

УДК 519.237

ИССЛЕДОВАНИЕ ВЗАИМОСВЯЗИ КОЭФФИЦИЕНТА РОЖДАЕМОСТИ И ИНДЕКСА ЦЕН В СУБЪЕКТАХ РФ С ПОМОЩЬЮ КВАНТИЛЬНОЙ РЕГРЕССИИ

Орлова И.В.

ФГБОУ ВО «Финансовый университет при Правительстве РФ», Москва, e-mail: IVOrlova@fa.ru

В настоящей статье исследуется взаимосвязь коэффициента рождаемости и индекса цен в субъектах РФ с помощью квантильной регрессии. Дополняя обычный регрессионный анализ, квантильная регрессия позволяет расширить спектр решаемых задач и тем самым получить дополнительную информацию из анализа исходной статистической информации об объекте исследования. Квантильная регрессия менее чувствительна к нарушению многочисленных предварительных требований, предъявляемых к исходной статистической информации в классическом регрессионном анализе. Применяя её при различных значениях уровня квантиля θ , мы можем получить набор квантильных регрессий вместо единственной, получаемой с помощью метода наименьших квадратов, и тем самым мы получаем более полную картину поведения объекта исследования. Существенное преимущество состоит также в том, что метод квантильной регрессии не является параметрическим методом. Кроме того, квантильная регрессия устойчива к выбросам в наблюдаемых регрессорах. Она также предоставляет возможность её использования при гетероскедастичности ошибок регрессии. Это позволяет не прибегать к различным ухищрениям для ослабления влияния гетероскедастичности на оценки, тем более что это не всегда возможно. Проведённое исследование позволяет сделать вывод, что регионы с относительно высокой рождаемостью в существенно большей степени реагируют на рост индекса цен, чем регионы с низкой рождаемостью. На концах распределения коэффициента рождаемости скорость изменения коэффициента рождаемости в зависимости от индекса цен сильно отличается от среднего значения, и при прогнозировании коэффициента рождаемости целесообразно учитывать, принадлежит ли регион к группе с высокой или низкой рождаемостью.

Ключевые слова: квантильная регрессия, коэффициент рождаемости, индекс цен

STUDY OF THE RELATIONSHIP BETWEEN THE BIRTH RATE AND THE PRICE INDEX IN THE SUBJECTS OF THE RUSSIAN FEDERATION USING QUANTILE REGRESSION

Orlova I.V.

Financial University under the Government of the Russian Federation, Moscow, e-mail: IVOrlova@fa.ru

This article examines the relationship between the birth rate and the price index in the subjects of the Russian Federation using quantile regression. In addition to the usual regression analysis, quantile regression allows you to expand the range of problems to be solved and thereby obtain additional information from the analysis of the initial statistical information about the object of study. Quantile regression is less sensitive to the violation of numerous preliminary requirements for the initial statistical information in classical regression analysis. Applying it for different values of the quantile level, we can get a set of quantile regressions instead of the single one obtained using the least squares method, and thus we get a more complete picture of the behavior of the object of study. A significant advantage is also that quantile regression is not a parametric method. In addition, quantile regression is resistant to outliers in the observed regressors. It also provides an opportunity to use it for heteroscedasticity of regression errors. This allows us not to resort to various tricks to weaken the influence of heteroscedasticity on estimates, especially since this is not always possible. The study allows us to conclude that regions with a relatively high birth rate are significantly more responsive to the growth of the price index than regions with a low birth rate. At the ends of the distribution of the birth rate, the rate of change in the birth rate depending on the price index is very different from the average value, and when predicting the birth rate, it is advisable to take into account whether the region belongs to a group with a high or low birth rate.

Keywords: quantile regression, birth rate, price index

Исследование взаимосвязи экономических показателей является актуальной задачей экономической науки. Ведущая роль в решении этой задачи принадлежит эконометрике. Возможности эконометрических исследований значительно расширяются с расширением возможностей вычислительной техники и развитием экономико-математических методов. В настоящее время наиболее широко применяются и развиваются методы регрессионного анализа, в котором на основе статистических данных об n наблюдениях исследуется условное среднее зависимой переменной Y от определяющих пере-

менных X_1, \dots, X_p , $E(Y|X)$. В предположении, что значение Y в i -м наблюдении y_i равно линейной комбинации x_{ij} значений X_j в этом наблюдении, $j = 1, \dots, p$, плюс случайный остаток ε_i со средним значением ноль, $y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$, условное среднее $E(y_i|x_i)$ равно

$$E(y_i|x_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad i = 1, \dots, n,$$

где $x_i = (1, x_{i1}, \dots, x_{ip})$ – i -я строка матрицы значений регрессоров X .

