DRAFT PAPER

Modeling land speculation with rural-urban land use transitions

Paper for the 49th congress of the European Regional Science Association

25-29 August 2009, Lodz, Poland

Jasper Dekkers[a,]* and Piet Rietveld[a]

[a] Department of Spatial Economics, VU University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands

Abstract

Spatial planning policies can create submarkets for land, leading to an artificial scarcity for certain types of land use (usually urban land use) and higher land prices in that submarket, which can eventually spill over to other submarkets. In a spatial context, this is mainly the case in urban fringes, where rural land owners often ask a higher price for their land because of higher expectations for their land to be converted into urban land use.

Various general land market models have been developed in the past decades, also to specifically model the Dutch land market. Different approaches have been tried with more or less success. In this paper we describe an alternative approach to model the Dutch rural land market. As a case study area we take the province of Noord-Holland.

In our approach we use the hedonic pricing method in combination with a linear probability model and weighted least squares. The results show that this approach has an answer for a number of technical-methodological issues that occur in the models that gave us the idea for this analysis. Further, although the level of explained variance is rather low when we try to explain what factors influence the probability of a rural parcel becoming urbanised, the factors in general show the good signs and are significant.

Keywords: Land market, land-use transition, hedonic pricing method, linear probability model, Noord-Holland

JEL classification: C21, C51, Q24, R11

CONCEPT

DRAFT PAPER

1. INTRODUCTION

Spatial planning policies can create segmented submarkets for land, leading to an artificial scarcity for certain types of land use (usually urban land use) and higher land prices in that submarket, which can eventually spill over to other submarkets. In a spatial context, this is mainly the case in urban fringes, where rural land owners often ask a higher price for their rural land because of higher expectations for their land to be converted into urban land use. The price for rural parcels in the urban fringe is directly related to the probability that the parcels will become urbanized in the (near) future, the so-called transition probability (see Figure 1).

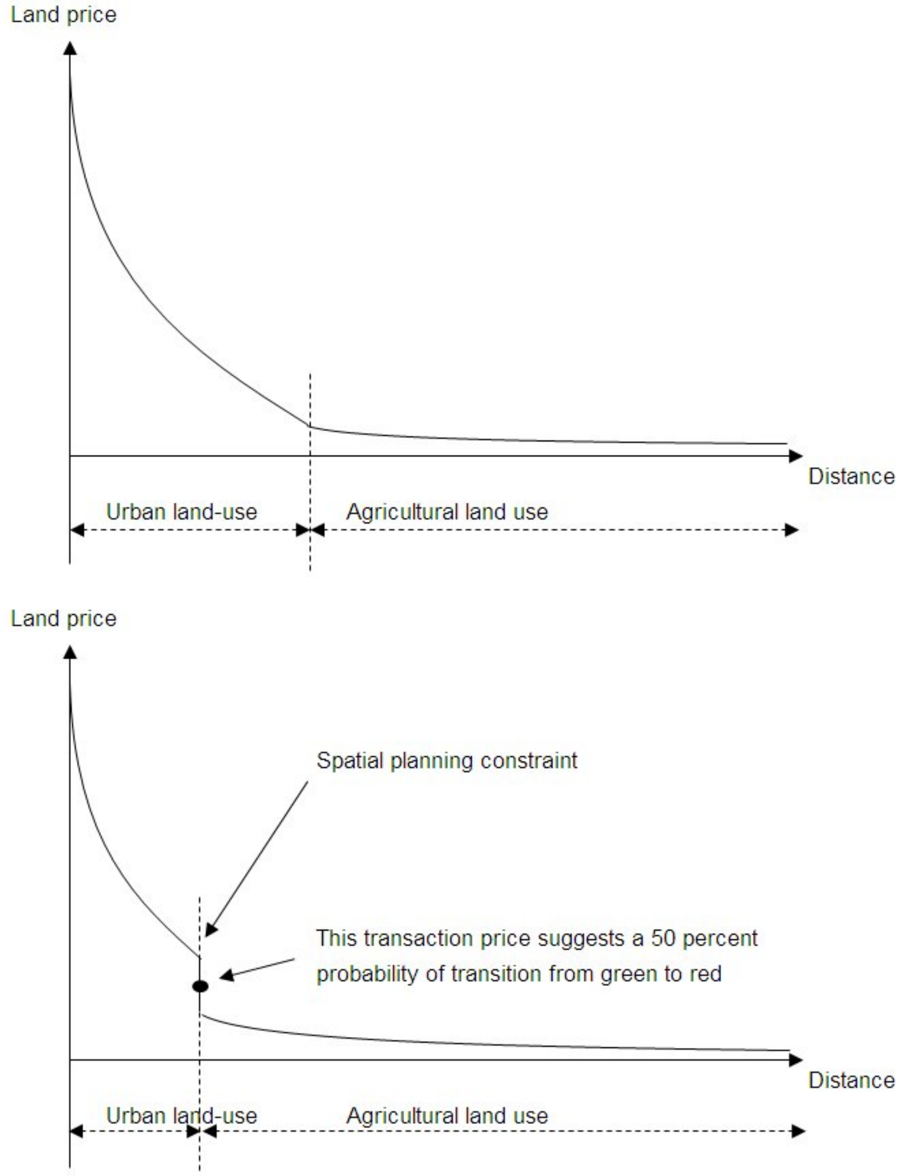

Figure 1. Relation between land price, land use and distance from city centre: (a) with no spatial planning constraint / (b) with spatial planning constraint.

In this paper, we will discuss two explanatory land market models for the Dutch land market. Both of them explain land prices or land prices per hectare and use both structural and spatial characteristics, including spatial planning constraints (Section 2). In Section 3 we then suggest an alternative modelling approach that combines the strong points of both models and explains the relation between transition probabilities and land prices in a more straightforward way, at least from an econometric point of view. Further, in Section 4 we apply this model to a case study, the province of Noord-Holland. And finally, in Section 5 we discuss the model outcomes and in Section 6 we then use the results from this model to derive a transition probability map for the entire province, so also for land that has not been traded, and we discuss how our method can be used in combination with the modelling of future land use change in order to support the spatial planning processes in the Netherlands. We end with some concluding remarks in Section 7.

2. TWO HEDONIC LAND MARKET MODELS

Both Buurman (2003) and Luijt et al. (2003) describe an explanatory model for the Dutch rural land market. Both models are filled with data from the InfoGroMa-database of the Government Service for Land and Water Management (DLG, part of the Ministry of Agriculture, Nature and Food Quality). This dataset is a subset of the Dutch cadastral database containing all transactions of parcels outside urban areas in the Netherlands. All transfers of ownership rights are being registered in this database. This immediately reveals one of the shortcomings of the database: it does not register options on parcels, meaning that options cannot be included in the analysis. This does not severely influence our analysis however, since the inclusion of the trading of options would just give us more information about the division of transaction profits (if any) among different actors, the final market price of a sold parcel is not influenced by the process of option trading.

Based on extensive literature research, both models include numerous actors and factors that are most likely to affect transaction prices of land. These can be divided into transaction characteristics (e.g. price, date of sale, type of buyer and seller), parcel characteristics (e.g. parcel size, soil quality) and spatial characteristics (e.g. accessibility measures, environmental measures, zoning designations). The general model is thus:

$$price_i = f(x_i) + \varepsilon_i$$

in which x_i represent all the transaction, parcel and spatial characteristics that influence the price and ε_i is the error term. Both models disregard transactions with one or more parcels with immobile property, since data on immobile property are not available.

