,07	331,49	296,87	329,77	409,25
Moran's I	0,0017	???	0,0025	0,0014	0,0023	0,0015	0,0013	0,0016	???
Standard normal deviate (z_I)	1,4374	???	1,9667	1,3604	1,8021	1,3261	1,3383	1,6282	???
Ramsey reset test (p-value)	0,8572	0,8554	0,8268	0,8755	0,8428	0,8845	0,8552	0,8445	0,8500
VIF, average value	5,83	7,66	5,62	6,13	5,91	5,56	8,16	31,45	40,43

Note: Results based on observations from the period 1997-2001, robust standard errors in parentheses. For all models involving local measures of spatial structure the values of the parameters σ_e and γ_e in Equation 3 are assumed to be given, equal to the values resulting from the estimation of the basic model ($\sigma_e = -0,1088$ and $\gamma_e = 1,0963$). Besides R^2 (and the adjusted R^2) we have included the log-likelihood value (L), the Average Prediction Error (APE $= \frac{\sum_i (|\hat{P}_i - P_i|)}{n}$, where $\hat{P}_i$ is the predicted price of house i, and n is the observed number of houses), and the Standardized Root Mean Square Error (SRMSE).

housing prices are affected by the intrazonal balance between workers and jobs. The relevant parameter estimate reflects only a marginal effect, and it is not significantly different from zero. The introduction of this variable does not lead to a significant increase in the goodness-of-fit, and it has practically no impact on the evaluation of other variables. The results are a bit more encouraging for LM4, corresponding to the hypothesis that house prices are affected by the intrazonal population. The relevant parameter is estimated to be positive, with a p-value of 0,053, but also this estimate reflects a relatively marginal effect. Once again, we have experimented with a large number of alternative model formulations incorporating the basic ideas underlying LM4 and LM5, without finding results worth reporting. We have for instance experimented with variables adjusting for variations in the spatial extension of the zones.

The results based on LM6 offer no support for the hypothesis that a high local labor market accessibility (measured by the variable RELACC in Table 2) contributes to the housing prices. This conclusion is somewhat modified in the case where the geography is subdivided into separate areas, represented by LM7 in Table 2. For urban and semi-urban areas the relevant parameter estimates are negative, contradicting the hypothesis that high local accessibility to job opportunities is considered to be attractive. This might for instance be due to negative externalities of residing close to industrial areas. The parameter estimates are not significantly different from zero, however, it is possible that we estimate the net effect of forces pulling in separate directions. For rural areas, on the contrary, our parameter estimate indicate that the local labor market accessibility (RELACC(R)) contributes positively to explain variations in housing prices. It is intuitively reasonable that households value local labor market accessibility especially in areas with a long distance to job opportunities in other parts of the region. The relevant parameter estimate of about 0,33, but it is not found to be significantly different from 0 at the 5% level of significance. As a measure of the accuracy of this parameter estimate the corresponding 95% confidence interval is (-0,05, 0,71), while the 95% confidence intervals of the parameter estimates related to RELACC(U), and RELACC(SU) are (-0,27, 0,08) and (-0,31, 0,06), respectively. In evaluating the accuracy of the parameter estimates, keep in mind that the number of observations is considerably lower in rural than in the other areas. The lack of significant results might also reflect the presence of harmful multicollinearity.

As could be expected, the denominators in the definition of local accessibility measures (see

Equation (6)), have approximately the same value in the three sub-areas. Hence, estimation results only change marginally if this average value is set equal to 1 for each sub-area. Notice also from Table 2 that the quadratic term in the distance deterrence function does not contribute significantly to explain housing prices in the case where local labor market accessibility is separately accounted for in the three kinds of subareas. It also follows from Table 2 that LM7 does not significantly contribute with a goodness-of-fit exceeding the results based on BM. The value of the likelihood ratio test statistic is 5,21, which means that the hypothesis that we cannot reject the hypothesis that the relevant model extension does not contribute to explain the spatial variation in house prices. The reported positive loglikelihood values are explained from the fact that the logarithm of house prices defines a function that is very flat for the relevant range of values, with correspondingly small variance (see Osland et al. 2005). As another test of the model extension, we obtain a F-statistic of .... for the null hypothesis that the three local measures of accessibility jointly have no effect on house prices. This value does not exceed the critical value of the $F_{3,2770}$ distribution, hence we cannot reject the null hypothesis.

