ASTIN BULLETIN, Vol. 29, No. 1, May 1999, 101–163.
Halmstad Prize 2001 for the best contribution to actuarial literature in 1999
Hachemeister Prize 2000 of the CAS for the best 1999 ASTIN paper
First version: Mai 21, 1997; 1. rev. April 25, 1998; 2. rev. April 30, 1999

ESTIMATING THE VALUE OF THE WINCAT COUPONS OF THE WINTERTHUR INSURANCE CONVERTIBLE BOND: A STUDY OF THE MODEL RISK*

UWE SCHMOCK

Mathematical Finance
Department of Mathematics, ETH Zürich
CH-8092 Zürich, Switzerland
schmock@math.ethz.ch

ABSTRACT. The three annual $2\frac{1}{4}\%$ interest coupons of the Winterthur Insurance convertible bond (face value CHF 4 700) will only be paid out if during their corresponding observation periods no major storm or hail storm on one single day damages at least 6 000 motor vehicles insured with Winterthur Insurance. Data for events, where storm or hail damaged more than 1 000 insured vehicles, are available for the last ten years. Using a constant-parameter model, the estimated discounted value of the three WINCAT coupons together is CHF 263.29. A conservative evaluation, which accounts for the standard deviation of the estimate, gives a coupon value of CHF 238.25. However, fitting models which admit a trend or a change-point, leads to substantially higher knock-out probabilities of the coupons. The estimated discounted values of the coupons can drop below the above conservative value; a conservative evaluation as above leads to substantially lower values. Hence, already the model uncertainty is higher than the standard deviations of the used estimators. This shows the dominance of the model risk. Consistency, dispersion, robustness and sensitivity of the models are analysed by a simulation study.

CONTENTS

1. Introduction 2
2. Presentation and discussion of the data 6
3. A critical review of a binomial model 8
4. Composite Poisson models with constant parameter 9
4.1. Bernoulli distribution for the knock-out events 11
4.2. Pareto distribution for the knock-out events 13
4.3. Generalised Pareto distribution for the knock-out events 18
5. Testing the constant-parameter Poisson model 21
5.1. Testing for over-dispersion 21
5.2. Testing for time-inhomogeneity 23

1991 *Mathematics Subject Classification.* 62P05 (primary); 90A09 (secondary).

Key words and phrases. WINCAT coupon, Winterthur Insurance, catastrophe bond, storm, hail, model risk, (generalised) Pareto distribution, generalised extreme value distribution, composite Poisson model, generalised linear model, change-point, peaks over threshold.

*Financial support by Crédit Suisse is gratefully acknowledged.

6. Fitting a generalised extreme value distribution 24
7. Composite Poisson models with a time-dependent parameter 26
7.1. Linear trend of the parameter 28
7.2. Log-linear trend of the parameter 30
7.3. Square-root linear trend of the parameter 32
7.4. Modified-linear trend of the parameter 33
7.5. Smooth transition of the parameter 33
8. Testing for a positive trend in the Poisson parameter 35
9. Composite Poisson model with a change-point 38
10. Peaks-over-threshold method 40
11. Comparison of the estimated values 44
12. Model robustness and sensitivity analysis 46
13. Model consistency and dispersion analysis 47
Acknowledgements 53
References 53

1. Introduction

The Swiss insurance company Winterthur Insurance has launched a three-year subordinated 2¼% convertible bond with so-called WINCAT coupons, where CAT is an abbreviation for catastrophe. This bond with a face value of CHF 4 700 may be converted into five Winterthur Insurance registered shares[1] at maturity (European-style option) between the 18th and 24th of February 2000. The annual interest coupon of 2¼% will *not* be paid out if on any one calendar day during the corresponding observation period for the coupon at least 6 000 motor vehicles insured with Winterthur in Switzerland are damaged by hail or storm (wind speeds of at least 75 km/h). If the number of insured motor vehicles changes by more than 10%, then the knock-out limit of 6 000 claims will be adjusted correspondingly.

Had Winterthur launched an identical fixed-rate convertible bond, then, according to Crédit Suisse First Boston's brochure [3], the coupon rate would have been around 0.76% lower (approximately 1.49%). In other words, the investor receives an annual yield premium of 0.76% for bearing a portion of Winterthur's damage-to-vehicles risk. This convertible bond is intended as an instrument to diversify portfolios. The WINCAT coupons are very suitable for this purpose, because storm and hail damages have only a very small correlation with traditional financial market risk. The European-style conversion right, however, strongly ties the bond to the financial market. It is the intention of Winterthur Insurance to test the Swiss capital market for such products, make investors acquainted with them, and obtain a partial reinsurance through the financial market by securitizing a portion of its damage-to-vehicles risk.

Within the range of designs of catastrophe bonds, the Winterthur Insurance convertible bond with WINCAT coupons "Hail" belongs to the more conservative ones, namely the principal-protected catastrophe bonds. Besides the pure catastrophe

[1] Due to the merger of Winterthur Insurance and Crédit Suisse Group on December 15th, 1997, the bond may be converted into 36.5 Crédit Suisse Group registered shares at maturity. Due to the conversion right and the rising market value of the Winterthur Insurance registered shares (see [20]), the convertible bond offered a good investment opportunity during its first few months.

bonds, where the coupons and the principal are at risk, another more conservative variant are the deferred catastrophe bonds, where no payment as such is at risk, but the payments may be deferred. This gives the issuer of such a bond an interest-free credit in case of a catastrophe.

Two guiding principles for specifying the conditions of the WINCAT coupons were simplicity and absence of moral hazard. For the purpose of reinsurance, it would have been interesting for Winterthur Insurance to include a knock-out limit connected to the total number of claims during an observation period. To reduce moral hazard, damage arising from a natural cause was chosen as the triggering event, and the knock-out limit is tied to the number of claims and not to the capital necessary to pay full indemnity to the insured. If an event with at least 6 000 claims occurs, then Winterthur Insurance saves the corresponding $2\frac{1}{4}\%$ coupon interest payment on 399.5 million Swiss francs, which makes CHF 8 988 750 at the corresponding coupon date. On the other hand, according to Winterthur Insurance, CHF 3 000 have to be paid out per claim on the average for motor vehicles damaged by storm or hail. Therefore, when an event with at least 6 000 claims occurs, Winterthur Insurance can expect to save up to 50 % by means of the WINCAT coupons—a profit from a knock-out event seems extremely unlikely. A possible problem with the knock-out limit can be borderline cases of events with about 6 000 claims when a few insured do not know the exact date of the damage (because they have been on holiday, for example). A way to moderate the severity of such a problem would be a linear reduction of the coupon interest rate from $2\frac{1}{4}\%$ to 0 % between 5 000 and 7 000 claims. However, such a specification would make the product more complex and the statistical analysis for the coupon pricing even more involved.

This study was made possible by the willingness of Winterthur Insurance to collect and *publish* the relevant available historical data on the web page [19] as well as in Crédit Suisse First Boston's brochure [3] and thereby to set standards in product transparency, fairness of pricing and investor education. This enables a scientific discussion of such products and their corresponding pricing methodologies, which in turn helps to enhance transparency and acceptance of such products. To satisfy this aim and to build up the confidence of investors, the various sources of risk of such new products should be made explicit to avoid unpleasant surprises. The present paper seeks to make a contribution in this direction with emphasis on education. Since convertible bonds are well-established securities in the market, a lot of information concerning Winterthur Insurance is contained in the legally binding prospectus [18], which helps the investor to judge the default risk and the possible profits from the European-style conversion right. However, no information (other than the exact legal specification) for estimating the knock-out probability of the WINCAT coupons is given in this legally binding prospectus; in particular, there is no historical data on the subject in the prospectus. Apparently, Winterthur Insurance and Crédit Suisse First Boston have been aware of this deficiency; hence their decision to publish [3] and to make the historical data available on the web page [19].

This paper will focus on estimating the risk arising from the WINCAT coupons, with emphasis on the model risk which is not addressed in [3]. For a discussion of the various disguises of model risk, we refer to [4]. Based on the available historical data, we shall present and work out several models and calculate the discounted value of the WINCAT coupons in every case for an easy comparison of the various results. For the pricing of the European-style option for converting the bond into

Year	Date	Event	Number of claims	Vehicles insured index	Adjusted claims
1987				1.248	
1988				1.204	
1989				1.161	
1990	27. Feb.	Storm	1 646	1.127	1 855
	30. June	Hail	1 395		1 572
1991	23. June	Hail	1 333	1.104	1 472
	6. July	Hail	1 114		1 230
1992	21. July	Hail	8 798	1.098	**9 660**
	31. July	Hail	1 085		1 191
	20. Aug.	Hail	1 253		1 376
	21. Aug.	Hail	1 733		1 903
1993	5. July	Hail	6 589	1.099	**7 241**
1994	2. June	Hail	4 802	1.086	5 215
	24. June	Hail	940		1 021
	18. July	Hail	992		1 077
	6. Aug.	Hail	2 460		2 672
	10. Aug.	Hail	2 820		3 063
1995	26. Jan.	Storm	1 167	1.067	1 245
	2. July	Hail	1 290		1 376
1996	20. June	Hail	1 262	1.000	1 262

TABLE 1.1. Claim numbers of past events causing over 1 000 adjusted claims as provided in [3] and [19]. During 1987–1989, such events did not occur. Since the number of motor vehicles insured with Winterthur tends to increase, former actual claim numbers are set into relation with the number of insured vehicles to obtain the number of adjusted claims.

Winterthur Insurance registered shares, we refer to [3]. We should mention here, that the current value of the call option depends on the knock-out probability of the last coupon, because the exercise price of the call option is either CHF 4 805.75 (face value of the bond plus last coupon), if the last coupon is paid, or simply the face value of CHF 4 700, if the last coupon is knocked out.

To estimate the risk of the WINCAT coupons, a 10-year history of damage claims is provided in [3] and [19], see Table 1.1. During this period, a total of 17 events with more than 1 000 damaged vehicles were registered. Of these events, 15 happened during the summer and two were winter storms. None of these events occurred between 1987 and 1989. Only two of the events, which happened on the 21^{st} of July 1992 and the 5^{th} of July 1993, caused at least 6 000 claims. Without any sophisticated modelling, this suggests a knock-out probability of 20%, i. e., the expectation of the annual coupon payment would be 80% of the $2\frac{1}{4}$% WINCAT coupon, which is an expected annual yield of 1.8%. Of course, as mentioned in [3, p. 11], this estimate has little statistical significance.

In Section 2 of this note, we present and briefly discuss the available historical data. Section 3 contains a critical review of a simple binomial model. In Section 4 we give a review of the constant-intensity model to estimate the discounted value of the WINCAT coupons. We discuss several distributions which can be used to obtain an estimate for the probability that an event causing more than 1 000 adjusted claims actually leads to the knock-out of the coupon. These distributions include the Bernoulli distribution, the Pareto distribution (used in [3]) and finally, as suggested by extreme value theory, the generalised Pareto distribution. According to [18] and Winterthur's web page [19], the length of the observation period for the first coupon is not an entire year as assumed in [3]; therefore we recalculate the discounted value of the WINCAT coupons also for the cases already considered in [3]. In Section 5 we test the constant-parameter model with respect to over-dispersion and time-inhomogeneity. Since the historical data set is small, we can calculate the corresponding probabilities under the null hypothesis exactly and do not need to utilise asymptotic results for these tests.

