# Spatial analysis of factors affecting Finnish farmland prices

**Perttu Pyykkönen**
**Pellervo Economic Research Institute**
**Eerikinkatu 28**
**00180 HELSINKI**
**FINLAND**
**perttu.pyykkonen@ptt.fi**

***Paper prepared for presentation at the 99th seminar of the EAAE***
***(European Association of Agricultural Economists),***
***'The Future of Rural Europe in the Global Agri-Food System', Copenhagen, Denmark,***
***August 24-27, 2005***

# SPATIAL ANALYSIS OF FACTORS AFFECTING LAND PRICES

**Abstract:** The purpose of the study was to find out the factors affecting farmland prices in Finland. A hedonic pricing model that takes into account the presence of spatial dependence was applied for a very large sales price data (6 281 observations). The results confirmed the importance of land quality and the structural change as determinants of land price. In addition, off-farm job possibilities seemed to have a significant role in determining land prices. The effect of government payments on land prices was also clear, even though the support does not explain very much of the differences in land prices between regions.

**Keywords:** land price, hedonic pricing, spatial econometrics

**JEL Classification:** Q11

# SPATIAL ANALYSIS OF FACTORS AFFECTING LAND PRICES

## 1 Introduction

Traditionally, land prices are explained to be an outcome of the expected future incomes (land rent). In this pure net present value model (NPV) the future land rents are capitalised into land prices. In the purest form, only the agricultural earnings are taken into account (e.g. Weersink et al 1999). However, there seem to be many other factors that affect farmland prices but agricultural earnings (e.g. Just and Miranowski 1993, Bureau 1998).

Both time series and cross-sectional data have been used in the land price analysis. In time series studies the focus is on the price fluctuation and in cross-sectional studies on the price level.

Traditionally, farmland prices have varied greatly between different regions in Finland. Joining the European Union in 1995 meant an enormous change in Finnish agricultural policy (Kettunen and Niemi 1994). The previous support, which mostly consisted of price support, was reduced and area-based support measures were introduced. These support measures do not necessarily take into account differences in land quality; hence, one might expect price differences between good and poor land to have decreased. However, land price differences between regions seem to have increased in recent years.

Taking into account the rapid structural change that is going on in Finnish agriculture, the issues of how and at what price resources, especially land, are transferred to continuing farmers are very important.

The purpose of the hedonic analysis in this study is to determine the factors that affect land prices and how these factors cause regional land price differences. Unlike other land value studies, the interpretation of infrastructure effects here concentrates on the non-farm job opportunities of farm families rather than on the non-farm demand for farmland. The spatial aspects of the econometrics will also be analysed.

## 2 Hedonic (spatial) model 1995-2002

### *2.1 Model*

Typical variables used in the estimation of land prices are land characteristics such as quality, parcel size and the length of the growing period. These affect the productivity of land, and thus the expected land rent, but do not reflect the total capability of land to produce income, since agricultural returns are increasingly affected by policy decisions. Thus, support measures are often included in the models.

Moreover, agricultural productivity models are inappropriate in capturing all the factors that affect land prices. Another important factor usually included in the models is non-farm demand. Hedonic models have therefore usually been extended by including such non-farm factors as distance to the nearest urban area, population density, population growth, housing and recreational values.

Thus, the basic model applied in this study is based on the following formula.

$$LP = \alpha + \beta X + \gamma Z + \delta GP + \varepsilon \qquad (1)$$

where

LP = vector of land prices
X = matrix of agricultural characteristics variables
Z = matrix of non-farm characteristics variables
GP = matrix of support variables
$\varepsilon$ = error term.

The model applied in this study is quite similar to those of Goodwin et al. (2002; 2003) models. In addition to their models, the spatial autocorrelation will be tested.

### *2.2 Spatial econometric issues*

Because of the spatial nature of land sales, the effect of local factors may not have been taken into account accurately enough in the basic hedonic model. The assumption that the errors for two neighbouring sales (or regions) are uncorrelated is difficult to justify. If the errors are correlated, OLS does not yield efficient estimates. Moreover, if the left hand side variables themselves are correlated, the OLS will also give biased estimates. Thus, spatial analysis is worth attempting (see Anselin 1999 and 2002, Pace et al 1998).

In this study it was possible to use the exact location based on GIS coordinates instead of the adjacency approach employed in most other spatial studies concerning land prices. This means that we can create a weight matrix that includes more accurate weights than only 1 or 0. The building of the weight matrix starts with counting an n*n matrix that includes the distances from each sale to each other sale. In order to build a real weight matrix, we have to modify the matrix structure so that the nearest sales are given the largest weights. A very common practice is to take the inverse of the distance, or the inverse of the squared distance as a weight. When the distances are long (as is the case in Finland) a large proportion of weights will be very small. Thus, it is perhaps not necessary to count every weight but, for instance, take into account only the N nearest sales, or only the sales inside some precise radius. However, this is more or less an empirical matter.