Вместо поиска условного среднего значения прогнозируемой переменной квантильная регрессия направлена на поиск ус-

ловной медианы и любых других условных квантилей. Дополняя обычный регрессионный анализ, квантильная регрессия позволяет расширить спектр решаемых задач и тем самым получить дополнительную информацию из анализа исходной статистической информации об объекте исследования. В отличие от классической линейной регрессии, квантильная регрессия является непараметрическим методом. Она менее чувствительна к нарушению многочисленных предварительных требований, предъявляемых к исходной статистической информации в классическом регрессионном анализе [1, 2].

Цель исследования: выявление и оценка изменения степени влияния индекса цен на коэффициент рождаемости в субъектах РФ в зависимости от уровня индекса цен с помощью квантильной регрессии.

Материалы и методы исследования

В исследовании использованы данные по субъектам РФ: коэффициент рождае-

мости (число родившихся на 1000 человек населения) и индекс потребительских цен. Источник – Федеральная служба государственной статистики [3, 4].

Мы можем предположить наличие линейной взаимосвязи между коэффициентом рождаемости и индексом цен. При этом также можем предположить наличие гетероскедастичности в остатках, которая зачастую наблюдается в моделях, построенных по пространственным данным.

На рис. 1 приведена диаграмма рассеяния коэффициента рождаемости от индекса цен, на которой в верхней части отчетливо видны «выбросы» – наблюдения, относящиеся к регионам, имеющим аномально высокие по сравнению с другими регионами значения коэффициента рождаемости [5].

Построив линейную регрессионную модель, проверим выполнение второй предпосылки МНК, а именно, условие гомоскедастичности с помощью теста Бреуша – Пагана (функция `bptest()` в R).

```
library(lmtest)
> fm<-lm(Y~X)
> summary(fm)
> # Бреуша-Пагана
> bptest(fm, varformula = ~X, studentize = FALSE)
```

```
Breusch-Pagan test
data: fm
BP = 9.5856, df = 1, p-value = 0.001961
```

Тестовая статистика имеет значение 9,59; p-value = 0,002 свидетельствуют о наличии гетероскедастичности, как и можно было предположить при работе с пространственными данными.

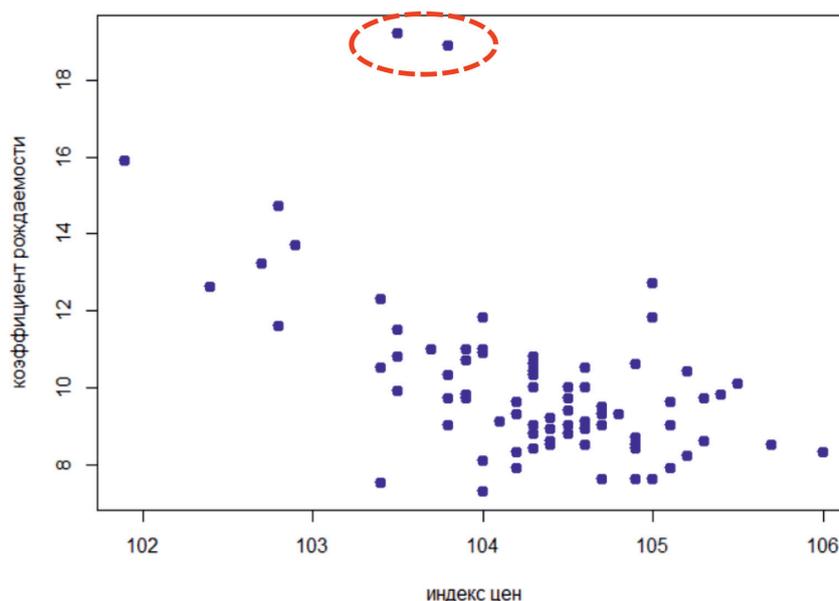


Рис. 1. Диаграмма рассеяния коэффициента рождаемости

Будем использовать метод квантильной регрессии, который относится к робастным методам, так как является устойчивым к отклонениям от предположений классических моделей. Метод квантильной регрессии устойчив к аномальным наблюдениям, «выбросам» и успешно применяется при неоднородности данных.