Only the most severe event within an observation period matters for the possible knock-out of a WINCAT coupon. In Section 6 we therefore fit a generalised extreme value distribution to the observed yearly maxima.

In Section 7 we present and discuss various models with a time-dependent parameter for the number of events with more than 1 000 adjusted claims. We shall give several reasons why there might be a trend in the data. An investor, who wants to take a possible trend into account, might use one of these models to estimate the discounted value of the WINCAT coupons. Alternatively, an investor, who prefers a constant-parameter model, can use one of the trend models to create a stress scenario for risk management. These trend models will lead to substantially lower estimates for the values of the WINCAT coupons. In the subsequent section we apply a permutation test to most of the trend models to test the null hypothesis, that there is no trend, and we explain why a permutation test is not adequate for the remaining model with a square-root linear trend.

In contrast to the continuous-trend models, there can also be a sudden change in the expected event frequency. Such a change-point model is presented in Section 9.

The composite Poisson models discussed in Sections 4–9 make use of the assumption that the event frequency is independent from the event severity, namely the adjusted numbers of claims arising from these events. The corresponding trend and change-point models take only a varying event frequency into account. The peaks-over-threshold method from extreme value theory, which we use in Section 10, provides a convenient way to model a possible trend in the event frequency as well as in the event severity. However, when choosing only one additional parameter for the time-inhomogeneous extension of the peaks-over-threshold model, then those two trends are coupled.

A short discussion of the various values of the WINCAT coupons is given in Section 11; see Table 11.1 for a comparison. The substantially different values indicate that the model uncertainty is the dominating risk for the evaluation of the discounted value of the WINCAT coupons.

In Section 12 we use a scenario technique to investigate the robustness and sensitivity of the various models with respect to new data. This is done by adding fictitious data for the year 1997 to Table 1.1, namely no event for a favourable scenario or a repetition of the four events from 1992 for a stress scenario. The

corresponding changes of the estimated coupon values are given in Table 12.1 for the models under consideration.[2]

In the last section we check the consistency of the models and investigate the dispersion of the estimated discounted coupon values by a simulation study. For every fitted model—under the assumption that it describes reality correctly—we generate 1 000 new random data sets according to the distribution specified by the fitted model. These data sets replace the actual observations recorded in Table 1.1, and we use the model to estimate the discounted coupon values based on the random data set. In this way we can check whether the model can recover its own features from the simulated data—in particular the mean and the median—and we can see how far the simulated coupon values deviate from the mean. This can help to determine conservative estimates of the coupon values for the models. The mean, the median, the standard deviation and the 15.9%-quantil for the various models are listed in Table 12.1. Instructive are also the histograms in Figures 13.1 and 13.2, showing the distributions of the estimated coupon values for some selected models.

If the knock-out probability P_{CAT} for the WINCAT coupons were known exactly, then a very small risk premium for the investor would suffice, because the investor has the freedom to invest only a small fraction of the capital in the Winterthur Insurance convertible bond thereby diversifying the risk. This small risk premium is the motivation for insurance companies to securitize their catastrophe risk. However, the true knock-out probability P_{CAT} is not known. Therefore, at various places in this paper, we follow the procedure used in [3] and add an estimated standard deviation $\hat{\sigma}(P_{\text{CAT}})$ to the estimated knock-out probability P_{CAT} to obtain a conservative upper estimate, thereby adding a risk premium for the investor to account for the uncertainty of P_{CAT}. We could elaborate on this point by using the entire estimated distribution of P_{CAT} and tilt it towards higher values (the paper [17] by G. G. Venter is interesting in this context). Taking investor-dependent utility functions and the current market price of risk into account, a more profound analysis might be possible than the one sketched above. However, since the estimated knock-out probabilities and the corresponding standard deviations will vary substantially with the models used, the model risk should also be taken into account, because it seems to be the dominating one in the present problem. There should be a coherent way to calculate an adequate risk premium which accounts for the variation of the estimated knock-out probability *and* the corresponding model risk. We leave it to future research to develop a rigorous mathematical basis for this purpose and to apply it to the present problem.

2. Presentation and discussion of the data

Whether a WINCAT coupon is paid on February 28$^{\text{th}}$ depends on the events happening during the corresponding observation period. These observation periods are specified on Winterthur's web page [19], see Table 2.1. The first observation period is shorter than a year so that there are always four months left between the end of the observation period and the coupon payment date. This provides enough time

[2]Actually, one hail storm in the area Entlebuch/Sarnen with 1 825 claims was recorded on the 11$^{\text{th}}$ of June 1997. Another hail storm hit the town of Lucerne on the 21$^{\text{th}}$ of July 1998 and caused 3 085 claims. There were no other events with more than 1 000 claims during the years 1997 and 1998. In this paper, however, we only use the information available at the time the bond was issued.

Coupon date	Relevant observation period
February 28, 1998	February 28, 1997 – October 31, 1997
February 28, 1999	November 1, 1997 – October 31, 1998
February 28, 2000	November 1, 1998 – October 31, 1999

TABLE 2.1. Observation periods for the WINCAT coupons according to [18] and the web page [19] of Winterthur Insurance.

to count the number of claims and to determine whether the corresponding coupon is knocked out. In the 10-year history of damage claims provided in [3] and [19], see Table 1.1, two events are not within the period from February 28$^{\text{th}}$ to October 31$^{\text{st}}$. This is relevant for the first coupon, we shall therefore always reduce the knock-out probability for the first coupon in a deterministic way (see Table 3.2) using

$$P_{\text{CAT}} = 1 - (1 - \tilde{P}_{\text{CAT}})^{15/17}, \tag{2.1}$$

where $\tilde{P}_{\text{CAT}}$ denotes here the knock-out probability if the observation period were a full year. Formula (2.1) is motivated by the Poisson models used in in following sections. It corresponds to reducing the Poisson parameter by the factor $^{15}/_{17}$, see the discussion in the introduction of Section 4 and the one of formula (4.9). By using (2.1), we neglect the fact that the number of events not occurring in the period from February 28$^{\text{th}}$ to October 31$^{\text{st}}$ is random as well. This simplification, however, is suggested by the lack of data and can be justified by the small influence of this $^{15}/_{17}$-correction (CHF 2.21 for $\tilde{P}_{\text{CAT}} = 20\%$, for example) when compared with the model uncertainty to be discussed. Furthermore, when analysing the adjusted claim numbers, we assume that the two numbers arising from the winter storms come from the same underlying distribution as the numbers arising from the hail storms. Again, this simplifying assumption is suggested by the small historical data set.

The number of claims arising from damage by storm or hail have to be set into relation with the number of vehicles insured with Winterthur in Switzerland. The statistical basis is 773 600 insured risks per year in 1996 or 744 764 insured motor vehicles on April 1$^{\text{st}}$, 1996. Note that many motor-cycles are only insured during the summer months. The above numbers include the motor vehicles insured with Neuenburger Schweizerische Allgemeine Versicherungsgesellschaft, which merged with Winterthur Insurance in 1997. The column *Vehicles insured index* in Table 1.1 gives the number of insured risks in 1996 divided by the number of insured risks for the respective year. The column *Adjusted claims* in Table 1.1 contains the claim numbers multiplied with the insured-vehicles index. Only events with more than 1 000 adjusted claims are shown in Table 1.1, because other historical data is not provided by Winterthur Insurance.

If the statistical basis changes by more than 10%, based on the number of insured motor vehicles on April 1$^{\text{st}}$, then, according to [18, Condition 2(e)], the knock-out limit of 6 000 claims will be adjusted accordingly, rounded to the nearest multiple of 100 claims. As the column *Vehicles insured index* of Table 1.1 shows, the statistical basis tends to increase, but it seems unlikely that it reaches the adjustment trigger of 10% within three years without a merger with another insurance company. Apparently, such a scenario slightly increases the risk of the investor. On the other hand, there was a recent change in the Swiss legislation concerning the mandatory

motor vehicle insurance, and new competitors are becoming active in the motor vehicles insurance market. Therefore, it is not clear whether a rising trend in the statistical basis will persist. For the further analysis in this paper, we assume that the statistical basis stays constant. It should be kept in mind however, that (depending on the model) the estimated coupon values in Table 11.1 can change by up to CHF 10 if the statistical basis changes by as much as $\pm 10\%$ already in the first observation period.

3. A critical review of a binomial model

To extract the relevant information from the historical data given in Table 1.1, we could use a simple model consisting of ten Bernoulli random variables X_{1987}, $X_{1988}, \ldots, X_{1996}$, where $X_y = 1$ means that an event with at least 6 000 adjusted claims happened in the observation period ending at October 31st of the year y. We set $X_y = 0$ otherwise. For the model we assume that these ten random variables are independent and identically distributed. We are interested in estimating the probability $p = \mathbb{P}(X_y = 1)$. An unbiased estimator of p is the empirical mean

$$\hat{p} = \frac{1}{10} \sum_{y=1987}^{1996} X_y. \tag{3.1}$$

The data of Table 1.1 leads to $\hat{p} = 0.2$, because there were two observation periods out of ten where an event with at least 6 000 adjusted claims happened. Using coupon knock-out probabilities of $P_{\text{CAT}}(1997) = 1 - (1 - 0.2)^{15/17} \approx 0.179$ for the first observation period and $P_{\text{CAT}}(1998) = P_{\text{CAT}}(1999) = 0.2$ for the following two years, and using the interest rate structure of Table 3.1, the discounted value of the three WINCAT coupons is calculated in Table 3.2.

Of course, the estimator in (3.1) can only lead to one of the eleven values in the set $\{0.0, 0.1, 0.2, \ldots, 0.9, 1.0\}$. Hence, to be realistic, we should not favour any specific value within the interval $[0.15, 0.25]$. A recalculation of Table 3.2 with the knock-out probabilities 15% and 25% gives CHF 259.08 and CHF 229.78, respectively, for the discounted value of the three WINCAT coupons.

From a statistical point of view we should also consider the standard deviation of the estimator in (3.1). This will give an impression of the quality of the estimator. Since the variance is given by

$$\sigma^2(\hat{p}) = \text{Var}\left(\frac{1}{10} \sum_{y=1987}^{1996} X_y\right) = \frac{p(1-p)}{10}, \tag{3.2}$$

we could follow statistical practice and use the estimated value $\hat{p} = 0.2$ for p to

Coupon	Interest rate	Discount factor
1.	1.87%	0.9816
2.	2.33%	0.9550
3.	2.57%	0.9267

TABLE 3.1. Assumptions regarding the interest-rate structure taken from [3]. The interest rates correspond to the zero-coupon yield on Swiss Confederation bonds plus a spread of 35 basis points.