The purpose of Buurman's model is to explain (spatial) differences in transaction prices of rural land parcels. A semi-logarithmic regression function has been estimated that relates the transaction values to the size of the transaction (the amount of land sold), the time of sale (in quarters, to correct for price increase in time), the parcel-, spatial-, and transaction characteristics. Translating the impact of parcel characteristics on the transaction value has been done by weighing the impact using parcel size. The resulting function specifies the contribution of the various features to the explanation of the transaction price:

$$\ln(price_i) = \ln(\beta_0) + \beta_1 \cdot \ln(size_i) + \beta_2 \cdot Time_i + \beta_3 \cdot ParcelChar_i \\ + \beta_4 \cdot SpatialChar_i + \beta_5 \cdot TransactionChar_i + \ln(\varepsilon_i)$$

The model has been applied in a case-study in the province of Noord-Brabant on over 6,000 transactions and performs well with an adjusted R^2 of 0.73 and all highly significant coefficients with the right signs.

Luijt et al. (2003) also describe a single equation logarithmic regression model for explaining land prices. They use slightly different parcel characteristics and do not include time and transaction characteristics. The model form is the same as that of Buurman:

$$\ln((price/ha)_i) = \ln(\beta_0) + \beta_1 \cdot \ln(size_i) + \beta_2 \cdot ParcelChar_i \\ + \beta_3 \cdot SpatialChar_i + \ln(\varepsilon_i)$$

A cross-section estimation of the model using 6,000 transactions nation-wide resulted in an adjusted R^2 of 0.29. This is considerably lower than the adjusted R^2 of Buurman's model. One explanation for this is the difference in dependent variables: The (adjusted) R^2 of a model tends to be substantially higher when explaining price sec instead of price per hectare or per square meter (ceteris paribus). We tested this assumption by analyzing the dataset of Buurman (2003) again. As mentioned, when explaining price sec, the model yields an adjusted R^2 of 0.73. When explaining price per hectare, and removing the explanatory variable ln(size) from the model, the adjusted R^2 drops to 0.43. This is probably caused by a decrease of total variance: the contribution of transaction size on the right side of the equation explains the difference between 0.43 and 0.73. Another reason for this result is probably the use of more explanatory variables (see Table 1).

Table 1. Overview of the variables used in both single-equation models

Characteristic	Buurman (2003)	Luijt *et al.* (2003)
Transaction price	Dependent variable	Dependent variable (per ha)
Transaction size	In square metres	In hectares
Time	Quarterly time dummies	-
Type of Buyer	Distinction between Relative, Farmer, National government and Municipality	-
Land is rented	Dummy (yes/no)	-
Land quality	Combination of soil type and ground water level	Soil quality according to Rental policy (pre-'95)
Distance to urban area	Distinction between urban areas and the Randstad (the highly urbanised western part of the country)	-
Spatial plans	Distinction: Built-up, Infra and Other	Built-up only
Regional policy	Provincial urban growth policy	-
Land use	Dominant land use types around sold parcels; distinction between Forest, Nature and Built-up	Distinction between Greenhouse horticulture and Land development areas
Agric. competition	-	Nr of strong agric. companies in the vicinity
Address density	-	Yes, source: CBS
Tax value (WOZ)	-	Value according to Law Immobile Property

A drawback of both models is the fact that not all characteristics are equally important for all buyers. For instance, project developers most likely do not value soil quality as much as farmers do. Therefore, Luijt et al. (2003) propose an alternative for the single equation model in the form of a two-step multiple equation model. The general model is re-written as follows:

$$price_i = \alpha_r + \beta_r \cdot x_{ri} + \varepsilon \qquad \text{in case of a red buyer, and}$$

$$price_i = \alpha_g + \beta_g \cdot x_{gi} + \varepsilon \qquad \text{in case of a green buyer}$$

In order to determine which equation to use on a transaction, transition probability must first be estimated using parcel and buyer characteristics. Given the fact that it is known from the dataset whether a buyer is red or green, in a first step the probability of a red buyer is estimated using parcel characteristics (influence of the nearest urban areas measured in terms of the Reilly-value which accounts for both distances to and sizes of nearest urban areas (Reilly, 1931); Spatial municipal plans for built-up and nature areas from the New Map of the Netherlands; Greenhouse horticulture, Recreation and Land Development areas):

$$prob_{\mathrm{ri}} = \frac{1}{1 + \mathrm{e}^{-(\beta_0 + \beta_1 \cdot \mathrm{X_i})}}$$

Transactions where $prob_{ri} \leq 0.5$ are supposed to have a green buyer involved. The price of such transactions is analysed in step 2 using green parcel characteristics (Potential agricultural profit per hectare; Agricultural competition measured by the number of strong agricultural companies in the vicinity; Soil quality according to Rental policy (pre-1995)) and:

$$\ln(price / ha)_{gi} = \ln(\beta_{g0}) + \beta_{g1} \cdot (ParcelChar_{gi}) + \beta_{g2} \cdot prob_{ri} + \varepsilon$$

The prices of transactions in which a red buyer is involved ($prob_{ri} > 0.5$) are also analysed in this step using red parcel characteristics (Address density; tax-value of property (WOZ-value), Transaction size) and $prob_{ri}$:

$$\ln(price / ha)_{ri} = \ln(\beta_{r0}) + \beta_{r1} \cdot (ParcelChar_{ri}) + \beta_{r2} \cdot prob_{ri} + \varepsilon$$

Graphically, this model can be depicted as follows:

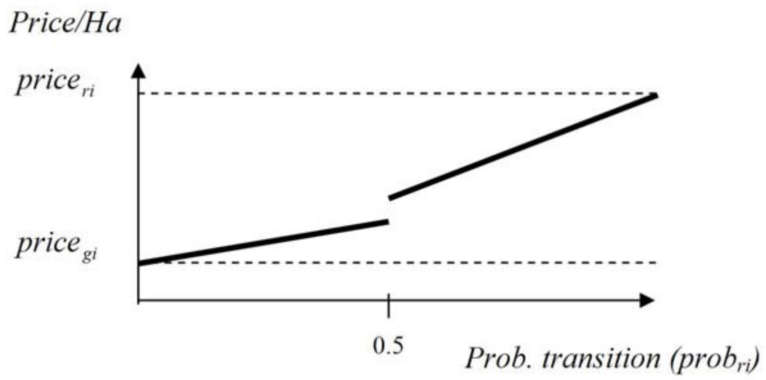

Figure 2. Two-step multiple equation transition probability model of Luijt *et al*. (2003).

An important drawback of using an artificial cut-off value at $prob_{ri} = 0.5$ is the uncertainty for land transactions with probabilities around this cut-off value of being allocated in the right equation thus reducing the explanatory power of the model around this point. Also, it is not guaranteed whether the two functions connect at all at $prob_{ri} = 0.5$. In the case study of Luijt et al. (2003), the results of the estimations show that the two functions certainly do not connect.

3. AN ALTERNATIVE APPROACH

In this section an alternative approach is developed. The strong points of both models (that of Buurman and that of Luijt) are being integrated in this model: Buurmans large set of explanatory variables and the probabilistic approach of Luijt c.s..

The aims of this so-called linear probability model are:

- Estimate hedonic price models for the green- and red submarkets separately, in order to do justice to the segmented nature of land markets;
- Compute transition probabilities from green to red in order to describe market expectations on observed transitions;
- Determine the contribution of explanatory variables to these transition probabilities;
- Use this model to predict transition probabilities for zones where no transactions took place.

