IOP Conf. Series: Journal of Physics: Conf. Series **803** (2017) 012150 doi:10.1088/1742-6596/803/1/012150

Finite-sample and asymptotic sign-based tests for parameters of non-linear quantile regression with Markov noise

M A Sirenko[1], P F Tarasenko[1] and M I Pushkarev[2]

[1] National Research Tomsk State University, 36, Lenin Ave., Tomsk, 634050, Russia
[2] National Research Tomsk Polytechnic University, 30, Lenin Ave., Tomsk, 634050, Russia
E-mail: sma@mail.tsu.ru

Abstract. One of the most noticeable features of sign-based statistical procedures is an opportunity to build an exact test for simple hypothesis testing of parameters in a regression model. In this article, we expanded a sing-based approach to the nonlinear case with dependent noise. The examined model is a multi-quantile regression, which makes it possible to test hypothesis not only of regression parameters, but of noise parameters as well.

1. Introduction

Sign-based statistical procedures [1-4] are known to be more robust for outliers than the least squares and to have a possibility to control a precise significance level for finite samples when testing a simple hypothesis. In this paper, the sign-based approach is extended to the case of a non-linear model with dependent noise. Thus, the model of a multi-quantile regression is considered [5, 6], which allows one to test hypotheses both of the parameters of the regression function, and the parameters of stationary Markov noise ε_t.

According to the sign-based approach, the residuals are substituted with indicators of their belonging to interquantile intervals $\bar{\boldsymbol{s}} = (s_1, \ldots, s_n)$, where s_t takes a finite number of values. The unknown parameters in this scheme are $\boldsymbol{\upsilon} = (\boldsymbol{\theta}', \boldsymbol{\mu}', \boldsymbol{Q}')'$, where vector $\boldsymbol{Q}$ contains linearly independent r-dimensional joint probabilities of the states of indicators $\{s_t\}$ generated by the process. Since the problem is considered in a nonparametric setting, each fixed values of parameters $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ correspond to a class of finite-dimensional distributions of initial process ε_t. However, we can show that for any continuous parameterization of finite-dimensional distributions, all the derivatives of the likelihood of indicators $\mathrm{P}(\bar{\boldsymbol{s}} \mid \boldsymbol{\upsilon})$ are expressed in the same way. In the problem of testing simple hypothesis $H_0 : \boldsymbol{\upsilon} = \boldsymbol{\upsilon}_0$, it gives the opportunity to build a test based on the principle of the maximal likelihood ratio.

In this paper, we consider the problem of calculating the critical values to provide the desired significance level with any accuracy for finite samples, as well as the critical values based on the asymptotic distribution of the test statistic.

The obtained tests can be used as a basis for estimating parameters $\boldsymbol{\upsilon}$ by the principle of maximal p-values [7], as well as for the development of tests for the linear hypothesis.

2. Problem statement

IOP Conf. Series: Journal of Physics: Conf. Series **803** (2017) 012150 doi:10.1088/1742-6596/803/1/012150

Let us consider a non-linear regression model with dependent noise:

$$y_t = g_t(\boldsymbol{\theta}) + \varepsilon_t,\ t = \overline{1,n}. \tag{1}$$

In this model, $g_t(\boldsymbol{\theta})$ is a continuously differentiable function of parameters $\boldsymbol{\theta} = (\theta_1,...,\theta_T)' \in \mathrm{R}^T$. Noise ε_t is an (r-1)-order stationary Markov process. One-dimensional distribution functions $\mathrm{P}\{\varepsilon_t < x\}$ are unknown and not necessarily equal for different t, but they have several coinciding quantiles with the same level.

Let a finite set of intervals $C_1(\boldsymbol{\mu}),\ldots,C_K(\boldsymbol{\mu})$ be a partition of R^1, $\mathrm{P}\{\varepsilon_t \in C_k(\boldsymbol{\mu})\} = p_k$, $k = \overline{1,K}$, probabilities p_k are known. Parameters $\boldsymbol{\mu}$ determine the width of the interquantile range and specify the scale of one-dimensional noise distribution. Boundaries of intervals $C_k(\boldsymbol{\mu}) = \langle c_{k-1}(\boldsymbol{\mu}),\ c_k(\boldsymbol{\mu})\rangle$ (here, an angle bracket means either open or closed) depend linearly on unknown parameters $\boldsymbol{\mu}$ and $c_k(\boldsymbol{\mu}) = a_k + \boldsymbol{d}_k'\boldsymbol{\mu}$, $k = \overline{1,K-1}$, $c_0(\boldsymbol{\mu}) = -\infty$, $c_K(\boldsymbol{\mu}) = +\infty$. Under these circumstances, a_k and $\boldsymbol{d}_k$ are fixed and valid parameters forming a set of $\left\{\boldsymbol{\mu} : (\boldsymbol{d}_k - \boldsymbol{d}_{k-1})'\boldsymbol{\mu} + (a_k - a_{k-1}) > 0,\ k = \overline{2,K-1}\right\}$.

