

УДК 51:004(045)

МОДЕЛИРОВАНИЕ ЭКСТРЕМАЛЬНЫХ ПОТЕРЬ В СТРАХОВАНИИ*

Калюжная В.О.,

студентка факультета прикладной математики и информационных технологий, Финансовый университет,
Москва, Россия
leralenina618@yandex.ru

Аннотация. Целью данной работы является изучение основных методик и проблем при моделировании экстремальных величин, которые изучаются в таком разделе статистики, как теория экстремальных значений. В работе представлены теоретические основы, необходимые при моделировании экстремально больших потерь в страховании. Помимо классических методов, которые также могут использоваться при построении и выборе моделей с экстремальными наблюдениями, существует ряд дополнительных оценочных и диагностических процедур. Если основной интерес заключается в изучении максимальных значений за период, то используется обобщенное распределение экстремальных значений и приложимые к нему методики оценки. Если же необходимо изучить хвост распределения, начиная с какого-либо конкретного значения, то используется обобщенное распределение Парето. В таком случае основной задачей является поиск порога – значения, с которого для моделирования можно использовать названное распределение. В работе рассмотрены различные методологии для его поиска, которые могут быть реализованы с помощью программной среды R. На реальных данных продемонстрированы основные процедуры и выявлены возникающие трудности моделирования.

Ключевые слова: теория экстремальных значений; обобщенное распределение экстремальных значений; теорема Фишера-Типпета; обобщенное распределение Парето; выбор порога; оценка Хилла; тяжелые хвосты; метод максимального правдоподобия

MODELING EXTREME LOSSES IN INSURANCE

Kalyuzhnaya V.O.,

student of the Faculty of Applied Mathematics and Information Technology, Financial University,
Moscow, Russia
leralenina618@yandex.ru

Abstract. This article is aimed to study the main techniques and issues while modeling large losses, which are studied in such section of statistics as the extreme value theory. The paper describes the theoretical foundations necessary for modeling the extremely large losses in insurance. Besides the classical methods that can be used to construct and select models with extreme observations, there are a number of additional evaluation and diagnostic procedures. If the primary interest is in studying values' maximum over a period, then a generalized extreme value distribution and applied estimate techniques are used. If it is necessary to study the tail of the distribution, starting from a particular value, then a generalized Pareto distribution is used. In this case, the main point is to determine a threshold – the value from which such distribution can be used for modeling. The paper considers various methodologies for its search, which can be implemented with the RStudio software environment. The basic procedures are demonstrated on real data and the emerging modeling difficulties are identified.

Keywords: extreme value theory; generalized extreme value distribution; the Fisher-Tippett theorem; the generalized Pareto distribution; threshold selection; Hill's estimate; heavy tails; maximum likelihood estimation

Научный руководитель: **Аль-Напор М.С.**, кандидат физико-математических наук, доцент, доцент Департамента анализа данных, принятия решений и финансовых технологий, Финансовый университет, Москва, Россия.

* Статья победителя IX Международного научного студенческого конгресса «Цифровая экономика: новая парадигма развития».

Введение

Статистическая теория экстремальных значений постепенно развивалась с начала XX в. Впервые у В. И. Борткевича (1922 г.) была сформулирована проблема нахождения распределения максимального значения в последовательности случайных величин. Далее существенный вклад был внесен работой R. Fisher и L.H.C. Tippett (1928 г.), где авторы указали принадлежность такого распределения к одному из трех типов. В работах L.H.E. von Mises (1936 г.) и Б. В. Гнеденко (1943 г.) были установлены условия сходимости к этим распределениям [1]. Однако ее неоспоримое значение в приложениях было признано только после книги E.J. Gumbel "Statistics of Extremes" 1957 г., которая и по сей день считается актуальным источником.

Исторически сложилось так, что главную ценность теория экстремальных значений представляла при изучении окружающей среды с расчетами уровней моря, скорости ветра, речного стока и т.д. В настоящее время теория экстремальных значений – одна наиболее важных статистических областей, широко используемая как в технических, так и в финансовых областях, страховании, при оценке риска и пр.

В отличие от выводов, относящихся к средним значениям случайной величины, где большие данные могут рассматриваться как выбросы и часто удаляются, в нашем исследовании именно «отбрасываемая» часть данных является основной и представляет интерес для изучения.

В последние несколько лет различные авторы (P. Embrechts, C. Klippelberg, A.J. McNeil) отметили, что теория имеет непосредственное отношение к моделированию экстремальных страховых потерь [2]. Использование теории экстремальных значений (ТЭЗ) в рамках изучения приложений к страхованию обосновывается тем фактом, что в некоторых случаях актуариев могут интересовать лишь большие потери, такие как при перестраховании [3]. Например, для эксцедентного типа перестрахования страховая компания производит выплаты в полном объеме по всем искам, размеры которых не превосходят некоторую сумму, называемую уровнем удержания, а в случае превышения этого уровня разница погашается

перестраховщиком. Знать форму распределения значений ниже этого нет необходимости, хотя важно учитывать вероятность того, что будет произведен перестраховочный платеж или среднее значение таких выплат.

В рамках ТЭЗ рассматриваются два типа потерь: к первому типу относят наибольшие (наименьшие) потери за определенный период времени (временной блок), ко второму – потери, превышающие определенный порог.

Первый тип представляет большой интерес при управлении операционными рисками, где количественная оценка экстремальных явлений является ключевой задачей. Также распределения экстремальных значений используются при моделировании максимальных значений индексов ценных бумаг, рассматриваемых на определенных промежутках времени.

Второй тип потерь более распространен в актуарной деятельности. Он имеет непосредственное отношение к расчетам в эксцедентном перестраховании.

Метод определения максимумов в заданных временных блоках – традиционный для анализа сезонных данных, например гидрологических, в то время как пороговый метод использует данные более эффективно за счет отсутствия «искусственного» разделения на блоки, и по этой причине чаще применяется во множестве различных приложений [4]. Он более эффективен и потому, что, рассматривая лишь одно максимальное значение в блоке в первом подходе, например в году, не учитываются подобные ему другие необходимые максимумы в этом же блоке. Однако в таком случае может возникнуть зависимость между эксцессами, т.е. нарушение предположения о независимости распределения потерь. Используя методику блочных максимумов, такую проблему можно обойти, разумно распределив данные по группам, что не всегда представляется возможным [5].