Several analytical steps are needed to obtain the desired results with this approach. These steps can be summarized as follows:

Summary of steps in modelling approach

Step 1a.	Find subset of transactions for which it is clear on an *a priori* basis that the new use of the parcels concerned is red. These are called 'red transactions'.
Step 1b.	Same for green transactions. This leads to a set of 'green transactions'.
Step 1c.	The remaining transactions are called 'uncertain transactions'.
Step 2a.	Estimate hedonic price model for subset of red transactions.
Step 2b.	Same for green transactions.
Step 3a.	Use the red price model estimated in *step 2a* to predict the value that *all* parcels (red, green, uncertain), would have according to red market conditions. This leads to predicted red prices for all parcels.
Step 3b.	Same for green price model. This leads to predicted green prices for all parcels.
Step 4.	Compute for all transactions the transition (from green to red) probability as $prob_{ri}$ as the ratio of the difference between the actual price minus the predicted green price divided by the difference between the predicted red price and the predicted green price.
Step 5.	Use a linear probability model to explain the values of the transition probability defined in *step 4* by means of a set of explanatory variables.
Step 6.	Compute the transition probability of all zones in the region concerned, so including the zones where no transactions took place.

Figure 3. Summary of steps in linear probability approach

We now first explain our approach step by step in more detail and then report our results from a case-study that we did to test this approach in the province of Noord-Holland (Section 4). We will again use the cadastral InfoGroMa-database we mentioned in Section 2, but we will now disregards family transactions. For Noord-Holland this database contains 2,685 parcels in 1,625 transactions.

Explanation of the proposed method

Suppose a potential buyer foresees a transition probability from green to red of $prob_{ri}$. Suppose that we know the prices of the parcels concerned when they would fall in the green market segment (*$price_{gi}$*) and when they would be in the red segment (*$price_{ri}$*). How much is the buyer willing to offer? Graphically, this model can be depicted as follows (linear variant):

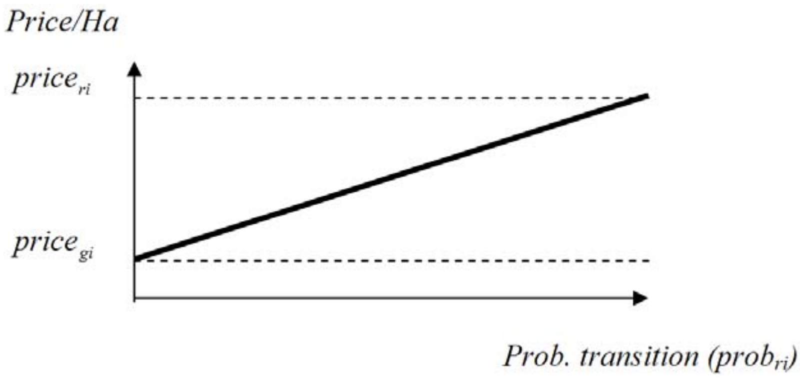

Figure 4. Single-equation transition probability model.

With a large spatial database of transactions, we can deduce $prob_{ri}$ from the observed $price_i$ and explain this transition probability by using parcel and transaction characteristics:

$$price_i = (1 - prob_{ri}) \cdot (\alpha_g + \beta_g \cdot x_{gi}) + prob_{ri} \cdot (\alpha_r + \beta_r \cdot x_{ri}) + \varepsilon_i$$

Where *g* is green, *r* is red, x_g represents variables that are relevant in a green setting (i.e. buyer is a farmer, soil quality, et cetera), and x_r represents variables that are relevant in a red setting (i.e. accessibility for commuting, et cetera).

Step 1a – Find red subset of transactions

The problem is that we do not know the transition probability in this formula. To solve this, we can select a subsample of transactions for which we know almost certain that the parcels will become developed, i.e. $prob_{ri} = 1$. In this case, we have chosen to take a sample of 10 percent of the transactions that have the highest price per square metre.

Step 1b – Find green subset of transactions

For the other transactions, we do not know $prob_{ri}$ with certainty. We can, however, in a similar approach as with step 1a take a sample of transactions where, according to expert judgement, $prob_{ri}$ is very close to zero, for instance because a parcel lies within the Ecological Main Structure and another policy restricting urbanisation applies in that area. Similarly as with the red sub sample, we have chosen to take a sample of 10 percent of the transactions that have the lowest price per square metre.

Table 2 gives some descriptive statistics for both the red and green subsamples. Realizing that with these subsamples the observed differences are not statistically significant, we see large differences in transaction price and transaction price per square metre between the two subsamples. Further, no leased land is found in the red subsample. Next, on average transactions in the red subsample lie closer to the Randstad and to urban areas and they more often lie within 200 metres of a main road. Finally, in the red subsample on average a higher share of the transactions lies within urban development plans or classify as land use types urban green or greenhouse horticulture. In contrast, for the green subsample on average a higher share of the transactions lies within the Ecological Main Structure. The same applies for buffer zones and for the land use type nature. For Belvedere policy zones, there virtually is no difference between the two subsamples.

Step 1c – Remaining transactions

The remaining 80 percent of the transactions are only used from step 3 onwards. Figure 5 displays four maps containing the locations of the sold parcels for the two sub samples, the remaining parcels and all parcels together. Amsterdam is located at the bottom of the maps, the harbor area is just visible. What is clearly visible is that the red parcels (the upper-right map in Figure 5) are in general concentrated near urban areas, while the green parcels (the lower-left map) are in general further away from urban areas.

Table 2. Summary statistics for the red and green subsamples related to step 1a and 1b. For each variable, the upper (lower) figure relates to the red (green) subsample.

Variable	Min.	Max.	Avg.	Std. dev.
Transaction characteristics				
Transaction price (x 1,000 Euro)	25.63	29,743	846.80	2,580
	1.53	277.58	42.19	58.12
Transaction price per square metre (Euro/m^2)	7.78	153.15	20.22	19.98
	0.16	0.67	0.50	0.13
Year 1998 (total number; red/green)				9 / 65
Year 1999 (total number; red/green)				34 / 37
Year 2000 (total number; red/green)				38 / 31
Year 2001 (total number; red/green)				45 / 6
Year 2002 (total number; red/green)				36 / 20
Leased land (0/1)	0.00	0.00	0.00	0.00
	0.00	1.00	0.56	0.50
Structural characteristics				
Surface area (x 1,000 m^2)	2.50	580.76	41.20	69.60
	3.12	569.20	78.67	101.03
Spatial characteristics				
Share of a transaction that is located in an urban	0.00	1.00	0.32	0.47
development zone (New Map of the Netherlands; %)	0.00	1.00	0.01	0.08
Distance to the Randstad (km)	0.00	77.29	16.02	17.79
	0.00	76.53	25.53	20.07
Distance to the nearest built-up area (km)	0.00	1.81	0.29	0.29
	0.00	3.36	0.65	0.55
Distance to main road < 200 metre (0/1)	0.00	1.00	0.15	0.36
	0.00	1.00	0.08	0.28
Share of a transaction's land use that is urban green (%)	0.00	1.00	0.57	0.23
	0.00	1.00	0.01	0.08
Share of a transaction's land use that is greenh. hortic. (%)	0.00	1.00	0.30	0.16
	0.00	0.00	0.00	0.00
Land development project (BBL/DLG) (0/1)	0.00	1.00	0.25	0.44
	0.00	1.00	0.35	0.48
Share of a transaction that is located in the provincial	0.00	1.00	0.25	0.44
ecological main structure (PEHS; %)	0.00	1.00	0.59	0.49
Share of a transaction that is located in a buffer zone (%)	0.00	1.00	0.12	0.32
	0.00	1.00	0.19	0.39
Share of a transaction that is located in a Belvedere policy	0.00	1.00	0.09	0.28
Zone (UNESCO world heritage area; %)	0.00	1.00	0.08	0.27
Share of a transaction's land use that is nature (%)	0.00	1.00	0.01	0.11
	0.00	1.00	0.06	0.23

Note: The subsamples are taken from the dataset with transaction data in Noord-Holland (period 1998-2002; N_{sample}=1,625; N_{red}=162; N_{green}=159)

