The suitability of hedonic models for cost-benefit analysis: Evidence from commuting flows

Arnstein Gjestland *, David Philip McArthur †, Liv Osland ‡ and Inge Thorsen§

Abstract

In this paper we compare two estimates of benefits arising from the construction of two new bridges in the south west of Norway. Our first estimate comes from a hedonic property value model. Rather than follow an approach which is strictly theoretically correct, we adopt a simple one-step approach. In order to investigate whether this simplified approach gives a reasonable estimate, we compare it to an estimate derived from a travel demand model. We find that both methods give very similar estimates, suggesting that the simplistic hedonic approach was reasonable.

1 Introduction

One important function of economists is to provide guidance to decision makers on whether a particular investment should be undertaken. Usually, this advice will be given to decision makers within the public sector. One of the main problems of appraising projects within the public sector is that many of the costs and benefits will come in the form of externalities which are not traded in markets. A number of approaches have been developed to value such costs and benefits (Haab and McConnell, 2002). However, it is difficult to be confident that reliable estimates have been obtained when one of these valuation procedures is followed since there is not usually anything to compare it to.

In this paper, we are concerned with capturing one of the non-market benefits of a road investment project. In particular, we are interested in valuing the changes in labour market

*Stord/Haugesund University College, Bjørnsonsgate 45, 5528 Haugesund, Norway. e-mail: arnstein.gjestland@hsh.no
†Stord/Haugesund University College. e-mail: david.mcarthur@hsh.no
‡Stord/Haugesund University College. e-mail: liv.osland@hsh.no
§Stord/Haugesund University College, e-mail: inge.thorsen@hsh.no

accessibility brought about by the opening of two new road bridges connecting two islands. We take two different approaches to the valuation. Firstly, we use an approach based on estimating the demand for commuting and then calculating the value of time savings for existing and new road users. Secondly, we use the hedonic property price approach to value this same amenity. A similar result from both indicates that either of the methods which we propose for measuring the benefit of improved labour market accessibility is appropriate.

The result that either method can be used ought not be surprising from theoretical perspective. When all of the assumptions of the underlying models are met, they should be measuring the same benefit. Of course, in reality, these ideal theoretical conditions are never met exactly. This leaves practitioners facing uncertainty about the best way to proceed. Both of the approaches adopted in this paper use a very simple approach to calculating the benefit, rather than the more demanding (and potentially infeasible) theoretically robust approach. When we find that both approaches give similar estimates under these conditions, it suggests that a simple approach to benefit estimation can give reasonable approximations.

Surprisingly little research of this type has been undertaken to our knowledge. It would be beneficial to practitioners to know how robust different valuation techniques are. It would also be useful to know when certain 'short-cuts' can be taken with methods without compromising the results too much. It is vital to know when such short-cuts are likely to seriously undermine any estimates made. We believe our empirical approach of verifying an estimate by using two different methods and data-sets is the best way to proceed to answer such questions.

The paper is structured as follows. Section 2 outlines the study area and the infrastructure problem which we consider. Section 3 provides the theoretical foundation for the benefit estimations which we carry out. We cover travel demand and hedonic models in this section. Section 4 provides some detail on how we estimate a commuting demand function and constructs a demand curve. This is then used to estimate the benefit of the infrastructure improvement. Section 5 presents the hedonic model we use and the housing data used to estimate it. This section also contains details on how we estimate the benefit arising from the infrastructure improvement. Section 6 compares the two estimates and provides some concluding remarks.

2 The study area

The study area consists of the islands of Stord and Bømlo located off the coast of south-west Norway, and can be seen in Figure 1. Stord can be characterised as a semi-urban area, whereas Bømlo and Fitjar have a more rural character. The business activities in the region are mainly industrial, and rely heavily on offshore, shipping, shipping-equipment, fish farming and the processing industry. The region lies between the city of Bergen in the north, Haugesund to the south and Stavanger still further south. Driving from Stavanger and Haugesund to Bergen requires crossing Stord. Prior to 2001, this journey required taking a car-ferry.

In 2001, a new road link, called Trekantsambandet (or the Triangular Connection), was opened. The project was part of the main coastal route between Kristiansand in the south of Norway and Trondheim in the North. The project consisted of an 8 km long subsea tunnel, two suspension bridges (linking Stord and Bømlo to each other and to the tunnel to the mainland), one road bridge and around 12 km of extra roads. The final cost of the project was around NOK 1 850 million. We will focus only on the part of the project linking the two islands, and only on commuting flows between the islands. Needless to say, the total demand for the infrastructure is far larger than the demand generated from the islands alone. To illustrate the magnitude of the change in travelling times to the islands' residents, the journey time between the administrative centres of Bømlo and Stord, Bremnes and Leirvik respectively, was reduced from 55 minutes to 32 minutes.

Since the opening of Trekantsambandet, there has been an increase in commuting between the islands. In 2000, there were 394 commuters going from Bømlo to Stord. This increased to 532 by 2009. The corresponding numbers for commuting in the opposite direction are 115 and 196. Total commuting between the islands has therefore increased by 43% over the period 2000-2009.

As should be clear at this stage, the islands provide an ideal geography for our study. The only way to travel between the two islands is by crossing Trekantsambandet. Because we are dealing with islands, the housing markets are clearly delimited and the vast majority of the commuters across Trekantsambandet also live on the islands. This should make it possible to measure the capitalisation of the benefits experienced by commuters into house prices.

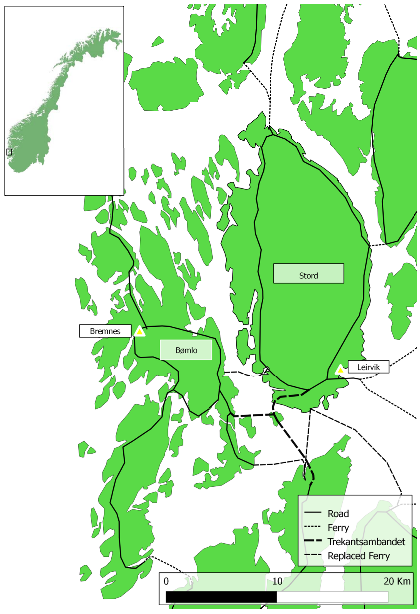

Figure 1: The study area showing the administrative centres of Stord and Bømlo and the transportation network.

