Housing Market Risks

Theebe, M.A.J.

Citation for published version (APA):
Besides being affected by macroeconomic factors influencing all residential properties, there are also factors at a micro-level that influence house values, but which are beyond the influence of homeowners. If the government decides to allow the construction of a nuclear plant on a formerly vacant industrial parcel, prices of surrounding properties are likely to be affected. There are many of these potential threats, like the construction of high-rise office buildings or the opening of a huge shopping mall, new waste sites or nuclear depositories, factories worsening air quality, declining neighborhood status, and many more.

The impact of these negative externalities, also referred to as Not-In-My-Backyards or NIMBYs, on house prices can be estimated explicitly. When appraising a residential property, an appraiser will include many characteristics of the house to derive a final estimate of the value. Hedonic regression tries to mimic the appraisal process in a reverse way, by deriving implicit prices of property characteristics from sales prices. Besides physical attributes, property values are also influenced by amenities, like view, nearby shops, or public transportation. Including of these externalities in the hedonic regression, will result in an estimate of their impact on property prices. In this way, the regression could estimate that the adjacent nuclear power plant adds a significantly negative value to the total property value. The hedonic approach is the most frequently applied technique to evaluate externalities.

With respect to externality analysis, the most common alternative to hedonic regression is contingent valuation. This method derives monetary values by means of surveys, in which persons are asked for their willingness to pay for a hypothetical situation. The greatest strengths of this technique compared to hedonic regression are less stringent data requirements and the lower dependence on data usage. The results of Smith and Mansfield (1998) and Brookshire et al. (1982) provide evidence in favor of this method, amongst others. However, the largest weakness is its subjectivity, as shown for instance by Schkade and Payne (1994) and Diamond and Hausman (1994). For many individuals it will be difficult to assign a value to objects. Especially when the examined object is intangible, like air quality or noise pollution, one might question the reliability of the assigned values. Moreover, there might be an important difference between hypothetical and actual decisions. Depending on the belief of the examined person whether the price actually has to be paid or not, the assigned value might be too low or too high,
respectively. Moreover, people tend to give socially desired answers, and answers could be forced in certain directions by framing, that is the way choices are positioned.

Although there might be situations in which a survey method is preferred to the hedonic approach, we consider it to be inferior for externality analysis when sufficient transaction data is available for hedonic regression. For the Dutch housing market, this is the case, and we analyze housing market externalities in the next two chapters by means of hedonic regression.

To illustrate the exact workings of externality analysis with use of hedonic regression, we discuss in section 6.1 how the impact of externalities on house prices can be measured, and we provide a review of the wide range of externality studies in section 6.2. However, using hedonic regression for NIMBY-studies could introduce potential problems regarding the reliability of the outcomes. More efficient estimates can be obtained by using so-called ‘spatial correlation’ techniques. The problems with hedonics and the principles of spatial correlation techniques are discussed in sections 6.3 and 6.4, respectively. In section 6.5, we construct a simulation to illustrate how using spatial autocorrelation really could improve NIMBY-analysis. Section 6.6 concludes this chapter.

### 6.1 Externalities and hedonic regression

The potential price impact of an externality can be dealt with in various ways, of which three approaches are explained in this section. These alternatives indicate the presence of an externality by means of dummy variables, by a variable reflecting distance or proximity, and by variables that are specific for the externality.

Some studies measure the potential impact of an externality by introducing a dummy variable indicating the nearby presence. In this case, the dummy will be 1 for a property if this property is situated within a specific distance of the externality, or if the externality is visible. With a negative externality, the estimated coefficient would be negative, indicating the externality has a negative impact on property prices. The number of dummy variables could be extended for several distance ranges, such that the presence could be nuanced; dummies might indicate a distance between 500 and 1000 feet, between 1000 and 1500 feet, and so forth.