On this occasion, the most interesting cases are the simplest ones: a symmetric two-quantile and three-quantile regression, when there is only one parameter $\boldsymbol{\mu}$ and it is equal to the half of the interquantile range. For two-quantile regression K=3, $c_1(\boldsymbol{\mu}) = -\mu$, $c_2(\boldsymbol{\mu}) = \mu$, p_1=p_3=p, p_2=1–2p. For three-quantile regression K=4, $c_1(\boldsymbol{\mu}) = -\mu$, $c_2(\boldsymbol{\mu}) = 0$, $c_3(\boldsymbol{\mu}) = \mu$, p_1=p_4=p, p_2=p_3=(1–2p)/2. Moreover, the one-quantile regression also fits the same model. In this case, parameters $\boldsymbol{\mu}$ are absent, K=2, and $c_1(\boldsymbol{\mu})$=0, p_1=p, p_2=1–p. In all these 3 models, probability p is given.

Let us introduce the following notation for joint probabilities:

$$\begin{gathered}\mathrm{P}^{(l)}(k_1,...,k_l) = \mathrm{P}\left\{\varepsilon_{t-l+1} \in C_{k_1}(\boldsymbol{\mu}),...,\varepsilon_t \in C_{k_l}(\boldsymbol{\mu})\right\}, l = \overline{1,r},\ k_1,...,k_r = \overline{1,K},\\ \tilde{\mathrm{P}}^{(l)} = \left\{\mathrm{P}^{(l)}(k_1,...,k_l) : k_1,...,k_l = \overline{1,K}\right\},\ l = \overline{1,r}.\end{gathered} \tag{2}$$

In particular, $\tilde{\mathrm{P}}^{(1)} = \{p_1,...,p_K\} = \{\mathrm{P}^{(1)}(1),...,\mathrm{P}^{(1)}(K)\}$. Here parameters $\boldsymbol{\theta}$, $\boldsymbol{\mu}$ and $\tilde{\mathrm{P}}^{(r)}$ are unknown, but $\tilde{\mathrm{P}}^{(r)}$ has linearly dependent probabilities. Let $\boldsymbol{Q}$ be a vector, formed by some set of linear independent probabilities from $\tilde{\mathrm{P}}^{(r)}$.

Let us consider structural transformation $j_l(i_1,...,i_l) = 1 + \sum_{j=1}^{l}(i_j - 1)K^{l-j}$, $l = \overline{1,r}$, which defines the correspondence between the set of r-dimensional probabilities $\tilde{\mathrm{P}}^{(l)}$ and one-dimensional vector $\mathrm{P}^{(l)}$ according to the following rule: $[\mathrm{P}^{(l)}]_{j_l(i_1,...,i_l)} = \mathrm{P}^{(l)}(i_1,...,i_l)$. Here and further, $[A]_j$ means the j-th row of a matrix or the j-th element of a vector. With structural matrix G and vector $\boldsymbol{D}$, the transition from independent probabilities $\boldsymbol{Q}$ to $\mathrm{P}^{(r)} = \boldsymbol{D} + G\boldsymbol{Q}$ can be made. This transition could take into consideration not only a normalization requirement, given one-dimensional probabilities, stationary condition, but also finite-dimensional distribution symmetry. In addition, further we will use structural matrices F_i, $i = \overline{1,r-1}$, which provide the transitions to lower level probabilities through $\mathrm{P}^{(i)} = F_i\mathrm{P}^{(i+1)}$.

Within this framework, let us denote true parameters by $\boldsymbol{\upsilon} = (\boldsymbol{\theta}',\boldsymbol{\mu}',\boldsymbol{Q}')'$ and hypothetical parameters by $\boldsymbol{\upsilon}_0 = (\boldsymbol{\theta}_0',\boldsymbol{\mu}_0',\boldsymbol{Q}_0')'$. Thereafter, we can formulate the problem of testing simple hypothesis H_0 about the parameters of models (1) and (2) against composite alternative H_1:

$$H_0 : \boldsymbol{\upsilon} = \boldsymbol{\upsilon}_0,\ H_1 : \upsilon \neq \upsilon_0 \tag{3}$$

For the construction of statistical procedures, we will use indicators $\overline{s} = (s_1,...,s_n)$. These indicators correspond to the numbers of the intervals, in which residuals $y_t - g_t(\boldsymbol{\theta}_0)$ fall, or $s_t = s_t(\boldsymbol{\theta}_0,\boldsymbol{\mu}_0) = s(y_t - g_t(\boldsymbol{\theta}_0),\boldsymbol{\mu}_0)$, where $s(\boldsymbol{u},\boldsymbol{\mu}_0) = k$ for $\boldsymbol{u} \in C_k(\boldsymbol{\mu}_0)$.

The statistics for hypothesis (3) testing will be built on the principle of the maximum likelihood ratio. It means that the critical region will contain only such parameters that provide the highest value of the gradient norm in the hypothetical point:

$$\nabla_{\boldsymbol{\upsilon}} \mathrm{L}(\bar{s} \mid \boldsymbol{\upsilon}, \boldsymbol{\upsilon}_0)\big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0} = \left.\frac{\nabla_{\boldsymbol{\upsilon}} \mathrm{P}(\bar{s}|\boldsymbol{\upsilon})}{\mathrm{P}(\bar{s}|\boldsymbol{\upsilon}_0)}\right|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0} \tag{4}$$

Here, $L(\bar{s}|\boldsymbol{\upsilon},\boldsymbol{\upsilon}_0) = \mathrm{P}(\bar{s}|\boldsymbol{\upsilon}) / \mathrm{P}(\bar{s}|\boldsymbol{\upsilon}_0)$ is the ratio function of indicator-based likelihood.

