

This is an electronic version of the following article Kostov, P. (2009) A spatial quantile regression hedonic model of agricultural land prices, *Spatial Economic Analysis*, 4(1), 53-72. The final definite published version is available online at: <http://dx.doi.org/10.1080/17421770802625957>.

## **A spatial quantile regression hedonic model of agricultural land prices**

Philip Kostov  
University of Central Lancashire

### ***Abstract***

Land price studies typically employ hedonic analysis to identify the impact of land characteristics on price. Owing to the spatial fixity of land however, the question of possible spatial dependence in agricultural land prices arises. The presence of spatial dependence in agricultural land prices can have serious consequences for the hedonic model analysis. Ignoring spatial autocorrelation can lead to biased estimates in land price hedonic models. We propose using a flexible quantile regression based estimation of the spatial lag hedonic model allowing for varying effects of the characteristics and more importantly varying degrees of spatial autocorrelation. Applying this approach to a sample of agricultural land sales in Northern Ireland we find that the market effectively consists of two relatively separate segments. The larger of these two segments conforms to the conventional hedonic model with no spatial lag dependence, while the smaller much thinner market segment exhibit considerable spatial lag dependence.

Key words: spatial lag, quantile regression, hedonic model  
JEL codes: C13, C14, C21, Q24.

### **Introduction**

The hedonic pricing model (Rosen, 1974) is an extremely popular choice in studies of house and land prices. Hedonic theory suggests that a product price consists of the sum of expenditures on a number of bundled product attributes, each of which has its own implicit price. Rosen (1974) developed theoretical basis and estimation strategy for identification of a model of differentiated (i.e. heterogeneous) goods. This model includes assumptions about perfect competition and establishes buyers and sellers schedules that lead to market equilibrium. In simple terms the equilibrium price function is established as a double envelope curve of the bid functions (demand from individual buyers) and offer functions (supply by individual sellers). In what follows we will only focus on the first stage of Rosen's (1974) method that involves estimating the hedonic pricing function in terms of the product attributes, i.e. of the implicit prices of the product characteristics. We will not consider the issues related to the identification of the hedonic model, i.e. the use of the hedonic pricing function in estimating a demand function. There is however a great deal of uncertainty concerning the functional form of the hedonic price function (Williams, 1989).

Theoretically, the curvature of the hedonic price function could be concave, convex or linear (Freeman, 1993). It is generally accepted that the hedonic pricing function is nonlinear.

When one deals with pricing spatially fixed assets, the issue of potential spatial dependence arises. Spatial dependence induces bias and/or inefficiency in conventional estimators. Furthermore there is an intricate interplay between functional form and spatial dependence in the sense that incorrect functional form assumptions can induce spurious spatial dependence. One can potentially avoid the problem of spurious spatial dependence by non-parametrically estimating the hedonic function. Due to the relatively small datasets used in empirical modeling, or the complexity of these models (e.g. large number of variables), this may often be infeasible or undesirable. We suggest using a spatial quantile regression, which is a semi-parametric estimation method characterised by parametric rate of convergence. This allows us to alleviate the potential problem of spurious spatial dependence, at a very low cost. The paper is organized as follows. The next section discusses the problems posed by non-linearity and spatial dependence and their intricate interplay. The methodological framework is then explained and compared to some alternatives. The following two sections present the data and results from an empirical application of the proposed procedure to Northern Ireland agricultural land prices. The discussion of these results highlights some merits of the proposed methodology.

### **Nonlinearity and spatial dependence**

The nonlinearity of the hedonic pricing function has strong theoretical underpinnings. Nesheim (2002) shows that nonlinearity is a robust feature of a hedonic economy with social interactions. This is important since, as discussed later, social interaction is one of the potential sources of spatial dependence. Even in the absence of social interaction however, nonlinearity is a generic property of equilibrium in the hedonic model, as proven by theorem 1 in Ekeland *et al.* (2004). Both Nesheim (2002) and Ekeland *et al.* (2004) derive non-linearity in a fully specified hedonic model, that includes demand function. The non-linearity in this case stems essentially from the imperfect substitutability of different attributes. To explain the latter, consider the more general Lancasterian model (Lancaster, 1966) which views products as bundles of attributes and only assigns prices to these bundles, but not on the attributes

themselves. Rosen's (1974) application of the hedonic model is essentially a simplified version of the Lancasterian model that assumes separability of the attributes' bundles thus resulting in demand for attributes instead of the more general Lancasterian demand for bundles of attributes. Assuming that separation form (or substitutability of attributes between) bundles is not costless will result in a non-linear pricing function, even if the pricing function for bundles of attributes is linear. While Lancaster (1966) does make such a linearity assumption, it is purely for simplification and illustration purposes (Lancaster, 1966: 135). Therefore considering the hedonic model as a special case of the Lancasterian model explicitly leads to a non-linear representation for the pricing function. Even abstracting from this, Rosen (1974) states that linearity of the hedonic pricing function will only hold under very restrictive arbitrage conditions. Furthermore he explicitly assumed that these do not hold (Rosen, 1974: 37).

Patton and McErlean (2003) argue that ignoring the problem of spatial autocorrelation in hedonic land pricing model can yield biased estimates. The exact effect depends on the type of spatial dependence and the model definition. Spatial dependence may for example arise simply because of land valuers' independent adoptions of similar land valuation practices. If so, the spatial dependence observed in our data does not reflect a truly spatial process, but merely spatial clustering of the sources of the behavior in question. This type of spatial clustering, known as spatial error model, produces (spatial) heterogeneity in the error terms. Hence ignoring this form of spatial dependence has the same implications as the violation of the homoscedasticity assumption in regression models. The parameter estimates remain consistent, but owing to the spatial heteroscedasticity, the estimated standard errors are biased downwards and this increases the occurrence of Type 1 errors when these standard errors are used for statistical inference. In principle, one may apply a general heteroscedasticity correction, but the resulting estimates will still be less efficient than these resulting from explicitly modeling the spatial error dependence, provided the latter is known. The source of this inefficiency is the fact that estimates based on general heteroscedasticity correction will ignore available information about the spatial nature of the error correlation.

Alternatively spatial dependence may be produced by diffusion process, when spatial spillovers cause the price of a land parcel to depend on the price of neighbouring parcels. This type of spatial dependence, known as spatial lag model, has much more serious implications. These are essentially the same as omitting a significant explanatory variable. The resulting estimates are generally biased and inconsistent. Under some choices of the spatial weighting matrix however, ignoring the spatial lag dependence can still lead to consistent estimates in linear models (Lee, 2002). An example for this is when the neighbourhood structure is such that it can be decomposed into data subsets in which every observation is neighbour to every other observation within the same data subset. In such cases, similarly to the spatial error dependence, non-spatial estimation is still consistent and unbiased, though inefficient. Maddison (2004) notes that spatial lag dependence may follow from a general model misspecification. This misspecification may be for example due to omitting significant explanatory variables that are spatially correlated. Alternatively such misspecification can result from an inappropriate functional form. When the functional form is unknown, even the very fact whether there is spatial autocorrelation becomes dependent on the particular assumptions regarding the functional form. Incorrect functional form is an important source of cross-sectional autocorrelation in errors (see Greene, 2003:192). Therefore the choice of the wrong functional form may spuriously induce spatial autocorrelation (which is a form of cross-sectional autocorrelation). Basile and Gress (2005) demonstrate this trade-off in an empirical application. They estimate semiparametric models in which they model non-parametrically the regression part retaining a fixed spatial lag parameter. They find that the spatial lag parameter estimate in their semi-parametric models is considerably smaller than in the parametric specifications. Thus incorrect functional form assumptions can 'increase' spatial dependence. Furthermore even if the spatial autocorrelation is not present in the 'true' model, incorrect functional form assumptions can 'create' it.