A widely used approach is nuance the externality presence by using a variable reflecting the distance between a property and the externality. With a negative externality, the estimated coefficient for distance is likely to be positive; each additional unit of distance causes the property price to be higher. However, just including a linear distance variable might cause distortion of the regression estimates. It is likely that, after some distance, the externality will no longer impact prices. For instance, a linear distance
variable implies that properties at a distance of 5,000 feet are affected twice as much as properties located 10,000 feet from the externality, while in practice there might be no effect after 5,000 feet at all. This issue can be dealt with in three different manners. The simplest solution is selecting only those properties located within a specific range from the externality. This impact range could be estimated empirically. A second option is to construct a proximity variable based on a threshold distance, like

\[
(6.1) \quad PROX = \text{MAX} [\text{THRESHOLD} - \text{DISTANCE}; 0]
\]

For properties located outside the indicated range, the externality is assumed to have no impact. In case of a negative externality, the coefficient for proximity must be negative. The major difference regarding the previous solution of selecting nearby properties is that information from properties outside the assumed impact range is not discarded now, but still used for estimation of implicit prices of other property characteristics. A third alternative to deal with a diminishing impact on property prices is to include a variable reflecting quadratic distance besides linear distance. In this way, each additional foot of distance rises or lowers property prices, but at a decreasing rate.

Finally, instead of using dummies or distance variables, the presence of an externality can be expressed by variables that are specific for the externality. For example, when studying an externality that is likely to cause much noise like airports or highways, one might measure the price impact by including variables reflecting noise levels. In that specific case, each additional decibel is assumed to lower property prices.

6.2 NIMBY literature review

Literally hundreds of studies have addressed the potential impact of externalities on housing prices, covering a broad area. Analyzed topics concern environmental aspects, traffic noise, factories, toxic or regular waste sites, and neighborhood characteristics. Due to this abundance, it is hardly possible to create a survey discussing all existing literature. Moreover, many of these studies are reviewed by other studies, grouped by externality type. Jackson (2001) and Boyle and Kiel (2001) provided a survey of environmental issues, and Nelson (1980, 1982) reviewed several studies concerning the impact of traffic noise on house prices. Farber (1998) discussed studies for undesirable land uses, and Kiel and Zabel (1996) reviewed studies regarding neighborhood composition. This section intends to illustrate the enormous variety of studied potential threats to property values.

Environmental aspects claim a large part of the NIMBY-literature. Many hedonic studies focused on the impact of air quality on house prices (Zabel and Kiel (2000), Graves et al (1988), and Murdoch and Thayer (1988)), in which air quality usually is expressed as a chemical composition, like the concentration of $\text{NO}_2$, $\text{SO}_2$ or $\text{SO}_3$. However, Boyle and Kiel (2001) concluded that the coefficients estimating air quality impact are insignificant
for most reviewed studies. On the other hand, many studies show that the quality of surrounding water does influence property prices, like Legget and Bockstael (2000), Michael et al (1996), and Mendelson et al (1992). Water quality is reflected in various ways, sometimes by continuous variables indicating pH-level or water clarity, or by dummy variables showing a quality rating. Another environmental aspect concerns properties located in environmentally sensitive areas, which face a larger risk of natural hazards. Murdoch, Singh and Thayer (1993) showed the negative impact of an earthquake on property values, while lower prices of properties located in a floodplain were studied by Furman et al (1991) and Shilling, Sirmans and Benjamin (1989).

A huge number of studies is devoted to the impact of traffic noise on house prices. The impact of road nuisance is studied by Hughes and Sirmans (1992) and Huang and Palmquist (2001), while aircraft nuisance related to house prices was analyzed by Collins and Evans (1994) and Levesque (1994), amongst others. Poon (1978) analyzed noise generated by railroads. In hedonic models, noise is frequently specified by a variable reflecting the noise level. A more extensive discussion of these studies is provided in Chapter 8.

The nearby presence of smelly, giant or even dangerous plants could be a serious threat to property values as well. Refineries (Flower and Ragas (1994)), nuclear power plants (Nelson (1981), Gamble and Downing (1982), Webb (1980)), electric utility power plants (Blomquist (1974)), incinerators (Kiel and McClain (1995)), lead smelters (Dale et al (1999)) and chemical plants (Carroll et al (1996)) are not popular with homeowners. Variables specifying the distance to the object or dummy variables reflecting location are used to capture the impact on property prices.

The negative house price impact of sites storing radioactive and other hazardous materials covers another part of literature. Examples are Kinnard and Geckler (1991), Michaels and Smith (1990), Smolen, Moore and Conway (1992), Kohhase (1991), and Kiel (1995). Also the impact of regular waste sites on surrounding houses received a lot of attention, like McClelland, Schulze and Hurz (1990), Thayer, Alberts, Rahmatian (1992), Ketkar (1992), and Reichert, Small and Mohanty (1992). Just as with the analysis of plants, distance variables or dummies are used.