3. The gradient of indicators likelihood

In searching for the right side of (4), we can use the following expression to get Taylor expansion $\mathrm{P}(\bar{s} \mid \boldsymbol{\upsilon}) = \mathrm{P}(\bar{s} \mid \boldsymbol{\upsilon}_0) + \nabla'_{\boldsymbol{\upsilon}}\mathrm{P}(\bar{s} \mid \boldsymbol{\upsilon}_0)(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0) + \mathrm{o}(\|\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0\|)$:

$$\mathrm{P}(s_1,\ldots,s_n|\boldsymbol{\upsilon}) = \mathrm{P}(s_1|\boldsymbol{\upsilon})\mathrm{P}(s_2|s_1,\boldsymbol{\upsilon})\cdots\mathrm{P}(s_{r-1}|s_1,s_2,\ldots,s_{r-2},\boldsymbol{\upsilon})\prod_{t=r}^{n}\mathrm{P}(s_t|s_{t-r+1},\ldots,s_{t-1},\boldsymbol{\upsilon}). \tag{5}$$

For this expansion, the continuously differentiable parametrization of finite-dimensional distributions by $\boldsymbol{\mu}$ and $\boldsymbol{Q}$ is needed. For an arbitrary parameterization, we have:

$$\begin{gathered}
\mathrm{P}(s_1|\boldsymbol{\upsilon}) = \mathrm{P}(s_1|\boldsymbol{\upsilon}_0) + \nabla'_{\boldsymbol{\upsilon}}\mathrm{P}(s_1|\boldsymbol{\upsilon}_0)(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0) + \mathrm{o}(\|\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0\|),\\
\mathrm{P}(s_2|s_1,\boldsymbol{\upsilon}) = \mathrm{P}(s_2|s_1,\boldsymbol{\upsilon}_0) + \nabla'_{\boldsymbol{\upsilon}}\mathrm{P}(s_2|s_1,\boldsymbol{\upsilon}_0)(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0) + \mathrm{o}(\|\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0\|),\\
. \qquad \ldots \qquad .\\
\mathrm{P}(s_t|s_{t-r+1},\ldots,s_{t-1},\boldsymbol{\upsilon}) = \mathrm{P}(s_t|s_{t-r+1},\ldots,s_{t-1},\boldsymbol{\upsilon}_0) + \nabla'_{\boldsymbol{\upsilon}}\mathrm{P}(s_t|s_{t-r+1},\ldots,s_{t-1},\boldsymbol{\upsilon}_0)\times\\
\times(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0) + \mathrm{o}(\|\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0\|),\ t=\overline{r,n}.
\end{gathered} \tag{6}$$

Substituting (6) in (5) and after regrouping of terms, we have:

$$\frac{\mathrm{P}(\bar{s}|\boldsymbol{\upsilon})}{\mathrm{P}(\bar{s}|\boldsymbol{\upsilon}_0)} = 1 + \left(\frac{\nabla'_{\boldsymbol{\upsilon}}\mathrm{P}(s_1|\boldsymbol{\upsilon}_0)}{\mathrm{P}(s_1|\boldsymbol{\upsilon}_0)} + \frac{\nabla'_{\boldsymbol{\upsilon}}\mathrm{P}(s_2|s_1,\boldsymbol{\upsilon}_0)}{\mathrm{P}(s_2|s_1,\boldsymbol{\upsilon}_0)} + \ldots + \frac{\nabla'_{\boldsymbol{\upsilon}}\mathrm{P}(s_{r-1}|s_1,\ldots,s_{r-2},\boldsymbol{\upsilon}_0)}{\mathrm{P}(s_{r-1}|s_1,\ldots,s_{r-2},\boldsymbol{\upsilon}_0)}\right.$$
$$\left. + \sum_{t=r}^{n}\frac{\nabla'_{\boldsymbol{\upsilon}}\mathrm{P}(s_t|s_{t-r+1},\ldots,s_{t-1},\boldsymbol{\upsilon}_0)}{\mathrm{P}(s_t|s_{t-r+1},\ldots,s_{t-1},\boldsymbol{\upsilon}_0)}\right)\times(\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0) + \mathrm{o}(|\boldsymbol{\upsilon}-\boldsymbol{\upsilon}_0\|).$$

