Welfare impacts of landscape dis-amenities: Comparative hedonic approaches

The presentation will compare and contrast two different hedonic approaches to quantify the impacts of dis-amenities from wind turbines. We use the same dataset on house sales in Denmark to explore i) the spatial interrelationships in the house market and ii) the patterns in socio-demographic neighbourhoods. For both analyses we explore the implications for hedonic valuation. Individual abstracts for the two analyses are given below.

I Spatial effects in hedonic models: The case of dis-amenities from wind turbines

Anders Fonnesbech-Wulf¹, Mette Termansen¹, Morten Fuglsang¹, Jim Smart¹, Jacob Ladenburg² and Berit Hasler¹

¹) Department of Environmental Science, Aarhus University.
²) Danish Institute of Governmental Research

Introduction

Using revealed preference methods to value landscape dis-amenities poses the challenge that there is often a high level of correlation between the independent variables. This is acknowledged in literature and often used in arguments for using statistically designed datasets, for example when applying choice experiments. Spatial datasets used for example in valuation studies using property markets pose additional challenges as the researcher may be unaware of important landscape characteristics impacting house prices. Due to the systematic spatial structure of the landscape data such structure has important implications for statistical analysis. Spatially correlated structures in geographical data are usually referred to as spatial heterogeneity. Furthermore, it can often not be ruled out that sales prices at one location may influence sales prices in the neighbourhood as buyers and sellers look at past sales to form expectations on property values. Ignoring empirical evidence for such interdependency the researcher runs the risk of reporting biased estimates of the impact of landscape amenities or dis-amenities and other property characteristics. In this paper we explore the empirical evidence for the policy relevance of both spatial heterogeneity and spatial interdependence in house price markets when estimating the potential welfare effects of proximity to wind turbines. We test this for a study in the Danish municipality of Randers, where there is a current debate on a proposal to expand the number of wind turbines.

Methodology

We test a range of econometric specifications based on different assumptions about the nature of the spatial relationship. The OLS specification is used as a reference point. The specification of the spatial hedonic models requires a definition of the spatial relationships being assessed. The spatial relationship is defined through the spatial weight matrix, $W$. We have no a priory assumption about a specific spatial structure of the data, except from the general assumption that spatially close observations are more related than spatially distant observations, all other aspects being equal. We therefore take an empirical approach to the definition of $W$. The approach is based on $k$-nearest neighbours regression and cross validation identifying the number of neighbours used to define which observations are considered potentially correlated in the housing market.
The spatial error correction model (SEM) is specified using the following standard specification for
the property price \( P \), where a spatial lag of the error term is the representation of spatial
heterogeneity (Anselin, 1988);

\[
P = X\beta + u = \lambda Wu + \varepsilon
\]

Where \( X \) represents the property characteristics, \( P \) is the logarithm of the property value and \( E[\varepsilon] = 0, E[u] = 0, E[\varepsilon \varepsilon'] = \sigma^2 I \) and \( E[uu'] \neq \sigma^2 I \n\)

The spatial auto-regressive model (SAR) is specified using a spatial lag of the response variable to
model the interaction between property values

\[
P = \rho WP + X\beta + \varepsilon
\]

Where \( E[\varepsilon] = 0, E[u] = 0 \) and \( E[\varepsilon \varepsilon'] = \sigma^2 I \n\)

Combining the two specifications allow us to take into account both spatial interdependencies. This
model is often referred to as the general spatial model GSM (Anselin, 1988). The GSM
specification uses two spatial weights matrixes \( W_1 \) and \( W_2 \).

\[
P = \rho W_1 P + X\beta + u = \lambda W_2 u + \varepsilon
\]

We have used the software made available by (LeSage, 2000) and implemented in MATLAB.