Finally, residential property values may be negatively influenced by the presence of other buildings, or the composition of the neighborhood. Studies are performed for the price impact of a high-rise office building (Thibodeau (1990)), a shopping center (Colwell, Gurjal and Coley (1985)), a church (Do, Wilbur and Short (1994) and Carrol, Claurette and Jensen (1996)) and mobile home parks (Munneke and Slawson (1999)). Neighborhood characteristics are also determinants for house prices, as shown for crime (Buck et al (1991), Thaler (1978) and Hellman and Naroff (1979), income (Marous (1996)), racial composition (Schnare (1976), Bailey (1966), and King and Mieszkowski (1973), and the presence of rental properties (Ko et al (1991). The impact of these
objects is measured in various ways; distance variables, distance dummies, and ratios reflecting neighborhood characteristics.

These are just a few references to illustrate the wide range of housing market externalities studied with hedonics. However, a hedonic regression will only yield reliable estimates if the 'correct' set of variables is selected. Otherwise, the regression might confuse the price impact of an unknown characteristic with the externality impact. In practice, however, not all relevant property characteristics are known, such that estimates of externality studies applying straightforward hedonics are not reliable.

6.3 The problem with hedonics

The conclusion whether an external factor is a significant determinant for property values depends on the significance level of the estimate. With hedonic regression, the selection of variables could influence this significance, which should be corrected for. So-called spatial autoregression techniques can yield more reliable conclusions regarding the NIMBY-coefficient, as we explain in this section.

Broadly speaking, property characteristics can be divided into two completely different categories. The first group consists of physical attributes, like square footage, presence of garage and maintenance. With use of these physical characteristics only, usually a major part of the variance in house prices could be explained; R-squares of 0.7 are not uncommon in empirical studies. Many housing characteristics are registered with brokerage offices, where information about physical attributes could be obtained. However, more difficulties will be found with the second category of characteristics: locational attributes. Distance to the city center or nearby presence of a park could drive house values up, just as neighborhood amenities and status. But this information could be captured by dummy variables defining nearby presence, by variables reflecting distance between the individual properties and the amenity, or by variables describing neighborhood characteristics like crime or income. This locational information is not readily available in most situations, and collecting or constructing the variables frequently meets practical problems. In an ideal situation, all relevant location characteristics influencing property prices are known and correctly specified. In practice, this will seldom be the case. As shown in Dubin (1988), amongst others, omitted variables will not affect the sign or magnitude of a parameter estimate, but its variance will be biased in a downward way. This might lead to incorrect interpretation of externality study results, since t-statistics could erroneously indicate significance.

In many hedonic studies, the inclusion of region dummies tries to solve this omitted variables problem. However, in order to be homogeneous, these regions must be rather small. For example, Pace, Barry and Sirmans (1998) state that typical neighborhoods
might have 10 to 20 houses turn over in a year. For the studies in Chapters 7 and 8, this implicates a regression model with ten thousands of region dummy variables. Moreover, this would really hamper the analysis of price effects of externalities, since it is very likely that a few specific neighborhood dummies will capture this externality effect, instead of the variable specifying the externality presence.

In addition, region dummies and variables reflecting neighborhood characteristics could be based on zip code areas, or based on other arbitrary boundaries that do not necessarily correspond with borders that separate homogenous areas. This might contribute to regression distortion as well. Therefore, incorrect and omitted variables hamper hedonic regression, a problem that needs to be solved in order to obtain reliable findings in a housing market externality analysis.

**Clustered residuals**

In an econometric way of speaking, the residuals following from a hedonic regression could reflect unused information. This can be visualized by mapping the residuals in a two-dimensional chart, with the axes specifying the location of the properties, with x- and y-coordinates. This map could reveal geographical clustering of residuals, as shown in Figure 6.1. For example, since Amsterdam properties are generally more expensive than Rotterdam properties, the residuals for Amsterdam house prices as a function of attributes will be positive, while Rotterdam residuals will be negative. Even the inclusion of many neighborhood characteristics would not solve this problem. To deal with this, so-called spatial autocorrelation is introduced in the real estate literature. Similar to autocorrelation in time series analysis, in which an observation depends on observations nearby in time, an observation in a cross-sectional analysis may depend on observations located nearby in space. This spatial dependence could be modeled as well. Instead of including preceding observations in time, as in time series regression, the cross section estimation explicitly incorporates values of nearby located properties. In this way, the estimation could use information from omitted location variables, since these variables might be capitalized into prices of adjacent properties. The empirical rationale for this econometric modeling is the sales comparison method used by property appraisers, in which values of properties of comparable quality and at comparable locations are incorporated explicitly in the derivation of the estimated property value.

