Прогноз квазипериодических временных рядов непараметрическими методами*

Е. Ю. Клочков
eklochov@gmail.com
Московский физико-технический институт, ФУПМ, каф. «Интеллектуальные системы»

В работе рассматривается непараметрический метод прогнозирования квазипериодических временных рядов. В качестве метода используется квантильная регрессия. Его преимущества в том что, несмотря на его простоту, он хорошо приближает многие из известных распределений. Предлагаемый метод тестируется на данных о продажах продуктов.

Ключевые слова: *квантиль, квантильная регрессия, линейное программирование.*

Quasiperiodic time series forecast using nonparametric methods*

Y. Y. Klochkov
Moscow Institute of Physics and Technology

This paper considers the nonparametric method of quasi-periodic time series forecasting. As a method the quantile regression is used. Its advantage is that, despite its simplicity, it is a good approximation of many known distributions. The proposed method is tested on data from the sales of products.

Keywords: *quintile, quantile regression, linear programming.*

Введение

Использование квантильной регрессии началось с так называемых «неэффективных статистик» с использованием нескольких квантилей для замены «на скорую руку» [6]. Позднее, Р. Коенкер и Г. Бассет-младший расширили понятие обычных квантилей в модели локализации до более общего класса линейных моделей, в которых условные квантили имеют линейную форму [4].

Мы называем число $\theta-$квантилью случайной величины, если с вероятностью θ случайная величина не превосходит это число. Например, $\frac{1}{2}-$ квантиль есть ничто иное как медиана. Было обнаружено что взвешенное среднее арифметическое $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$ квантилей с коэффициентами 0.3, 0.4, 0.3 имеет асимптотическую эффективность почти восемьдесят процентов для нормального, лапласова, логистического распределений и распределения Коши [3]. Тем самым, отвлекаясь от конкретного вида распределения, квантильная регрессия захватывает широкий класс задач.

В работе [2] для прогнозирования временного ряда применяется медианная регрессия. В предположени, что значение временного ряда зависит от нескольких предшествующих ему, рассматриваемых в качестве признака, мы пронаблюдаем как с возрастанием числа признаков изменяется ошибка. Настоящий метод также применяется для расчета компенсации работникам [5], оценивания нетто-премий страховыми компаниями [1] и других задач.

Научный руководитель В. В. Стрижов

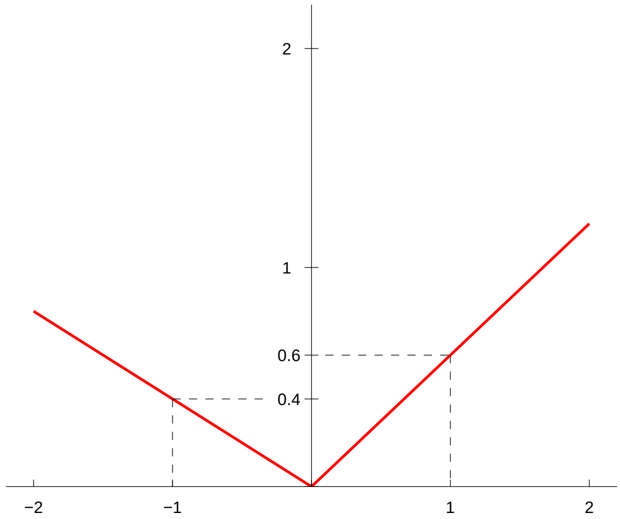


Рис. 1. График функции $\rho_{0.6}(y)$.

Постановка задачи

Имеется выборка $\{y_i, \mathbf{x}_i\}_{i=1}^t$, где $\mathbf{x}_i$ — $k \times 1$ вектор независимых признаков. Мы предполагаем, что условные квантили определяются соотношением

$$\mathrm{Quant}_\theta(y_i|\mathbf{x}_i) = \mathbf{x}_i^T\mathbf{b} + \varepsilon_i. \tag{1}$$

Оценка $\mathbf{b}_\theta$ определяется из соотношения

$$\hat{\mathbf{b}}_\theta = \arg\min_{\mathbf{b}\in\mathbb{R}^k} \sum_{i:y_i \geqslant \mathbf{x}_i^T\mathbf{b}} \theta|y_i - \mathbf{x}_i^T\mathbf{b}| + \sum_{i:y_i < \mathbf{x}_i^T\mathbf{b}} (1-\theta)|y_i - \mathbf{x}_i^T\mathbf{b}|. \tag{2}$$

Линейная модель (2) была представлена Коенкером и Бассетом [4], как обобщение простой квантили.