When the weight matrix is built we have to specify the proper alternative of the possible spatial models. There are two basic alternatives: the spatial lag model and the spatial error model. Actually there is a third alternative (general spatial model) that combines the spatial lag and the spatial error models (LeSage 1999). However, the estimation would require the use of two different weight matrices, and it is rarely used in practice. The spatial lag model corrects the estimation in the case where the dependent variables are correlated with each other, whereas the spatial error model corrects the possible correlation in error terms.

The spatial lag model can be written:

$$LP = \alpha + \rho WLP + \beta X + \gamma Z + \delta GP + \varepsilon \tag{2}$$

where W is the spatial weight matrix.

The intuition behind this specification is that when a land sale occurs the trading partners' knowledge of other sales in the neighbourhood may affect the price. The spatial lag term is correlated with disturbances. Consequently, this means that the spatial lag term must be treated as an endogenous variable and OLS is no more an unbiased and consistent estimator. Instead, ML or some other proper estimation method that takes into account this endogeneity has to be used.

The spatial error model is a special case of regression with a non-spherical error term, in which the off-diagonal elements of the covariance matrix express the structure of spatial dependence. This being the case, the OLS results are still unbiased but no more efficient. If there is spatial correlation in the error this means that the disturbance term is actually:

$$\varepsilon = \lambda W\varepsilon + u \tag{3}$$

where the u is white noise. Modifying this a little we get:

$$\varepsilon = (I - \lambda W)^{-1} u \quad (4)$$

Incorporating this into the basic model we can write the spatial error model as follows:

$$LP = \alpha + \lambda WLP + \beta X - \lambda W\beta X + \delta Z - \lambda W\delta Z + \gamma GP - \lambda W\delta GP + \varepsilon \quad (5)$$

Thus, the model is actually a spatial lag model with an additional set of spatially lagged exogenous variables (WX, WZ, WGP) and a set of k non-linear constraints on the coefficient. The intuitive interpretation for the spatial error model is that there are some local characteristics that cannot be captured by the explanatory variables but that affect the price level.

The most commonly used specification test for spatial autocorrelation is Moran's I-test. Moran's I is similar but not equivalent to a correlation coefficient and is not centred around 0. Inference is based on a standardized z-value that follows a normal distribution. However, Moran's I is not a very strong test. In principle, it tests the spatial error, but according to Anselin and Rey (1991) it picks up a range of misspecification errors such as non-normality and heteroskedasticity of residuals, as well as spatial lag dependence. When spatial regressions are estimated by maximum likelihood (ML), the Lagrange Multiplier (LM) test can be used. This is an asymptotic test that follows a $X^2$ distribution with one degree of freedom. Both types of spatial model have their own LM test. Burridge (1980) presented the test for spatial error, and Anselin (1988) for the spatial lag. The joint use of these tests also provides the best guidance with respect to the choice of the proper alternative (Anselin and Rey 1991). If the test proposes the spatial lag model we can further test the presence of spatial error dependence using the LM test in order to decide whether the joint model of lag and error is needed (LeSage 1999).

All these tests assume the normality and homoskedasticity of the residuals. If these assumptions are violated the tests are not valid or at least one should be cautious when interpreting the results. Since diagnostic problems are very common with spatial data of this type there are some other tests that are not so sensitive to these assumptions.

Firstly, in order to be able to decide which spatial specification would be the correct one, robust versions of the LM tests have also been developed (Anselin et al. 1996). This means that the robust LM lag test is robust to spatial error and vice versa. The following testing procedure is therefore suggested for choosing the correct model: firstly perform the non-robust LM tests, and if only one of them is statistically significant choose it. However, if both are significant, continue with robust LM tests. Then choose the one that is more significant.

Secondly, Kelejian and Robinson (1992) proposed a robust test that tells whether there is spatial autocorrelation in the error. This statistic is obtained from an auxiliary regression of cross products of residuals and cross products of the explanatory variables (collected in a matrix Z with P columns). The cross products are for all pairs of observations for which a nonzero correlation is postulated (each pair only once, total number of pairs $h_N$). This test is a large sample test and follows a $X^2$ distribution with P degrees of freedom.

### *2.3 Data*

The land price data were collected from the price statistics of the National Land Survey (NLS). They are published at the regional level twice a year, but for this study the original data were obtained for each land transfer.

The original data received from the NLS consisted of 6 511 arms-length transfers of arable land from 1995 to 2002. After removing duplicates (20) and the observations with some missing data on other variables (66), the data consisted of 6 425 observations. In addition, in the preliminary hedonic regressions 144 more observations were removed, since they were deemed to be clear outliers. The criterion for detecting an observation outside the final data was that the residual was more than three times the standard deviation of the disturbance term (e.g. Osborne and Overbay 2004). The outliers were mainly sales that were either very low priced (usually less than 500 €/ha) or very high priced (more than 10 000 €/ha). Thus, the final data used in the further analysis consisted of 6 281 observations.