Напомним, что для любых θ , таких, что $0 < \theta < 1$, 100θ -процентный квантиль q_θ представляет собой пороговое значение, для которого случайная величина может принять меньшее значение с вероятностью θ . Функция распределения случайной величины в точке q_θ принимает значение, равное θ , $F(q_\theta) = \theta$.

Квантиль $q(\theta)$ можно представить как решение задачи оптимизации, минимизирующее взвешенную сумму модулей отклонений случайной величины Y от квантиля [6]

$$q(\theta) = \operatorname{argmi} n_c E[\rho_\theta(Y - c)], \quad (1)$$

где $\rho_\theta(\cdot)$ обозначает следующую функцию:

$$\rho_\theta(y) = (\theta - I(y < 0))y = ((1 - \theta)I(y < 0) + \theta I(y \geq 0))y.$$

Индикаторная функция $I(y < 0)$ принимает значение 1, если $y < 0$ и значение 0, если $y \geq 0$, $I(y < 0) = \begin{cases} 1, & \text{если } y < 0 \\ 0, & \text{если } y \geq 0 \end{cases}$.

Исходя из этого, величина $\rho_\theta(y - c)$ принимает значение $(1 - \theta)|y - c|$, если $y < c$, и значение $\theta|y - c|$, если $y \geq c$. Функция $\rho_\theta(z)$ является асимметричной функцией от модуля z . При этом вес $(1 - \theta)$ присваивается отрицательным z , а вес θ присваивается положительным z .

Квантильная регрессия предполагает, что условный квантиль порядка θ случайной величины Y линейно зависит от регрессоров $X_j, j = 1, \dots, p$,

$$q_\theta(y_i | \mathbf{x}_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}. \quad (2)$$

Квантильная регрессия позволяет получить иную, по сравнению с классической регрессией, информацию о поведении эндогенной переменной. Применяя её при различных значениях уровня квантиля θ , мы можем получить набор квантильных регрессий вместо единственной, получаемой с помощью метода наименьших квадратов, и тем самым мы получаем более полную картину поведения исследуемого показателя. Существенное преимущество состоит также в том, что метод квантильной регрессии не является параметрическим методом. Кроме того, квантильная регрессия устойчива к выбросам в наблюдаемых регрессорах. Она также предоставляет возможность её использования при гетероскедастичности ошибок регрессии. Это позволяет не прибегать к различным ухищрениям для ослабления влияния гетероскедастичности на оценки, тем более что это не всегда возможно.

Результаты исследования и их обсуждение

Рассмотрим задачу моделирования влияния индекса цен на коэффициент рождаемости в субъектах Российской Федерации с помощью квантильной регрессии в среде R и в Gretl.

Первая модель была получена в Gretl для квантиля 0,9 и варианта «Вычислить стандартные ошибки» при отключённом «Робастные стандартные ошибки / интервалы».

Модель 1: Квантильная оценка, использованы наблюдения 1-82

Зависимая переменная: Y

tau = 0,9

	Коэффициент	Ст. ошибка	t-статистика	P-значение	
const	193,497	45,9010	4,216	<0,0001	***
X	-1,74286	0,440091	-3,960	0,0002	***
Медиана зав. перемен	9,650000		Ст. откл. зав. перемен	2,143853	
Сумма модулей ошибок	169,9286		Сумма кв. остатков	487,2063	

Уравнение квантильной регрессии порядка $\tau = 0,9$ имеет вид

$$q_\tau(y_i | \mathbf{x}_i) = 193,497 - 1,74286x_i.$$

Все коэффициенты модели значимы. Выборочный коэффициент регрессии $\hat{\beta}_1$ при x_i отрицательный, $\hat{\beta}_1 = -1,74286$. Это означает, что для 10% регионов, в которых коэффициент рождаемости находится выше предсказанного по 90%-му уравнению квантильной регрессии, повышение индекса цен на единицу приводит в среднем к падению коэффициента рождаемости на 1,74.