Coupon	Principle	Interest	Discount factor	P_{CAT}	Value
1.	4 700	2 1/4 %	0.9816	17.9 %	CHF 85.25
2.	4 700	2 1/4 %	0.9550	20 %	CHF 80.79
3.	4 700	2 1/4 %	0.9267	20 %	CHF 78.40
			Discounted value of the three WINCAT coupons:		CHF 244.44

TABLE 3.2. Calculation of the discounted value of the three WINCAT coupons for the estimate $\hat{p} = 0.2$. The three discount factors are taken from Table 3.1. The product of the principle, the coupon interest rate and the discount factor is multiplied with the probability $(1 - P_{\text{CAT}})$ that the corresponding coupon is not knocked out. The $^{15}/_{17}$-correction according to (2.1) was applied to the knock-out probability of the first coupon to take care of its shorter observation period given in Table 2.1.

obtain an estimate for the variance $\sigma^2(\hat{p})$. This would mean to use $\hat{p}(1-\hat{p})/10$ as the estimator. In this binomial model, however, a short calculation shows that

$$\mathbb{E}\left[\frac{\hat{p}(1-\hat{p})}{10}\right] = \frac{10-1}{10} \cdot \frac{p(1-p)}{10},$$

hence we would underestimate the variance in (3.2) by a factor 9/10. Therefore, we estimate the variance of $\hat{p}$ by the unbiased estimator

$$\hat{\sigma}^2(\hat{p}) = \frac{10}{9} \cdot \frac{\hat{p}(1-\hat{p})}{10} \tag{3.3}$$

to obtain $\hat{\sigma}(\hat{p}) = \sqrt{0.2 \cdot 0.8/9} \approx 0.13$ for the standard deviation of $\hat{p}$. For a conservative estimate we may use $P_{\text{CAT}} = \hat{p} + \hat{\sigma}(\hat{p}) \approx 0.33$ as the knock-out probability. A recalculation of Table 3.2 with this knock-out probability leads to CHF 205.24.[3]

The empirical mean in (3.1) is a minimal sufficient estimator for the knock-out probability p *in this model* [10, Chapter 1, Problem 17], hence we have done our best within this model. We cannot expect more from this model, because it uses the data of Table 1.1 very inefficiently. Already in the first step, the data is reduced to ten yes/no decisions (10 bit of information). By taking the mean in (3.1), this information is further reduced by ignoring the order of the ten yes/no decisions, leading to one out of eleven possible numbers. This is less then 4 bit of information. Having gone through this bottleneck, not much can be done with a statistical examination afterwards.

4. Composite Poisson models with constant parameter

To extract more data from Table 1.1 than in the previous section, we shall review several composite Poisson models. The one in Subsection 4.2 was used for the analysis in [3]. For every calendar day in an observation period there is a slight chance of a major storm or hail storm causing more than 1 000 adjusted claims. The data of Table 1.1 as well as common knowledge suggest that this slight chance varies with the season: In Switzerland, a storm is more likely to occur in late

[3]All numerical calculations for this paper were done with the software package *Mathematica*. Only rounded numbers are given in the text, but for subsequent calculations machine precision of the numbers is used. Values in Swiss francs are given up to $^1/_{100}$, although not all given digits are necessarily significant.

autumn or winter than in any other season while hail storms usually occur in summer. If the dependence between the different days is sufficiently weak, then the Poisson limit theorem suggests that a Poisson random variable might be a good approximation for the number of those events within an observation period, which cause more than 1 000 adjusted claims. Note that Table 1.1 records hail storms for August 20$^{\text{th}}$, 1992, and the following day, hence the assumption of "sufficiently weak dependence" has to be kept in mind. Such two-day events can arise artificially from a single storm due to the dividing line at midnight, or they can arise due to weather conditions favouring a hail storm on two consecutive days. The use of a compound Poisson model however, which allows us to model such two-day events conveniently, does not seem to be appropriate here, because a single observation is not sufficient for a reliable estimate of the corresponding parameter. Concerning Poisson approximation, we refer to Barbour, Holst and Janson [1].

The seasonal dependence mentioned above is also the reason why we have chosen the exponent $^{15}/_{17}$ in the correction formula (2.1). We think that this exponent based on the available data is more appropriate than the exponent $^{2}/_{3}$ based on the length of the shorter first observation period given in Table 2.1.

The Poisson distribution with parameter $\lambda \geq 0$ is defined by

$$\text{Poisson}_\lambda(k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad \text{for } k \in \mathbb{N}_0. \tag{4.1}$$

Let the random variable N_y describe the number of days within the observation period ending in year $y \in \{1987, \ldots, 1996\}$, on which more than 1 000 adjusted claims arose from damage by storm or hail. We assume that these ten random variables are independent and that each of them has a Poisson distribution with the same parameter $\lambda \geq 0$. Since $\mathbb{E}[N_y] = \lambda$, the empirical mean

$$\lambda_{1000}^{\text{const}} = \frac{1}{10} \sum_{y=1987}^{1996} N_y \tag{4.2}$$

is an unbiased estimator for λ, which is also sufficient [10, Section 1.9, Example 16]. Table 1.1 contains $m = 17$ events within the $n = 10$ observation periods, hence

$$\lambda_{1000}^{\text{const}} = \frac{m}{n} = \frac{17}{10} = 1.7\,. \tag{4.3}$$

Figure 4.1 contains an illustration of the counting data and this empirical mean. Since $\text{Var}(N_y) = \lambda$, the variance of the estimator $\lambda_{1000}^{\text{const}}$ in (4.2) is λ/n with $n = 10$, hence the estimated standard deviation of $\lambda_{1000}^{\text{const}}$ is

$$\hat{\sigma}(\lambda_{1000}^{\text{const}}) = \sqrt{\frac{\lambda_{1000}^{\text{const}}}{n}} = \sqrt{\frac{m/n}{n}} = \frac{\sqrt{m}}{n} = \frac{\sqrt{17}}{10} \approx 0.41\,. \tag{4.4}$$

It remains to determine the probability that an event, which causes more than 1 000 adjusted claims, actually causes at least 6 000 adjusted claims and therefore leads to the knock-out of the corresponding WINCAT coupon. For this purpose we shall consider a sequence $\{X_k\}_{k\in\mathbb{N}}$ of independent, identically distributed random variables, where X_k describes the severity of the k^{th} event. We always assume that the sequence $\{X_k\}_{k\in\mathbb{N}}$ is independent of $N_{1987}, \ldots, N_{1996}$. The random variables $X_1, \ldots, X_{N_{1987}}$ are used to describe the severity of the events in 1987, the variables $X_{N_{1987}+1}, \ldots, X_{N_{1987}+N_{1988}}$ those in 1988 and so on. In the following subsections we consider three different distributions for the random variables $\{X_k\}_{k\in\mathbb{N}}$.

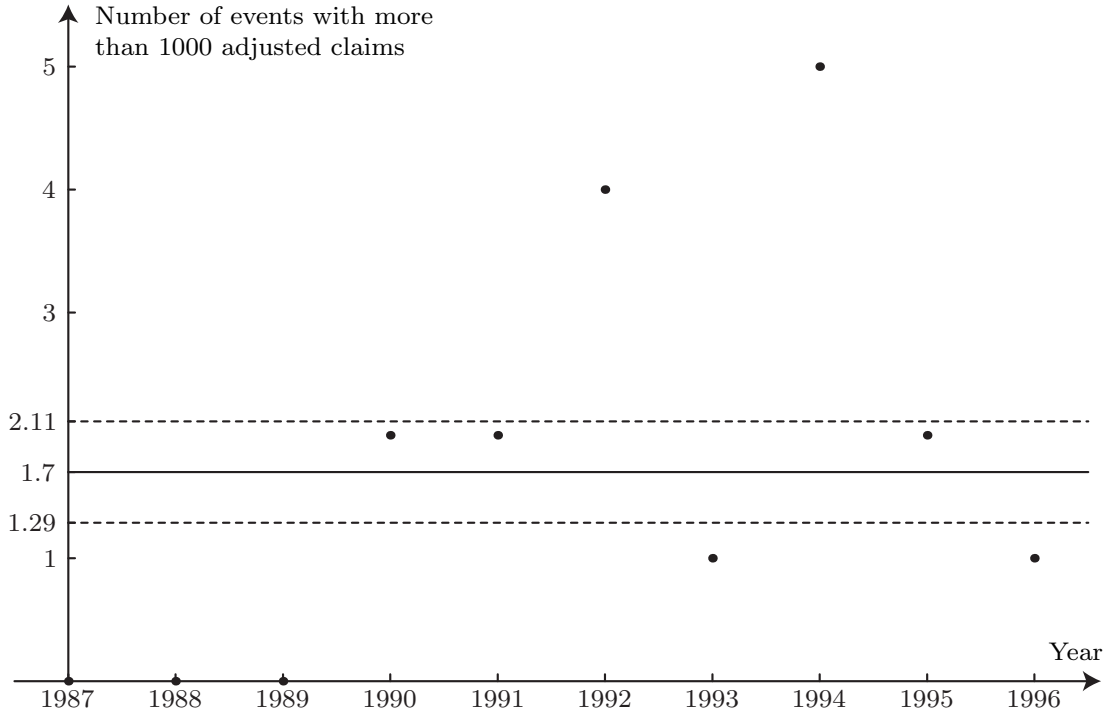

FIGURE 4.1. Observed number of events in the ten observation periods November 1st to October 31st causing more than 1 000 adjusted claims. The empirical mean of $\lambda_{1000}^{\text{const}} = 1.7$ events per observation period is also shown. The dashed lines indicate the estimated standard deviation 0.41 of the estimator $\lambda_{1000}^{\text{const}}$ given by (4.4). The estimated standard deviation for the distribution of the observations is $(\lambda_{1000}^{\text{const}})^{1/2} = \sqrt{1.7} \approx 1.30$.

4.1. **Bernoulli distribution for the knock-out events.** In this subsection we introduce a simple model to describe events with more than 1 000 adjusted claims, which actually cause at least 6 000 adjusted claims; meaning that they lead to a knock-out of the WINCAT coupon. For this purpose we introduce Bernoulli random variables $X_1, \ldots, X_m$ for the $m = 17$ events, where $X_k = 1$ means that event number $k \in \{1, \ldots, m\}$ caused at least 6 000 adjusted claims. We set $X_k = 0$ otherwise. We assume that $X_1, \ldots, X_m$ are independent and identically distributed. Proceeding as in Section 3, we can estimate the probability $p_{6000} = \mathbb{P}(X_k = 1)$ by the unbiased empirical mean

$$\hat{p}_{6000} = \frac{1}{m} \sum_{k=1}^{m} X_k. \tag{4.5}$$

The data of Table 1.1 leads to $\hat{p}_{6000} = 2/m = {}^2\!/_{17} \approx 0.118$. An analysis similar to (3.2) and (3.3) gives the estimate

$$\hat{\sigma}(\hat{p}_{6000}) = \sqrt{\frac{m}{m-1} \frac{\hat{p}_{6000}(1-\hat{p}_{6000})}{m}} = \sqrt{\frac{{}^2\!/_{17} \cdot (1 - {}^2\!/_{17})}{16}} \approx 0.081 \tag{4.6}$$

for the standard deviation of $\hat{p}_{6000}$.

