Investigating the Determinants of Finnish Agricultural Land Prices Using Generalised Additive Model

Valtiala, J.[1], Ovaska, S.[2] & Sipiläinen, T.[1]

[1]Department of Economics and Management, P.O. Box 27, FI – 00014
University of Helsinki, Finland

[2]Natural Resources Institute Finland (Luke), Latokartanonkaari 9, FI – 00790
Helsinki, Finland

Paper prepared for presentation at the 165. EAAE Seminar
'Agricultural Land Markets –
Recent Developments, Efficiency and Regulation'

Humboldt-Universität zu Berlin, Germany, April 4-5, 2019

Abstract

The agricultural sector in Finland has witnessed a rapid structural change since the Finnish EU-accession. At the same time, agricultural land prices have increased considerably. Using a hedonic pricing model, we investigated the characteristics affecting the prices of parcels sold. We analysed a dataset consisting of over a thousand additional agricultural land transactions and discovered several regional and production related characteristics affecting prices. The generalised additive modelling framework enabled estimation of a regional price level as a smooth trend surface. The model captured spatial dependency in the prices while retaining the sensible interpretation of the results.

Keywords: agricultural land prices, hedonic pricing model, spatial fixed effect, spatial Durbin model

JEL Code: Q15, C23

1. Introduction

Numerous researchers have surveyed the determinants of land prices in the U.S. and EU. In the mid-2000s, Huang et al. (2006) noted that the studies to that date had emphasised agricultural factors instead of non-agricultural ones. The role of non-agricultural factors has recently attracted more attention. Borchers, Ifft and Kuethe (2014) discuss the complex set of factors underlying land prices as the agricultural use value of land insufficiently explains the prices in the U.S. Both European and American studies have generally found urban proximity an important determinant (Sklenicka et al., 2013; Abelairas-Etxebarria & Astorkiza, 2012; Plantinga et al., 2002; Cavailhès & Wavresky, 2003; Delbecq, Kuethe, & Borchers, 2014; Guiling, Brorsen, & Doye, 2009). However, it is not always obvious when this proximity to a city or smaller town reflects distance to various customers and processors and when it reflects an alternative land use value, for example for housing or industrial use. Sometimes the proximity of customers has an obvious impact. Henderson and Gloy (2009) show that the proximity of ethanol plants had an impact on land prices in the U.S. great plains. Some studies have found particular environmental amenities influential (Bastian et al. 2002; Uematsu, Khanal, & Mishra 2013; Wasson et al. 2013), and some macroeconomic factors, especially interest rates, also drive prices (Hallam, Machado, & Rapsomanikis, 1992; Burns et al., 2018; Alwokuse & Duke, 2006).

Agricultural factors must be considered, although the income-generating potential of agricultural land is not always explicitly stated. Some studies find soil quality important (Maddison, 2000; Huang et al., 2006; Sklenicka et al., 2013; Nilsson & Johansson, 2013) as quality influences yields and farm income. However, studies disagree about whether farm economics influences agricultural land prices. Burns et al. (2018) found farm revenues insignificant in the U.S. and Hallam et al. (1992) in England and Wales, whereas Alwokuse and Duke (2006) as well as Devadoss and Manchu (2007) discovered a causality in U.S. land prices. However, the relationship between land prices and farm income may not be direct. Livanis et al. (2006) discovered that an increase in farm income increases U.S. land prices, but they also found that urban proximity increased both land values and farm income, overstating the impact of the income. Agricultural policy generally increases prices through subsidy capitalisation (Latruffe & Le Mouël, 2009; Ciaian, Kancs, & Espinosa, 2017). However, differences between payment types and research approaches do exist. Feichtinger and Salhofer (2011) discovered through meta-regression analysis that decoupled direct payments increased the total capitalisation rate while agri-environmental payments decreased it. On the other hand, Czyżewski, Przekota, and Pocta-Wajda (2017) and Karlsson and Nilsson (2014) found European single area payments had an insignificant or even negative effect on land prices. Furthermore, Pyykkönen (2006) found that areas with higher animal density had higher agricultural land prices, indicating a demand for additional land to spread manure on and meet the requirements of agri-environmental subsidies.

The aim of the present article is to research the determination of Finnish agricultural land prices at the parcel level. Peltola et al. (2006) noted that the hedonic price models used in their study left half of the variation in Finnish agricultural land prices unexplained. However, while they acknowledged the risk of bias in the coefficients due to spatial autocorrelation, they considered unexplained variation as random with no effect on the coefficients. As agricultural land prices vary greatly nationally and regionally across EU member states (Eurostat 2018), there is a growing need for additional domestic research with regional breakdown. Patton and McErlean (2003) emphasized the importance of the spatial effect, noting that only a few studies to that date had considered the issue. Currently, adjusting the model for dependency between observations close to each other and for spatial autocorrelation has become a standard. Most of the studies have applied different forms of spatial regression and spatial fixed effects (SFE). However, both approaches have faced justified criticism. Anselin and Arribas-Bel (2013) demonstrate that SFE does not generally remove spatial autocorrelation, noting inter alia, that the use of SFE becomes problematic when the delineation of a spatial effect is not clear. On the other hand, von Graevenitz and Panduro (2016) argue against using parametric spatial regression in the hedonic pricing model (HPM) applications. They consider the common interpretations of the spatial lags inadequate in the HPM context, showing that the spatial correlation parameter may be meaningless or even absurd when the HPM framework is strictly applied.

Trend surface models have not been subject to such criticism and remain generally unexploited in agricultural land price analysis. Barnard et al. (1997) made an early contribution concerning trend surface models. We contribute to the agricultural land literature by applying a generalised additive model. We also estimate SFE and spatial regression models for comparison. The rest of the article consists of four sections. The second section discusses the estimation framework, estimation method, and empirical model applied. It also presents the study data. The third section presents the results, and the fourth section draws the conclusions.

2. Materials and methods

2.1. Estimation framework

HPM serves as the empirical framework for the study. According to Rosen's theory (1974), a land parcel is essentially a collection of value-bearing attributes, like soil quality and proximity to the nearest urban area. In technical terms, the price of a heterogeneous product is a function of different attributes: $P_i = f(z_1, z_2, \dots z_n)$. Differentiating the function with respect to an individual attribute gives an implicit price of that attribute, $\frac{\partial f}{\partial z_n}$, i.e., its marginal effect. The hedonic pricing schedule envelopes multiple buyers' bid functions and sellers' offer functions thus determining the price of an attribute given optimal amounts of other characteristics. The framework assumes that a single seller or buyer on the market cannot affect prices. HPM has proven a viable framework for the analysis when the objective has simply been to evaluate the various characteristics of parcels. Several empirical studies researching the determinants of agricultural land prices have applied HPM, which reflects the pragmatic value of the framework.