After some early attempts, for example Dubin (1988), spatial autocorrelation was adapted more and more in real estate literature. Examples are Pace and Gilley (1997), Pace et al.
(1998) and Pavlov (2000), all of whom applied spatial statistics in property price studies. These publications show very convincing evidence of the importance of spatial autocorrelation when modeling real estate prices, through significant improvements of regression performance. Since spatial autocorrelation got included in commercial statistical packages and is even freely distributed as freeware add-ons or programming code, one could not possibly ignore spatial autocorrelation when performing hedonic regressions for the analysis of real estate markets. A more extensive discussion of the techniques is provided in the next section.

6.4 Spatial autoregression explained

Normally speaking, the specification of a hedonic regression would be

\[ P = X\beta + u \]

in which the vector \( P \), consisting of prices of \( n \) properties, is regressed on matrix \( X \), containing \( k \) characteristics for each of the \( n \) properties. These characteristics may be both physical and locational. The results of this regression are implicit prices for each of the \( n \) attributes, captured in vector \( \beta \). This vector could be written as a function of the characteristics and the property prices:

\[ \beta = (X'X)^{-1}X'P \]

The \( n \) regression residuals, which are the differences between the actual property prices and the prices predicted by the estimated attribute prices, are contained by vector \( u \). The smaller these residuals, the better the true prices are explained by the modeled characteristics. Ideally, the residuals do not show regularities or systematic differences with respect to each other. If they do exhibit regularities, there is still some information left that could be used to improve regression performance. After all, removing these regularities with another regression specification will lower the new residuals; these lower residuals will raise regression performance statistics. However, it is very likely that residuals from a standard hedonic regression do show regularities; if locational information is omitted or incorrectly specified, properties located nearby each other will show similar regression residuals. Some areas will show mostly negative residuals, in other areas positive residuals might dominate. This spatial inter-dependency of these residuals could be expressed by means of a variance-covariance matrix \( V \):

\[ \text{var}(u) = \sigma^2 V \]

The \( \sigma^2 \) represents the constant part of the covariances. If the residuals were not dependent on each other and if all variances were constant, matrix \( V \) would be an Identity matrix, a matrix filled with zeros except for the ones at the main diagonal. This is assumed by OLS-regression.
Regularities in residuals
The previous paragraph argued that this OLS-assumption is likely to be violated if using property prices with hedonic regression. The standard cure in econometric theory for this violated OLS assumption is to transform both the prices and the property characteristics with a specific matrix, $L$, such that the residuals from a new regression behave neatly. The hedonic specification would then look like

$$(6.5) \quad LP = LX\beta + Lu$$

This procedure is called Generalized Least Squares (GLS). The transformation matrix $L$ follows automatically from rewriting the estimation for the vector with implicit prices as

$$(6.6) \quad \beta_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}P$$

or, if one defines $V^{-1}$ as $L'L$:

$$(6.7) \quad \beta_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}P$$

This matrix $V$ from Equation (6.7) is equal to the $V$ from Equation (6.4), so matrix $L$ could be derived from the variance-covariance matrix of the standard hedonic regression residuals. In this way, the regularities in the residuals are now explicitly incorporated into the new estimation of the implicit prices. The implicit prices vector could therefore be calculated as

$$(6.8) \quad \beta_{EGLS} = \left(X'\hat{V}^{-1}X\right)^{-1}X'\hat{V}^{-1}P$$

So, by estimating the implicit prices for housing attributes as in Equation (6.8), the regularity problem could be solved.

So far, this is standard textbook econometrics. However, spatial autocorrelation makes specific assumptions regarding the structure of matrix $V$. If residuals are spatially clustered, one residual will be very similar to a residual located nearby in space. The larger the spatial distance between the residuals, the lower the dependence will be, as is the case in the illustration of a two-dimensional residual plot shown above. This assumption is used by spatial regression.

The regularity of residuals over space, as illustrated in Figure 6.1, can be used to model the variance-covariance matrix $V$ of Equation (6.7). In that case, the element of $V$ at row $i$ and column $j$ would be a direct function of the distance between house $i$ and house $j$.