3 Theoretical considerations

The aim of this paper is establish whether we obtain the same estimate of the benefits arising from a transportation network investment using two different methods. This section first outlines the basic theory underlying the estimation of benefits from a travel demand function.

3.1 Estimating the direct benefits of commuting

The estimation of the direct benefits from a change in the transportation network, such as the opening of Trekantsambandet, is based on a standard microeconomic framework. The first stage is to estimate a demand curve. There are various ways to accomplish this. In this paper, we use the popular gravity model approach (Sen and Smith, 1995). Other options are of course available (Ortuzar and Willumsen, 2011). A demand curve for trips between two locations is shown in Figure 4.

In Figure 4 we begin at a situation where the generalised cost of travel across a link is given by P_0. At this level, the demand for trips is T_0. The consumer surplus for these road users is given by the area ABP_0. Assume now that an investment is made which reduces the generalised cost of travel to P_1. From a welfare perspective there are two effects which must be considered.

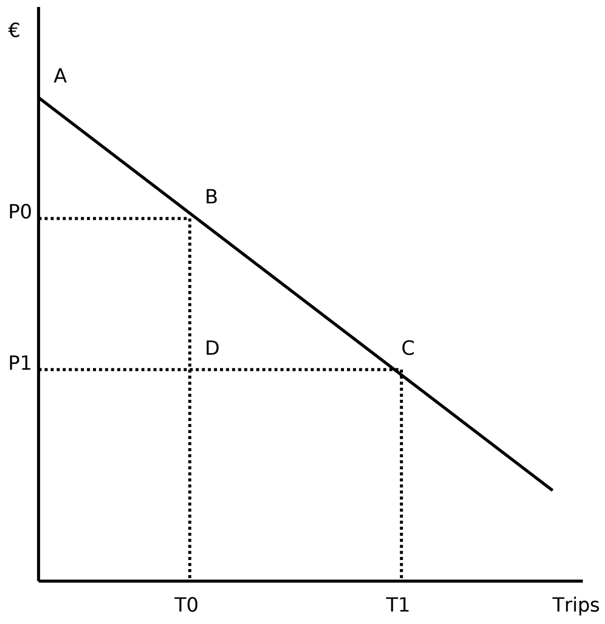

Figure 2: Demand for trips across a link as a function of the generalised cost of travel.

Firstly, the users who made the T_0 trips can now do so at a lower cost. This increases their consumer surplus by the amount P_0BDP_1. The additional effect is the *induced demand*, i.e. the increase in demand from T_0 to T_1. These users were not willing to travel at the previous price, but are willing to do so at the new price. The consumer surplus for these users is lower than that for the previous users. The consumer surplus from the new users is given by the area BCD. The total change in consumer surplus caused by the change is therefore given by P_0BCP_1.

Figure 4 can also be used to show the so-called 'rule-of-a-half' (Small and Verhoef, 2007, p. 183). When the demand for trips is linear, the area BCD is simply half the number of new users multiplied by the change in price. This 'rule of thumb' gives reasonable approximations when the demand curve is close to linear. Even when the demand curve is non-linear, the rule works well for small changes in quantity. Using the rule of half the change in consumer surplus resulting from a reduction in the generalised cost can be calculated as follows:

$$\Delta CS = \frac{1}{2} \cdot (T_0 + T_1) \cdot (P_0 - P_1) \tag{1}$$

More generally, the change in consumer surplus can be written as:

$$\Delta CS = \int_{P_0}^{P_1} T(P)\, dP \tag{2}$$

In this current paper, we model commuting flows rather than total traffic flows. Our hypothesis is that it should be possible to 'see' this change in consumer surplus capitalise into the housing market. Before moving onto specifics we turn our attention to the literature concerning capitalisation and house prices.

3.2 Non-market valuation using the hedonic method

In this paper we use a hedonic model to evaluate the benefits of a new infrastructure project. This method is used extensively in the valuation of a range of public goods which are not directly bought and sold in any market, see for example Kuminoff et al. (2010) for an overview of a range of empirical studies, Wilhelmsson (2000) for a survey of the assumptions for using the hedonic price function to estimate the total benefit to society of an amenity improvement, and Nguyen-Hoang and Yinger (2011) for a presentation of updated theoretical issues.

The hedonic model is based on Rosen (1974) and it is founded on the idea that various goods are composed of a number of attributes that provide utility to the consumers, and that each of these attributes has an implicit price. The primary goal of the hedonic theory is to explain how the hedonic price function is a result of the interaction between suppliers and demanders in the market for the heterogeneous commodity. This interaction is different from the ordinary supply and demand analyses, since the hedonic function encompasses the so called "bid functions" on the demand side and the "offer functions" on the supply side for each individual attribute of the heterogeneous good.

Following Rosen (1974), consumers have maximized their utility when the partial derivative of the bid function with respect to each attribute is equal to its implicit price. The implicit price is the partial derivative of the exogenous hedonic price function with respect to each attribute. This marginal bid is interpreted as the maximum amount a given household is willing to pay (WTP) for a partial marginal increase in a given attribute. An alternative way of thinking about this is as a compensated demand curve for a given attribute, given constant utility and income level.

The majority of the empirical research using the hedonic model aims at obtaining values on the implicit prices for various attributes. Important reasons are that the amount of data needed is relatively small and the implicit prices reveal important information on the values of a range of non-market goods. These factors are hence important motivations for this study. In the hedonic literature, this approach is called Rosen's first step. The hedonic price function is, however, not a demand relationship. To obtain a theoretically correct estimate of marginal willingness to pay, Rosen's second step is used. The second step uses the information from the first step in addition to a range of household characteristics, and aims to estimate the marginal bid-function of households at various levels of the given attribute. The interpretation of the function is an inverse compensated demand relationship. If all consumers are identical with respect to their utility structure, while the suppliers are different, then the equilibrium implicit prices can be interpreted as the marginal willingness to pay for specific attributes (Rosen, 1974). This special case is, however, not observed in practice. Brown and Rosen (1982) analyse the second step, and some important problems with this step are summarized in Ekeland et al. (2002). A broad and updated summary of the main analytical approaches to obtain information about consumer preferences through the second step is found in Taylor (2008).

In this paper, we are going to study to what extent one estimated implicit price can be used as a measure of willingness to pay for improvement in road infrastructure. There is an agreement among researchers that the hedonic price function may reveal information about the value of non-traded goods, given that this value capitalizes into house prices. What we want to study in this paper is how close this estimate is to a more commonly used measure of willingness to pay. This is an important topic given that many papers use the implicit prices as measures or approximations of willingness to pay.