In this study, we apply generalized additive model (GAM) to capture unobserved spatial heterogeneity in Finnish farmland prices. We also estimate two additional models applying standard methods, namely, SFE and spatial regression, for comparison. Spatial dependency arises when the value of an observation depends on the values of neighbouring or nearby observations (LeSage & Pace 2009). Ignoring spatial dependency when it actually exists causes bias and inefficiency in the estimates. GAMs capture spatial correlation in land prices by estimating the function (spline) of prices and location as a smooth surface. This is essentially a non-constant fixed effect throughout the study area. Compared to spatial fixed-effects modelling, splines provide a great deal of flexibility and capture spatial correlation better because of the data-driven estimation process. However, SFE naturally capture spatial dependency perfectly if the true data generating process is really such that spatial correlation is strictly bordered into an area and the correlation between observations is constant inside the area. In land price applications this is usually not the case. Anselin and Arribas-Bel (2013) provide a thorough discussion of the inability to SFE in remove spatial correlation. They demonstrate that in fact SFE does not generally remove spatial autocorrelation.

Being an ordinary least squares regression and therefore the simplest alternative, SFE provides a starting- point for the analysis. The present analysis applies Finnish NUTS3 regions as fixed effects. In the territorial NUTS classification of the EU, Finland has 19 NUTS3 regions, thus implying 18 parameters to estimate. In the case of spatial regression, we estimate spatial Durbin model (SDM) as shown in equation (1). Other types of spatial model do exist, but, as LeSage and Pace (2009) note, SDM controls spatial dependency in the dependent variable and in residuals. In fact, Le Sage and Pace (2009) argue that the spatial econometric literature has focused too much on comparison procedures for optimal spatial model selection and ignored SDM to a great extent. Furthermore, they prefer concentrating on capturing spatial dependency in the dependent variable than in the residuals because the former corrects estimates from bias.

$$\boldsymbol{y} = \rho W \boldsymbol{y} + X(\boldsymbol{\beta} + \alpha) + WX(-\rho\alpha) + \boldsymbol{\varepsilon} \tag{1}$$

Equation (1), vector $\boldsymbol{y}$ is a dependent variable, X is a matrix of independent variables, $\boldsymbol{\beta}$ a vector of regression coefficients, α a scalar, ρ a coefficient of spatial autocorrelation, $\boldsymbol{\varepsilon}$ an iid error, and W is the weight matrix including weighted distances between observations. SDM essentially includes spatial lags of explanatory and the dependent variables besides explanatory variables. We applied the Gaussian weighting scheme, $e^{\left(-\left(\frac{d}{50}\right)^2\right)}$, for distances d in W. This implies that the closest observations gain relatively large weight and the furthest relatively less weight. The number 50 indicates the critical distance such that within a 50 kilometre radius the spatial correlation between observations is relatively high and lower beyond that. Since we do not know the true correlation structure, this weighting scheme should be a relatively realistic and generic assumption.

GAMs include both strictly parametric and smooth data driven model components. Essentially, GAMs are generalized linear models with smooth functions of explanatory variables as additive components (Hastie & Tibshirani, 1986).

$$g\big(E(Y_i)\big) = \sum X_i \beta_i + \sum f_i(Z_i) \tag{2}$$

Equation (2) represents the basic form of GAM where $g(\cdot)$ is a link function, $E(Y_i)$ are the expected values of the dependent variable Y_i, $\sum X_i \beta_i$ parametric components with covariates X_i and the corresponding parameters β_i, and $\sum f_i(Z_i)$ are data driven components with covariates Z_i described by smooth functions $f_i(\cdot)$. Although not apparent in (2), smooth functions may be functions of multiple variables.

The estimation process applies penalised regression splines to determine the outlying dependencies. Penalising aims to balance between the smoothness and flexibility of the function. We applied thin plate splines to estimate the non-parametric components of the empirical model.

Thin plate splines do not share the weaknesses of other spline bases as they smooth any number of variables and require neither selecting knot placements nor basis functions (Wood, 2017). The property of multivariate smoothing is of great importance in this study because one of the regressor components in the empirical model is a multivariate function of latitude and longitude. Thin plate spline smoothing minimises squared differences by finding the optimal function $\hat{f}$ as shown in equation (3).

$$\|\boldsymbol{y} - \boldsymbol{f}\|^2 + \phi P_{lk}(f) \tag{3}$$

Component $\|\boldsymbol{y} - \boldsymbol{f}\|^2$ expresses squared differences with $\boldsymbol{f} = \left(f(\boldsymbol{z_1}), f(\boldsymbol{z_2}), \dots, f(\boldsymbol{z_n})\right)'$, and $\phi P_{lk}(f)$ represents the penalising component with $P_{lk} = \int \dots \int \sum_k \frac{k!}{\tau_1! \dots \tau_l} \left(\frac{\partial^k f}{\partial z_1^{\tau_1} \dots \partial z_k^{\tau_k}}\right)^2 dz_1 \dots dz_k$ and ϕ the smoothing parameter which determines the trade-off between flexibility and smoothness (Wood, 2017).

We implemented the estimation with R software (R core team, 2017) using the mgcv package (Wood, 2003; 2004; 2011; 2017; Wood, Pya, & Säfken, 2016) to estimate GAMs and the spdep package (Bivand & Piras, 2015; Bivand, Hauke, & Kossowski, 2013) and the gstat package (Pebesma, 2004; Gräler, Pebesma, & Heuvelink, 2016) for spatial regression and semivariograms. The mgcv package applies penalized iteratively re-weighted least squares to estimate GAMs and generalized cross validation by default to estimate the smoothing parameter ϕ.

$$GCV = n \frac{D(\hat{\gamma})}{\left(n - tr(\boldsymbol{A})\right)^2} \tag{4}$$

Laid out in equation (3), the generalized cross validation criterion consists of three terms. $D(\hat{\gamma})$ is the model deviance given the estimated parameters $\hat{\gamma}$, term n the number of observations, $tr(\boldsymbol{A})$ the effective degrees of freedom, essentially, the trace of the model influence matrix $\boldsymbol{A}$.

2.2. Data

The National Land Survey Finland registers statistical information on real estate transactions and provides the Official Purchase Price Register (KHR) which formed the basis of the dataset used. The dataset consists of land transactions made between 2008 and 2015 and the first half of 2016. Not all the farmland transactions during the period were included. The transactions had to cover at least two hectares of land and include no buildings and no more than 0.3 ha of other types of land. In addition, the transactions in which land was given away for free and transactions between relatives were removed from the register. With these criteria, the price should reflect a free market price for additional agricultural land well. Figure 3 presents the prices per hectare. On the shores of the Baltic Sea, there is generally more a favorable climate, better soil quality, more farms, and greater population density. Consequently, the prices are generally higher and most of the transactions have been made there. Lakes and forests dominate the central part of the country, and accordingly relatively less agricultural land and farms exist. This may explain the small number of transactions in the area.