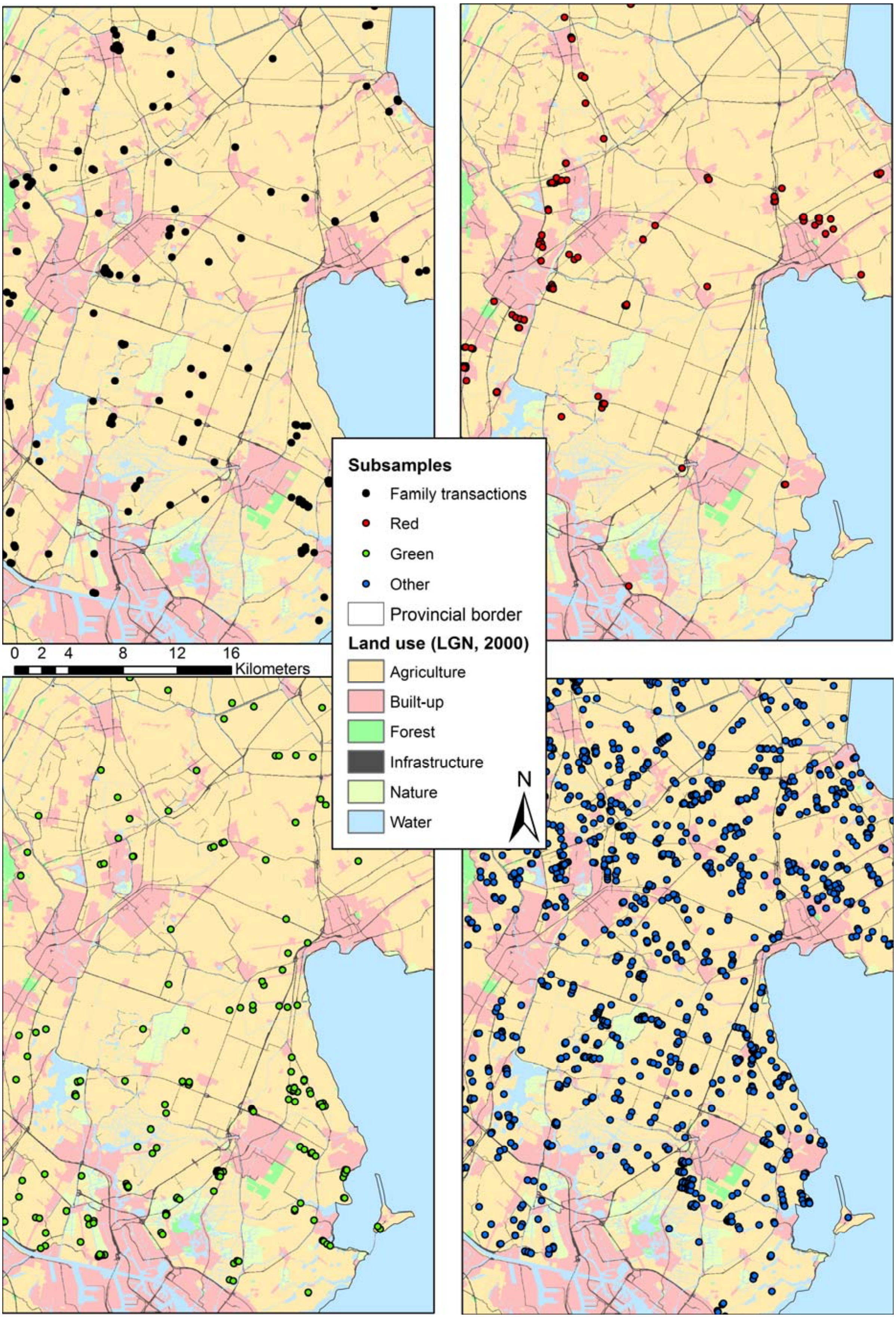

Figure 5. locations all geocoded rural land transactions without immobile property in the period 1998-2002 and land use in 2000 in a part of the province of Noord-Holland.

Step 2a and 2b– Estimate hedonic model for red and green transactions

Since for the red sample selected in step 1a we know that $prob_{ri} = 1$, this means we can simplify the basic equation (Section 1) and estimate for the red transactions:

$$price_i = \alpha_r + \beta_r \cdot x_{ri} + \varepsilon_i$$

This leads to estimates $\hat{\alpha}_r$ and $\hat{\beta}_r$.

In the case of the green transactions we know that $prob_{ri} = 0$, so we can again simplify the basic equation and estimate:

$$price_i = \alpha_g + \beta_g \cdot x_{gi} + \varepsilon_i$$

This leads to estimates $\hat{\alpha}_g$ and $\hat{\beta}_g$.

Steps 3a and 3b – Estimate red and green transaction prices

Next, once we know $\hat{\alpha}_r$ and $\hat{\beta}_r$ and $\hat{\alpha}_g$ and $\hat{\beta}_g$, we can compute:

$$\widehat{\mathrm{price}}_{\mathrm{ri}} = \hat{\alpha}_{\mathrm{r}} + \hat{\beta}_{\mathrm{r}} \cdot \chi_{\mathrm{ri}} \quad \text{and} \quad \widehat{\mathrm{price}}_{\mathrm{gi}} = \hat{\alpha}_{\mathrm{g}} + \hat{\beta}_{\mathrm{g}} \cdot \chi_{\mathrm{gi}}$$

in which $\hat{\beta}_r$ and $\hat{\beta}_g$ are vectors where the number of columns (N) equals the number of explanatory variables in the matrices χ_{ri} and χ_{gi}.

Step 4 – Estimate transition probability

In the following step of the analysis, for the remaining parcels i (where $0 < prob_{ri} < 1$) we have:

$$price_i = (1 - prob_{ri}) \cdot \widehat{price}_{gi} + \varepsilon_{gi} + prob_{ri} \cdot \widehat{price}_{ri_i} + \varepsilon_{ri}$$

After rewriting this formula, we can compute $prob_{ri}$ using $\mathrm{price_i}$ and the earlier determined vectors of $\widehat{price}_{ri}$ and $\widehat{price}_{gi}$ (Figure 6):

$$prob_{ri} = \frac{price_{\mathrm{i}} - \widehat{price}_{\mathrm{gi}} - (\varepsilon_{gi} + \varepsilon_{ri})}{\widehat{price}_{\mathrm{ri}} - \widehat{price}_{\mathrm{gi}}}$$

leading to $\widehat{\mathrm{prob}}_{\mathrm{ri}}$ assuming $\varepsilon_{gi} = 0$ and $\varepsilon_{ri} = 0$.

Step 5 – Explain transition probability

Finally, we can estimate the factors that have an impact on $prob_{ri}$. Let these factors be denoted as z_i (mainly overlapping with χ_i, only the surface and time dummy variables are excluded). We have to take into account the fact that there can be observations where $price_i < \widehat{price}_{gi}$ or $price_i > \widehat{price}_{ri}$. This can be related to the situation where $\varepsilon_i \neq 0$. Plausible causes for the price dropping below the green price can be for instance soil contamination. In case the value rises above the red, urban price, we can think of for instance overestimation of parcel value by a buyer. In order to include these 'special cases' in the estimation, we use a linear probability model in combination with weighted least squares (WLS):

$$\text{prob}_{ri} = \frac{\hat{A}_i - (\varepsilon_{gi} + \varepsilon_{ri})}{\hat{\text{B}}_\text{i}} = \text{c}\, z_\text{i} + \mu_i$$

Where:

$$\hat{A}_i = \widehat{price}_\text{i} - \widehat{price}_\text{gi}\,,\ \hat{\text{B}}_\text{i} = \widehat{price}_\text{ri} - \widehat{price}_\text{gi} \text{ and } \quad \mu_i = \frac{(\varepsilon_{gi} + \varepsilon_{ri})}{\hat{\text{B}}_i}$$

If we ignore $-(\varepsilon_{gi} + \varepsilon_{ri})$ in the first formula above for the moment, we get:

$$\text{prob}_\text{ri} = \frac{\hat{A}_i}{\hat{\text{B}}_\text{i}} = \text{c}\, \text{z}_\text{i} + \frac{1}{\hat{\text{B}}_i}$$

Where c is the vector to be estimated and z_i is a vector of all the model variables. For the WLS-analysis we can estimate the weights on this set of variables by calculating $1/\hat{\text{B}}_\text{i}$. This method allows values of prob_ri above 1 and below 0. WLS is a method of regression, similar to ordinary least squares (OLS) in that it uses the same minimization of the sum of the residuals:

$$S = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

However, instead of weighting all observations equally, they are weighted such that observations with a greater weight contribute more to the fit:

$$\frac{1}{\hat{B}_i} = \frac{1}{\hat{price}_{ri} - \hat{price}_{gi}}$$

Step 6 – Compute transition probability for all zones

In this final step we can use the WLS model coefficients in combination with the underlying spatial data sets of the explanatory variables to calculate the transition probability for all zones in the region concerned. This can be done by rasterizing all data sets on, for instance, a 25 metre grid.