From this, we can obtain the following expression:

$$\nabla_{\boldsymbol{\upsilon}}\mathrm{L}(\bar{s}|\boldsymbol{\upsilon},\boldsymbol{\upsilon}_0)\big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0} = \frac{\nabla_{\boldsymbol{\upsilon}}\mathrm{P}(s_1|\boldsymbol{\upsilon}_0)}{\mathrm{P}(s_1|\boldsymbol{\upsilon}_0)} + \frac{\nabla_{\boldsymbol{\upsilon}}\mathrm{P}(s_2|s_1,\boldsymbol{\upsilon}_0)}{\mathrm{P}(s_2|s_1,\boldsymbol{\upsilon}_0)} + \ldots + \frac{\nabla_{\boldsymbol{\upsilon}}\mathrm{P}(s_{r-1}|s_{r-2},\ldots,s_1,\boldsymbol{\upsilon}_0)}{\mathrm{P}(s_{r-1}|s_{r-2},\ldots,s_1,\boldsymbol{\upsilon}_0)} + \sum_{t=r}^{n}\frac{\nabla_{\boldsymbol{\upsilon}}\mathrm{P}(s_t|s_1,\ldots,s_{t-1},\boldsymbol{\upsilon}_0)}{\mathrm{P}(s_t|s_1,\ldots,s_{t-1},\boldsymbol{\upsilon}_0)}.$$

The last expression can be transformed in such a way that the gradient of the likelihood ratio will be expressed in terms of *r* and (*r*-1)-order joint probabilities:

$$\nabla_{\boldsymbol{\upsilon}}\mathrm{L}(\bar{s} \mid \boldsymbol{\upsilon},\boldsymbol{\upsilon}_0)\big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0} = \sum_{t=r}^{n}\frac{\nabla_{\boldsymbol{\upsilon}}\mathrm{P}(s_{t-r+1},\ldots,s_t|\boldsymbol{\upsilon}_0)}{\mathrm{P}(s_{t-r+1},\ldots,s_t|\boldsymbol{\upsilon}_0)} - \sum_{t=r+1}^{n}\frac{\nabla_{\boldsymbol{\upsilon}}\mathrm{P}(s_{t-r+1},\ldots,s_{t-1}|\boldsymbol{\upsilon}_0)}{\mathrm{P}(s_{t-r+1},\ldots,s_{t-1}|\boldsymbol{\upsilon}_0)}. \tag{7}$$

Now let us turn to the problem of determining gradients $\nabla_{\boldsymbol{\upsilon}}\mathrm{P}(s_{t-r+1},\ldots,s_t|\boldsymbol{\upsilon})\big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0}$ and $\nabla_{\boldsymbol{\upsilon}}\mathrm{P}(s_{t-r+1},\ldots,s_{t-1}|\boldsymbol{\upsilon})\big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0}$. This task requires obtaining derivatives of each parameter in $\boldsymbol{\upsilon}$. Despite the fact that given values $\boldsymbol{\upsilon}$ correspond to a set of distributions on $\mathrm{P}(\bar{s}|\boldsymbol{\upsilon})$, the following theorem shows that for an arbitrary continuous differentiable parametrization of distribution $\mathrm{P}(\bar{s}|\boldsymbol{\upsilon})$ by parameters $\boldsymbol{\mu}$ and $\boldsymbol{Q}$, the gradients expression does not depend on the parametrization method. Therefore, they can be used for further assessments concerning local changes of the likelihood function.

Theorem 1. Let *r*-dimensional continuous distribution density $f_t\left(x_1,...,x_r\right)$ of random variables $\varepsilon_{t-r+1},...,\varepsilon_t$ exist, where $t=\overline{r,n}$. Then, for any continuous differentiable parameterization of distribution $\mathrm{P}\left(\overline{s}\middle|\boldsymbol{\upsilon}\right)$ by parameters $\boldsymbol{\mu}$ and $\boldsymbol{Q}$, the following types of expressions hold:

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}\mathrm{P}(s_{t-r+1},...,s_t\left|\boldsymbol{\upsilon}\right.)\Big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0} &= \sum_{i=1}^{r}\Big[\mathrm{P}_{r,i}(C_{s_{t-r+1}},...,C_{s_t}\left|c_{s_{t-r+i}-1}\right.)f_{t-r+i}(c_{s_{t-r+i}-1}) \\ &-\mathrm{P}_{r,i}(C_{s_{t-r+1}},...,C_{s_t}\left|c_{s_{t-r+i}}\right.)f_{t-r+i}(c_{s_{t-r+i}})\Big]\cdot\nabla_{\boldsymbol{\theta}}g_{t-r+i}(\boldsymbol{\theta}_0).\end{aligned} \tag{8}$$

$$\begin{aligned}\nabla_{\boldsymbol{\theta}}\mathrm{P}(s_{t-r+1},...,s_{t-1}\left|\boldsymbol{\upsilon}\right.)\Big|_{\boldsymbol{\upsilon}=|\boldsymbol{\upsilon}_0} &= \sum_{i=1}^{r-1}\Big[\mathrm{P}_{r-1,i}(C_{s_{t-r+1}},...,C_{s_{t-1}}\left|c_{s_{t-r+i}-1}\right.)f_{t-r+i}(c_{s_{t-r+i}-1}) \\ &-\mathrm{P}_{r-1,i}(C_{s_{t-r+1}},...,C_{s_{t-1}}\left|c_{s_{t-r+i}}\right.)f_{t-r+i}(c_{s_{t-r+i}})\Big]\cdot\nabla_{\boldsymbol{\theta}}g_{t-r+i}(\boldsymbol{\theta}_0).\end{aligned} \tag{9}$$