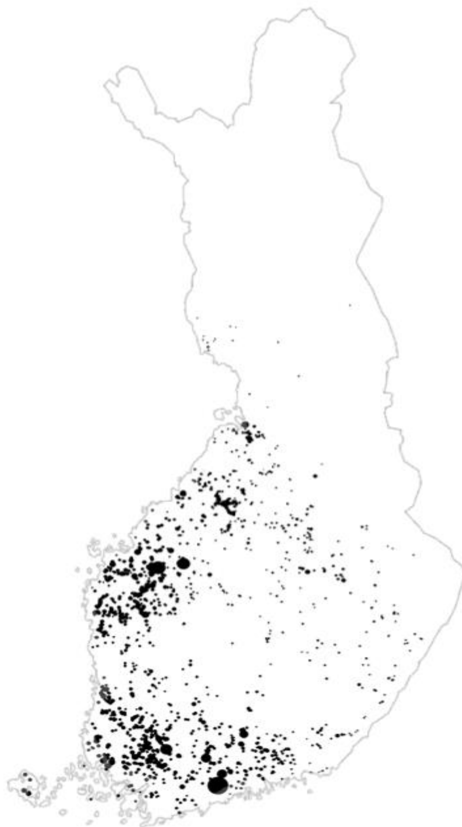


Figure 3. Agricultural land prices per hectare in Finland. A larger dot represents a higher price per hectare

Beside the total price and size of a traded parcel, the KHR includes additional information concerning each transaction. From this additional information, we used data on whether agricultural land is bordered by water, its status in a land use plan (agriculture or forestry, residential buildings, etc.), land use plan type (general plan, detailed shore plan, or no plan), buyer type (private person, government, limited company, etc.), and coordinates (ETRS-TM35FIN) for the geometric centre of a parcel. Water next to a parcel could be considered as an option for cheap irrigation or, probably less likely, as a value-bearing environmental amenity. Water next to the parcel may also cause tighter requirements for buffer strips and spraying distances included in the conditions for agri-environmental subsidies. Many studies have put a special focus on valuing irrigation water or have otherwise included irrigation in a hedonic model (Buck, Auffhammer, & Sunding, 2014; Barnard et al., 1997). The land use plan determines the planned purpose of an area, and purposes other than agriculture or forestry may be construction, conservation, or quarrying. Therefore, it reflects the value of alternative land use directly. Similarly, Czyżewski et al. (2017) included a permission-to-build variable in their hedonic model, a variable which was statistically significant.

The National Land Survey Finland's Terrain data registry (MTJ) supplements the KHR data. This includes information on distances from parcels' geometric midpoints to roads, other municipal centres, and cities. Information on an administrative agricultural land parcel is usually collected yearly in physical terms, or in the case of dividing a physical parcel between crops, on parcel terms as farmers apply for areal subsidies. Authorities also define several physical parcel characteristics such as administrative size used for areal payments and eligibility for different types of subsidy. Variables selected from MTJ were eligibility for compensation of agri-environmental actions and LFA-subsidies, the proportion of agricultural land in a municipality, and the relative amount of manure phosphorus at municipality level. Each parcel has an identification number in administrative registers. We matched identification numbers with the coordinates given in KHR and found 3,519 matches out of 5,085 transactions. KHR includes only the total area of agricultural land sold, but it does not provide information on how many parcels were sold. We selected 1,418 transactions in which the total area sold corresponded to the administrative area.

We recorded drainage type and parcel shape for 701 observations from aerial photographs, because such data was not readily available. These were considered important variables to include in the model. For non-checked parcels, it was assumed that a parcel has subsurface drainage. This assumption can be justified by noting that only 20% of the parcels in Finland have open ditches. Palmquist and Danielson (1989), for example, have examined the impact of drainage in their hedonic farmland price study. Land shape was classified into two groups according to the optimality of shape. The first class consisted of square and rectangular shapes representing the optimal shape, while the second class included all other shapes. Czyżewski et al. (2017), for example, have considered the impact of parcel shape on price. Parcels were assumed to be of the second class in non-checked parcels.

Table 1 presents the variables used in the empirical analysis.

Variable	**Description**	**Descriptive statistics**			
Continuous variables		*Min*	*Max*	*Mean*	*St.dev*
PRICE	Total price of land transaction in euros	1,700	380,800	45,082.84	41,421.93
SIZE	Size of traded parcel in hectares	2	24	4.688	2.959
YEAR	Fixed effect for transaction year	2008	2016		
LATITUDE	Northern coordinate	6659544	7369014		
LONGITUDE	Eastern coordinate	103873	689490		
Dichotomous variables		*Share of observations with value 1*			
BUYER	Dummy indicating whether buyer is public body or some other	3.1%			
PAYMENT	Dummy indicating whether parcel was not eligible for both LFA and environmental payments year before trade	3.1%			

DITCH	Dummy indicating whether there are open drainage ditches on parcel	10.3%		
WATER	Dummy indicating whether parcel is located next to water	6.9%		
SHAPE	Dummy indicating whether parcel is rectangular	17.6%		
USE	Dummy variable indicating whether parcel is located in area with more detailed land use plan	16.3%		
PURPOSE	Dummy variable indicating whether primary purpose of area in land use plan is other than agriculture or forestry	2.2%		
Categorical variables		*Number of groups*	*Smallest group*	*Largest group*
ROAD	Distance to nearest road in metres	4	0 – 500	≥ 2,000
URBAN	Distance to nearest village or urban area in metres	4	0 – 2,500	≥ 10,000
SHARE	Share of agricultural land from total area in municipality	4	0 – 9	≥ 24
PHOSPHORUS	Amount of manure phosphorus from animal farms divided by hectares of agricultural land in municipal area	4	0 – 4	≥ 9
REGION	Finnish NUTS3-level areas. Totaling 19 provinces	19		

Metric units in some variables were categorized to make the interpretation more convenient. Furthermore, the scale was truncated in the case of URBAN and ROAD up to ten and three kilometres which guided the formation of different distance groups. Grouping of observations in the case of PHOSPHORUS and SHARE was somewhat arbitrary. However, since the main purpose was simply to distinguish a possible trend between classes, a grouping criterion was not considered particularly critical. Sample quantiles determined the four groups in SHARE and PHOSPHORUS, while grouping was a progressive measure in the case of URBAN and ROAD to maintain a sensible interpretation of the classes. Of 1,418 observations, two parcels were given away for free, two lacked information on WATER, and two on REGION, so these observations were removed. Therefore, the total number of observations was 1,412.

3. Results and discussion

Since graphical inspection indicated that PRICE and SIZE variables were log-normal, we applied logarithmic transformations for these variables. The residuals in the initial model fit contained some gross outliers which may bias estimates and reduce the efficiency of the estimator. We removed twenty observations in which studentized residuals exceeded three in absolute value. The removal did increase the significance of the estimates without any significant changes in the estimates, indicating the robustness of the results. Graphical inspection shows that, after the removal, the residuals were distributed approximately normally, the pattern in the residual–fitted value plot did not indicate a severe heteroskedasticity problem, and the fitted–observed value plot indicated a fairly good model specification (Appendix 1). However, the semivariogram plot suggested slight spatial autocorrelation (Figure 4). Estimating SDM and GAM to control for spatial dependency was therefore necessary.

