VERSITA

10.2478/v10103-012-0040-8

GRAŻYNA TRZPIOT[*]

Spatial Quantile Regression

Abstract

In a number of applications, a crucial problem consists in describing and analyzing the influence of a vector X_i of covariates on some real-valued response variable Y_i. In the present context, where the observations are made over a collection of sites, this study is more difficult, due to the complexity of the possible spatial dependence among the various sites. In this paper, instead of spatial mean regression, we thus consider the spatial quantile regression functions. Quantile regression has been considered in a spatial context. The main aim of this paper is to incorporate quantile regression and spatial econometric modeling. Substantial variation exists across quantiles, suggesting that ordinary regression is insufficient on its own. Quantile estimates of a spatial-lag model show considerable spatial dependence in the different parts of the distribution.

1. Introduction

1.1 Linear Regression - introduction

Linear regression is the standard tool for many empirical studies. When the relationship between a dependent variable, *y*, and a set of explanatory variables, ***X***, can be written as $y = X\beta + u$, a simple ordinary least squares

[*] Ph.D., Professor at the University of Economics in Katowice

(OLS) regression of y on X can provide unbiased estimates of the parameters, β, and a predicted value, $\hat{y} = X\hat{\beta}$.

This heavy reliance on linear regression models has been carried over to the analysis of spatial data. The most commonly used spatial model adds a weighted average of nearby values for the dependent variable to the list of explanatory variables:

$$y = \rho WY + X\beta + u\,. \tag{1}$$

In this model, $\boldsymbol{W}$ is a "spatial weight" matrix that specifies the relationships between observations. W is a "spatial weight matrix" with rows that sum to one and zeros on the diagonals, and ρ is a parameter measuring the strength of the relationship.

The model can be useful when X does not fully account for the tendency for the dependent variable to be highly correlated over space, so that nearby values of Y provide significant explanatory power. The endogeneity of WY poses challenges for estimation. Most empirical applications are based on maximum likelihood estimation of the model under the assumption of normally distributed errors. Other approaches are based on instrumental variables (IV) estimation, usually with spatially lagged values of X (such as WX and WWX) as instruments for WY. Several researchers have used the spatial AR model as the basis for quantile regressions in which both ρ and β are allowed to vary across quantiles.

Typical specifications of the spatial weight matrix are based on first-order contiguity when the data are drawn from geographic units such as counties or census tracts. Though the approach is used less commonly for point data, typical specifications are similar in that the spatial weights are assumed to decline rapidly with distance between observations.

Predicted values are then based on

(1) the structural model

$$\hat{Y}(\tau) = \hat{\rho} WY + X\hat{\beta}$$

(2) the reduced form

$$\hat{Y}(\tau) = (I - \hat{\rho} W)^{-1} X\hat{\beta}$$

(3) or a decomposition into "trend" and"signal" components

$$\hat{Y}(\tau) = X\hat{\beta} + \hat{\rho} W(I - \hat{\rho} W)^{-1} X\hat{\beta}$$

Spatial effects generally appear as noise around a spatial trend that looks much like the predicted values from an OLS regression of Y on X. The objective

of a regression analysis is to estimate the coefficients, ρ and β, and to obtain predictions of y at given values of $\boldsymbol{X}$.

Regression analysis is not well suited to explaining the *distribution* of a variable. When the predicted values from a regression are $\hat{y} = X\hat{\beta}$, then the distribution of the predicted values simply mimics the distribution of the variables in X. The implied effect of a change in one of the explanatory variables is to cause a *parallel* shift of $\hat{y}$ by an amount determined by the variable's estimated coefficient. Though a parallel shift may be reasonable in some cases, it is limitation that a research may not want to impose beforehand.

1.2. Quantile Regression

Quantile regression is a method for estimating functional relations between variables for all portions of probability distribution. Typically a response variable Y is some function of predictor variable X. Regression application focus in estimating rates of changes in the mean of the response variable distribution as some function of a set of predictor variables. In the other words the function is defined for the expected value of Y conditional X, $E(\mathrm{Y}|\mathrm{X})$. Regression analysis gave incomplete picture of the relationships between variables especially for regression models with heterogeneous variances.