Определение 1. *Пусть задана случайная величина ξ. Тогда для $\theta \in (0,1)$ назовем θ-квантилью этой случайной велечины такое число x_θ, что*

$$\begin{cases} \mathsf{P}(\xi \leqslant x_\theta) \geqslant \theta, \\ \mathsf{P}(\xi < x_\theta) \leqslant \theta. \end{cases} \tag{3}$$

Как известно, медиана обладает наименьшим математическим ожиданием модуля отклонения. Аналогично можно показать, что при произвольном θ, положив функцию потери $\rho_\theta(y) \triangleq y(\theta - [y<0])$, соответствующая квантиль находится решением задачи

$$\arg\min_{u\in\mathbb{R}} \mathsf{M}\rho_\theta(\xi - u). \tag{4}$$

Это можно легко проверить, продифференцировав $\mathsf{M}\rho_\theta(\xi - u)$ по u. Действительно, распишем выражение $\mathsf{M}\rho_\theta(\xi - y)$ в виде суммы двух интегралов

$$\mathsf{M}\rho_\theta(\xi - u) = (1-\theta)\int_{-\infty}^{u} (u-y)dF_\xi(y) +$$

$$+\theta\int_{u}^{\infty} (y-u)dF_\xi(y)$$

и продифференцируем по u. Получим

$$(1-\theta)\int_{-\infty}^{u} dF_\xi(y) - \theta\int_{u}^{\infty} dF_\xi(y) = 0,$$

откуда,

$$(1-\theta)F_\xi(u) - \theta(1-F_\xi(u)) = 0,$$

$$F_\xi(u) = \theta,$$

что и соответствует определению квантили в случае, когда F непрерывная строго возрастающая.

При прогнозировании искомое значение y сопоставляется со случайной величиной ξ. Тогда для конечного набора значений выражение (4) переписывается в виде

$$\arg\min_{u\in\mathbb{R}} \sum_{i=1}^{t} \rho_\theta(y_i - u). \tag{5}$$

Эта задача обобщается до описаной выше в случае, когда ищется условная квантиль.

Представление в виде задачи ЛП

Задача (2) сводится к линейному программированию. У нас есть обучающая выборка $(y_i, \mathbf{x}_i)_{i=1}^t$. Из признаков составим матрицу $X = ||\mathbf{x}_1\, \mathbf{x}_2\, \ldots\, \mathbf{x}_t||$. Положим

$$\mathbf{r} = y - X^T\mathbf{b}$$

тогда,

$$r_i = y_i - \mathbf{x}_i^T\mathbf{b},\ i = \overline{1,\, t}.$$

Перепишем задачу (2) в виде

$$\hat{\mathbf{b}}_\theta = \arg\min_{\mathbf{b}\in\mathbb{R}^k} \sum_{i=1}^{t} \theta(y_i - \mathbf{x}_i^T\mathbf{b})_+ + (1-\theta)(y_i - \mathbf{x}_i^T\mathbf{b})_-,$$

где $(\cdot)_+$ и $(\cdot)_-$ — положительная и отрицательная части числа, соответственно. Тогда положив $\mathbf{r} = \mathbf{r}^+ - \mathbf{r}^-$, где $\mathbf{r}^+$, $\mathbf{r}^-$ – положительная и отрицательная части вектора $\mathbf{r}$, получаем задачу ЛП

$$\begin{cases} \theta\mathbf{e}^T\mathbf{r}^+ + (1-\theta)\mathbf{e}^T\mathbf{r}^- \to \min \\ X^T\mathbf{b} + \mathbf{r}^+ - \mathbf{r}^- = \mathbf{y} \\ (\mathbf{b}, \mathbf{r}^+, \mathbf{r}^-) \in \mathbb{R}^k \times {\mathbb{R}_+}^{2t}, \end{cases} \tag{6}$$