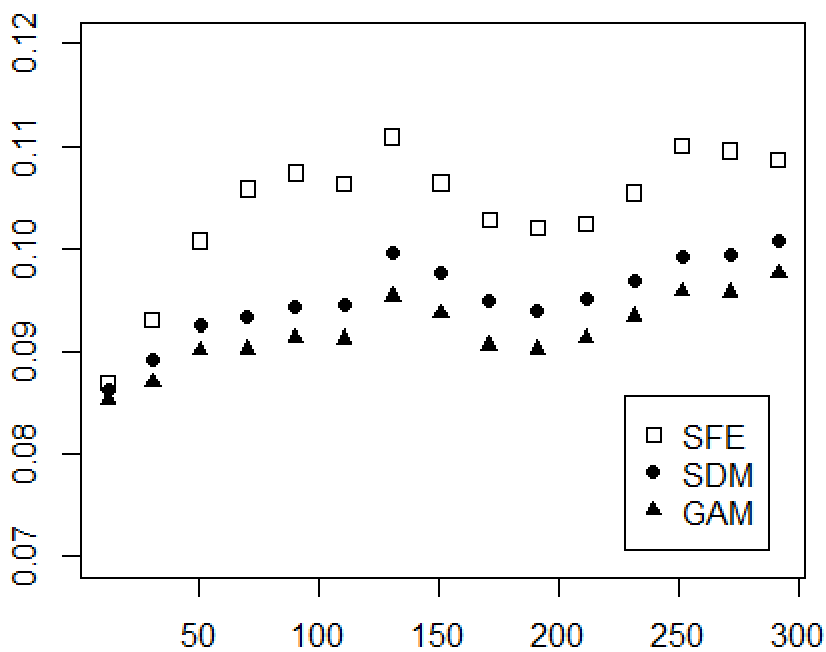


Figure 4. Semivariograms of different models. Vertical axis shows semivariance and horizontal axis shows distance in kilometres.

Appendix 2 tabulates the estimation results of the initial SFE, SFE without outliers, and SDM. The estimates in three different models were closely similar to each other without any qualitative differences. The major difference was in the ability to remove spatial autocorrelation. According to semivariogram plots (Figure 3), SDM and GAM work well in this respect. In fact, the outcome contradicts the result of Dormann et al. (2007) that spatial regression removed spatial autocorrelation while GAM did not. This contradiction becomes less surprising when considering the differences in datasets. Dormann et al. (2007) use simulated data with a known spatial correlation structure while the real data in the present study is highly complex in terms of spatial correlation and other dependencies, which are not all spatial by nature. The comparison between the two models lacks an objective measure for which model should be preferred. The overall fit is a somewhat meaningless criterion simply because a trend surface model (GAM, kriging, or other polynomial surface) could easily provide a better fit by adjusting the penalizing criterion.

The interpretation of spatial GAM estimates is fairly straightforward, which is not the case in SDM estimates. As noted by von Graevenitz and Panduro (2015), the proper interpretation of spatial lag in the HPM context would imply temporal convergence of prices and thus temporal adjustments in the hedonic pricing schedule as well. This is not plausible in the static context of the present study. Furthermore, they note that spatially lagged variables are not fixed regressors but rather outcome variables which further complicates the interpretation of the results. It is noteworthy that GAM does not assume a type of spatial dependency as SDM does, but the estimation for that part is data-driven. The general dependency is clearly visible in the trend surface spline of GAM.

Considering the discussion above, we prefer the GAM specification. Table 2 shows the results. The parametric terms of the model have the expected signs, and the variables seem to explain the price well in general. From the control variables, individual years differed significantly from the reference year 2008 thus capturing time-varying effects, inflation and policy related expectations inter alia. However, transactions made by public bodies do not seem to differ from those made by private buyers.

Table 2. Coefficient estimates and corresponding standard errors

	Coefficient estimate	Standard error
INTERCEPT	8.614	0.061 ***
SIZE	1.122	0.018 ***
$URBAN_2$	-0.103	0.041 *
$URBAN_3$	-0.106	0.037 **
$URBAN_4$	-0.138	0.038 ***
$ROAD_2$	-0.054	0.023 *
$ROAD_3$	-0.034	0.023
$ROAD_4$	-0.080	0.028 **
$SHARE_2$	0.059	0.028 *
$SHARE_3$	0.075	0.031 *
$SHARE_4$	0.204	0.038 ***
$PHOSPHORUS_2$	0.024	0.034
$PHOSPHORUS_3$	-0.012	0.036
$PHOSPHORUS_4$	0.096	0.038 *
WATER	0.051	0.034
SHAPE	0.101	0.023 ***
DITCH	-0.153	0.029 ***
PAYMENT	-0.288	0.050 ***
PURPOSE	0.772	0.078 ***
USE	0.097	0.025 ***
$YEAR_2$	0.109	0.044 *
$YEAR_3$	0.125	0.038 **
$YEAR_4$	0.172	0.038 ***

$YEAR_5$	0.254	0.039 ***
$YEAR_6$	0.280	0.038 ***
$YEAR_7$	0.285	0.037 ***
$YEAR_8$	0.308	0.038 ***
$YEAR_9$	0.300	0.043 ***
BUYER	-0.074	0.065
Adjusted R2	0.838	

SHAPE and DITCH are variables with a somewhat direct impact on the efficiency of production, and both were statistically highly significant. The significance of DITCH reflects the value of the subsurface drainage investment and the overall need for effective drainage in Finnish conditions. Palmquist and Danielson (1989) showed that drainage on wet soils causes a 34% increase in average in land value. In the present study, parcels with open drainage had a lower price by some 14%. On the other hand, parcels with more optimal shape had around an 11% price premium. Larger parcels also commanded a small premium compared to smaller parcels. The coefficient of SIZE shows that the price increases slightly more than the size. The statistical significance of SHAPE, DITCH, and SIZE supports the view that the structural change affects agricultural land prices considerably. These characteristics have higher value as growing farms aim to maximize the field labour efficiency.

The proximity of a road affects prices, but the impact becomes more significant when the closest road lies relatively far away, further than two kilometres. This result was expected as parcel accessibility in general has an undisputable impact on production efficiency. The high significance of PAYMENT indicates the capitalisation of agri-environmental and less favourable area payments. It should be noted that the vast majority of the parcels were eligible for the payments and only around 3.1% were not. Still, this relatively small group was sufficient to reveal the capitalisation effect of subsidies. The price of the non-eligible parcels was almost 25% lower than the parcels eligible for the payments. This contrasts with the finding of Feichtinger and Salhofer (2011). Although cautious with his result, Pyykkönen (2006) found LFA payment significant in 2000-02 data but not environmental payments. Thus, it is possible that only the LFA payment capitalized into land values presently, but this study cannot reach a definite conclusion on this issue. The current dataset did not allow distinguishing the LFA payment and environmental payments due to lack of deviating observations.