Quantile regression was developed as an extension of the linear model for estimating rate of change in all parts of the distribution of response variables. The estimates are semi parametric in the sense that no parametric distributional form (eg. normal, Poisson, negative binominal, etc.) is assumed for the random error part of the model ε, although a parametric form is assumed for the deterministic portion of the model (eg. $\beta_0 X_0 + \beta_1 X_1$). The conditional quantiles denoted by $Q_y(\tau|\mathrm{X})$ are the inverse of the conditional cumulative distribution function of the response variable $\mathrm{F}_y^{-1}(\tau|\mathrm{X})$, where $\tau \in [0, 1]$ denotes quantile rank.

The quantile model posits the τ^{th} quantile of Y conditional on x to be,

$$\mathrm{Q}(\tau|\mathrm{x}) = \alpha(\tau) + \mathrm{x}\beta(\tau),\ 0 < \tau < 1. \tag{2}$$

If $\beta(\tau)$ is a constant β, the model reduces to the standard conditional expectation model, $\mathrm{E}(\mathrm{Y}|\mathrm{x}) = \alpha + \mathrm{x}\beta$, with constant variance errors. When $\beta(\tau)$ depends on τ, the model allows the distribution of Y to depend on x in different ways at different parts of the distribution. The traditional linear model can be viewed as a summary of all the quantile effects; that is, $\int Q(\tau|x)d\tau = E(Y|x)$.

Under this interpretation, traditional analysis loses information due to its aggregation of possibly disparate quantile effects. Many different quantile paths, for example, can lead to $\beta_k = 0$. On the one hand, $\beta_k = 0$ can mean x_k does not matter – does not affect the distribution of Y. But it can also mean there are important but compensating quantile effects relating Y and x.

Quantile regression is much better suited to analyzing questions involving changes in the distribution of a dependent variable. Quantile regressions allow for separate effects of an explanatory variable on different points of the dependent variable distribution. The coefficient estimates are then frequently interpreted as being analogous to standard linear regression estimates, albeit for different points in the distribution of the dependent variable (Trzpiot 2008, 2009 a, b, c, 2010, 2011 a, b).

It is less commonly recognized that quantile regression can produce estimates of changes in the full distribution of the dependent variable when the values of the explanatory variables change. The set of coefficients produced for independent variables imply a change in the full distribution dependent variables.

Special issues do not necessarily arise when estimating quantile regressions using spatial data. Several researchers have proposed variants of the spatial autoregressive (AR) model, $y = \rho WY + X\beta + u$, for quantile analysis.

These procedures treat $\boldsymbol{WY}$ as just another endogenous explanatory variable. The spatial AR model may not necessarily be the best choice for spatial modeling, particularly for large data sets comprising individual geographic points rather than large zones or tracts. In situations where the distribution of the dependent variable changes smoothly over space, a nonparametric procedure may be a much better approach.

2. Distribution of the Dependent Variable

In general, the conditional quantile function for y given a set of variables X can be written:

$$Q_y(\tau|X) = X\beta(\tau|X), \quad 0 < \tau < 1. \tag{3}$$

Usually, we have limited our attention to a small number of values for the quantile, τ. Focusing on that values provides useful information about the distribution of the dependent variable given values of X, but it certainly does not provide a complete picture of the full distribution of y.

One way to use quantile regression estimates to simulate the distribution of the dependent variable is to draw randomly from possible values of and then estimate a separate quantile regression for each value of τ. For example, we might draw 1000 values of τ from a uniform distribution ranging from 0 to 1, i.e., $\tau \sim$ U(0, 1). If we let *J* represent the number of draws from the U(0,1) distribution, then we have:

$$\hat{Q}_y(\tau_j \mid X) = X\hat{\beta}(\tau_j \mid X) \quad j = 1,\ldots, J \tag{4}$$

With *J* estimates of the conditional quantile in hand, a standard kernel density function can be applied to $X\hat{\beta}(\tau_j \mid X)$ to estimate the density function for the dependent variable. Since quantile estimates are generally fairly smooth across, drawing multiple values of τ from a U(0,1) distribution is a very inefficient way of constructing the density function. Using a limited range of value for τ is more efficient. For example, we might restrict the estimates to τ, , or a still more limited set of values for that provides good coverage of the set of permissible values for τ. Since quantile estimates are likely to have very high variances at extreme values of τ such as 0,01 or 0,99, it generally is a good idea to trim the extreme observations if a grid of values is used for τ.

