A C T A U N I V E R S I T A T I S L O D Z I E N S I S
FOLIA OECONOMICA 286, 2013

*Grażyna Trzpiot**

BAYESIAN SPATIAL QUANTILE REGRESSION

Abstract. In this paper we present a Bayesian spatial model quantile regression. We develop a spatial quantile regression model that does not assume normality and allows the covariates to affect the entire conditional distribution, rather than just the mean. The conditional distribution is allowed to vary from site-to-site and is smoothed with a spatial prior.

Key words: quantile regression, spatial quantile regression, bayesian spatial model.

I. INTRODUCTION

The aim of this paper is to present a Bayesian spatial model quantile regression. We describe a spatial quantile regression model that does not assume normality and allows the covariates to affect the entire conditional distribution, rather than just the mean. The conditional distribution is allowed to vary from site-to-site and is smoothed with a spatial prior. We knew from the literature quantile regression model (Koenker 1978, 2001). The standard model approach is to estimate the effect of the covariates separately for a few quantile levels by minimizing an objective function. This approach is known due to computational convenience and theoretical properties. This approach performs separate analyses for each quantile level of interest. As a result, the quantile estimates can cross that for a particular combination of covariates the estimated quantile levels are non-increasing, which can cause problems for prediction.

II. QUANTILE REGRESSION

We began from density regression, and let $f(y \mid x)$ be the density of some covariates x. There are several models for the conditional distribution, Bayesian models. Many of these models are infinite mixtures with mixture probabilities that depend on x. Although these models are quite flexible, one drawback is the difficulty in interpreting the effects of each covariate, for example, whether there is a statistically significant time trend in the distribution's upper tail probability.

As a compromise between general Bayesian density regression and the usual

* Prof., Department of Demography and Economics Statistics, University of Economics in Katowice.

additive mean regression, we describe a Bayesian spatial quantile regression model. Quantile regression models the distribution's quantiles as additive functions of the predictors. This additive structure permits inference on the effect of individual covariates on the response's quantiles. The additive structure also permits density regression for high dimensional x.

Let y_i be the dependent variable and $X_i = (X_{i1},\dots,X_{ip})'$ be the covariates. Quantile regression models y_i 's conditional density based on its quantile function (inverse CDF) $q(\tau \mid X_i; s_i)$, defined as $P\{y_i < q(\tau \mid X_i)\} = \tau \in [0; 1]$.

We model $q(\tau \mid X_i;)$ as

$$q(\tau \mid X_i), = X_i'\beta(\tau), \tag{1}$$

where $\beta(\tau) = (\beta_1(\tau),\dots,\beta_p(\tau))'$ are the coefficients for the τ^{th} quantile level.

The quantile regression estimation is a solution of the following optimization problem:

$$\hat{\beta}(\tau_k) = \arg\min_{\beta} \sum_{y_i > X_i'\beta} \tau_k \left| y_i - X_i'\beta \right| + \sum_{y_i < X_i'\beta} (\tau_k - 1)\left| y_i - X_i'\beta \right|. \tag{2}$$

In this method different quantiles are analyzed separately. The distribution of the error term is unspecified. Quantile regression methods and they applications on the financial markets can be found in Trzpiot (6-11) among others.

II. MODEL FOR THE QUANTILE PROCESS

We begin modeling the quantile function by ignoring spatial location and assuming the model with the intercept, for that $X_i = 1$. In this case, the quantile function can be reduced to $q(\tau) = \beta(\tau)$. The process $\beta(\tau)$ must be constructed so that $q(\tau)$ is non-decreasing in τ. We can notice

$$\beta(\tau) = \sum_{m=1}^{M} B_m(\tau)\alpha_m , \tag{3}$$

where M is the number of basis functions, $B_m(\tau)$ is a known basis function of τ, α_m are unknown coefficients that determine the shape of the quantile function. We use Bernstein basis polynomials

$$B_m(\tau) = \binom{M}{m} \tau^m (1-\tau)^{M-m} . \quad (4)$$

An attractive feature of these basis functions is that if $\alpha_m \geq \alpha_{m-1}$ for all $m > 1$, then $\beta(\tau)$, and thus $q(\tau)$, is an increasing function of τ. This reduces the complicated monotonicity constraint to a sequence of constraints $\delta_m = \alpha_m - \alpha_{m-1} \geq 0$, for $m = 2, \ldots, M$.