In addition to the sales price (PRICE), the data include information about the parcel size (SIZE), and whether a lake, river or sea borders the parcel (DWATER). In order to be able to apply the spatial econometric technique, the GIS coordinates of each land sale were also obtained, allowing the building of a spatial weight matrix based on actual distances.

A possible problem with this dataset is that it covers only a small proportion of the transfers of arable land in Finland. Firstly, a considerable proportion of the transfers take place in generation changes. Another reason for the small coverage is that transfers often include forest or buildings. Nevertheless, the data is quite large, as can be seen in Table 1.

Table 1. The number of representative sales and the amount of transferred land in different years.

| | Number of sales | Total area, ha | Average size of the sale, ha |
|---|---|---|---|
| 1995 | 666 | 4 047 | 6.08 |
| 1996 | 656 | 4 050 | 6.17 |
| 1997 | 704 | 4 707 | 6.69 |
| 1998 | 912 | 5 665 | 6.21 |
| 1999 | 873 | 5 851 | 6.70 |
| 2000 | 741 | 4 556 | 6.15 |
| 2001 | 893 | 5 727 | 6.41 |
| 2002 | 836 | 5 593 | 6.69 |

Incorporating municipality level data from different sources extended this price data. The variables that try to capture the effect of differences in land quality (in addition to DWATER) are the average yield in the region (YIELD). Since the production conditions vary considerably between regions in Finland and not all crops are suitable for production everywhere, the barley yield is chosen to represent the yield potential. Barley can be grown almost everywhere, and it is the most common grain in Finland. Unfortunately, the data was available only at the province level. The average of the years 1995-2000 was used as a proxy for this land quality measure. The third variable capturing the land quality differences was the length of the thermic growth period (number of days when the temperature is more than 5 °C) (THERMIC). The value of the nearest meteorological station (of the 63 stations) was taken to represent the variable in each municipality. The average distance to the nearest station was 31 kilometres.

The support variables were gathered from the integrated administrative control system of the Ministry of Agriculture and Forestry (IACS). Only land-based support was included in the analysis. The support was divided into four groups. The first covered crop support based on the EU's common agricultural policy (CAP). The second group was the support for less favourable areas (LFA). The third was based on the environmental scheme in which about 95% of farmers participate in Finland (ENV). The fourth group consisted of purely national land-based support based on articles 141 and 142 in the Finnish Accession Treaty (NAT). All the variables were calculated by dividing the total amount of the named support (paid in each municipality) by the total area of arable land (in each municipality) in each of the years in the research period.

The third set of variables was connected to the farm structure in the region. The first variable was the farm density (FARMDEN), which was calculated by dividing the number of farms in each municipality in each of the years by the total land area of the municipality. This variable reflects the number of potential land purchasers in the neighbourhood. The second variable tries to capture the effect of structural change the agriculture. Since the environmental scheme and other forms of support regulate the land area needed for manure in animal husbandry, a variable termed manure density (MANDEN) was calculated (see also Vukina and Wossink 2000). The calculation was based on the numbers of different animals in each municipality. The manure each animal group produces was calculated based on normative values. The amount of phosphorous ($P_2O_5$) in each of the manure groups was then calculated, again based on normative values. The amount of phosphorous was chosen as the indicator of manure density since it is the most restricting nutrient according to the environmental scheme. After this, the phosphorous amounts were summed to obtain an aggregate measure of manure density in each municipality. The third structural variable was the proportion of agricultural income in farm households (PARTTIME). The data were obtained from taxation statistics. The last structural variable was the proportion of special crops (SPECROP) in the region. This was calculated as the relative share of the potato and sugar beet area from the total area under cultivation. Potatoes and sugar beet are very intensive crops, and they have special criteria for the land for cultivation. The data source was the IACS register.

The fourth set of variables covered the non-farm factors that may affect land prices. Variables such as population density (POPDENS) are usually included in land value models to indicate the urban pressure and the non-farm demand for land. Nevertheless, as Finland is a large and generally sparsely populated country, the urban pressure as such may not be a very relevant factor. What is more, the proportion of agricultural land is indeed very small compared to other EU countries. Hence, the non-farm demand for farmland for future development cannot be very large. However, the population density, together with the unemployment factor (UNEMPL) and the proportion of agricultural labour from the total labour (AGRLAB), works as a proxy for the job opportunities and availability of services in the neighbouring area. This is a very relevant factor considering the part-time nature of Finnish family farms, since less than half of their net income is derived from agriculture. The better the non-farm job opportunities are, the more reliably the farm family can regard the future of their agricultural production. Hence, the willingness to pay more for the additional land is greater. These three variables were calculated based on municipal statistics (Statistical Centre).