С математической точки зрения наблюдаемые потери представляют собой набор независимых одинаково распределенных неотрицательных случайных величин, функция распределения которых известна или же подлежит поиску и оценке. Строго говоря, основная работа ограничивается рассмотрением правого хвоста распределения, а теория экстремальных

значений – основной инструмент для изучения данного вопроса.

Распределения экстремальных значений

К семейству распределений экстремальных значений обычно относят следующие три типа семейств: распределение Гумбеля, Фреше и Вейбулла. Каждое из них появилось независимо от других, однако все они являются представителями более общего семейства – обобщенного распределения экстремальных значений:

$$F(x) = G_{\gamma, \mu, \theta}(x) = \exp \left[- \left(1 + \gamma \frac{x - \mu}{\theta} \right)^{\frac{1}{\gamma}} \right]. \quad (1)$$

Согласно второму замечательному преде-

лу, предел величины $(1 + \gamma x)^{\frac{1}{\gamma}}$ при $\gamma \rightarrow 0$ равен $\exp(x)$. Отсюда ясно, что $G_0(x)$ – это распределение Гумбеля, функция $G_\gamma(x)$ при $\gamma > 0$ имеет форму распределения Фреше, а при $\gamma < 0$ – форму распределения Вейбулла. Параметр γ называется индексом экстремальных значений и играет ключевую роль в поведении правого хвоста.

Основной теоремой в теории экстремальных значений является теорема Фишера-Типпета, главный смысл которой заключается в стремлении максимума нормированных величин к обобщенному распределению экстремальных значений:

$$\left[F \left(\frac{x - b_n}{a_n} \right) \right]^n \rightarrow G(x).$$

Обобщенное распределение Парето

Наряду с вышеописанными распределениями следует рассмотреть обобщенное распределение Парето (ОРП, GPD), тесно связанное с распределениями экстремальных значений. Распределения данного семейства используются для изучения превышений величин над заданным порогом. Обозначив функцию распределения через $W(x)$, связь с обобщенным

распределением экстремальных значений имеет вид:

$$W(x) = 1 + \ln G(x).$$

В семейство обобщенных распределений Парето входят три вида распределений: экспоненциальное, Парето и бета. Обобщенным распределением Парето является семейство распределений, включающее указанные распределения как частные случаи и имеет вид:

$$F(x) = W_{\gamma, \theta}(x) = 1 - \left(1 + \gamma \frac{x}{\theta} \right)^{\frac{1}{\gamma}}. \quad (2)$$

Поскольку предельное значение выражения $\left(1 + \gamma \frac{x}{\theta} \right)^{\frac{1}{\gamma}}$ при $\gamma \rightarrow 0$ это $e^{\frac{x}{\theta}}$, поэтому W_0 –

экспоненциальная функция распределения. Когда значение параметра γ (или, что эквивалентно, параметра α) положительно, функция распределения $W_{\gamma, \theta}(x)$ имеет вид распределения Парето. В зависимости от тяжести хвостов: ОРП является распределением с легким хвостом $\gamma < 0$, тяжелым $\gamma > 0$ и экспоненциальным $\gamma = 0$.

Рассмотрим теперь условное распределение выборки из n величин (где n довольно большое), распределенных по какому-либо закону, превышающих некоторый порог d и называемых эксцессами. Вторая теорема в теории экстремальных значений теорема Пикандса-Балкема-де Хаана гласит, что правый хвост распределения превышений над порогом (эксцессов) сходится по форме к одному из трех видов обобщенного распределения Парето [6].

Оценки в моделях экстремальных значений

Моделирование экстремальных событий обычно требует нахождения компромисса между тем, чтобы быть как можно ближе к асимптотическим результатам, т.е. разумным образом использовать результаты теорем Фишера-Типпета и Пикандса-Балкема-Де Хаана и одновременно пытаться оставить достаточное для моделирования количество данных с целью получения более надежных оценок.

В случае распределения максимума метод максимального правдоподобия позволяет получать наилучшие оценки. При рассмотрении распределения эксцессов до применения метода максимального правдоподобия следует использовать специальные методики для поиска порога, с помощью которых можно достигнуть наилучших результатов в моделировании.

Оценка максимального правдоподобия для параметров экстремальных распределений

Для n наблюдений отрицательная логарифмическая функция правдоподобия имеет вид [6]

$$-l(\gamma, \mu, \theta) = n \ln \theta + (1 + \gamma) \sum_{j=1}^n y_j + \sum_{j=1}^n \exp(-y_j), \quad (3)$$

где

$$y_j = \frac{1}{\gamma} \ln \left(1 + \gamma \frac{x_j - \mu}{\theta} \right).$$

Минимизация (3) может быть произведена любым стандартным способом оптимизации, однако получившаяся функция неустойчива в областях, удаленных от глобального максимума и, следовательно, численная оптимизация может потерпеть неудачу. Поэтому необходимо найти подходящие начальные значения, такие как, например, оценки метода моментов для распределения Гумбеля

$$\theta = \frac{\sqrt{6}\sigma(X)}{\pi}, \quad \mu = E(X) - \gamma\xi,$$

где $\xi = 0,5772157$ – константа Эйлера-Маскерони, или значения, найденные итерационным методом [6]:

$$\hat{\mu} = -\hat{\theta} \ln \left[\frac{1}{n} \sum_{j=1}^n \exp \left(-\frac{x_j}{\hat{\theta}} \right) \right], \quad (4)$$

$$\hat{\theta} = \frac{\sum_{j=1}^n x_j}{n} - \frac{\sum_{j=1}^n x_j \exp \left(-\frac{x_j}{\hat{\theta}} \right)}{\sum_{j=1}^n \exp \left(-\frac{x_j}{\hat{\theta}} \right)}. \quad (5)$$

Поскольку оценка $\hat{\theta}$ (5) не зависит от $\hat{\mu}$, то уравнение для масштаба можно решить итеративно, начав с какого-либо положительного первоначального значения $\hat{\theta}_0$. Затем результат подставляется в (4), чтобы получить $\hat{\mu}$. Полученные оценки параметров $\hat{\theta}$ и $\hat{\mu}$ (вместе с $\gamma = 0$) являются необходимыми стартовыми значениями для численной минимизации отрицательного логарифмического правдоподобия.