4. SETTING UP THE CASE STUDY

Description of the study area

When looking at the land market in Noord-Holland (Figure 6), we see many submarkets in this province. It contains much agricultural areas to which forms of nature protection policies and/or nature development plans apply. Also, there is much horticulture and flower bulb land, which is relatively expensive. Then there is the national airport, Schiphol Amsterdam airport that is assumed to have a large impact on land use and prices of its surrounding areas. We think that it is just this heterogeneity in land uses and prices that will be an interesting setting in which to apply the model.
The surface of the province of Noord-Holland covers 4,059 square kilometres, of which 2,657 square kilometres of land surface. This equals 7.8 percent of the land surface of the Netherlands. In 2000, 2.5 million people were living in Noord-Holland, which makes it the second province of the Netherlands with regard to the number of inhabitants. Approximately 19 percent of the national Gross Domestic Product (GDP) is being generated in Noord-Holland (CBS, 2000). Table 3 shows that the Commercial services sector is relatively important in Noord-Holland and that the Manufacturing and 'Agriculture, forestry and fishery' sectors are relatively less important.
Like in the rest of the Netherlands, in Noord-Holland agricultural land use is deceasing and urban land uses and also nature are increasing. Compared to the Netherlands, the province of Noord-Holland has less forest and nature and more built-up area.

Figure 6. The province of Noord-Holland.

Table 3. Production structure of Noord-Holland and the Netherlands (Source: CBS, 2000)

	Gross added value 1999 (%)	
	Noord-Holland	the Netherlands
Agriculture, forestry and fisheries	1.6	2.8
Industry	17.0	25.5
Commercial services	59.3	49.0
Government and healthcare	22.2	22.8
Total	100.0	100.0

Selecting model variables

The variables we use overlap to some extent with the variables we used in our explanatory land - market model for the province of Noord-Holland as described in Dekkers et al. (2004). That model is the same model as Buurman (2003) used. The difference is that our current analysis not exclusively focuses on the explanation of rural land prices, but rather on rural-urban transition probability. Therefore, we include factors that we expect to have an influence on this transition. First, we include transport noise as a factor in the red explanatory hedonic pricing model by defining a dummy-variable that has the value 1 when a parcel is located less than 200 metres away from a main road. Then, we also include information on land use: built-up, urban green and greenhouse horticulture in the red model, and nature in the green model. Finally, in the green explanatory model we take into account whether or not parcels lie in a buffer zone (RNP, 1958) or in a Belvedere/UNESCO zone (OCW *et al.*, 1999). We expect these policies to negatively influence land prices because they restrict agricultural use to a more or less extent. We tested various model and variable specifications. We choose a semi-logarithmic model specification since this is the most widely used form. We also tested other variables related to, for instance, soil quality, accessibility and spatial policies aimed at nature and/or recreation development, but these factors were not significant.

5. ESTIMATION RESULTS

The estimation results (Table 4) give us confidence for both the red and the green model: the coefficients show the expected signs and most of them are significant. First, in the red model we see that parcels that are closer to urban areas have a higher price. Next, parcels that are less than 200 metres away from a main road have a lower price than parcels located further away, the most probable reason being the experience of transport noise in this area. Further, when the land use of a transaction is urban green or greenhouse horticulture, this positively influences parcel prices. In the green model we see that leased land is cheaper, as is land that is in a land development project area of BBL/DLG or in the provincial ecological main structure. And when a parcel lies in a buffer zone, we indeed see a negative influence on the land price, confirming our expectation. The coefficient of the Belvedere policy zone also has a negative sign, but this variable is not significant.

Table 4. Results for the subsample estimates in the formulas related to step 2a and 2b

	Red (N=162)		Green (N=159)	
Variable	Coeff.	Sign.	Coeff.	Sign.
LN(price) – dependent variable	-	-	-	-
Constant	2.817	***	-0.990	***
LN(surface)	0.979	***	1.044	***
Year 1999	0.260		-0.029	
Year 2000	0.312		0.081	
Year 2001	0.305		-0.043	
Year 2002	0.205		0.177	**
Share of a transaction that is located in an urban development zone (New Map of the Netherlands; %)	0.195	*		
Distance to the Randstad (km)	-0.003			
Distance to the nearest built-up area (km)	-0.370	**		
Distance to main road < 200 metre	-0.306	**		
Share of a transaction’s land use that is urban green (%)	0.364	*		
Share of a transaction’s land use that is greenh. hortic. (%)	0.206			
Leased land			-0.156	***
Land development project (BBL/DLG)			-0.097	*
Share of a transaction that is located in the provincial ecological main structure (PEHS; %)			-0.111	*
Share of a transaction that is located in a buffer zone (%)			-0.211	***
Share of a transaction that is located in a Belvedere policy zone (UNESCO world heritage area; %)			-0.072	
Share of a transaction’s land use that is nature (%)			-0.147	
s.e. of regression	0.585		0.276	
R-square	0.799		0.955	
Adjusted R-square	0.785		0.977	

Note: *** = significant at 0.01; ** = significant at 0.05; * = significant at 0.10

We subsequently use the coefficients of the red and green model for the estimation of red and green transaction prices and for the computation of the rural-urban transition probability. Two graphs (Figures 7 and 8) show the transaction prices (X-axis) plotted against the estimated prices (Y-axis) for red respectively green transactions. The red model somewhat overestimates land prices. Further, the slope of the trendline is below 45 degrees, which is to a large extent caused by the six outlier observations with real land prices ranging between 65 and 154 euro/m^2. The green model performs quite well with the slope of the trendline being near 45 degrees.

Figure 9 shows that most calculated transition probabilities lie between 0 and 1, meaning that the WLS estimation does not have to correct for a large number of probabilities outside this range: there are no probabilities below 0 and only 3.5 percent of the observation's probabilities is above 1. The regular OLS estimation explaining $prob_{ri}$ should therefore not differ too much from the WLS estimation. We do not include the time-dummies in the WLS estimation since they do not explain transition probability in any way. In fact, for the time being, in our WLS model we disregard the timing of when a rural-urban transition will occur.