If N is a random variable with the Poisson distribution given by (4.1) describing the number of events, and if independently of everything else we perform a Bernoulli experiment with success probability $p_{6000} \in [0, 1]$ for each of the N events, then an

elementary exercise shows that the resulting number of successful events has a Poisson distribution with parameter $p_{6000}\lambda$. Therefore, under the above assumptions, the number of events per observation period leading to at least 6 000 adjusted claims has a Poisson distribution. An estimate for the corresponding Poisson parameter is

$$\lambda_{6000}^{\text{const}} = \hat{p}_{6000} \cdot \lambda_{1000}^{\text{const}} = \frac{2}{m} \cdot \frac{m}{n} = \frac{2}{10} = 0.2 \,. \tag{4.7}$$

The probability that no such event happens, is given by $\exp(-\lambda_{6000}^{\text{const}})$, see (4.1) with $k = 0$. Hence, the estimated knock-out probability is

$$P_{\text{CAT}} = 1 - \exp(-\lambda_{6000}^{\text{const}}) = 1 - \exp(-0.2) \approx 0.181 \,. \tag{4.8}$$

A recalculation of Table 3.2 with this value of P_{CAT} leads to a discounted value of CHF 249.93 for the three WINCAT coupons.

To estimate the knock-out probability of the first WINCAT coupon, we have to replace $\lambda_{1000}^{\text{const}} = {}^{17}\!/_{10}$ from (4.3) by $\lambda_{1000}^{\text{const}} = {}^{15}\!/_{10}$, because only 15 events are recorded in Table 1.1 for the period from February 28$^{\text{th}}$ to October 31$^{\text{st}}$. This leads via (4.7) and (4.8) to

$$\begin{aligned} P_{\text{CAT}} &= 1 - \exp\bigl(-\hat{p}_{6000} \cdot \lambda_{1000}^{\text{const}}\bigr) \\ &= 1 - \exp\Bigl(-\frac{2}{17} \cdot \frac{15}{10}\Bigr) = 1 - \exp\Bigl(-\frac{15}{17} \cdot 0.2\Bigr) \approx 0.162, \end{aligned} \tag{4.9}$$

which is exactly the same result as the one obtained by applying the correction formula (2.1) to the result of (4.8).

The variance of the estimator $\lambda_{6000}^{\text{const}}$ is not easily computable from the variances of $\hat{p}_{6000}$ and $\lambda_{1000}^{\text{const}}$, because these two estimators are dependent (knowing $\lambda_{1000}^{\text{const}}$ restricts the set of possible values for $\hat{p}_{6000}$). According to our model assumptions however, we have observations from $n = 10$ independent Poisson random variables available, which describe the number of events in each of the ten observation periods leading to at least 6 000 adjusted claims. Similar to (4.2) and (4.4), we therefore see that the estimator (4.7) for $\lambda_{6000}^{\text{const}}$ is unbiased and that

$$\hat{\sigma}(\lambda_{6000}^{\text{const}}) = \sqrt{\lambda_{6000}^{\text{const}}/n} = \sqrt{0.2/10} \approx 0.141 \,.$$

For a conservative estimate of the knock-out probability we might use

$$P_{\text{CAT}} = 1 - \exp\bigl(-\lambda_{6000}^{\text{const}} - \hat{\sigma}(\lambda_{6000}^{\text{const}})\bigr) \approx 1 - \exp(-0.341) \approx 0.289 \,.$$

A recalculation of Table 3.2 with this value of P_{CAT} leads to CHF 218.24 for the discounted value of the three WINCAT coupons.

There is a methodical problem with the approach in this subsection so far. We are mainly interested in an unbiased estimator for the knock-out probability P_{CAT}. The unbiasedness of the estimator $\lambda_{6000}^{\text{const}}$ for a model specific parameter is not of primary concern. To elaborate on this point, let $N_{6000,n}$ be the number of events with at least 6 000 adjusted claims within $n = 10$ observation periods. According to our model assumptions, $N_{6000,n}$ has a Poisson distribution with parameter $np\lambda$, where λ is the intensity for the number of events per observation period with more than 1 000 adjusted claims, and $p = p_{6000}$ is the "success" probability for the following Bernoulli experiment indicating whether actually at least 6 000 adjusted claims arise from the event. The estimator (4.8) corresponds to

$$P_{\text{CAT}} = 1 - \exp(-N_{6000,n}/n) \tag{4.10}$$

with $n = 10$. Calculating the expectation gives

$$\begin{aligned}\mathbb{E}[1 - \exp(-N_{6000,n}/n)] &= 1 - \sum_{k=0}^{\infty} e^{-k/n}\, \frac{(np\lambda)^k}{k!} e^{-np\lambda} \\ &= 1 - \sum_{k=0}^{\infty} \frac{(np\lambda e^{-1/n})^k}{k!} e^{-np\lambda} \\ &= 1 - \exp\bigl(-(1 - e^{-1/n})np\lambda\bigr),\end{aligned} \tag{4.11}$$

which is different from $1 - \exp(-p\lambda)$, hence (4.10) is biased. Multiplying $N_{6000,n}$ in (4.10) by the correction factor $\log\bigl(\frac{n}{n-1}\bigr)^n$ leads to the estimator

$$P_{\text{CAT}} = 1 - \Bigl(1 - \frac{1}{n}\Bigr)^{N_{6000,n}} \tag{4.12}$$

with expectation $1 - \exp(-p\lambda)$ as a calculation similar to (4.11) shows. Hence the estimator (4.12) is unbiased. Since $n = 10$ and $N_{6000,10} = 2$ by Table 1.1, we obtain

$$P_{\text{CAT}} = 1 - \Bigl(\frac{9}{10}\Bigr)^2 = 0.19\,. \tag{4.13}$$

The corresponding recalculation of Table 3.2 gives CHF 247.37 for the discounted value of the three WINCAT coupons.

For the variance of the estimator in (4.12) we obtain after a short calculation similar to (4.11)

$$\operatorname{Var}(P_{\text{CAT}}) = \mathbb{E}\biggl[\Bigl(1 - \frac{1}{n}\Bigr)^{2N_{6000,n}}\biggr] - e^{-2p\lambda} = e^{-2p\lambda}(e^{p\lambda/n} - 1).$$

Using the estimate $\lambda_{6000}^{\text{const}} = 0.2$ for $p\lambda$ from (4.7) and $(9/10)^2$ for $e^{-p\lambda}$ from (4.13), we obtain

$$\hat{\sigma}(P_{\text{CAT}}) = e^{-p\lambda}\sqrt{e^{p\lambda/n} - 1} \approx e^{-p\lambda}\sqrt{p\lambda/n} \approx \Bigl(\frac{9}{10}\Bigr)^2 \sqrt{0.02} \approx 0.115\,. \tag{4.14}$$

A recalculation of Table 3.2 with the conservative knock-out probability $P_{\text{CAT}} + \hat{\sigma}(P_{\text{CAT}}) \approx 0.305$ gives CHF 213.73 for the discounted value of the WINCAT coupons.

The estimated standard deviation in (4.14) is slightly smaller than the one in the simple binomial model calculated via (3.3). This indicates that in our case the composite Poisson model of this subsection leads only to a slight improvement. Indeed, the estimator (4.13) for the knock-out probability uses only the information that two events within the ten years caused at least 6 000 adjusted claims. Since the model of this subsection allows these two events to happen in the same year, the estimated knock-out probability in (4.13) is 1% lower than the one in the binomial model. If the two events with at least 6 000 adjusted claims had actually happened in the same year and not in consecutive ones, the discrepancy in the estimated knock-out probabilities would be 9%, because the estimate in the binomial model of Section 3 would drop from 20% to 10%. In this respect the composite Poisson model of this subsection is more robust than the binomial one.

4.2. **Pareto distribution for the knock-out events.** The binomial model of Section 3 and the corresponding composite Poisson model of Subsection 4.1 do not use the adjusted claim numbers recorded in Table 1.1. For the benefit of a better estimate of p_{6000}, let us incorporate these numbers into the model. The step function in Figure 4.2 is the empirical distribution function of the adjusted claim

numbers from Table 1.1. A heavy-tailed distribution of common use is the Pareto distribution, its distribution function is given by

$$\mathrm{Pareto}_{a,b}(x) = \begin{cases} 1-(a/x)^b & \text{for } x \geq a, \\ 0 & \text{for } x < a, \end{cases} \tag{4.15}$$

where a and b are strictly positive parameters. The Pareto distribution is used in [3] to model the number of adjusted claims per event given that more than 1 000 adjusted claims arise from the event. We choose the threshold $a = 1\,000$, because only such events are contained in Table 1.1. At first glance it might look as if we make a conceptual mistake by fitting the distribution of an apparently integer-valued random variable by a distribution having a density. However, the involved numbers from the last column of Table 1.1 are sufficiently large for such an approximation and, in addition, they are actually rounded numbers arising as the product of the number of claims and the vehicles insured index. Therefore, the use of a continuous distribution function should not cause an intellectual problem (see Section 6 and the end of Section 13 however).

To fit the empirical distribution with a Pareto distribution as in Figure 4.2, we need an estimator for the exponent b. If a random variable X has a Pareto distribution with parameters a and b, then $Y \equiv \log(X/a)$ satisfies

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X \leq ae^y) = 1 - \Big(\frac{a}{ae^y}\Big)^b = 1 - e^{-by}, \qquad y \geq 0,$$

which means that Y has an exponential distribution with expectation $\mathbb{E}[Y] = 1/b$. Hence, if the independent random variables $X_1, \ldots, X_m$ with a Pareto distribution given by (4.15) describe the adjusted number of claims for the m events, then the random variables $Y_1, \ldots, Y_m$ with $Y_k \equiv \log(X_k/a)$ are independent and exponentially distributed. Their empirical mean $(1/m)\sum_{k=1}^m Y_k$ is an unbiased estimator for $1/b$. This suggests to estimate b by the reciprocal value

$$\frac{m}{\sum_{k=1}^m Y_k} = \frac{m}{\sum_{k=1}^m \log(X_k/a)}. \tag{4.16}$$

Another way to derive this estimator is to consider the likelihood function

$$L_m(b) \equiv \prod_{k=1}^m \frac{b}{X_k}\Big(\frac{a}{X_k}\Big)^b, \qquad b > 0, \tag{4.17}$$

which is the product of the densities of the Pareto distribution (4.15) evaluated at $X_1, \ldots, X_m$. By differentiating the logarithm of L_m, we find that b given by (4.16) maximises L_m, hence (4.16) is also the maximum-likelihood estimator for b.

