A two-step procedure for estimating spatial error quantile regression models

Elena Semerikova, Faculty of Economic Sciences, Higher School of Economics University, Moskov (Russia)
Giuseppe Arbia, Faculty of Economics, Catholic University of the Sacred Heart, Milan (Italy)

Abstract
Conditional quantile regression has been considered in the spatial econometric literature as a robust alternative to the standard linear regression when data are characterized by non-normality. They are also complementary to the conditional mean model in that they provide a complete representation of the conditional distribution of Y given the predictors and not only their conditional mean. So far the literature concentrated on conditional quantile spatial lag specifications which incorporate spatial correlation in the form of a spatially lagged variable included in the list of predictors. This paper aims at exploring the other side of the moon by specifying conditional spatial quantile models in the form of a spatial error correlation and by suggesting feasible estimators of the parameters involved. We propose a procedure analogous to the feasible GLS suggested for the standard linear models, we illustrate its use on real data and we examine its small sample properties through a set of Monte Carlo experiment.

Keywords: *Spatial error quantile model; Efficiency*

1. Introduction

Conditional quantile regression models (Koenker and Bassett, 1978; Koenker, 2005; Hao and Naiman, 2007) provide a robust alternative to the standard linear regression when data are characterized by non-normality features such as skewness, heavy tails and outliers. Since then, the spatial econometric literature has concentrated (with few remarkable exceptions, see Dai and Tian, 2019) on conditional quantile spatial lag specifications (McMillen, 2013; Su and Yang, 2011) which incorporate substantive spatial correlation in the form of a spatially lagged variable included in the list of predictors. However, spatial correlation can often manifest itself also in the form of a spatial lag in the error term which captures the effects of omitted variables, a situation typical of many individual studies where some of the predictors maybe not accessible because they are protected by confidentiality. This paper aims at exploring the other side of the moon of spatial conditional quantile by specifying models in the form of a spatial correlation present in the error term and by suggesting feasible estimators of the parameters.

2. A two-step estimation procedure for a spatial error quantile regression

In order to remove the negative effects on estimators' properties induced by the presence of a positive residual spatial correlation, in the area of quantile regression the spatial econometric literature, as said, has concentrated on the spatial lag specification (McMillen, 2013). Here we consider the spatial error alternative. In the case of a linear regression on the mean value, a spatial error model (SEM) is specified as follows:

$$y = x\beta + u \qquad (4)$$
$$u = \rho W u + \epsilon \qquad (5)$$
$$\epsilon \overset{iid}{\sim} N(0, \sigma_\epsilon^2) \qquad (6)$$

From Equation (5) we have that:

$$(I - \rho W)u = \epsilon \qquad (7)$$

and, hence, from Equation (4) we have:

$$(I - \rho W)y = (I - \rho W)x\beta + \epsilon \qquad (8)$$

where the residuals are now "cleaned" from the spatial correlation and can be assumed to be independently distributed.
Our proposal is to follow the same strategy for a quantile regression adopting the following two-step procedure. In the first step, we postulate the following system of equations (for a related approach see Chernozhukov and Hansen, 2005 and Su and Yang, 2007):

$$Q_y(\tau|x) = x\beta \tag{9}$$

$$u = Q_y(\tau|x) - x\beta \tag{10}$$

$$u = \rho W u + \epsilon \tag{11}$$

$$\epsilon \overset{iid}{\sim} \frac{\chi_2^2 - 2}{2} \tag{12}$$

where in Equation (12), to capture the presence of skewness in the residuals, the error term ϵ is assumed i.i.d. distributed as a (standardized) Chi-square with 2 degrees of freedom. We will refer to this model as the *Spatial Error Quantile Model* (SEQM). In order to implement our estimation strategy, we can estimate the ρ parameter in Equation (11) using a GMM procedure. The GMM approach is particularly useful in this context because, as it is well known, it is a distribution-free procedure particularly efficient when normality cannot be assumed. We assume that we can estimate the parameter ρ using the same moment conditions suggested by Keleijan and Prucha (1978) (see also Arbia, 2014 for further details). Let us refer to the estimator thus obtained as to $\hat{\rho}$. We can then define the transformed variables as follows:

$$y^* = (I - \hat{\rho}W)y \text{ and } x^* = (I - \hat{\rho}W)x \tag{13}$$

and we can derive the optimal parameters' values with a linear programming optimization of the equation:

$$Q_{y*}(\tau|x^*) = x^*\beta \tag{14}$$

3. A Monte Carlo study

The example presented in the previous section shows that the suggested two-step procedure might be effective in removing residual spatial correlation from the quantile regressions thus overcoming the problems of reduced efficiency in the parameters' estimators. In this section, in order to explore the small sample properties of this procedure, we will report the results of a simulation study. In our experiments, we considered a sample of n = 100 data laid on a 10-by-10 regular square lattice grid generated through the following set of equations:

$$y = x\beta + u \tag{15}$$
$$u = \rho W u + \epsilon \tag{16}$$
$$\epsilon \sim \frac{\chi_2^2 - 2}{2} \tag{17}$$

with the vector $\beta \equiv [\beta_1, \beta_2]$ including a constant term, W a rook's case defined weight matrix and all the other symbols previously defined. From Equation (16) we have:

$$u = (I - \rho W)^{-1}\epsilon \tag{18}$$

Hence, from Equation (22) we obtain:

$$y = x\beta + (I - \rho W)^{-1}\epsilon \tag{19}$$

where we generate ϵ as described in Equation (17) and x as an independent N(0,1). We further set $\beta_1 = 10$ and $\beta_2 = 0.5$. A single realization of such a process is reported in Figure 6 which displays the OLS and the median regression.

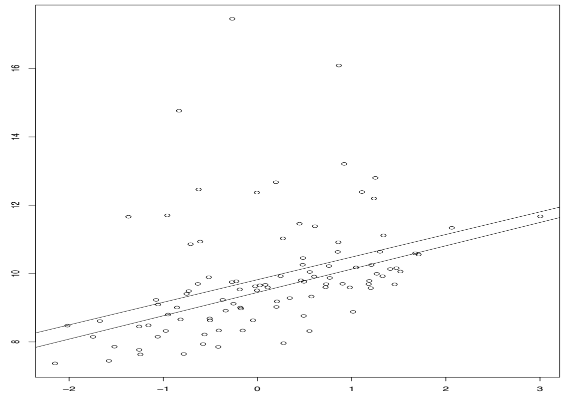

Figure 6: OLS and median regression in one single run of the Monte Carlo simulation.

We then fit to the simulated data the traditional conditional quantile regression and the quantile regression on the transformed variables as shown in Equation (13), using a linear programming optimization with $\hat{\rho}$ estimated with the GMM procedure suggested by Keleijan and Prucha (1978).

Conclusions

This paper aims at introducing a procedure to fit conditional quantile regression models to spatial data using the spatial error specification. While the effects of the suggested procedures still need to be properly investigated, the first results are encouraging showing more efficient estimates of the model's parameters and of the hypothesis testing procedures on residuals' spatial correlation.

Future developments include the possibility of estimating from the data the degrees of freedom of the chi-squared distribution in the case of skewed datasets, and the treatment of other forms of non-normality (kurtosis, multi-modoalities etc.).

References

Arbia. G. (2014) *A primer for spatial econometrics: with applications in R*. Palgrave Mac Millan.

Arbia. G.. Espa. G. and Giuliani. D. (2022) *Spatial microeconometrics*. Routledge.

Dai. X. Li. E. and Tian. M. (2019): Quantile regression for varying coefficient spatial error models. *Communications in Statistics - Theory and Methods*. DOI: 10.1080/03610926.2019.1667396.

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

Koenker R. and Bassett (1978) Quantile regression. *Econometrica.* 46 33-30.

McMillen. D. P. (2013) *Quantile Regression for Spatial Data*. Springer-Verlag

Moran. P. A. P. (1950) Notes on Continuous Stochastic Phenomena. *Biometrika* 37 (1): 17–23.

Murakami and Seya (2020) Spatially filtering unconditional quantile regression: application to a hedonic analysis. *Environmetrics*. 30. 5. https://doi.org/10.1002/env.2556

Su. L. Yang. Z. (2007) Instrumental Variable Quantile Estimation of Spatial Autoregressive Models. *Working Papers* 05-2007. Singapore Management University. School of Economics.