Finally, local characteristics of spatial structure are combined in a more general model formulation. We have experimented by many combinations of variables. The set of characteristics underlying model LM8 in Table 2 is based on the selection of characteristics that proved to contribute significantly, or nearly significantly, in separate representations of local spatial structure. It follows from the table that parameter estimates do not change considerably compared to the experiments with separate representations of the variables. The standard errors, however, are inflated, probably by the presence of multicollinearity. Even if the distance from subcenter 2 (SUB2DIST) is the only local characteristic that contributes significantly to explain spatial variations in house prices, all goodness-of-fit measures are improved compared to the alternative model specifications. Notice in particular that the value of the likelihood ratio test statistic is $(2 \cdot (320,48 - 314,21) \approx)$ 12,5 when LM8 is compared to LM1. This exceeds the critical value (11,07) of a chi square distribution with 5 degrees of freedom at the 5 percent level of significance.

The VIF-values reported in Table 2 indicate how much the variances of the estimated coefficients are inflated by multicollinearity. Kennedy (2003) suggests that VIF>10 indicates harmful collinearity. Hence, it follows from our results that the introduction of local measures of labor market accessibility (in LM7) potentially causes harmful collinearity, reflecting the fact that a considerable part of the sample variation in the relevant accessibility variables are explained by the other independent variables in the hedonic regression model. This might be one reason why parameter estimates related to local labor market accessibility in urban and semi-urban areas have not come out with statistically significant signs, despite our large number of observations.

According to the reported values of the White test statistic the hypothesis of homoscedasticity is rejected in all model specifications in Table 2. Still, we find that the robust estimator of variance produce results that deviate only marginally from estimates based on the ordinary least square estimator.

The positive values of the Moran's I indicate positive spatial autocorrelation. The Moran's I is calculated from a binary row standardized weight matrix, see for instance Anselin (2002), where zones are defined as neighbors if they have a common border. All houses within a zone are also neighbors, while a house is not a neighbor to itself. The standard normal deviate z_I is constructed from values of the mean and the variance of the Moran statistic (Anselin (1988). The null hypothesis of no spatial autocorrelation in the residuals is rejected at the 5% significance level if $z_I > 1,645$. According to the results in Table 2 this hypothesis cannot be rejected for model LM1. This significant tendency of positive spatial autocorrelation disappears when subcenter 1 (Bryne) is omitted from the model specification. Without entering into other details , this results in a model specification where $z_I = 1,5848$.

The Ramsey reset test is usually referred to as an omitted variable test (see for instance Davidson and MacKinnon (1993)), and is also used to detect incorrect functional form (see for instance Wooldridge (2002)). The reported p-values mean that we cannot, at any relevant level of significance, reject the hypothesis that the model is correctly specified.

5.2 An interpretation of the results based on the model formulation incorporating a measure of the relative local labor market accessibility

In comparing the results based on model LM7 to the results based on model BM in Table 2, notice first that the estimated impact of non-spatial attributes are relatively invariant with respect to the introduction of the local labor market accessibility measures. The estimated impact of spatially defined characteristics are not invariant in this respect, however. The estimate of the parameter related to RURLOT is considerably less accurate, and the estimated sensitivity of housing prices with respect to variations in regional accessibility and the distance from the cbd has changed as a result of collinearity between independent variables. The 95% confidence interval for β was, for instance, (..., ...) for model BM, and (..., ...) for model LM7.

