

А. А. Кудрявцев¹

канд. экон. наук, доцент кафедры управления рисками и страхования Санкт-Петербургского государственного университета

АНАЛИЗ ВОЗМОЖНОСТИ ПРИМЕНЕНИЯ НЕПАРАМЕТРИЧЕСКОЙ РЕГРЕССИИ К ПРОГНОЗИРОВАНИЮ РЕЗЕРВА ПРОИЗОШЕДШИХ, НО НЕЗАЯВЛЕННЫХ УБЫТКОВ

Введение

Необходимость резерва произошедших, но незаявленных убытков связана с задержками в информировании страховой компании о произошедших страховых случаях, в связи с чем у нее могут возникать дополнительные (иногда достаточно большие) обязательства после окончания действия соответствующего договора. Поэтому оценка таких резервов является одной из основных проблем управления страховой организацией. Ей уделяется большое внимание на различных уровнях управления — актуарном, бухгалтерском, андеррайтинговом и т. д.

К настоящему времени разработано несколько десятков методик прогнозирования резервов произошедших, но незаявленных убытков. Обзор указанных методик можно найти, например, в монографии Г. Тейлора (Taylor, 2000). Большая часть из них представляет собой непараметрические оценки прироста накопленных убытков по группе договоров, заключенных в один и тот же период времени. Такой подход был заложен еще в 60-х гг. прошлого века (см. Benedikt, 1969; Masterson, 1962) и к сегодняшнему дню является основным в актуарном анализе. Его достоинствами являются относительная простота и гибкость. К недостаткам следует отнести слишком большое число степеней свободы, а также ограниченную возможность учета трендов и изменений бизнес-среды (в частности, влияние изменения поведения страхователей или дизайнера страхового продукта на размер резерва).

В рамках этой парадигмы сложилось стандартное представление исходных данных в виде таблицы, по строкам которой идут периоды возникновения убытков, а по столбцам — периоды урегулирования. Таким образом, календарное время измеряется с левой верхней ячейки до правой нижней. Соответственно наблюдаемые данные располагаются в верхней треугольной части таблицы. Подобное представление данных принято называть «треугольным». В таблице представлены реальные данные о поквартальных выплатах (в млн руб.) одной из страховых компаний Санкт-Петербурга по портфелю каско за 2005—2007 гг.

Стремление преодолеть указанные недостатки привело к разработке альтернативных подходов. В первую очередь в середине 1980-х гг. были предложены

¹ Электронный адрес: kudr@AK1122.spb.edu.

Таблица

Пример традиционного, «треугольного» представления данных

Период возникновения убытков	Период урегулирования										
	1	2	3	4	5	6	7	8	9	10	11
1	30,70	86,98	92,38	93,66	94,06	94,41	96,77	96,79	96,97	97,00	97,03
2	29,79	75,81	80,82	82,27	82,93	83,23	83,28	83,46	83,63	83,70	
3	25,75	78,35	85,68	88,55	89,24	89,77	90,19	90,32	90,38		
4	27,70	71,04	83,96	86,31	88,21	88,69	88,69	88,72			
5	14,25	60,12	69,58	73,71	74,61	75,78	75,84				
6	17,62	53,10	59,02	60,44	61,28	61,73					
7	11,17	45,45	55,20	58,88	61,64						
8	13,82	51,13	63,11	68,24							
9	15,49	53,72	64,40								
10	13,16	72,06									
11	15,48										

параметрические методы (Sherman, 1984). Подход Шермана заключался в подгонке параметрической кривой, позволяющей прогнозировать будущие накопленные убытки. В наиболее развитой форме данный подход был представлен в работе Рамсея (Ramsay, 2005).