The results show that the price decreases as distance to the nearest urban area grows. As discussed in the introduction, urban proximity may reflect the value of land in some other use or proximity of customers. The latter is the more likely explanation in a sparsely populated country like Finland. From the other factors, a more detailed land use plan and a land use purpose other than agriculture does increase prices. Although the exceptional observations in the case of PURPOSE are few, around 2.2% from the observations, these parcels commanded significantly higher prices. The impact is indeed considerable as these parcels yielded more than twice the price, while the estimated impact for USE remains around nine percent. This was an expected outcome. Since PURPOSE reflects conversion of the land, those parcels were bought for some purpose other than farming. However, the interpretation for USE is much more complicated. A more detailed land use plan may indicate upcoming improvements in infrastructure or an option to build on a parcel in the future. In the study by Czyżewski et al. (2017), an option to build increased the total price by 10%, almost the same as the USE coefficient here.

Proximity to water does not seem to affect prices. However, since there were relatively few parcels lying close to water, and there was no further information on the type of water, whether fresh or sea water, the result should be taken with a slight reservation. The effect of SHARE is statistically significant, and the impact grows considerably as the portion of agricultural land becomes large. Because a greater amount of agricultural land in a municipal area generally implies more specialisation in agriculture in the region, there is also likely to be more demand and competition for additional land. The preconditions for farming are also usually more favourable in these regions. As in Pyykkönen (2006), our results also support the importance of animal density in land prices. The impact of PHOSPHORUS is positive and significant in the group with the highest levels of manure phosphorus which, therefore, indicates increased demand for land in areas dominated by animal production.

Figure 5 shows the estimated trend surface. Shades and contours show the logarithmic level of price at a particular point on the map. The figure covers almost whole country excluding Lapland, the northernmost region of Finland. The origin in the figure corresponds to approximately Aland, which is the most southwestern region of the country. The figure exhibits an expected pattern. The prices generally decrease towards east and north. This is due to various factors, such as soil quality and climate, which make preconditions for farming worse. In general, these factors are somewhat indistinguishable, so it would be difficult to isolate the effect of climate from, say, population density. The figure also reveals that the price level is lower in the middle of the country. This finding becomes questionable recalling that only a few transactions are located in the central part of the country (Figure 3).

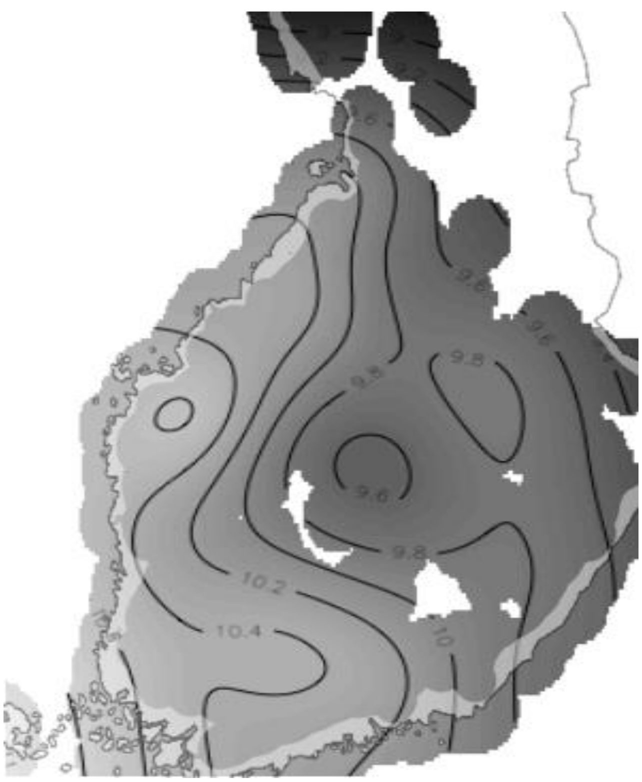


Figure 5. Graphical illustration of price as function of coordinates. Contours and shading illustrate log-level of price

4. Conclusions

We investigated factors affecting additional arable land prices in Finland using GAM. Out of three different estimation strategies, we preferred GAM because it removed spatial autocorrelation and provided results that were easy to interpret. Most of the variables under consideration in hedonic price model were statistically significant, and the large number of observations and generally well-behaved residuals increases the general reliability of the coefficient values. Open ditch drainage and non-optimal parcel shape were associated with lower prices which indicates growing willingness to improve the efficiency of farming. Furthermore, another land use purpose than agriculture and non-eligibility for (environmental compensation and LFA) subsidy payments had a considerable effect. As for the capitalisation of subsidies, the result suggests that a farmer must pay for the capitalization of subsidy rights when buying agricultural land. We also found that the prices are higher in regions with higher animal density and a greater proportion of agricultural land. This result indicates that structural change which affects the local demand for land has an important role in determining Finnish agricultural land prices.

Some dummy variables had very few deviating observations, and the dataset was incomplete in the case of shape and drainage. This may have affected the results, but it should be noted that the data were generally representative and the number of observations was large. Additionally, the incompleteness of data in cases of shape and drainage tends to underestimate the actual association. Therefore, the present results should reflect the reality.

The present study examined factors affecting the prices of individual parcels. However, we do not know much about the factors affecting the fluctuation of Finnish farmland prices over time. This is an important research question for the future. Another future direction would be to examine exceptional prices closely. The present study removed these as outliers because the interest was centred in the average case. However, extreme values could reveal some completely new information on the performance of land markets. Furthermore, detailed information about the characters of buyers and sellers could reveal new insights into the price differences not explained by parcel characteristics. Perhaps, agricultural land market is not a competitive market at all.