Гипотеза о том, равняется ли параметр γ нулю, может быть протестирована с использованием теста отношения правдоподобия. Согласно ему удвоенная разница между функциями правдоподобия следует распределению хи-квадрат с числом степеней свободы, равным разнице между количеством параметров. Тест отношения правдоподобия не учитывает размер выборки, поэтому полезным является критерий Байеса-Шварца (SBC) [5, с. 156], который при ранжировании моделей от более простой к сложной рекомендует сравнивать разницу логарифмических функций правдоподобия со

значением $\frac{r}{2} \ln(n)$, где r – разница в количестве оценочных параметров, n – размер выборки. Таким образом, добавление одного параметра требует увеличения логарифмического правдоподобия на $0,5 \ln(n)$. Для больших размеров выборки требуется большее увеличение, но оно не пропорционально самому размеру выборки.

Оценка максимального правдоподобия для параметров распределения Парето

Отрицательная логарифмическая функция правдоподобия

$$-l(\gamma, \mu, \theta) = n \ln \theta + (1 + \gamma) \sum_{j=1}^n y_j, \quad (6)$$

где

$$y_i = \frac{1}{\gamma} \ln \left(1 + \gamma \frac{x_i - \mu}{\theta} \right).$$

Оценки максимального правдоподобия двух параметров θ и γ (μ обычно задается заранее) обобщенного распределения Парето получаются путем минимизации отрицательной ЛФП относительно θ и γ , начальные значения

для которой находятся обычно методом моментов. Гипотеза о том, что $\gamma = 0$ также может быть формально проверена с использованием теста отношения правдоподобия или критерия Байеса-Шварца.

Оценка параметра формы в распределении Парето. Оценка Хилла

Пусть имеется любое распределение, хвост которого ведет себя согласно Парето распределению, т.е. хвостовая вероятность задается выражением

$\left(\frac{x}{\theta}\right)^{-\alpha}$. Для формирования оценки параметра α

введем значение порога d , после которого хвост эквивалентен распределению Парето. Тогда оценка параметра формы имеет вид

$$\hat{\alpha} = \left(\frac{1}{n} \sum_{j=1}^n \ln \frac{x_j}{d} \right)^{-1}. \quad (7)$$

В случае, когда порог неизвестен, можно воспользоваться построением графика оценки

$$\hat{\alpha}_{Hill}(k) = \left(\frac{1}{k} \sum_{j=1}^k \ln \frac{x_{j,n}}{x_{k+1,n}} \right)^{-1}. \quad (8)$$

Оценка Хилла (8) будет представлять собой последовательность оценок, рассчитанных для каждого k . Ее можно использовать, когда приблизительно выполняются условия $n \rightarrow \infty$ или

$$\frac{k}{n} \rightarrow 0.$$

Относительно выбора числа k существует множество различных методик, которые делятся на два основных направления. Первая группа методов основана на асимптотических предположениях, вторая состоит из предложений практиков. Последние более специфичны по своей природе, но некоторые из них справляются достаточно хорошо. Их иногда называют эвристическими методами.

Для малых значений k дисперсия оценки Хилла относительно высока, а по мере увеличения она спадает и начинается рост смещения. Из полученных соотношений P. Hall и A. H. Welsh был сделан вывод, что главная

задача заключается в нахождении такого значения, при котором балансируются указанные характеристики. Для некоторых параметрических распределений можно найти аналитический вид зависимости оценки смещения и дисперсии от порога [7].

Основным методом для поиска порога, описываемым в рамках асимптотической теории, является метод, основанный на минимизации асимптотической средней квадратичной ошибки оценки (MSE), например хвостового индекса (J. Beirlant), хвостовых вероятностей (P. Hall, I. Weissman, J. Danielsson) или экстремальных квантилей (A. Ferreira) [8, с. 57]. Для выборок конечного объема такие методы бывают порой неэффективны. Формулы поиска оценок в таких методологиях крайне громоздки, по этой причине не будут выноситься на рассмотрение в данной работе. В R встроено несколько функций с различными оценками, которыми автор работы будет пользоваться далее.

Несмотря на наличие серьезной асимптотической теории, в экономической практике часто прибегают и к эвристическим правилам. Они основаны на нахождении той области графика, где «стабилизируется» оценка Хилла, и значение k выбирается как наибольшее значение наблюдения (т.е. наименьшее значение порога) при ее стабилизации. Метод нахождения устойчивой области на графике путем наблюдения называется «техникой глазного яблока» (от англ. Eye-Ball method).

Самой простой методикой можно назвать поиск такого значения k , для которого в заданном окне последующие оценки $\hat{\alpha}(k)$ будут принадлежать определенному коридору.

Другая, чуть более усложненная методика, описывается в [6] и имеет вид

$$k_{eye}^* = \min \left\{ k \in \{2, \dots, n-w\} \mid h < \frac{1}{w} \sum_{i=1}^w I \left\{ \begin{array}{l} \hat{\alpha}(k+i) < \\ < \hat{\alpha}(k) \pm \varepsilon \end{array} \right\} \right\},$$

где w — это размер движущегося окна, который обычно составляет 1% от длины выборки. Параметр ε задает разрешенный диапазон границ для $\hat{\alpha}(k)$, в пределах которого должны находиться все $[\hat{\alpha}(k+1), \dots, \hat{\alpha}(k+w)]$. Не менее $h\%$ оценок должно быть в пределах диапазона для получения k , который будет рас-