Since we have no a priori expectation to the spatial structure of the data, we estimate the weight
matrices using \( knn \)-regression (Härdle, 1990). To estimate the first weight matrix used in SEM, SAR and as \( W_1 \) in GSM we consider the spatial lag term \( \lambda W \) in the SEM and \( \rho W \) in the SAR and
GSM) as a spatial smoother and establish the number of neighbours by cross validation on
deviations from the mean property value. The idea is that the number \( (k) \) of neighbours that results
in the smallest mean squared error is the optimal number for description of the spatial structure. To
estimate the second weight matrix \( W_2 \) in the GSM our point of departure is the SAR model. In this
model we implement the \( knn \)-regression and cross validation procedure on the residuals of the SAR
regression. We can thus obtain the number of neighbours that best predicts the value of the residual
in observation \( i \) and therefore best describes the spatial structure of the residuals. To ensure
identification of the two different lag terms in the GSM, we use a slightly different definition of \( W_2 \)
than \( W_1 \). Where \( W_1 \) is defined as \( W_1 = I/d_{ij} \) if observation \( i \) and \( j \) are neighbours and 0 otherwise, \( W_2 \)
is defined as \( W_2 = I \) if observation \( i \) and \( j \) are neighbours. Neighbours are defined by relative
proximity measured by the Euclidean distance between property \( i \) and property \( j \) \( (d_{ij}) \). After
estimation of the matrices \( W_1 \) and \( W_2 \) the matrices are normalised to have a row sum of unity for
ease of interpretation. That is: \( W_1 \) is a weighted average weighted by the distance between the
neighbouring observations while \( W_2 \) is an un-weighted average.

**Data**

The data used for the analysis are all sales from 2006-2009 of single family detached houses in the
Danish municipality Randers covering an area of 800.14 km\(^2\) of mixed urban and rural dwellings.
Excluding observations with incomplete data records, 2918 house remains for the analysis. Data on
house characteristics as well as environmental amenities and dis-amenities are included in the analysis (Table 1).

Table 1: Variables used in the analysis

<table>
<thead>
<tr>
<th>Name</th>
<th>Definition</th>
<th>Min</th>
<th>Mean</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>Indexed house sales price (DKK in 000’s)(^1)</td>
<td>44</td>
<td>1,568</td>
<td>9,232</td>
</tr>
<tr>
<td>H_Area</td>
<td>Area of living space (m(^2))(^2)</td>
<td>48</td>
<td>143</td>
<td>422</td>
</tr>
<tr>
<td>P_Area</td>
<td>Area of the plot(^2) (m(^2))</td>
<td>72</td>
<td>1,409</td>
<td>47,417</td>
</tr>
<tr>
<td>Rooms</td>
<td>Number of rooms(^2)</td>
<td>1</td>
<td>4.8</td>
<td>15</td>
</tr>
<tr>
<td>Age</td>
<td>The number of years since the house was built(^2) (years)</td>
<td>0</td>
<td>54.9</td>
<td>332</td>
</tr>
<tr>
<td>Flat</td>
<td>Dummy variable equal 1 for houses with flat roof, 0 otherwise(^2)</td>
<td>0</td>
<td>0.02</td>
<td>1</td>
</tr>
<tr>
<td>Bath</td>
<td>Number of bathrooms(^2)</td>
<td>1</td>
<td>1.3</td>
<td>4</td>
</tr>
<tr>
<td>Toilet</td>
<td>Number of toilets(^2)</td>
<td>1</td>
<td>1.6</td>
<td>4</td>
</tr>
<tr>
<td>Road</td>
<td>Distance to road over 6 meters wide(^3) (m)</td>
<td>10</td>
<td>549.2</td>
<td>5,826</td>
</tr>
<tr>
<td>Lake</td>
<td>Distance to Nearest Lake(^3) (m)</td>
<td>20</td>
<td>481</td>
<td>1,595</td>
</tr>
<tr>
<td>Forest</td>
<td>Distance to nearest forest(^3) (m)</td>
<td>10</td>
<td>303.9</td>
<td>1,306</td>
</tr>
<tr>
<td>Recreation</td>
<td>Distance to nearest site of recreation(^3) (m)</td>
<td>10</td>
<td>331.5</td>
<td>2,990</td>
</tr>
<tr>
<td>Wind Turbine Park</td>
<td>Dummy variable equal 1 if the nearest wind turbine is a part of a windmill park, 0 otherwise</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zone</td>
<td>Zone = 1 for properties located in an urban area(^4), otherwise Zone = 0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wind Turbine</td>
<td>Distance to nearest wind turbine(^3) (m)</td>
<td>190</td>
<td>2925.9</td>
<td>6,879</td>
</tr>
</tbody>
</table>

\(^1\) Selling prices are discounted using the regional house sales index (Danmarks Statistik, 2010).
\(^2\) Building and dwelling register (BBR), 2010; \(^3\) Distances calculated as Euclidean distances using the TOP10DK landcover map 2009 version; \(^4\) Urban and Rural zones are defined by By- og landskabsstyrelsen, 2007

**Results**

The number of neighbours used to specify the spatial weight matrix is identified as the number of neighbours, \(k\) minimising the mean squared prediction error (Figure 1).
The number of neighbours used in the specification of $W$ are therefore 20.