Based on the model formulation BM, that incorporates both traveling time from the cbd and the gravity based labor market accessibility measure S_j, Osland and Thorsen (2005) suggested a distinction between an urban attraction effect and a labor market accessibility effect. The effect of labor market accessibility is captured through the introduction of a gravity based accessibility measure, while the falling housing price gradient reflects households evaluation of urban amenities. The housing price gradients in Figure ??? are based on the specification of a standard house. The standard house is defined as not being rebuilt, it has a garage, it is not located in the rural areas, and the price refers to the year 2000. Lotsize, age, living area and the number of toilets are given by their average values. The values of the local and regional labor market accessibility indices are set equal to zero. The dashed line in the figure reflects the urban attraction effect in the case where parameter estimates are based on BM, while the solid line is based on LM7. The two lines represent a difference in predicted housing prices of .... 1998-NOK. Hence, the level of the urban attraction effect depends on the formulation of local accessibility measures. LM7 can be argued to offer a more reliable prediction of the urban attraction effect, since the parameter estimates resulting from BM is biased, due to the effect of omitted variables.

Similar considerations apply for the evaluation of the regional labor market accessibility effect. The two accessibility gradients in Figure ??? are based on the assumption that the standard house is located in the center of Stavanger. The dashed line in Figure ??? is based on M6, while the solid line refers to the case where the area-specific local accessibility measures are accounted for. The introduction of the area-specific accessibility measures has no significant impact on how regional accessibility to job opportunities are predicted to influence house prices.

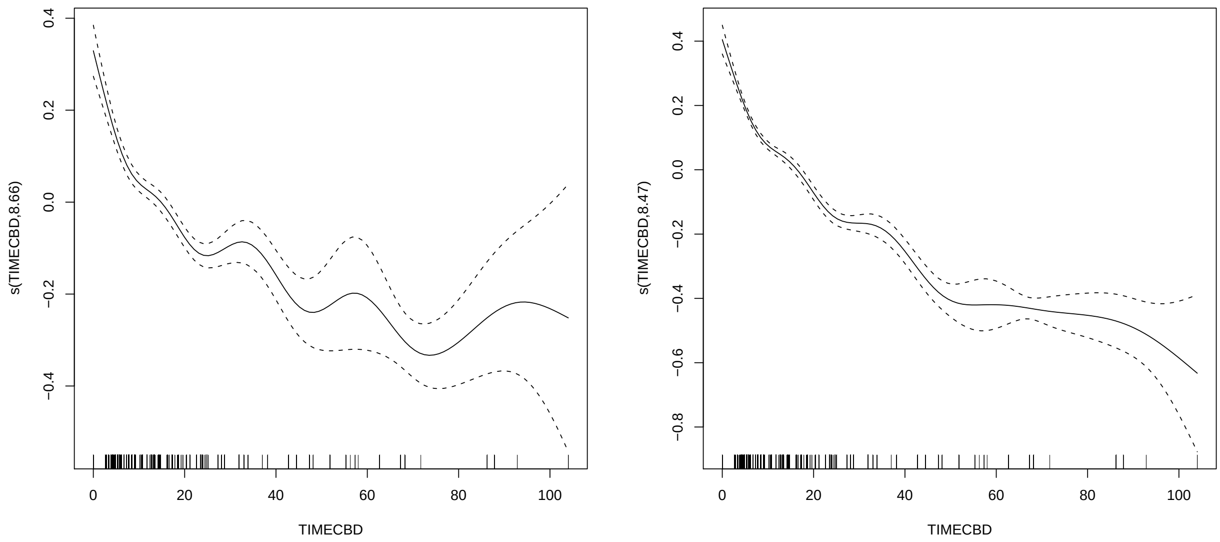
In rural areas the elasticity of house prices with respect to variations in the local labor market accessibility index is estimated to be 0,32, while the elasticities are estimated to be considerably lower, and negative, in urban and semi-urban areas. Figure ??? offers an illustration of how variations in this index are predicted to affect house prices within each of the sub-areas. The values of the index are represented on the horizontal axis. The lengths of the lines reflect the span of observed index values in the respective areas. Each line refers to a standard house, where the distance to the cbd and the value of the regional labor market accessibility are set equal to the observed mean value for the relevant sub-area. According to the figure .....