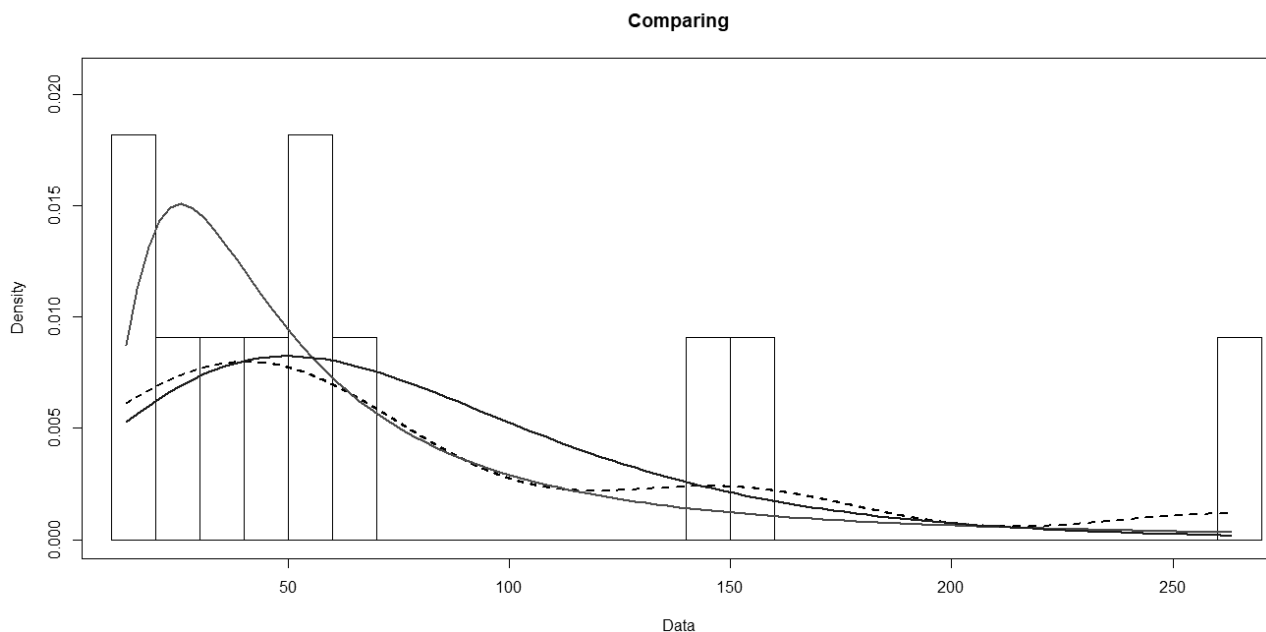


Рис. 1. Сравнение экстремальных моделей

Источник: построено автором.

смагиваться в качестве возможного. Здесь h обычно составляет около 90% и ε выбирается равным 0,3.

Эвристические правила просты в применении, но несколько произвольны, что ставит под сомнение адекватность их реального использования.

График средних избыточных потерь

Полезной графической процедурой при моделировании правого хвоста является построение графика средних избыточных потерь

$$\widehat{e}(d) = \frac{\sum_{j=1}^n (x_j - d)_+}{\sum_{j=1}^n I_{\{x_j > d\}}} \quad (9)$$

и является суммой всех превышений над заданным порогом, деленной на их количество.

Интерпретация графика средних избыточных потерь подробно разбирается у J. Beirlant, P. Vynckier, J. L. Teugels [9], P. Embrechts, R. V. Hogg и S. A. Klugman. Кратко для положительного параметра формы график должен выглядеть как линейно возрастающая функция, для нулевого — как горизонтальная прямая, для

отрицательного — как убывающая прямая. Использовать любые численные оценки данного графика нецелесообразно.

Применение экстремального анализа к данным

Для работы с экстремальным анализом были выбраны данные о потерях при пожарах, зафиксированных датской страховой компанией, состоящие из 2167 наблюдений. Для использования обобщенного распределения экстремальных значений из данных рассматриваются лишь блочные максимумы. В данном случае в качестве блоков были выбраны года, поэтому окончательная выборка состоит из 11 элементов, на основе которых и будет проводиться анализ для получения наилучшей модели.

Рассмотрим возникающие различия оптимизации в двух методах определения начальных значений. Для обобщенной модели оптимизация логарифмической функции правдоподобия для обеих методик дает практически идентичные результаты: количество итераций составило 36 и 37 шагов для каждого из методов соответственно.

В случае, если будет выбрана модель Гумбеля, результаты оптимизации для двух методик сильно отличаются скоростью сходимости:

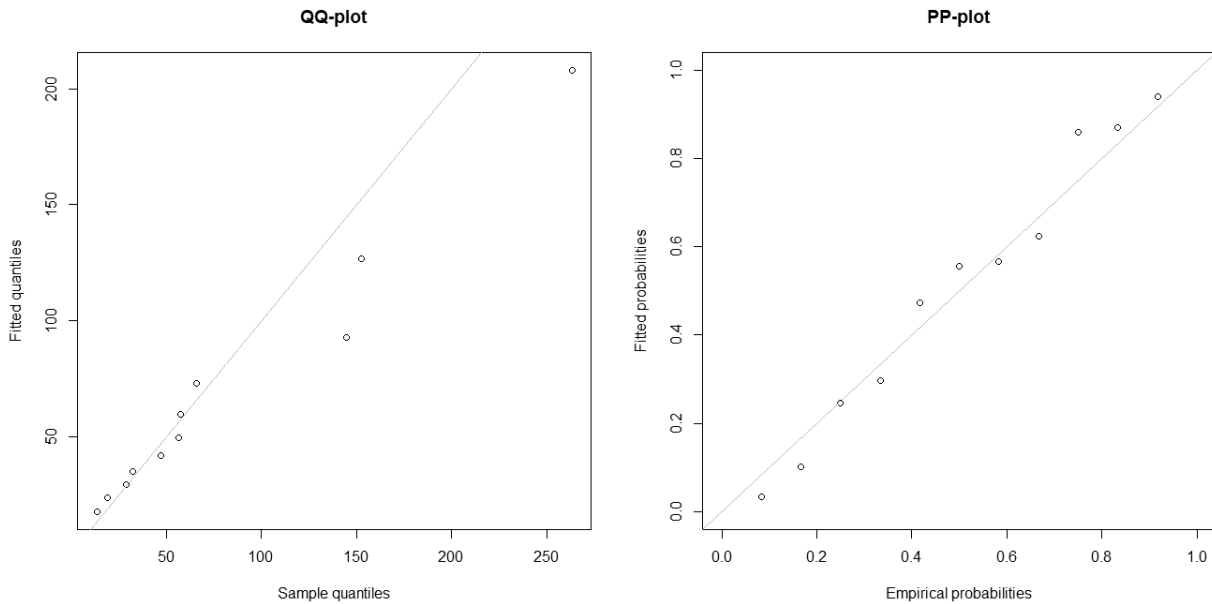


Рис. 2. Проверка качества модели, основанной на обобщенном распределении экстремальных значений

Источник: построено автором.