Let us calculate the expectation of the estimator in (4.16). The sum $\sum_{k=1}^m Y_k$ has a gamma distribution with parameters m and b, meaning that

$$\mathbb{P}\Big(\sum_{k=1}^m Y_k \leq y\Big) = \int_0^y \frac{b}{\Gamma(m)}(bt)^{m-1}e^{-bt}\,dt, \qquad y \geq 0.$$

This fact is easily proved by an induction on m, because the convolution of the exponential density and the gamma density of parameter m leads to the gamma density of parameter $m+1$:

$$\int_0^t be^{-b(t-s)} \cdot \frac{b}{\Gamma(m)}(bs)^{m-1}e^{-bs}\,ds = \frac{b}{m\Gamma(m)}(bt)^m e^{-bt}, \qquad t \geq 0,$$

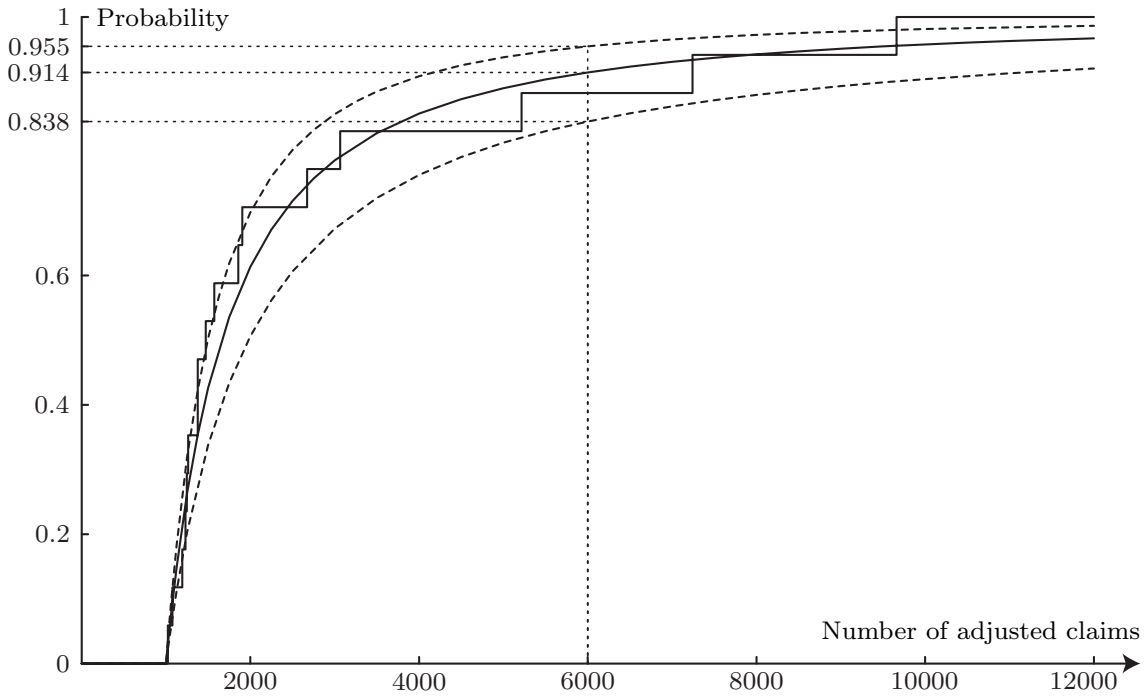

FIGURE 4.2. The step function is the empirical distribution of the number of (adjusted) claims per event, given that more than 1 000 claims arise from the event. Also shown is the fitted Pareto distribution (4.15) with $a = 1\,000$ and $b = \hat{b}_{17}$, where $\hat{b}_{17} \approx 1.37$ is the maximum-likelihood estimate, corrected with the factor $(m-1)/m$ for $m = 17$ to eliminate the bias. The estimated probability, that an event with at least 1 000 claims causes at most 6 000 claims, is around 0.914. Two additional Pareto distributions (dashed curves) illustrate the estimated standard deviation of $\hat{b}_{17}$. The lower dashed curve corresponds to $\hat{b}_{17} - \hat{\sigma}(\hat{b}_{17}) \approx 1.02$, the upper one to $\hat{b}_{17} + \hat{\sigma}(\hat{b}_{17}) \approx 1.73$.

and the gamma function satisfies $\Gamma(m+1) = m\Gamma(m)$. Calculating the expectation of (4.16) for $m \geq 2$ shows that

$$\begin{aligned} \mathbb{E}\left[\frac{m}{\sum_{k=1}^{m} \log(X_k/a)}\right] &= \int_0^\infty \frac{m}{t} \cdot \frac{b}{\Gamma(m)} (bt)^{m-1} e^{-bt}\, dt \\ &= \frac{mb}{m-1} \int_0^\infty \frac{b}{\Gamma(m-1)} (bt)^{m-2} e^{-bt}\, dt = \frac{m}{m-1} b. \end{aligned} \tag{4.18}$$

This means that the estimator in (4.16) underestimates the tail of the Pareto distribution. To obtain an unbiased estimator for b, we therefore have to use

$$\hat{b}_m = \frac{m-1}{\sum_{k=1}^{m} \log(X_k/a)} \tag{4.19}$$

instead of (4.16). The data from the last column of Table 1.1 leads to

$$\hat{b}_{17} \approx 1.37\ . \tag{4.20}$$

The Pareto distribution with this value is shown in Figure 4.2.

A calculation similar to (4.18) leads to

$$\mathrm{Var}(\hat{b}_m) = \frac{b^2}{m-2} \tag{4.21}$$

for all $m \geq 3$. Therefore, $\hat{\sigma}(\hat{b}_m) = \hat{b}_m/\sqrt{m-2}$ is an unbiased estimator for the standard deviation; using the numerical value from (4.20) gives

$$\hat{\sigma}(\hat{b}_{17}) \approx 1.37/\sqrt{15} \approx 0.35\,. \tag{4.22}$$

The Pareto distributions with $\hat{b}_{17} \pm \hat{\sigma}(\hat{b}_{17})$ are shown as dashed curves in Figure 4.2.

Using $\hat{b}_m$ for the parameter of the Pareto distribution (4.15), we obtain the estimator

$$\hat{p}_{6000} = 1 - \mathrm{Pareto}_{1000,\hat{b}_m}(6000) = 6^{-\hat{b}_m} \tag{4.23}$$

for the probability that an event, which causes more than 1 000 adjusted claims, actually causes at least 6 000 adjusted claims. The numerical value $\hat{b}_{17} \approx 1.37$ from (4.20) leads to

$$\hat{p}_{6000} \approx 6^{-1.37} \approx 0.0857\,. \tag{4.24}$$

Considering the two Pareto distributions corresponding to $\hat{b}_{17} - \hat{\sigma}(\hat{b}_{17}) \approx 1.02$ and $\hat{b}_{17} + \hat{\sigma}(\hat{b}_{17}) \approx 1.73$ (see Figure 4.2), we obtain via (4.23) the asymmetric interval

$$[6^{-1.73}, 6^{-1.02}] \approx [0.045, 0.162] \tag{4.25}$$

around the estimate $\hat{p}_{6000} \approx 0.0857$ as an indication of the standard deviation. This is an improvement compared to the interval $[0.037, 0.199]$ arising from the Bernoulli distribution via (4.6).

Following the approach in [3], we recalculate the estimate (4.7) for the Poisson parameter $\lambda^{\mathrm{const}}_{6000}$ describing the number of knock-out events per observation period using $\lambda^{\mathrm{const}}_{1000} = 1.7$ from (4.3) and $\hat{p}_{6000} \approx 0.0857$ from (4.24). We obtain $\lambda^{\mathrm{const}}_{6000} = \hat{p}_{6000} \cdot \lambda^{\mathrm{const}}_{1000} \approx 0.1457$. As in (4.8), the estimated knock-out probability is

$$P_{\mathrm{CAT}} = 1 - \exp\bigl(-\lambda^{\mathrm{const}}_{6000}\bigr) = 1 - \exp\bigl(-\hat{p}_{6000} \cdot \lambda^{\mathrm{const}}_{1000}\bigr) \approx 0.1356\,. \tag{4.26}$$

A recalculation of Table 3.2 with this value of P_{CAT} leads to a discounted value of CHF 263.29 for the three WINCAT coupons.

To get a rough estimate of the standard deviation of the knock-out probability in (4.26), consider it as a function of the two parameters $\hat{b}_{17}$ and $\hat{\lambda} \equiv \lambda^{\mathrm{const}}_{1000}$:

$$P_{\mathrm{CAT}}(\hat{b}_{17}, \hat{\lambda}) = 1 - \exp\bigl(-6^{-\hat{b}_{17}}\hat{\lambda}\bigr).$$

Using the approximating plane in (b, λ) and thereby neglecting all higher order terms in the Taylor expansion, we get

$$P_{\mathrm{CAT}}(\hat{b}_{17}, \hat{\lambda}) \approx P_{\mathrm{CAT}}(b, \lambda) + \frac{\partial P_{\mathrm{CAT}}}{\partial b}(b, \lambda)(\hat{b}_{17} - b) + \frac{\partial P_{\mathrm{CAT}}}{\partial \lambda}(b, \lambda)(\hat{\lambda} - \lambda).$$

Since $\hat{b}_{17}$ and $\hat{\lambda}$ are unbiased, we obtain for the variance

$$\begin{aligned}\mathrm{Var}\bigl(P_{\mathrm{CAT}}(\hat{b}_{17}, \hat{\lambda})\bigr) \approx{}& \Bigl(\frac{\partial P_{\mathrm{CAT}}}{\partial b}(b, \lambda)\Bigr)^2 \mathrm{Var}(\hat{b}_{17}) + \Bigl(\frac{\partial P_{\mathrm{CAT}}}{\partial \lambda}(b, \lambda)\Bigr)^2 \mathrm{Var}(\hat{\lambda}) \\ &+ \frac{\partial P_{\mathrm{CAT}}}{\partial b}(b, \lambda)\, \frac{\partial P_{\mathrm{CAT}}}{\partial \lambda}(b, \lambda)\, \mathbb{E}[(\hat{b}_{17} - b)(\hat{\lambda} - \lambda)].\end{aligned}$$

The two estimators $\hat{b}_m$ and $\hat{\lambda} = \lambda^{\mathrm{const}}_{1000}$ are certainly not independent, because the observed number m of events determines $\lambda^{\mathrm{const}}_{1000}$ via (4.3) and the variance of $\hat{b}_m$ via (4.21). However, $\hat{b}_{17}$ and $\hat{\lambda}$ are independent and therefore uncorrelated, meaning

that $\mathbb{E}[(\hat{b}_{17} - b)(\hat{\lambda} - \lambda)] = 0$. Evaluating the partial derivatives of the knock-out probability P_{CAT} at the estimated point $(\hat{b}_{17}, \hat{\lambda})$ instead of (b, λ), and using the estimated standard deviations from (4.22) and (4.4) instead of $(\text{Var}(\hat{b}_{17}))^{1/2}$ and $(\text{Var}(\hat{\lambda}))^{1/2}$, we obtain the approximation

$$\begin{aligned}\hat{\sigma}(P_{\text{CAT}}(\hat{b}_{17}, \hat{\lambda})) &\approx \sqrt{\Big(\frac{\partial P_{\text{CAT}}}{\partial b}(\hat{b}_{17}, \hat{\lambda})\Big)^2 \hat{\sigma}^2(\hat{b}_{17}) + \Big(\frac{\partial P_{\text{CAT}}}{\partial \lambda}(\hat{b}_{17}, \hat{\lambda})\Big)^2 \hat{\sigma}^2(\hat{\lambda})} \\ &\approx 0.086\,. \end{aligned} \tag{4.27}$$

From (4.26) and (4.27) we obtain $P_{\text{CAT}}(\hat{b}_{17}, \hat{\lambda}) + \hat{\sigma}(P_{\text{CAT}}(\hat{b}_{17}, \hat{\lambda})) \approx 0.221$ as a conservative estimate of the knock-out probability. A recalculation of Table 3.2 leads to a discounted value of CHF 238.25 for the three WINCAT coupons. Due to these calculations, in [3] the rounded knock-out probability of 0.25 is considered to be a conservative estimate, leading to a discounted value of CHF 229.78.[4] This value is supposed to include a risk premium for the investor because the standard deviation of the knock-out probability is added and the result rounded in a conservative way.