Acknowledgements

This study was funded by the Ministry of Agriculture and Forestry of Finland. We want to thank professor Sami Myyrä for his valuable expertise, contribution, and important role as our early phase ideas were refined to a funded research plan

References

Abelairas-Etxebarria, P., & Astorkiza, I. (2012). Farmland prices and land-use changes in periurban protected natural areas. *Land Use Policy, 29*, 674–683. doi:10.1016/j.landusepol.2011.11.003

Anselin, L., & Arribas-Bel, D. (2013). Spatial fixed effects and spatial dependence in a single cross-section. *Papers in Regional Science, 92*, 3–17. doi:10.1111/j.1435-5957.2012.00480.x

Awokuse, T. O., & Duke, J. M. (2006). The causal structure of land price determinants. *Canadian Journal of Agricultural Economics*, *54*, 227–245. doi:10.1111/j.1744-7976.2006.00047.x

Barnard, C. H., Whittaker, G., Westenbarger, D., & Ahearn, M. (1997). Evidence of capitalization of direct government payments into U.S. cropland values. *American Journal of Agricultural Economics*, *79*, 1642–1650. doi:10.2307/1244396

Bivand, R., & Piras, G. (2015). Comparing Implementations of Estimation Methods for Spatial Econometrics. *Journal of Statistical Software*, *63*(18), 1–36. doi:10.18637/jss.v063.i18

Bivand, R. S., Hauke, J., & Kossowski, T. (2013). Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods. *Geographical Analysis*, *45*, 150–179. doi:10.1111/gean.12008

Borchers, A., Ifft, J., & Kuethe, T. (2014). Linking the price of agricultural land to use values and amenities. *American Journal of Agricultural Economics*, *96*, 1307–1320. doi:10.1093/ajae/aau041

Buck, S., Auffhammer, M., & Sunding, D. (2014). Land markets and the value of water: Hedonic analysis using repeat sales of farmland. *American Journal of Agricultural Economics*, *96*, 953–969. doi:10.1093/ajae/aau013

Burns, C., Key, N., Tulman, S., Borchers, A., & Weber, J. (2018). *Farmland values, land ownership, and returns to farmland, 2000-2016* (Economic Research Report No. 245). Retrieved from U.S. Department of Agriculture website: https://www.ers.usda.gov/webdocs/publications/87524/err-245.pdf?v=0

Cavailhès, J., & Wavresky, P. (2003). Urban influences on periurban farmland prices. *European Review of Agricultural Economics*, *30*, 333–357. doi:10.1093/erae/30.3.333

Ciaian, P., Kancs, A., & Espinosa, M. (2017). The impact of the 2013 CAP reform on the decoupled payments' capitalisation into land values. *Journal of Agricultural Economics*, *69*, 306–337. doi:10.1111/1477-9552.12253

Czyżewski, B., Przekota, G., & Pocta-Wajda, A. (2017). The incidence of agricultural policy on the land market in Poland: Two-dimensional and multilevel analysis. *Land Use Policy*, *63*, 174–185. doi:10.1016/j.landusepol.2017.01.016

Delbecq, B. A., Kuethe, T. H., & Borchers, A. M. (2014). Identifying the extent of the urban fringe and its impact on agricultural land values. *Land Economics*, *90*, 587–600. doi:10.3368/le.90.4.587

Devadoss, S., & Manchu, V. (2007). A comprehensive analysis of farmland value determination: a county-level analysis. *Applied Economics*, *39*, 2323–2330. doi:10.1080/00036840600675687

Dormann, C. F., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., . . . Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, *30*, 609–628. doi:10.1111/j.2007.0906-7590.05171.x

Eurostat. (2018). *Agricultural land prices by region* [Data table]. Retrieved from http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=apri_lprc&lang=en

Feichtinger, P., & Salhofer, K. (2011). *The valuation of agricultural land and the influence of government payments* (Factor markets working paper No. 10). Retrieved from Centre for European Policy Studies website: https://www.ceps.eu/system/files/book/2011/12/Factor%20Markets%20WP%20No%2010%20on%20Agricultural%20Land%20and%20Influence%20of%20Government%20Payments%20D16.pdf

Gräler, B., Pebesma, E., & Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal*, *8*(1), 204–218. Retrieved from https://journal.r-project.org/archive/2016/RJ-2016-014/RJ-2016-014.pdf

Guiling, P., Brorsen, B. W., & Doye, D. (2009). Effect of urban proximity on agricultural land values. *Land Economics*, *85*, 252–264. doi:10.3368/le.85.2.252

Hallam, D., Machado, F., & Rapsomanikis, G. (1992). Co-integration analysis and the determinants of land prices. *Journal of agricultural economics*, *43*, 28–37. doi:10.1111/j.1477-9552.1992.tb00195.x

Hastie, T., & Tibshirani, R. (1986). Generalized additive models. *Statistical Science 1*, 297–318.doi: 10.1214/ss/1177013604

Huang, H., Miller, G. Y., Sherrick, B. J., & Gómez, M. I. (2006). Factors influencing Illinois farmland values. *American Journal of Agricultural Economics*, *88*, 458–470. doi:10.1111/j.1467-8276.2006.00871.x

Karlsson, J., & Nilsson, P. (2014). Capitalisation of single farm payment on farm price: An analysis of Swedish farm prices using farm-level data. *European Review of Agricultural Economics*, *41*, 279–300. doi:10.1093/erae/jbt021

Latruffe, L., & Le Mouël, C. (2009). Capitalization of government support in agricultural land prices: What do we know?. *Journal of Economic Surveys*, *23*, 659–691. doi:10.1111/j.1467-6419.2009.00575.x

LeSage, J., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Boca Raton, FL: Chapman & Hall/CRC.

Livianis, G., Moss, C. B., Breneman, V. E., & Nehring, R. F. (2006). Urban sprawl and farmland prices. *American Journal of Agricultural Economics*, *88*, 915–929. doi:10.1111/j.1467-8276.2006.00906.x

Maddison, D. (2000). A hedonic analysis of agricultural land prices in England and Wales. *European Review of Agricultural Economics*, *27*, 519–532. doi:10.1093/erae/27.4.519

National Land Survey of Finland. (2018). Properties consisting of cultivated land only / unbuilt, more than 2 ha [Data table]. Retrieved from https://khr.maanmittauslaitos.fi/tilastopalvelu/rest/API/kiinteistokauppojen-tilastopalvelu.html?lang=en&v=1.2.0#t41g4_x__x_Maa

Nilsson, P., & Johansson, S. (2013). Location determinants of agricultural land prices. *Jahrbuch für Regionalwissenschaft*, *33*, 1–21. doi:10.1007/s10037-012-0071-4

Palmquist, R. B., & Danielson, L. E. (1989). A hedonic study of the effects of erosion control and drainage on farmland values. *American journal of agricultural economics*, *71*, 55–62. doi:10.2307/1241774

Patton, M., & McErlean, S. (2003). Spatial effects within the agricultural land market in Northern Ireland. *Journal of Agricultural Economics*, *54*, 35–54. doi:10.1111/j.1477-9552.2003.tb00047.x

Pebesma, E. J. (2004). Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, *30*: 683–691. doi:10.1016/j.cageo.2004.03.012

Plantinga, A. J., & Miller, D. J. (2001). Agricultural land values and the value of rights to future land development. *Land Economics*, *77*, 56–67. doi:10.2307/3146980

Plantinga, A. J., Lubowski, R. N., & Stavins, R. N. (2002). The effects of potential land development on agricultural land prices. *Journal of Urban Economics*, *52*, 561–581. doi:10.1016/S0094-1190(02)00503-X

Pyykkönen, P. (2006). *Factors affecting farmland prices in Finland* (Doctoral dissertation, University of Helsinki, Helsinki, Finland). Retrieved from http://ethesis.helsinki.fi/julkaisut/maa/talou/vk/pyykkonen/factorsa.pdf