6 Results based on semi-parametric approaches

The results presented above are based on approaches where all the variables are represented through parametric specifications in the model formulation. As an alternative local peaks and valleys in housing prices can be identified in semi-parametric approaches, see for instance Clapp (2003). In this section we consider model formulations where the predictor TIMECBD is the only variable that enters through a non-parametric smooth function. This represents a very flexible approach, since it imposes no a priori parametric assumptions on the smoothed function. The method is hence useful when the aim is to study potential underlying parametric structure in the data. Venables and Ripley (1997) show that smoothing splines adapt better to general smooth curves compared to for instance polynomials or lowess.

Figure 3 illustrates the results of semi-parametric model formulations based on our data. The term $\beta \log \text{TIMECBD}$ in the model formulation (4) is substituted by a smoothing function $s(\text{TIMECBD})$. Part a) of the figure refers to a model formulation that corresponds to the basic model in all other respects than the specification of distance from the cbd, while the left part of the figure is based on a model formulation where labor market accessibility is not accounted for.

The plots in Figure 3 is estimated by using the mgcv (version1.3-12) package in R. This package uses a variant of generalized additive models (GAM), see for instance Hastie and Tibshirani (1990) for a comprehensive review. In this case, penalized regression smoothing splines are estimated, by way of maximum likelihood estimation. Penalizing implies a compromise between fit and amount of smoothness (Venables and Ripley 1997). The degree of smoothing is automatically chosen by generalized cross validation (GCV), see Wood (2000) and Wood (2001). This means that the estimated degrees of freedom for the smooth is chosen so that the GCV-score is minimized. The estimated degrees of freedom are indicated at the vertical axis of the plots in this subsection. The approach underlying Figure 3 allows a maximum of 10 degrees of freedom.



a) All variables in the basic model are incorporated

b) The variable representing labor market accessibility is ignored in the model formulation

Figure 3: Illustrations of semi-parametric approaches to estimate the relationship between (logarithmic) housing prices and the distance from the cbd. The distance from the cbd is represented by a smoothing function s(TIMECBD) in the model formulation.

Increasing this maximum to for instance 20 leads to plots with a similar pattern, except for the most peripheral areas of the region, where the plots become more irregular. This is primarily due to the relatively small number of observations from those areas, and we have not reported results based on such more flexible non-parametric representation of distance from the cbd.

In Figure 3 the solid lines represent the smooth function, that is the predicted value of the dependent variable as a function of variations in TIMECBD. The dashed curves delimit approximate 95% confidence intervals of the smooth function. Following Wood (2001), the smooth is given an average value of zero. The y-axis hence shows how this predictor causes the dependent variable to alter round its mean.