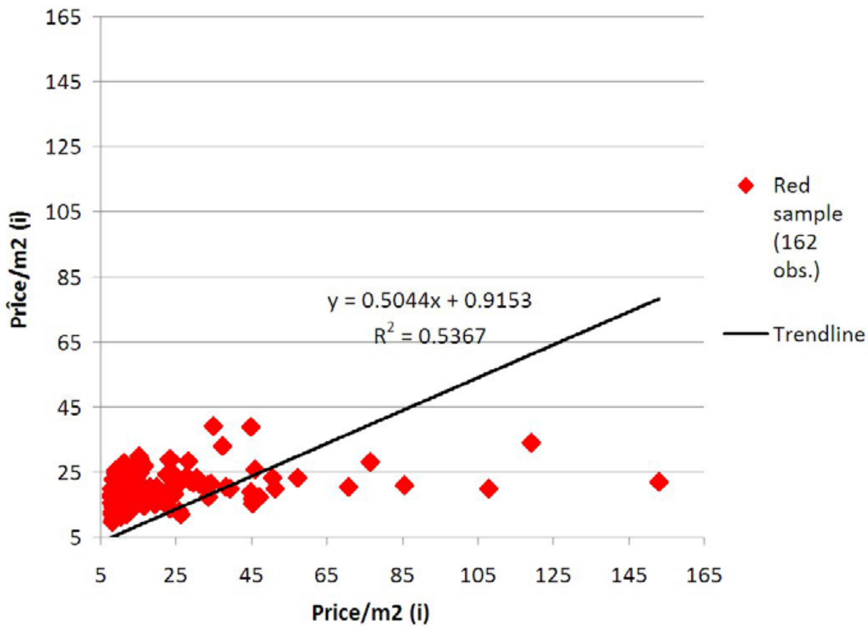

Figure 7. Plot of real versus estimated transaction price/m^2 for the red subsample.

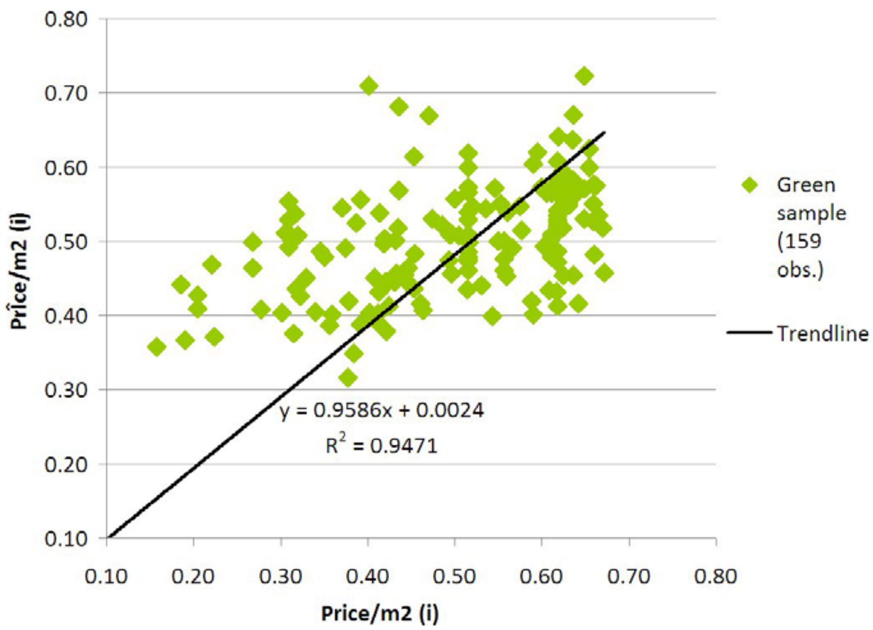

Figure 8. Plot of real versus estimated transaction price/m^2 for the green subsample.

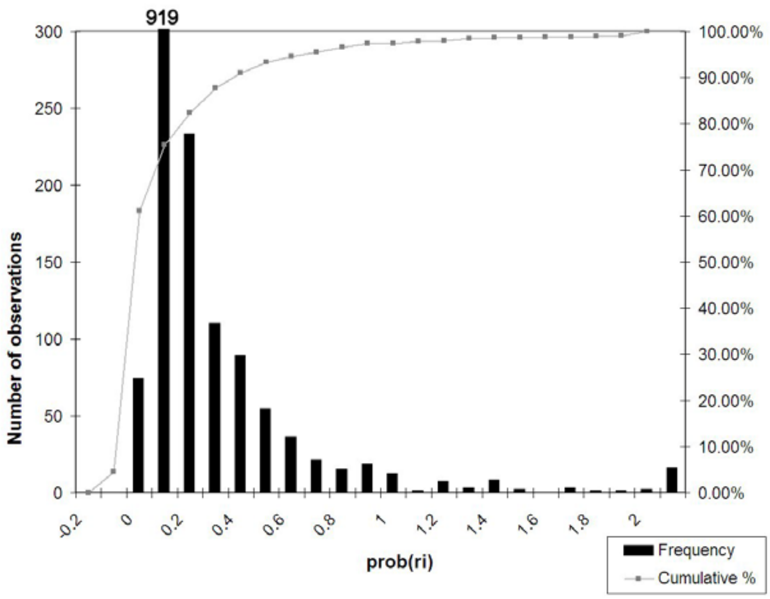

Figure 9. Histogram of calculated rural-urban transition probabilities.

As the results of the OLS and WLS analysis of vector c in Table 5 show, they do indeed not differ much. The results show that the coefficients of the 'red variables' explain the variables well: when a parcel lies in an area with urban development plans, the probability of a green parcel to become urbanised ($prob_{ri}$) is positively influenced; the further away from the Randstad, the lower the $prob_{ri}$. The same goes for the distance to the nearest built-up area, although this effect is not significant. Further, when a transaction takes place in an area where existing land use is predominantly built-up, urban green or greenhouse horticulture, the transition probability is influenced positively to a more or less extent.

The ‘green variables’ tell us that when land is leased and/or when parcels are in an area where the Government Service for Land and Water Management of the Netherlands (DLG; a department of the Netherlands Ministry for Agriculture, Nature Management and Fisheries (LNV)) has land development plans, this negatively influences $prob_{ri}$. Also, parcel lying in the provincial ecological main structure, have a lower transition probability. Next, the influence of a buffer zone on $prob_{ri}$ is negative. The buffer zone policy has a strong protective value. When the buffer zones where established in the 1960s, the original intent was that agriculture would be a driving force in these areas, in combination with nature and recreational developments. Over the years, the national government has changed it strategy more and more toward actively acquiring land for nature development. Next to that, buffer zones are in general not very large areas, and they are to be found in areas between major urban agglomerations where urbanisation pressures are already high. Scale increase for agricultural businesses is hard to realise given all these reasons, therefore these areas have less potential for modern farming then elsewhere. A same kind of reasoning applies on the Belvedere/UNESCO areas: this policy also has a highly protective value, farming activities have difficulty expanding here and multiple restrictions apply. Therefore the transition probability is influenced negatively.

Table 5. Results for the WLS estimation of vector c following formulas 17 and 18

	Vector c (OLS) (all 1,625 obs.)		Vector c (WLS) (all 1,625 obs.)[1]	
Variable	Coeff.	Sign.	Coeff.	Sign.
Prob(ri) – dependent variable	-	-	-	-
Constant	0.379	***	0.213	***
Urban development plans (New Map of the Netherlands)	0.358	***	0.185	***
Distance to the Randstad (km)	-0.004	***	-0.002	***
Distance to the nearest built-up area (km)	-0.044	**	-0.022	*
Distance to main road < 200 metre	0.032		0.022	
Share of a transaction’s land use that is urban green (%)	0.049		0.016	
Share of a transaction’s land use that is greenh. hortic. (%)	0.243	*	0.118	
Leased land	-0.137	***	-0.080	***
Land development project (BBL/DLG)	-0.067	***	-0.038	***
Share of a transaction that is located in the provincial ecological main structure (PEHS; %)	-0.051	**	-0.030	**
Share of a transaction that is located in a buffer zone (%)	-0.112	***	-0.066	***
Share of a transaction that is located in a Belvedere policy zone (UNESCO world heritage area; %)	-0.133	***	-0.075	***
Share of a transaction’s land use that is nature (%)	-0.064		-0.036	
s.e. of regression	0.384		0.384	
R-square	0.294		0.296	
Adjusted R-square	0.289		0.290	

Note: *** = significant at 0.01; ** = significant at 0.05; * = significant at 0.10

6. APPLICATION OF MODEL RESULTS

Calculation of a transition probability map

We can use the model results to calculate a transition probability map for all parcels, sold and not sold, in the entire province of Noord-Holland (see step 6 in figure 3). The map legend in Figure 10 shows that the range of probability values is rather low, going from -2.58 to a maximum of 0.32 where we would expect more values between 0 and 1. An explanation for these low probability lies in the fact that our model does not account for the large impact of *uncertainty* of land use change *over time*: a buyer buys a parcel now, then has to wait for, for example, ten years before a decision is taken whether or not a parcel is allowed to change from rural to urban land use. So next to uncertainty whether the change will occur, also the fact that a buyer determines his offer using a discount rate lowers his bid price and thus the transition probability values in our model.