McMillen (2003: 208–209) notes that “tests for spatial autocorrelation also detect functional form misspecification, heteroskedasticity, and the effects of missing variables that are correlated over space”. The causes of spatial autocorrelation in the hedonic model are therefore rather complex. If it is caused by missing variables or functional form misspecification, it is a statistical problem and it does not affect the

validity of the hedonic model assumptions. We will not discuss the impact of missing variables here. Although including spatial lag can to some extent approximate the effect of such missing variables, explicitly incorporating them leads to considerable efficiency gains. For example Anselin and Le Gallo (2006) use kriging to interpolate air quality and include in house price model. Their results show dramatic improvement over the reference model in which this variable is missing (and is therefore approximated by the spatial dependence). In what follows we will implicitly assume away the potential problem of missing variables.

Without entering into details about what are the many possible causes for the spatial error model, we note that in principle spatial clustering is not inconsistent with the assumptions of the hedonic model. A spatial lag model on the other hand will be inconsistent with these assumptions. For example spatial diffusion process could be generated by interactions between agents, in the form of social norms, neighbourhood effects, and peer group effects. Such processes would involve some form of social interaction that is in general at odds with the assumptions of perfectly competitive markets. Therefore if one takes every effort to ensure the problems of missing variables and incorrect functional specifications are avoided, the spatial lag dependence in the hedonic model could indicate deviations from the assumption of perfectly competitive markets.

If the hedonic model is estimated non-parametrically, then we exclude the possibility of incorrect functional form contributing to finding spatial autocorrelation. Note however that due to the slower rate of convergence of nonparametric estimators, this requires larger datasets. When the size of the dataset or the complexity of the model make such an approach infeasible, we suggest using a quantile regression specification. The quantile regression is essentially a semi-parametric model, which means that although it cannot fully guarantee to eliminate, it will at least alleviate the potential impacts of the functional form assumption on spatial dependence. Furthermore since the conventional linear programming type of quantile regression estimator (Koenker and Bassett, 1978) is characterised by the parametric rate of convergence (the inverse of the square root of the sample size), this approach can be applied to relatively small datasets.

## Methodology

The linear spatial lag model has the following form:

$$y = \lambda W y + X \beta + u \quad (1)$$

where  $\lambda$  is the spatial lag parameter,  $W$  is a known spatial weights matrix,  $W y$  is the spatially lagged dependent variable,  $X$  is a matrix containing in its rows the values of explanatory variables,  $\beta$  is a vector of unknown regression parameters, and  $u$  is a vector of random disturbance terms.

Since the spatially lagged dependent variable is present on the right hand side of (1), the ordinary least squares (OLS) estimator is usually inconsistent. There are two main types of estimators for the spatial lag model that have been extensively studied and used in the literature. These are the maximum likelihood or quasi maximum likelihood estimator (see e.g. Anselin, 1988) and the generalized method of moment estimator (see Kelejian and Prucha (1998, 1999) and Lee (2003, 2007)). Both these estimators employ the assumption that the disturbances  $u$  are independent and identically distributed (i.i.d.). The latter assumption may however be too restrictive. Social interactions for example may cause the variance of the aggregated level data be inflated (Lin and Lee, 2006). Furthermore Kelejian and Prucha (2006) argue that spatial units are often heterogeneous in important characteristics such as size. As a result, the above estimators are inconsistent. The general method of moments method have been extended to allow for heteroscedasticity (Lin and Lee, 2006 and Kelejian and Prucha, 2006).

Quantile regression is an important method for modelling heterogenous effects of variables on a response and at the same time taking into account unobserved heterogeneity and allowing for heteroscedasticity among the disturbances (Koenker, 2005). Note that the latter can essentially include any forms of spatial error dependence. The quantile regression generalisation of the (linear) spatial lag model could be written as:

$$y = \lambda(\tau) W y + X \beta(\tau) + u \quad (2)$$

In contrast to model (1) the spatial lag parameter  $\lambda(\tau)$  and the vector of regression parameters  $\beta(\tau)$  are  $\tau$ -dependent, where  $\tau$  is the corresponding quantile of the dependent variable. Note that the resulting model is essentially a nonlinear one. Since this specification allows the spatial parameter  $\lambda(\tau)$  to be dependent on  $\tau$ , allows for a different degree of spatial dependence at different points of the response distribution. For example we could have spatial lag dependence only present in some parts of the distribution of the dependent variable, but not in other. The linear quantile regression can usually be considered an approximation to a more general non-linear model. Thus it could alleviate the problems arising from potential functional misspecification. Furthermore, the quantile estimators are robust and much less sensitive to outliers. Note that the quantile regression model does not make any distributional assumptions concerning the residuals  $u$ . One may impose some restrictions on the residuals, e.g. assuming they are i.i.d., but such assumptions will only affect the variance-covariance matrix of the estimates, not the estimates themselves.

Since the spatial lagged variable is present on the right hand side of (2) then, similarly to the mean case, the conventional quantile regression estimator of Koenker and Bassett (1978) will in general be inconsistent. Accounting for endogeneity via instrumental quantile regression estimation however should circumvent this problem. Such an approach, using spatially lagged independent variables, could be viewed as direct quantile regression generalisation of two stage least squares (2SLS) estimator of Kelejian and Prucha (1998).

Amemiya (1982) and Powell (1983) discuss such a two stage approach applicable to the median regression model, which is a type of quantile regression with  $\tau=0.5$ . Kim and Muller (2004) extend this approach to the general quantile regression setting. Zietz *et al.* (2008) apply the above method in a spatial quantile regression model of house prices, using the spatially lagged independent variables as instruments. Other alternative estimators for quantile regression with endogenous regressors have been suggested in Chen and Portnoy (1996), Lee (2004) and Ma and Koenker (2006). Abadie *et al.* (2002) proposed an estimator, which solves a convex programming problem with first step nonparametric estimation of a nuisance function and is therefore computationally more demanding. Here we will use the instrumental



variable quantile regression (IVQR) method of Chernozhukov and Hansen (2006). Below we briefly describe it.

In principle this method assumes i.i.d. data, and as discussed above such an assumption could be questionable for spatial data. Recently Su and Yang (2007) have extended this method to allow instrumental variables quantile regression estimation of the spatial lag model. It can be defined as follows. Let us first assume that the quantile of interest  $\tau$  is given. Then:

1. For any value of  $\lambda$  in a pre-determined set of values run an ordinary quantile regression (using the given quantile  $\tau$ ) of the spatially filtered dependent variable  $y - \lambda W y$  on the dependent variables (i.e.  $X$ ) and instruments (i.e.  $WX$  in this case).
2. Minimise the norm of the vector of coefficient estimates for the instrument in the above quantile regression to obtain the IVQR estimator for  $\lambda$ . The intuition behind this step is provided by Chernozhukov and Hansen (2006) and can be explained as follows. If we have a single instrumental variable, the value of  $\lambda$  in the previous step that yields an estimate of zero for the coefficient of this single instrumental variable in the first step is the appropriate value for  $\lambda$ . When there is more than one instrument, the minimisation of the norm of the vector of their parameters estimates provides a generalisation of the above logic. See Chernozhukov and Hansen (2006) and Su and Yang (2007) for more details.
3. For the optimal value  $\hat{\lambda}$  from the previous step, run an ordinary quantile regression of the spatially filtered dependent variable  $y - \hat{\lambda} W y$  on the dependent variables and instruments to obtain the IVQR estimates for the parameters of the dependent variables.

