A Spatial Bayesian Hedonic Pricing Model of Farmland Values

Cotteleer G. [1], Stobbe T. [2] and Van Kooten G.C. [1,2]

[1] Agricultural Economics and Rural Policy, , Wageningen University, Wageningen, The Netherlands
[2] Department of Economics, University of Victoria, Victoria, B.C., Canada

Abstract— **In 1973, British Columbia created the Agricultural Land Reserve (ALR) to protect farmland from development. This study investigates whether the ALR has been effective near the city of Victoria. Therefore, we employ a GIS-based hedonic pricing model and quantify ALR specific measures. Bayesian Model Averaging in combination with Markov Chain Monte Carlo Model Composition are used to address specification uncertainty. Results show that zoning schemes are partly credible. Zoned farmland sells for lower prices than other farmland. However, farmland located closer to the city of Victoria is priced higher and hobby farmers pay higher prices than conventional farmers.**

Keywords— **Farmland prices, Bayesian Model Averaging, Hedonic pricing.**

I. INTRODUCTION

As cities grow and spread into the countryside, agricultural land is often the first victim of urban development. Despite programs and laws to protect agriculture, farmland prices in the rural-urban interface have increased significantly, often beyond the reach of farmers wishing to enter the sector or expand their operations. Because land prices are driven by the development and not agricultural potential of land, farming near urban areas becomes more difficult both financially and logistically.

In the current study, we examine the effect of urban encroachment on farming near Victoria, the capital of British Columbia, Canada's westernmost province. BC's agricultural land is limited, with the most productive land located near the most-rapidly growing urban centers – Vancouver, Victoria and Kelowna in the Okanagan Valley in the Interior. To protect the 1.1% of the Province considered prime farmland from development, the government created the Agricultural Land Reserve (ALR) in 1973. The ALR is a zoning ordinance that prevents agricultural land from being subdivided or used for non-agricultural purposes without permission from the Agricultural Land Commission (ALC). The ALR permits only one dwelling per parcel, which is intended to serve as a farmer's residence.

Speculation by developers and purchases of farmland for residential purposes (rural estates) are the main factors that drive up agricultural land prices near urban centers. We seek to determine empirically whether speculation in anticipation of changing land designation is happening on ALR land.

We employ a GIS-based hedonic pricing model to quantify ALR specific measures and investigate characteristics that contribute to farmland prices near the urban fringe. We also employ spatial econometric techniques that take into account spatial dependencies that are not incorporated as covariates in the hedonic pricing model. The problem with spatial econometric techniques is that they require a priori specification of a weighting matrix of spatial relations between observations, although choice of a specific relationship is arbitrary (Anselin, 1988). Another problem is that there is little in the way of theory to guide the choice of the covariates to be included in the hedonic pricing model. This means that there is both parameter uncertainty and uncertainty in the choice of the spatial weighting matrix.

Our objective is, therefore, to investigate whether the ALR has been effective in preserving farmland near Victoria, but in a way that resolves uncertainty in the application of the spatial hedonic pricing model. To address the latter issue, we apply Bayesian Model Averaging in combination with Markov Chain Monte Carlo Model Composition (MC^3) to deal with model uncertainty. The benefit of Bayesian Model Averaging is that it does not assume there is only one correct model specification; rather, final parameter estimates are weighted averages based on a whole range of possible model specifications, including different explanatory variables and different specifications of the weighting matrix. Furthermore, the MC^3

framework makes sure that model specifications with high posterior probabilities are taken into account in the weighted averages.

Although the MC^3 framework has been extended to spatial econometric models by LeSage and Parent (2007), and LeSage and Fischer (2007), the current research explicitly incorporates the selection of different specifications of the weighting matrix (based on nearest neighbors, distances and spatiotemporal patterns) in both MC^3 procedures for the spatial lag and error dependence models. To our knowledge, this extension of the MC^3 procedure constitutes an additional contribution of our research.