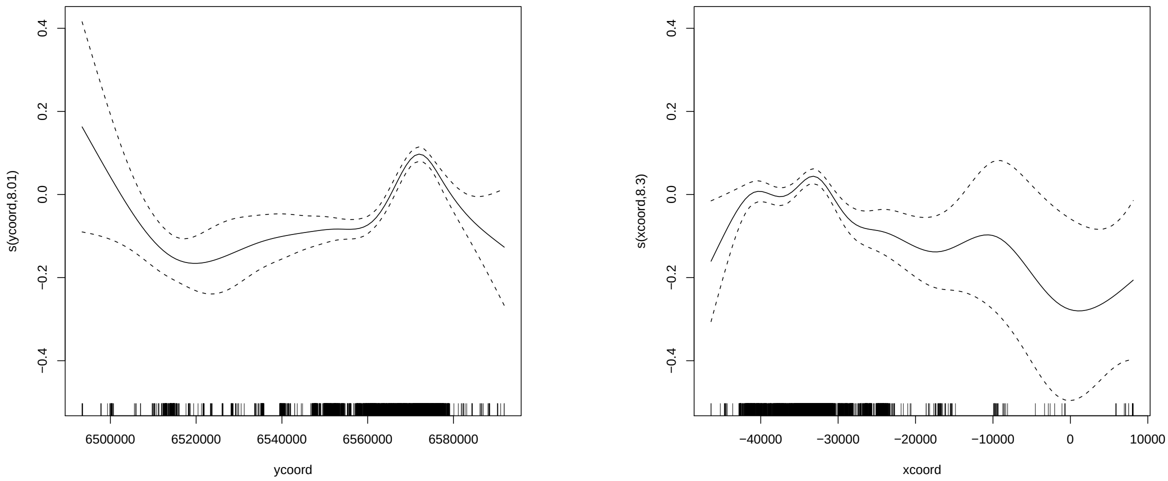
The semi-parametric approaches underlying the plots in Figure 3 have only marginal impact on the estimated coefficients related to the variables that still enter parametrically in the model specification. Compared to the basic model (BM) the adjusted R^2 increases somewhat in the semi-parametric model specification where all variables in the basic model are incorporated (from 0,7396 to 0,7410), while it is somewhat lower in the model specification where labor market accessibility is not accounted for (0,7390). The increased fit resulting from the GAM model comes at the expense of the degrees of freedom. As indicated at the vertical axis in part a) of Figure 3 the degrees of freedom used for the smoothing function are 8,66. This means that the degrees of freedom used for the GAM model is 21,66, while the number of parameters to be estimated in the basic model is 15. The GCV-score (see for instance Wood (2001)) is also slightly lower in the GAM model, but the differences are very small (BM: 0,0478, GAM: 0,0477). Notice that not even a flexible non-parametric representation of TIMECBD adds more to the explanatory power than a simple measure of labor market accessibility in a parametric approach.

According to the plot in part a) of Figure 3 the confidence is very narrow at locations close to the cbd, whereas the confidence bands are much wider for peripheral locations, where there are fewer observations, located further apart from each other. According to this plot a local peak seems to exist in a distance of around 32 minutes from the cbd (Bryne), while no other statistically significant irregularities are evident for the rest of the estimated path. Hence, the figure reveals no other clear hypotheses of local variables that should be included in an appropriate explanation of spatial variation in housing prices.

The gradient in the right part of Figure 3 incorporates both the urban attraction effect and the labor market accessibility effect, since labor market accessibility is not explicitly accounted for in the model formulation. In this case the irregularities to some degree are smoothed out, and the housing price gradient is predicted to be steeper than in the case where the labor market accessibility is accounted for through a separate measure. The distance from the cbd and labor market accessibility are strongly negatively correlated. It is intuitively reasonable that the gradient becomes more irregular and flatter, with a wider confidence band, in part a) of the figure, where some of the effect of variations in distance is captured through the introduction of the labor market accessibility measure.

As another data-mining attempt to search for possible local irregularities we have used a semi-parametric approach to study direction-specific housing price gradients. Figure 4 illustrates how housing prices vary in the south-north direction (y-coordinates), and the west-east direction (x-coordinates). The terms $\beta \log \text{TIMECBD}_i + \beta_q (\log \text{TIMECBD}_i)^2$ in the model formulation (4) are substituted by a smoothing function $s(\text{ycoord})$ in part a) of Figure 4, and by $s(\text{xcoord})$ in part b) of the figure. In all other respects the underlying model correspond to the basic model.

Part a) of the figure clearly reflects the urban attraction effect originating from Stavanger in the north. The peak in the south are explained by some outliers in the sparsely populated municipality of Sokndal. Those outliers probably are related to sales of houses in locations that are attractive for holiday purposes. In addition, however, the path corresponding to the southern parts of the region might also to some degree reflect adjustments to local variations originating from the center of Egersund. Part b) of the figure shows that house prices tend to fall somewhat when moving from the more densely populated areas in the west, to the more sparsely populated areas in the east.



a) Distance represented by s(ycoord); south-north   b) Distance represented by s(xcoord); west-east

Figure 4: Results based on a semi-parametric approach to study direction-specific variations in house prices.