Before turning our attention to a generalised Pareto distribution for the knock-out events, let us conclude this subsection with some supplementary considerations concerning the biasedness of the estimators for p_{6000} and P_{CAT}. First note that $\lambda_{1000}^{\text{const}}$ from (4.2) and $\hat{b}_{17}$ from (4.19) are unbiased estimators for the two model parameters λ and b, but this does not imply that $\hat{p}_{6000}$ and P_{CAT}, given by (4.23) and (4.26), respectively, are unbiased. The arguments leading to the unbiased estimator (4.12) in the case of the Bernoulli distribution for the knock-out probability in Subsection 4.1 suggest that the estimator

$$P_{\text{CAT}} = 1 - \Big(1 - \frac{\hat{p}_{6000}}{n}\Big)^{N_{1000,n}} = 1 - \Big(1 - \frac{6^{-\hat{b}_{17}}}{n}\Big)^{N_{1000,n}} \tag{4.28}$$

is a small improvement, because this would be an unbiased estimator for P_{CAT} if $\hat{b}_{17}$ were non-random. Here the random variable $N_{1000,n}$ denotes the number of events with more than 1 000 adjusted claims within the n observation periods. Recall that $N_{1000,n}$ has a Poisson distribution with parameter $n\lambda$. Substituting our estimate $\hat{b}_{17} \approx 1.37$ from (4.20) and $N_{1000,n} = 17$ for the $n = 10$ observation periods into (4.28) leads to $P_{\text{CAT}} \approx 0.1361$, which gives a discounted value of CHF 263.13 for the three WINCAT coupons. This is a decrease of only CHF 0.16 compared to the value arising from (4.26).

If we consider $N_{1000,10} = 17$ as non-random and replace $\hat{b}_{17} \approx 1.37$ from (4.20) by $\hat{b}_{17} - \hat{\sigma}(\hat{b}_{17}) \approx 1.02$ in the estimator (4.28) to find a conservative estimate, we get $P_{\text{CAT}} \approx 0.242$, which via Table 3.2 leads to CHF 232.14 for the discounted value of the three WINCAT coupons. Note that this knock-out probability is about 0.02 larger than the one obtained from (4.27) and is already very close to the conservatively rounded value of 0.25 from [3].

An examination of the above model reveals that the conditional distribution of the estimator $\hat{b}_m$ given $m = N_{1000,n}$ is only specified in the case $N_{1000,n} \geq 2$. Furthermore, (4.21) shows that $\hat{b}_m$ does not have a variance unless $m = N_{1000,n} \geq 3$. Hence, the above approach of fitting the empirical distribution of the adjusted claim numbers by a Pareto distribution is applicable only in the case of appropriate data

[4] In [3] a discounted value of CHF 227.09 is actually derived, because the $^{15}/_{17}$-correction for the first observation period is not taken into account.

sets. Such an a priori exclusion of certain data sets already introduces a bias which suggests that unbiasedness for estimators like (4.26) or (4.28) is a problematical notion. Maybe a notion of conditional unbiasedness would be more appropriate. This means in our case that one would like to have estimators for P_{CAT} such that the conditional expectation given $\hat{p}_{6000}\, N_{1000,n} \geq 1$, for example, is the right one.

4.3. **Generalised Pareto distribution for the knock-out events.** In Subsection 4.2, we did not give a theoretical argument in favour of the Pareto distribution in addition to the desire to pick a heavy-tailed distribution. Let us use an idea from extreme value theory to overcome this deficiency. It will turn out that we should use a generalised version of the Pareto distribution defined in (4.15).

Let $X_1, \ldots, X_k$ denote the adjusted number of claims arising from k events. We shall assume that $X_1, \ldots, X_k$ are independent and distributed according to a heavy-tailed distribution function. We are only interested in those numbers which exceed a certain threshold a, which is 1 000 in our case. This means we are interested in the excess distribution function

$$F_a(x) = \mathbb{P}(X_1 - a \leq x \,|\, X_1 > a), \qquad x \in \mathbb{R}.$$

Extreme value theory essentially says the following in our case [6, Section 3.4]: If the original distribution function of $X_1, \ldots, X_k$ is heavy-tailed, then the excess distribution functions $\{F_a\}_{a>0}$ can be better and better approximated (with respect to the supremum norm) by generalised Pareto distributions of the form

$$G_{\xi,\tau_a}(x) = \begin{cases} 1 - (1 + \xi x/\tau_a)^{-1/\xi} & \text{for } x \geq 0, \\ 0 & \text{for } x < 0, \end{cases}$$

as the threshold a tends to infinity. Here ξ is a strictly positive[5] shape parameter and the scale parameter $\tau_a > 0$ varies with the threshold a. This suggests that we should try to fit the empirical distribution function of the observations exceeding the threshold a by a distribution function of the form

$$G_{a,\xi,\tau}(x) = \begin{cases} 1 - (1 + \xi(x-a)/\tau)^{-1/\xi} & \text{for } x \geq a, \\ 0 & \text{for } x < a. \end{cases} \tag{4.29}$$

Note that in the heavy-tailed case $\xi > 0$, the (shifted) generalised Pareto distribution (4.29) with $\tau = a\xi$ reduces to the Pareto distribution (4.15) with $b = 1/\xi$. Hence, $G_{a,\xi,\tau}$ gives us the freedom of the additional scale parameter τ.

Before fitting a generalised Pareto distribution function to the observations, an exploratory data analysis should be done, see [6, Chapter 6], to check the assumption of a heavy-tailed distribution and to determine a suitable threshold. However, since there are only $m = 17$ observations available in Table 1.1, there seems to be no point in choosing a higher threshold than $a = 1000$ in our case, because the historical data set is quite small already. The assumption of a heavy tail is (at least partially) supported by Figure 4.4.

The log-likelihood function for the $m = 17$ observations originating from a generalised Pareto distribution is

$$l(\xi, \tau) = -m \log \tau - \left(1 + \frac{1}{\xi}\right) \sum_{k=1}^{m} \log\left(1 + \xi \frac{X_k - 1000}{\tau}\right), \quad \xi > 0,\, \tau > 0. \tag{4.30}$$

[5]The cases $\xi = 0$ and $\xi < 0$ are discussed in Section 13, see (13.1) and (13.2).

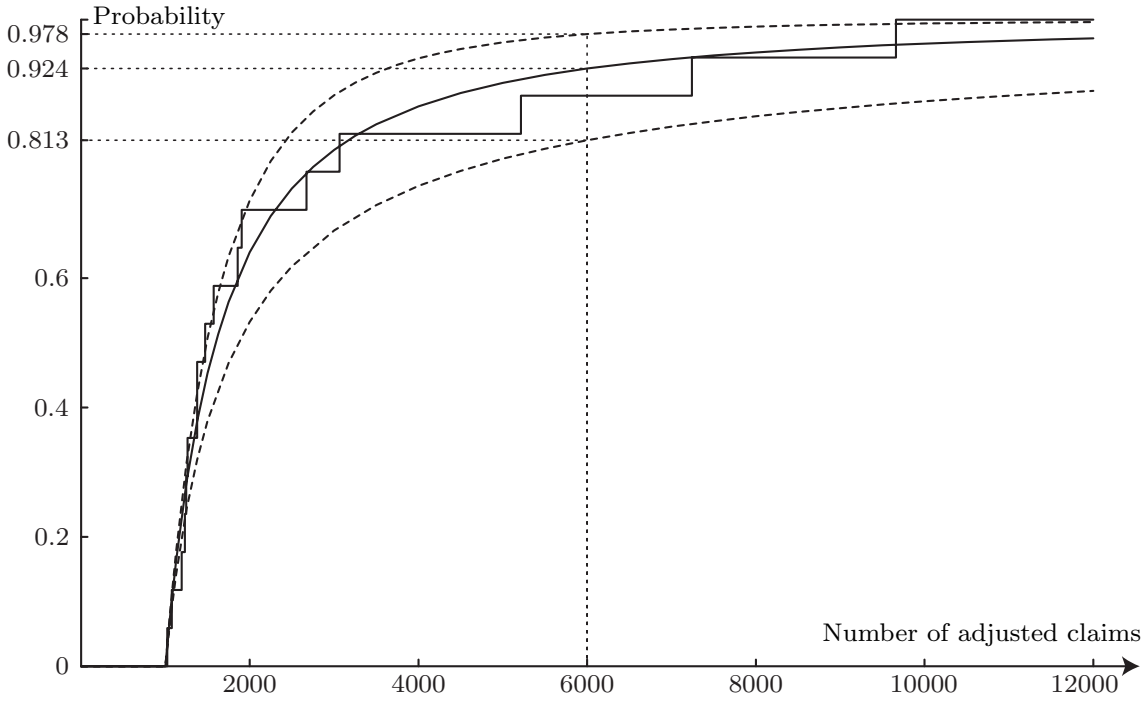

FIGURE 4.3. The empirical distribution of the number of adjusted claims per event (solid step function) and the fitted generalised Pareto distribution (solid curve) with threshold $a = 1000$, estimated exponent $1/\hat{\xi} \approx 1.38$ and estimated scale parameter $\hat{\tau} \approx 660.7$. The estimated probability that an event with at least 1 000 claims causes at most 6 000 claims, is approximately 0.924, and $[0.813, 0.978]$ is an approximate 68 %-confidence interval for this probability. The two dashed curves are generalised Pareto distributions chosen such that they indicate the standard deviation of the estimated probability for at most 6 000 claims.