в первом случае итераций понадобилось 10, во втором – 637, что говорит о неэффективности итеративного поиска для распределений с параметром формы, близким к нулю.

По результатам экспериментов на смоделированных данных для метода 2 также выявлено, что скорость достижения минимума для небольших выборок и положительного параметра формы у него ниже, а для больших выборок и при отрицательном параметре формы – гораздо выше.

Изобразим графически (рис. 1) полученные результаты с учетом двух вариантов параметров формы (пунктирная линия означает эмпирическую плотность, светлая – плотность обобщенного распределения экстремальных значений с параметром формы 0,6384, более темная – плотность распределения Гумбеля, т.е. с параметром формы 0).

Метод отношения правдоподобия говорит о необходимости выбора трехпараметрической модели. Этот же вывод подтверждается результатом выбора между двумя моделями с различным количеством параметров критерием Байеса-Шварца.

В качестве проверки качества модели критерий Колмогорова-Смирнова и критерий Андерсона-Дарлинга показали себя неэффективно ввиду малого объема выборки, поэтому для

проверки адекватности используются графики: qq-сравнение эмпирических и теоретических квантилей и pp-сравнение теоретических и эмпирических вероятностей (рис. 2).

По графикам видно, что модель достаточно хорошо описывает реальные данные. Итак, для максимальных годовых выплат по пожарам получена модель с параметрами (округление до двух знаков после запятой) $\mu = 37,79$, $\theta = 28,94$, $\gamma = 0,64$:

$$F(x) = \exp \left[- \left(1 + 0,64 \frac{x - 37,79}{28,94} \right)^{-\frac{1}{0,64}} \right]. \quad (10)$$

На основе модели можно получать среднее значение максимальных годовых выплат по пожарам, вероятности их наступления и другие статистически важные показатели, что позволит актуарию рассчитывать премии и резервы, учитывая данные, которые ранее представляли собой случайные выбросы и могли бы серьезно повлиять на финансовую стабильность страховой компании.

Для работы со вторым направлением в экстремальном анализе методикой пороговых значений будет использоваться исходный набор данных.

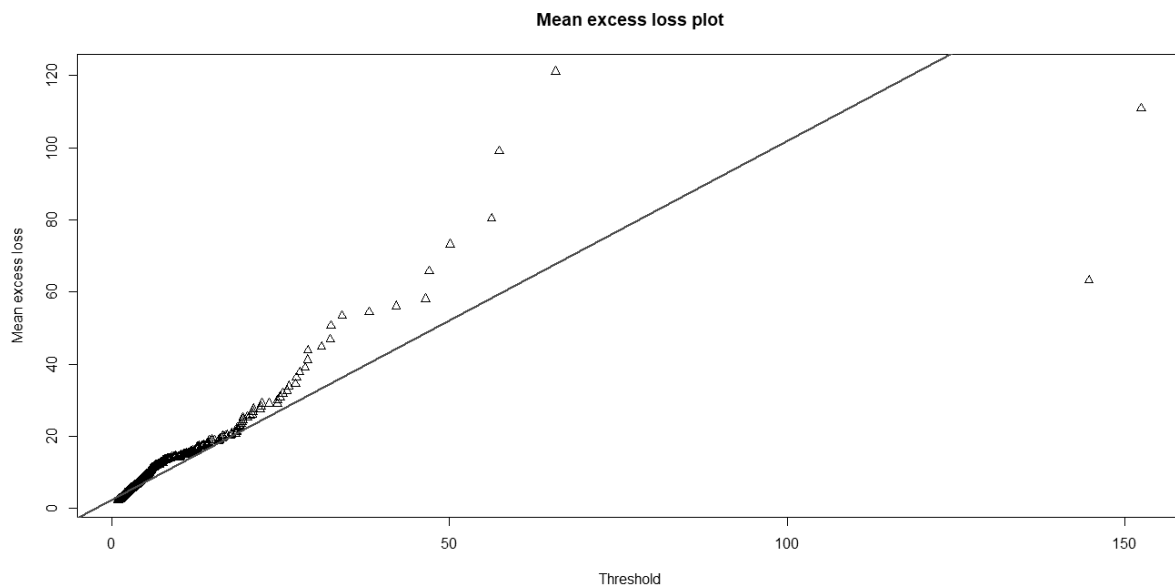


Рис. 3. График средних избыточных потерь

Источник: построено автором.

Для начала анализа с использованием обобщенного распределения Парето (ОРП), рассмотрим график средних избыточных потерь (рис. 3).