3. The Effect of a Discrete Change in an Explanatory Variable

Quantile regression estimates can have interesting implications for the distribution of *y* values even in a model with a single explanatory variable. Consider a model with *k* explanatory variable in addition to the intercept. After estimating quantile regressions for *J* quantiles, the predicted values for quantile are simply τ_j:

$$\hat{Q}_y(\tau_j \mid X) = \hat{\beta}_0(\tau_j) + \hat{\beta}_1(\tau_j)x_1 + ... + \hat{\beta}_k(\tau_j)x_k, \quad j = 1,\ldots,J \tag{5}$$

When simplified the notation by replacing $\beta(\tau_j \mid X)$ with $\beta(\tau_j)$, but it should be clear that the estimates depend on the observed values of X. Even in the single-explanatory case where k = 1, the implied effect of changing from to produces *J* separate values for

$$\hat{Q}_y(\tau_j \mid X, x_1 = \delta_0) = \hat{\beta}_0(\tau_j) + \hat{\beta}_1(\tau_j)\delta_0 + ... + \hat{\beta}_k(\tau_j)x_k, \; j = 1,\ldots,J \tag{6}$$

$$\hat{Q}_y(\tau_j \mid X, x_1 = \delta_0) = \hat{\beta}_0(\tau_j) + \hat{\beta}_1(\tau_j)\delta_1 + ... + \hat{\beta}_k(\tau_j)x_k\ ,\ j = 1,\ldots,J \qquad (7)$$

With *J* quantiles and *n* observations, equation (6) and (7) imply *nJ* values for the conditional quantile functions. Since $\hat{\beta}_1(\tau_j)$ is not constant, the conditional quantile functions imply a full distribution of values for *y* even when x_1 is the only variable in the model.

4. Quantile version of the Spatial AR Model

The analysis up to this point has not been explicitly spatial. Although the explanatory variables might include measures of access to various amenities such as a city's central business districts, parks, or lakes, nothing yet is unique to the analysis of spatial data. Several attempts have been made to adapt the standard spatial autoregressive (AR) model for quantile regression.

The spatial AR model adds a weighted average of nearby values of the dependent variable to the list of explanatory variables. The model is written $y = \rho WY + X\beta + u$, where ***X*** is the *nxk* matrix of explanatory variables, ***Y*** is the dependent variable, and ***W*** is an *nxn* matrix specifying the spatial relationship between each value of ***Y*** and its neighbors.

Suppose the observations represent census tracts. If each tract is contiguous to four other tracts, then W_{ij} = ¼ for each of the four tracts that is contiguous to observation *i*, and W_{ij} = 0 for all other values of *j*. In this example, each of the n elements of ***WY*** is simply the average, for each observation, of the four neighboring values of ***Y***. More generally, if observation *i* is contiguous to other tracts, then $W_{ij} = 1/n_i$ for the tracts that are contiguous to observation *i*, and W_{ij} = 0 otherwise. For point data, ***WY*** might form a weighted average of the nearest *K* neighbors, or the weights might decline with distance. ***WY*** is clearly an endogenous variable. Indeed, one interpretation of ***WY*** is that it is the set of predicted values from kernel regressions of ***Y*** on the set of geographic coordinates. For example, suppose we were to write $y_i = f(lo_i, la_i) + u_i$. If we use a rectangular kernel with a very small window size e.g., the four closest observations – then the cross-validation version of the kernel regression estimator is

$$\hat{y}_i = \frac{1}{n_i}\sum_{j=1}^{n} I_j y_j\ , \qquad (8)$$

where I_j indicates that observation *j* is one of the nearest neighbors to observation *i*, and indicates the number of observations that are being given weight when constructing the estimate for observation *i*. Not surprising, adding the predicted value of ***Y*** as an explanatory variable for ***Y*** often produces highly significant results.

Although the spatial lag variable, ***WY***, is formally equivalent to a kernel regression, the approaches could hardly be more different in spirit.

The spatial AR model is based on an assumption that the researcher can truly specify the full spatial relationship between all of the observations. After specifying the entire path by which each of *n* observations can influence all of the other observations, all that is left is to determine the strength of the relationship by estimating ρ.