The many alternatives to incorporate this spatial dependency in the new estimation of the vector $\beta$ could be classified in two categories. At one hand, the so-called lattice models directly approach the variance-covariance matrix with use of calculated distances or functions of these distances for each pair of two houses, based on a system of X- and Y-coordinates in space. On the other hand, matrix $\hat{V}$ (not the inverse $\hat{V}^{-1}$) might be
modeled by estimating the general correlation between houses as a function of distance. This procedure is applied by geostatistical models.

Lattice models
The former category is most widely used in real estate and probably the most intuitive. The clearest way of incorporating residuals from nearby located properties is by adjusting Equation (6.2) to

\[(6.9) \quad P = X\beta + \rho W (P - X\beta) + \epsilon\]

although a more appropriate notation would be

\[(6.10) \quad P = X\beta + u, \quad \text{with} \ E[u_i | X_i] = 0, \quad u_i = \rho \sum_{j \neq i} w_i^j u_j + \epsilon_i, \quad \text{and} \quad w_i^j = f(\text{dist}_{i,j})\]

In these equations, besides being dependent on property characteristics, the prices also depend on the difference between true prices and predicted prices \(X\beta\), so the regular regression residuals. To make sure the regression only considers the residuals from nearby properties, the residuals are weighted by weight matrix \(W\). The element on row \(i\) and column \(j\) is a function of the distance between property \(i\) and property \(j\). This matrix must be specified in advance, and could be constructed by many alternative methods, as described below. The parameter \(\rho\) in Equation (6.10) specifies the degree of spatial dependency. The higher this parameter, the stronger is the dependency of property prices on nearby located properties.

Rewriting Equation (6.9) yields

\[(6.11) \quad (I - \rho W)P = (I - \rho W)X\beta + \epsilon\]

In this equation, \(I\) represents an Identity matrix. Scaling the dependent and independent variables by \((I-\rho W)\) is similar to what happens in Equation (6.5) with matrix \(L\). Hence, analogous to (6.6) and (6.7), and with use of spatial autocorrelation, the estimates of the implicit property prices could be written as

\[(6.12) \quad \beta_{SAR} = (X'\Psi^{-1}X)^{-1}X'\Psi^{-1}P\]

in which the inverse of the variance-covariance matrix is defined as

\[(6.13) \quad \Psi_{SAR}^{-1} = (I - \rho W)(I - \rho W)\]

comparable with the specification of \(V^{-1}\) as the product of matrices \(L'\) and \(L\). So, this is a special situation of the Generalized Least Squares estimation as in standard literature, with now the variance-covariance matrix specified as a function of spatial distance between two observations.
The estimator of Equation (6.13) is indicated as $\beta_{\text{SAR}}$. SAR stands for simultaneous autoregression. This is just one alternative for direct modeling of the variance-covariance matrix. For example, CAR models errors with use of conditional autoregression, in which the expectation and the variance of the dependent variable are made conditional on surrounding observations. Equation (6.9) might then look like

\begin{equation}
(6.14) \quad P = X\beta + (I - \rho W)^{1/2} \varepsilon = X\beta + \rho W(Y - X\beta) + (I - \rho W)^{1/2} \varepsilon
\end{equation}

or

\begin{equation}
(6.15) \quad (I - \rho W)^{1/2} P = (I - \rho W)^{1/2} X\beta + \varepsilon
\end{equation}

With transformation matrix $L$ now specified as $(I - \rho W)^{1/2}$ and a symmetric weight matrix, conditional autoregression models specify the variance-covariance matrix $\Psi^{-1}$ as

\begin{equation}
(6.16) \quad \Psi_{\text{CAR}}^{-1} = (I - \rho W)
\end{equation}

So, one could deal with spatial autocorrelation by modeling the variance-covariance matrix used in the GLS-estimator with use of the spatial distances between individual observations. These distances are used to construct the weight matrix $W$. This matrix specifies the degree of price influence of nearby properties.

**Weight matrix of lattice models**

There are many alternatives to specify the weight matrix, of which the simplest one is the nearest neighbor specification. In this case, all elements of the weight matrix will be zero, except if there is no observation closer. If observations $i$ and $j$ are nearest neighbors, element $w_{ij}$ will be 1. This could be extended to $m$ neighbors. Note that this binary matrix does not need to be symmetric; an outlier in space might find another specific observation to be its nearest neighbor, but this other observation could find its nearest neighbor in close proximity.