II. A BAYESIAN APPROACH TO HEDONIC PRICING MODEL SPECIFICATION

To investigate the impact of BC's Agricultural Land Reserve (ALR) and such things as land fragmentation on farmland prices, we specify a hedonic pricing model (see (Rosen, 1974). Given the spatial nature of the data, it is important to incorporate spatial dependence in the model. Spatial dependence can be incorporated as spatial lag or spatial error dependence. A general formulation that includes both is (Anselin, 1988):

$$P = \alpha\iota + \rho W1P + X\beta + u,$$

$$\text{with } u = \lambda W2u + \varepsilon \text{ and } \varepsilon \sim N(0, \sigma 2I), \qquad [1]$$

where P is a vector of property prices, X is a matrix of property characteristics, β is a vector of associated coefficients to be estimated, α is a constant to be estimated and ι an associated vector of ones, ε is a vector of error terms; W_1 and W_2 are spatial weighting matrices. The spatial weights are specified a priori between all pairs of observations. In our model, where each observation i corresponds to a farmland sales transaction, each element w_{ij} weights the degree of spatial dependence according to the proximity or distance between parcel i and any other parcel j; ρ is the coefficient of the spatial lag dependence structure; and λ is the coefficient in a spatial autoregressive structure for the error term. When λ=0 and $\rho\neq0$, (1) represents the Spatial Autoregressive (SAR) model. If ρ=0 and $\lambda\neq0$, we have the Spatial Error Model (SEM).

Lacking guidance regarding the choice of a weighting matrix, we specify a variety of different types: Several variations employ binary weights, two are based on distances, and two are based on spatiotemporal patterns. In the case of binary weights, an element in the weighting matrix equals one if two observations are considered to be neighbors and zero if not.

Because there is uncertainty about which weighting matrix and set of explanatory variables to use in our hedonic pricing model, we employ Bayesian techniques that allow us to specify posterior model probabilities for each specific model we wish to consider. These model probabilities tell us how likely it is that a given model is the correct one. Rather than basing parameter estimates only on the model with the highest posterior probability, we use Bayesian Model Averaging and weight the estimates of the whole range of potential models with the posterior model probabilities, which are given by (Koop, 2003):

$$p(M_i \mid y) = \frac{p(y \mid M_i)\, p(M_i)}{\sum_{m=1}^{M} p(y \mid M_m)\, p(M_m)} \qquad [2]$$

where $p(y|M_i)$ is the marginal likelihood that model M_i is the correct one and $p(M_i)$ are the prior model probabilities. If, a priori, the researcher considers each model to be equally likely, all prior model probabilities are equal to $1/M$, where M is the total number of models to be considered. In this case the posterior model probabilities are determined only by the marginal likelihoods. The marginal likelihood for model i is (Koop, 2003):

$$p(y \mid M_i) = \int p(y \mid \theta, M_i)\, p(\theta \mid M_i) d\theta, \qquad [3]$$

where $p(y|\theta,M_i)$ is the likelihood and $p(\theta|M_i)$ is the prior for the parameter vector θ. In our case, θ includes either [α, β, σ^2, λ] or [α, β, σ^2, ρ], depending on whether one considers the spatial error or lag model. The specifications of the marginal likelihoods for the spatial lag and error dependence models are provided in LeSage and Parent (2007).

To derive the posterior model probabilities, we need to consider each possible model specification. With k

potential explanatory variables and δ potential specifications of the weighting matrix, there are $2^k \times \delta$ models to consider, which is practically infeasible. (For example, with k=21 and δ=6, there are 12,582,912 models to consider.) Therefore, we use Markov Chain Monte Carlo Model Composition (Madigan, et al., 1995). The stochastic process generated by MC^3 explores regions of the model space with high posterior model probabilities. The number of iterations in the MC^3 procedure is pre-specified. At the start of the Markov chain, a regression model is chosen at random. Suppose the current model is M_i. The model that is proposed in the next step of the chain has either one variable more than the current model ('birth step'), one variable less than M_i ('death step'), or one variable of M_i replaced by a variable not currently in the model ('move step'). The proposed model M_j is then compared to the current model M_i and the probability of acceptance is given by:

$$p(\text{accept new model}) = \min\left[1, \frac{p(M_j \mid y)}{p(M_i \mid y)}\right] \qquad [4]$$

A random draw using the probability from [4] of accepting the new model and not accepting it determines whether the new model indeed replaces the old, whether M_j replaces M_i.

