

УДК: 658.01

Абдураманов Р.А.

Національний технічний університет України «КПІ»

Пасенченко Ю.А.

Національний технічний університет України «КПІ»

МОДЕЛИРОВАНИЕ НЕТТО-ПРЕМИЙ С ПРИМЕНЕНИЕМ МЕТОДА КВАНТИЛЬНОЙ РЕГРЕССИИ

MODELING NET-PREMIUMS USING THE METHODS OF QUANTILE RE- GRESSION

Регресійні моделі – популярний інструмент для задач тарифікації в умовах неоднорідних портфелів. Незважаючи на це, класичні регресійні методи мають ряд недоліків, які значно обмежують сферу їх використання. Стаття присвячена альтернативному підходу оцінювання квантильної регресії. Метод квантильної регресії дає змогу подолати недоліки класичної регресійної моделі.

Регрессионные модели – популярный инструмент для задач тарификации в условиях неоднородных портфелей. Тем не менее, классические регрессионные методы имеют ряд недостатков, которые значительно ограничивают сферу их применения. Статья посвящена альтернативному подходу оценивания квантильной регрессии. Метод квантильной регрессии позволяет преодолеть недостатки классической регрессионной модели.

Regression models are popular tool for rate-making in the context of heterogeneous portfolios. Nevertheless, classical regression methods have some disad-

vantages which significantly restrict their sphere of using. The paper is devoted to an alternative approach – quantile regression. The quantile regression method allows overcome disadvantages of classical regression.

Ключевые слова: регрессионные модели, доверительный интервал, тарификация, квантильная регрессия.

Вступление. Устойчивость и достаточный уровень платежеспособности страховой компании определяется достаточной стоимостью услуг, которые она предоставляет на рынке, за счет которых должна обеспечить покрытие будущего ущерба, нормативных издержек страховой компании, а также обеспечить нормальную прибыль. Таким образом, основная задача страховщика заключается в оценке стоимости предоставляемых услуг, то есть оценки страховой премии. Данная статья посвящена вопросам статистического оценивания нетто-премии, которая, собственно, и предназначена для возмещения будущего чистого ущерба, а также применению альтернативного метода оценивания – метода квантильной регрессии [2, 6, 7, 9].

Задача оценки нетто-премий рассматривается в работах М. Томаса [1], М. Де-нюйт [4] и Дж. Дасн [5], в которых авторы рассматривают как теоретические, так и практические аспекты оценки нетто-премий. Классические методы оценивания ограничены выполнением жестких предположений относительно анализируемых данных. В более поздних работах для оценки нетто-премий предполагается использовать обобщенные линейные модели Д. Андерсона [3] и П. МакКулага [8]. Данный инструмент в большей степени согласуется с основной идеей оценки нетто-премий и учитывает экономическую специфику самой задачи, но так же требует выполнения основных предположений.

Постановка задачи. Основной целью данной статьи является предложение принципиально нового подхода к оценке нетто-премий, который позволяет преодолеть основные недостатки классических регрессионных методов. Данный метод основан на методе квантильной регрессии.

Важно обратить внимание на то, что он не только позволяет решить ряд технических проблем, но и хорошо согласуется с идеей оценки нетто-премии как квантили распределения выплат.

Методология. В качестве методологической базы исследования использовалась эконометрическая теория, теория математической статистики, а также инструменты численных методов.

Практическая реализация рассматриваемых методов была реализована в среде Maple.

Результаты исследования. Для расчета и научного обоснования размера нетто-премии используют так называемый принцип страховой эквивалентности, который увязывает нетто-премии с оценкой будущих выплат. При этом стандартный подход состоит в выделении ожидаемой нетто-премии, соответствующей средним выплатам по портфелю однородных рисков в отсутствие выбросов (то есть событий катастрофического характера), и рисковой надбавки, обеспечивающей страховщику запас надежности на случай превышения фактических выплат над ожидаемыми. Иногда нетто-премию предлагается рассчитывать как квантиль распределения выплат. Такой подход имеет свои преимущества и недостатки, но он хорошо объясняет необходимость рисковой надбавки.

