Лев Костенко, Руслан Хасанов

Сглаживание таблиц смертности: подбор функциональной зависимости и гладкое локальное сглаживание.

Вероятностная модель оставшейся продолжительности жизни является основным математическим инструментарием актуария, работающего в страховании жизни или пенсионном фонде. Базовым методом для построения такой модели является использование таблицы смертности вкупе с понятием условной вероятности. Под таблицей смертности обычно понимают упорядоченный по возрасту ряд чисел, характеризующий порядок вымирания поколения людей. Данный объект обычно строится на основе статистического материала, а значит, может содержать в себе случайные ошибки, обусловленные процессом сбора и обработки статистических данных. Интуитивно понятно, что вероятности смерти индивида должны быть гладкой функцией от возраста. Для того чтобы не переносить отмеченные выше случайные ошибки на результаты актуарных расчетов – величины страховых премий, пенсионных взносов, страховых резервов и обязательств по выплате пенсий, рекомендуется использовать сглаженные таблицы смертности. Под сглаживанием обычно имеют в виду математические методы построения новой таблицы смертности, задающей близкую к исходной модель оставшейся продолжительности жизни, и такой, что вероятности смерти обладают желаемому свойству гладкости год от года.

На следующей иллюстрации приведен график зависимости вероятностей смерти в течение года в зависимости от возраста, определенных на основе таблицы смертности Российской Федерации 2015 года для мужчин:

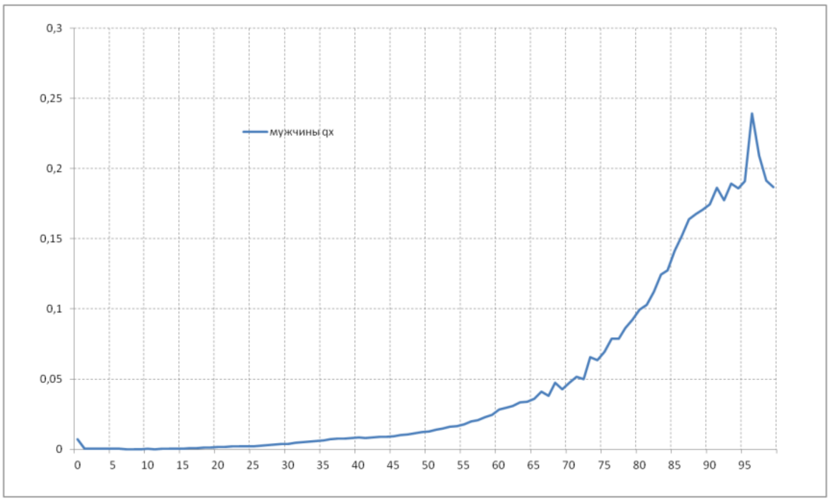

Рис. 1

Помимо явного дефекта негладкости на картинке видно, что для возрастов более 95 лет вероятности смерти начинают убывать, что интуитивно также не представляется корректным.

Данный эффект можно объяснить скудностью статистической информации для людей старших возрастов.

Стандартным методом сглаживания таблиц смертности является выбор аналитической зависимости вероятности смерти от года и дальнейший подбор параметров выбранной функции с целью минимизировать разницу сглаженной и первичной таблиц. Классической здесь является экспоненциальная зависимость $q_x = A + BC^x$, широко известная в литературе как модель Мейкхема. Если осуществлять подбор параметров в данной модели так, чтобы минимизировать сумму квадратов отклонений в целочисленных точках, то получается результат, изображенный на следующем графике.