## 3 Results

The dependent variable in the applied model in this study is the sales price of land (€/ha). Summary statistics and the expected signs of the independent variables are presented in Table 2.

Table 2. Descriptive statistics of the variables, and the expected signs of the explanatory variables in the estimation.

| | 1995-1999 | | 2000-2002 | | |
|---|---|---|---|---|---|
| | Mean | Std. dev. | Mean | Std. dev. | Exp. sign |
| PRICE | 3131.771 | 1721.563 | 4229.588 | 2347.046 | |
| SIZE | 6.381 | 5.704 | 6.427 | 6.313 | - |
| DWATER | 0.078 | 0.269 | 0.104 | 0.305 | + |
| YIELD | 3276.694 | 212.499 | 3284.664 | 201.935 | + |
| THERMIC | 159.645 | 8.906 | 159.647 | 8.648 | + |
| CAP | 101.290 | 42.367 | 153.103 | 48.136 | + |
| LFA | 136.668 | 50.956 | 192.765 | 19.616 | + |
| ENV | 99.768 | 32.093 | 113.451 | 7.901 | + |
| NAT | 44.760 | 15.320 | 54.672 | 21.949 | + |
| FARMDEN | 0.703 | 0.366 | 0.606 | 0.314 | + |
| MANDEN | 5.780 | 2.362 | 5.381 | 2.459 | + |
| PARTTIME | 0.384 | 0.093 | 0.406 | 0.095 | - |
| SPECROP | 0.029 | 0.056 | 0.028 | 0.056 | + |
| AGRLAB | 0.184 | 0.091 | 0.149 | 0.079 | - |
| POPDENS | 20.007 | 37.394 | 19.722 | 34.741 | + |
| UNEMPL | 0.168 | 0.047 | 0.129 | 0.042 | - |

Based on a Box-Cox analysis (see e.g. Greene 1993 or Pindyck and Rubinfeld 1991) the semi-log specification was chosen for the analysis. This analysis was performed as a maximum likelihood estimation of the total data (6.281 observations) with the whole set of independent variables described in Table 2. Thus, unlike the independent variables the sales price is transformed into a logarithmic form in further analyses.

The analyses begin with the basic OLS-estimation using the total data. In addition to the independent variables shown in Table 2, the year dummies were included into the model. This was done in order to capture a possible trend in the land prices. The results of this estimation are presented in the first column of Table 3.

In the next step, the analysis was continued by testing the effect of changing policy regime in 2000. This involved a Chow test, which allows us to decide whether the structural change in the data leads to different coefficients (Greene 1993). In this case, the data were divided into two sub datasets: the first covered the years 1995-1999 (3 811 observations) and the latter the Agenda years 2000-2002 (2 740 observations). The value of the Chow test, 6.34 (with 16, 6249 degrees of freedom), was highly significant, indicating that the coefficients differ in different policy regimes. The results of this analysis are presented in Table 3.

Table 3. Regression results of the estimated semi-log models according to the policy regime. The dependent variable is the logarithm of the sales price. *)

| | 1995-2002 | | | 1995-1999 | | | 2000-2002 | | |
|---|---|---|---|---|---|---|---|---|---|
| | Coeff. | t-value | P | Coeff. | t-value | p | Coeff. | t-value | p |
| Intercept | 4.34698 | 19.09 | *** | 4.34327 | 14.51 | *** | 3.22844 | 7.18 | *** |
| SIZE | -0.00649 | -7.16 | *** | -0.006950 | -5.74 | *** | -0.005571 | -4.11 | *** |
| DWATER | 0.085143 | 4.51 | *** | 0.081190 | 3.19 | ** | 0.090135 | 3.22 | ** |
| YIELD | 0.000332 | 8.63 | *** | 0.000310 | 5.92 | *** | 0.000275 | 4.28 | *** |
| THERMIC | 0.009176 | 8.38 | *** | 0.011520 | 7.26 | *** | 0.010990 | 6.61 | *** |
| CAP | 0.003885 | 9.50 | *** | 0.003077 | 5.66 | *** | 0.005297 | 7.86 | *** |
| LFA | 0.002392 | 6.19 | *** | 0.001536 | 3.01 | ** | 0.006073 | 7.59 | *** |
| ENV | 0.001447 | 2.76 | ** | 0.000395 | 0.53 | - | -0.000518 | -0.29 | - |
| NAT | 0.000119 | 0.26 | - | 0.001241 | 1.85 | - | 0.000810 | 1.09 | - |
| FARMDEN | 0.391603 | 14.89 | *** | 0.370316 | 11.46 | *** | 0.452809 | 9.76 | *** |
| MANDEN | 0.037746 | 10.48 | *** | 0.031550 | 6.49 | *** | 0.044050 | 7.22 | *** |
| PARTTIME | -0.11427 | -1.32 | - | -0.230009 | -2.12 | * | 0.174225 | 1.20 | - |
| SPECROP | 1.0455 | 8.71 | *** | 0.812302 | 5.23 | *** | 1.44159 | 7.43 | *** |
| AGRLAB | -0.45449 | -5.49 | *** | -0.351191 | -3.49 | *** | -0.632613 | -4.38 | *** |
| POPDENS | 0.001274 | 7.72 | *** | 0.001094 | 5.31 | *** | 0.001576 | 5.77 | *** |
| UNEMPL | -1.69213 | -10.14 | *** | -1.54876 | -7.44 | *** | -1.88321 | -6.64 | *** |
| | | | | | | | | | |
| Adj. $R^2$ | 0.505 | | | 0.449 | | | 0.526 | | |
| F | 292.01 | | *** | 164.18 | | *** | 162.12 | | *** |
| Likelihood | -3450.34 | | | -2072.90 | | | -1337.83 | | |
| Multicollinearity | 177.75 | | | 183.718 | | | 211.90 | | |
| Jarque-Bera | 63.98 | | *** | 30.51 | | *** | 28.22 | | *** |
| Koenker-Basset | 116.74 | | *** | 70.20 | | *** | 89.80 | | *** |