The hedonic specifications using this weights matrix test the significance of individual characteristics of property values (Table 2).

Table 2: Parameter estimates of the alternative specifications

<table>
<thead>
<tr>
<th>Variables</th>
<th>OLS</th>
<th>SEM</th>
<th>SAR</th>
<th>GSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>13.469**</td>
<td>13.502**</td>
<td>8.108**</td>
<td>8.145**</td>
</tr>
<tr>
<td>$H_{\text{Area}}$</td>
<td>0.00264**</td>
<td>0.00294**</td>
<td>0.00267**</td>
<td>0.00277**</td>
</tr>
<tr>
<td>$P_{\text{Area}}$</td>
<td>0.00002**</td>
<td>0.00002**</td>
<td>0.00002**</td>
<td>0.00002**</td>
</tr>
<tr>
<td>Rooms</td>
<td>0.03308**</td>
<td>0.02719**</td>
<td>0.02718**</td>
<td>0.02610**</td>
</tr>
<tr>
<td>Age</td>
<td>-0.00366**</td>
<td>-0.00331**</td>
<td>-0.00291**</td>
<td>-0.00292**</td>
</tr>
<tr>
<td>Flat</td>
<td>0.10257</td>
<td>0.04831</td>
<td>0.02718</td>
<td>0.02554</td>
</tr>
<tr>
<td>Bath</td>
<td>0.08890**</td>
<td>0.07459**</td>
<td>0.06735**</td>
<td>0.06644**</td>
</tr>
<tr>
<td>Toilet</td>
<td>0.04126*</td>
<td>0.01823</td>
<td>0.02815</td>
<td>0.02347</td>
</tr>
<tr>
<td>Road</td>
<td>-0.00013**</td>
<td>-0.00013**</td>
<td>-0.00008**</td>
<td>-0.00007**</td>
</tr>
<tr>
<td>Lake</td>
<td>0.00010**</td>
<td>0.00008</td>
<td>0.00007*</td>
<td>0.00006</td>
</tr>
<tr>
<td>Forest</td>
<td>-0.00016**</td>
<td>-0.00013*</td>
<td>-0.00007</td>
<td>-0.00006</td>
</tr>
<tr>
<td>Recreation</td>
<td>0.00002</td>
<td>0.00001</td>
<td>-0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>Wind Turbine Park</td>
<td>-0.07480</td>
<td>-0.09026</td>
<td>-0.07268</td>
<td>-0.08684</td>
</tr>
<tr>
<td>Zone</td>
<td>0.19368**</td>
<td>0.16184**</td>
<td>0.10374**</td>
<td>0.10054**</td>
</tr>
<tr>
<td>Wind Turbine</td>
<td>0.000014*</td>
<td>0.000020**</td>
<td>0.000010</td>
<td>0.000012</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-</td>
<td>-</td>
<td>0.38497**</td>
<td>0.38197**</td>
</tr>
</tbody>
</table>
Table 3 Likelihood-ratio tests between alternative specifications

<table>
<thead>
<tr>
<th></th>
<th>OLS</th>
<th>SEM</th>
<th>SAR</th>
<th>GSM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-likelihood</td>
<td>-1866.4</td>
<td>-767.9</td>
<td>-754.7</td>
<td>-751.1</td>
</tr>
<tr>
<td>LR-test against OLS</td>
<td>-</td>
<td>2197.02</td>
<td>2223.48</td>
<td>2230.58</td>
</tr>
<tr>
<td>LR-test against SEM</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>33.56</td>
</tr>
<tr>
<td>LR-test against SAR</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>7.1</td>
</tr>
</tbody>
</table>