These constraints are sufficient, but not necessary, to ensure an increasing function. As is typical for semiparametric models (Reich B.J., Fuentes M., 2007), for finite M this model does not span the entire class of continuous monotonic functions. However, as M increases, the Bernstein polynomials basis with these constraints induces a prior with dense support on the space of continuous monotone functions from $[0; 1] \to R$ (Chang et al., 2007).

Since the constraints on $\alpha = (\alpha_1, \ldots, \alpha_M)$ are expressed in terms of the difference between adjacent terms, we reparameterize to $\delta_1 = \alpha_1$ and $\delta_m = \alpha_m - \alpha_{m-1}$ for $m = 2, \ldots, M$. The original basis function coefficients are then $\alpha_m = \sum_{l=1}^{m} \delta_l$.

Following Cai and Dunson (2008), we ensure the quantile constraint by introducing a latent unconstrained variable δ_m^* and taking $\delta_1 = \delta_m^*$ and

$$\delta_m = \begin{cases} \delta_m^*, \delta_m^* \geq 0 \\ 0, \ \delta_m^* < 0 \end{cases} \quad (5)$$

for $m > 1$.

The δ_m^* have independent normal priors $\delta_m^* \sim N(\bar{\delta}_m(\Theta), \sigma^2)$, with unknown hyperparameters Θ. We can use $\bar{\delta}_m(\Theta)$ in a center the quantile process on a parametric distribution $f_0(y \mid \Theta)$, for example, a $N(\mu_0; \sigma^2)$, random variable with $\Theta = (\mu_0; \sigma_0)$.

Letting $q_0(\tau \mid \Theta)$ be the quantile function of $f_0(y \mid \Theta)$, the $\bar{\delta}_m(\Theta)$ are then chosen so that

$$q_0(\tau|\Theta) \approx \sum_{m=1}^{M} B_m(\tau)\bar{\alpha}_m(\Theta) , \quad (6)$$

where

$$\overline{\alpha}_m(\Theta) = \sum_{l=1}^{m} \overline{\delta}_l(\Theta)\,.$$

The $\overline{\delta}_m(\Theta)$ are chosen to correspond to the following ridge regression estimator:

$$(\overline{\delta}_1(\Theta),...,\overline{\delta}_M(\Theta))' = \arg\min_{d} \sum_{k=1}^{K} \left(q_0(\tau|\Theta) - \sum_{m=1}^{M} B_m(\tau_k) \left[\sum_{l=1}^{m} d_l \right] \right)^2 + \lambda \sum_{m=1}^{M} d_m^2, \quad (7)$$

where $d_m > 0$ for $m > 1$, $\{\tau_1, ..., \tau_K\}$ is a dense grid on (0,1). We find that simple parametric quantile curves can often be approximated almost perfectly with fewer than M terms. Therefore, several combinations of d give essentially the same fit, including some undesirable solutions with negative values for elements of $\overline{\delta}$. As $\sigma \to 0$, the quantile functions are increasing shrunk towards the parametric quantile function $q_0(\tau | \Theta)$, and the likelihood is similar to $f_0(y | \Theta)$.

III. MODEL FOR THE QUANTILE PROCESS WITH COVARIATES

Next possibility to model the quantile process is to add covariates. Then the conditional quantile function we can denote as

$$q(\tau|X_i) = X_i'\beta(\tau) = \sum_{j=1}^{p} X_{ij}\beta_j(\tau)\,. \quad (8)$$

Like in previous rules, the quantile curves are modeled using Bernstein basis polynomials

$$\beta_j(\tau) = \sum_{m=1}^{M} B_m(\tau)\alpha_{jm}\,, \quad (9)$$

where α_{jm} are unknown coefficients. The processes $\beta_j(\tau)$ must be constructed so that $q(\tau|X_i)$ is nondecreasing in τ for all X_i. Collecting terms with common basis functions gives

$$X_i'\beta(\tau) = \sum_{m=1}^{M} B_m(\tau)\theta_m(X_i), \tag{10}$$

where $\theta_m(X_i) = \sum_{j=1}^{p} X_{ij}\alpha_{jm}$. Therefore, if $\theta_m(X_i) \geq \theta_{m-1}(X_i)$ for all $m > 1$, then $X_i'\beta(\tau)$ and thus $q(\tau|X_i)$ is is an increasing function of τ.