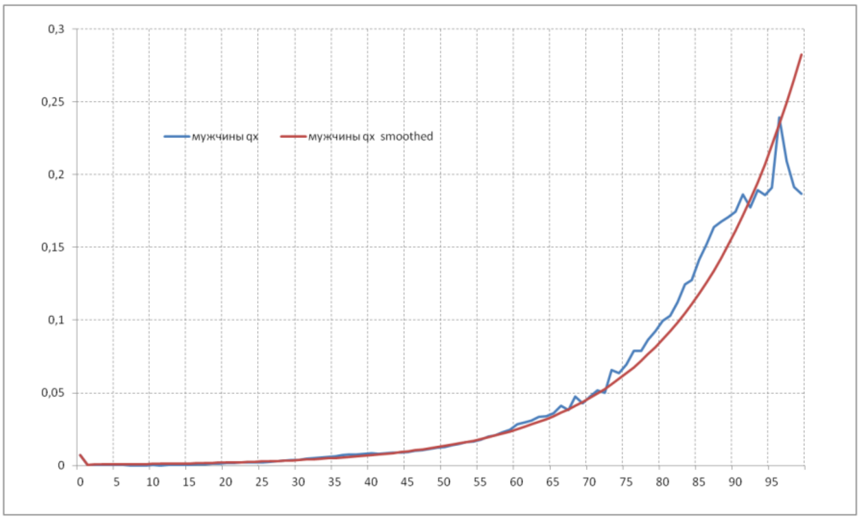

Рис.2

Согласно современным представлениям о характере функциональной зависимости уровней смертности на старших возрастах рост вероятностей смертности замедляется и уже не носит экспоненциальный вид. В связи с этим использование модели Мейкхема не дает качественный результат сглаживания, что хорошо видно на Рисунке 2. Для решения данной проблемы возможно модифицировать указанную модель, предполагая полиномиальную зависимость смертности для старших возрастов и склеивая ее с экспонентой на остальном временной интервале. Чтобы получилась гладкая кривая, значение производных двух функций в точки склеивания должны быть равны. В случае склеивания двух функций итоговая сглаживающая функция будет иметь следующий вид:

$$y = 1(x < x_0) * f_1(a_1, \dots, a_n; x) + 1(x \geq x_0) * f_2(b_1, \dots, b_m; x) \quad (1)$$

Где $1(\mathrm{A})$ принимает значение 1 при выполнении условия «А» и 0 в противном случае.

Для поиска необходимых коэффициентов так же, как и в предыдущем случае, можно применить метод наименьших квадратов, однако в этом случае необходимо рассмотреть дополнительные уравнения связи. Таким образом, для поиска коэффициентов полинома третей степени система функций будет иметь следующий вид:

$$\begin{cases} \sum\left(y_i - b_3x_i^3 - b_2x_i^2 - b_1x_i - b_0\right)^2 \to min \\ \exp(a_1x_0 + a_0) = b_3x_i^3 - b_2x_i^2 - b_1x_i - b_0 \\ a_1 * \exp(a_1x_0 + a_0) = 3b_3x_0^2 - 2b_2x_0 - b_1 \end{cases} \tag{2}$$

Где точка $(x_0; y_0)$ является точкой склеивания. Значение x_0 может выбираться с учетом минимизации суммы квадратов отклонений итоговой функции (1) в целочисленных точках. Результаты применения данного алгоритма изображены на следующей иллюстрации.

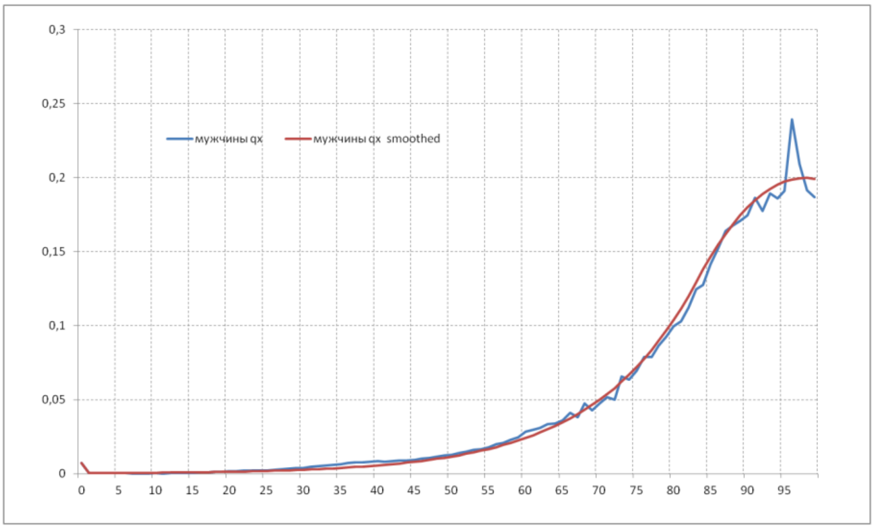

Рис.3

Альтернативой для сглаживания путем подбора функциональной зависимости является локальное сглаживание. Данный метод состоит в том, что сглаженные вероятности смерти определяются как некоторая функция от исходных вероятностей смерти для соседних возрастов. Данная процедура позволяет нивелировать случайные ошибки, содержащиеся в исходных вероятностях, а также избавиться от привязки к конкретному виду предполагаемой функциональной зависимости.