This procedure is defined as above for a given quantile, but it could be re-run over any other quantile. Note also that it is very simple to program, as long as a linear quantile regression routine is available. One simply needs to define a sufficiently detailed grid of values for  $\lambda$ , run the corresponding quantile regression problems over this grid, minimise the norm in step 2 (i.e. calculate it and find the value of  $\lambda$  that minimises it), and then take the corresponding estimates for the dependent variables from the quantile regression for the optimal value of  $\lambda$  in step 1.



When computing the norm of the vector of coefficient estimates in step two, it is normalised by some positive definite matrix. In the just identified case considered in Chernozhukov and Hansen (2006) and Su and Yang (2007), the choice of such positive definite matrix does not matter, e.g. one can just set it to an identity matrix. There are some asymptotic efficiency considerations that suggest using the inverse of the variance–covariance matrix of the parameters estimated in step 1. In this case the resulting norm to be minimized in step 2 can be interpreted as a Wald test statistic testing whether the coefficients for the instruments are jointly zero. When the asymptotic confidence intervals are replaced by confidence intervals based on inverting the above mentioned Wald statistic, this results in finite sample inferential procedure that is robust to weak or partial identification (Chernozhukov *et al.*, 2007; Chernozhukov and Hansen, 2008).

### **Comparison with other methods**

First we will compare the used methodology with the alternative quantile regression estimator of Kim and Muller (2004), which effectively means to the implementation of Zietz *et al.* (2008). Kim and Muller’s (2004) two–step estimator is computationally simpler. It only requires two consecutive quantile regressions. The method employed here is essentially single step estimation, but carries out a search over a set of values for the spatial lag parameter and thus requires a separate quantile regression to be estimated for each value in this range. Thus from a practical implementation point of view the two-step method of Kim and Muller (2004) seems preferable. Note however that these two methods have a totally different approach to controlling for endogeneity. Using the terminology of Blundell and Powell (2003), the two stage quantile regression uses the so called ‘fitted values’ approach, replacing the dependent variable and the endogenous spatially lagged dependent variable by the fitted values from the first stage. The method employed here, on the other hand can be viewed as a generalised method of moments approach, which in the same terminology is an

‘instrumental variables’ approach. The orthogonality conditions for the instruments are actively imposed by the minimisation in step 2. Note that while in the two stage approach these conditions are assumed and need to hold, here we essentially try to impose them. In principle step 2 tries to obtain estimates that closely conform to these moment conditions. In the case of weak identification for example, this may not be possible. Then the indirect approach of Chernozhukov and Hansen (2008) based on inverting the Wald test statistic still ensures a valid finite sample inference. Irrespective of the strength of the chosen instruments the availability of finite sample inference (alongside the asymptotic one) is another advantage when one is dealing with relatively small datasets. The other alternative approaches to controlling endogeneity in quantile regression models, namely the ‘fitted values’ approach (as in Kim and Muller, 2004) and the ‘control function’ approach (see Lee, 2004) only rely on asymptotic results. The availability of finite sample inference is a major advantage when dealing with relatively small datasets. One of the motivations for proposing the spatial lag quantile regression model is the possibility for varying (across the sample) degree of spatial dependence. In this case it would not be unreasonable to expect that the degree of identification of the spatial dependence via the instrumental variable will also be varying. It is therefore important to have a method that accounts for the possibility of (locally) weak instruments.

Alternatively, the proposed approach could be compared to more general semi-parametric additive models. Additive modelling is an attractive option for hedonic models since it allows for an implicit price interpretations of the results. In such an approach the coefficients will be typically assumed to vary relatively smoothly in some sense with the values of some underlying variables. In general, this variation could be defined with regard to the variables in question (i.e. the effect of each

additive component varying with its own values), or they may vary with the values of some other variables, as in functional coefficients models. A particular type of such model is the geographically weighted regression (Fotheringham *et al.*, 1997, 2002) in which all regression coefficients are varying across space. In general some theoretical considerations can suggest what the effect modifier should be. For example the market segmentation work in housing (Goodman and Thibodeau, 2003) suggests that markets can become segmented according to the age of the properties. This naturally leads to a varying coefficient representation. One could however argue that often the actual effect modifier may be unobserved or unknown. In such cases a spatially varying relationship can essentially approximate such effects. This line of reasoning provides some justification for using the geographically weighted regression framework. The varying coefficients framework is obviously quite versatile, since one can incorporate different forms of spatial and non-spatial variation in the same model. All these methods however require larger datasets because the corresponding estimators converge at a non-parametric rate, and therefore are not applicable to relatively small datasets. In a typical hedonic pricing model, it would not be unreasonable to expect the effects of the hedonic attributes to be ordered in the sense that they could be expected to be either stable or mostly increasing (decreasing) in price. If this is the case the linear quantile regression will provide reasonable approximation to the non-linear relationships. In quantile regression we represent the effects of the hedonic attributes with regard to the price. If these effects are mostly increasing (or decreasing) in price, simply plotting them should, in many cases, provide useful guide to whether one can suspect market segmentation. The quantile regression framework is however more robust than the standard varying coefficients model, which is essentially based on some form of smoothing. In addition to the

robustness to outliers in the direction of the dependent variable and lack of distributional assumptions, the quantile regression method does not do smoothing in the conventional sense. Although the results would be ordered with regard to the quantiles of the dependent variable, there is no requirement that these effects are smooth in any sense. Thus the quantile regression model should be able to consistently estimate different effects without the need to pre-specify how the coefficients need to vary. For example the geographically weighted regression assumes that all coefficients vary in space. When this is not the case the model will be misspecified and the estimated coefficients will in general be biased. When the linear quantile restriction is inappropriate, the standard quantile regression estimators will similarly be biased. In this case different alternative non-parametric quantile methods are available. These do essentially employ some form of smoothing within the linear quantile estimator. Note however that the resulting estimators still share all the advantages of the conventional linear quantile regression estimator. Additionally we can view some of these estimators as weighed versions of the conventional linear quantile estimator. This is particularly easy to see for a locally linear or locally polynomial quantile regression, where the smoothing required to determine the local estimator produces implicit weights with regard to the conventional one. Similarly for spline based quantile regression which is essentially linearised with regard to the spline basis evaluations model, the evaluation of the spline basis functions provides implicit weights with regard to the original variables. Therefore extending the present estimator to non-parametric setting is relatively straightforward. The inferential apparatus for non-parametric endogenous quantile regression however will require some attention and up to our knowledge no one has so far proposed reliable inference procedures in this setting.

## **Data**

We use a dataset consisting of 197 agricultural land sales in Northern Ireland. This data is analysed in Patton and McErlean (2003, 2004). The data were collected by a mail survey from buyers of agricultural land. The names and addresses of buyers of agricultural land were obtained from the Valuation and Lands Agency. The survey design followed Dilman's 'Total design Method' and targeted the entire population of agricultural land transactions throughout Northern Ireland between September 1996 and June 1999. Such a dataset could be considered relatively small to allow an efficient non-parametric estimation, but it is nevertheless large enough to apply the proposed methodology.