Modelling future land use

The understanding of the functioning of the land market and the occurrence of current land use, especially in the urban-rural transition zone, can help to improve our ability to model future land-use change. The quantification of land-use change through the use of land-use models is very important for evaluating the effects of spatial policy (MNP, 2004; Borsboom-van Beurden *et al*., 2005). There are plenty land use models available for simulating land-use change. Most models only simulate urban or rural land-use types. Because of the increasing overlap between policy fields in the Netherlands and the importance of rural-urban transitions in land-use change, it is desirable to have models that are suitable for integrated scenario analyses. For the Netherlands, several of these integrated models are available, one of them being the Land Use Scanner (Hilferink and Rietveld, 1999). This economics-oriented probabilistic model uses a logit-function to simulate demand for and supply of land in an iterative process. The way current and future developments, spatial policies and spatial pressure on land are modelled is subject to scientific debate. The general consensus is that there is room for improvement in both practical and fundamental parts of the model. Practical improvements can be made in particular with regard to sensitivity analysis of results to the scale or resolution of modelling, calibration and validation issues and a more fundamental improvement can be the restructuring or enhancement of the economics-theoretical foundations of the model.

Because of the big difference between processes in the land market and factors that affect land-use change, strengthening this economics-theoretical link is quite a challenge. Some land-use models have a solid theoretical foundation, see for instance Anas (1982). But many land-use models, in particular models that integrate urban and rural land use, lack such a basis (see a.o. Lee Jr, 1973; Wegener,

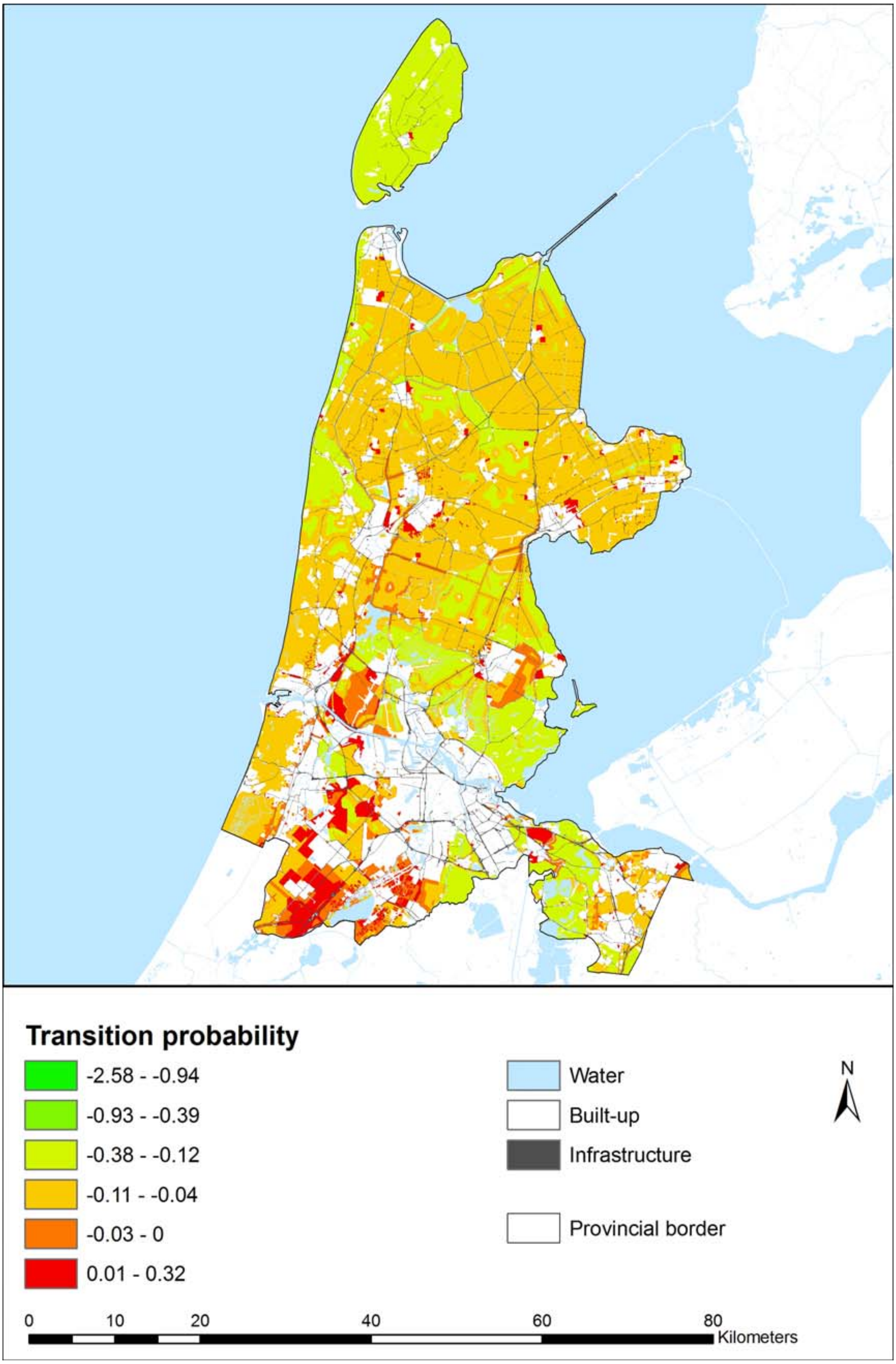

Figure10. Rural-urban transition probability in the province of Noord-Holland.

1994). However, we think that economic land market models can provide the necessary theoretical backing for these models. Consider the bid-rent theory and how, based on Alonso's urban model (Alonso, 1964), McFadden (1978) developed an empirically more practical *stochastic* maximization model. If we apply this to land-use this means that the model computes for each location the *probability* of a certain land-use type by comparing its utility with the utility of all other possible land-uses on that location. This type of model also forms the foundation for the integral (i.e. including both urban and rural land use types) land-use model Land Use Scanner.

In the doubly constrained logit model that the Land Use Scanner uses, the expected amount of land in cell c that will be used for land-use type j can be formulated as:

$$M_{cj} = a_j \cdot b_c \cdot \exp(\beta \cdot s_{cj}) \qquad (1)$$

in which:

M_{cj} is the amount of land in cell c expected to be used for land-use type j;

a_j is the demand balancing factor (condition 1) that ensures that the total amount of allocated land for land-use type j equals the sectoral claim;

b_c is the supply balancing factor (condition 2) that makes sure the total amount of allocated land in cell c does not exceed the amount of land that is available for that particular cell;

β is a parameter that allows for the tuning of the model. A high value for β makes the suitability more important in the allocation and will lead to a more mixed land use pattern; and

s_{cj} is the suitability of cell c for land-use type j, based on its physical properties, operative policies and neighbourhood relations.

For a more extensive description of the Land Use Scanner model, we refer to Dekkers and Koomen (2007).