This procedure for proposing new models is extended by LeSage and Fischer (2007) to include uncertainty with respect to the choice of the spatial weighting matrix in the MC^3 procedure. However, only different numbers and types of nearest neighbor based weighting matrices are included in their procedure. As indicated above, we specify six different weighting matrices (two binary, two distance based, and two spatiotemporal). We extend their selection procedure by employing the MC^3 procedure that considers six different weighting matrices.

We begin the MC^3 procedure by considering a regression model with a randomly selected weighting matrix and randomly selected variables. Next we use 100,000 iterations to determine posterior model probabilities for each of the models visited during one of the 100,000 iterations. Each iteration involves the following steps:

Current model: M_{i}

Step 1: Toss a fair die with two sides 1s, two sides 2s and two sides 3s

Outcome	Decision
1.	Exclude variable from model at random
2.	Add at random a new explanatory variable not currently in model
3.	Drop current explanatory variable at random from model; replace with randomly chosen explanatory variable not now in model

Choose new model M_{j} over M_{i} with probability given by (4).

Step 2: Toss a coin

Outcome	Decision
Heads	Retain current weighting matrix (retain model M_j or M_i)
Tails	Choose new weighting matrix at random from those not currently in model (Choose new model M_{j+} over M_j or M_i with probability given by (4).

Model for next iteration: M_{m} = one of ($M_{\mathrm{j+}}$, M_{j}, M_{i}) is chosen with some probability.

Based on the MC^3 procedure, for each variable we can calculate the probabilities that this variable should be included in the model. Inclusion probabilities for variables are calculated as the number of times a variable is included in a model that was accepted divided by the total number of iterations (draws). This differs from the inclusion probabilities in LeSage and Parent (2007). They base the inclusion probabilities on the number of times a variable is included in each unique proposed model. We argue that our measure better reflects the inclusion probabilities for two reasons: Although they might be unique, proposed models can be rejected and, therefore, they do not always have high posterior model probabilities. Further, we rather base our estimate on the total number of draws, instead of the number of unique proposed models.

III. DATA AND VARIABLES

Our study area is the Saanich Peninsula of southern Vancouver Island, a rich agricultural area just north of Victoria. We use 533 observations of farmland parcels that were sold in the period 1974 (the year following creation of the ALR) to 2006. The data include all ‘single cash’ transactions but exclude sales that incorporated more than one parcel. A dummy variable (‘vacant land’) is used to distinguish between properties that do or do not have substantial structures, such as farmhouses, barns, poultry and milking facilities, etc. Only parcels were selected that could be linked to all fifteen datasets we used, so that for each observation all explanatory variables were available. Finally, if properties were sold more than once, we included only the most recent transaction in our analysis, because the structure of our weighting matrices cannot handle multiple sales of the same property.

The different data sets come from the B.C. Ministry of Agriculture and Lands, the B.C. Assessment Authority, other government agencies, and private sources. The GIS-based hedonic pricing model uses the per hectare market value of land as the dependent variable; the covariates include size of the farmland parcel, type of farm, topographical features of the land, a fragmentation index, distance to Victoria, an ALR dummy variable and the number of hectares excluded from the ALR each year.

IV. EMPIRICAL RESULTS AND DISCUSSION

The Bayesian model averaged estimates are not based on all unique models visited in each of the 100,000 iterations. Means and t-statistics for the coefficients are only calculated for the 1000 models with the highest marginal likelihoods in the spatial lag specifications and the 200 ‘best’ models in the spatial error specifications. The reason that less models are used for the spatial error specifications is that it is simply too time consuming to calculate the means and dispersion measures for more than 200 models – the combination of 200 models and 5000 draws per model took about 60 hours. For the spatial lag specifications, the combination of 1000 models and 10,000 draws per model takes about 10 hours. For the spatial lag specifications, 100,000 draws in the MC3 procedure produces 18,164 unique models. For the spatial error specifications we find 8,535 unique models in 100,000 draws.