Whether and when the estimated implicit prices can be interpreted as marginal willingness to pay is described in Freeman (2003); Palmquist (1992*b*,*a*). We summarize important results from their research below:

As mentioned above, in equilibrium the implicit prices will be equal to the marginal values of the bid functions. In this way, the implicit prices can correspond to individual marginal willingness to pay (MWTP). Consequently, the value of a marginal change in the amount of a public good is the sum of the marginal willingness to pay for each affected individual (n in all) Freeman (2003, p. 373):

$$w_q = \sum_{i=1}^{n} b_i = \sum_{i=1}^{n} \left(\frac{\partial P}{\partial q} \right)_i \tag{3}$$

Following the notation from Freeman (2003), w_q is the total marginal welfare change, and

b_i is the individual's MWTP. The total price of the heterogeneous good housing is given by P, while q denotes the relevant attribute that is only marginally changed.

For non-marginal changes, the matters could be more complicated, mainly because we frequently do not have knowledge about bid-functions. However, if it is reasonable to assume that the hedonic price function does not change, and if the number of houses affected is small compared to the total market, Palmquist (1992*b*) and Freeman (2003) show that the "hedonic price function can be used to predict the changes in the prices of affected properties. Benefits are exactly measured by the increase in the values of the affected properties, and knowledge of the marginal bid functions is not required" (Freeman, 2003, p. 378). If the change in the attribute or public good increase overall house prices, for each house owner, this increase captures the upper bound of the total benefit of the amenity improvement for the individual (Freeman, 2003; Palmquist, 1992*a*).

4 The commuting model

As mentioned in Section 3.1, the first stage in the estimation of the change consumer surplus resulting from the opening of Trekantsambandet is to estimate a demand curve. Various options are available, but we choose to use a doubly-constrained gravity model. The approach is popular and has been shown by Anas (1983) to give identical results to the multinomial logit model. This means it is possible to give a random utility interpretation to the parameters in the doubly-constrained gravity model. The specific model formulation we use is described in detail in McArthur et al. (2010). We will present the key features of the model here. A basic gravity model of commuting flows can be expressed as below.

$$T_{ij} = O_i D_j \exp(-\beta d_{ij})$$

Here,

T_{ij} is the estimated number of commuters from origin i to destination j

O_i is the observed number of commuting trips originating from zone i

D_j is the observed number of commuting trips terminating in zone j

d_{ij} is travelling time from origin i to destination j

The underlying idea is that the interaction (i.e. the number of trips) between two zones will be proportional to the size of the zones and inversely proportional to the spatial separation between them. The spatial separation can be measured in a number of ways, e.g. travel time, generalised cost, and can be modelled using a number of different functional forms. However, there are several deficiencies in this basic version of the model. Several adjustments must be made to obtain reliable parameter estimates. Firstly, the basic model presented is an unconstrained model i.e. there is nothing stopping there being more trips originating from a zone than people living there, or more trips terminating in a zone than there are jobs there. This is nonsensical and can be corrected for by including two balancing factors in the model, A_i and B_i.

Another weakness in the model highlighted by Sheppard (1979), is that the basic model fails to account for spatial structure. For example, destinations which clustered together may be disproportionately attractive/unattractive depending on whether agglomeration benefits or congestion externalities exist. If this is the case, failure to adequately account for the spatial structure results in a misspecified model. This problem can be avoided by including an accessibility measure in the model (Fotheringham, 1983).

$$S_{ij} = \sum_{\substack{k=1 \\ k \neq i, k \neq j}}^{w} D_k e^{(-\beta d_{ij})} \tag{4}$$

Here, w is the number of potential destinations. If n denotes the number of destinations for which there is observed interaction from origin i, then $w \geq n$. The standard reference for this kind of accessibility measure is Hansen (1959), and it can be interpreted as a job opportunity density measure (see Gitlesen and Thorsen, 2000). The inclusion of this accessibility term means that the model can be termed a competing destinations model.

The model used in this paper includes several more refinements. Due to the importance of pecuniary costs (in the form of road tolls and ferry prices) in our study region, we include a term, c_{ij}, which measures the pecuniary costs incurred when commuting between zones i and j. We also include a term to account for the benefit of living and working in the same zone, $\mu\delta_i j$ (where δ is the Kronecker delta). This has been found to play an significant role in explaining commuting patterns (Thorsen and Gitlesen, 1998). Local labour market effects are accounted for by measuring the number of jobs relative to workers in each zone. Adding all of these elements together gives the following model specification.

$$T_{ij} = A_i O_i B_j D_j S_{ij}^{\rho} \left(O_i^{\alpha_1} D_j^{\alpha_2}\right)^{\delta_{ij}} e^{(-\beta d_{ij} - \sigma c_{ij} + \mu \delta_{ij})} \quad (5)$$

$$A_i = \left[\sum_j B_j D_j S_{ij}^{\rho} \left(O_i^{\alpha_1} D_j^{\alpha_2}\right)^{\delta_{ij}} e^{(-\beta d_{ij} - \sigma c_{ij} + \mu \delta_{ij})}\right]^{-1} \quad (6)$$

$$B_j = \left[\sum_i A_i O_i S_{ij}^{\rho} \left(O_i^{\alpha_1} D_j^{\alpha_2}\right)^{\delta_{ij}} e^{(-\beta d_{ij} - \sigma c_{ij} + \mu \delta_{ij})}\right]^{-1} \quad (7)$$

$$S_{ij} = \sum_{\substack{k=1 \\ k \neq i, k \neq j}}^{w} D_k e^{(-\beta d_{ij} - \sigma c_{ij} + \mu \delta_{ij})} \quad (8)$$

The parameters are estimated simultaneously by the method of maximum likelihood. Maximum likelihood was found through an irregular simplex iteration sequence (Nelder and Mead, 1965). Standard errors were estimated by numerical derivation. The data used to estimate these parameters come from the study region outlined in Section 2 as well as the surrounding area. The reason for including the surrounding area is to increase the number of observations and to have sufficient variation in travel times and pecuniary costs to obtain reliable parameter estimates. More information on this expanded study area can be found in McArthur et al. (2010). The parameter estimates we obtain can be found in Table 1.

Table 1: Parameter estimates based on the model which is specified by Equation (5).