R Core Team. (2017). R (3.4.3 (2017-11-30)) [A language and environment for statistical computing]. Vienna, Austria: R Foundation for Statistical Computing.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, *82*, 34–55. doi:10.1086/260169

Sklenicka, P., Molnarova, K., Pixova, K. C., & Salek, M. E. (2013). Factors affecting farmland prices in the Czech Republic. *Land Use Policy*, *30*, 130–136. doi:10.1016/j.landusepol.2012.03.005

Uematsu, H., Khanal, A. R., & Mishra, A. K. The impact of natural amenity on farmland values: A quantile regression approach. *Land Use Policy*, *33*, 151–160. doi:10.1016/j.landusepol.2012.12.007

Von Graevenitz, K., & Panduro, T. E. (2015). An alternative to the standard spatial econometric approaches in hedonic house price models. *Land Economics*, *91*, 386–409. doi:10.3368/le.91.2.386

Wasson, J. R., McLeod, D. M., Bastian, C. T., & Rashford, B. S. (2013). The Effects of Environmental Amenities on Agricultural Land Values. *Land Economics*, *89*, 466–478. doi:10.3368/le.89.3.466

Wood, S. N. (2003). Thin-plate regression splines. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *65*, 95–114. doi:10.1111/1467-9868.00374

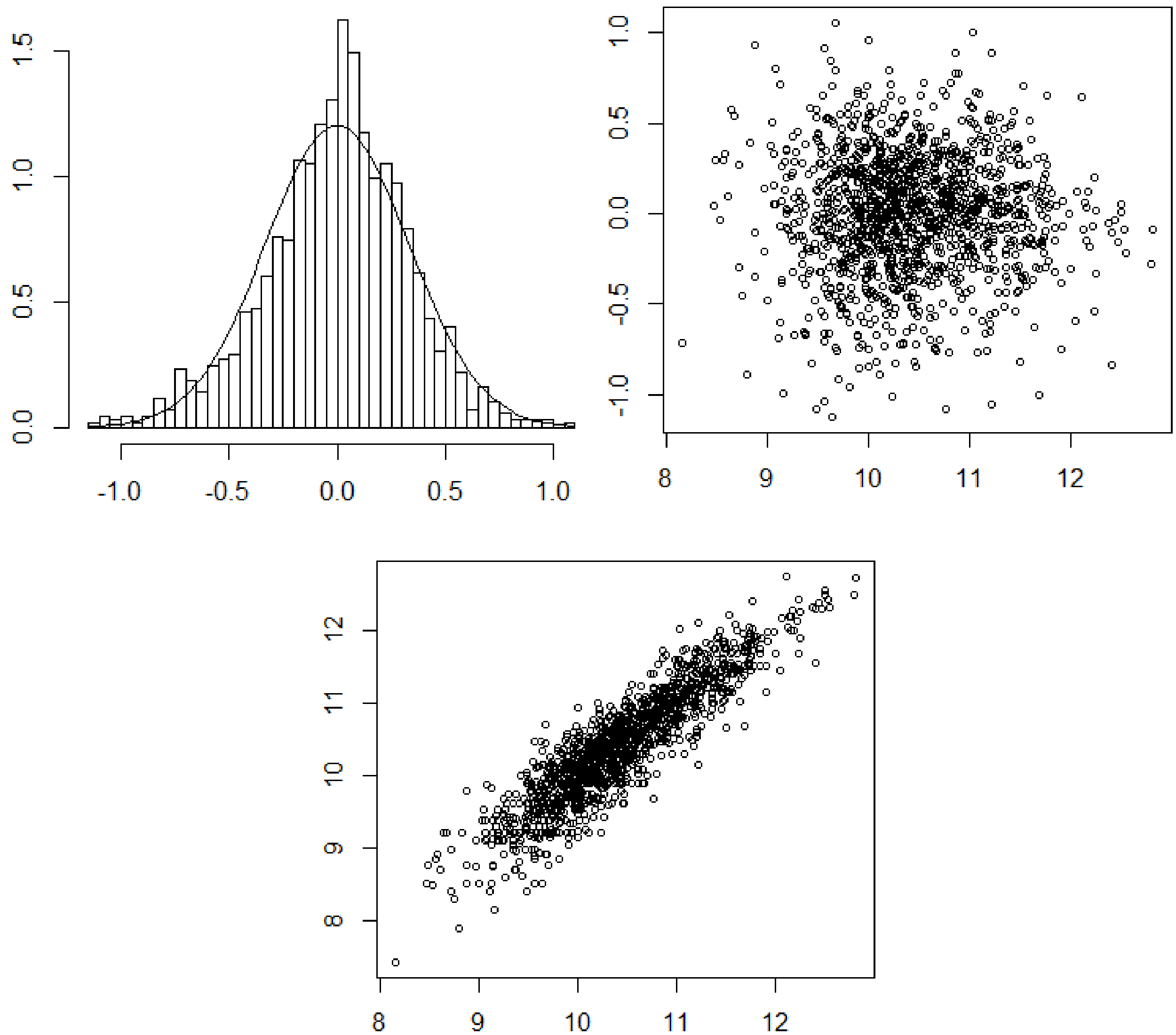
Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association*, *99*, 673–686. doi:10.1198/016214504000000980

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *73*, 3–36. doi:10.1111/j.1467-9868.2010.00749.x

Wood, S. N. (2017). *Generalized Additive Models: An Introduction with R* (2nd edition). Boca Raton, FL: Chapman and Hall/CRC.

Wood S. N., Pya, N., & Säfken, B. (2016). Smoothing parameter and model selection for general smooth models (with discussion). *Journal of the American Statistical Association*, *111*, 1548–1575. doi:10.1080/01621459.2016.1180986

Appendix A. Diagnostic plots



Top-left panel presents residual histogram with normal curve. Top-right panel shows residuals plotted against observed logarithmic values, and bottom panel presents observed values plotted against fitted values, both on log-scale.