*) Time dummies are not reported. In the regime models all the dummies were positive, and increased in time in general (the only exception being the dummy for the year 1997). Dummies for years 1998, 1999, and 2002 were also statistically significant.

The overall fit seems to be quite good. However, there also seem to be considerable diagnostic problems. Firstly, the normality assumption of the disturbance does not hold (Jarque-Bera test).

Compared to the preliminary analyses (removing the outliers and taking the logarithm of the dependent variable) the situation has improved but still one has to be cautious when interpreting the results of other diagnostic tests. More diagnostic problems may arise when we look at the multicollinearity condition number and the heteroskedasticity test.

Based on analysis of tolerances the multicollinearity problems are mainly associated with support variables CAP and LFA. This is quite natural, since by definition CAP support is based on yield levels but at much rougher level than the variable YIELD, which is also used in the analysis. Similarly, LFA is collinear with YIELD and both of them are collinear with some infrastructure variables as well as with the thermic growth period (THERMIC).

As a consequence of the heteroskedasticity, the diagnostics in the OLS regression ($R^2$, F, t) may be misleading. In order to correct the effect of heteroskedasticity we should count the corrected variances for coefficients by, for instance, White's (1980) method. However, the heteroskedasticity tests are also sensitive to factors such as spatial dependence. Hence, no correction to the model diagnostics is made at this stage of the analysis, but we will return to this as well as to the multicollinearity problem in the further analyses.

In the next step we check if there is spatial dependence in the data. Based on the Chow test, the data is divided into two sub datasets in further analyses. In order to be able to test the spatial dependence, we have to create a weight matrix for spatial weights. In this study, we restrict the spatial dependence to extend only to a distance of 100 kilometres. In order to weight the nearest observation more, we use the inverse of the squared distance as a spatial weight. After calculating these weights they are row-standardized.

The results of the spatial dependence tests are presented in Table 4. Since both of the non-robust LM tests were highly significant in both of the periods, the robust versions of LM tests were used in the choice of the correct model specification. Thus, only the robust versions of LM tests are presented.

Table 4. Spatial dependence tests in the models.

| | 1995-1999 | probability | 2000-2002 | probability |
|---|---|---|---|---|
| Moran's I (error) | 17.55 | 0.000 | 12.55 | 0.000 |
| Kelejian-Robinson (error) | 223.01 | 0.000 | 309.87 | 0.000 |
| LM (error) | 0.23 | 0.634 | 9.70 | 0.002 |
| LM (lag) | 14.38 | 0.000 | 49.87 | 0.000 |

The tests clearly indicate that there is spatial dependence in the data in both periods. The LM tests also clearly suggest that the spatial lag specification is the correct model specification. This means that left hand side variables (i.e. land prices) are correlated with each other. This causes a severe econometric problem that must be taken into account. On the other hand, this means that the characteristics variables, even though many of them are not at farm level, seem to capture quite well the differences in local characteristics that cause price differences.

Thus, in the next step the spatial lag specification of the model is used. Now, the OLS is no longer a consistent estimation method since there is a spatially-lagged variable on the right hand side of the model (i.e. endogenous variable). Instead, we can use the maximum likelihood estimation method or the instrumental variables. The results of the spatial models based on maximum likelihood estimation are presented in Table 5.