Inserting the data from the last column of Table 1.1, we can calculate the maximum-likelihood estimator $(\hat{\xi}, \hat{\tau})$ numerically, i. e., we can search for the point $(\hat{\xi}, \hat{\tau})$ which maximises l. As starting values for the numerical iteration procedure, we can choose $\xi = 1/\hat{b}_m$ and $\tau = a/\hat{b}_m$, where $\hat{b}_m$ is the estimator (4.19) for the Pareto distribution, or we can use a probability-weighted moment approach (see [6, Section 6.3.2 and page 358]) to obtain a priori estimates for ξ and τ. We find

$$\hat{\xi} \approx 0.7243 \qquad \text{and} \qquad \hat{\tau} \approx 660.7, \tag{4.31}$$

hence

$$\hat{p}_{6000} = 1 - G_{1000,\hat{\xi},\hat{\tau}}(6000) \approx 1 - 0.92425 = 0.07575\ . \tag{4.32}$$

The corresponding fit of the empirical distribution with a generalised Pareto one is shown in Figures 4.3 and 4.4. A calculation as in (4.28) gives the estimate

$$P_{\text{CAT}} = 1 - \Big(1 - \frac{\hat{p}_{6000}}{10}\Big)^{N_{1000,10}} \approx 0.121,$$

which leads via a recalculation of Table 3.2 to a discounted value of CHF 267.48.

For comparison with the earlier results on the standard deviation of $\hat{p}_{6000}$ in the case of the Bernoulli distribution for the knock-out events in (4.6) and for the corresponding case of the Pareto distribution in (4.25), we would like to give

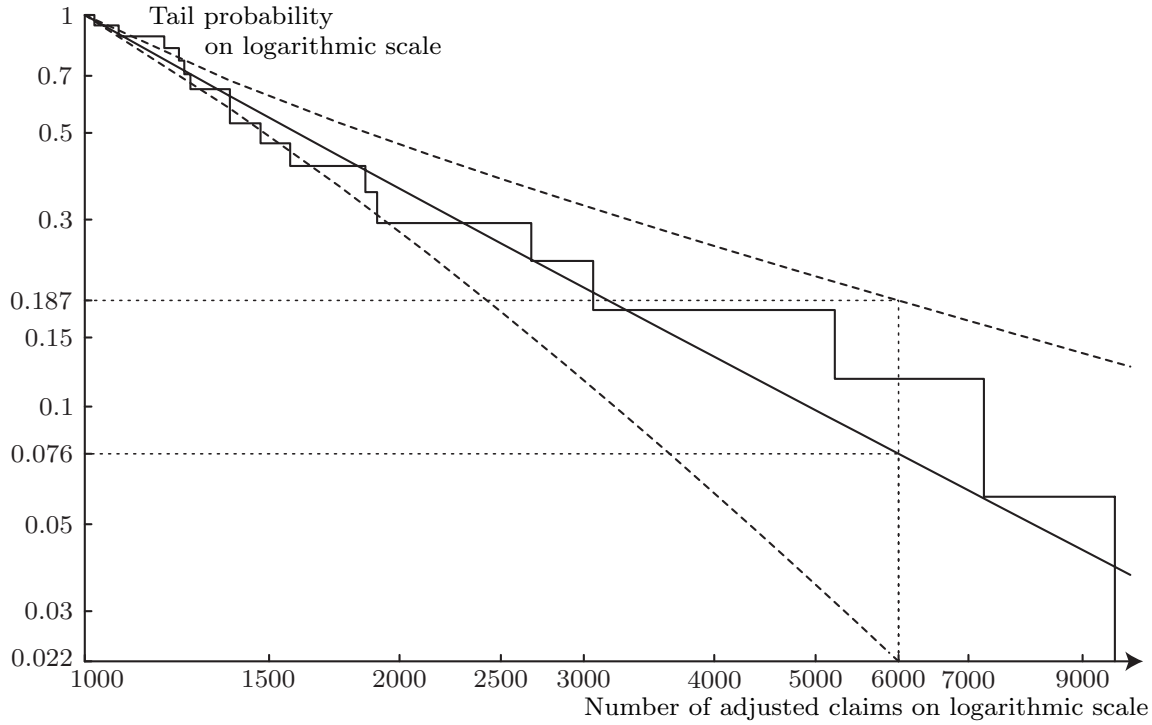

FIGURE 4.4. This is Figure 4.3 on log-log scale to magnify the important part. Instead of the distribution functions, the corresponding tail probabilities are shown. Pareto distribution functions defined by (4.15) would give straight lines in this log-log plot. The estimated generalised-Pareto fit $x \mapsto 1 - G_{1000,\hat{\xi},\hat{\tau}}(x)$ is close to a straight line because $a\hat{\xi}/\hat{\tau} \approx 1.096$ is quite close to one. The estimates $\hat{p}^-_{6000} \approx 0.022$, $\hat{p}_{6000} \approx 0.0757$ and $\hat{p}^+_{6000} \approx 0.187$ are shown. This figure supports the model assumption, that the adjusted claim numbers follow a heavy-tailed distribution.

again an estimate for the standard deviation of $\hat{p}_{6000}$. This does not seem to be possible by analytical means, however. Therefore, we prefer to construct an interval $[\hat{p}^-_{6000}, \hat{p}^+_{6000}]$ around the estimated value $\hat{p}_{6000} \approx 0.0757$ from (4.32), which can serve as the region for accepting the null hypothesis $p = \hat{p}_{6000}$ at a 68%-confidence level when using the log-likelihood ratio statistic. We choose the 68% level, because this is the probability that a normally distributed random variable with mean μ and variance $\sigma^2 > 0$ takes its value in the interval $[\mu - \sigma, \mu + \sigma]$. As log-likelihood ratio statistic, also called deviance, we use

$$D(\xi, \tau) = 2l(\hat{\xi}, \hat{\tau}) - 2l(\xi, \tau), \qquad \xi > 0,\ \tau > 0. \tag{4.33}$$

We want to determine the smallest interval $[\hat{p}^-_{6000}, \hat{p}^+_{6000}]$ such that

$$\begin{aligned} &\big\{\, (\xi, \tau) \in (0, \infty)^2 \;\big|\; D(\xi, \tau) \le \chi^2_{2,0.32} \,\big\} \\ &\qquad\qquad \subset \big\{\, (\xi, \tau) \in (0, \infty)^2 \;\big|\; 1 - G_{1000,\xi,\tau}(6000) \in [\hat{p}^-_{6000}, \hat{p}^+_{6000}] \,\big\}, \end{aligned} \tag{4.34}$$

where $\chi^2_{2,0.32} \approx 2.30$ denotes the 32%-quantile of the chi-squared distribution with two degrees of freedom. In other words: We are looking for the smallest probability $\hat{p}^-_{6000}$ and the largest probability $\hat{p}^+_{6000}$, which can arise from generalised Pareto distributions with parameters (ξ, τ) close to $(\hat{\xi}, \hat{\tau})$ in the sense that the deviance

$D(\xi,\tau)$ does not exceed the 32%-quantile $\chi^2_{2,0.32}$ of the χ^2_2-distribution. This choice for the upper bound of the deviance $D(\xi,\tau)$ is based on the asymptotic normality of the maximum-likelihood estimators, see for example [10, Section 8.8]. According to [12, Appendix A], the approximation of the distribution of the deviance by the chi-squared distribution is often quite accurate for small numbers of observations, even when the normal approximation for the parameter estimates is unsatisfactory. When compared to methods using the second derivatives of the log-likelihood function at the estimated point $(\hat\xi,\hat\tau)$, the log-likelihood ratio statistic has the advantage of being able to give asymmetric confidence intervals and thereby being less prejudiced. This is useful in our case, because we don't want to obtain negative estimates for $\hat p^-_{6000}$, for example. It should be kept in mind that (4.34) is in general a strict inclusion, hence $[\hat p^-_{6000},\hat p^+_{6000}]$ can correspond to a higher confidence level than 68%. This is problematical for larger numbers of parameters, because the confidence intervals get too large. Bootstrap methods are an alternative in this case.

Note that the interval $[\hat p^-_{6000},\hat p^+_{6000}]$ in (4.34) does not depend on the parametrisation arising from $(\xi,\tau)\mapsto G_{1000,\xi,\tau}$ in (4.29). We can use this observation to change to an advantageous parametrisation which reduces the amount of numerical calculations necessary to determine the above acceptance interval. Since the equation $p = 1 - G_{1000,\xi,\tau}(6000)$ can be solved for τ yielding

$$\tau(\xi,p) = \frac{5000\xi}{p^{-\xi}-1},$$

we can use p itself as a parameter by changing the parametrisation from (4.29) to $(\xi,p)\mapsto G_{1000,\xi,\tau(\xi,p)}$. Rewriting the inclusion (4.34) with this parametrisation yields $\big\{\,(\xi,p)\in(0,\infty)\times(0,1)\;\big|\;D(\xi,\tau(\xi,p))\le\chi^2_{2,0.32}\,\big\}\subset(0,\infty)\times[\hat p^-_{6000},\hat p^+_{6000}]$. Numerical calculations lead to $[\hat p^-_{6000},\hat p^+_{6000}]\approx[0.022,0.187]$, the corresponding exponents $\hat\xi^-\approx 0.355772$ and $\hat\xi^+\approx 1.396$ are the only ones with a deviance less or equal to the quantile $\chi^2_{2,0.32}$. The shifted generalised-Pareto distribution functions

$$x\mapsto G_{1000,\hat\xi^-,\hat\tau^-}(x)\qquad\text{with}\qquad \hat\tau^- = \tau(\hat\xi^-,\hat p^-_{6000})\approx 620.3$$

and

$$x\mapsto G_{1000,\hat\xi^+,\hat\tau^+}(x)\qquad\text{with}\qquad \hat\tau^+ = \tau(\hat\xi^+,\hat p^+_{6000})\approx 740.6$$

are shown in Figures 4.3 and 4.4. If we consider the number $N_{1000,10}=17$ as non-random and use $\hat p^+_{6000}\approx 0.187$ instead of $\hat p_{6000}$, a calculation as in (4.28) leads to the conservative estimate $P_{\text{CAT}}\approx 0.274$. A recalculation of Table 3.2 gives a discounted value of CHF 222.75 for the three WINCAT coupons.

5. Testing the constant-parameter Poisson model

5.1. **Testing for over-dispersion.** In Section 4 the number of events with at least 1 000 adjusted claims per observation period is modelled by ten independent random variables N_y for the years $y\in\{1987,\dots,1996\}$, each one having the same Poisson distribution (4.1) with parameter $\lambda\ge 0$. Since the expectation and the variance of the Poisson distribution are equal to the parameter λ, the empirical mean $\lambda^{\text{const}}_{1000}$ of $N_{1987},\dots,N_{1996}$ was used in (4.2) as an unbiased estimator for the expectation and the variance. However, if we don't want to rely on the assumption of a Poisson distribution when investigating the variance (but keep the assumption

that $N_{1987}, \dots, N_{1996}$ are independent and identically distributed), then we should estimate the variance $\sigma_N^2 = \mathrm{Var}(N_y)$ by the unbiased estimator

$$\hat{\sigma}_N^2 = \frac{1}{9} \sum_{y=1987}^{1996} (N_y - \hat{\mu}_N)^2 \qquad \text{with} \qquad \hat{\mu}_N = \frac{1}{10} \sum_{y=1987}^{1996} N_y \,.$$

The data of Table 1.1 leads to $\hat{\sigma}_N^2 = 2.9$, which yields the standard deviation

$$\hat{\sigma}(\hat{\mu}_N) = \sqrt{\hat{\sigma}_N^2/10} = \sqrt{29}/10 \approx 0.54 \tag{5.1}$$

for the empirical mean $\hat{\mu}_N$ of $N_{1987}, \dots, N_{1996}$. Note that $\hat{\sigma}_N^2 = 2.9$ is quite a bit larger than $\lambda_{1000}^{\mathrm{const}} = 1.7$ from (4.3). This observation raises the question whether the data of Table 1.1 exhibits over-dispersion, meaning in our case that the variance of $N_{1987}, \dots, N_{1996}$ is actually larger than the mean. Such an over-dispersion can arise, for example, from a Poisson parameter λ which is itself a random variable. In the present case, global weather conditions could have determined different values for λ in the ten observation periods. See, e. g., [12] for a discussion of over-dispersion.