Likelihood-ratio tests between the alternative specifications are presented in table 3. The tests suggest that the general spatial model is the best specification as the critical values in a $\chi^2$-distribution are 6.64 for one restriction and 9.21 for two restrictions respectively. Thus the GSM model performs significantly better than the alternative specifications on a 1% significance level. The SAR model does not perform quite as well as the GSM model, but better than the two models that do not include a lag of the response variable. This corroborates the evidence for a spatial autoregressive process and hence spatial dependence. As demonstrated in appendix B the presence of spatial dependence will produce upward biased parameter estimates if not modelled. Inspection of the parameter estimates in Table 2 confirms this theoretical result. An important consequence of this result is that the explicit modelling of the spatial dependence turns the parameter for the distance to wind turbines from significant in the OLS and SEM specifications to insignificant in the SAR and GSM specifications. Failing to model the spatial dependence would thus have led us to conclude that the object of the valuation study was a significant factor in determining the property values while the inclusion of the spatial autoregressive process suggests that this conclusion is a result of the biasing influence of spatial dependence.

Conclusions

The analysis shows that the results related to the dis-amenity of wind turbine are highly dependent on the choice of model specification. The results indicate that spatial effects cannot be ignored in the property markets and that the general model specification including both a spatial error component and a spatial lag gives a superior explanation of the data. The OLS specification is the poorest description of the data. The results are supported by theory suggesting that the significant negative parameter estimates for the dis-amenity from proximity to wind turbines in the standard OLS model is likely to be due misspecification of the model.

References


Appendix A
Properties of the OLS-estimator in the presence of spatial heterogeneity

Unbiased estimates of the parameters \( \beta \) in the spatial error model can be obtained using OLS if the number of observations is sufficiently large to ensure \( E[\varepsilon] = 0 \). This can be demonstrated as follows

\[
\hat{\beta}_{SEM} = (X'X)^{-1}(X'(X\beta + (I_n - \lambda W)^{-1}\varepsilon))
\]
\[
= (X'X)^{-1}(X'X\beta + (I_n - \lambda W)^{-1}\varepsilon X')
\]
\[
= \hat{\beta}_{OLS} + (X'X)^{-1}(I_n - \lambda W)^{-1}\varepsilon X'
\]
\[
= \hat{\beta}_{OLS} \quad \text{if} \quad E[\varepsilon] = 0
\]

Estimation of a spatial error process with OLS can result in inconsistent estimates of the variance of the estimates of \( \beta \) and thus have consequences for inference in the model. This is demonstrated below

\[
E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(X'X)^{-1}X'u'u'X(X'X)^{-1}]
\]
\[
= (X'X)^{-1}X'E[u'u']X(X'X)^{-1}
\]

The spatially dependent error term can be found by

\[
u = \lambda Wu + \varepsilon
\]
\[
u = (I_n - \lambda W)^{-1}\varepsilon
\]

Combining these results we can derive the covariance matrix

\[
E[u'u'] = \varepsilon(I_n - \lambda W)^{-1}\varepsilon'(I_n - \lambda W')^{-1}
\]
\[
= \sigma^2((I_n - \lambda W)(I_n - \lambda W'))^{-1}
\]
\[
= \sigma^2Q \quad \text{where} \quad Q = ((I_n - \lambda W)(I_n - \lambda W'))^{-1}
\]

Finally this result can be used to obtain

\[
E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = E[(X'X)^{-1}X'u'u'X(X'X)^{-1}]
\]
\[
= (X'X)^{-1}X'E[u'u']X(X'X)^{-1}
\]
\[
= (X'X)^{-1}X'\sigma^2QX(X'X)^{-1}
\]
\[
= \sigma^2_{OLS}(X'QX(X'X)^{-1})
\]
\[
= \sigma^2_{OLS}Q
\]

This will only correspond to the OLS estimator in the trivial case of \( Q = I \) which presupposes \( \lambda = 0 \). Thus the variance on the estimates of \( \beta \) will only be estimated correctly with OLS in the absence of spatial heterogeneity.
Appendix B