Table 5. The results of the spatial models according to the policy regime. ML estimation. *)

| | 1995-1999 | | | 2000-2002 | | |
|---|---|---|---|---|---|---|
| | Coeff. | z-value | P | Coeff. | z-value | P |
| Intercept | 2.94219 | 10.00 | *** | 1.96264 | 4.51 | *** |
| SIZE | -0.006462 | -5.65 | *** | -0.005373 | -4.19 | *** |
| DWATER | 0.082648 | 3.45 | *** | 0.094218 | 3.56 | *** |
| YIELD | 0.000167 | 3.35 | *** | 0.000149 | 2.42 | * |
| THERMIC | 0.00630213 | 4.16 | *** | 0.006933 | 4.34 | *** |
| CAP | 0.002148 | 4.16 | *** | 0.003931 | 6.09 | *** |
| LFA | 0.001232 | 2.55 | * | 0.004334 | 5.68 | *** |
| ENV | 0.000684 | 0.98 | - | 0.000328 | 0.20 | - |
| NAT | 0.000383 | 0.61 | - | 0.000603 | 0.86 | - |
| FARMDEN | 0.25983 | 8.37 | *** | 0.310971 | 6.95 | *** |
| MANDEN | 0.022251 | 4.81 | *** | 0.030390 | 5.21 | *** |
| PARTTIME | -0.128898 | -1.26 | - | 0.198938 | 1.44 | - |
| SPEC | 0.56067 | 3.81 | *** | 1.02946 | 5.55 | *** |
| AGRLAB | -0.261276 | -2.75 | ** | -0.426853 | -3.11 | ** |
| POPDENS | 0.000908 | 4.66 | *** | 0.001341 | 5.18 | *** |
| UNEMPL | -1.17317 | -5.93 | *** | -1.30035 | -4.81 | *** |
| P | 0.362542 | 18.83 | *** | 0.344618 | 14.51 | *** |
| | | | | | | |
| Likelihood | -1923.79 | | | -1247.78 | | |
| Breusch-Pagan | 78.88 | | *** | 119.21 | | *** |
| LM test for remaining spatial error | 1.25 | | - | 1.85 | | - |

*) Instead of the Student's t distribution a standard normal distribution is used in evaluating the significance of the model coefficients. Thus, z-values are counterparts to the t-values in the OLS regression.

The likelihood ratios are clearly greater than in the analyses ignoring the spatial effect. This and the great significance of the ρ parameter confirm the need to use spatial specification. Moreover, the small values of LM tests for the remaining spatial dependence in the error term (1.25 for 1995-1999 and 1.85 for 2000-2002) indicate that the spatial lag specification properly takes the spatial dependence into account.

Since there is a possible problem with non-normality of the errors (and with the heteroskedasticity, even though the spatial dependency has been taken into account), the same analysis was estimated by using instrumental variables, which provide a more robust estimation method. The principle of instrumental variables estimation is based on the existence of a set of instruments that are strongly correlated with the original variables, but asymptotically uncorrelated with the error term. Kelejian and Robinson (1992) have shown that a series of spatially lagged exogenous variables are the proper set in spatial models. This set was also used in this study in the two-stage least-squares analysis. The results of this estimation are presented in Table 6.

Table 6. The results of the spatial models according to the policy regime. Instrument variable estimation (2SLS).

| | 1995-1999 | | | 2000-2002 | | |
|---|---|---|---|---|---|---|
| | Coeff. | z-value | p | Coeff. | z-value | P |
| Intercept | 2.852840 | 6.70 | *** | 1.161060 | 2.34 | * |
| SIZE | -0.006431 | -5.59 | *** | -0.005248 | -4.12 | *** |
| DWATER | 0.082741 | 3.45 | *** | 0.096804 | 3.68 | *** |
| YIELD | 0.000158 | 2.66 | ** | 0.000068 | 1.04 | - |
| THERMIC | 0.005969 | 3.12 | ** | 0.004364 | 2.47 | * |
| CAP | 0.002089 | 3.76 | *** | 0.003066 | 4.43 | *** |
| LFA | 0.001212 | 2.49 | * | 0.003233 | 3.89 | *** |
| ENV | 0.000703 | 1.00 | - | 0.000864 | 0.52 | - |
| NAT | 0.000328 | 0.50 | - | 0.000471 | 0.67 | - |
| FARMDEN | 0.252784 | 6.40 | *** | 0.221151 | 4.22 | *** |
| MANDEN | 0.021658 | 4.28 | *** | 0.021739 | 3.41 | *** |
| PARTTIME | -0.122449 | -1.17 | - | 0.214588 | 1.57 | - |
| SPEC | 0.544623 | 3.46 | *** | 0.768487 | 3.83 | *** |
| AGRLAB | -0.255541 | -2.63 | ** | -0.296554 | -2.08 | * |
| POPDENS | 0.000896 | 4.50 | *** | 0.001192 | 4.56 | *** |
| UNEMPL | -1.149220 | -5.36 | *** | -0.931254 | -3.19 | *** |
| P | 0.385662 | 4.68 | *** | 0.562848 | 7.98 | *** |
| | | | | | | |
| LM test for remaining spatial error | 0.032 | | - | 11.25 | | *** |

In general, the results remain very much the same as in the ML estimation. The signs and significances of the parameters do not change. However, there are some changes in the magnitude of the parameters. In the 1995-1999 subset the changes are very small. The values of the significant coefficients do not generally differ by more than 6% from each other. The significances also remain very much the same.