$$\begin{aligned}\nabla_{\boldsymbol{\mu}}\mathrm{P}(s_{t-r+1}(\boldsymbol{\theta}_0,\boldsymbol{\mu}_0),...,s_t(\boldsymbol{\theta}_0,\boldsymbol{\mu}_0)\left|\boldsymbol{\upsilon}\right.)\Big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0} &= \sum_{i=1}^{r}\Big[\mathrm{P}_{r,i}(C_{s_{t-r+1}},...,C_{s_t}\left|c_{s_{t-r+1}-1}\right.)f_{t-r+i}(c_{s_{t-r+i}-1})\boldsymbol{d}_{s_{t-r+i}-1} \\ &-\mathrm{P}_{r,i}(C_{s_{t-r+1}},...,C_{s_t}\left|c_{s_{t-r+1}}\right.)f_{t-r+i}(c_{s_{t-r+i}})\boldsymbol{d}_{s_{t-r+i}}\Big].\end{aligned} \tag{10}$$

$$\begin{aligned}\nabla_{\boldsymbol{\mu}}\mathrm{P}(s_{t-r+1}(\boldsymbol{\theta}_0,\boldsymbol{\mu}_0),...,s_{t-1}(\boldsymbol{\theta}_0,\boldsymbol{\mu}_0)\left|\boldsymbol{\upsilon}\right.)\Big|_{\upsilon=\upsilon_0} &= \sum_{i=1}^{r-1}\Big[\mathrm{P}_{r-1,i}(C_{s_{t-r+1}},...,C_{s_{t-1}}\left|c_{s_{t-r+1}-1}\right.)f_{t-r+i}(c_{s_{t-r+i}-1})d_{s_{t-r+i}-1} \\ &-\mathrm{P}_{r-1,i}(C_{s_{t-r+1}},...,C_{s_{t-1}}\left|c_{s_{t-r+1}}\right.)f_{t-r+i}(c_{s_{t-r+i}})d_{s_{t-r+i}}\Big].\end{aligned} \tag{11}$$

$$\nabla_{\mathrm{Q}}\mathrm{P}\left(s_{t-r+1}\left(\theta_0,\mu_0\right),...,s_t\left(\theta_0,\mu_0\right)\middle|\upsilon\right)\Big|_{\upsilon=\upsilon_0} = [\mathrm{G}]'_{j_r(s_{t-r+1},...,s_t)}, \tag{12}$$

$$\nabla_{\boldsymbol{Q}}\mathrm{P}\left(s_{t-r+1}\left(\boldsymbol{\theta}_0,\boldsymbol{\mu}_0\right),...,s_{t-1}\left(\boldsymbol{\theta}_0,\boldsymbol{\mu}_0\right)\middle|\boldsymbol{\upsilon}\right)\Big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0} = [F_{r-1}G]'_{j_{r-1}(s_{t-r+1},...,s_{t-1})}. \tag{13}$$

Here, $f_t(\cdot)$ is a density function of ε_t, $\mathrm{P}_{r,i}\left(C_{s_{t-r+1}},...,C_{s_t}\middle|c_k\right)$ is a transitional probability of vector's fall $\left\|\varepsilon_s: s=\overline{t-r+1,t}; s\neq t-r+i\right\|$ into the parallelepiped, formed by intervals $\left\{C_{s_j}: j=\overline{t-r+1,t}; j\neq t-r+i\right\}$ under the following condition: $\varepsilon_{t-r+i}=c_k$.

4. Sign-based tests

Formally, the required test for hypothesis (3) has the following form:

$$\left\|\nabla_{\boldsymbol{\upsilon}}L(\overline{s}|\boldsymbol{\upsilon},\boldsymbol{\upsilon}_0)\Big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0}\right\|^2 > \text{const}, \tag{14}$$

where $\|\cdot\|$ is an appropriate vector norm and equations (7)-(13) define test statistics. It is obvious that, if we take $p_1=\ldots=p_K=1/K$ and hypothetical parameters $\boldsymbol{Q}_0$ determine the *r*-dimensional uniform distribution (i.e. $[\boldsymbol{Q}_0]_i=K^{-r}$), then test (17) will be locally the most powerful against any linear one-dimensional ony-sided alternative, since likelihood ratio denominator $\mathrm{P}\left(\overline{s}\middle|\boldsymbol{\upsilon}_0\right)$ turns into a constant. In other cases, we can be driven by logical relevance of the maximum likelihood ratio principle.

As an alternative for (17), we can use the test in the following form:

$$\left\|\nabla_{\boldsymbol{\upsilon}}\mathrm{P}(\overline{s}\,|\,\boldsymbol{\upsilon})\Big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0}\right\|^2 = \left\|\mathrm{P}(\overline{s}\,|\,\boldsymbol{\upsilon}_0)\cdot\nabla_{\boldsymbol{\upsilon}}\mathrm{L}(\overline{s}\,|\,\boldsymbol{\upsilon},\boldsymbol{\upsilon}_0)\Big|_{\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0}\right\|^2 > \text{const}, \tag{15}$$

which is locally the most powerful against any linear one-dimensional one-side alternative.