Properties of the OLS-estimator in the presence of spatial dependence

OLS is not an unbiased estimator of the parameters in the presence of spatial dependence. The estimates of the spatial autoregressive model will take the form

\[ \hat{\beta}_{\text{SAR}} = (X' X)^{-1} X' (I_n - \rho W) P \]
\[ \hat{\beta}_{\text{SAR}} = (X' X)^{-1} X' P - \rho (X' X)^{-1} X' WP \]
\[ \hat{\beta}_{\text{SAR}} = \hat{\beta}_{\text{OLS}} - \rho (X' X)^{-1} X' WP \]
\[ \hat{\beta}_{\text{OLS}} = \hat{\beta}_{\text{SAR}} + \rho (X' X)^{-1} X' WP \]

The OLS estimator is thus biased in the presence of spatial dependence and will overestimate the parameters compared to the unbiased estimator \( \hat{\beta}_{\text{SAR}} \)
II) Welfare effects of environmental amenities and dis-amenities: A segmented market hedonic approach in Denmark

Jim Smart¹, Mette Termansen¹, Anna Pill-Dam², Jacob Ladenburg³, Anders Fonnesbech-Wulf¹, Morten Fuglsang¹ and Berit Hasler¹

1) Department of Environmental Science, Aarhus University., 2) Department of Business, Aarhus University
3) Danish Institute of Governmental Research

Introduction

Efforts to meet renewable energy targets have increased the demand for on- and off-shore wind turbines. Local opposition to such developments, and particularly to on-shore wind farms, often argues that while renewable energy provides benefits at the national scale, it also imposes considerable costs for local residents. The main adverse local effects of wind turbines are perceived to be negative impacts on the scenic value of the landscape and localised noise pollution associated with energy generation. Several studies have therefore sought to quantify the extent of such impacts. Most Danish studies have used stated preference approaches in this regard and have indicated substantial welfare losses from wind turbine developments. To date, few studies have applied revealed preference approaches to this issue, although Jordal-Jørgensen (1995) did estimate a hedonic price function using a small sample of 79 house sales and found a significant effect on the house price of proximity to wind turbines. Two of the main limitations of using a hedonic approach to value landscape dis-amenities are the empirical challenges posed by the high level of interdependencies among the explanatory variables and the fact that extant on-shore wind turbine locations reflect local planning decisions. Furthermore, preferences towards wind turbines could also influence purchasers’ choice of residential location, and preferences may therefore be systematically correlated with the distance to wind turbines. The hedonic data sample is therefore likely to have less desirable statistical properties than a designed choice experiment. However, stated responses in hypothetical markets may exaggerate the level of response which would actually be observed in a real market (Kahneman et al 1999).

In this paper we explore the use of a two-stage hedonic valuation approach at the regional scale to estimate potential welfare effects from locating wind turbines in proximity to residential areas. The analysis follows a similar approach to Day, Bateman & Lake (2007), who study the economic value of ‘peace and quiet’ in a major urban centre in the U.K. The analysis involves 3 steps: i) market segmentation, ii) estimation of hedonic price functions, and iii) estimation of welfare effects.

Methodology

Hedonic pricing is a well established, two-stage methodology for estimating marginal impacts on house price (stage one) and welfare consistent, transferable estimates of the value of changes in the provision of environmental amenities and dis-amenities (stage two) (Rosen 1974, Cheshire & Sheppard 1998, Zabel & Kiel 2000, Day, Bateman & Lake 2007). Here we seek to apply both stages of this method to a comprehensive dataset comprising more than 16,000 sales of individual family homes in Zealand, Lolland and Falster in 2005, GIS-determined proximity distances from these properties to lakes, woods, roads and wind turbines, the national grid locations of the properties (Kort & Matrikelstyrelsen 2002), and the socio-demographic characteristics of the ‘large
neighbourhoods’ (Piil Damm & Schultz-Nielsen 2008) within which the properties are situated. The data has been made available by the Rockwool Research Foundation.

Following developments in the literature (Day, Bateman & Lake 2007), multi-dimensional model-based clustering is used to identify separate segments in the property market in Zealand, Lolland and Falster within which house buyers with particular socio-demographic characteristics look to purchase properties with particular combinations of attributes, across a particular range of spatial locations. Model-based clustering seeks to classify observations (the properties sold) into an optimum number of distinct, but possibly overlapping, clusters within a P-dimensional space spanned by relevant attributes and characteristics of the observations. Here we derive an 8-dimensional model-based clustering of the 2005 housing market in Zealand, Lolland and Falster using house floor area, garden area, property age, national grid location, average disposable income in the ‘large neighbourhood’ within which a property is situated, and two factors representing ‘housing density’ and ‘household composition’ from a factor analysis of ‘large neighbourhood’ socio-demographics as the 8 dimensions across which clusters are identified.