На этой основе был создан регрессионный вариант прогнозирования резервов. Он, по-видимому, был впервые предложен Б. Ценвиртом (B. Zehnwirth), но результаты не были им опубликованы. Первые публикации принадлежат группе британских актуариев, силами которых подобные методики сейчас и развиваются (см. England, Verrall, 2002; Renshaw, Verrall, 1998), а также (Taylor, 2000). Суть данного подхода заключается в оценке параметрической формулы типа

$$E[X_{ij} | X_{i1}, \dots, X_{i,j-1}] = g(i, j), \quad (1)$$

где X_{ij} — совокупные убытки, оплаченные (или, как вариант, заявленные) в период урегулирования j по страховым случаям, произошедшим в период i ; $g(i, j)$ — нелинейная параметрическая функция (на практике часто экспонента от линейной или аддитивной функции).

Иными словами, по известным данным («верхнему треугольнику») задается поверхность, отражающая особенности процесса урегулирования. Эта поверхность естественным образом распространяется на ячейки с неизвестными величинами убытков, что и обеспечивает построение прогноза.

Достоинствами такого подхода является упрощение анализа (аналитик имеет дело с небольшим числом параметров заданного класса кривых), а также расширение возможностей для выявления и моделирования трендов и других изменений в данных. Тем не менее к существенным недостаткам следует отнести сложность (а иногда и невозможность) подгонки гладкой кривой к реальным данным. Кроме того, большие сомнения вызывает качество оценки. Действительно, функция $g(i, j)$ подбирается по $n(n+1)/2$ наблюдениям (n — размерность таблицы данных), а прогноз нужен для $n(n-1)/2$ ячеек.

Таким образом, несмотря на большое количество разработанных методов, все они имеют существенные недостатки, снижающие точность прогнозирования. Следовательно, необходимо разрабатывать новые подходы, свободные от упомянутых недостатков. Это объясняет цель данной статьи — анализ возможности применения непараметрической регрессии к прогнозированию резерва произошедших, но незаявленных убытков.

Структура данных и требования к модели непараметрической регрессии

Метод непараметрической регрессии хорошо известен (Györfi et al., 2002; Härdle, 1990). Его суть состоит в поточечном расчете наилучшей линии, описывающей регрессионную закономерность. Результат часто можно представить в виде взвешенной суммы наблюдаемых значений. Различные методы, как правило, отличаются критерием минимизации ошибки и некоторыми другими техническими деталями. Ряд алгоритмов предполагает использование классических регрессионных методов в качестве вспомогательного инструмента. Основное преимущество состоит в том, что не нужно делать предположения о параметрическом классе функций и в большинстве случаев о типе распределения ошибок. При этом подход достаточно гибок, чтобы учесть тренды и иные изменения в статистике. Тем не менее «платой» за отсутствие параметрического класса будет большое число степеней свободы, что можно считать определенным недостатком.

Очевидно, использование данного подхода для прогнозирования величины резервов произошедших, но незаявленных убытков позволяет достичь компромисса между традиционным непараметрическим подходом и параметрической регрессией. С одной стороны, можно построить регрессионную линию, удобную для анализа изменений резерва во времени, с другой — сохраняется традиционный непараметрический подход, в частности использование среднего взвешенного (но с другими весами). При этом обеспечивается достаточно полный учет всех особенностей процесса урегулирования.

Непараметрическая регрессия уже использовалась для оценки резерва произошедших, но незаявленных убытков (Antonio, Beirlant, 2008). Однако в указанной работе непараметрическая регрессия использовалась исключительно как технический инструмент, а не как основное средство прогнозирования. Кроме того, использовался только один подход — метод регрессионных сплайнов без обсуждения возможных альтернатив.

Большое значение для выбора подходящих методов непараметрической регрессии имеет «треугольное» представление данных. В силу того что для актуария важно сравнить различия по периодам возникновения убытков (по строкам таблицы), для которых объем наблюдений не одинаков, имеет смысл рассматривать статистическую конфигурацию как лонгитюдинальную (структурированную во времени). В подобной ситуации данные о прошлых агрегированных убытках задаются в следующей форме:

$$(j, y_{ij}), j = 1, 2, \dots, n - i + 1, i = 1, 2, \dots, n.$$

Кроме того, при выборе методов непараметрической регрессии важно обратить внимание на различия между моделями с фиксированным эффектом и моделями со смешивающим эффектом. В отличие от первых вторые позволяют анализировать различия между периодами возникновения (когортами) убытков, но требуют более значительного времени для получения результата с той же точностью. Тем не менее в моделях со смешивающим эффектом преимущества непараметрической регрессии проявляются сильнее. Модели непараметрической регрессии со смешивающим эффектом для лонгитюдинальных данных можно найти, например, в (Wu, Zhang, 2006).