Если страховой портфель неоднороден, то есть застрахованные объекты, так или иначе, отличаются друг от друга (что практически всегда имеет место на практике), тогда стараются выделить факторы риска (ковариаты), объясняющие эти различия. В подобной ситуации нетто-премия оценивается на основе условного распределения выплат (при условии на значения ковариат), что фактически позволяет дифференцировать портфель на более однородные подпортфели, для каждого из которых имеет место свое условное распределение.

На практике для анализа неоднородных портфелей часто используют регрессионные методы. К сожалению, классические регрессионные модели не вполне согласуются с особенностями объекта исследования, то есть со спецификой реальных страховых портфелей. Прежде всего, они ориентируются на услов-

ные математические ожидания (в нашем случае – оценки ожидаемой нетто-премии), оставляя без внимания проблему обоснования рисков надбавок. Кроме того, часто игнорируется характер используемых данных. В частности, страховые портфели характеризуются:

- наличием событий катастрофического характера;
- зависимостью застрахованных объектов (кумуляцией рисков),
- дефицитом информации для проверки адекватности избранной модели.

Классические регрессионные модели не устойчивы к выбросам, часто требуют независимости наблюдений и предполагают достаточно большой объем выборки. Поэтому их следует использовать с большой осторожностью и с привлечением дополнительной количественной и качественной информации.

Тем не менее, классические регрессионные методы не позволяют получить полностью однородные группы, так как даже после разбиения на основе факторов риска (ковариат) остается довольно высокая степень остаточной неоднородности. В этой ситуации наилучшей моделью будущих выплат будет смесь распределений, которая не очень хорошо согласуется с указанными подходами, требующими, как правило, «красивых» распределений.

На практике для оценки нетто-премий широко распространены линейные регрессионные модели, но иногда применяются и нелинейные подходы. Общий вид модели можно представить следующим образом:

$$y = g^{-1}(X\beta). \quad (1)$$

Она позволяет учесть влияние описательных характеристик застрахованного объекта (факторов риска, или ковариат) на величину страховых выплат. Различные подходы, по существу, отличаются выбором функции $g(\cdot)$ и характером учета ошибки.

Так, для классических моделей, основанных на методе наименьших квадратов, соотношение (1) принимает вид

$$g(y_i) = \mathbf{x}_i\beta + \varepsilon_i \quad \text{или} \quad E[g(\tilde{y}_i) | \tilde{\mathbf{x}}_i = \mathbf{x}_i] = \mathbf{x}_i\beta,$$

что хорошо согласуется с идеей линеаризации. Стандартное предположение состоит в том, что ошибки ε_i подчинены нормальному распределению с нулевыми математическими ожиданиями и подходящими дисперсиями σ_i^2 . Это означает жесткую фиксацию распределения случайной величины $\tilde{y}_i$. Квантиль распределения оценивается по формуле

$$Quant_{\theta}(\tilde{y}_i | \mathbf{x}_i) = g^{-1}(\sigma_i \cdot z_{\theta} + \mathbf{x}_i \boldsymbol{\beta}). \quad (2)$$

Предпосылки о независимости наблюдений, а также об одинаковом распределении ($\sigma_i^2 = \text{const}$) можно до определенной степени ослабить.

В качестве функции $g(\cdot)$ можно выбрать любую дифференцируемую монотонную функцию, так что подобный выбор часто осуществляется на основе содержательных соображений. В частности, в страховой практике популярен выбор логарифмической функции $g(\cdot)$, приводящей к зависимости вида

$$y = e^{\mathbf{x}\boldsymbol{\beta} + \varepsilon} = e^{\beta_0} \cdot e^{x_1\beta_1} \cdot \dots \cdot e^{x_k\beta_k} \cdot e^{\varepsilon},$$

где e^{β_0} – оценка премии по базовому договору, а $e^{x_j\beta_j}$ – мультипликативные поправочные коэффициенты, зависящие от группы риска j , к которой принадлежит застрахованный объект. Эта модель хорошо согласуется с практикой ценообразования на страховые продукты и страхового андеррайтинга, что объясняет ее популярность. Тем не менее, данный вариант имеет смысл по построению только при подчинении зависимой переменной логнормальному распределению.