	Parameter value	Standard error
$\hat{\beta}$	0.064495	0.001087
$\hat{\sigma}$	0.024402	0.000757
$\hat{\mu}$	4.079679	0.170076
$\hat{\alpha}_1$	0.082746	0.045424
$\hat{\alpha}_2$	-0.584802	0.032903
$\hat{\rho}$	-0.075938	0.037708

We now have sufficient information to construct a demand curve for commuting trips between Stord and Bømlo. The curve is constructed by aggregating the zones on the Stord and then on

Bømlo, and then systematically varying the time taken to pass between the islands. Prior to the opening of Trekantsambandet when a ferry had to be taken to travel to the other island, this part of the journey took around 25 minutes (including waiting time). When the bridges opened, this was cut to 6 minutes. The ferry price became the road toll, so the pecuniary cost of crossing between the islands remained constant. We also hold the cost of working and living in different zones constant.

In order to be able to calculate the consumer surplus, we convert time to money. The inclusion of the pecuniary costs in the model allows us to value 1 minute of time from our own data rather having to rely on an external estimate. To begin with, we take the ratio of the travel time parameter to the pecuniary cost parameter i.e. $\frac{0.065}{0.024}$. This means that a one minute reduction in travelling time corresponds to a reduction of approximately 2.64 NOK (€0.33) in pecuniary costs. In Norway, some commuting expenses are tax deductible. After making some adjustments for this, we estimate a value of time of 1.90 NOK (€0.24) per minute. To construct a generalised cost we then add the pecuniary costs to the monetised value of time. When taking a ferry, commuters[1] had to purchase a ticket for 40.8 NOK. When the bridge opened, a toll of 45 NOK had to be paid (in 2001 prices). The demand curve constructed from this process is given below.

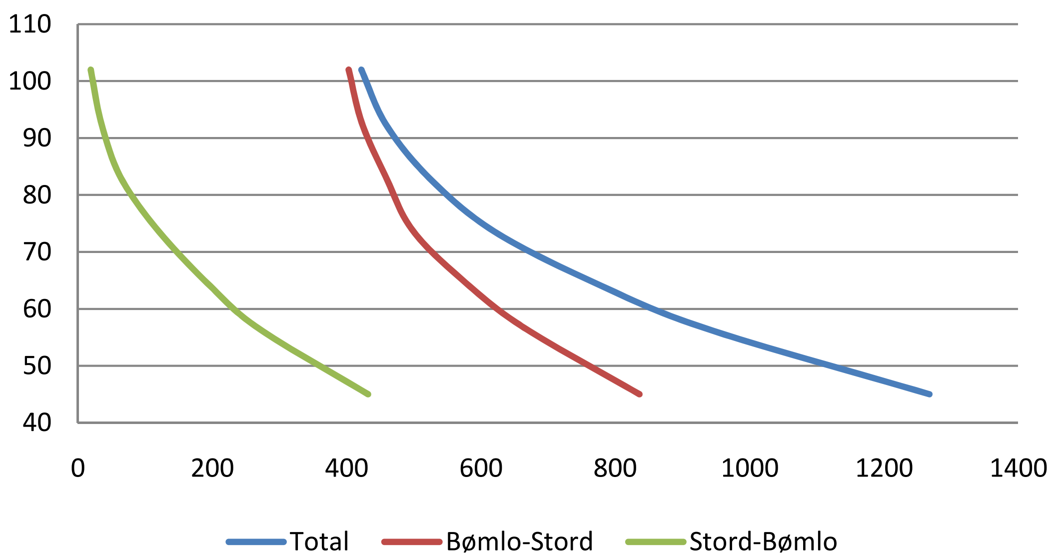

Figure 3: Demand for commuting across between Stord and Bømlo as a function the generalised cost of travel measured in 2006 NOK.

[1]Commuters could receive a discount of 40% if they prepurchased journeys. We assume this discounted price was paid by the commuters.

In Figure 4 shows that for a generalised cost of 98.5 NOK, there are 458 trips predicted. After Trekantsambandet, this cost was reduced to 62.4 NOK. At this cost, a total of 939 trips are predicted. This means that 458 commuters directly benefited from the reduction in journey time. In addition, 481 new commuters were encouraged onto the link. Rather than apply the rule of half, we fit a demand function so that we can analytically solve for the change in consumer surplus. We fit an exponential model and obtain the parameters shown in Equation (9).

$$\hat{T}_{ij} = 2826.1e^{-0.02*P} \tag{9}$$

This gives an R^2 value of 0.97. We then calculate the consumer surplus as shown in Equation (10).

$$\Delta CS = \int_{P_0}^{P_1} \hat{T}_{ij}(P)\, dP = \int_{62.4}^{98.5} 3178.546e^{-0.02*P}\, dP = 24,233 \text{ NOK} \tag{10}$$

This gives the change in consumer surplus for all road users on a one-way trip. We work with the Marshallian demand curve which accounts only for substitution effects and not income effects. This means that we are measuring consumer surplus rather than the more theoretically appealing measures of compensating variation (CV) or equivalent variation (EV) which can be derived from the Hicksian demand curve. In practice, consumer surplus is by far the most used measure. Partly this is because the CV and EV measures are more difficult to calculate. The CS will usually lie somewhere between the CV and EV measures. When income effects are small, as is usually the case with transport projects, the measures will provide similar results (De Jong et al., 2007).

Having calculated the change in consumer surplus for all road users on a one-way trip, a number of steps need to be taken to calculate the present value of this change. Firstly, we multiply the change in consumer surplus by two to get the CS for a return trip. This is then multiplied by a standard 230-day working year. This benefit is assumed to accrue every year for the next 100 years, and this benefit is discounted at a rate of 8%, which is the rate recommended by the Norwegian Ministry of Transport and Communications. This given a present value of 150,424,208 NOK (€18,889,265) in 2006 prices.

In this calculation, we assume that all workers travel alone and travel by car. In our study area, we believe this to be a reasonable assumption. There is minimal public transport available and the rural character of the region means that the car is the dominant mode of transport.

5 The hedonic model

The second approach we take to benefit estimation is the hedonic approach (Rosen, 1974). The general formulation of the hedonic price function used here is specified as:

$$P_{it} = f(X_{sit}, X_{lit}, T_{it}) \tag{11}$$

Here, P_{it} is the price of house i in year t, X_{sit} is the value of the structural attributes of each house, X_{lit} represents the locational characteristics and T_{it} is a set of time dummies indicating in which year the house was sold.