A difference between the Land Use Scanner and our modelling approach is that where the Land Use Scanner can model up to 15 land use types[1], our model only discerns urban versus rural land use. But our transition probability model is also a stochastic model that calculates probabilities. In the Land Use Scanner, the values of the different suitability factors for each land use type are determined through expert judgment. More recently, there have been some calibration and validation attempts of the model. In this process the suitability values were derived based on historical developments and and subsequently applied in an extrapolation of current trends to determine the probabilities for future

[1] The Land Use Scanner actually can model much more than 15 land use types, but modelling more than 15 land use types is not considered to generate realistic empirical results.

land use (Loonen and Koomen, 2008). An interesting application of our model results would be to insert the WLS model variables and their respective coefficients as suitability factors and values for urban and rural land use. Next to the WLS model variables and coefficients, the maps generated in step 6 for all model variables also need to be included in the Land Use Scanner. We can then compare the probable land use change as calculated by the model with the actual land use change that occurred. When we also carry out two similar calculations using (a) expert judgement and (b) calibration-validation techniques, we can then learn more about what land use changes do and do not occur in all three simulations. By comparing the similarities and difference in the model outcomes, we can increase our understanding of what (spatial) factors are more certain and likely to occur compared to others.

7. CONCLUDING REMARKS

This paper discusses attempts to model the possible occurrence of spillover effects between rural-urban submarkets in the urban fringe. These submarkets are created by spatial planning policies restricting urban land development.

Earlier attempts to model the Dutch land market using hedonic pricing techniques only try to explain what (spatial and non-spatial) factors contribute to land prices and/or try to explain the price difference between rural parcels with an agricultural or urban designation. In this paper we take these approaches one step further and develop a model that uses the hedonic pricing method in combination with a linear probability model and weighted least squares. This model uses the factors that explain difference in land prices to analyse the probability of rural land becoming urbanised.

The results show that our approach has an answer for a number of technical-methodological issues that occur in the models that gave us the idea for this analysis: by not directly including different buyers in the analysis we get around the problem that all characteristics are equally important for all buyers. Next to that, the artificial cut-off between separate ‘red’ en ‘green’ explanatory models at $prob_{ri}$=0.5 in the model of Luijt *et al.* (2003) is removed. Further, although the level of explained variance is not that high when we try to explain what factors influence the probability of a rural parcel becoming urbanised, the factors in the WLS estimation of vector c show the good signs and are in general significant.

Recommendations

Several recommendations to improve our model can be made. First, we expect that the explanatory power can be improved by, for instance, adding more explanatory variables. Second, the nature-related spatial planning variables (based on the New Map of the Netherlands data from 2002) can be re-examined and cleaned further by evaluating the status of the different project plans at the time of their inclusion in the data. It is possible that the positive effect for the year 2002 is in some way related to the fact that all the New Map of the Netherlands-based data relates to this year. Third, we should try to include data on soil quality in the green model. Fourth, and perhaps most importantly, in a new version of the model we should consider including the timing of a transition occurring, i.e. how long does a buyer probably have to wait before the decision is taken whether or not the expected transition will take place? Procedures to change the destination of a parcel can take multiple years, depending amongst others on whether or not stakeholders make formal protests against such a change. So explicitly including the factor time in the model is empirically relevant.

ACKNOWLEDGEMENTS

The authors would like to thank the National Institute for Public Health and the Environment (RIVM) and the Netherlands Environmental Assessment Agency (MNP), the BSIK-programme ‘Ruimte voor Geo-Informatie’ (www.rgi.nl) and Habiforum (www.habiforum.nl) for partially funding this research. Next, the authors thank the Government Service for Land and Water Management of the Netherlands (DLG), a department of the Netherlands Ministry for Agriculture, Nature Management and Fisheries (LNV), for supplying the land transaction data. The authors also thank Jan Luijt and Tom Kuhlman of the Agricultural Economic Institute (LEI) of Wageningen University and Research Centre for the discussions about land market submarkets, modelling and data issues. And finally, the authors thank Gerard Kooman, Edwin Bleijinga, Rik Heskes en Ton van Bart of the Province of Noord-Holland for the discussions about the land market in their province.

REFERENCES

Alonso, W.A. (1964) *Location and land use: toward a general theory of land rent*, Harvard University Press, Cambridge.

Anas, A. (1982) *Residential location markets and urban transportation: Economic theory, econometrics and policy analysis with discrete choice models*, Academic Press, New York

Borsboom-van Beurden, J.A.M., W.T. Boersma, A.A. Bouwman, L.E.M. Crommentuijn, J.E.C. Dekkers and E. Koomen (2005), *Ruimtelijke Beelden - Visualisatie van een veranderd Nederland in 2030*, Milieu- en Natuurplanbureau, Bilthoven.

Buurman, J.J.G., (2003), *Rural Land Markets: a spatial explanatory model*, Dissertation, Amsterdam, Vrije Universiteit.

Buurman, J.J.G., A.v.d. Brink, H.J. Scholten and J.E.C. Dekkers (2004), Ruimtelijk beleid en grondprijzen in het landelijk gebied, Tijdschrift voor Sociaalwetenschappelijk onderzoek in de Landbouw, 19 (2): 73-90.

CBS (2000) *Statline*, Centraal Bureau voor de Statistiek, www.statline.nl, last visited: June 15, 2005.

Dekkers, J.E.C. and Koomen, E. (2007), Land-use simulation for water management: application of the Land Use Scanner model in two large-scale scenario-studies, in: Koomen, E., Stillwell, J.C.H., Bakema, A., Scholten, H.J. (eds.), *Modelling land-use change; progress and applications, GeoJournal Library*, Springer, Dordrecht.

Dekkers, J.E.C., P. Rietveld, A. van den Brink and H.J. Scholten (2004), Exploring the land market in the province of Noord-Holland using a spatial regression model, *paper for the 44th ERSA-congress*, Porto, Portugal, August 25-29.

Hilferink, M. and P. Rietveld (1999), Land Use Scanner: An integrated GIS based model for long term projections of land use in urban and rural areas, *Journal of Geographical Systems*, 1 (2): pp.155-177.

Lee Jr, D.B. (1973) Requiem for Large-Scale Models, *Journal of the American Institute of Planners*, 39(3): 163-178.

Loonen, W. and Koomen, E. (2008), Calibration and validation of the Land Use Scanner allocation algorithms, *MNP rapport*, Bilthoven.

Luijt, J., J.W. Kuhlman and J. Pilkes (2003), Agrarische grondprijzen onder stedelijke druk – Stedelijke optiewaarde en agrarische gebruikswaarde afhankelijk van ligging, *NPB Werkdocument 2003/15*, Den Haag, LEI.

McFadden, D. (1978) Modelling the choice of residential location, *In Spatial Interaction Theory and Planning Models*, Karlqvist and others. North Holland Publ., Amsterdam

MNP (2004), *Milieu- en natuureffecten Nota Ruimte*, RIVM, Bilthoven.

OCW, LNV, VROM and V&W (1999) *Nota Belvedere*, Ministerie van Onderwijs, Cultuur en Wetenschap, Ministerie van Landbouw, Natuurbeheer en Visserij, Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieu and Ministerie van Verkeer en Waterstaat, SdU, Den Haag

Reilly, W.J., 1931. The Law of Retail Gravitation, New York.

Ricardo, D. (1817), *On the Principles of Political Economy and Taxation*, John Murray, London.

RNP (1958) *De ontwikkeling van het Westen des lands*. Werkcommissie Westen des Lands van de Rijksdienst voor het Nationale Plan (RNP), Staatsdrukkerij, The Hague

Wegener, M. (1994) Operational Urban Models: State of the Art, *Journal of the American Planning Association*, 60(1): 17-29