Both the Bayes factor and the significance of the coefficient for spatial dependence indicate that SEM specifications are preferred over SAR specifications. The Bayes factor is often used to compare two model specifications assuming that prior model probabilities are the same. Therefore, we only present the results for the SEM specification. Based on the MC3 procedure, we can conclude that the spatial error structure is best described by the distance-based weighting matrices.

The specifications of the five models with the highest posterior model probabilities resulting from the MC3 procedures are provided in Table 2. In this table, ones indicate the inclusion of a certain variable or weighting matrix and zeros indicate exclusion. Posterior model probabilities for the five ‘best’ models and probabilities for the inclusion of each of the variables and spatial weighting matrices are also presented in Table 1. The Bayesian model averaged means and t-statistics for β, $\sigma 2$ and λ are provided in Table 2.

For both the spatial lag and error specifications, the models that included only the variables lot size, GDP and vacant land are preferred over larger models that include more variables. In general, smaller models with fewer covariates have higher posterior model probabilities than larger models with more covariates. This is similar to our findings (see Table 1). This partly explains why the estimated means for the coefficients are only significant for the variables lot size, vacant land (=0 if a significant structure exists on the property) and GDP. In case a variable is not included in a model, implicitly the estimated mean of the coefficient and t-statistic for that covariate will be set to zero. However, we found that coefficients of variables with low probabilities of being included can be highly significant in some of the model specifications.

Table 1: Spatial error MC3 model selection information (100,000 draws and 8535 unique models)

Variables	M1	M2	M3	M4	M5	Variable probabilities
ALR	0	0	0	0	0	0.0274
ALR boundary	0	0	0	0	0	0.0342
Distance to ALR boundary (km)	0	0	0	0	0	0.0058
ALR excluded ha	0	0	0	0	0	0.0283
Fragmentation index	0	0	0	0	0	0.0168
Grain	0	0	0	0	1	0.0910
Vegetable	0	0	0	1	0	0.0699
Tree fruit	0	0	0	0	0	0.0155
Small fruit	0	0	0	0	0	0.0410
Cows	0	0	0	0	0	0.0185
Poultry	0	0	0	0	0	0.0179
Vacant land	1	1	1	1	1	0.5029
Log of distance (km) to Victoria City Hall	0	0	0	0	0	0.0370
Log of distance (km) to Victoria airport	0	0	0	0	0	0.0047
Log of nearest distance (km) to Patricia Bay highway	0	0	0	0	0	0.0086
GDP	1	1	1	1	1	0.9999
Interest rates	0	0	1	0	0	0.0751
Maximum elevation in meters	0	0	0	0	0	0.0045
Average difference elevation level (Δ m/ha)	0	1	0	0	0	0.1027
Log of lot size (ha)	1	1	1	1	1	0.9998
Hobby farm	0	0	0	0	0	0.0222
W 5 nearest neighbors	0	0	0	0	0	0.0132
W Delaunay	0	0	0	0	0	0.0016
W distances	1	1	1	1	1	0.9852
W squared distances	0	0	0	0	0	0.0000
W distances temporal	0	0	0	0	0	0.0000
W squared distances temporal	0	0	0	0	0	0.0000
Model probabilities	0.153	0.060	0.042	0.029	0.027	

We conclude that farmland parcel sizes are important in explaining prices per ha. The log of parcel size is highly significant ($p<0.01$) and has a negative effect on the log of prices per ha. This is contrary to the expectation that farmers seek to acquire large properties to realize economies of scale because larger parcels have higher productivity levels than small ones (Cavailhes and Wavresky, 2003). There are several explanations for this result. First, average parcel size is only 3.76 ha, so the likelihood that economies of scale are an issue is small. Another reason for this unexpected result is that, when agricultural land is purchased for development purposes in expectation that it will be excluded from the ALR in the future, its value is sometimes negatively related to the size of the parcel. The reason is that the costs of subdividing land increase relative to benefits as the size of the parcel increases (Colwell and Munneke, 1999).