In contrast, nonparametric and semi-parametric regressions involve far less difficult. We could easily write the model in semiparametric form as

$$y_i = f(lo_i, la_i) + X_i\beta + u_i \tag{9}$$

or in the conditionally parametric form

$$y_i = X_i\beta(lo_i, la_i) + u_i . \tag{10}$$

The spatial AR model is based on the assumption that the researcher can specify a simple parametric function that accounts for both the relationship between ***X*** and ***Y*** and the entire spatial relationship between all observations. Nonparametric approaches are based on an assumption that the researcher can correctly specify the variables that influence ***Y***, but they allow for local variation in the marginal effect of ***X*** on ***Y***. The spatial AR model and its variants may be useful in situations where the objective is to estimate a causal relationship between ***Y*** and neighboring values of the dependent variable.

4.1 Quantile Regression with an Endogenous Explanatory Variable

The spatial AR model is most commonly estimated by maximizing the log-likelihood function that is implied under the assumption of normally distributed errors. An alternative approach based on the generalized method of moments method allows the model to be estimated using a variant of two-stage least squares (2SLS). In the first stage, the endogenous variable, ***WY***, is regressed on a set of instruments.

The predicted value of $\boldsymbol{WY}$ is then used as an explanatory variable in the second stage regression. Most researchers use $\boldsymbol{X}$, $\boldsymbol{WX}$, and, sometimes, additional orders of the spatial lags such as $\boldsymbol{WWX}$ as instruments for $\boldsymbol{WY}$.

Though this method can work well when the goal is to estimate a standard regression, quantile regression may be more complex because instrumental variables are needed for $\boldsymbol{WY}$ when estimating a regression for each quantile, τ.

Two methods have been used to form the instrumental variables needed these quantile regressions. The simpler version was proposed by Kim and Muller (2004). Their approach is a straightforward extension of 2SLS.

For each value of τ, they first estimate a quantile regression for $\boldsymbol{WY}$ using the set of instruments (e.g., $\boldsymbol{X}$ and $\boldsymbol{WX}$) as explanatory variables. The predicted values from the quantile regression are $\hat{WY}(\tau)$. In the second stage, they estimate another quantile regression for the same value of τ, this time with $\boldsymbol{Y}$ as the dependent variable and $\boldsymbol{X}$ and $\hat{WY}(\tau)$ as the explanatory variables.

Only 10 quantile regressions are needed to estimate the model for 5 quantiles (e.g., τ = 0,10, 0,25, 0,50, 0,75, 0,90). Zietz et al. (2008) and Liao and Wang (2012) use this approach to estimate quantile versions of the spatial AR model. They use bootstrap procedures to construct standard error estimates.

Though somewhat more complicated, the Chernozhukov and Hansen (2006) approach may be more robust than the Kim and Muller (2004) approach because it does not require that the same quantile be used in both stages of the procedure.

An additional advantage is that Chernozhukov and Hansen present a covariance matrix estimate that is easy to construct.

In the version describe here, the predicted values, $\hat{WY}$, from an OLS regression of $\boldsymbol{WY}$ on the instruments are used as the instrumental variable for $\boldsymbol{WY}$. This instrumental variable is then used as an explanatory variable for a series of quantile regressions $\boldsymbol{Y} - \rho\boldsymbol{WY}$ of on $\boldsymbol{X}$ and $\hat{WY}$ The same quantile, τ, is used for each of the regressions, while a grid of alternative values is used for ρ. The estimated value of ρ is the value that produces the coefficient on $\hat{WY}$ that is closest to zero. After finding $\hat{\rho}$ the estimated values of β are calculated by a quantile regression of $\boldsymbol{Y} - \hat{\rho}\boldsymbol{WY}$ on X. The motivation behind this estimator is a property of two-stage least squares: when instruments are chosen optimally, the coefficient on $\hat{WY}$ will be zero when both the actual variable, $\boldsymbol{WY}$, and the instrumental variable are included in a regression.

Standard error estimates are easy to construct for the Chernozhukov and Hansen method. Let e represent the residuals from the quantile regression of $\boldsymbol{Y} - \hat{\rho}\boldsymbol{WY}$ on $\boldsymbol{X}$, and define $I(|e_i| < h)/(2h)$, where h is a constant bandwidth.