Базовую модель представить с помощью следующих равенств:

$$y_i(t) = \alpha(t) + \beta_i(t) + \varepsilon_i(t), i = 1, 2, \dots, n, \quad (2)$$

где $y_i(t)$ — кривая развития убытков для i -го периода возникновения убытков; $\alpha(t)$ — функция генерального среднего, отражающая основной тренд; $\beta_i(t)$ —

поправочная функция, специфичная для i -го периода возникновения убытков; $\varepsilon_i(t)$ — функция ошибок.

Для модели со смешивающим эффектом $\beta_i(t) \equiv 0$, так что $\alpha(t)$ иногда называют функцией фиксированного эффекта, а $\beta_i(t)$ — функциями смешивающего эффекта. Хотя актуария интересуют точки $t = j$, можно получить и другие точки кривых, что удобно для проведения анализа.

Предварительный анализ показал, что выбор ядра не имеет большого значения. Это можно объяснить тем, что «опорные» точки $t = j$ расположены достаточно редко. Существенно большее влияние на прогноз имеет выбор самого алгоритма расчетов. В частности, время расчетов может различаться на порядок. Размах доверительного интервала также может отличаться в несколько раз.

Пример расчета

Для иллюстрации применения непараметрической регрессии для прогнозирования резерва произошедших, но незаявленных убытков использованы данные таблицы. Хотя имеется множество различных методов непараметрической регрессии, в данной статье будут рассмотрены только относительно простые методы.

Первый из них — «наивный» метод локального линейного ядра — относится к моделям с фиксированным эффектом. Он определяется следующей формулой (с учетом структуры данных, представленной в предыдущем пункте):

$$\hat{\alpha}(t) = \frac{\sum_{i=1}^n \sum_{j=1}^{n-i+1} K_h(j-t) y_{ij}}{\sum_{i=1}^n \sum_{j=1}^{n-i+1} K_h(j-t)},$$

где $K_h(\cdot)$ — соответствующее ядро.

Данная формула позволяет определить только общий тренд. Расчеты на основе гауссовского ядра дают результат, представленный на рис. 1 (кривая развития убытков показана сплошной линией с выделением целочисленных точек, границы 95%-ного доверительного интервала даны штриховой линией).

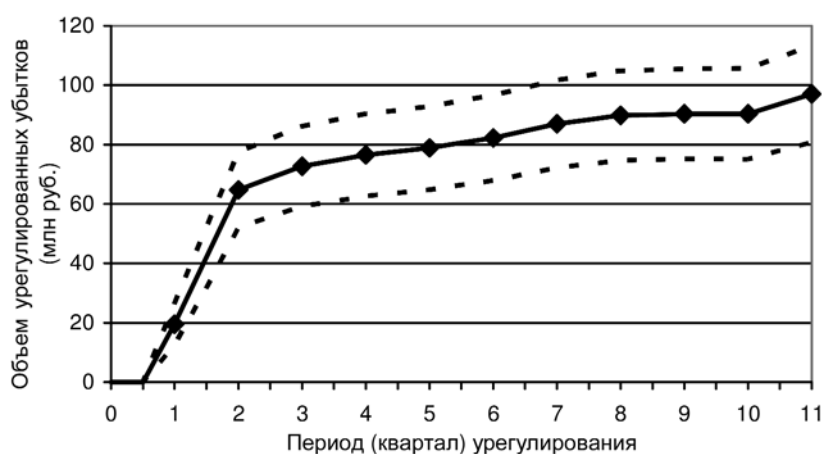


Рис. 1. Кривая развития убытков, полученная «наивным» методом локального линейного ядра

Результат не очень точен: размах доверительного интервала (за исключением первого периода) составляет от 13 до 16,7% от максимального убытка. Тем не менее время расчетов довольно мало.