Another approach, resulting in a symmetric weight matrix, is specifying element $w_{ij}$ as 1 if the distance between observations $i$ and $j$ is smaller than a chosen limit. Just as with the asymmetric alternative, observations located nearby each other will have higher weights than other observations.

Alternatively, the weight matrix could be created with use of a non-binary approach. The elements could be specified as a continuous function of the distance between two observations, like $w_{ij} = 1 / \{\text{distance}_{ij}^C\}$, in which $C$ is a constant. The inverse relationship between weights and distance make sure that observations in close proximity will have higher weights, as is assumed by spatial autoregression.

Alternatively, as used by Bailey and Gatell (1995), another weight matrix could be constructed with use of so-called Delaunay-triangles. Instead of looking for a specific
number of nearest neighbors, this algorithm will create groups of on average six adjacent observations. This information is used to construct a binary weight matrix. With Delaunay triangles, no parameter $\rho$ needs to be specified for determination of the rate of geometric decay in weights, as is the case with nearest-neighbor approaches.

**Geostatistical models**

Lattice models make use of the grid pattern formed by discrete spatial coordinates. In real estate, the exact location of a property is indicated by two coordinates, longitude and latitude. These coordinates are usually expressed on a discrete scale. The number of possible locations is therefore countable, and could be plotted as a lattice. In geostatistical models, however, location could be defined on a continuous scale. Spatial information is then dealt with in a completely different way. Matrix $\tilde{V}$ (not the inverse $\tilde{V}^{-1}$) is modeled by estimating the general correlation between houses as a function of distance, instead of creating weight matrices with distance functions for each pair of houses. Since we will not use geostatistical models in our estimations, we will not discuss these methods.

Dubin (1998) mentions that there is no consensus amongst real estate researchers regarding which alternative provides the best representation of the spatial dependency in the housing market. All approaches have been used in publications. Since all results depend on the a priori specified spatial structure she concludes this to be problematic. In practice, the researcher has to find the structure yielding the best regression performance. In this thesis, we will use both CAR and SAR to model the regression equation, and specify the weight matrix with nearest neighbors and Delaunay triangles, to test for the most convenient specifications.

To summarize, omitted variables will not affect the sign or magnitude of a parameter estimate, so OLS will not lead to inconsistent estimates. However, variances are likely to be biased in a downward way, leading to artificially raised significance indicators. In case of analyzing NIMBYs, the significance is of major importance: does an externality influence property prices, or not? Spatial autocorrelation produces more efficient estimates, and will therefore be more useful when analyzing housing market externalities.

### 6.5 Simulation of a NIMBY-study using spatial techniques

To our knowledge, no NIMBY-study has applied spatial techniques, although the previously cited studies showed how these techniques could improve hedonic regression. To show the workings and the usefulness of the combination of externality analysis and spatial techniques, we created a simulation. In this simulation, spatial regression is superior to alternative regression methods if it yields better regression performance statistics than alternative regression methods. The simulation is meant to give more insights into the calculations of Chapters 7 and 8, and to show that spatial techniques
produce better regression performance statistics. Moreover, parameter estimates estimated with spatial techniques appear to be closer to the simulated values.

We simulate prices of 10,000 properties as a function of one physical attribute (square meters of living area), one locational attribute (location value), and one NIMBY-factor. The surface of living area is simulated in between 20 and 180 square meters, with an average of 100. We value each square meter at € 800. The location value is simulated by constructing a map with location values, as shown in Figure 6.2. This map has 3 favorable locations, for which location prices are highest. X- and Y-coordinates are simulated for each property, and linked to the map to obtain a location value. Since the map is constructed with a formula, exact location values can be calculated. Simulated location values range from about € 66,000 to € 211,000, with a mean of € 168,500.

**Figure 6.2: Simulated location values.**

Coordinates of an artificial NIMBY-factor, like the presence of a toxic waste site, are simulated as well. Properties directly adjacent to the waste site are assumed to sell at a huge discount of € 150,000 compared to properties located at a distance of 10 units. With each unit of distance, the discount declines with € 15,000, such that there is no NIMBY-effect if the distance to the NIMBY exceeds 10 units, the threshold distance. Adding this discount to the simulated location values yields the map shown in Figure 6.3.
Figure 6.3: Simulated location values, with negative externality.