7 Concluding remarks

The main result in this paper is that the incorporation of local spatial structure characteristics only marginally improves the goodness-of fit compared to the results following from a basic model where such local characteristics are not accounted for. In the basic model spatial structure is represented by two globally defined measures, and according to our results distance from the cbd and labor market accessibility capture most of the spatial variation in house prices. In fact, an adequate functional representation of the distance from the cbd results in satisfying values of goodness-of-fit indices, even if labor market accessibility is not explicitly accounted for through a separate variable. This does not mean that the labor market accessibility measure only marginally contributes to explain spatial variation in house prices. As reported in Osland and Thorsen (2005) the incorporation of this variable leads to a distinction between two substantial effects in the determination of house prices: the urban attraction effect and the labor market accessibility effect. A model specification where only distance from the cbd is accounted for is biased, despite the fact that this variable satisfactorily captures the aggregate impact of the two effects.

Similarly, local spatial structure characteristics might contribute to explain spatial variation in house prices, despite the fact that they do not improve the goodness-of-fit to an appreciable extent. According to our results the specification of subcenters outside the central parts of the region contributes significantly to explain spatial variation in house prices. Our results support a hypothesis that it is in particular important to account for subcenters that are located in a long distance from the central parts of the region. This corresponds to the hypothesis that the impact of variations in distance from the subcenter is positively related to the distance from the cbd. We also find that spatial variation in house prices is significantly influenced by a variable representing the administrative centers in the most centrally located municipalities of the region, and our results offer some indication that house prices are positively related to the size of the intrazonal population.

We have also proposed to account for the position of a zone through a measure of relative labor market accessibility. This measure is based on comparing the values of the labor market accessibility measure of a zone to the corresponding values in neighboring zones (with a common border). Our results on this measure give no support for the hypothesis that a high local labor market accessibility contributes positively to house prices. The results are somewhat more encouraging in the case where the geography is subdivided into an urban, a semi-urban, and a rural area, however. Still, none of the estimated parameters differ significantly from zero at the 5% level of significance. We still find this kind of local labor market accessibility measures to be appealing, and leave further experiments on other data sets for future research.

Even if the local variables introduced do not contribute considerably to an overall explanation of systematic spatial variation in house prices, they are potentially important if the ambition is to predict prices at specific locations, like for instance, Egersund.

As an alternative attempt to find possible systematic spatial variations in the house price gradient originating from the cbd, we have also experimented with semi-parametric approaches, where distance from the cbd is the only variable that enters through a non-parametric specification. Except from a identification of relevant subcenters those experiments do not suggest alternative local measures of spatial structure. We also find that the semi-parametric approaches do not outperform the parametric alternatives for our data sets. Still, we find semi-parametric approaches useful, for instance in revealing systematic direction-specific variations in house prices.

By studying the residuals we have identified a few zones where our models lead to considerable over/under predictions in house prices. One possible approach to improve goodness-of-fit is to introduce dummy-variables for such zones. We have refrained from such approaches, however, since our main ambition has been to identify how general, rather than location-specific, spatial structure characteristics affect house prices.

Despite some positive empirical findings in experimenting with local spatial structure characteristics, our results lead to the conclusion that distance from the cbd, in combination with a regionally defined labor market accessibility measure, explain the major part of systematic spatial variations in house prices. Our experiments support a hypothesis that variations in house prices are primarily due to a distinction between urban attraction and regional labor market accessibility, and they also support the hypothesis that the region we consider has a distinct monocentric pattern.