По наклону графика можно судить о положительности параметра формы, т.е. о тяжести хвостов распределения. За исключением па-

ры-тройки крупнейших значений можно четко наблюдать прямую линию, которая говорит о том, что ОРП может обеспечить разумное соответствие большей части набора данных, возможно даже всех (проверка этого предположения будет приведена далее). По графику средних избыточных потерь непросто опреде-

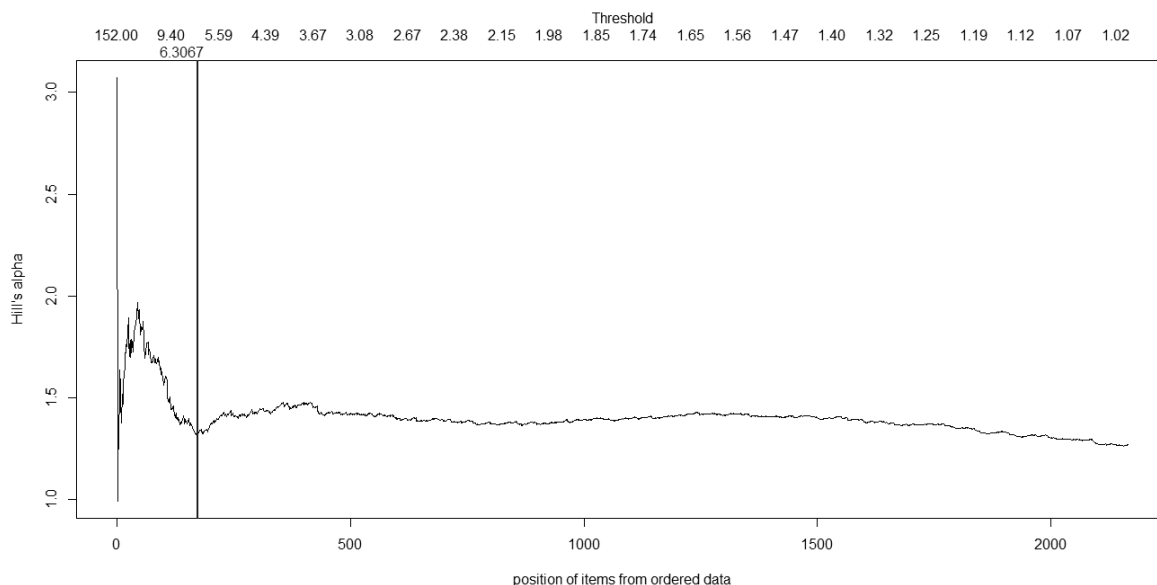


Рис. 4. График оценки Хилла для коридора длиной 21 и шириной 0,03

Источник: построено автором.

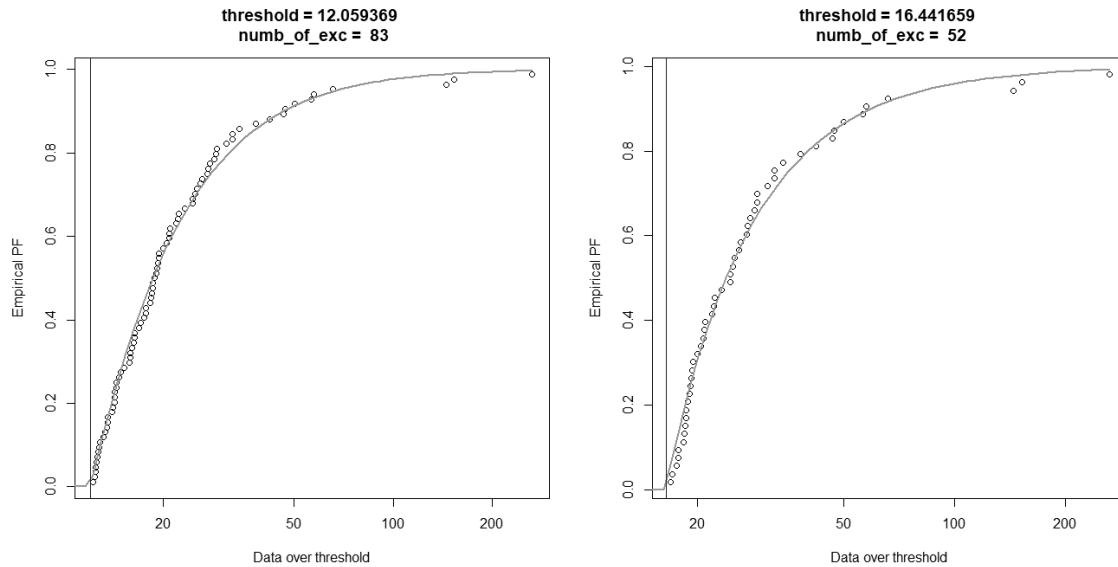


Рис. 5. Сравнение ЭФР и модельной ФР для разных значений порога

Источник: построено автором.

лечь конкретное значение порога, но можно предположить, что он будет ниже 20.

С целью получения конкретного значения порога построим график оценки Хилла (рис. 4), которая является самой популярной техникой в процессе нахождения необходимой величины.

Поиск подходящего порога будет производиться с использованием оценок, кратко описанных выше и реализованных в пакетах `laeken` и `tea`. Для сравнения получившихся порогов используется графический анализ модели и некоторые тестовые характеристики. В качестве предполагаемых порогов будут рассмотрены 6 различных величин.

Первые применяемые значения порогов найдены с помощью двух основных методик в рамках асимптотического и эвристического подходов: на основе смещения (A Bias-based procedure for Choosing the Optimal Threshold) согласно A. Guillou и P. Hall и Eye-Ball method из J. Danielsson, M. Ergun и L. de Haan в [7]. Рассмотрим наглядное представление сравнения эмпирической и моделируемой ФР (рис. 5) для найденных порогов.

Будут рассматриваться и другие варианты, такие как: отсутствие порога – 0; порог, найденный упрощенной методикой определения с коридором длиной 21 и шириной 0,03; по-

роги, найденные с помощью функции `dAMSE` (minimizing the AMSE criterion with respect to k) по алгоритму, изложенному F. Caeiro и M. I. Gomes в [10] и `minAMSE` (Weighted asymptotic mean squared error estimator) согласно J. Beirlant, P. Vynckier, J. L. Teugels [9, с. 296]. Визуализация сравнения функций распределения отображена на рис. 6.

По графикам совершенно точно можно исключить нулевой порог, что говорит о том, что все данные не подчинены распределению Парето, в то время как график средних избыточных потерь выдвинул предположение о возможном нулевом пороге по причине, вероятно, большого количества мало отличающихся друг от друга небольших значений выборки. Вопрос о значении, после которого хвост наилучшим образом описывается ОРП, остается открытым. По оставшимся графикам тяжело понять необходимую величину, однако стоит отметить, что слишком высокий или низкий порог порождают очень маленький или большой объем набора превышений, что может не лучшим образом сказаться на дальнейшем моделировании.