Simulated property values, based on simulated property size, location values and the NIMBY-impact. Darker colors indicate lower values, bright colors reflect high values. Clearly visible are the three centers with values declining with distance, and the consequences of introducing a negative externality.
With a simulated constant of € 150,000, and the physical, locational and NIMBY attributes defined as before, the simulated house prices are calculated as:

\begin{equation}
\text{PRICE}_j = 150,000 + 800 \times \text{SIZE} + \Phi(X_j, Y_j) - 15,000 \times \text{PROX}_j
\end{equation}

The proximity \( \text{PROX} \), is calculated as a function of the straight-line distance between the property and the border of the NIMBY-impact area, as

\begin{equation}
\text{PROX}_j = \max \left[ \text{THRESHOLD} - \sqrt{(X_j - X_{\text{NIMBY}})^2 + (Y_j - Y_{\text{NIMBY}})^2}; 0 \right]
\end{equation}

The resulting prices are shown in Figure 6.4. Since we simulate reality, the simulated threshold distance of 10 units for which the externality does not influence property prices anymore should be found empirically.

**Simulation Results**

The simulated property values are estimated as a function of the physical and locational characteristic and the proximity of the NIMBY-object. Threshold distances of 9, 10 and 11 units are used for estimation. Results are shown in Table 6.1.

**Estimation without location information**

We start by performing a Maximum Likelihood estimation of the price regressed on property size and the NIMBY-proximity, thereby ignoring location values. As is shown in the first panel of Table 6.1, estimations of the implicit price per square meter are very close to the simulated price of € 800. Estimated constants are approximately 318,000. This is almost equal to the simulated constant of 150,000 plus the mean location value of 168,500. The best estimation for the ‘true’ NIMBY-impact of 15,000 per distance unit concerns the smallest impact-area, and the regression performance statistics suggest this is the best specification. Denote, however, that the simulated threshold was 10. Moreover, it is likely that part of the location value effect is captured by the estimate for the NIMBY-factor. The NIMBY is situated in an area with above average location values, causing the estimated NIMBY impact to be less negative.

**Estimation with location information by means of region dummies**

The second estimation uses region dummies, to capture the effect of different locations on property values. We created 16 equally sized regions of each 25 by 25 units of distance, and one dummy variable for each of these regions. Dummy variables show value 1 for a specific property if it is situated within the region. Simulated properties are scattered almost equally over all 16 regions. Adding dummies for 15 regions to the regression yields the results shown by panel B of Table 6.1. Note that the estimated implicit price per square meter is even closer to the simulated price of € 800 than in the first panel, and Likelihood Ratios are higher. Estimations of the constant now include the simulated constant of € 150,000 and the average location value for the region that was omitted from the regression because of singularity reasons. The regression performance statistics show that inclusion of location information really improves estimation.
Table 6.1: Estimation of simulated implicit prices of property attributes and the NIMBY-proximity.¹

### Panel A: Estimation without region dummies

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Threshold=9</th>
<th>LR</th>
<th>Threshold=10</th>
<th>LR</th>
<th>Threshold=11</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>317,744.4</td>
<td>32,463.9</td>
<td>317,843.1</td>
<td>32,432.8</td>
<td>317,919.0</td>
<td>32,372.1</td>
</tr>
<tr>
<td>SIZE</td>
<td>801.7</td>
<td>10,592.6</td>
<td>801.4</td>
<td>10,568.0</td>
<td>801.1</td>
<td>10,524.2</td>
</tr>
<tr>
<td>PROX</td>
<td>-13,159.3</td>
<td>703.2</td>
<td>-10,406.0</td>
<td>671.7</td>
<td>-8,228.9</td>
<td>612.2</td>
</tr>
</tbody>
</table>

*Adj R², rmse, Loglik*

### Panel B: Estimation with region dummies

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Threshold=9</th>
<th>LR</th>
<th>Threshold=10</th>
<th>LR</th>
<th>Threshold=11</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>280,085.6</td>
<td>28,756.8</td>
<td>280,117.2</td>
<td>28,731.2</td>
<td>280,156.5</td>
<td>28,623.3</td>
</tr>
<tr>
<td>SIZE</td>
<td>801.1</td>
<td>19,116.0</td>
<td>800.8</td>
<td>19,084.1</td>
<td>800.3</td>
<td>18,975.7</td>
</tr>
<tr>
<td>PROX</td>
<td>-14,992.2</td>
<td>1,976.0</td>
<td>-12,487.4</td>
<td>1,946.6</td>
<td>-10,477.8</td>
<td>1,829.3</td>
</tr>
</tbody>
</table>