However, in the 2000-2002 subset the values of the significant coefficients are considerably (from 2% up to 37%) smaller. The value of the ρ parameter increases respectively. Thus, we can suspect that due to the diagnostic problems (especially heteroskedasticity) the spatial dependency is underestimated in the ML estimation. The remaining spatial dependence in the error term also suggests that the spatial lag model does not necessarily take into account all the spatial dependence. Instead, a general spatial model should, perhaps, be used. Unfortunately, the program used in the analyses allows only a rough estimation of general spatial model proposed by Kelejian and Prucha (1998). Furthermore, the application of a general model (i.e. including spatial error in the spatial lag model in question) would probably not considerably affect the levels of coefficient estimates.

In the last stage, an attempt was made to drop some of the most correlating variables. As mentioned before, the multicollinearity is mostly associated with the support variables. Dropping these variables had little effect on the results. The multicollinearity considerably decreased, but measured in terms of condition number (still over 100) there are still some problems. Dropping the time dummies further decreased the multicollinearity, but not much. However, the results of the other parameters seemed to remain quite robust. Looking at the log-likelihood values as well as information criteria (Aiken and Schwartz) also suggests a preference for the original analysis with support variables and time dummies included in the model.

Now, we turn to interpreting the results. The interpretation is mainly based on the instrumental variable estimation presented in Table 6. The negative coefficient of the sales size (SIZE) in both periods confirms the common result of lower transaction costs for larger sales sizes. However, the economic meaning of this result is quite small, since we can calculate that at the mean level a one hectare increase in the sales size would decrease the price per hectare by only about 20€. The dummy variable (DWATER) indicates that the irrigation possibility (or some recreational value) increases the land price by nearly 10% at the mean level. In the earlier period (1995-1999) the effect seems to be about 260€, and in the latter period 410€ at mean level. This may also be a reflection of the fact that land is usually of better quality by a river than in the middle of a forest.

The two variables controlling the productivity effects (yield level and length of thermic growth period) are both positive and statistically significant with the exception of yield in the latter period. The YIELD variable correlates with the support variables, and dropping the support variables makes these pure hedonic variables statistically more significant. The implicit price of the length of the growing period seems to be about 20€/ha per one day increase in the length of the growing period.

The interpretation of support variables is somewhat difficult due to the multicollinearity. Moreover, only the CAP and LFA variables are statistically significant. However, the significance in the latter period may be connected to the insignificance of the YIELD variable. Thus we should be very cautious in interpreting these results. In spite of this, elasticities of support measures (calculated at the mean value level) as well as the yield elasticity are presented in Table 7.

Table 7. Elasticities of yield and support variables with respect to sales price (calculated at the mean value level).

| | ML estimation | | IV estimation | |
|---|---|---|---|---|
| | 1995-1999 | 2000-2002 | 1995-1999 | 2000-2002 |
| Yield | 0.547 | 0.489 | 0.518 | (0.223) |
| CAP | 0.218 | 0.602 | 0.212 | 0.469 |
| LFA | 0.168 | 0.835 | 0.166 | 0.623 |

When roughly estimating the effect of a 1€ increase in support or in market returns (based on the value of the yield, i.e. 0.11€/kg), the income sources seem to have a slightly different effect on the land price in the 1995-1999 data. The discount rate for market income is 21-22%, and for CAP support around 15%. Based in the OLS regression the discount rates would be almost the same for both income sources (11% for market income, and 10% for CAP-support). However, in the 2000-2002 data the discount rates seem to be considerably lower for CAP support, from 6% to 8% depending on the estimation method, whereas the discount rates of market income seem to remain at very high level (17% or 38%). The results for LFA support seem to be quite similar compared to the CAP support.

The next set of variables (farm density, manure density, proportion of special crops, and part-time level) controls the agricultural structure. Farm density is significantly positive, indicating that the greater the number of potential buyers for the specific parcel, the higher the price. Calculated at the mean value this means that doubling the farm density would cause an increase in the land price of about 560€/ha.

The positive and significant sign for manure density reflects structural change and the increased demand for additional land due to environmental pressure. Again, calculated at the mean value, doubling the manure density would increase the land price by 400-500€/ha. The proportion of special crops (sugar beet and potato) in the region also has a very strong effect on land prices. The effect also seems to be growing. The effect of the part-time variable was expected to be negative. However, in the latter period the sign is positive, but the coefficients are not statistically significant in either of the research periods.

The fourth variable set consists of infrastructure variables. They are all significant and of the expected sign. The less important agriculture is in a region, the higher the land prices seem to be. This variable is also correlated with the part-time variable in the previous set, which may explain the non-significance of the part-time variable. An increase in the unemployment rate seems to decrease land prices, but the effect calculated at the mean level is relatively small (a one percentage point increase in the unemployment rate would only decrease the land price by 40€/ha).