The precise specification of the function is guided both by what data are available as well as previous research on housing markets in neighbouring regions (Osland and Thorsen, 2008; Osland, 2010; McArthur et al., Forthcoming). Based on this previous work, a log-log specification is chosen where the natural logarithms of the dependent and continuous independent variables are used. All the dummy variables are included in levels.

We have include three types of spatially related attributes. The first variable is the travelling time to the central business district, which is interpreted as an urban attraction effect. Secondly, we measure accessibility to the labour market (Osland and Thorsen, 2008). The third spatially related variable is the Euclidean distance to the coastline.

It is not immediately obvious how labour market accessibility should be measured, and a variety of measures are available. One popular approach to the measurement of accessibility is provided by Hansen (1959). He uses a gravity-based measure of potential employment opportunities. In general, the labour market accessibility of a particular zone is given by:

$$\text{AccessPC}_i = \sum_{j=1}^{w} D_j \exp\left(-\beta d_{ij} - \gamma c_{ij}\right) \tag{12}$$

where Access_i is the accessibility of zone i, D_j is the number of job opportunities in zone j, d_{ij} is the travelling time between the zones and c_{ij} are the pecuniary costs incurred when commuting from i to j. Before we can implement these measures we need to obtain parameter estimates for β and γ. There are a number of ways in which this can be achieved. Osland and Thorsen

(2008) estimate the accessibility measure parameters within the hedonic house price model. In this paper, it is possible to do this for the measure accounting only for distance, but not for the measure including pecuniary costs. This is due to insufficient variation in these costs in our sample. We therefore adopt an alternative approach and utilise the commuting data from McArthur et al. (2010), outlined in Section 4.

The other type of locational variable we measure, urban attraction, is captured by introducing the travelling time to the central business district (CBD). In some cases, the location of the CBD of an area will be obvious. In the case we consider in this paper, the picture is more complicated. On the island of Stord, it is obvious that the centre is located in the town of Leirvik. On Bømlo, there is no clear CBD. The population has traditionally been relatively dispersed over the island in small villages. One hypothesis is that Leirvik also represents the CBD for the island of Bømlo. A number of tests were conducted to determine the best way to proceed. Firstly, each post code on Bømlo was selected as a CBD on Bømlo and entered into the hedonic function, one at a time. The influence of distance to the location of the new bridge and to Leirvik on Stord was also tested in this way. The results indicated that it was the distance to the new bridge and the distance to Leirvik which gave the most significant results. Out of these two, the distance to Leirvik gave the highest R^2. We therefore treat Leirvik as the CBD for the entire study region.

In order to study the existence of non-linearity in the variables included in the hedonic function, a semiparametric least squares regression procedure was used as an exploratory tool in the modelling process, see Wood (2006) and Osland (2010) for details. Each continuous variable has been included in a semiparametric smooth, one at a time, where all the other variables were included as traditional parametric regressors. The result from this procedure showed that the age variable should be squared and that the lot size variable was only weakly significant in our study area.

5.1 Spatio-Temporal stability

One important question is whether there is a difference in the estimated implicit prices across space. Given the existence of a transportation barrier at least prior to Trekantsambandet, we are interested in whether Stord and Bømlo could be considered submarkets. The literature in the field of submarkets is voluminous. See, for example, Goodman (1981); Rothenberg et al.

(1991); Bourassa et al. (1999); Watkins (2001); Bourassa et al. (2003); Goodman and Thibodeau (2003); Wilhelmsson (2004); Jones et al. (2004); Goodman and Thibodeau (2007); Islam and Asami (2009); Osland (2010) for reviews and discussions of various theoretical and empirical approaches on the issue of submarkets.

Malon Straszheim was among the first to raise the question of market segmentation when estimating hedonic price functions (Freeman, 1979). These submarkets give rise to changing implicit prices and/or intercept terms of the hedonic price function. As will be explained in the data section, we focus on one house type. Our main issue related to submarkets is, hence, the possibility of spatially varying parameters. Freeman (1979, p. 163) argues, however, that if the structure of demand and supply are the same across regions, even with the existence of barriers to mobility, differences in implicit prices or price levels may be arbitraged away. Freeman's description is relevant for the very homogeneous study area that is considered here: houses on Bømlo could be good substitutes for houses located on Stord.

To anticipate events of our empirical analysis on these matters: A range of tests have been performed on the data to explore the existence of a submarket on the island Bømlo. Clear-cut conclusions have to some extent been difficult to draw because of problems with multicollinearity. Overall, however, the evidence goes in the directions of the coefficients in the hedonic price function being stable across the two islands. Most of the variation in house prices between the two islands are explained by differences in the values of the included attributes and not by spatially varying implicit prices. This conclusion is supported by results in McArthur et al. (Forthcoming). This analysis shows that estimated hedonic house price parameters in neighbouring regions are transferable. In that paper this is the case even when considering clearly separate but adjacent housing and labour market areas.

5.2 The housing data

The data used in this paper consists of single family detached houses sold between 1992-2008. This time span is necessary in order to have enough observations to estimate a hedonic model. It is also useful since it gives us several years both before (prior to 2001) and after the opening of Trekantsambandet. The reason for using single family houses is that this is the overall dominating type of house in the area.

The data from 1992-2001 comes from the period before Trekantsambandet was opened and consists of 663 observations. The data from this period has mainly been provided by Statistics Norway. They are based on a questionnaire which was sent to everybody who has bought a freeholder dwelling in the given period. Information on house price, the size of the house and lot, number of toilets, whether it has a garage, the age and time of sale was gathered from this data source. According to Statistics Norway, this information is very reliable - 80% of the questionnaires were returned and the data have been used to construct the Norwegian house price index for the relevant period.

In addition we need the post codes and coordinates for each house. This was gathered via an online Real Estate Registry which contains information from the national land register of Norway. Data on these variables is complete for all observations. This register also provides information on whether the house has been sold on the market and whether it is characterised as a holiday home. If we are going to study the importance of labour market accessibility, we need to exclude the holiday homes from our sample. In all, these two features are important to be able to exclude observations that do not belong to the population we are studying.

Information on the size of the house, lot size, age, type of house and the existence of a garage can be found in both registers. When this information was missing on individual observations in the data from the questionnaire, it was retrieved from the land register. In both registers there are observations with missing information on independent variables. These observations have not been included in the sample. In all we have lost 92 observations (43 from Bømlo and 46 from the island of Stord) because of missing information on some independent variables.