Appendix B. Estimation results

Table B.1. SFE estimation results with original dataset

	Coefficient estimate	Standard error
INTERCEPT	8.720	0.083 ***
SIZE	1.106	0.021 ***
$URBAN_2$	-0.115	0.049 *
$URBAN_3$	-0.109	0.044 *
$URBAN_4$	-0.171	0.045 ***
$ROAD_2$	-0.072	0.027 **
$ROAD_3$	-0.054	0.027 *
$ROAD_4$	-0.142	0.033 ***
$SHARE_2$	0.061	0.033
$SHARE_3$	0.167	0.036 ***
$SHARE_4$	0.305	0.039 ***
$PHOSPHORUS_2$	0.021	0.038
$PHOSPHORUS_3$	-0.039	0.039
$PHOSPHORUS_4$	0.081	0.038 *
WATER	0.058	0.041
SHAPE	0.084	0.028 **
DITCH	-0.112	0.035 **
PAYMENT	-0.329	0.058 ***
PURPOSE	0.627	0.092 ***
USE	0.093	0.030 **
$REGION_2$	0.105	0.063
$REGION_3$	0.016	0.066
$REGION_4$	0.066	0.073
$REGION_5$	-0.053	0.073
$REGION_6$	-0.057	0.081
$REGION_7$	-0.265	0.074 ***
$REGION_8$	-0.431	0.086 ***
$REGION_9$	-0.573	0.103 ***
$REGION_{10}$	-0.531	0.076 ***
$REGION_{11}$	-0.707	0.090 ***
$REGION_{12}$	-0.515	0.093 ***
$REGION_{13}$	0.051	0.064
$REGION_{14}$	0.207	0.066 **
$REGION_{15}$	-0.143	0.098
$REGION_{16}$	-0.282	0.063 ***
$REGION_{17}$	-0.763	0.159 ***
$REGION_{18}$	-1.183	0.129 ***
$REGION_{19}$	0.679	0.203 ***
$YEAR_2$	0.125	0.053 *

$YEAR_3$	0.134	0.046 **
$YEAR_4$	0.172	0.046 ***
$YEAR_5$	0.266	0.047 ***
$YEAR_6$	0.309	0.046 ***
$YEAR_7$	0.322	0.044 ***
$YEAR_8$	0.322	0.045 ***
$YEAR_9$	0.306	0.052 ***
BUYER	0.028	0.078
Adjusted R2	0.773	

Table B.2. SFE results without outliers

	Coefficient estimate	Standard error
INTERCEPT	8.721	0.074 ***
SIZE	1.115	0.018 ***
$URBAN_2$	-0.125	0.043 **
$URBAN_3$	-0.123	0.039 **
$URBAN_4$	-0.181	0.040 ***
$ROAD_2$	-0.063	0.024 **
$ROAD_3$	-0.038	0.024
$ROAD_4$	-0.086	0.029 **
$SHARE_2$	0.051	0.029
$SHARE_3$	0.149	0.032 ***
$SHARE_4$	0.280	0.035 ***
$PHOSPHORUS_2$	0.018	0.033
$PHOSPHORUS_3$	-0.017	0.034
$PHOSPHORUS_4$	0.088	0.033 **
WATER	0.054	0.036
SHAPE	0.091	0.024 ***
DITCH	-0.165	0.031 ***
PAYMENT	-0.289	0.052 ***
PURPOSE	0.737	0.083 ***
USE	0.089	0.026 ***
$REGION_2$	0.107	0.055
$REGION_3$	0.042	0.058
$REGION_4$	0.048	0.064
$REGION_5$	-0.036	0.064
$REGION_6$	-0.076	0.072
$REGION_7$	-0.225	0.065 ***
$REGION_8$	-0.433	0.075 ***
$REGION_9$	-0.583	0.090 ***
$REGION_{10}$	-0.516	0.066 ***

$REGION_{11}$	-0.720	0.079 ***
$REGION_{12}$	-0.529	0.082 ***
$REGION_{13}$	0.061	0.056
$REGION_{14}$	0.216	0.058 ***
$REGION_{15}$	-0.087	0.087
$REGION_{16}$	-0.245	0.056 ***
$REGION_{17}$	-1.016	0.148 ***
$REGION_{18}$	-1.268	0.116 ***
$REGION_{19}$	0.680	0.177 ***
$YEAR_2$	0.102	0.047 *
$YEAR_3$	0.128	0.040 **
$YEAR_4$	0.175	0.040 ***
$YEAR_5$	0.246	0.041 ***
$YEAR_6$	0.280	0.040 ***
$YEAR_7$	0.306	0.039 ***
$YEAR_8$	0.324	0.040 ***
$YEAR_9$	0.288	0.045 ***
BUYER	-0.061	0.069
Adjusted R2	0.817	

Table B.3. SDM results

	Coefficient estimate	Standard error
INTERCEPT	4.537	1.118 ***
SIZE	1.116	0.018 ***
$URBAN_2$	-0.147	0.042 ***
$URBAN_3$	-0.127	0.038 ***
$URBAN_4$	-0.172	0.039 ***
$ROAD_2$	-0.047	0.023 *
$ROAD_3$	-0.027	0.023
$ROAD_4$	-0.080	0.028 **
$SHARE_2$	0.023	0.029
$SHARE_3$	0.055	0.032
$SHARE_4$	0.167	0.038 ***
$PHOSPHORUS_2$	0.024	0.036
$PHOSPHORUS_3$	-0.020	0.037
$PHOSPHORUS_4$	0.054	0.039
WATER	0.023	0.034
SHAPE	0.100	0.023 ***
DITCH	-0.163	0.029 ***
PAYMENT	-0.276	0.051 ***
PURPOSE	0.776	0.080 ***

USE	0.086	0.025 ***
$YEAR_2$	0.106	0.045 *
$YEAR_3$	0.110	0.039 **
$YEAR_4$	0.153	0.039 ***
$YEAR_5$	0.231	0.039 ***
$YEAR_6$	0.283	0.039 ***
$YEAR_7$	0.295	0.038 ***
$YEAR_8$	0.320	0.039 ***
$YEAR_9$	0.278	0.043 ***
BUYER	-0.067	0.066
Lag.SIZE	-0.854	0.238 ***
Lag.$URBAN_2$	-2.250	0.730 **
Lag.$URBAN_3$	-1.693	0.639 **
Lag.$URBAN_4$	-2.151	0.568 ***
Lag.$ROAD_2$	0.573	0.333
Lag.$ROAD_3$	0.449	0.276
Lag.$ROAD_4$	-0.266	0.288
Lag.$SHARE_2$	0.152	0.154
Lag.$SHARE_3$	0.143	0.162
Lag.$SHARE_4$	0.144	0.141
Lag.$PHOSPHORUS_2$	-0.589	0.174 ***
Lag.$PHOSPHORUS_3$	-0.193	0.161
Lag.$PHOSPHORUS_4$	-0.101	0.123
Lag.WATER	-1.109	0.464 *
Lag.SHAPE	0.101	0.390
Lag.DITCH	-0.790	0.310 *
Lag.PAYMENT	1.684	0.698 *
Lag.PURPOSE	-1.028	1.566
Lag.USE	-0.391	0.227
Lag.$YEAR_2$	0.095	0.845
Lag.$YEAR_3$	-0.683	0.828
Lag.$YEAR_4$	-0.903	0.804
Lag.$YEAR_5$	-0.427	0.589
Lag.$YEAR_6$	0.625	0.748
Lag.$YEAR_7$	0.621	0.657
Lag.$YEAR_8$	0.383	0.720
Lag.$YEAR_9$	-1.880	0.742 *
Lag.BUYER	0.926	1.157
Lag.PRICE (rho)	0.705	0.065 ***
Nagelkerke pseudo R2	0.836	