Non-land items in each transaction were excluded from the total sales price to yield a 'pure' land price. Additionally transactions between family members, as well purchases for non-agricultural purpose were excluded in order to make the dataset as close to the assumptions of the pure hedonic model as possible. Price per acre was deflated using a retail price index because of the time span over which it was collected. The variables used are the same as in Patton and McErlean (2004). These are listed below.

Acreage is measured in number of acres and represents the size of the land plot. Land quality score is measured using values close to 1 to represent good quality land and values close to 7 to represent poor quality land. The land quality score is the variable most closely representing the productive capacity of agricultural land. It is however not possible to perfectly capture the land quality in a single variable. Another land quality proxy used in the study is the drainage score. It measures the drainage

property of the land and is measured on a scale from 1 to 10. Similarly to the land quality score smaller values imply higher drainage capability. Most of the Northern Ireland agricultural land is actually grazing land. Owing to the significant amount of rainfall in the region, poor drainage would mean the corresponding land plots would be unusable for livestock grazing for prolonged periods of time and hence will be less productive. Similarly the drainage score would affect the productivity of the arable land. Therefore for Northern Ireland lower drainage score implies better land quality. The other land quality variable is the number of dairy cows per hectare. It reflects the grass growing capacity of grazing land that may not be properly captured by the other two land quality variables. Since such missing variables (like temperature) would in general be spatially correlated, the inclusion of dairy cows per hectare should contribute to reducing the possibility to find spatial autocorrelation. The choice of dairy cows instead of livestock units is determined by the fact that dairy cows require better land than other livestock and in the period under study due to the higher profitability of dairying, better land plots were generally 'reserved' for dairying.

Access to road is self explanatory indicator variable. The Distance to nearest urban area is measured in metres and is computed using GIS procedures. Finally the potential site indicates whether, according to the buyer, there is potential building site included within the land parcel. Patton and McErlean (2003, 2004) consistently find that this variable is insignificant, which is hardly surprising, since it is designed to capture the influence of non-agricultural factors, while only land to be used for agricultural purposes is included in the sample. Furthermore, it represents a subjective buyers' view, which, as long as the sellers do not share it, would not affect the price. Some descriptive statistics are presented in table 1.

## Results

First, for illustration purposes we present estimation results for the linear model. Table 2 shows the results from several spatial dependence tests. We present these for both a linear and a log-linear model. We use an inverse squared distance spatial weighting matrix. The standard LM tests for both forms of spatial dependence are significant. Similarly the DLR (double length artificial regression) tests introduced by Baltagi and Li (2000), which are similar to the LM tests but have better small sample properties, are both significant. When the robust (to the presence of the other form of spatial dependence, for more details see Anselin, 1988) LM tests are applied however, only the one for spatial lag one is significant. Thus one could conclude that there is spatial lag dependence but not spatial error dependence.

### Insert Table 2.

The portmanteau test is essentially a joint version of the robust spatial lag LM test and the standard spatial error LM test. It tests whether both forms of spatial autocorrelation are present. Note that it is highly significant thus rejecting the null. This is consistent with the robust LM tests above and suggests that only spatial lag is present.

The last test is the spatial Durbin test. It exploits the fact that the spatial lag representation can nest within itself spatial error dependence (the so called spatial Durbin model). It is essentially an LR test on the general spatial Durbin model against the spatial error model and tests whether the restrictions implied by the latter are valid. The spatial Durbin test statistic is insignificant. This is at odds with the previous tests, because it suggests that the spatial error restriction cannot be rejected. The



apparent contradiction between the tests for different forms of spatial dependence, indicates a misspecification problem. Since the spatial lag autocorrelation can account for (spatially correlated) omitted variables, one should be inclined to deduce that a likely source of this misspecification is the functional form assumption. Note that we reach the same conclusion about the loglog functional form.

A summary of the estimation results for the linear spatial lag model is presented in Table 3. The two stage least squares (2SLS) estimation is implemented following Kelejian and Prucha (1998) by using spatially lagged independent variables as instruments for the spatially lagged dependent variable.

### **Insert Table 3.**

What is remarkable about these results is that the spatial lag parameter is considerably different between the ML and the 2SLS estimators, which also can indicate some problems with the functional representation.

Now we proceed to the results from the IVQR model, presented on figures 1-8. We estimate the whole quantile process which produces separate coefficients estimates for every observation in the sample. We use an equidistant grid over the interval  $[-1.5, 1.5]$  with increments of 0.01 to search for the optimal values of the spatial lag coefficient. The coefficient estimates for all variables are plotted together with their 95% confidence bounds. We omit the estimates for the intercept in the model, since it is not readily interpretable. For comparison reasons we also plot the 2SLS estimates with robust standard errors and their corresponding 95% confidence intervals. Due to the heteroscedasticity correction, applied in the latter estimates, they should be the linear model estimates most comparable to the quantile regression results. Additionally we plot both the asymptotic and the finite sample confidence intervals for the IVQR estimates.

Note that we view the quantile regression as a semiparametric model and thus use a graphical representation for the results, as it is customary for non and semi-parametric estimation. In most quantile regression applications, only a small subset of quantile

regressions is estimated and these are presented in a tabular form. For example Zietz *et al.* (2008) estimate 9 separate quantile regressions. Here we advocate for the use of the quantile regression models as a semiparametric alternative. This means estimating the whole quantile regression process where possible. In this case due to the rather small dataset, this involves estimating only 197 separate quantile regressions. When the dataset is large, this may not be practical. For example Zietz *et al.* (2008) use a dataset consisting of 1366 observations. In such cases a regular grid at e.g. every percentile could provide a reasonable approximation.

At first sight one may notice that the confidence intervals for the quantile regression estimates are comparable, in terms of size, to the confidence intervals for the corresponding 2SLS estimates. The finite sample inference approach of Chernozhukov and Hansen (2008) generally produces wider confidence intervals when compared to the asymptotic inference method.

Let us first consider the land quality variables. These are land quality score, drainage score and dairy cows per hectare. The land quality score coefficients are negative indicating that better quality land (i.e. lower quality score) is valued more (see Figure 1). Note however that the coefficients for the land quality scores are insignificant for the higher quantiles i.e. for the more expensive land parcels. One may also notice that there is some significant difference between the asymptotic and the finite sample inference results for these higher quantiles. Owing to weaker identification, inverting the corresponding Wald tests at the higher quantiles produces considerably wider confidence intervals.

The drainage score coefficients are also negative which conforms to the expectations (Figure 2). The coefficient estimates are broadly similar to the parametric specification, except at the lower and the higher quantiles. Yet again at the higher quantiles the finite sample inference method produces considerably wider confidence intervals. In this case finite sample IVQR inference yields insignificant coefficient estimates at the higher quantiles, in contrast to the asymptotic one.

The coefficients for dairy cows per hectare (Figure 3) are all insignificantly different from zero which is also consistent to the mean model. The coefficients for the potential site (Figure 4) are also insignificant, as in the mean model. There are however some observations at the very high quantiles for which the effect of potential

site is significantly positive. In principle the dataset is constructed based on purely agricultural land sales and this should be expected to exclude the effect of non-agricultural pressures on the price. This in general makes the coefficients of potential site, which is measured by the responses of the buyers, which may not shared by the sellers, insignificant. Whenever such pressures are not excluded from the dataset however, one could expect that this variable would have significant positive impact on the price. Since such non-agricultural opportunities would in general be more profitable than purely agricultural use of this land, they will only be pronounced in the more expensive parcels of land, i.e. in the higher quantiles. This is exactly the result we obtain.