To specify our prior for the α_{jm} to ensure monotonicity, we assume that $X_{i1} = 1$ for the intercept and the remaining covariates are suitably scaled so that $X_{ij} \in [0; 1]$ for $j > 1$. Since the constraints are written in terms of the difference between adjacent terms, we reparameterize to $\delta_{j1} = \alpha_{j1}$ and $\delta_{jm} = \alpha_{jm} - \alpha_{jm-1}$ for $m = 2, \ldots, M$. We ensure the quantile constraint by introducing latent unconstrained variable $\delta^*_{jm} \sim \mathrm{N}(\bar{\delta}_{jm}(\Theta); \sigma_j^2)$ and taking

$$\delta_{jm} = \begin{cases} \delta^*_{jm}, & \delta^*_{1m} + \sum_{j=2}^{p} I(\delta^*_{jm} < 0)\delta^*_{jm} \geq 0 \\ 0, & otherwise \end{cases} \tag{11}$$

for all $j = 1, \ldots, p$ and $m = 1, \ldots, M$. Recalling $X_{i1} = 1$ and $X_{ij} \in [0; 1]$ for $j = 2, \ldots, p$, and thus $X_{ij}\delta_{jm} \geq X_{ij}I(\delta_{jm} < 0)\,\delta_{jm} \geq I(\delta_{jm} < 0)\,\delta_{jm}$ for $j > 1$,

$$\theta_m(X_i) - \theta_{m-1}(X_i) = \sum_{j=1}^{p} X_{ij}\alpha_{jm} \geq \delta_{1m} + \sum_{j=2}^{p} X_{ij}I(\delta_{jm} < 0)\delta_{jm} \geq \delta_{1m} + \sum_{j=2}^{p} I(\delta_{jm} < 0)\delta_{jm} \geq 0 \tag{12}$$

for all X_i, giving a valid quantile process. As in previous section we center the intercept curve on a parametric quantile function $q_0(\Theta)$.

The remaining coefficients have $\delta^*_{jm}(\Theta) = 0$ for $j > 1$. Although this model is quite flexible, we have assumed that the quantile process is a linear function of the covariates, simplifying interpretation. In some applications the linear quantile relationship may be overly-restrictive. In this case, transformations of the original predictors such as interactions or basis functions can be added to give a more flexible model.

The linear relationship between the predictors and the response is not invariant to transformations of the response. To alleviate some sensitivity to transformations, it may be possible to develop a nonlinear model for $q(\tau \mid X_i)$, so that $q(\tau \mid X_i)$ and $T(q(\tau \mid X_i))$ span the same class of functions (and therefore response distributions) for a class of transformations T.

IV. MODEL FOR THE SPATIAL QUANTILE PROCESS WITH COVARIATES

Let y_i be the observed data time and space location $(t; s)_i$, and denote the time and spatial location of the i^{th} observation as t_i and s_i, respectively.

Our goal is in estimating the conditional density of y_i as a function of s_i and covariates $X_i = (X_{i1}, \ldots, X_{ip})'$,where $X_{i1} = 1$ for the intercept.

In particular, we would like to study the conditions that lead to extreme value depending on time.

Given our interest in extreme events and return levels, we model y_i's conditional density based on its quantile (inverse CDF) function $q(\tau \mid X_i; s_i)$, which is defined so that $P\{y_i < q(\tau \mid X_i; s_i)\} = \tau$, $\tau \in [0; 1]$.