Finally, since ALR land cannot be subdivided without going through the Agricultural Land Commission, the negative coefficient on parcel size suggests that much of the land in the Saanich Peninsula is bought for the purpose of rural estates and hobby farms. In British Columbia, property taxes that are some 70% lower apply to land classified as ‘farm status’ than to equivalent land that is not in this category. The revenue threshold for attaining farm class status is quite low: The property must generate an annual gross income of $2500 or more at least once every two years if the farm is between 0.8 and 4.0 ha in size. For properties less than 0.8 ha, the gross income threshold is $10,000, while it is $2,500 plus 5 per cent of the property’s assessed value if the farm exceed 4 ha. As most buyers would not be farmers, an increase in property size much beyond the 0.8 ha threshold, and especially beyond 4 ha, would be viewed negatively.

Table 2: Spatial error Bayesian model averaging estimates (5000 draws, 500 burn-in draws, based on top 200 models)

Variables	Averaged coefficients	Averaged t-statistics
ALR	-0.004743	-0.084630
ALR boundary	-0.004144	-0.090991
Distance to ALR boundary (km)	-0.000674	-0.009470
ALR excluded ha	0.000141	0.040854
Fragmentation index	0.000079	0.010276
Grain	-0.021561	-0.303633
Vegetable	-0.023208	-0.282190
Tree fruit	0.000043	0.000593
Small fruit	0.010847	0.112284
Cows	0.001779	0.022456
Poultry	-0.001762	-0.018536
Vacant land	-0.193862	-2.172357
Log of distance (km) to Victoria City Hall	-0.010133	-0.106383
Log of distance (km) to Victoria airport	0.000145	0.002221
Log of nearest distance (km) to Patricia Bay highway	0.000172	0.008841
GDP	0.961483	23.534174
Interest rates	-0.026511	-0.442759
Maximum elevation (m)	0.000002	0.002452
Average difference elevation level (Δ m/ha)	0.002059	0.536199
Log of lot size (ha)	-0.560305	-21.125527
Hobby farm	0.002496	0.038247
λ	0.152495	377.060343
R-squared	0.651867	
Adjusted R-squared	0.650252	

We hypothesized that land within the ALR would be valued higher than land outside the ALR if farmland preservation is expected to be permanent. We test this hypothesis with the ALR-dummy and conclude that land located within the ALR sells at a lower price than that outside the ALR, but this result is not significant. This suggests that speculation is taking place on at least some ALR land. However, it could also be that, since farmland outside and in the ALR is increasingly used for large rural estates, there is little difference between prices as the effect of ALR zoning has been negated to a large extent.

Regarding the credibility of the ALR, we also tested whether increased exclusions of land from the ALR resulted in greater speculation. As expected, the estimated coefficient on this variable is positive, suggesting that, as more land is excluded from the ALR, land values are higher, which is suggestive of speculation. However, this effect is again not statistically significant when averaged over all models.

We also test the hypothesis that, if zoning within the ALR is credible, ALR land close to the edges of the ALR will sell for less than ALR land in the ALR interior, due to negative urban spillovers. All the indicators we use to test this hypothesis (dummy for parcels at the ALR boundary, distance to the ALR boundary and the fragmentation index) point in the same direction. All estimates coefficients support the hypothesis that the ALR boundary is credible, none of the results can be considered statistically significant. The variability with respect to these variables again indicates that the ALR boundary is only credible for a small subset of land in the ALR.

Macro-economic variables are important in the model because the data span a period of more than 30 years. Prices are expected to rise and fall jointly with macro-economic changes. For example, we find that farmland prices rise significantly ($p<0.01$) with increasing GDP. As the country's GDP increases, people are wealthier and able to spend some of their additional income on land purchases, increasing the demand for land and thus its price. Furthermore, as interest (and mortgage) rates increase, borrowing is less affordable and the demand for property declines (and property prices fall), but not significantly.

Not surprisingly, vacant land is significantly ($p<0.05$) less valuable than land that has no structures on it. While this result is partly accounted for by the fact that productive farm enterprises would require some structures, it is primarily driven by the existence of a residence on the property. A residence substantially increases the value of the land, but not by as much as might be expected. That is, farmland without a residence remains much more valuable than its use in agriculture would suggest.