В меню **Модель 1** при выборе **Анализ/Матрица коэффициентов ковариации** была получена автоковариационная матрица, на главной диагонали которой находятся дисперсии коэффициентов регрессии:

	const	x
const	2106,9	-20,2
x	-20,2	0,19368

Вектор оценок коэффициентов квантильной регрессии $\hat{\beta}$ имеет асимптотически нормальное распределение. В протоколе модели 1 приведены стандартные ошибки параметров s_{β_j} ($s_{\beta_1} = \sqrt{0,19368} = 0,440091$) и соответствующие t-статистики. Далее доверительные границы параметров получаются в виде $\hat{\beta}_j \pm t^* s_{\beta_j}$, где t^* – критическое значение для заданного уровня значимости. Эти доверительные границы можно получить при выборе в меню **Окно модели/Анализ/Доверительные интервалы для коэффициентов**:

$$t(80, 0,025) = 1,990$$

Переменная	коэффициент	95 доверительный интервал
const	193,497	(102,151, 284,843)
x	-1,74286	(-2,61867, -0,867049)

Gretl и R предоставляют возможность получения доверительных интервалов $\hat{\beta}_j$ разными способами. Коэффициенты квантильной регрессии для всех методов одинаковые. Различаются лишь 5%-ные доверительные интервалы коэффициентов. Они получены при разных допущениях относительно отклонений от регрессии. В основном рассматриваются 3 варианта предположений относительно отклонений:

независимые одинаково распределённые (i.i.d), не независимые одинаково распределённые (n.i.d), независимые неодинаково распределённые (i.ni.d). К тому же предоставляется возможность выбора робастных или неробастных методов оценки. Выбор того или иного допущения предоставляется пользователю.

В Gretl можно задать не один квантиль, а список квантилей. В этом случае будут вычислены оценки коэффициентов квантильной регрессии, их стандартные ошибки и доверительные интервалы для каждого из указанных значений.

При выдаче протокола **Наблюдаемые и расчетные значения** некоторые наблюдения помечаются * – это означает, что ошибка квантильной регрессии для этого наблюдения выходит за пределы 2,5 единиц стандартных ошибок. В нашем случае – это 3 наблюдения: 42 (Чеченская Республика) и 63 (Республика Тыва) имеют коэффициент рождаемости существенно выше предсказанного для данной группы регионов с высокой рождаемостью, а для 46-го наблюдения (Республика Мордовия) коэффициент рождаемости значительно ниже ожидаемого. Это явно аномальные наблюдения. Для классической регрессии они бы сильно повлияли на результаты моделирования, а для квантильной регрессии не имеет значения, насколько велико отклонение u_i от ожидаемого значения, имеет значение только факт принадлежности наблюдения к данной группе.

Далее воспользуемся возможностью построения серии моделей квантильной регрессии с различными значениями квантилей, указав квантили от 0,1 до 0,9 с шагом 0,1 (Модель 3).

Модель 3: Квантильная оценка, использованы наблюдения 1-82

Зависимая переменная: Y

Асимптотические стандартные ошибки считаются независимыми и одинаково распределёнными

	tau	Коэффициент	Ст. ошибка	t-статистика
const	0,100	127,600	32,0240	3,98452
	0,200	139,275	23,3508	5,96446
	0,300	139,625	17,0092	8,20881
	0,400	154,378	23,1615	6,66528
	0,500	137,300	29,7506	4,61503
	0,600	159,143	27,6001	5,76603
	0,700	155,306	27,8035	5,58583
	0,800	167,250	34,2478	4,88352
	0,900	193,497	45,9010	4,21553
	x	0,100	-1,14286	0,307040
0,200		-1,25000	0,223883	-5,58327
0,300		-1,25000	0,163081	-7,66491
0,400		-1,38889	0,222068	-6,25434
0,500		-1,22222	0,285244	-4,28483
0,600		-1,42857	0,264625	-5,39848
0,700		-1,38889	0,266575	-5,21013
0,800		-1,50000	0,328362	-4,56813
0,900		-1,74286	0,440091	-3,96022