As a consequence of a broader strategy change by Statistics Norway towards use of national registers in statistics production (Longva et al., 1998), the questionnaire census of home buyers was discontinued from 2002. For the period 2002–2009 we were forced to use data from the national land register. This register only contains the last sale of a property and the provided information on some variables also vary with municipality since the data gathering process is decentralised. To compensate for this shortcoming we also used data from an online real estate site (finn.no)[2], containing nearly 50% of all sales. This last data source was used to include previous sales and also complement for missing information in the land registry. To our knowledge this is the most complete data set of house sales covering this area.

In both samples there were a few observations with very large lots. These tended to have

[2]The data were provided by Statistics Norway.

buildings that were very old or very small. In all 6 observations with lots larger than $10000m^2$ were removed from the two samples.

In order to estimate the accessibility measures and the distance to the administrative centre of the region, information on distances between post codes was utilised. Distances were measured by travelling time by car using a shortest route algorithm, accounting for speed limits and road categories. Journeys between Bømlo and Stord before Trekantsambandet involved taking a ferry. In this case expected waiting time and travelling time on the ferry is accounted for. The matrices were prepared by the Norwegian Mapping authority. The number of jobs in each post code was also measured in order to estimate an accessibility measure. These data are based on the Employer-Employee register. They were provided to us by Statistics Norway and relate to the year 2006. Descriptive statistics for our data are presented in Table 2.

	Entire Sample		Bømlo		Stord	
	Before	After	Before	After	Before	After
Price (in 1998 NOK)	781,387	1,777,429	641,709	1,437,636	847,960	2,036,054
	(389,762)	(904,636)	(318,917)	(721,055)	(402,942)	(944.535)
Price per sq m	4,805	10,876	4,165	8,406	5,111	12,755
	(2,330)	(6,781)	(2,114)	(4,451)	(2,368)	(7,604)
Age	30.73	28.64	33.99	33.77	29.18	24.73
	(24.92)	(22.45)	(27.62)	(23.49)	(23.40)	(20.82)
TimeCBD	23.49	19.04	54.36	30.03	8.77	10.67
	(22.69)	(13.31)	(8.32)	(8.86)	(7.47)	(9.50)
Area	171.67	178.08	164.31	180.71	175.18	176.07
	(67.94)	(61.86)	(61.13)	(57.25)	(70.74)	(65.14)
LotSize	1,251	1,140	1,492	1,353	1,136	979
	(883)	(783)	(1,066)	(906)	(755)	(630)
WC	2.22		1.96		2.35	
	(0.84)		(0.79)		(0.84)	
Garage	0.45	0.42	0.44	0.38	0.46	0.45
	(0.50)	(0.49)	(0.50)	(0.49)	(0.50)	(0.50)
Dist2Coastt		538.70		398.04		625.25
		(508.99)		(379.35)		(557.22)
BoatHouse		0.01		0.02		0.00
		(0.09)		(0.12)		(0.05)
n	663	752	214	325	449	427

Table 2: Descriptive statistics for the data.

5.3 The hedonic results

We present three models in this section. For our first two models, we estimate separate equations for before and after the opening of Trekantsambandet. For our third model, we pool the data. The models are predicated on the assumption that the parameters are constant across space. Given a log-log specification of our models, the parameters related to the continuous variables are interpreted as elasticities and the parameter related to the dummy variables are interpreted as percentage changes (when multiplied by 100). In the Before models, 1992 is the base year for the time-dummies. In the After models 2002 is the base year.

The results found in Table 3 show that all coefficients have the expected sign and are significant. One exception is found for lotsize which is not significant in the 'After' model. One probable substantial interpretation of this result is that people in our rural or semi-urban study area do not value increases in lot-size even though a lot is valued in itself.

Another feature of the results is that almost all the estimated parameters are stable both before and after the opening of Trekantsambandet. This is the case even though the number of variables vary between the models. The parameter related to labour market accessibility is also stable over the two time periods. A Wald test on the equality of the estimated parameter in the before and after models was performed. This null hypothesis could not be rejected, with a p-value 0.68.

We have also tested the stability of the parameter related to distance to CBD in the Before and After model. In this case the null hypothesis of equality could be rejected with a p-value of 0.012. It should be noted, however, that the correlation between distance to CBD and labour market accessibility is -0.85 in the overall sample, so it could be a coincidence that it is the parameter related to distance to CBD that is varying and not the price-elasticity related to labour market accessibility. Excluding distance to CBD from the price functions shows that the parameters of the labour market accessibility variable is highly stable. The estimated parameter is 0.2566 in the Before model, and 0.3089 in the After model. So the estimated price elasticity is stable, but the market value of accessibility seems to be higher in the After model. The results from Table 3 indicate that it is distance to CBD that has changed its impact and not the labour market accessibility variable.

To further test stability over time, we pool the data from both time periods. In this way we increase the number of observations which may reduce the problem of multicollinearity. The results confirmed the analysis above and is reported in Table 3. The parameter related to lot-size has changed over time as has the parameter related to distance to the CBD. When interacting the labour market accessibility variable with a dummy variable getting the value of 1 before, else 0, the interaction variable proved not to be significant. This result was the case when including the same type of interaction variables for lot-size and distance to CBD and without these two interaction variables. The null-hypothesis of no significance of the interacted labour