где $\mathbf{e} = (1, 1, \ldots, 1)^T$. Легко понять, что ${r^+}_i$ и ${r^-}_i$ не могут иметь ненулевые значения одновременно, поэтому задачи (2) и (6) эквивалентны. На выходе мы получаем решение $(\mathbf{b}, \mathbf{r}^+, \mathbf{r}^-)$, из которого нам нужен вектор $\mathbf{b}$.

Задачу также можно представить в каноническом виде

$$\begin{cases} \theta\mathbf{e}^T\mathbf{r}^+ + (1-\theta)\mathbf{e}^T\mathbf{r}^- \to \min \\ X^T\mathbf{b}^+ - X^T\mathbf{b}^- + \mathbf{r}^+ - \mathbf{r}^- = \mathbf{y} \\ (\mathbf{b}^+, \mathbf{b}^-, \mathbf{r}^+, \mathbf{r}^-) \in {\mathbb{R}_+}^{2k+2t}, \end{cases} \tag{7}$$

что на практике ускоряет работу алгоритма. Здесь в качестве решения задачи (2) возвращаем $\mathbf{b} = \mathbf{b}^+ - \mathbf{b}^-$.

Прогнозирование временных рядов

При прогнозировании временных рядов мы предполагаем, что каждое следующее значение временного ряда зависит от предыдущих. Тем самым в качестве признаков мы выбираем k предыдущих значений ряда: дан ряд $\{y_i\}_{i=1}^t$ — мы строим выборку $\{y_i, \mathbf{x}_i\}_{i=k+1}^t$, где

$$\mathbf{x}_i = (y_{i-1}, \ldots, y_{i-k})^T. \tag{8}$$

Решение представляется в виде

$$y_\theta = b_1 y_t + b_2 y_{t-1} + \ldots + b_k y_{t-k+1} \tag{9}$$

Так же может оказаться полезным положить в качестве $k+1$ -го признака единицу. Тогда решением будет

$$y_\theta = b_0 + b_1 y_t + b_2 y_{t-1} + \ldots + b_k y_{t-k+1} \tag{10}$$

В качестве функции потерь будем использовать $L(\tilde{y}, y) = |\tilde{y} - y|$, где $\tilde{y}$ – предсказание, y – известное значение выборки. Как уже отмечалось, наименьшим отклонением суммы модулей обладает медиана, поэтому в качестве предсказания будем брать $\tilde{y}_i = \mathbf{x}_i^T \mathbf{b}_{0.5}$.

Вычислительный эксперимент

Для вычислительного эксперимента используются данные о продажах красных вин в Австралии по месяцам с января 1980 по июнь 1994. Данные можно считать приближенно периодическими с периодом в 12 месяцев.