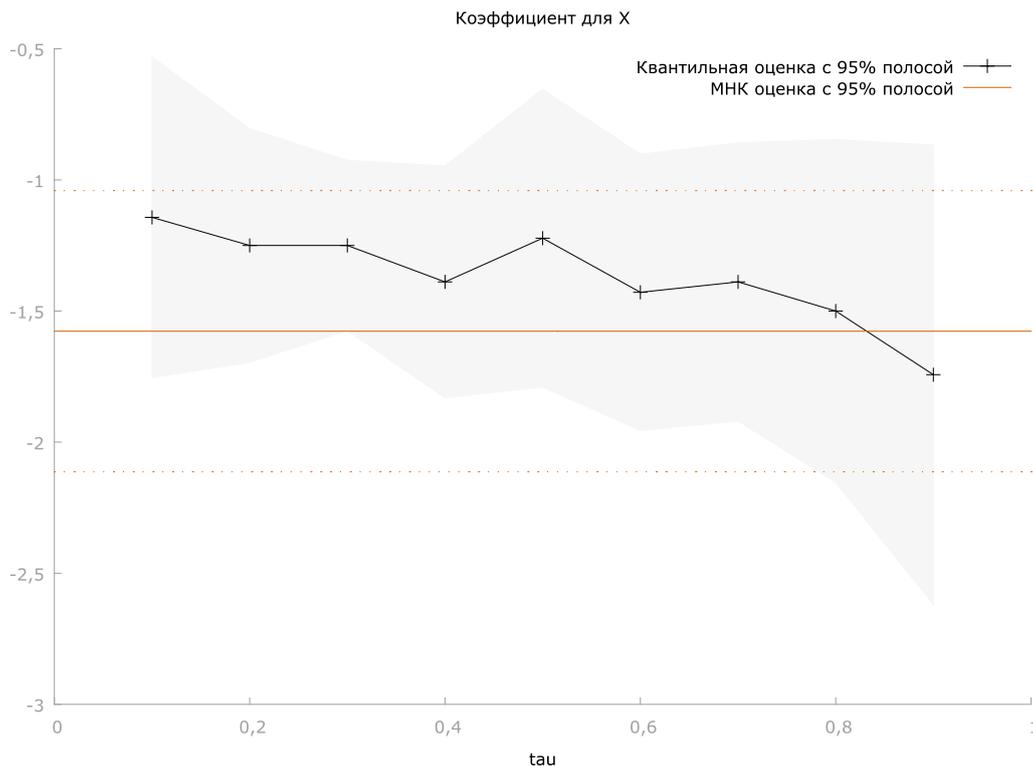


Рис. 2. Изменение значений коэффициента $\hat{\beta}_1$

Изменение значений коэффициента $\hat{\beta}_1$ моделей в зависимости от значения квантиля τ приведено на рис. 2.

Коэффициент $\hat{\beta}_1$ всех квантильных регрессий меньше нуля. Это означает, что все квантили с ростом индекса цен убывают, причём чем больше τ , тем скорость убывания больше. Для 10% регионов, в которых коэффициент рождаемости выше 90% условного квантиля, вычисленного по уравнению квантильной регрессии с уровнем квантиля $\tau = 0,9$, скорость убывания границы подмножества, определяемой 90%-ным условным квантилем коэффициента рождаемости, равна $\hat{\beta}_1 = -1,74$, а для 10% регионов с условно низким коэффициентом рождаемости, лежащим ниже вычисленного по уравнению 10%-ной квантильной регрессии значения, убывание квантиля при возрастании индекса цен на единицу составляет $\hat{\beta}_1 = -1,14$, то есть при увеличении индекса цен на единицу 10%-ный квантиль коэффициента рождаемости, отделяющий регионы с низкой рождаемостью от остальных регионов, уменьшится на 1,14. Это означает, что регионы с относительно высокой рождаемостью в существенно большей степени реагируют на рост индекса цен. По уравнению ре-

грессии, полученному с помощью метода наименьших квадратов (МНК), скорость убывания среднего значения коэффициента рождаемости составляет $-1,577$, то есть при увеличении индекса цен на 1 коэффициент рождаемости уменьшится в среднем на $-1,577$. Однако, как видим, на концах распределения коэффициента рождаемости скорость изменения коэффициента рождаемости сильно отличается от среднего значения и при прогнозировании коэффициента рождаемости целесообразно учитывать, принадлежит ли регион к группе с высокой или низкой рождаемостью.