Following Banfield & Raftery (1993), Fraley & Raftery (1998) and Fraley & Raftery (2010), it is assumed the observed distribution of property sales across the 8-dimensional space can be described by a Gaussian mixture model:

$$\prod_{i=1}^{n} \sum_{k=1}^{G} \tau_k \phi_k \left( x_i \mid \mu_k, \Sigma_k \right)$$

where \(x\) represents the data (of which there are \(n\) data points in the full dataset), \(G\) is the number of clusters in the model, \(\tau_k\) is the probability that a particular observation belongs to the \(k\)th cluster, and \(\phi_k \left( x_i \mid \mu_k, \Sigma_k \right)\) is the probability distribution of the \(k\)th cluster within P-space. Clusters are assumed to be centred at their means \(\mu_k\), and the covariances \(\Sigma_k\) describe the geometric shapes of the clusters within P-space. The covariance matrices for the different clusters can be re-parameterised through the eigenvalue decomposition:

$$\Sigma_k = \lambda_k D_k A_k D_k' \quad (k = 1, 2, \ldots, G \text{ clusters})$$

where \(\lambda_k\) determines cluster volume, \(D_k\) determines cluster orientation and \(A_k\) determines cluster shape (Banfield & Raftery 1993). This decomposition enables models with different restrictions to be fitted to the data via likelihood maximisation using expectation-maximisation (E-M) methods (Fraley & Raftery 1998). Maximised likelihoods from models which impose different restrictions on different total numbers if clusters, e.g. all \(G=12\) clusters must have the same volume: \(\lambda_k = \lambda \ \forall \ k\), or all \(G=16\) clusters must have the same volume and the same orientation: \(\lambda_k = \lambda \ \forall \ k\ and \ D_k = D \ \forall \ k\), can be compared via the Bayesian Information Criterion (BIC) (Schwartz 1987) to determine which model provides the best balance between complexity and the ability to describe the actual distribution of observations through P-space.

E-M optimisation can be computationally intensive for large multi-dimensional datasets so a c.2000 cluster initialisation was used to reduce execution time. This initial clustering was derived by ‘peeling’ and ‘pruning’ a minimum spanning tree (MST) as described by Posse (2001). Model-based clustering was implemented in R using the nnclust (for the MST: Lumley 2010) and mclust

Separate (stage one) hedonic price functions were estimated for each market segment identified by model-based clustering. These segment-specific hedonic price functions were estimated initially using OLS on a semi-log functional form:

\[
\ln\left( P_i \right) = (X_i + e_i(x, y))\beta_i + e_i \quad i = 1, 2 \ldots N
\]

where \( i \) identifies market segment, \( P_i \) is the \( M_i \times 1 \) vector containing the sale prices of properties in segment \( i \), \( X_i \) is an \( M_i \times K_i \) matrix of explanatory variables comprising property attributes, socio-demographic characteristics of the neighbourhood in which the property is located and GIS-derived distances to the nearest wind turbine, recreational facility, lake and road for each property in segment \( i \), \( e_i(x, y) \) is a \( M_i \times 2 \) matrix reporting the coordinate grid location of each property in segment \( i \), and \( \beta_i \) is the \((K_i + 2) \times 1\) vector of parameters to be estimated and \( e_i \) is the \( M_i \times 1 \) vector of residuals.

Paired Wald tests are used to determine whether the hedonic price functions for all 20 property market segments are significantly different from one another. Where hedonic price functions are not significantly different market segments are combined and the hedonic price functions re-estimated. This process continues until a set of statistically distinct segments and their corresponding hedonic price functions have been identified.