	Before		After		Pooled	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Area	0.2772	6.17	0.2662	5.77	0.3019	9.43
LotSize	0.0666	2.57	-0.01808	-0.60	-0.0134	-0.45
BeforeLotSize					0.0701	1.86
Age	0.2442	2.66	0.2253	4.01	0.2500	5.24
Age^2	-0.0820	-5.17	-0.0741	-6.94	-0.0828	-9.51
WC	0.3179	5.17				
Garage	0.0927	3.20	0.0639	2.31	0.0905	4.61
TimeCBD	-0.0383	-2.01	-0.0975	-4.52	-0.0915	-4.81
BeforeTimeCBD					0.0372	2.14
ACCESS	0.1951	5.28	0.1789	4.52	0.1813	6.51
β	0.0788		0.0788		0.0788	
γ	0.0164		0.0164		0.0164	
BoatHouse			0.3592	2.12		
Dist2Coast	-0.328	-2.79	0.0382	2.42	-0.0366	-3.69
Year93	0.0336	0.41			0.0231	0.28
Year94	0.1177	1.49			0.1169	1.42
Year95	0.3494	4.75			0.3409	4.56
Year96	0.3848	5.55			0.3694	5.16
Year97	0.6268	8.76			0.5290	7.44
Year98	0.7929	10.79			0.6707	9.32
Year99	0.8494	12.18			0.7552	10.85
Year00	0.9880	13.37			0.8798	12.48
Year01	1.0484	14.69			0.9375	13.36
Year03			0.0453	0.56	1.5081	5.54
Year04			0.1457	1.97	1.5563	5.78
Year05			0.2260	3.00	1.6650	6.35
Year06			0.3715	4.88	1.7390	6.54
Year07			0.5811	7.79	1.8915	7.20
Year08			0.7488	10.34	2.0978	7.96
Year09			0.7930	10.95	2.2551	8.59
Constant	9.4011	21.12	11.7044	23.56	2.2988	8.72
R^2	0.6988		0.5693		0.7547	
$\bar{R}^2$	0.6904		0.5600		0.7499	
ℓ	-156.93		-302.1970		-488.3675	
RESET	0.0048		0.4776		0.0001	
Mean VIF	4.03		3.59		21.47	
Moran's I	0.8011		0.1624		0.0.1379	
Z	3.1776		6.4964		7.5008	
RLM-lag	0.8231		2.5950		0.7791	
RLM-error	2.2564		5.2373		28.1535	
n	663		752		1415	

Table 3: Hedonic models for before and after the opening of Trekantsambandet and with different measures of labour market accessibility. t-values are computed based on White-adjusted standard errors. ℓ denotes log-likelihood value. The null hypothesis of no spatial effects is rejected at the 5% level if the reported Z-value of the Moran's I is higher than 1.645 (Anselin, 1988). RLM denotes robust lagrange multiplier and the critical value of the RLM-test statistics is 3.84 at the 5% signficance level. In the reported test statistics for spatial effects, weights so that $k = 3$ is used, since this provides the highest log-likelihood value of the spatial error model.

market accessibility variable could not be rejected. In all mentioned cases, we achieved p-values of around 0.2. The final model results are presented in a pooled model variant in Table 3.

All the reported models have been tested for spatial effects (Anselin, 1988). The weight matrices use the structure of k-nearest neighbour. The k-nearest neighbour is chosen on the basis of metric distances. Moreover, chosen the weight matrices are symmetric and have been row-standardised. The Moran's I test show that there exists significant spatial autocorrelation in the residuals, and the subsequent robust LM-tests (RLM) indicate that the spatial error specification is the correct one. Following Florax and Nijkamp (2003) the robust LM-tests correct for local lag or error dependence. The non-robust variants of these tests provide the same conclusions. The estimated parameters of the spatial error models are as good as equal to the ones estimated by ordinary least squares. This results indicates that the documented ordinary least squares models are consistent. In Table 3 we present the spatial error models when pooling the datasets.

To sum up, our final model is a pooled model with two parameters varying over time and all other parameters are constant over time. The price-elasticity of access to jobs is constant and has not changed during the study period. This may be due to the fact that the pecuniary costs the commuters have to pay if they live on Bømlo and work on Stord, or vice verse are the same. The implicit price elasticity of the distance to CBD-variable has changed, however. All else equal, the CBD-house price gradient is steeper after Trekantsambandet. In addition to these two changes in implicit price elasticities, the values of the spatial interaction variables themselves have changed, and it is to this topic we now turn with a focus on the labour market accessibility variable.

5.4 Calculating the benefit of labour market accessibility

In calculating the total benefit from the improvement in labour market accessibility, we follow the approach of Gravel et al. (2006). Our approach is to calculate the value of a 'standard' house in each of the postal codes. To do this, we take the average structural attributes from each post code. In addition, we define the house as being sold in 2001 and having a garage. We then calculate the price of this house in each post code using the values of accessibility from before the opening of Trekantsambandet, and then again using the values after it was opened. We use the difference between these two values to represent the standard household's willingness to pay for the transport infrastructure.

To move from the willingness to pay of a single household to the total willingness to pay, we multiply the value of the project to houses in each postal zone by the number of houses in that zone[3]. This is a 'rough-and-ready' approach since our equation applies to single family homes, although we apply it to all housing types here. We believe this to be a reasonable approximation since single family homes dominate our study area. The total willingness to pay calculated in this way should represent the present value of the benefit and should therefore be comparable with the results from the commuting model. In 1998 prices, we calculate a total benefit of 127,003,509 NOK (€15,948,251).

6 Conclusion

We have estimated the benefit arising from the improved labour market accessibility brought about by the opening of Trekantsambandet using two different methods. The traditional approach using commuting benefits gave an estimate of 150,424,208 NOK (€18,889,265) in 2006 prices. The hedonic house price approach gave an estimate of 127,003,509 NOK (€15,948,251) in 1998 prices. Inflating this using the CPI to 2006 prices gives an estimate of 149,483,131 NOK. This gives a difference of only 941,077 NOK between the two estimates.

Our aim of this paper was to provide the first step in a research agenda aimed at examining empirically the impact of taking certain theoretical short-cuts. In this paper, a number of short-cuts have been taken to simplify the benefit estimation process. With regard to the hedonic model, we dispense with Rosen's second step, which is difficult to implement due to data requirements and instead use only the far more convenient, and more widely applied, first step. In the commuting model, we estimate the consumer surplus from the Marshallian demand curve. We assume, as do most applied studies, that income effects are negligible and that the Marshalian and Hicksian demand curves coincide. We find that both of these methods give almost identical benefit estimates, suggesting that the short-cuts taken have had little to no practical impact.

The next step in this research agenda is to replicate the study in different geographical areas. With numerous studies, it will be possible to enter into meta-analysis and to try to understand

[3]Data on the number of houses in each zone were provided by the Norwegian Mapping Authority and were accessed through the Norge Digitalt portal (www.norgedigitalt.no)

what factors influence the result. For instance, is there something special about the region we study here which gives the result we find, or could it be found somewhere else? Once we understand what determines the success or failure of a particular short-cut, we can provide practical guidance to practitioners about how to proceed. Such advice can simplify the process of non-market valuation in some cases, and prevent unjustified abuse of the methods in others.