The effect of access to road (Figure 5) which is highly significant in the mean regression however is not significant for most observations in the IVQR estimates. It seems to be significantly positive for the lower (according to the asymptotic IVQR confidence intervals only) and higher quantiles in our sample.

Distance to urban area has a significant negative effect consistent with expectations and with the mean model (Figure 6). The quantile regression coefficients however show considerable variability compared to the linear estimates. Additionally there is considerable difference between the asymptotic and the finite inference confidence intervals. The nature the effect of distance to urban area on the price of agricultural land is complex. In principle the desirability of land parcels depends on their accessibility. This feature is proxied here by the indicator Access to road, but it also depends on the nature of the local infrastructure, i.e. quality of the road links, access to input and output markets etc. Obviously distance to urban area is a very imprecise proxy for these characteristics. For these reasons it only weakly identifies the endogenous spatial variation. This results in significant differences between asymptotic and finite sample inference.

The coefficients of acreage are not significant except for the extreme low quantiles and for the higher quantiles (Figure 7). The result for the extreme low quantiles may be due to the unreliability of the conventional quantile regression estimates at extreme quantiles and for this reason we will not comment on it. The considerable number of significant negative effects at the higher quantiles however suggests that there is a price discount for higher acreages in the most expensive parcels of land. The large

differences between asymptotic and finite sample inference are probably due to the highly nonlinear nature of the effects of acreage, which is not sufficiently well approximated by the linear quantile regression representation.

The difference between asymptotic and finite sample confidence intervals is most pronounced for the spatial lag coefficient. While the asymptotic inference discovers spatial lag dependence over most of the sample, the finite sample inference only finds evidence for spatial lag dependence in the higher quantiles of the dependent variable (Figure 8). At these higher quantiles we had the significant effect of access to road, acreage and potential site and the loss of significance of land quality score. Remembering that in general spatial lag dependence is inconsistent with the pure hedonic model, our results suggest market segmentation where the higher quantiles, in contrast to the rest of the sample deviate from the pure hedonic model. Note that similarly to the significance of the spatial lag coefficient, the other high quantile effects also suggest some kind of deviation from perfect competition. One may say that the hedonic model essentially breaks down at the higher quantiles, because none of the three land quantity variables is significant. The reason why the pure hedonic model breaks down for the higher quantiles is also obvious. If we abstract for a moment from the potential site effect, this is likely to be the best agricultural land, which as discussed earlier is in short supply in Northern Ireland. The latter means that the market for such land will be much thinner with the potential effects of creating distortions and deviations from the purely competitive market. As for the potential site cases, then due to the nature of residential planning, there could be spatial spillovers. One can formally test whether the model is different at the higher quantiles. To illustrate this we present in table 4 Wald-type tests for equality of slopes (i.e. the coefficients), constructed along the lines of Basset and Koenker (1982).

#### **Insert Table 4**

We are essentially testing whether the model in the upper quantiles, represented here by the 0.95<sup>th</sup> quantile is the same as the median model (the 0.5<sup>th</sup> quantile) and in the lower quantiles (represented here by the 0.05<sup>th</sup> quantile). Both the joint and the individual (for separate quantile regression coefficients) tests are presented. The joint tests are highly significant demonstrating the difference in the quantile regression model in the upper quantiles and the rest of the sample. The individual tests show

which coefficients contribute to this difference. Note however, that these tests have an auxiliary function. The primary point of interest here is not exactly how different are the estimated quantile regression coefficients, but their statistical significance, which as explained earlier, allows us to effectively split the sample into two qualitatively different segments. Therefore we are mainly interested in the joint tests as supporting evidence for the discovered market segmentation.

A peculiar characteristic of the Northern Ireland (and Ireland) land market is the conacre system, under which land is only rented on a short-term basis of up to 11 months. This system effectively creates information about the productive characteristics of agricultural land. The main stakeholders have to some extent directly or indirectly access to this type of information and therefore this contributes to a more efficient land pricing. We can however hypothesise that due to its scarcity the best agricultural land is rarely available for conacre rental. This means that it is much more difficult for the interested buyers to reliably assess its productive ability. Our results suggest that this is indeed the case, since at the higher quantiles we discover significant deviations from the fully competitive hedonic model. Owing to the small size of the agricultural land market in Northern Ireland, a signaling system, such as the conacre one, is instrumental in facilitating more efficient market pricing. It helps reduce market inefficiencies. Therefore a transition towards a longer term based rental system, as in Great Britain, can be expected to impact negatively on the land market in Northern Ireland.

## **Conclusions**

We have applied a spatial lag quantile regression to a hedonic land prices model. In this way we allow for varying effects of the hedonic characteristics and more importantly varying degrees of spatial autocorrelation. We apply this approach to a sample of agricultural land sales in Northern Ireland. Due to the parametric rate of convergence of the quantile regression estimator the estimated confidence intervals compare favourably to those from a parametric spatial lag model. Therefore the proposed spatial quantile regression generalizes the linear spatial lag model at a relatively low cost and is applicable to small samples. Finite sample inference, robust to weak identification, is available. We demonstrate how the finite sample inference

can lead to qualitatively different results, in particular with regard to the spatial dependence. Our results suggest that the agricultural land market in Northern Ireland effectively consist of two segments. The larger of these two segments conforms to the conventional hedonic model with no spatial lag dependence, while the smaller much thinner market segment exhibits considerable spatial lag dependence. Although we use a linear quantile regression that cannot fully overcome the potential pitfalls of a functional misspecification, it is essentially a semi-parametric approach that is much more flexible than the conventional parametric modeling. Additionally the linear quantile regression has been extensively studied and provides tools for a fully parametric inference. Nevertheless, the approach could, if desired, be potentially extended to a more general non-parametric setting.

#### **References:**

- Abadie, A., J. Angrist and G. Imbens (2002) Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings, *Econometrica* 70, 91-117.
- Amemiya, T. (1982) Two stage least absolute deviations estimators, *Econometrica*, 50, 689-711.
- Anselin, L. (1988) *Spatial Econometrics: Methods and Models*, Kluwer Academic Press.
- Anselin, L. and J. Le Gallo. (2006). Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects, *Spatial Economic Analysis* 1 (1), 31-52).
- Baltagi, B. H. and D. Li (2001) Double length artificial regressions for testing spatial dependence, *Econometric Reviews*, 20(1), 31-40.
- Basile R. and B. Gress (2005) Semi-parametric Spatial Auto-covariance Models of Regional Growth Behaviour in Europe, *Région et Développement*, 21, 93-118.
- Bassett, G. and R. Koenker (1982). Tests of Linear Hypotheses and L1 Estimation, *Econometrica* 50, 1577–1583.
- Blundell, R. and J.L. Powell (2003) Endogeneity in semiparametric and nonparametric regression models. in: Dewatripont, M., Hansen, L.P. and Turnovsky,

This is an electronic version of the following article Kostov, P. (2009) A spatial quantile regression hedonic model of agricultural land prices, *Spatial Economic Analysis*, 4(1), 53-72. The final definite published version is available online at: <http://dx.doi.org/10.1080/17421770802625957>.

S.J. (eds.), *Advances in Economics and Econometrics: Theory and Applications*, Eighth World Congress, vol. II, Cambridge University Press, Cambridge, pp. 312–357.

Chen, L. and S. Portnoy (1996) Two-stage regression quantiles and two-stage trimmed least squares estimators for structural equation models, *Communications in Statistics, Theory and Methods*, 25, 1005-1032.

Chernozhukov, V. and C. Hansen (2006) Instrumental quantile regression inference for structural and treatment effect models, *Journal of Econometrics* 127, 491-525.