Define $\Phi_i = f_i \hat{WY}_i$ and $Z_i = f_i X_i$. Then the covariance matrix for $\hat{\theta} = (\hat{\rho}, \hat{\beta})$ is:

$$V(\hat{\theta}) = J(\tau)^{-1} S(\tau) J(\tau)^{-1'} \tag{11}$$

where

$$J(\tau) = \begin{bmatrix} \Phi' WY & \Phi' X \\ Z' WY & Z' X \end{bmatrix} \text{ and } \quad S(\tau) = \tau(1-\tau) \begin{bmatrix} \hat{WY}' WY & \hat{WY}' X \\ X' WY & X' X \end{bmatrix}$$

As is the case for any instrumental variables (IV) estimator, the estimates from either approach can be sensitive to the choice of instruments. However, an important advantage of the IV approach over maximum likelihood estimation, which is commonly used for the non-quantile version of the spatial AR model, is that that there is no need to invert the nxn matrix $(I - \rho W)^{-1}$ when estimating the model.

It may still prove necessary to invert large matrices when constructing predicted values for *Y*. Let $\hat{Y}(\tau)$ denote the set of predicted values of the dependent variable for quantile τ.

Three procedures are often used to construct $\hat{Y}(\tau)$:

quantile version of the structural model

$$\hat{Y}(\tau) = \hat{\rho}(\tau) WY + X\hat{\beta}(\tau) \tag{12}$$

quantile version of the reduced form

$$\hat{Y}(\tau) = (I - \hat{\rho}(\tau) W)^{-1} X\hat{\beta}(\tau) \tag{13}$$

quantile version of a decomposition into "trend" and"signal" components

$$\hat{Y}(\tau) = X\hat{\beta}(\tau) + \hat{\rho}(\tau) W (I - \hat{\rho}(\tau) W)^{-1} X\hat{\beta}(\tau) \tag{14}$$

Though the first procedure may be viewed as cheating because it uses actual values of ***WY*** to predict ***Y***, it is commonly used for standard linear simultaneous equations models.

The second version follows directly from the original model specification: the equation implies $Y = (I - \rho W)^{-1}(X\beta + u)$, from which equation (13) follows by setting *u* = 0. Finally, equation (14) is derived by noting that equation (13) also provides a way to estimate ***WY*** for the expression given in equation (12):

$$\hat{WY}(\tau) = W(I - \hat{\rho}(\tau) W)^{-1} X\hat{\beta}(\tau) \tag{15}$$

5 Conditionally Parametric Quantile Regression

Nonparametric approaches can be adapted to quantile regression models. In the case of a single explanatory variable, x, all that is necessary to make the model nonparametric is to add a kernel weight function $K((x-x_t)/h)$ when estimating a quantile regression for a target point x_t. After estimating the function for a series of target points, the estimates can then be interpolated to all values of x. The nonparametric approach is a flexible way to add nonlinearity to the estimated quantile regressions. Although nonparametric approaches can potentially be applied to variables with many explanatory variables, the variance of the estimated can become very high when there are more than two or three explanatory variables.

The problem can be simplified significantly in the case of spatial data sets. It is usually reasonable to assume that a simple linear model fits the data well in small geographic areas. The difficulty is in specifying a global parametric model that fits the data well across a large region. The spatial AR model allows for local variation around a global parametric trend. In contrast, a conditionally parametric approach allows for flexible trends by allowing the coefficients to vary smoothly over space. Nonparametric models can be hard to summarize because they produce separate estimates for every observation. As we have seen, this is not a problem for quantile regression models because the easiest way to interpret the results is to present sets of density estimates showing how the distribution of the dependent variable changes as the values of individual explanatory variable change. As a result, it is just as easy to estimate nonparametric quantile regressions as it is to use a linear approach.

5.1 CPAR Quantile Regression for Spatial Data

Consider the conditional quantile function $Q_y(\tau|X) = X\beta(\tau|X)$, in which the dependent variable is a linear function of a set of explanatory variables, X. Now suppose that we want to allow the coefficients to vary over space. Using *lo* and *la* to represent the geographic coordinates, we can write the conditional quantile function as:

$$Q_y(\tau|X, lo, la) = X\beta(\tau|X, lo, la) \tag{16}$$

Although it is possible to include *lo* and *la* as explanatory variables – in which case they are part of the X matrix – the more common approach is to keep

them separate. Also, note that *lo* and *la* can represent any geographic coordinate system rather than just longitude and latitude.