Другой подход – обобщенные линейные модели – предполагает аддитивную ошибку в (1):

$$y_i = g^{-1}(\mathbf{X}\boldsymbol{\beta}) + \eta_i \quad \text{или} \quad E[\tilde{y}_i | \tilde{\mathbf{x}}_i = \mathbf{x}_i] = g^{-1}(\mathbf{x}_i \boldsymbol{\beta}).$$

При этом распределения ошибок выбираются из более широкого класса, а именно – из экспоненциального класса параметрических семейств распределений. Наилучшая оценка коэффициентов регрессии достигается на основе максимизации функции правдоподобия. Функция $g(\cdot)$ часто подбирается таким образом, чтобы, по возможности, упростить решение упомянутой задачи максимизации. Так, при построении обобщенной линейной модели на базе

гамма-распределения (так называемая гамма-регрессия) можно взять логарифмическую функцию с выходом на мультипликативную модель.

Для определения квантили в моделях такого рода следует использовать информацию о предполагаемом распределении. В частности, для гамма-регрессии следует решить уравнение

$$\frac{1}{\Gamma(\alpha)} \int_0^{y_\theta \cdot \alpha \exp(-\mathbf{x}\boldsymbol{\beta})} t^{\alpha-1} e^{-t} dt = \theta \quad (3)$$

относительно y_θ , где α – параметр гамма-распределения.

Альтернативный подход – метод квантильной регрессии, который позволяет непосредственно получить значения квантилей. Это позволяет находить оценки нетто-премий непосредственно, не разбивая данную задачу на две части: оценку ожидаемой нетто-премии и рискованной надбавки. Кроме того, важным преимуществом метода является возможность непосредственного задания значения квантили на уровне, для которого необходимо получить оценки, так как он выступает параметром модели.

Определенным недостатком данного подхода является невозможность оценки математического ожидания зависимой переменной. Тем не менее, ее можно заменить оценкой медианы.

Метод квантильной регрессии не обязательно привязан к линейным зависимостям. Его легко модифицировать путем введения нелинейного преобразования. В частности, введение функции $g(\cdot)$ в соотношение (1) позволит его привести к виду

$$Quant_\theta(y_i | \mathbf{x}_i) = g^{-1}(\mathbf{x}_i \boldsymbol{\beta}_\theta).$$

На основе последней формулы также можно построить мультипликативную модель.

До сих пор обсуждались распределения выплат, построенные на основе статистики прошлых выплат (возможно, с необходимыми поправками на изменение тенденций развития страхового портфеля). Тем не менее, не по всем договорам страхового портфеля будут иметь место страховые выплаты. Менеджеров страховых компаний и актуариев интересует распределение выплат

до наступления страхового случая, а нередко и до заключения договора. Поэтому соответствующее распределение выплат должно иметь скачок в нулевой точке для моделирования отсутствия страхового случая. Иными словами, все рассмотренные методы должны быть соответственно изменены для учета указанной особенности статистики выплат.

В рамках подобной модификации распределение случайной величины $\tilde{y}$ должно рассматриваться как усеченное распределение. Для конструирования «полного» распределения случайной величины Y добавим событие «отсутствие выплат» с вероятностью скачка $p = \Pr[Y = 0]$ и перенормируем полный ансамбль вероятностей. Тогда для нетривиальных значений $y > 0$ получим формулу $F_Y(y) = p + (1 - p) \cdot F_{\tilde{y}}(y)$, которая важна для оценки квантилей

$$y_\theta = F_{\tilde{y}}^{-1}\left(\frac{\theta - p}{1 - p}\right).$$

Таким образом, определение нетто-премии предполагает измерение квантили уровня

$$\theta^* = \frac{\theta - p}{1 - p} \quad (4)$$

случайной величины $\tilde{y}$, оцененной по статистике выплат.