Marginal effects on the hedonic price functions are market specific and therefore unsuitable for transfer between locations (markets) or for direct use in welfare estimation. The second stage of our hedonic analysis will aim to produce welfare consistent, transferable estimates of the changes in social value which would result from – in this case – construction of additional on-shore wind turbines and the provision of additional facilities for recreation in general and for lake-related amenities. The marginal prices of these environmental (dis-)amenities will be produced for each property in the dataset by differentiating the relevant segment-specific hedonic price function, at conditions appropriate to that property (location, property attributes, neighbourhood characteristics and distances to relevant environmental (dis-)amenities). These property- and segment-specific prices then provide the data from which pseudo-Marshallian demand functions for the environmental amenities will be estimated. Demand function estimation will use instrumental variables to allow for the endogeneity which exists between the level of the environmental amenity ‘chosen’ when a particular property is purchased and the marginal price of that amenity, as expressed in the curvature at that point on the hedonic price function for the relevant property market segment. Once a pseudo-Marshallian demand function has been established, the area under that function between two levels of ‘provision’ for the environmental (dis-)amenity concerned provides a measure of the corresponding welfare change.

**Results**

Model-based clustering suggested that a 20-cluster solution in which cluster cores were constrained to have the same volume, shape and orientation (in the 8-dimensions), provided a good compromise between fit and complexity. This solution was used as the property market segmentation for subsequent hedonic analysis. Figure 1 shows five of these property market segments plotted in
geographic (Easting, Northing) space. Here a ‘market segment’ is taken to represent properties of particular styles, in particular types of neighbourhood, within a particular geographic range. Note that the geographic range of the market segments which emerge from model-based clustering can differ substantially (Figure 1, segments 7 & 8) and that market segments can be overlapping (Figure 1, segments 11 & 16) or relatively separate (Figure 1, segments 1 & 7) when projected onto geographic space.

Segment-specific hedonic price functions show the expected effects of house attributes (*ceteris paribus* sale price increases as house area and garden area increase, increases in increments for additional toilets and bathrooms, decreases as living area expands in the attic storey and initially decreases, then increases again, as the age of the property increases) and a strong ‘Copenhagen’ location effect. In addition the hedonic price functions are influenced by neighbourhood socio-demographics (*ceteris paribus* sale price increases as western ethnicity in the neighbourhood increases, and decreases as crime in the neighbourhood increases) and the distance to environmental and communication amenities / dis-amenities (*ceteris paribus* sale price decreases as the distance to recreational facilities increases, decreases as the distance to a road increases and increases as the distance from a wind turbine increases). Results from the second stage hedonic analysis are still being produced.

**Conclusions**

Results produced so far suggest that the property market in Zealand, Lolland and Falster can be portrayed as a number of separate market segments, some of which may show considerable overlap in geographic (Easting, Northing) space (Figure 1), but which are nevertheless distinct in the other important dimensions which characterise house purchase: attributes of the properties themselves and features of the neighbourhoods within which those properties are located. Model-based clustering provides a statistically rigorous and computationally feasible approach for identifying property market segments within data sets of the size examined here (c. 16,000 observations).

Statistically distinct hedonic price functions have been estimated for the separate property market segments. These hedonic functions show the expected *ceteris paribus* responses to property attributes, property location and to neighbourhood characteristics. The *ceteris paribus* influences of environmental (dis-)amenities were also evident, particularly in the hedonic price functions of market segments containing higher numbers of transactions (typically > c.2000). Where present, property price responses to environmental attributes conformed to a priori expectations.

Results from the second stage of hedonic analysis are still awaited. The strength and consistency of the marginal effects which the environmental (dis-)amenities exert over the hedonic price functions, together with the overall size of the dataset, suggest that it should be feasible to estimate pseudo-Marshallian demand functions for these environmental (dis-)amenities. This would then enable theoretically consistent, transferable welfare estimates to be produced for changes in the provision of the environmental (dis-)amenities concerned.

Once the relevant pseudo-Marshallian demand functions are established, the spatial nature of this analysis and the richness of the spatially-referenced dataset should enable ‘welfare contours’ to be constructed to identify locations in Zealand, Lolland and Falster where it would appear to be particularly advantageous to locate new environmental amenities (recreational facilities or lake-
related amenities), or disadvantageous to locate environmental dis-amenities (additional on-shore wind turbines).

References


Lumley T. (2010) *Minimum spanning trees and clusters: Package ’nnclust’.* tlumley@u.washington.edu


Figure 1: Illustrative segments of the property market in Zealand, Lolland and Falster plotted in geographic (Easting, Northing) space.