We try to model $q(\tau \mid X_i; s_i)$, as

$$q(\tau \mid X_i; s_i), = X_i'\beta(\tau, s_i), \tag{13}$$

where $\beta(\tau, s_i) = (\beta_1(\tau, s_i), \ldots, \beta_p(\tau, s_i))$ are the spatially dependent coefficients for the τ^{th} quantile level. Directly modeling the quantile function makes explicit the effect of each covariate on the probability of an extreme value.

We know that the quantile process is different at each spatial location,

$$\beta_j(\tau, s) = \sum_{m=1}^{M} B_m(\tau)\alpha_{jm}(s), \tag{14}$$

where $\alpha_{jm}(s)$ are spatially-varying basis function coefficients. We enforce the monotonicity constraint at each spatial location by introducing latent Gaussian parameters $\delta^*_{jm}(s)$. The latent parameters relate to the basis function coefficients as $\alpha_{jm}(s) = \sum_{l=1}^{m} \delta_{jl}(s)$ and

$$\delta_{jm}(s) = \begin{cases} \delta^*_{jm}(s),\ \delta^*_{1m}(s) + \sum_{j=2}^{p} I(\delta^*_{jm}(s) < 0)\delta^*_{jm}(s) \geq 0 \\ 0, \quad otherwise \end{cases} \tag{15}$$

for all $j = 1, \ldots, p$ and $m = 1, \ldots, M$.

To encourage the conditional density functions to vary smoothly across space we model the δ^*_{jm} (s) as spatial processes. The δ^*_{jm} (s) are independent (over j and m) Gaussian spatial processes with mean $E(\delta^*_{jm}(s)) = \delta^*_{jm}(\Theta)$ and exponential spatial covariance

$$\text{Cov}(\delta^*_{jm}(s);\ \delta^*_{jm}(s')) = \sigma^2_j \exp(-\|s - s'\| / \rho_j), \tag{16}$$

where σ^2_j is the variance of δ^*_{jm} (s) and ρ_j determines the range of the spatial correlation function.

We have different models as special cases of the last model.

Setting $\beta_j(\tau;\ s) \equiv \beta_j$ for all τ, s, and $j > 1$ gives the usual linear regression model with location shifted by $\sum_{j=2}^{p} X_{ij}\beta_j$ and residual density determined by $\beta_1(\tau, s)$.

Another possibility is setting $\beta_j(\tau;\ s) \equiv \beta$ (s) for all τ and $j > 1$ gives the spatially dependent coefficients model where the effect of X_j on the mean varies across space via the spatial process β_j(s).

Allowing $\beta_j(\tau;\ s)$ to vary with s and τ relaxes the assumption that the covariates simply affect the mean response, and gives a density regression model where the covariates are allowed to affect the shape of the response distribution. In particular, the covariates can have different effects on the center ($\tau = 0{:}5$) and tails ($\tau \approx 0$ and $\tau \approx 1$) of the density.

V. ESTIMATION FOR BAYESIAN SPATIAL QUANTILE REGRESSION

Spatial quantile regression model can be implemented efficiently for moderately sized data sets. To approximate this model, we use a two stage approach. We first perform separate quantile regression at each site for a grid of quantile levels to obtain estimates of the quantile process and their asymptotic covariance. In the second stage, we analyze these initial estimates using the Bayesian spatial model for the quantile process.

The usual quantile regression estimate (Koenker, 2005) for quantile level τ_k and spatial location s is $\hat{\beta}(\tau_k, s_i) = (\hat{\beta}_1(\tau_k, s_i), \ldots, \hat{\beta}_p(\tau_k, s_i))$

$$\hat{\beta}(\tau_k, s) = \arg\min_{\beta} \sum_{s_i = s, y_i > X_i'\beta} \tau_k \left| y_i - X_i'\beta \right| + \sum_{s_i = s, y_i < X_i'\beta} (\tau_k - 1) \left| y_i - X_i'\beta \right|. \tag{17}$$