V. CONCLUSIONS

In this study, we were particularly interested in determining whether B.C.'s Agricultural Land Reserve was perceived to be an effective instrument for preserving farmland. We used spatial hedonic pricing models to investigate this question. We also wished to resolve the uncertainty of the choice of explanatory variables and the spatial weighting matrix in our model. Therefore, we used Markov Chain Monte Carlo Model Composition in combination with Bayesian model averaging to resolve this model uncertainty. Although basic model uncertainty could be resolved using these methods, we found they had some drawbacks as well. First, these methods are time consuming, although greater computing power partly addresses this issue. Further, these methods seem to results in lower bounds on the estimated means and t-statistics of the coefficients of interest. However, with more specific prior information this issue might also be partly resolved.

Using these techniques, we could nonetheless draw conclusions about which variables have high and low inclusion probabilities. Lot size, GDP and vacant land were very important in explaining farmland prices. Furthermore, we learned that our data are better described by a spatial error process than a spatial lag process, and that the inverse squared distance weighting matrix best describes this spatial error process.

With respect to the credibility of the ALR, we conclude that speculation is likely an important phenomenon, affecting at least part of the ALR, even though the estimated signs all support the hypothesis that the ALR is credible. For example, ALR land is sold for less than land outside the ALR, land at the ALR boundary sells for less, and farmland that is more fragmented and farther away from the heart of the ALR sells for less. However, these findings are not very robust, as none of these estimates are statistically significant and the inclusion probabilities for these variables are all very low. Therefore, we can conclude that the ALR is only partly credible, with speculation taking place at least on some parcels. Furthermore, smaller parcels are sold for higher prices per ha than larger parcels, indicating that economies of scale in agriculture do not appear to play a role.

An alternative explanation is that the higher prices per ha signify that farmland is most likely bought for residential purposes by those craving a rural lifestyle in close proximity to a large urban area. To some extent, it is possible that the requirements for obtaining farm class status and thereby lower property taxes may, counter-intuitively, be working against agricultural preservation in BC. As smaller farmland parcels are clearly preferred by buyers, the low threshold for achieving farm tax status makes it cheaper to own a large rural estate rather than an urban residential lot. A landowner does not need to be a professional or efficient farmer, but can simply be a hobby farmer. By raising the threshold or implementing other hurdles to achieving farm status, the government could reduce the desirability of living on large rural estates, but perhaps to the detriment of serious agricultural producers.

Overall, it appears that high prices for small farm properties and inexperienced farmer-buyers bode ill for sustaining viable commercial agriculture on the urban fringe. It may also hinder preservation of open space in the longer run if such open space is being protected under the guise of preserving farmland for agricultural purposes only.

REFERENCES

1. Anselin, L. "Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity." Geographical Analysis 20, no. 1(1988): 1-17.
2. Anselin, L. Spatial Econometrics: Methods and Models. Dordrecht: Kluwer Academic Publishers, 1988.

3. Cavailhes, J., and P. Wavresky. "Urban influences on periurban farmland prices." European Review of Agricultural Economics 30, no. 3(2003): 333-357.
4. Colwell, P. F., and H. J. Munneke. "Land prices and land assembly in the CBD." Journal of Real Estate Finance and Economics 18, no. 2(1999): 163-180.
5. Koop, G. Bayesian Econometrics. First ed. Chichester: John Wiley & Sons, Ltd, 2003.
6. LeSage, J., and M. M. Fischer (2007) Spatial Growth Regressions: Model Specification, Estimation and Interpretation. Fitzwilliam College, Cambridge.
7. LeSage, J., and O. Parent. "Bayesian Model Averaging for Spatial Econometric Models." Geographical Analysis 39(2007): 241-267.
8. Madigan, D., J. York, and D. Allard. "Bayesian Graphical Models for Discrete Data." International Statisitcal Review 63, no. 2(1995): 215-232.
9. Rosen, S. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." Journal of Political Economy 82, no. 1(1974): 34-55.

Use macro [author address] to enter the address of the corresponding author:

- Author: Geerte Cotteleer
- Institute: Wageningen University
- Street: Hollandseweg 1
- City: Wageningen
- Country: The Netherlands
- Email: geerte.cotteleer@wur.nl