Для окончательного выбора произведем расчет статистик и p -значений для критериев Колмогорова-Смирнова и Андерсона-Дарлингга:

	Thr = 0	Thr = 2,9565	Thr = 6,3067	Thr = 12,0594	Thr = 16,4417	Thr = 23,2839
KS_stat	0,3124	0,0299	0,0350	0,0572	0,0868	0,1036
KS_p	0,0000	0,7157	0,9833	0,9486	0,7959	0,9052
AD_stat	208,3141	0,5213	0,2979	0,3095	0,4600	0,3336
AD_p	0,0000	0,7251	0,9396	0,9307	0,7871	0,9102

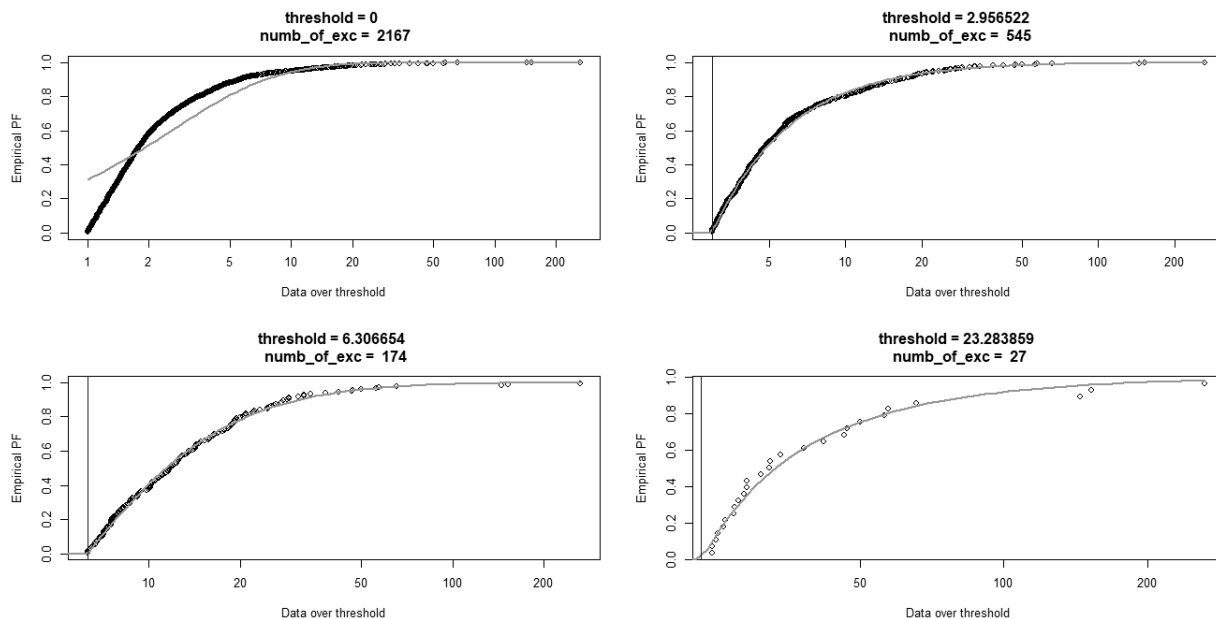


Рис. 6. Сравнение ЭФР и модельной ФР для разных значений порога

Источник: построено автором.

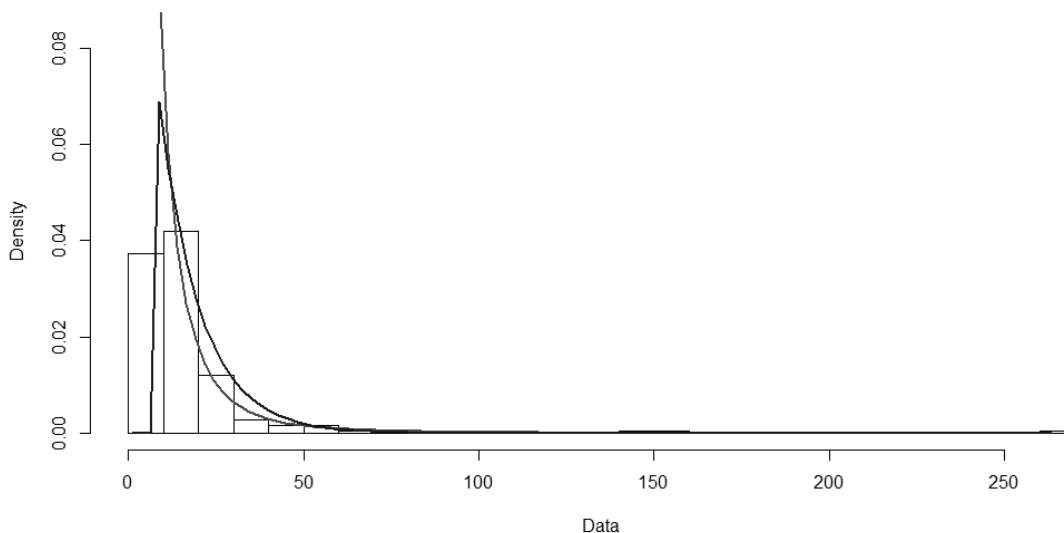


Рис. 7. Сравнение моделей, основанных на обобщенном распределении Парето и экспоненциальном распределении

Источник: построено автором.