Результатами исследования стали результаты имитационного анализа, который базируется на серии смоделированных выборок данных. Каждая выборка состоит из ста смоделированных убытков. Пусть i – номер выборки, j – номер значения в выборке, тогда y_{ij} – смоделированный убыток. Каждая выборка состоит из "хороших" и "плохих" рисков. Примерно 30 процентов каждой выборки "плохие" риски (с большим значением среднего убытка). "Плохие" риски помечены специальным бинарным индикатором $x_{ij} = 1$, когда бинарный индикатор $x_{ij} = 0$ указывает на "хорошие" риски. Зависимость линейная относительно выбранной модели:

$$y_{ij} = \alpha + \beta x_{ij} + \varepsilon_{ij} \text{ or } Quant_\theta(y_{ij} | x_{ij}) = \alpha + \beta x_{ij} + F^{-1}(\theta).$$

В математических терминах набор данных представляет собой смесь распределений. Все "хорошие" риски – это независимая реализация значений

одной функции распределения, а все "плохие" риски независимая реализация другой функции распределения, того же типа. Эффективный метод оценивания должен разделить между собой "хорошие" и "плохие" риски, и прогнозировать соответствующие квантили достаточно хорошо. Число значений (убытков) в выборке равно 100, количество выборок равно 250.

В исследовании используются четыре различных типа смесей распределения. Первый тип, смесь нормального (Гауссова) распределения $\tilde{y}^{good} \sim N(300;150)$ и $\tilde{y}^{bad} \sim N(450;150)$. Второй тип, смесь Гамма-распределения $\tilde{y}^{good} \sim \text{Gamma}(4;75)$ и $\tilde{y}^{bad} \sim \text{Gamma}(9;50)$. Данные распределения случайных величин имеют одинаковые математические ожидания и дисперсии, с небольшой асимметрией. Таким образом, мы оценим значимость наличия в распределениях асимметрии для оценок. Третий тип, тоже смесь Гамма-распределения, но с гораздо большей дисперсией $\tilde{y}^{good} \sim \text{Gamma}(0.334;900)$ и $\tilde{y}^{bad} \sim \text{Gamma}(0.334;1352)$. Четвертый тип, смесь Парето распределения $\tilde{y}^{good} \sim \text{Pareto}(3;600)$ и $\tilde{y}^{bad} \sim \text{Pareto}(3;900)$. Математическое ожидание и дисперсия последних двух распределений почти одинакова. Такой набор смесей распределений позволит оценить распределения с различной степенью тяжести хвоста.

Цель заключается в оценке точности метода квантильной регрессии с помощью имитационного анализа. В рамках исследования требуется оценить третью квантиль (0,75 стоимости) как оценку нетто-премии. Основные характеристики данных приведены в таблице.

Таблица

Основные характеристики полученных данных

Тип риска	Распределение	Мат. ожидание	Стандартное отклонение	Третья квантиль
“хороший”	Нормальное (Гауссово)	300.0	150.00	401.17
	Гамма с небольшой диспе-	300.0	150.00	383.21

	рсии			
	Гамма с большой дисперсией	300.6	520.13	358.62
	Парето	300.0	519.61	352.44
“плохой”	Нормальное (Гауссово)	450.0	150.00	551.17
	Гамма с небольшой дисперсией	450.0	150.00	540.12
	Гамма с большой дисперсией	451.6	781.36	538.72
	Парето	450.0	779.42	528.66

Полученные данные анализируются с помощью трех регрессионных подходов: метод наименьших квадратов, обобщенная линейная модель и квантильная регрессия. Так как, исходные параметры распределений почти одинаковы, различия в оценках будут объяснены качеством используемых методов оценивания.