## 4 Conclusions

Firstly, the analysis clearly showed that ignoring the spatial dependence may lead to incorrect results. The general interpretation concerning the signs of the coefficients as well as the significance of the results does not necessarily change very much compared to the OLS estimation, but the spatial analysis estimation is much more efficient. However, the parameter estimates may change considerably. The need to take into account the spatial dependence is, therefore, of crucial importance when spatial data are concerned.

Secondly, as expected, land quality as well as area-based support measures positively affected land prices. The support clearly affects land prices since it has a major role in creating land rent. Moreover, the very rough analysis on discount rates shows that discount rates for support variables seem to be lower than for market income, especially in the latter period from 2000-2002. However, due to the diagnostic problems one has to be very cautious in interpreting this result. Moreover, dropping the support variables decreases the power of the estimation very little, indicating that the support does not seem to explain very much of the variation in land prices between regions.

Thirdly, structural differences between regions and the structural change in agriculture seemed to have a major affect on land prices. The more farms there are in a region the more potential buyers there are, and the land price increases. In addition, investments in animal husbandry and the concentration of production seem to affect land prices. The effect comes from two sources. Growing farms need more land for their manure, and the proportion of retiring farmers may be lower. Thus, there is both increased demand and decreased supply for farmland.

Finally, infrastructure also had a very important role in determining the price level of agricultural land. If other industries are prospering in the surrounding area, agricultural viability also seems to improve. The non-farm opportunities offered to farm families make continuing and developing farming more tempting. The variables used in the analysis usually reflect the non-farm demand for farmland, but in Finnish conditions the importance of this is probably smaller than in countries where the proportion of farmland is much greater. Taking into account the part-time nature of Finnish agriculture, the explanation connected to the off-farm job possibilities is more relevant.

## References

Anselin, L. 1988. Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geographical Analysis 20:1-17.

Anselin, L. 1999. Spatial Econometrics. University of Texas. School of Social Sciences. 30 p.

Anselin, L. 2002. Under the Hood. Issues in the Specification and Interpretation of Spatial Regression Models. University of Illinois. The Regional Economics Applications Laboratory. Working paper 02-T-9.

Anselin, L. and Rey, S. 1991. Properties of tests for spatial error autocorrelation in the presence of endogenous regressors. Geographical Analysis 23:112-131.

Bureau, J.-C. 1998. Farm asset values and agricultural support. OECD. AGR/CA/APM(96)13/REV1.

Burridge, P. 1980. On the Cliff-Ord test for spatial autocorrelation. Journal of the Royal Statistical Society B 42:107-108.

Goodwin, B. K., Mishra, A. K. and Ortalo-Magné, F. N. 2002. Differentiated Policy Impacts on Agricultural Land Values. Paper presented at USDA Conference on Government Policy and Farmland Markets: Implications of the New Economy. Washington May 2002.

Goodwin, B. K., Mishra, A. K. and Ortalo-Magné, F. N. 2003. What's Wrong with Our Models of Agricultural Land Values? American Journal of Agricultural Economics 85:744-752.

Greene, W. H. 1993. Econometric analysis. Second edition. Macmillan.

Just, R. E. and Miranowski, J. A. 1993. Understanding Farmland Price Changes. American Journal of Agricultural Economics 75:156-168.

Kelejian, H. H. and Robinson, D. P. 1992. Spatial autocorrelation: a new computationally simple test with an application to per capita country police expenditures. Regional Science and Urban Economics 22:317-333.

Kettunen, L. and Niemi, J. 1994. The EU settlement of Finnish agriculture and national support. Agricutural Economic Research Institute Publication 75:1-88.

LeSage, J. P. 1999. Spatial Econometrics. Available at http://www.rri.wvu.edu/WebBook//LeSage/spatial/spatial.html.

Osborne, J. W. and Overbay, A. 2004. The power of outliers (and why researchers should always check for them). Practical Assesment, Research & Evaluation. 9(6). Retrieved from http://PAREonline.net.

Pace, R. K., Barry, R. and Sirmans, C. F. 1998. Spatial Statistics and Real Estate. Journal of Real Estate Finance and Economics 17:5-13.

Phipps, T. 2003. Discussion of Agricultural Land Values, Government Payments, and Production. American Journal of Agricultural Economics 85:770-771.

Pindyck, R. S. and Rubinfeld, D. L. 1991. Econometric Models & Economic Forecasts. Third Edition. McGraw-Hill.

Vukina, T. and Wossink, A. 2000. Environmental Policies and Agricultural Land Values: Evidence from the Dutch Nutrient Quota System. Land Economics 76:413-429.

Weersink, A., Clark, S., Turvey, C. G. and Sarker R. 1999. The Effect of Agricultural Policy on Farmland Values. Land Economics 75:425-439.

White, H. 1980. A heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48:817-838.