Chernozhukov, V., C. Hansen and M. Jansson (2007) Inference approaches for instrumental variable quantile regression, *Economics Letters* 95, 272–277.

Chernozhukov, V. and C. Hansen (2008) Instrumental variable quantile regression: A robust inference approach, *Journal of Econometrics* 142, 379–398.

Ekeland, I., J. Heckman and L. Nesheim (2004) Identification and estimation of hedonic models, *Journal of Political Economy*, 112(S1), S60–S109.

Fotheringham, A. S., M.E. Charlton and C. Brunsdon (1997) "Two Techniques for Exploring Non-stationarity in Geographical Data", *Geographical Systems*, 4, 59-82.

Fotheringham, A.S., C. Brunsdon and M.E. Charlton (2002) *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.

Freeman, M. (1993) *The Measurement of Environmental and Resource Values: Theory and Method*, Resources for the Future, Washington, DC.

Goodman A. C. and T. G. Thibodeau (2003) Housing Market Segmentation and Hedonic Prediction Accuracy, *Journal of Housing Economics* 12, 181-201.

Greene, W. (2003) *Econometric Analysis*, 3rd Edition, Prentice Hall, Englewood Cliffs.

Kelejian, H.H. and I.R. Prucha. (1998) A generalized spatial two stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics* 17, 99-121.

Kelejian, H. H. and I.R. Prucha (1999) A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* 40, 509-533.



This is an electronic version of the following article Kostov, P. (2009) A spatial quantile regression hedonic model of agricultural land prices, *Spatial Economic Analysis*, 4(1), 53-72. The final definite published version is available online at: <http://dx.doi.org/10.1080/17421770802625957>.

Kelejian, H. H. and I.R. Prucha (2006) Specification and estimation of spatial autoregressive models with autoregressive and heteroscedasticity disturbances. Working paper, University of Maryland.

Kim, T. H. and C. Muller (2004). Two-stage quantile regression when the first stage is based on quantile regression, *Econometrics Journal*, 7, 218-231.

Koenker, R. (2005) *Quantile Regression*, Cambridge: Cambridge University Press.

Koenker, R. and G. Bassett (1978) Regression Quantiles, *Econometrica* 46, 33-50.

Lee, L.F. (2002) Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models, *Econometric Theory* 18, 252-277.

Lancaster, K. J. (1966) A new approach to consumer theory. *Journal of Political Economy*, 74, 132–157.

Lee, L-F. (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances, *Econometric Reviews* 22, 307-335.

Lee, L-F. (2007) GMM and 2SLS estimation of mixed regressive, spatial autoregressive models, *Journal of Econometrics* 137, 489-514.

Lee, S. (2004) Endogeneity in Quantile Regression Models: A Control Function Approach, Institute for Fiscal Studies, CEMMAP Working Paper CWP08/04.

Lin, X. and L-F. Lee (2006) GMM estimation of spatial autoregressive models with unknown heteroscedasticity, Working paper, Department of Economics, Ohio State University.

Ma, L. and R. Koenker (2006) Quantile regression methods for recursive structural equation models, *Journal of Econometrics*, 134(2), 471-506.

Maddison, D. (2004) Spatial effects within the agricultural land market in Northern Ireland: a comment, *Journal of Agricultural Economics*, 55(1): 123-125.

McMillen D.P. (2003) Spatial autocorrelation or model misspecification? *International Regional Science Review* 26: 208–217.

This is an electronic version of the following article Kostov, P. (2009) A spatial quantile regression hedonic model of agricultural land prices, *Spatial Economic Analysis*, 4(1), 53-72. The final definite published version is available online at: <http://dx.doi.org/10.1080/17421770802625957>.

Nesheim, L. (2002) Equilibrium sorting of heterogeneous consumers across locations: Theory and empirical implications, Working Paper No. CWP08/02, Centre for Microdata Methods and Practice.

Patton, M. and S. McErlean (2003) Spatial effects within the agricultural land market in Northern Ireland, *Journal of Agricultural Economics*, 54(1): 35-54.

Patton, M. and S. McErlean (2004) Spatial effects within the agricultural land market in Northern Ireland: a reply, *Journal of Agricultural Economics*, 55(1): 127-133.

Powell, J.L. (1983). The asymptotic normality of two-stage least absolute deviations estimators, *Econometrica*. 50, 1569-1575.

Rosen, S. (1974) Hedonic Markets and Implicit Prices: Product Differentiation and Pure Competition. *Journal of Political Economy*, 82: 34–55.

Su, L. and Z. Yang (2007) Instrumental Variable Quantile Estimation of Spatial Autoregressive Models, Working paper 05-2007, Singapore Management University.

Williams, G. (1989) The effects of MRT on land values, *Journal of Valuation*, 8(1), 7-23.

Zietz, J., E. Zietz and S. Sirmans (2008) Determinants of House Prices: A Quantile Regression Approach, *Journal of Real Estate Finance and Economics*, forthcoming.

Table 1. Descriptive statistics

Variable	Units	Mean	Standard deviation	Minimum	Maximum
Price	£/ha	3559.26	1551.65	413.96	9609.72
Land quality Score	scale 1-7	3.66	0.78	2.00	5.06
Drainage score	scale 1-10	3.01	3.09	1.00	10.00
Dairy cows per hectare		0.36	0.19	0.01	0.99
Distance to Urban area	meters	4507.86	3757.72	64.67	30438.40
Acreage	acres	27.39	33.28	0.90	296.00
Access to road	indicator	0.53	0.50	0.00	1.00
Potential site	indicator	0.32	0.47	0.00	1.00

Table 2. Spatial autocorrelation tests

Test	Linear model		Loglog model	
	Test statistic	P-value	Test statistic	P-value
Standard spatial error LM test	6.196	0.013	8.634	0.003
Standard spatial lag LM test	9.162	0.002	13.072	0.000
DLR spatial error test	5.673	0.017	7.658	0.006
DLR spatial lag test	8.666	0.003	12.513	0.000
Robust spatial error LM test	0.232	0.630	0.073	0.787
Robust spatial lag LM test	3.198	0.074	4.511	0.034
Portmanteau test	9.394	0.009	13.145	0.001
Spatial Durbin Test	9.823	0.199	10.830	0.146

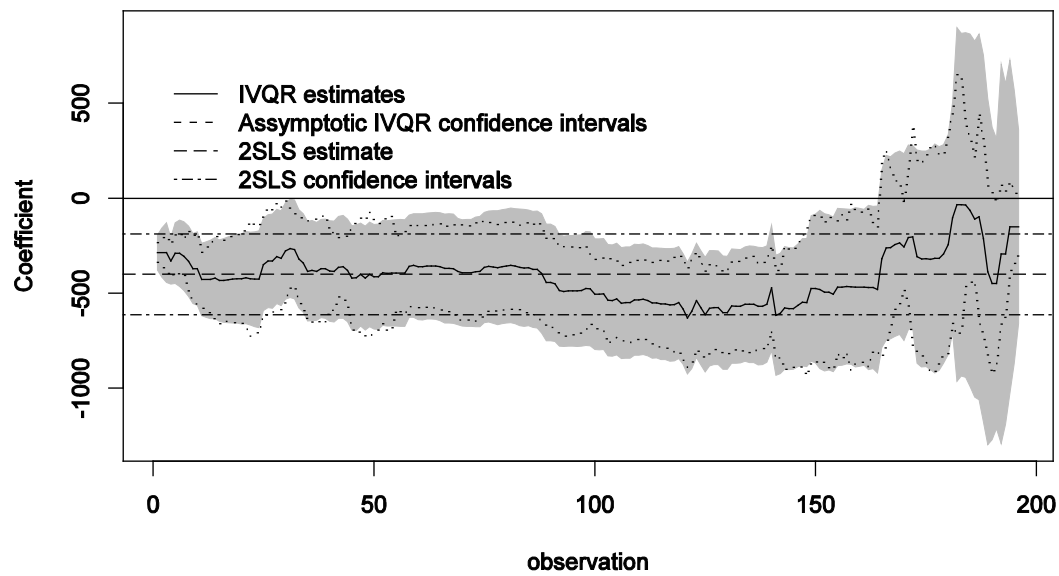
Table 3. Estimates for the linear spatial lag model