Concerning tests (17) and (18), there are still some questions to be answered. First of all, in (8)-(11), there are several unknown variables, which require reasonable replacements. Secondly, it is necessary to show how critical values will be determined and which vector norms should be used.

Let us start from the first question. To replace unknown values in (17) by some observed values, we can use the same principle which was used in [8].

For instance, ignoring some depending effects, we can replace $\mathrm{P}_{r,i}(C_{s_{t-r+1}},\ldots,C_{s_t}\,|\,c_k)$ by $\mathrm{P}^{(r,i)}(k_1,\ldots,k_r)=\sum_{s=1}^{K}\mathrm{P}^{(r)}(k_1,\ldots,k_{i-1},s,k_{i+1},\ldots,k_r)$. Thereafter, $\left(f_t(c_{k-1})-f_t(c_k)\right)/p_k$ and $\left(f_t(c_{k-1})\boldsymbol{d}_{k-1}-f_t(c_k)\boldsymbol{d}_k\right)/p_k$ will still be unknown. Their replacement by special scores $B_j=\left\{b_j(k):k=\overline{1,K}\right\}$, $j=\overline{1,r}$, is discussed in detail in [8]. For example, for K=2 (quantile regression), B_1={–1/p, 1/(1–p)}. For K=3 (symmetric two-quantile regression), B_1={–1, 0, 1}, B_2={1, –2p/(1–2p), 1}. For K=3 (symmetric three-quantile regression), B_1={–A, –α, α, A}, B_2={A, –1, –1, A}, where A=(1–2p)/2p and α is an a priori guess about value $\left(f_t(0)-f_t(\boldsymbol{\mu})\right)/f_t(\boldsymbol{\mu})$.

One of the most important features of the scores is their zero mean $\sum_{k=1}^{K}B_1(k)\,p_k=0$, $\sum_{k=1}^{K}B_2(k)\,p_k=0$ and zero covariation $\sum_{k=1}^{K}B_1(k)B_2(k)\,p_k=0$.

Replacements in (17) and (18) lead to obtaining tests in the following form:

$$\left\|\boldsymbol{\xi}_n(\overline{\boldsymbol{s}}\,|\,\boldsymbol{\upsilon}_0)\right\|^2>\mathrm{const}, \tag{16}$$

$$\left\|\mathrm{P}(\overline{\boldsymbol{s}}\,|\,\boldsymbol{\upsilon}_0)\cdot\boldsymbol{\xi}_n(\overline{\boldsymbol{s}}\,|\,\boldsymbol{\upsilon}_0)\right\|^2>\mathrm{const}. \tag{17}$$

Here, $\boldsymbol{\xi}_n(\overline{\boldsymbol{s}}\,|\,\boldsymbol{\upsilon}_0)$ is a vector statistic, which is a modified and normalized likelihood ratio gradient. Also, this statistic can be written in the following form:

$$\boldsymbol{\xi}_n(\overline{\boldsymbol{s}}\,|\,\boldsymbol{\upsilon})=n^{-1/2}\sum_{t=r}^{n}\sum_{i=1}^{r}W_{t,i}(\overline{\boldsymbol{s}}\,|\,\boldsymbol{\upsilon}),$$

$$\boldsymbol{W}_{t,i}(\overline{\boldsymbol{s}}|\boldsymbol{\upsilon})=\begin{bmatrix}\nabla_{\boldsymbol{\theta}}g_{t-r+i}(\boldsymbol{\theta})R_{t,1}^{(r,i)}(\boldsymbol{\upsilon})\\ R_{t,2}^{(r,i)}(\boldsymbol{\upsilon})\\ \dfrac{[G]'_{j_r(s_{t-r+1}(\boldsymbol{\upsilon}),\ldots,s_t(\boldsymbol{\upsilon}))}}{\mathrm{P}^{(r)}(s_{t-r+1}(\boldsymbol{\upsilon}),\ldots,s_t(\boldsymbol{\upsilon}))}-(1-\delta_{t,r})\dfrac{[F_{r-1}G]'_{j_{r-1}(s_{t-r+1}(\boldsymbol{\upsilon}),\ldots,s_{t-1}(\boldsymbol{\upsilon}))}}{\mathrm{P}^{(r-1)}(s_{t-r+1}(\boldsymbol{\upsilon}),\ldots,s_{t-1}(\boldsymbol{\upsilon}))}\end{bmatrix}, \tag{18}$$

$$R_{t,l}^{(r,i)}(\boldsymbol{\upsilon})=B_l(s_{t-r+i})\left(L_t^{(r,i)}(\boldsymbol{\upsilon})-(1-\delta_{t,r})(1-\delta_{i,r})L_{t-1}^{(r-1,i)}(\boldsymbol{\upsilon})\right),$$

$$L_t^{(r,i)}(\boldsymbol{\upsilon})=\frac{\mathrm{P}^{(r,i)}\left(s_{t-r+1}(\boldsymbol{\upsilon}),\ldots,s_t(\boldsymbol{\upsilon})\right)\mathrm{P}^{(1)}\left(s_{t-r+i}(\boldsymbol{\upsilon})\right)}{\mathrm{P}^{(r)}\left(s_{t-r+1}(\boldsymbol{\upsilon}),\ldots,s_t(\boldsymbol{\upsilon})\right)}.$$