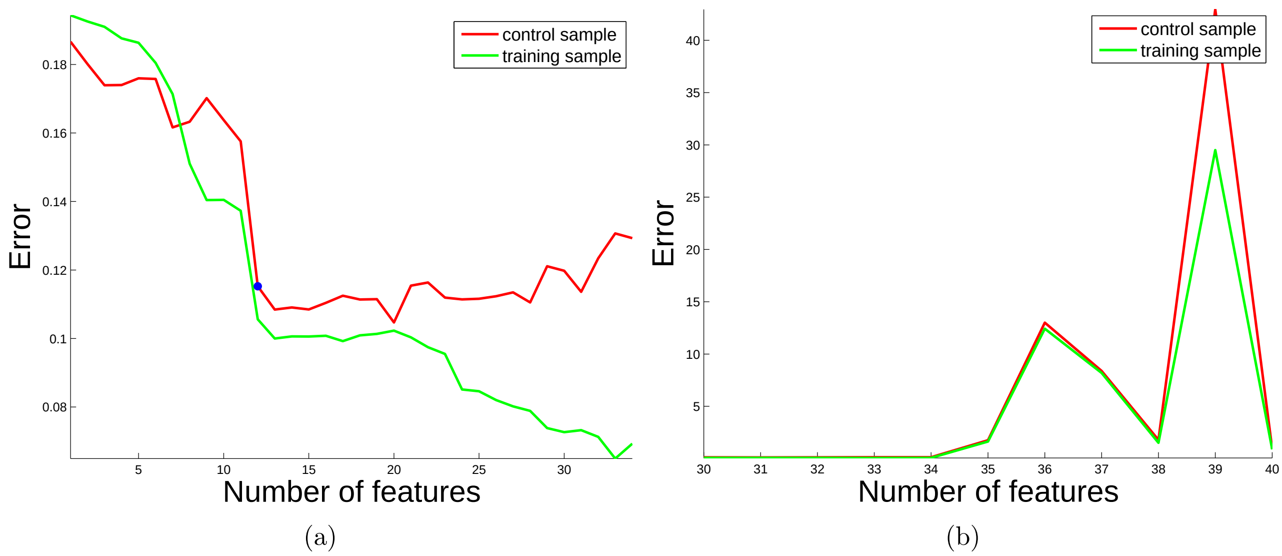
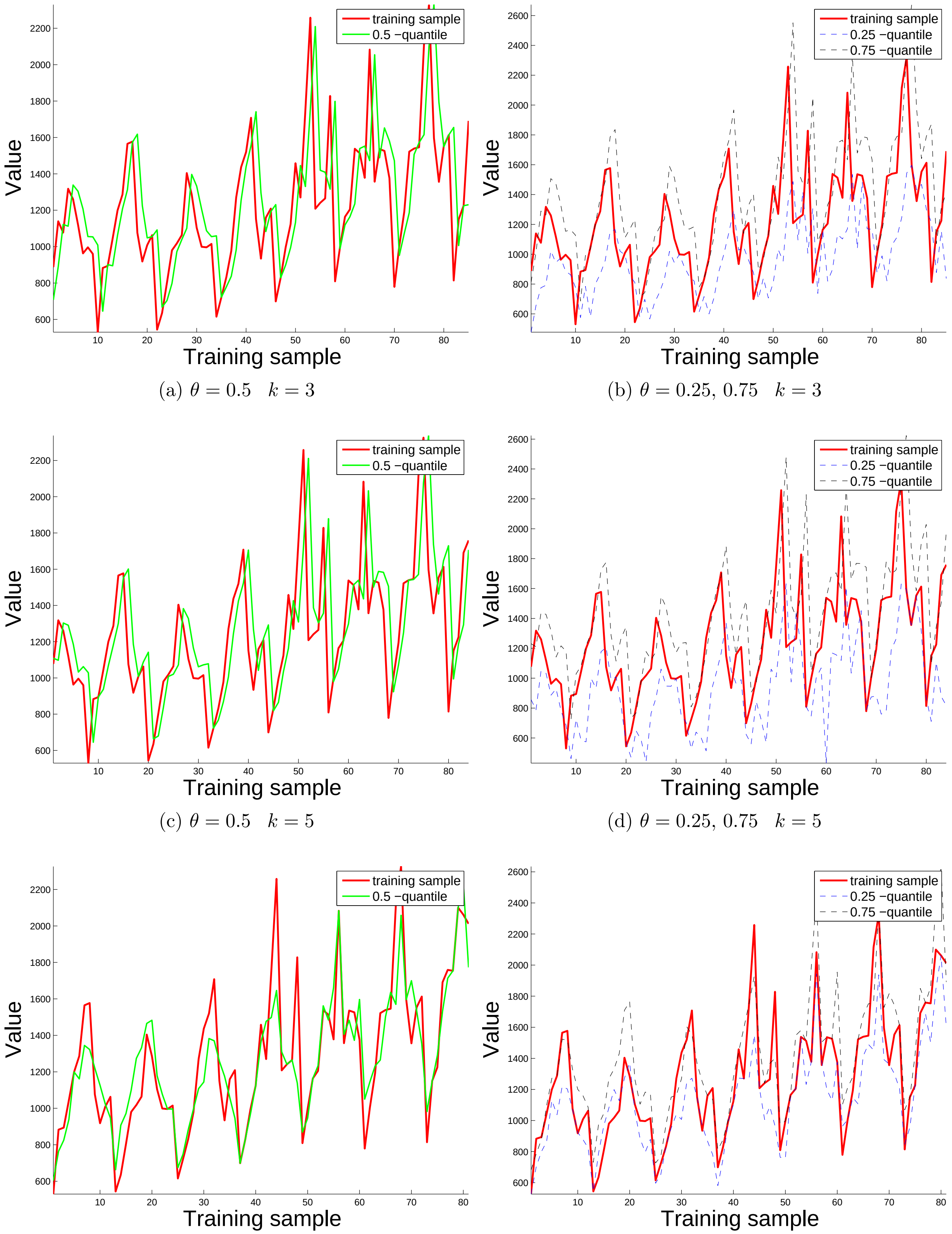