Ниже приведены графики квантильных регрессий при τ , равных 0,1; 0,25; 0,5; 0,75; 0,9 и график классической регрессии для условного среднего Y .

С помощью такого рода графиков можно строить доверительные границы для предсказанных значений y_i . Графики всех квантильных регрессий почти параллельны. Третий сверху график относится к классическому уравнению регрессии. Угловой коэффициент графика смещён, вероятно, под влиянием аномальных наблюдений, и потому этот график пересекается с графиком медианной регрессии. В данном случае прогнозировать значения y_i предпочтительней по медианной регрессии.

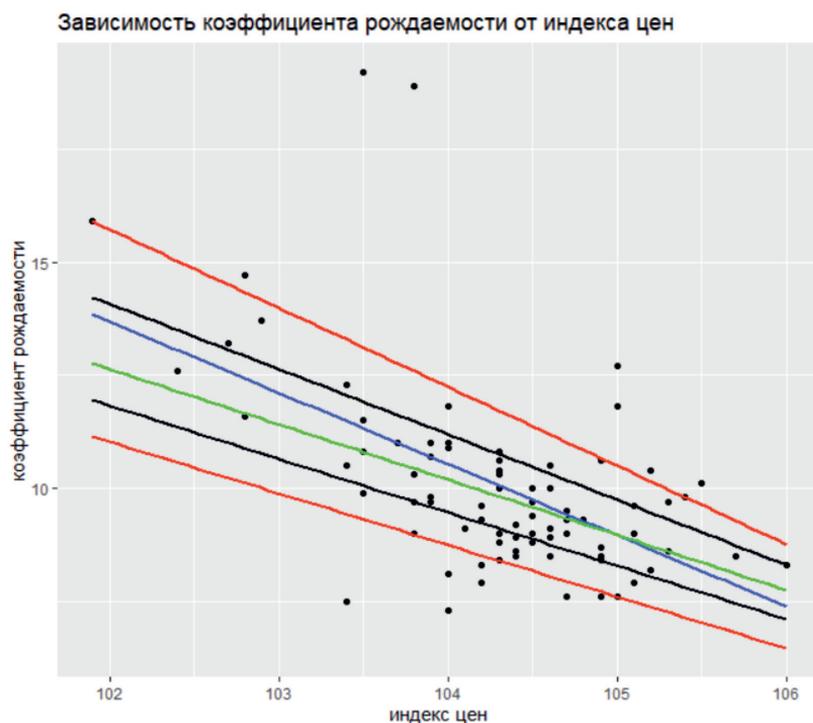


Рис. 3. Графики квантильных регрессий при пяти различных τ и график классической регрессии

Заключение

В заключение ещё раз подчеркнём, что квантильная и классическая регрессии не конкурируют между собой, а взаимно дополняют друг друга. Они отвечают на разные вопросы. Если в центре внимания классической регрессии лежит математическое ожидание условного среднего в зависимости от того, какое значение примет определённый набор параметров, называемых регрессорами, то в квантильной регрессии изучается математическое ожидание условных квантилей случайной величины в зависимости от того, какое значение примет набор регрессоров.

Список литературы

1. Kwark Noh-Sun, Lee Changhyun. Asymmetric effects of financial conditions on GDP growth in Korea: A quantile regression analysis, *Economic Modelling*, Elsevier. 2021. Vol. 94(C). P. 351–369.
2. Koenker R. *Quantile regression*. Cambridge University Press, NY, 2005. 368 p. DOI: 10.1017/CBO9780511754098.
3. Федеральная служба государственной статистики Регионы России. Социально-экономические показатели [Электронный ресурс]. URL: https://rosstat.gov.ru/bgd/regl/b19_14p/Main.htm (дата обращения: 20.03.2021).
4. Коэффициенты рождаемости, смертности, естественного прироста [Электронный ресурс]. URL: https://gks.ru/free_doc/2019/demo/t1_3.xlsx (дата обращения: 20.03.2021).
5. Орлова И.В. Выявление аномальных наблюдений при анализе динамики безработицы в Российской Федерации // *Фундаментальные исследования*. 2020. № 5. С. 142–148.
6. Hao L., Naiman D. *Quantile Regression*. Sage Publications, California, US. 2007. 107 p. DOI: 10.4135/9781412985550.