Appendix A

Table 3: Zonal data

Zone	Working population	Jobs	Observations	Relative access.
Rennesøy				
1	725	552	16	0,8946
2	98	24	4	0,9346
3	354	145	5	0,9267
4	127	23	4	0,9388
Randaberg				
5	3748	2195	89	1,0403
Stavanger				
6	328	4961	12	1,1390
7	95	4058	1	1,1331
8	769	1736	11	1,1140
9	688	1586	36	1,1322
10	1021	328	47	1,1343
11	1177	1630	41	1,1292
12	863	3905	23	1,1245
13	1125	1398	21	1,1277
14	555	2339	34	1,1319
15	1274	2864	41	1,1214
16	1382	396	26	1,1138
17	1518	4695	8	1,1262
18	1151	2141	29	1,1032
19	1750	407	47	1,0856
20	1637	392	16	1,1254
21	1777	1751	102	1,1029
22	2367	1627	40	1,1029
23	1340	627	45	1,1057
24	959	226	33	1,1018
25	846	271	16	1,1202
26	1042	341	27	1,1028
27	1001	132	23	1,1021
28	997	254	46	1,0930
29	1662	239	42	1,0777
30	945	1746	29	1,0707
31	1212	630	28	1,1118
32	2436	11309	10	1,1154
33	1719	529	44	1,0937
34	760	930	24	1,1147
35	240	583	4	1,0925
36	999	101	35	1,0677
37	919	147	28	1,0703
38	284	14	14	1,0622
39	1106	338	16	1,0550
40	1169	110	22	1,0506
41	4674	968	135	1,0642
42	237	37	13	0,7849
43	92	11	1	0,8779
Sola				
44	893	83	34	1,0961
45	2925	6178	70	1,0825
46	945	115	34	1,0902
47	497	63	22	0,9935
48	514	131	11	1,0236
49	2681	5423	74	1,0519
Sandnes				
50	1215	4870	22	1,1073
51	1338	1506	43	1,0900
52	1090	218	16	0,9432
53	371	147	8	1,0458
54	1383	240	57	0,9348
55	1150	302	40	0,9308
56	543	214	4	1,0501
57	788	6151	25	1,1017
58	1592	570	55	1,1014
59	651	1515	10	1,0871
60	678	207	19	1,1012
61	1280	175	10	1,0795
62	1911	307	53	1,0795
63	966	1355	23	1,1012
64	824	537	21	1,0830
65	737	276	6	1,0627
66	1010	787	22	1,0684
67	979	380	21	1,0670
68	914	49	10	1,0746
69	960	574	25	1,0791
70	1198	477	23	1,0474
71	942	253	13	1,0180
72	668	240	24	1,0245
73	21	3	3	0,5834
Klepp				
74	429	158	5	0,9335
75	3034	2043	72	1,0093
76	1047	1502	16	1,0111
77	340	208	2	0,9911
78	1457	457	10	1,0015
Gjesdal				
79	3354	1760	129	1,0046
80	336	184	16	0,8392
81	362	353	1	0,6896
Time				
82	5148	4343	93	0,9792
83	383	123	5	0,9036
84	1457	457	27	1,0015
Hå				
85	1493	1106	35	0,8704
86	1021	525	12	0,8149
87	348	81	6	0,7830
88	376	289	10	0,7491
89	2795	2511	62	0,9074
Bjerkreim				
90	395	213	8	0,7926
91	540	511	8	0,8143
Eigersund				
92	4612	4830	148	0,8825
93	367	97	7	0,7448
94	342	106	1	0,7472
Lund				
95	742	920	10	0,7219
96	235	45	2	0,5864
97	152	53	1	0,6349
Sokndal				
98	1125	916	21	0,7294
99	17	1	3	0,5308

Note: The relative accessibility is found by dividing S_j (see Equation 3) by the mean value of this measure for all the zones.

Appendix B

Table 4: Descriptive housing market statistics

	OBSERVATIONS	MEAN	STD. DEV.	MINIMUM	MAXIMUM
STAVANGER					
SALEPRICE	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					
SALEPRICE	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					
SALEPRICE	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					
SALEPRICE	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, 2002, *Under the hood". Issues in the specification and interpretation of spatial regression models*, *Agricultural Economics*, **27**, 247-267.