Рис. 2. Относительная ошибка на медиане, в зависимости от числа признаков.

Выборка, описанная в (8) разделена на две равные части — обучающая и контрольная. Обучающая используется для вычислений, контрольная — для оценки точности модели. На рис. 2 показана зависимость относительной ошибки от числа значений временного ряда, выбираемых в качестве признаков. Можно увидеть резкий спад на 12-ти признаках

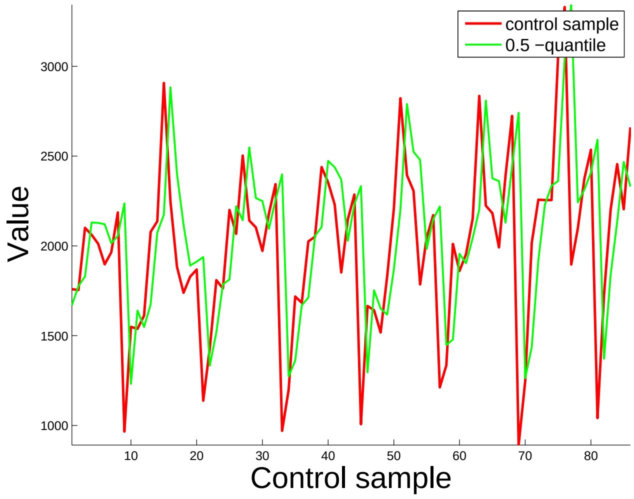


(a) $\theta = 0.5 \quad k = 3$ (b) $\theta = 0.25,\ 0.75 \quad k = 3$

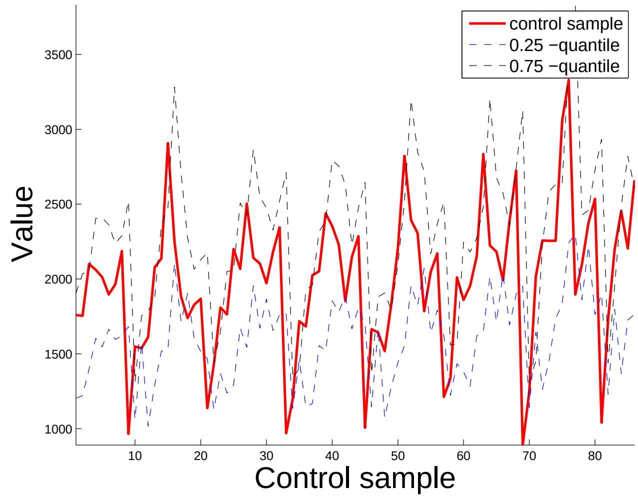
(c) $\theta = 0.5 \quad k = 5$ (d) $\theta = 0.25,\ 0.75 \quad k = 5$