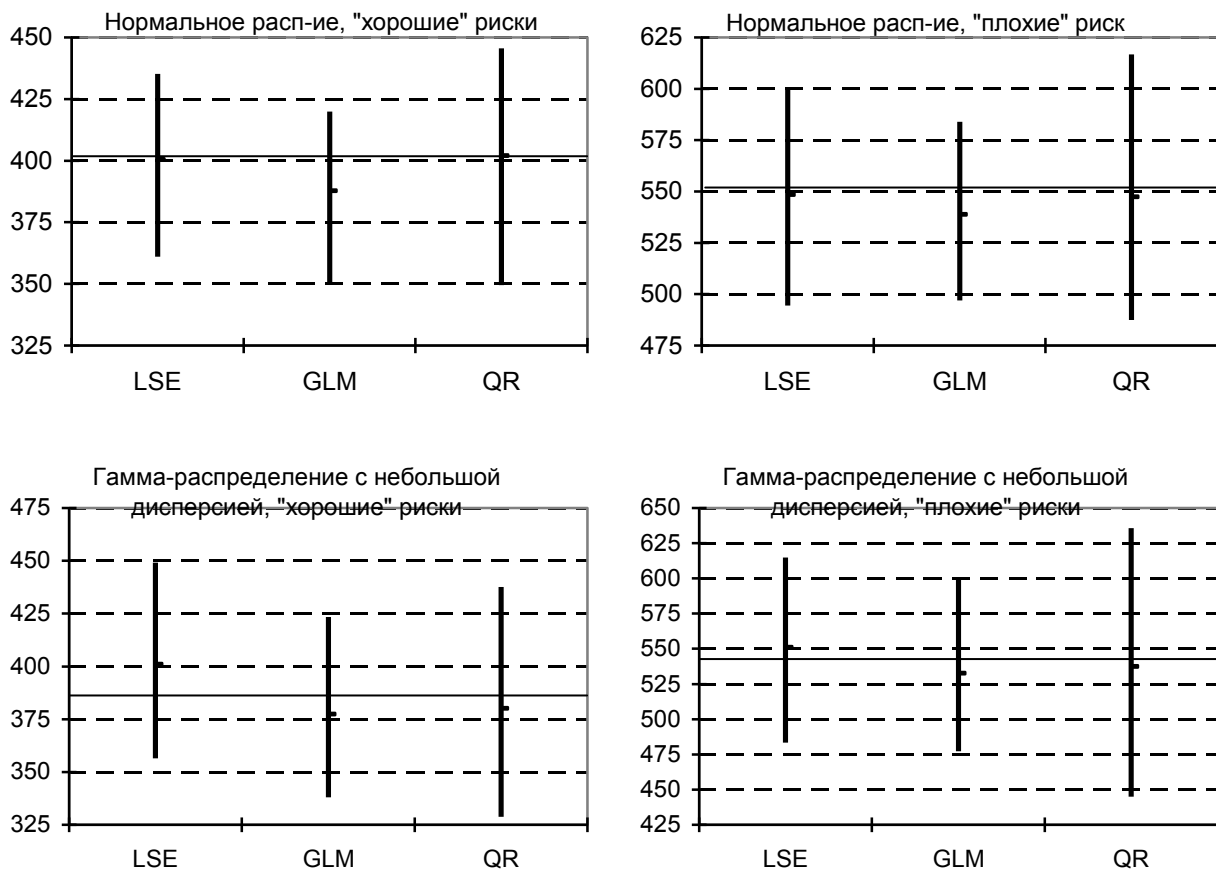
Первые два метода оценивают математическое ожидание, тогда как метод квантильной регрессии может оценить непосредственно третью квартиль. Соответственно оценка третьей квартили, по методу наименьших квадратов и обобщенной линейной модели, должна оцениваться с учетом предположения относительно закона распределения оцениваемых данных. Такой подход соответствует поведению актуария, который не знает точного распределения анализируемых данных и делает предположения относительно используемых методов оценивания.