It is noticeable that if $[\boldsymbol{Q}_0]_i=K^{-r}$, test (20) will still be locally the most powerful against any linear one-dimensional one-sided alternative, even after replacement of the unknown variable by scores. However, it is possible only with an additional condition that these alternatives differ only by parameters $\boldsymbol{\theta}$, or $\boldsymbol{\mu}$, or $\boldsymbol{Q}$. Plus, simulations show that the behavior of this test under alternatives is not always satisfactory. In addition, there are some difficulties in searching its asymptotic critical values. Therefore, we recommend to use test (19).

Now we can turn to the second mentioned question concerning choosing critical values and methods of defining vector statistics norms. As in the case of independent errors in [1,8], for hypothesis (3), we can build test in forms (19) and (20) with the exact significance value. This is possible due to the fact that under hypothesis, the distributions of statistics $\left\|\boldsymbol{\xi}_n(\overline{\boldsymbol{s}}\,|\,\boldsymbol{\upsilon}_0)\right\|^2$ and $\left\|\mathrm{P}(\overline{\boldsymbol{s}}\,|\,\boldsymbol{\upsilon}_0)\cdot\boldsymbol{\xi}_n(\overline{\boldsymbol{s}}\,|\,\boldsymbol{\upsilon}_0)\right\|^2$ coincide with the distribution of random variable $\left\|\boldsymbol{\xi}_n(\overline{\boldsymbol{\eta}}\,|\,\boldsymbol{\upsilon}_0)\right\|^2$ and $\left\|\mathrm{P}(\overline{\boldsymbol{\eta}}\,|\,\boldsymbol{\upsilon}_0)\cdot\boldsymbol{\xi}_n(\overline{\boldsymbol{\eta}}\,|\,\boldsymbol{\upsilon}_0)\right\|^2$, respectively, where random vector $\overline{\boldsymbol{\eta}}=(\eta_1,\ldots,\eta_n)'$ is composed of the sequence of random values, which are (r-1)-order Markov chain ($\eta_t\in\{1,2,\ldots,K\}$) with given finite-dimensional probabilities $\mathrm{P}\left(k_1,...,k_r \middle| \boldsymbol{\upsilon}_0\right)$. As a result, using the Monte-Carlo method, the percentage points of test statistics distributions (19) and (20) can be defined with any accuracy. It is noticeable that, in the determination of $\boldsymbol{\xi}_n(\bar{\boldsymbol{\eta}} \mid \boldsymbol{\upsilon}_0)$ parameters, $\boldsymbol{\mu}_0$ is not used, whereas $\boldsymbol{\theta}_0$ is used only in the determination of gradients $\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta}_0)$ in (21), whereas parameters $\boldsymbol{Q}_0$ influence only the values of probabilities $\mathrm{P}^{(r)}$, $\mathrm{P}^{(r-1)}$, $\mathrm{P}^{(r,i)}$, $\mathrm{P}^{(r-1,i)}$.

In case of a large number of observations *n*, it is better to use the asymptotic critical values. For test (19), it is possible to determine them through asymptotic normality of statistics distribution $\boldsymbol{\xi}_n(\bar{\boldsymbol{s}} \mid \boldsymbol{\upsilon}_0)$ under hypothesis. For this purpose, we can use generalization of theorem 7.7.9 from [9] in case of the sequence of random vectors, which could be easily obtained by applying theorem 7.7.7 from [9].

Lemma 1.

Let numerical sequence $\{a_t : t \geq 1\}$ and a sequence of random vectors $\{\boldsymbol{z}_t : t \geq 1\}$ satisfy the following conditions.

1. There is integer number *m*>0 so that for any *n* and $t_1,...,t_n$ $(0 < t_1 < ... < t_n)$, the sets of random variables $\{z_{t_1},...,z_{t_n}\}$ and $\{\boldsymbol{z}_1,...,\boldsymbol{z}_{t_1-m-1},\boldsymbol{z}_{t_n+m+1},...\}$ are mutually independent.
2. $\mathrm{M}z_t = 0$, $t = 1,2,...$.
3. $\mathrm{M}\|\boldsymbol{z}_t\|^{2+\delta} < M$ for some *M* and $\delta > 0$, $t = 1,2,...$.
4. $|a_t| < L$ for each *t* and some *L*>0.
5. There is limiting matrix $\Sigma = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\sum_{s=1}^{T} a_t a_s \mathrm{M}\boldsymbol{z}_t\boldsymbol{z}_s'$.

Then, random vector $T^{-1/2}\sum_{t=1}^{T} a_t \boldsymbol{z}_t$ converges distributionally to $N(0,\Sigma)$.

Theorem 2.