Даже на примере такой неточной оценки кривой урегулирования видно, что она недостаточно гладкая для использования какой-либо параметрической кривой. Это общее достоинство применения непараметрической регрессии.

В качестве альтернативы представлен другой метод — метод локальной константы на основе смешивающего эффекта. Он позволяет оценить и общий тренд, и поправочные функции, специфичные для периодов возникновения убытков. Метод задается следующим образом:

$$\hat{\alpha}(t) = \frac{\sum_{i=1}^n \sum_{j=1}^{n-i+1} w_{ij}(t) y_{ij}}{\sum_{i=1}^n \sum_{j=1}^{n-i+1} w_{ij}(t)} \quad \text{и} \quad \hat{\beta}_i(t) = \frac{\sum_{j=1}^{n-i+1} w_{ij}(t) y_{ij}}{\sum_{j=1}^{n-i+1} w_{ij}(t)} - \hat{\alpha}(t),$$

где $w_{ij}(t)$ — весовая функция, определяемая формулой

$$w_{ij}(t) = \frac{\delta^2(t) K_h(j-t)/\sigma_j^2}{1 + \delta^2(t) \sum_{k=1}^{n-i+1} K_h(k-t)/\sigma_k^2},$$

$\delta^2(t)$ — межкогортная дисперсионная функция; σ_j^2 — внутрикогортная дисперсионная функция, специфичная для j -го периода возникновения (когорты) убытков.

Результаты расчетов на основе гауссовского ядра приведены на рис. 2.

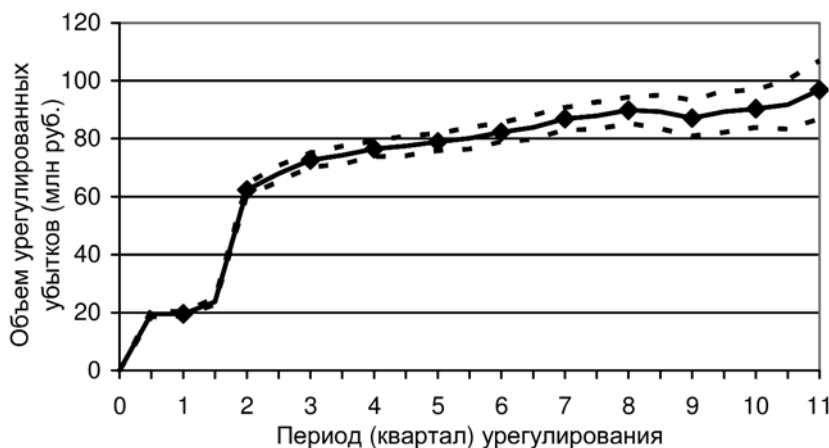


Рис. 2. Кривая развития убытков (общего тренда), полученная методом локальной константы на основе смешивающего эффекта

При анализе сразу бросается в глаза «изгиб» функции в окрестности точки 1,5. Тем не менее актуария интересуют в первую очередь целочисленные точки $t = j$. А они в принципе соответствуют оценкам, полученным предыдущим методом. При этом точность прогноза существенно выше (как правило, полуразмах составляет 1—5% от максимального ущерба).

Кроме того, интересно сравнить различия кривых развития убытков, специфичные для периодов возникновения убытков. Они представлены на рис. 3.