All that is necessary to estimate a nonparametric version of equation (16) is to specify a kernel weight function that indicates the weight given an observation with coordinates (*lo*, *la*) when estimate the function at a target point (lo_t, la_t). One approach is to use a simple product kernel:

$$K\left(\left(lo - lo_t\right)/h_1, \left(la - la_t\right)/h_2\right) \tag{17}$$

A more commonly used alternative is to make the weights depend simply on the straight-line distance between each observation and the target point, d_t:

$$K(d_t / h) \tag{18}$$

The kernel weight function in equation (19) draws a circle around the target point to form the weights. Although, equation 18 is slightly more general, there is little difference between the two in practice. With *J* different quantiles, the set of estimated coefficients for explanatory variable *k*, $\hat{\beta}_k$, is an nxJ matrix. With K explanatory variables in addition to the intercept, the *n*x*J* matrix of quantile predictions is

$$\hat{y} = \hat{\beta}_0 + \sum_{k=1}^{K} x_k \hat{\beta}_k \tag{20}$$

These predictions can be used to calculate density functions for predicted values of the dependent variable for arbitrary values of the explanatory variables. Suppose we want to evaluate the model at $x_1 = \delta$. Then the *n*x*J* set of predicted values is simply

$$\hat{y} = \hat{\beta}_0 + \delta x_1 + \sum_{k=2}^{K} x_k \hat{\beta}_k \,. \tag{21}$$

The calculations can be repeated for other values of δ and for other explanatory variables. The results can then be summarized using estimated kernel density functions.

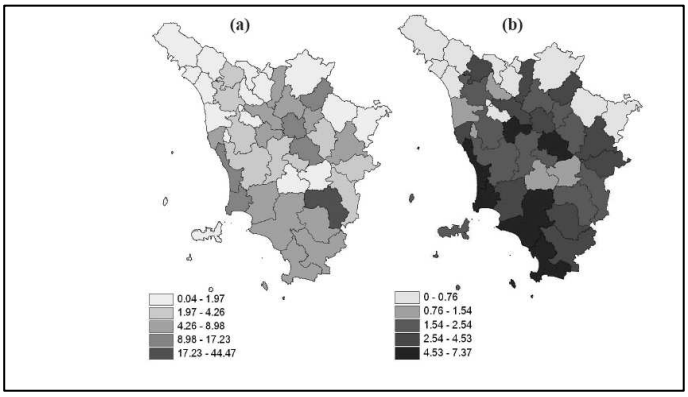
Model is analogous to conditionally parametric (CPAR) local linear regression. The estimation procedure involves estimating separate quantile regressions for various target points, with more weight placed on observations that are close to the target. Unlike a fully nonparametric approach, the CPAR

approach produces coefficient estimates for the explanatory variables. But unlike the spatial AR version of quantile regression, the estimated coefficients vary over space. The CPAR approach is less sensitive to model misspecification than the fully parametric spatial AR approach, and it accounts for local variation in an overall spatial trend. The approach is well suited for quantile analysis in situations where the distribution of the dependent variable is, for example, highly skewed in some locations, tightly clustered in others, while all the time varying smoothly over space. Moreover, the CPAR approach does not require the specification of a large (n *x* n) spatial weight matrix, making it amenable to large data sets.

We present results (Chambers, Pratesi, Salvati, Tzavidis 2005) obtained for the estimation spatial distribution of the mean and median production of olives per farm LES. The data are from Farm Structure Survey (2003). ***Z*** the incidence matrix of dimensions 2508 farm per 42 LESs. The neighborhood structure **W is** defined as follows: spatial weight w_{ij}, is 1 if area shares an edge with *j* and 0 otherwise.

The median map is intensive to the presence a few big farms that raise the medium level of production as a consequence the spatial distribution of the median is more homogenous.

Figure 1 a) Mean b) median production of olives



Source: Chambers R. & all (2005).

6. Conclusions

In summary we can say that the classic paper for quantile regression is Koenker and Bassett (1978). Koenker (2005) presents an extensive examination of the econometric theory related to a wide variety of quantile models. Buschinsky (1998) helped popularize the use of quantile regression analysis on the distribution of wages. The spatial AR version of the quantile model relies on approaches developed by Chernozhukov and Hansen (2006) and Kim and Muller (2004). The approaches have been applied to studies of house prices by Kostov (2009), Liao and Wang (2012) and Zeitz et al. (2008). The studies rely on the IV approach for estimating the spatial AR model. Nonparametric versions of quantile models relies heavily on Koenker work. Splines are also a potential alternative to kernel smoothing; it was done in Koenker and Mizera (2004). The use of nonparametric methods for spatial models has been forced by the invention of new terms by geographers for procedures that have already been used extensively in statistics and economics.