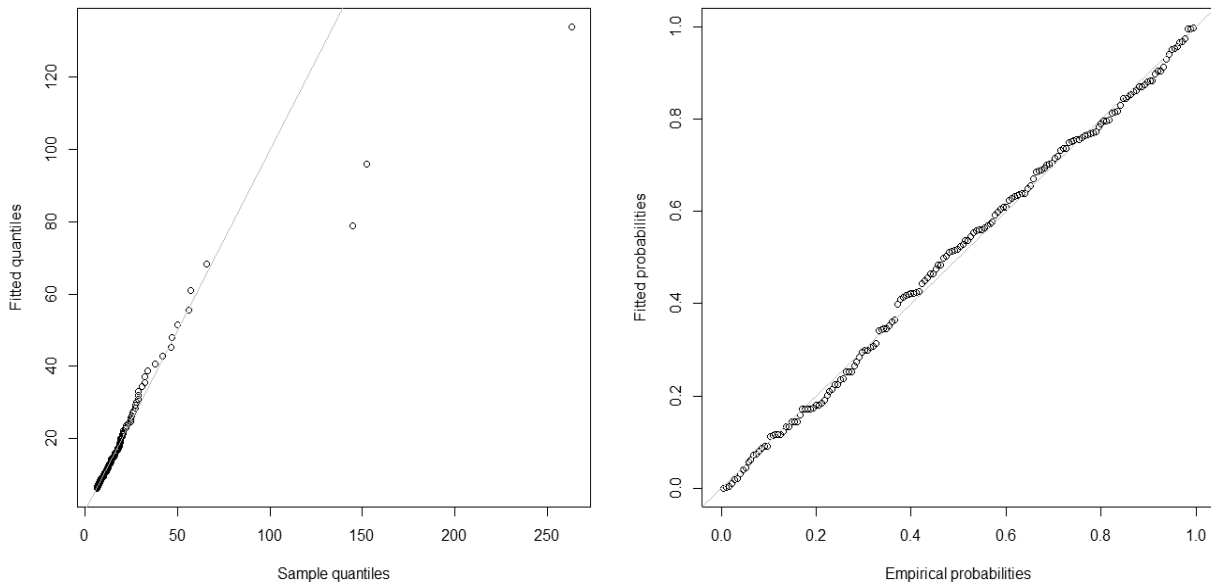


Рис. 8. Проверка качества выбранной модели, основанной на обобщенном распределении Парето

Источник: построено автором.

Следует найти порог, для которого тестовые статистики наименьшие, а р-значения наибольшие. Согласно тесту Колмогорова-Смирнова такие условия выполняются для 6,3067 и 12,0594, что подтверждается и тестом Андерсона-Дарлинга. В результате в качестве наилучшего выбирается порог, равный 6,3067, который был найден с помощью наиболее простой эвристической методикой. Данное значение используется для дальнейшего оценивания параметров модели.

Получим оценки для двух случаев: положительного и нулевого параметров формы, и далее сравним модели (рис. 7).

Тест отношения правдоподобия и Байеса-Шварца указали на необходимость учитывать параметр формы в модели.

Для проверки качества также воспользуемся (помимо тестовых статистик) графиками квантилей и вероятностями (рис. 8), которые наглядно демонстрируют адекватность модели.

Значит, с учетом порога 6,31 получаем модель, которая может использоваться для анализа хвостовых вероятностей, средних значений превышений и др.:

$$F(x) = 1 - \left(1 + 0,45 \frac{x - 6,31}{6,25} \right)^{-\frac{1}{0,45}}. \quad (11)$$

Заключение

С помощью теории экстремальных значений, метода максимального правдоподобия и других специальных методик есть возможность создавать модели, которые позволяют учитывать нередко возникающие в страховой сфере экстремально большие потери. При их использовании следует учитывать обнаруженные проблемы и сложности.

В процессе изучения моделирования максимумов были выявлены проблемы оптимизации функции правдоподобия: в зависимости от предполагаемой модели, объема данных или тяжести хвоста оптимизация проходит с различной скоростью для двух рассмотренных вариантов задания начальных значений.

При исследовании эксцессов или превышений над порогом было показано, что модели с порогами, найденными согласно сложным методикам, основанным на серьезной асимптотической теории, далеко не всегда превосходят в качестве те, что получены интуитивно понятными графическими способами.

Учитывая вышеуказанные особенности моделирования, в ходе работы построены модели, благодаря которым можно получать более полную информацию о потерях для корректного определения актуарием необходимых страховых премий и резервов.

Список источников

1. Джонсон Н.Л., Коц С., Балакришнан Н. Теория вероятностных распределений. Ч. 2. Одномерные непрерывные распределения: пер. с англ. О.И. Волковой, М.С. Стригуновой, В. Кокотушкиным. М.: БИНОМ; 2012. 600 с.
2. *Krivavych Y.* Large Loss Distributions: probabilistic properties, EVT tools, maximum entropy characterization; 2000. 20 p.
3. *Embrechts P.* Modelling Extremal Events for Insurance and Finance. P. Embrechts, C. Kluppelberg, T. Mikosch. Springer-Verlag; 2008. 650 p.
4. *McNeil A.J.* Estimating the Tails of Loss Severity Distributions using Extreme Value Theory. Alexander John McNeil. Astin Bulletin; 1997. 21 p.
5. *Charpentier A.* Computational Actuarial Science with R. A. Charpentier, John M. Chambers, Duncan Temple Lang, Torsten Hothorn, Hadley Wickham. New York: Chapman and Hall/CRC; 2014. 656 p.
6. *Klugman S.A.* Loss Models: From Data to Decisions. Stuart A. Klugman, Harry H. Panjer, Gordon E. Willmot. 3rd ed. A John Wiley & Sons, Inc., Publication; 1998. 783 p.
7. *Danielsson J.* Tail Index Estimation: Quantile Driven Threshold Selection. Jon Danielsson, Lerby M. Ergun, Laurens de Haan, Casper G. de Vries. 2016. 74 p.
8. *Bader B.* Automated, Efficient, and Practical Extreme Value Analysis with Environmental Applications. Brian Bader; 2016. 174 p.
9. *Beirlant J.* Excess functions and estimation of the extreme-value index. J. Beirlant, P. Vynckier, J. Teugels. *Bernoulli*. 1996;2(4):293–318.
10. *Caeiro F.* Direct reduction of bias of the classical hill estimator. Frederico Caeiro, M. Ivette Gomes, Dinis Pestana. *Revstat. Statistical Journal*. 2005;(2):113–136.