This estimate is easily obtained from the quantreg package in R and is consistent for the true quantile function and has asymptotic covariance (Koenker, 2005) where n_s is the number of observations at site s,

$$\operatorname{cov}\left[\sqrt{n_s}(\hat{\beta}_1(\tau_k, s_i), \ldots, \hat{\beta}_p(\tau_k, s_i)), \sqrt{n_s}(\hat{\beta}_1(\tau_l, s_i), \ldots, \hat{\beta}_p(\tau_l, s_i))\right] = H(\tau_k)^{-1} J(\tau_k, \tau_l) H(\tau_l)^{-1}, \tag{18}$$

where n_s is the number of observations at site s, $H(\tau) = \lim_{n_s \to \infty} \frac{1}{n_s} \sum_{i=1}^{n_s} X_i X_i' f_i(X_i'\hat{\beta}(\tau))$, $J(\tau_k, \tau_l) = [\tau_k \wedge \tau_l - \tau_k \tau_l] n_s^{-1} \sum X_i X_i'$.

Although consistent as the number of observations at a given site goes to *1*, these estimates are not smooth over space or quantile level, and do not ensure a non-crossing quantile function for all X. Therefore, we smooth these initial estimates using the spatial model for the quantile process

$$\hat{\beta}(s_i) = (\hat{\beta}_1(\tau_1, s_i), \ldots, \hat{\beta}_1(\tau_K, s_i), \hat{\beta}_2(\tau_1, s_i), \ldots, \hat{\beta}_p(\tau_K, s_i)).$$
$$\text{and } \operatorname{cov}(\hat{\beta}(s_i)) = \Sigma_i \tag{19}$$

We fit the mode

$$\hat{\beta}(s_i) \sim \mathrm{N}(\beta(s_i), \Sigma_i), \tag{20}$$

where the elements of $\beta(s_i) = (\beta_1(\tau_1, s_i), \ldots, \beta_1(\tau_K, s_i), \beta_2(\tau_1, s_i), \ldots, \beta_p(\tau_K, s_i))$' are functions of Bernstein basis polynomials.

This approximation gives reduction in computational time because the dimension of the response is reduced from the number of observations at each site to the number of quantile levels in the approximation, and the posteriors for the parameters that define are fully-conjugate allowing for Gibbs updates and rapid convergence.

REFERENCES

Cai B, Dunson D. B. (2007), Bayesian multivariate isotonic regression splines: Applications to carcinogenicity studies. *Journal of the American Statistical Association*, 102: 1158-1171.

Chang I, Chien L, Hsiung CA, Wen C, Wu Y (2007). Shape restricted regression with random Bernstein polynomials. *IMS Lecture Notes {Monograph Series*, 54: 187-202.

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

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

Reich B.J., Fuentes M. (2007), A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields, *Annals of Applied Statistics*, 1: 249-264.

Trzpiot G. (2009), Quantile Regression Model versus Factor Model Estimation, *Financial Investments and Insurances – World Trends and Polish Market*, University of Economics in Wrocław 60. 469-479.

Trzpiot G. (2009), Application weighted VaR in capital allocation, *Polish Journal of Environmental Studies*. Olsztyn. 18. 5B. 203-208.

Trzpiot G. (2009), Estimation methods for quantile regression, *Economics Studies 53*, University of Economics in Katowice. 81-90.

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. 28. 61-76.

Trzpiot G. (2011) Bayesian Quantile Regression, *Economics Studies 65*, University of Economics in Katowice, 33-44.

Trzpiot G. (2012) Spatial Quantile Regression, *Comparative Economic Research. Central and Eastern Europe*, Vol. 15, No 4, 265-279, Wydawnictwo Uniwersytetu Łódzkiego.

Grażyna Trzpiot

BAYESOWSKA 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.

Klasyczne podejście stosowane do takich problemów wykorzystuje założenie o skończonej wartości oczekiwanej zmiennych Y, wówczas przestrzenna funkcja regresji jest dobrze określona i dostarcza informacji o zależności zmiennej Y od zmiennych X. W tej pracy, w miejsce przestrzenna funkcja regresji wykorzystującej średnią, rozpatrzymy przestrzenna regresję kwantylową. Regresja kwantylowa zostanie omówiona w przestrzennym kontekście. Semiparametryczny model bayesowski i jego estymacja jest głównym celem tej pracy. 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.