[5] Clapp, J M, 2003, “A semiparametric Method for Valuing Residential Locations: Application to Automated Valuation”, *Journal of Real Estate Finance and Economics*, 27:3, 303-320.

[6] Davidson R and J G MacKinnon, 2003, *Estimation and inferences in econometrics*, Oxford University Press.

[7] 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.

[8] Fotheringham A S, 1983, “A new set of spatial-interaction models: the theory of competing destinations”, *Environment and Planning A*, **15** 15-36.

[9] Fotheringham A S, 1988, “Consumer store choice and choice-set definition”, *Marketing Science* **7** 299-310.

[10] Gjestland, A, I Thorsen, and J. Ubøe, 2006,

[11] Guiliano, G, and K A Small, 1991, "Subcenters in the Los Angeles region”, *Regional Science and Urban Economics*, **21**, 163-182.

[12] Handy, S L, 1993, "Regional versus local accessibility: implications for nonwork travel”, *Transportation Reasearch Record*, 1400, 58-66.

[13] Handy, S L, and D A Niemeier, 1997, "Measuring accessibility: an exploration of issues and alternatives", *Environment and Planning A*, **29**, 1175-1194.

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

[15] Hastie, T and R J Tibshirani, 1990, *Generalized Additive Models*, London: Chapman and Hall.

[16] Heikkila 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.

[17] Kennedy P, 2003, *A Guide to Econometrics*, The MIT Press.

[18] Li M M and H J Brown, 1980, "Micro-neighbourhood externalities and hedonic prices", *Land Economics* **56** 125-140.

[19] McDonald, J F, 1987, "The identification of urban employment subcenters", *Journal of Urban Economics*, **21**, 242-258.

[20] Osland, L, I Thorsen, and J P Gitlesen, 2005, "Housing price gradients in a geography with one dominating center", Working papers in economics, Department of Economics, University of Bergen, no. 06/05.

[21] Osland, L and I Thorsen, 2005, "Effects of housing prices of urban attraction and labor market accessibility", Working papers in economics, Department of Economics, University of Bergen, no. 17/05.

[22] Pellegrini P A, Fotheringham A S, 1999, "Intermetropolitan migration and hierarchical destination choice: a disaggregate analysis from the US Public Use Microdata Samples", *Environment and Planning A* **31** 1093-1118

[23] Quigley J M, 1982, "Nonlinear budget constraints and consumer demand: An application to public programs for residential housing", *Journal of Urban Economic* **12**, 177-201.

[24] Richardson, H W,1988, "Monocentric vs. Policentric Models: The Future of Urban Economics in Regional Science" *The Annals of Regional Science*, vol. 22, issue 2, 1-12

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

[26] Theriault M, F Des Rosiers, and F Joerin, 2003, ”Modelling accessibility to urban services using fuzzy logic: a comparative analysis of two methods”, Paper submitted for presentation at the 10th European Real Estate Society Conference, Helsinki, Finland.

[27] Thorsen, I, and Gitlesen J P, 2000, “A competing destinations approach to modeling commuting flows: a theoretical interpretation and an empirical application of the model”, *Environment and Planning A* **32** 2057-2074

[28] Venables, W N and B D Ripley, 1997, *Modern Applied Statistics with S-Plus*, Springer

[29] Waddell, P, B J L Berry, and I Hoch, 1993, ”Residential property values in a multinodal urban area: new evidence on the implicit price of location, *Journal of Real Estate Finance and Economics*, **22** 829-833

[30] Wood, S N, 2000, ”Modelling and smoothing parameter estimation with multiple quadratic penalties”, *Journal of the Royal Statistical Society B*, 62, 413-428.

[31] Wood, S N, 2001, *mgcv: GAMs and Generalized Ridge Regression for R*, R News, Vol. ?.

[32] Wooldridge, J M, 2002, ”Econometric analysis of cross section and panel data”, The MIT Press.

[33] Yinger, J, 1979, ”Estimating the relationship between location and the price of housing”, *Journal of Regional Science*, **19**, 271-289.