Let hypothesis (3) and the following conditions be fulfilled:

1. $\|\nabla_{\boldsymbol{\theta}} g_t(\boldsymbol{\theta})\| < L$ for each *t* and some *L*>0.
2. Limiting matrix $V = \lim_{n\to\infty} V_n$, where $V_n = \mathrm{M}\boldsymbol{\xi}_n(\bar{\boldsymbol{s}} \mid \boldsymbol{\upsilon})\boldsymbol{\xi}_n'(\bar{\boldsymbol{s}} \mid \boldsymbol{\upsilon})$ exists.
3. $\mathrm{P}^{(r)}(k_1,...,k_r)>0$ for each $k_1,...,k_r$.

Then random vector $\boldsymbol{\xi}_n(\bar{\boldsymbol{s}} \mid \boldsymbol{\upsilon})$ converges distributionally to *N*(0,*V*), whereas random value $\boldsymbol{\zeta}_n^2 = \boldsymbol{\xi}_n'(\bar{\boldsymbol{s}} \mid \boldsymbol{\upsilon})V_n^{-1}\boldsymbol{\xi}_n(\bar{\boldsymbol{s}} \mid \boldsymbol{\upsilon})$ converges distributionally to χ_q^2, where $q = \dim \boldsymbol{\upsilon}$.

5. Conclusion

In this article, we have obtained exact and asymptotic sign-based tests for simple hypothesis (3) $H_0 : \boldsymbol{\upsilon} = \boldsymbol{\upsilon}_0$ concerning the parameters of multi-quantile regression model (1) with stationary Markov noise ε_t,

In theorem 1, we showed that despite the nonparametric problem statement, we can obtain expressions for the gradient of likelihood for signs $\nabla \mathrm{P}(\bar{\boldsymbol{s}}|\boldsymbol{\upsilon})$. In this case, these expressions for the gradient do not depend on the parametrization method of distribution $\mathrm{P}(\bar{\boldsymbol{s}}|\boldsymbol{\upsilon})$. Thereafter, we used this fact to obtain tests (19) and (20) based on vector statistics $\boldsymbol{\xi}_n(\bar{\boldsymbol{s}} \mid \boldsymbol{\upsilon}_0)$.

In the article, we examined the question concerning obtainment of critical values, which achieve the necessary significance level with any accuracy on finite-samples. Further, we also examined such critical values which are based on asymptotic distribution of test statistics (theorem 2).

As a result, we recommended the test based on the principle of the maximum likelihood ratio with the following critical area:

$$\boldsymbol{\zeta}_n^2 = \boldsymbol{\xi}_n'(\bar{\boldsymbol{s}} \mid \boldsymbol{\upsilon}_0)V_n^{-1}\boldsymbol{\xi}_n(\bar{\boldsymbol{s}} \mid \boldsymbol{\upsilon}_0) > \mathrm{const},$$

where $V_n = V_n(\boldsymbol{\theta}, \boldsymbol{Q}) = \mathrm{M}\boldsymbol{\xi}_n(\overline{\boldsymbol{s}} \mid \boldsymbol{\upsilon})\boldsymbol{\xi}_n^{'}(\overline{\boldsymbol{s}} \mid \boldsymbol{\upsilon})$. This test has chi square asymptotic distribution.

The obtained test can be treated as a basis for the estimation procedure of parameters $\boldsymbol{\upsilon}$ based on maximum asymptotic *p*-value principle [7], i.e.,

$$\boldsymbol{\upsilon}_n = \arg\min_{\boldsymbol{\upsilon}_0} \boldsymbol{\zeta}_n^2(\boldsymbol{\upsilon}_0) .$$

For practical purposes, hypothesis testing is interesting when the hypothesis has the form of $H_0 : [\boldsymbol{\theta}]_j = [\boldsymbol{\theta}_0]_j$ and the rest parameters are nuisance. It is true especially for linear models. For such hypothesis, which can be treated as linear, we can use a two-stage testing procedure. This approach is closely examined in [1, 8] in the context of sign procedures.

References

[1] Boldin M V, Simonova G I and Tyurin Yu N 1997 *Sign-based Methods in Linear Statistical Models* (Providence: AMS Translations of mathematical monographs)
[2] Coudin E and Dufour J 2009 Finite-sample distribution-free inference in linear median regressions under heteroscedasticity and non-linear dependence of unknown form *Econometrics J.* **12** 19–49
[3] Tarassenko P F, Tarima S S, Zhuravlev A V and Singh S 2015 On sign-based regression quantiles *J. of Statist. Comput. and Simul.* **85** 1420–41
[4] Dufour J and Kiviet J 1998 Exact inference methods for first-order autoregressive distributed lag models *Econometrica* **82** 79–104
[5] Koenker R and Bassett G 1978 Regression quantiles *Econometrica* **46** 349–359
[6] Chen L A, Tran L T and Lin L C 2004 Symmetric regression quantile and its application to robust estimation for the non-linear regression model *J. Statist. Plann. Infer.* **126** 423–440
[7] Hodges J L Jr and Lehmann E L 1963 Estimates of location based on rank tests *J. Ann Math Statist* **34** 528–611
[8] Tarassenko P F 2005 *Indikatornyj Statisticheskij Analiz* (Tomsk: Tomsk State University Press)
[9] Anderson T W 1971 *The Statistical Analysis of Time Series* (New Jersey: John Wiley & Sons, Inc.)