(e) $\theta = 0.5 \quad k = 12$ (f) $\theta = 0.25,\ 0.75 \quad k = 12$

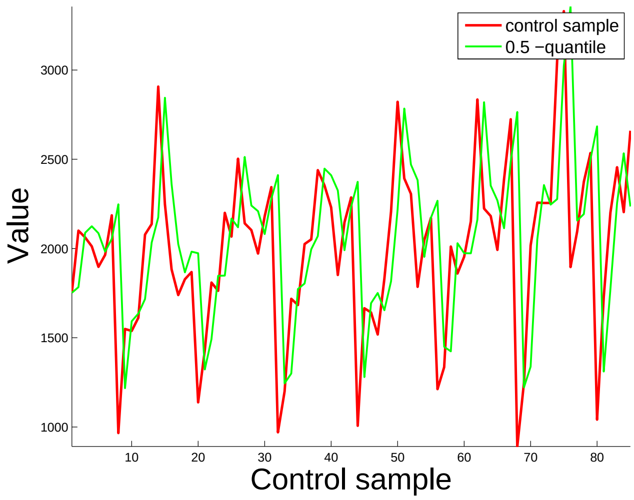
Рис. 3. Медиана и 0.25- и 0.75- квантили на обучающей выборке, число признаков $k = 3,\ 5,\ 12$