	Maximum Likelihood		2SLS		2SLS with robust standard errors	
	Coefficient	P. Level	Coefficient	P. Level	Coefficient	P. Level
(Intercept)	4767.305	0.000	4252.019	0.000	4347.197	0.000
Acreage	-5.706	0.029	-5.798	0.030	-4.880	0.019
Land Quality Score	-390.885	0.001	-373.219	0.003	-400.163	0.000
Distance to urban area	-0.117	0.000	-0.112	0.000	-0.111	0.000
Potential site	151.355	0.412	144.823	0.441	52.146	0.781
Drainage score	-152.566	0.000	-148.192	0.000	-144.184	0.000
Dairy cows per hectare	476.226	0.309	379.453	0.441	591.815	0.222
Access to road	462.365	0.009	473.426	0.009	411.528	0.011
Rho	0.247	0.002	0.370	0.036	0.347	0.034

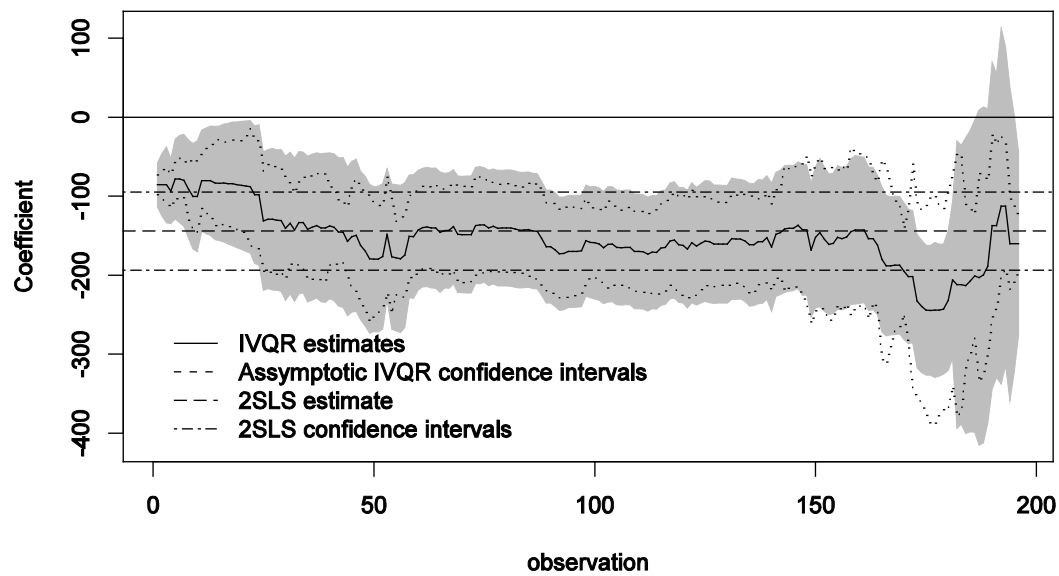
Table 4. Wald tests for equality of slopes (0.95 against 0.5 and 0.05 quantile)

	against the 0.5 quantile		against the 0.05 quantile	
	Test statistic	P value	Test statistic	P value
Joint test	2.691	0.007	2.996	0.003
Individual tests				
Spatial lag	3.748	0.054	5.926	0.015
Acreage	0.810	0.369	0.593	0.442
Land Quality Score	3.652	0.057	1.994	0.159
Distance to urban area	1.227	0.269	1.234	0.267
Potential site	0.000	0.993	0.052	0.820
Drainage score	0.850	0.357	3.499	0.062
Dairy cows per hectare	0.370	0.544	0.409	0.523
Access to road	7.706	0.006	3.233	0.073

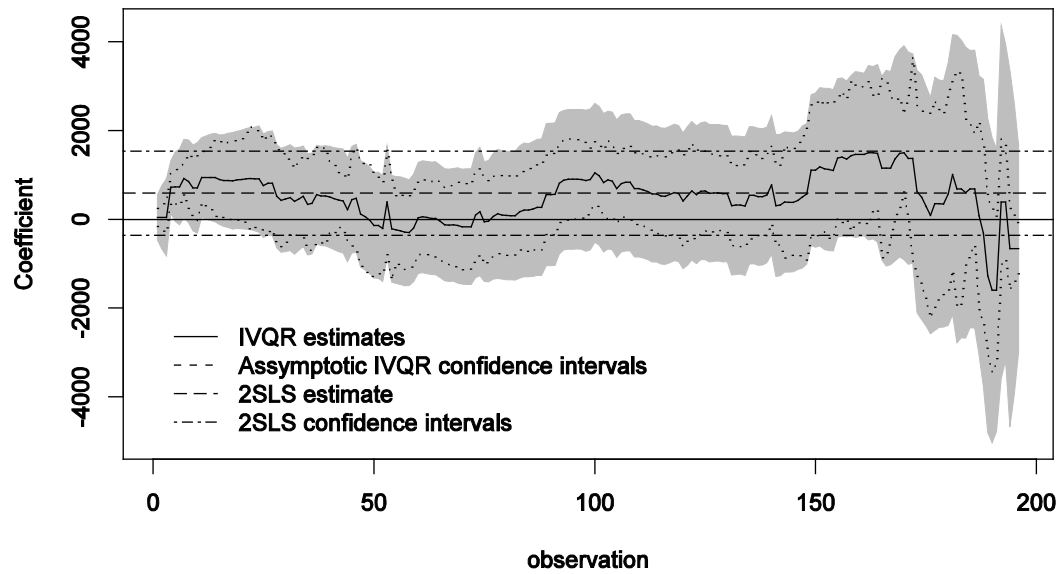
**Figure.1. Coefficients and 95% confidence intervals for Land quality score**