References

Anas, A. (1983), 'Discrete choice theory, information theory and the multinomial logit and gravity models', *Transportation Research Part B: Methodological* **17**(1), 13–23.

Anselin, L. (1988), *Spatial econometrics: methods and models*, Springer.

Bourassa, S. C., Hamelink, F., Hoesli, M. and MacGregor, B. D. (1999), 'Defining housing submarkets', *Journal of Housing Economics* **8**(2), 160–183.

Bourassa, S. C., Hoesli, M. and Peng, V. S. (2003), 'Do housing submarkets really matter?', *Journal of Housing Economics* **12**(1), 12–28.

Brown, J. and Rosen, H. (1982), 'On the estimation of structural hedonic price models', *Econometrics* **50**(3), 765–769.

De Jong, G., Daly, A., Pieters, M. and Van der Hoorn, T. (2007), 'The logsum as an evaluation measure: Review of the literature and new results', *Transportation Research Part A: Policy and Practice* **41**(9), 874–889.

Ekeland, I., Heckman, J. and Nesheim, L. (2002), 'Identifying hedonic models', *The American Economic Review* **92**(2), 304–309.

Florax, R. and Nijkamp, P. (2003), 'Misspecification in linear spatial regression models', *Tinbegen Institute Discussion Paper (TI 2003-081/3)* .

Fotheringham, A. (1983), 'A new set of spatial-interaction models: the theory of competing destinations', *Environment and Planning A* **15**(1), 15–36.

Freeman, A. (2003), *The measurement of environmental and resource values: theory and methods*, RFF press.

Freeman, A. M. (1979), 'Hedonic prices, property values and measuring environmental benefits: a survey of the issues', *The Scandinavian Journal of Economics* **81**(2), 154–173.

Gitlesen, J. and Thorsen, I. (2000), 'A competing destinations approach to modeling commuting flows: a theoretical interpretation and an empirical application of the model', *Environment and Planning A* **32**(11), 2057–2074.

Goodman, A. C. (1981), 'Housing submarkets within urban areas: Definitions and evidence', *Journal of Regional Science* **21**(2), 175–185.

Goodman, A. C. and Thibodeau, T. G. (2003), 'Housing market segmentation and hedonic prediction accuracy', *Journal of Housing Economics* **12**(3), 181–201.

Goodman, A. C. and Thibodeau, T. G. (2007), 'The spatial proximity of metropolitan area housing submarkets', *Real Estate Economics* **35**(2), 209–232.

Gravel, N., Michelangeli, A. and Trannoy, A. (2006), 'Measuring the social value of local public goods: an empirical analysis within paris metropolitan area', *Applied Economics* **38**(16), 1945–1961.

Haab, T. and McConnell, K. (2002), *Valuing environmental and natural resources: the econometrics of non-market valuation*, Edward Elgar Publishing.

Hansen, W. (1959), 'How accessibility shapes land use', *Journal of the American Planning Association* **25**(2), 73–76.

Islam, K. S. and Asami, Y. (2009), 'Housing market segmentation: A review', *Review of Urban & Regional Development Studies* **21**(2-3), 93–109.

Jones, C., Leishman, C. and Watkins, C. (2004), 'Intra-urban migration and housing submarkets: theory and evidence', *Housing studies* **19**(2), 269–283.

Kuminoff, N., Parmeter, C. and Pope, J. (2010), 'Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?', *Journal of Environmental Economics and Management* **60**(3), 145–160.

Longva, S., Thomsen, I. and Severeide, P. I. (1998), ‘Reducing costs of censuses in norway through use of administrative registers’, *International Statistical Review / Revue Internationale de Statistique* **66**(2), pp. 223–234.

McArthur, D., Kleppe, G., Thorsen, I. and Ubøe, J. (2010), ‘The impact of pecuniary costs on commuting flows’, *NHH Dept. of Finance & Management Science Discussion Paper No. 2010/4* .

McArthur, D., Osland, L. and I., T. (Forthcoming), ‘Spatial transferability of hedonic house price functions’, *Regional Studies* .

Nelder, J. and Mead, R. (1965), ‘A simplex method for function minimization’, *The computer journal* **7**(4), 308.

Nguyen-Hoang, P. and Yinger, J. (2011), ‘The capitalization of school quality into house values: A review’, *Journal of Housing Economics* **20**(1), 30–48.

Ortuzar, J. and Willumsen, L. (2011), *Modelling transport*, Wiley.

Osland, L. (2010), ‘An application of spatial econometrics in relation to hedonic house price modelling’, *Journal of Real Estate Research* **32**(3), 289–320.

Osland, L. and Thorsen, I. (2008), ‘Effects on housing prices of urban attraction and labor-market accessibility’, *Environment and Planning A* **40**(10), 2490–2509.

Palmquist, R. (1992*a*), ‘A note on transaction costs, moving costs, and benefit measurement’, *Journal of Urban Economics* **32**(1), 40–44.

Palmquist, R. (1992*b*), ‘Valuing localized externalities’, *Journal of Urban Economics* **31**(1), 59–68.

Rosen, S. (1974), ‘Hedonic prices and implicit markets: product differentiation in pure competition’, *The Journal of Political Economy* **82**(1), 34–55.

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

Sen, A. and Smith, T. (1995), *Gravity models of spatial interaction behavior*, Springer-Verlag.

Sheppard, E. (1979), ‘Gravity parameter estimation’, *Geographical Analysis* **11**(2), 120–132.

Small, K. and Verhoef, E. (2007), *The economics of urban transportation*, Psychology Press.

Taylor, L. (2008), ‘Theoretical foundations and empirical developments in hedonic modeling’, *Hedonic Methods in Housing Markets* pp. 15–37.

Thorsen, I. and Gitlesen, J. (1998), ‘Empirical evaluation of alternative model specifications to predict commuting flows’, *Journal of Regional Science* **38**(2), 273–292.

Watkins, C. A. (2001), ‘The definition and identification of housing submarkets’, *Environment and Planning A* **33**(12), 2235–2254.

Wilhelmsson, M. (2000), ‘The impact of traffic noise on the values of single-family houses’, *Journal of Environmental Planning and Management* **43**(6), 799–815.

Wilhelmsson, M. (2004), ‘A method to derive housing sub-markets and reduce spatial dependency’, *Property Management* **22**(4), 276–288.

Wood, S. N. (2006), *Generalized additive models: an introduction with R*, Chapman & Hall, CRC Press.