References

Buchinsky M. (1998), Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research, Journal of Human Resources, 33

Chambers R., Pratesi M., Salvati N., Tzavidis N. (2005), Small Area Estimation: spatial information and M- quantile regression to estimate the average production of olive per farm, Wp 279, Dipatimento di Statistica e Matematica Applicata all'Economia, Universita di Pisa

Chernozhukov V., Hansen Ch.. (2006), Instrumental Quantile Regression Inference for Structural and Treatment Effect Models, Journal of Econometrics, 132

Cleveland, W. S. (1994), Coplots, Nonparametric Regression, and Conditionally Parametric Fits. In T.W. Anderson, K.T. Fant, and I. Olkin (Eds.), Multivariate Analysis and its Applications, Hayward: Institute of Mathematical Statistics.

Kim T.-H., Muller Ch.. (2004), Two-Stage Quantile Regression when the First Stage is Based on Quantile Regression, Econometrics Journal, 7

Koenker R., Basset B., (1978), Regression Quantiles, Econometrica, Vol 46

Koenker R., Ng P., (2005), Inequality Constrained Quantile Regression, The Indian Journal of Statistics, Vol. 67

Koenker R., Hallock K. F. (2001), Quantile Regression, Journal of Economic Perspectives, 15

Koenker R., Mizera I. (2004). Penalized Triograms: Total Variation Regularization for Bivariate Smoothing, Journal of the Royal Statistical Society: Series B, 66

Kostov, P. (2009), A Spatial Quantile Regression Hedonic Model of Agricultural Land Prices, Spatial Economic Analysis, 4

Liao, Wen-Chi and Xizhu Wang (2012), "Hedonic House Prices and Spatial Quantile Regression," Journal of Housing Economics, 21, 16-27

Mcmillan. D. P. (2010), Issues in Spatial Data Analysis, Journal of Regional Science, 119-141.

Trzpiot G., (2008), The Implementation of Quantile Regression Methodology in VaR Estimation "Studies and Researches of Faculty of Economics and Management University of Szczecin",

Trzpiot G., (2009 a), Quantile Regression Model versus Factor Model Estimation, in: Financial Investments and Insurances", University of Economics in Wrocław, Vol 60

Trzpiot G., (2009 b), Application weighted VaR in capital allocation, Polish Journal of Environmental Studies, Olsztyn, Vol 18, 5B

Trzpiot G., (2009 c), Estimation methods for quantile regression, Economics Studies 53, Karol Adamiecki University of Economics in Katowice

Trzpiot G., (2010), Quantile Regression Model of Return Rate Relation – Volatility for Some Warsaw Stock Exchange Indexes, (in Polish), Finances, Financial Markets and Insurances. Capital Market, University of Szczecin, Vol 28, 61-76

Trzpiot G. (2011a). Bayesian Quantile Regression, Studia Ekonomiczne, Zeszyty Naukowe nr 65, Uniwersytet Ekonomiczny w Katowicach, 33-44

Trzpiot G. (2011b). Some tests for quantile regression models, Acta Universitatis Lodziensis, Łódź, Folia Economica, 255, 125-135

Zeitz J., Zietz E. N., Sirmans G. S. (2008) Determinants of House Prices: A Quantile Regression Approach, Journal of Real Estate Finance and Economics, 37

Streszczenie

PRZESTRZENNA REGRESJA KWANTYLOWA

W wielu zastosowaniach, podstawowym problemem jest opis i analiza wpływu wektora skorelowanych zmiennych objaśniających X na zmienna objaśnianą Y. W przypadku, gdy obserwacje badanych zmiennych są dodatkowo rozmieszczone przestrzennie, zadanie jest jeszcze trudniejsze, ponieważ mamy dodatkowe zależności, wynikające ze zmienności przestrzennej.

W tej pracy, w miejsce przestrzennej regresji wykorzystującej średnią, rozpatrzymy przestrzenna regresję kwantylową. Regresja kwantylowa zostanie omówiona w przestrzennym kontekście. Głównym celem pracy jest wskazanie na

możliwości powiązania metodologii regresji kwantylowej i ekonometrycznego modelowania przestrzennego. Dodatkowe zasoby informacji o zmienności otrzymujemy badając kwantyle, wychodząc poza tradycyjny opis klasycznej regresji. Estymacja kwantylowa w modelu przestrzennym uwydatnia zależności przestrzenne dla różnych fragmentów rozważanych rozkładów.