*Adj R², rmse, Loglik*

### Panel C: Estimation with spatial autocorrelation (CAR, Delaunay)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Threshold=9</th>
<th>LR</th>
<th>Threshold=10</th>
<th>LR</th>
<th>Threshold=11</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>317,751.8</td>
<td>46,360.9</td>
<td>317,913.7</td>
<td>46,396.6</td>
<td>318,011.3</td>
<td>46,191.0</td>
</tr>
<tr>
<td>SIZE</td>
<td>799.5</td>
<td>41,573.2</td>
<td>799.5</td>
<td>41,620.2</td>
<td>799.4</td>
<td>41,436.6</td>
</tr>
<tr>
<td>PROX</td>
<td>-14,190.9</td>
<td>1,395.3</td>
<td>-12,310.7</td>
<td>1,441.9</td>
<td>-9,945.4</td>
<td>1,257.2</td>
</tr>
</tbody>
</table>

*Adj R², rmse, Loglik*

### Panel D: Estimation with spatial autocorrelation (SAR, Delaunay)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Threshold=9</th>
<th>LR</th>
<th>Threshold=10</th>
<th>LR</th>
<th>Threshold=11</th>
<th>LR</th>
</tr>
</thead>
<tbody>
<tr>
<td>CONSTANT</td>
<td>317,943.7</td>
<td>22,702.3</td>
<td>318,210.6</td>
<td>23,900.2</td>
<td>318,222.7</td>
<td>22,574.5</td>
</tr>
<tr>
<td>SIZE</td>
<td>800.0</td>
<td>69,552.3</td>
<td>800.0</td>
<td>70,864.7</td>
<td>799.9</td>
<td>69,398.9</td>
</tr>
<tr>
<td>PROX</td>
<td>-14,053.5</td>
<td>3,396.1</td>
<td>-14,599.5</td>
<td>4,706.9</td>
<td>-11,467.9</td>
<td>3,237.5</td>
</tr>
</tbody>
</table>

*Adj R², rmse, Loglik*

¹Estimation of simulated property values as a function of physical and locational attributes. All estimates are significantly different from zero, as Likelihood Ratio-tests all specify a probability of zero for achieving a higher χ². Simulated coefficients are 800 for SIZE, -15,000 for PROX, a constant of 150,000, and a threshold of 10. Results represent best out 5 simulations of 10,000 observations each.
precision; root mean squared errors are almost half compared to the previous estimation. Note that the price for the NIMBY-factor is closer to the simulated € 15,000. However, performance statistics again indicate an incorrect threshold distance of 9 units to be the most convenient.

**Estimation with location information by means of spatial autocorrelation**

Panels C and D of Table 6.1 show what happens if spatial autocorrelation is used. Now, property prices are regressed on just the property size and the NIMBY-proximity, so without region dummies. We use two alternative specifications, both using Delaunay-triangles to specify the weight matrix; while the so-called CAR-model shows important improvements compared to the estimation with region dummies, the SAR-model is even better. Root mean square errors are less than 4 percent of those resulting from the first regression without location information. Both the CAR- and SAR-model reveal the correct threshold level of 10. Moreover, the SAR-model yields the exact simulated implicit prices per square meter, and an implicit price per unit of proximity close to the simulated € 15,000. Note that fine-tuning of the SAR-specification by other weight matrix specifications could even improve the estimation performance.

**Conclusions for the simulation**

Even though capturing locational information in region dummies results in considerable enhancements of regression performance, this procedure is inferior to the use of spatial autocorrelation. Our results unambiguously show that spatial autocorrelation improves both the performance of the estimation process. It might therefore be very well suited for analysis of the price impact of negative externalities.

**6.6 Summary and conclusions**

In this chapter, we explained how hedonic regression can be used to estimate the impact of externalities on house prices, and discussed how methods applying spatial information could improve this analysis. Although studies showed how spatial techniques can improve hedonic regression, this combination has not been used in real estate literature. A simulation illustrated the principles of the spatial techniques applied in the next two chapters. These techniques yielded much better regression performance statistics, due to the more efficient estimates.