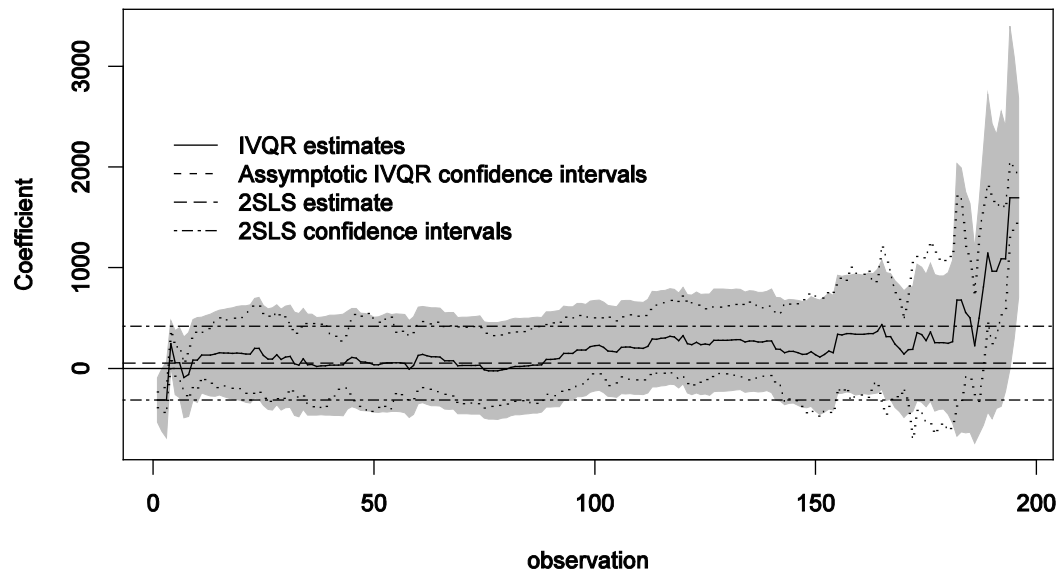
**Figure.2. Coefficients and 95% confidence intervals for Drainage score**



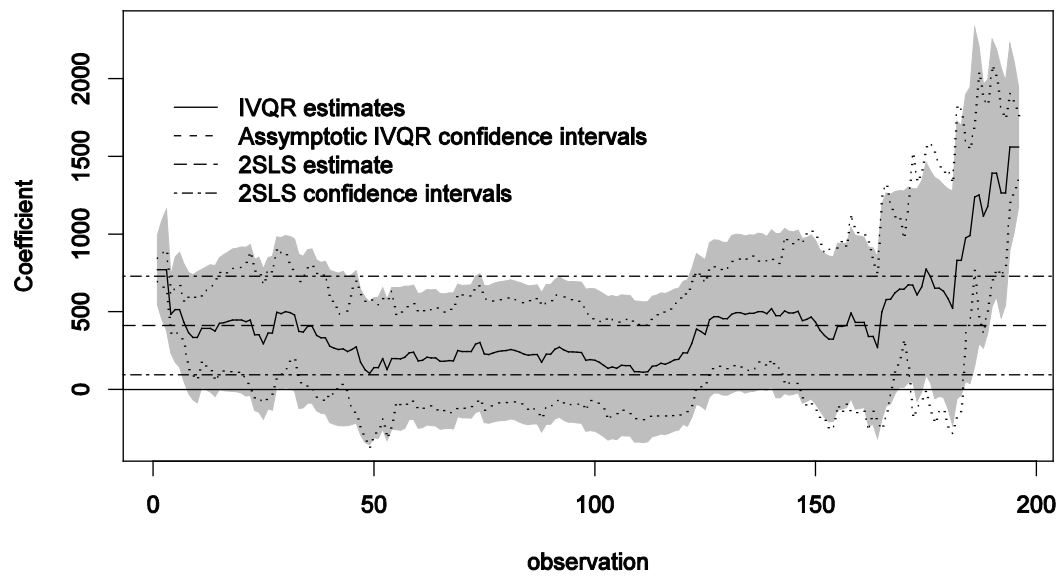
**Figure.3. Coefficients and 95% confidence intervals for Dairy cows per hectare**



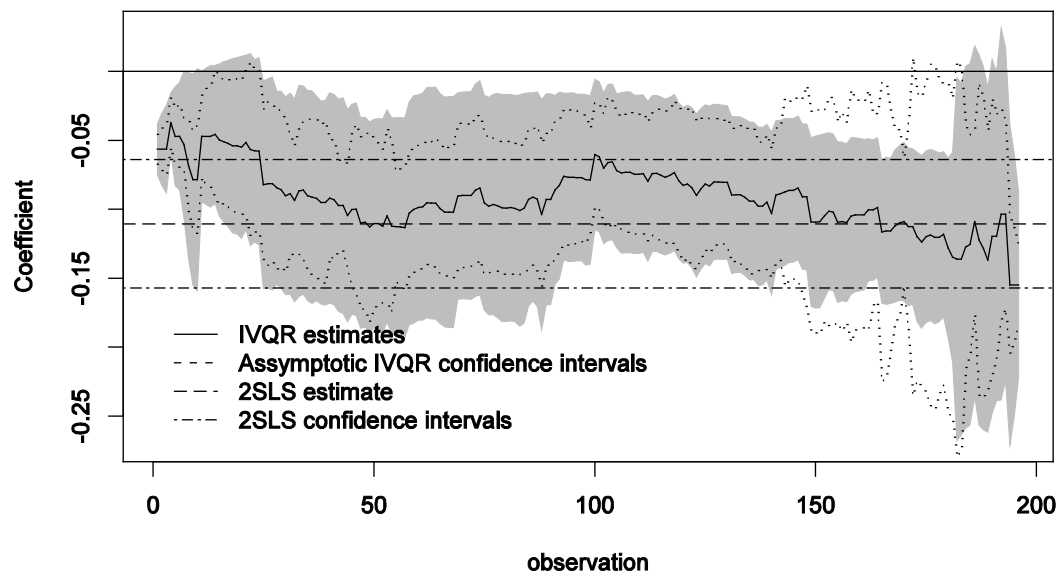
**Figure.4. Coefficients and 95% confidence intervals for Potential site**



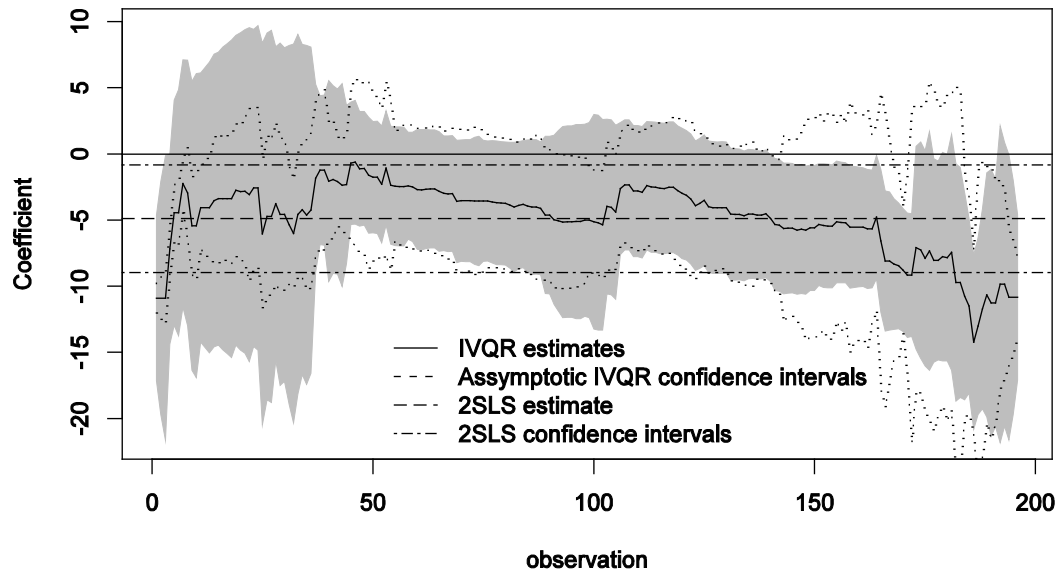
**Figure.5. Coefficients and 95% confidence intervals for Access to road**



**Figure.6. Coefficients and 95% confidence intervals for Distance to urban area**



**Figure.7. Coefficients and 95% confidence intervals for Acreage**



**Figure.8. Coefficients and 95% confidence intervals for Spatial lag**