Самым простым методом локального сглаживания является простое локальное усреднение исходных вероятностей. В литературе данный метод известен как «Фильтр 1/n». Здесь *n* - количество соседних точек, участвующих в усреднении. Как правило, данный параметр выбирают нечетным числом, и усреднение проводится симметрично относительно выбранной исходной точки. Результаты применения описанной процедуры в случае *n*=3 (*n* – количество соседних точек, участвующих в усреднении) для таблицы смертности Российской Федерации 2015 года для мужчин приведены на следующей картинке:

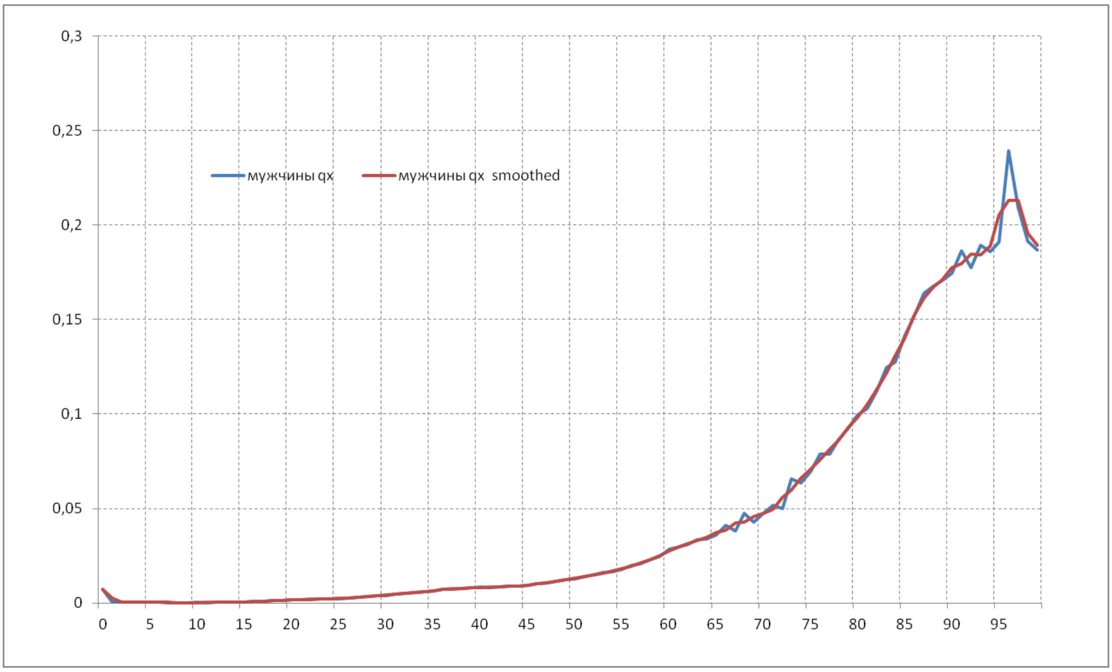

Рис. 4

Отмеченный метод сглаживания прост в практической реализации и хорошо работает на уже достаточно гладких исходных данных. Однако, в случае, если исходные данные содержат значимую статистическую погрешность, то метод приводит к эффекту, известному как “over smoothing” – сглаженный график имеет много перегибов, чтобы «успеть» за исходным.

Если q_x - исходные вероятности смерти индивида возраста $x = 0,\ldots,100$, то сглаженную величину q_x^1, полученную методом простого локального усреднения, можно представить в виде

$$q_x^1 = \sum_{i=0}^{100} l_i(x) q_i \,, \qquad (3)$$

где величины $l_i(x)$ равны $\dfrac{1}{n}$ для $|x - i| \le \dfrac{n-1}{2}$ и нулю иначе. Можно предположить, что если выбирать коэффициенты $l_i(x)$ гладкими, то результат сглаживания также будет гладким.

Классическим в теории локального сглаживания является метод ядерного сглаживания Надарая-Ватсона, согласно которому сглаженные величины q_x представляются в виде (3), где коэффициенты $l_i(x)$ задаются равенством

$$l_i(x) = \frac{K\left(\dfrac{x-i}{h}\right)}{\sum_{j=0}^{100} K\left(\dfrac{x-j}{h}\right)}. \qquad (4)$$

Здесь h – параметр масштаба, а K – некоторая неотрицательная функция, называемая ядром и удовлетворяющая следующим свойствам:

$$\int K(x)dx = 1, \ \int xK(x)dx = 0, \ \int x^2 K(x)dx > 0.$$

Если в качестве функции K взять индикаторную функцию $I(x)$ отрезка [0,1], то получается метод простого локального сглаживания с параметром $2h+1$. Другим широко используемым примером функции K является нормальное (гауссовское) ядро:

$$K(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}.$$

Другими известными ядрами являются ядро Епанечникова $K(x)=\frac{3}{4}(1-x^2)I(x)$, а также трикубическое ядро $K(x)=\frac{70}{81}(1-|x|^3)^3 I(x)$, однако для сглаживания таблиц смертности указанные три ядра дают примерно одинаковые результаты. Параметр h выбирается как значение, минимизирующее функционал

$$r^1(h)=\sum_{i=0}^{100}\left(\frac{q_i^1-q_i}{1-l_i(i)}\right)^2.$$

Если применить описанный метод для гауссовского ядра к уже упомянутой таблице смертности мужчин Российской Федерации 2015 года, то получается следующий результат:

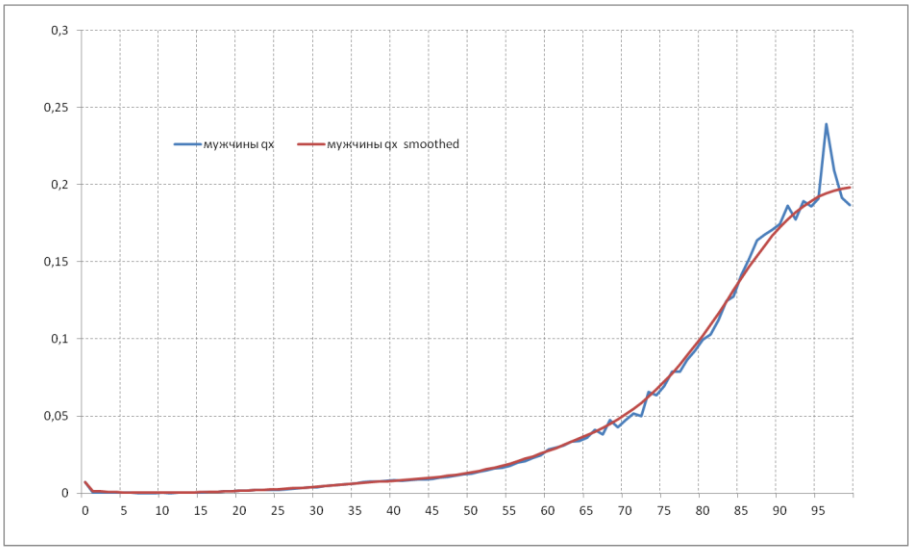

Рис. 5

Зависимость значений функционала $r^1(h)$ имеет следующий вид:

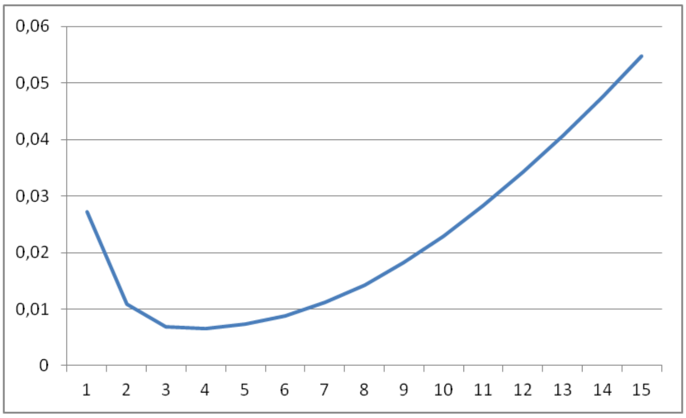

Рис. 6

В данном случае минимум достигается в точке $h=4$. Для задачи сглаживания таблиц смертности параметр h, как правило, принимает значения от 2 до 6.

Выбор метода сглаживания зависит от того, для каких целей будет использоваться сглаженная таблица смертности. В случае использования таблицы смертности для расчета обязательств негосударственного пенсионного фонда более приемлемым можно считать алгоритм, дающий более консервативную таблицу для пенсионных возрастов с наименьшим отклонением от исходной таблицы по выбранному параметру. Для примера ниже приведена сравнительная таблица ожидаемых возрастов остаточной продолжительности жизни для таблицы смертности мужчин Российской Федерации 2015 года, рассчитанных по исходной таблице, а также по сглаженным двумя описанными выше методами таблицам: подбором параметров функциональной зависимости (экспонента + полином) и методом гладкого локально сглаживания с гауссовским ядром:

	Ожидаемый возраст остаточной продолжительности жизни, лет		
возраст	**Исходная таблица смертности**	**Таблица смертности, сглаженная методом подбора функциональной зависимости (экспонента+полином)**	**Таблица смертности, сглаженная методом гладкого локального усреднения**
55	19,08	19,31	18,91
56	18,42	18,63	18,27
57	17,79	17,97	17,64
58	17,16	17,31	17,02
59	16,54	16,67	16,42
60	15,95	16,04	15,83
61	15,40	15,42	15,27
62	14,85	14,82	14,71
63	14,31	14,23	14,17
64	13,79	13,66	13,64
65	13,26	13,09	13,12
66	12,74	12,55	12,61
67	12,27	12,01	12,10
68	11,74	11,50	11,61
69	11,30	10,99	11,12
70	10,78	10,50	10,64