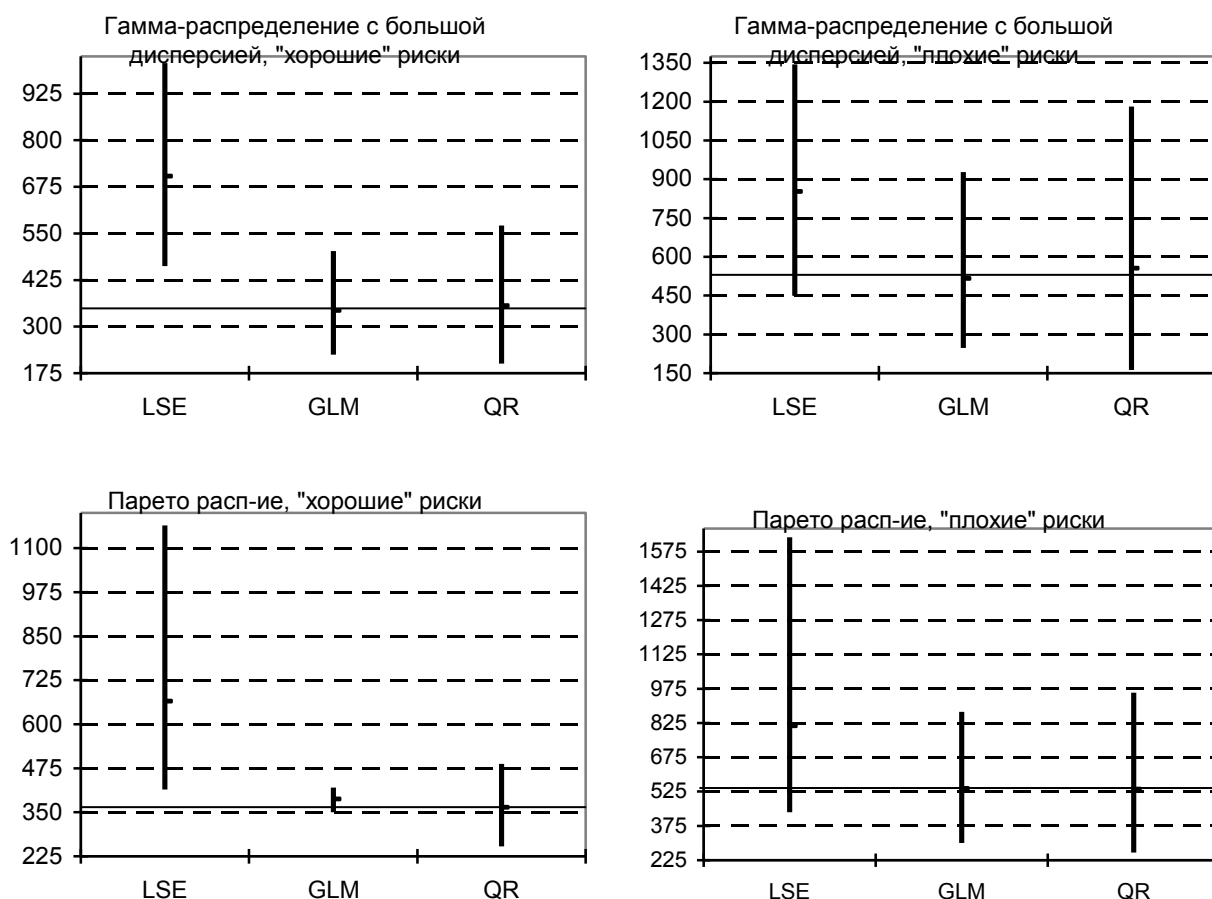
Для метода наименьших квадратов делается предположение о нормальном распределении анализируемых данных. Третья квартиль оценивается по следующей формуле:

$$E[y_{ij}|x_{ij}] + z_{0.75}\sigma^2,$$

где $E[\tilde{y}_{ij}|x_{ij}]$ математическое ожидание, $E[y_{ij}|x_{ij} = 0] = \alpha$ и $E[y_{ij}|x_{ij} = 1] = \alpha + \beta$, $z_{0.75}$ – это третья квартиль стандартного нормального распределения (равная 0,67449), и σ^2 – выборочная оценка стандартного отклонения. Так как Гамма-распределение используется для обобщенных линейных моделей (так называемая гамма регрессия) оценка третьей квартили основывается на функции распределения с соответствующими значениями параметров распределения. Так же, с помощью бутстреп метода были оценены соответствующие доверительные интервалы для сравнения методов.

Результаты моделирования показаны на рисунке.





Оценки и соответствующие им доверительные интервалы

Метод квантильной регрессии немного хуже метода наименьших квадратов для нормального распределения и распределения с небольшой асимметрией, но МНК очевидно хуже для распределений с большей асимметрией. В результате метод квантильной регрессии гораздо лучше в подобных ситуациях.

Обобщенная линейная модель лучше, чем квантильная регрессия для всех экспериментов. Данный вопрос требует дополнительного изучения, хотя есть предположение о недостаточной точности полученных оценок методом обобщенной линейной модели и точности его применения. Кроме того, предложенный в данной работе метод квантильной регрессии свободен от предположения относительно закона распределения исследуемых данных. Это может быть преимуществом, если распределение имеет более тяжелые хвосты, чем в данном исследовании.

Выводы. Научная новизна статьи заключается в использовании метода квантильной регрессии в задаче оценки нетто-премий. Основным преимуществом использования метода квантильной регрессии, по сравнению с классическими методами, является преодоление недостатков классических подходов.

Дополнительно, в рамках данного исследования был проведен имитационный анализ, который для ряда базовых распределений, характерных на практике, позволил сравнить методы между собой и дать ряд практических рекомендаций при решении задачи оценки нетто-премий.

Дальнейшие исследования могут быть направлены на практическую апробацию предложенного метода в рамках поставленной задачи, используя данные страховых компаний, работающих на рынке Украины.

Литература

1. Мак Т. Математика рискованного страхования / Т. Мак. – М.: Олимп-Бизнес, 2005. – 432 с.
2. Abduramanov R. The method of quantile regression, a new approach to actuarial mathematics / R. Abduramanov, A. Kudryavtsev // Insurance: Mathematics and Economics: 11th International Congress, 10-12 July, 2007: book of abstracts. – Piraeus, 2007. – P. 56–57.
3. Anderson D. A Practitioner's Guide to Generalized Linear Models / D. Anderson // CAS Discussion Paper Program. – N.Y.: Wiley, 2004. – P. 1–115.
4. Denuit M. Actuarial theory for dependent risks: measures, orders and models / M. Denuit, J. Dhaene, M. Goovaerts, R. Kaas. – Chichester: Wiley, 2005. – 440 p.
5. Dhaene J., On the structure of premium principles under point wise comonotonicity / J. Dhaene, A. Kukush, M. Pupashenko // Theory of Stochastic Processes. – 2006. – № 3–4. – P. 27–45.
6. Kenneth Q. Z. Direct Use of Regression Quantiles to Construct Confidence Sets in Linear Models / Q. Z. Kenneth, L. P. Stephen // The Annals of Statistics. – 1996. – № 1. – P. 287–306.
7. Koenker R. Regression Quantiles / R. Koenker, G. Bassett Jr. // Econometrica. – 1978. – № 1. – P. 33–50.

8. McCullagh P. Generalized Linear Models / P. McCullagh, J.A. Nelder. – N.Y.: CRC Press, 1989. – 511 p.
9. Huber P. J. Robust Statistics. / P. J. Huber. – N. Y.: Wiley, 1981. – 316 p.