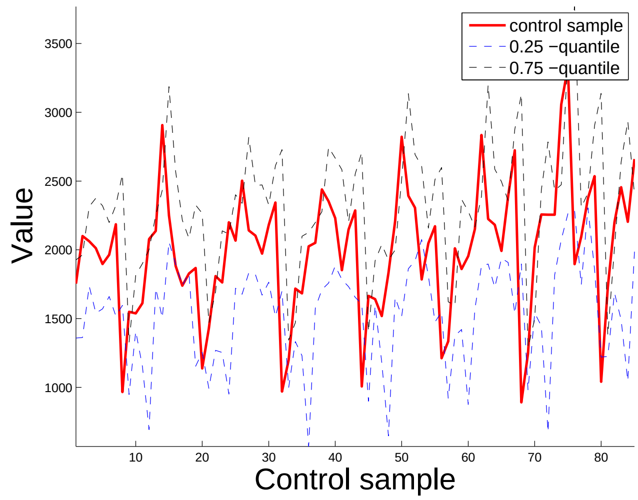
(a) $\theta = 0.5 \quad k = 3$



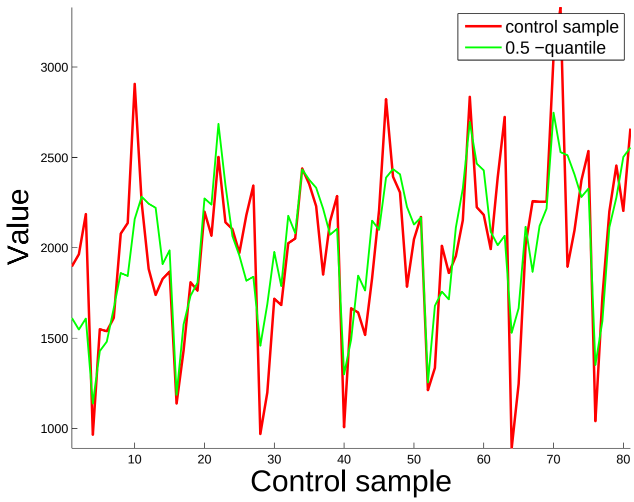
(b) $\theta = 0.25,\, 0.75 \quad k = 3$



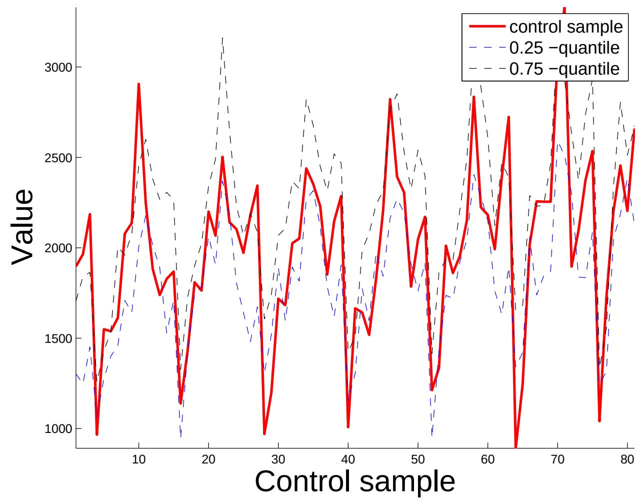
(c) $\theta = 0.5 \quad k = 5$



(d) $\theta = 0.25,\, 0.75 \quad k = 5$



(e) $\theta = 0.5 \quad k = 12$



(f) $\theta = 0.25,\, 0.75 \quad k = 12$

Рис. 4. Медиана и 0.25- и 0.75- квантили на контрольной выборке, число признаков $k = 3,\, 5,\, 12$

(точка выделена синим цветом), что соответствует одному году. На рис. 3 и 4 показаны результаты вычислений на обучающей и контрольной выборках, соответственно. По оси абсцисс откладываются месяцы, по оси ординат величины продаж, соответствующие месяцам. Полученные результаты сравниваются с известными значениями выборки при различных числах признаков k. Из графиков видно, что при $k = 3$ и 5 спады и подъемы сдвинуты, при $k = 12$ этот эффект пропадает. Коэффициенты вектора $\mathbf{b}$ для этого случая приведены в таблице 1.

θ	1	2	3	4	5	6	7	8	9	10	11	12
0.25	0.52	-0.08	0.19	0.24	0.11	-0.11	-0.18	-0.03	-0.02	0.23	0.02	0.05
0.5	0.63	0.05	0.2	0.07	0.03	-0.1	-0.07	0.01	0.01	0.15	0.07	0.01
0.75	0.82	0.09	0.02	0.01	0.14	-0.09	-0.06	-0.03	0.09	0.00	0.09	0.08

Таблица 1. Векторы $\mathbf{b}_{0.25}, \mathbf{b}_{0.5}, \mathbf{b}_{0.75}$

Как видно, значительно больше коэффициент при признаке, соответствующем значению ряда предыдущего года того же месяца. Тем не менее, квантили с меньшим числом признаков также хорошо приближают искомое значение — как видно из рис. 2, на 3-ех, 5-ти и 12-ти признаках ошибки соответственно, 0.19, 0.16 и 0.11 . На рис. 2(b) показана ошибка при k от 30 до 40. На графике видно, что при большом числе k ошибка сильно возрастает как на контрольной выборке (control sample), так и на обучающей (training sample). Поэтому это связано не с переобучением, а с тем, что метод оптимизации (7) расходится при больших входных данных.

Заключение

В работе рассматривалась задача прогнозирования временного ряда с использованием квантильной регрессии. Эксперимент показал, что даже при небольшом числе признаков метод дает хорошие результаты. Кроме того было отмечено, что при наблюдении за ошибкой при варьировании числа признаков, происходит резкий спад на периоде квазипериодического ряда. В ходе эксперимента переобучение не наблюдалось.

Литература

[1] *Абдурманов, Р.* Применение квантильной регрессии для оценки нетто-премий / Р. Абдурманов // *Ломоносов.* — 2008.

[2] *Литвинов, И.* Прогнозирование объемов продаж новых товаров (отчет) / И. Литвинов // *MachineLearning.ru.* — 2009.

[3] *Постникова, Е.* — Квантильная регрессия. — Master's thesis, НГУ, 2000.

[4] *Koenker, R.* Quantile regression / R. Koenker, J. Gilbert Bassett // *Econometrics.* — 1978. — Vol. 46, no. 1. — Pp. 33–50.

[5] *Koenker, R.* Quantile regression / R. Koenker, K. F. Hallock // *Journal of Economics Prospects.* — 2001. — Vol. 15, no. 4. — Pp. 143–156.

[6] *Mosteller, F.* On some useful "inefficient" statistics / F. Mosteller // *The Annals of Mathematical Statistics.* — Dec, 1946. — Vol. 17, no. 4. — Pp. 377–408.