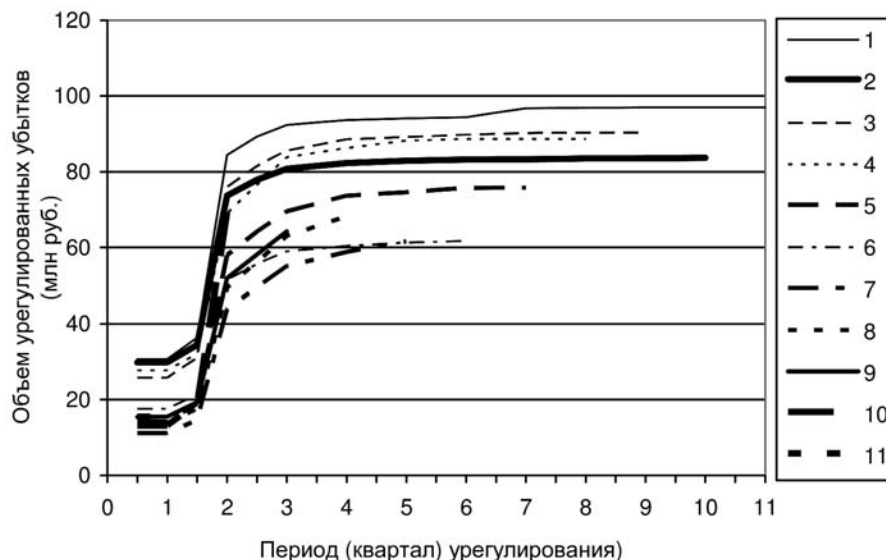


Рис. 3. Специфические кривые развития убытков, полученные методом локальной константы на основе смешивающего эффекта, для различных периодов возникновения (когорт) убытков

Различия между развитием убытков для различных периодов их возникновения выглядят заметными. Это означает, что использование только общего тренда (функции фиксированного эффекта) дает более грубые оценки. Действительно, использование функции фиксированного эффекта для обоих методов дает оценку в 218 млн руб., тогда как применение функций смешивающего эффекта для метода локальной константы позволяет снизить эту величину до 192 млн руб. за счет более точного учета поведения когорт урегулируемых убытков, что составляет примерно 88% от первой цифры.

Выводы

Методы непараметрической регрессии, предлагаемые для оценки резерва произошедших, но незаявленных убытков, представляются определенным компромиссом между двумя существующими группами методик — непараметрическим подходом и параметрической регрессией. Это позволит использовать при анализе преимущества обеих групп.

При построении алгоритма прогнозирования резерва следует учитывать структуру и другие особенности исходных данных. В частности, модель непараметрической регрессии со смешивающим эффектом выглядит предпочтительнее, несмотря на более продолжительное время расчетов.

Предварительное тестирование на реальных данных показывает высокую эффективность предложенного подхода. Непараметрическая регрессия как метод оценки резервов, несомненно, обладает рядом достоинств и большим потенциалом. Тем не менее особенности его применения требуют дополнительных исследований.

Источники

Antonio K., Beirlant J. Issues in claims reserving and credibility: a semiparametric approach with mixed models // *Journal of Risk and Insurance*. 2008. N 3. P. 643–676.

- Benedikt V.* Estimating Incurred Claims // *ASTIN Bulletin*. 1969. Vol. 5. Pt. II. P. 210—212.
- England P. D., Verrall R. J.* Stochastic Claims Reserving in General Insurance // *British Actuarial Journal*. 2002. Vol. 8. N 3. P. 443—518.
- Györfi L.* et al. *A Distribution-Free Theory of Nonparametric Regression*. N. Y., 2002.
- Härdle W.* *Applied Nonparametric Regression*. Cambridge, 1990.
- Masterson N.E. Problems in Motor Insurance — Claim Reserves // *ASTIN Bulletin*. 1962. Vol. 2. Pt. 1. P. 152—160.
- Ramsay C. M.* A new method of estimating loss reserves // *Proceedings of the Casualty Actuarial Society*. 2005. Vol. 92. P. 462—484.
- Renshaw A. E., Verrall R. J.* A stochastic model underlying the chain-ladder technique // *British Actuarial Journal*. 1998. Vol. 4. N 4. P. 903—923.
- Sherman R. E.* Extrapolating, smoothing, and interpolating development factors // *Proceedings of the Casualty Actuarial Society*. 1984. Vol. 71. P. 122—155.
- Taylor G.* *Loss Reserving: An Actuarial Perspective*. Boston, 2000.
- Wu H., Zhang J.-T.* *Nonparametric Regression Methods for Longitudinal Data Analysis: Mixed-Effects Modelling Approach*. Hoboken, N. J., 2006.