To investigate this question of over-dispersion, let us consider the possibility that a large variance as above, namely $\hat{\sigma}_N^2 \geq 2.9$, happened by chance. This means that we want to calculate the conditional probability $\mathbb{P}(\, \hat{\sigma}_N^2 \geq 2.9 \,|\, \hat{\mu}_N = 1.7 \,)$ under the null hypothesis that $N_{1987}, \dots, N_{1996}$ are independent and distributed according to (4.1) with an unknown Poisson parameter $\lambda > 0$. The small number of observations and their small values make it feasible to calculate the above conditional probability exactly. Under the null hypothesis, the sum $N_{1987} + \cdots + N_{1996}$ has a Poisson distribution with parameter 10λ and we obtain

$$\begin{aligned} &\mathbb{P}(\, N_y = n_y \text{ for every } y \in \{1987, \dots, 1996\} \,|\, \hat{\mu}_N = 1.7 \,) \\ &\qquad = \Bigg(\prod_{y=1987}^{1996} \frac{\lambda^{n_y}}{n_y!} e^{-\lambda} \Bigg) \Bigg/ \frac{(10\lambda)^{17}}{17!} e^{-10\lambda} = \frac{17!}{10^{17}} \prod_{y=1987}^{1996} \frac{1}{n_y!} \end{aligned} \tag{5.2}$$

for every tuple $(n_{1987}, \dots, n_{1996}) \in \mathbb{N}_0^{10}$ with $n_{1987} + \cdots + n_{1996} = 17$. Note that the conditional probability in (5.2) does not depend on the unknown parameter $\lambda > 0$. For every tuple in (5.2), there are

$$\frac{10!}{\prod_{i=0}^{17} (\#\{\, y \in \{1987, \dots, 1996\} \text{ with } n_y = i \,\})!}$$

different rearrangements of the tuple; all of these lead to the same probability in (5.2). A small program,[6] which considers all possible tuples for (5.2) satisfying $n_{y+1} \leq n_y$ for all $y \in \{1987, \dots, 1995\}$, finds 267 such tuples and yields

$$\mathbb{P}(\, \hat{\sigma}_N^2 \geq 2.9 \,|\, \hat{\mu}_N = 1.7 \,) \approx 0.0889 \,. \tag{5.3}$$

While this one-sided test does not show a significant deviation from the Poisson distribution on the 5%-level, it is certainly more conservative to use the standard deviation $\hat{\sigma}(\hat{\mu}_N) \approx 0.54$ from (5.1) instead of $\hat{\sigma}(\lambda_{1000}^{\mathrm{const}}) \approx 0.41$ from (4.4) to take the possibility of over-dispersion into account. Combining this result with the fitted Pareto distribution for the knock-out events (see Subsection 4.2), the

[6] The *Mathematica* command `NumberOfPartitions[17]` from the standard add-on package `DiscreteMath`Combinatorica`` shows that there are 297 partitions of 17 altogether, hence the running time of the program will be acceptable. Unnecessary loops in the program can be avoided by using the condition $(1996 - y)n_{y+1} \geq 17 - (n_{1987} + \cdots + n_y)$ for $y \in \{1987, \dots, 1995\}$.

analogue of (4.27) for the approximation of the standard deviation of the knock-out probability gives $\hat{\sigma}(P_{\text{CAT}}(\hat{b}_{17}, \hat{\lambda})) \approx 0.0893$. Together with (4.26) we obtain $P_{\text{CAT}}(\hat{b}, \hat{\lambda}) + \hat{\sigma}(P_{\text{CAT}}(\hat{b}, \hat{\lambda})) \approx 0.225$ as a conservative estimate of the knock-out probability. A recalculation of Table 3.2 leads to a conservative discounted value of CHF 237.15 for the three WINCAT coupons. This is only CHF 1.10 below the conservative value CHF 238.25 derived from (4.27).

It is possible to test the assumption of a Poisson distribution further by choosing an explicit alternative like a negative binomial distribution and considering the corresponding Neyman–Pearson test. In addition, we could choose a preferred measure of discrepancy for distributions and apply model selection criteria to come to a decision about the underlying distribution. In this paper, however, we want to pursue a different route, namely a possible deterministic time-inhomogeneity of the distribution of the numbers $N_{1987}, \ldots, N_{1996}$ of events per observation period. Concerning model selection in the case of independent and identically distributed random variables, we therefore refer the reader to [11], in particular to [11, Example 4.4.3], where the Poisson and the negative binomial distribution are the alternatives.

5.2. **Testing for time-inhomogeneity.** When looking at Figure 4.1 which shows the number of events in the ten observation periods causing more than 1 000 claims, we can ask whether there is something special about the order of the ten observations; in particular, whether the assumption of an identical distribution for the random variables $N_{1987}, \ldots, N_{1996}$ is justified.

Starting from $(0, 0, 0, 1, 1, 2, 2, 2, 4, 5)$, namely the ten observations in increasing order, we need 38 successive transpositions of adjacent entries of the tuple to rearrange it in decreasing order. To rearrange the observed tuple $(0, 0, 0, 2, 2, 4, 1, 5, 2, 1)$ into decreasing order, we need 28 successive transpositions of adjacent entries:

$$\begin{aligned}
(0,0,0,2,2,4,1,5,2,1) &\to (2,2,4,1,5,2,1,0,0,0) \quad 21 \text{ transpositions} \\
&\to (2,2,4,5,2,1,1,0,0,0) \quad 2 \text{ transpositions} \\
&\to (4,5,2,2,2,1,1,0,0,0) \quad 4 \text{ transpositions} \\
&\to (5,4,2,2,2,1,1,0,0,0) \quad 1 \text{ transposition}
\end{aligned} \tag{5.4}$$

Since the number of 28 transpositions is well above the half of 38, we can use this observation for a permutation test to find out whether the data shows a tendency to be arranged in increasing order.

Under the assumption that the ten observations are given by ten exchangeable, $\mathbb{N}_0$-valued random variables $N_{1987}, \ldots, N_{1996}$, every permutation of the ten observations has the same probability. If $N_{1987}, \ldots, N_{1996}$ are independent and identically distributed, then exchangeability follows. For every one of the

$$\frac{10!}{3!\,3!\,2!} = 50\,400 \tag{5.5}$$

different permutations of the ten observations, we can count the required number of successive transpositions of adjacent entries to obtain the decreasing order given by the tuple $(5, 4, 2, 2, 2, 1, 1, 0, 0, 0)$. This number is always between zero and 38. Figure 5.1 shows the resulting distribution function of this number.

Under the null hypothesis where all permutations of the ten observations have the same probability, only for 2 953 permutations out of 50 400, about 5.86 % of them, 28 or more transpositions of adjacent entries are needed to reach the decreasingly

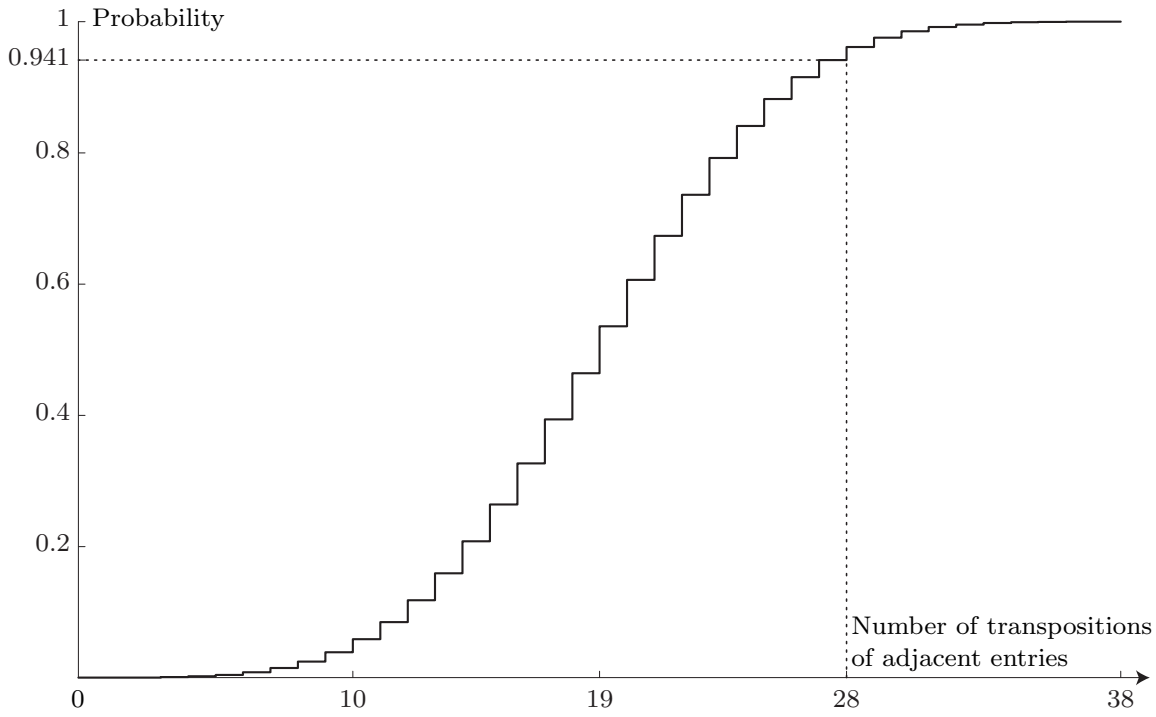

FIGURE 5.1. Distribution function of the number of successive transpositions of adjacent entries necessary to order a random permutation of the ten observations into decreasing order. For the observed data, 28 transpositions are necessary. At least 28 transpositions are necessary for about 5.86% of all permutations.

ordered tuple. (If there were a substantially higher number of permutations than 50 400, then a suitable number of random permutations would have to be generated in order to get an estimate for this percentage.)

Note that for the permutation test of this subsection we do not assume that the distribution of $N_{1987}, \ldots, N_{1996}$ lies in a certain class; in particular, the test is parameter-free. Furthermore, the test does not depend on the actual numbers but merely on their relative order or ranks; an observation like $(0, 0, 0, 3, 3, 4, 1, 7, 3, 1)$ would give the same test result. For such a distribution-free test and just ten observations, 5.86% is a remarkable result. However, as we can see from (5.4), it is mainly caused by the position of the three zero observations.

6. Fitting a generalised extreme value distribution

For a knock-out of a WINCAT coupon, only the most severe event within the corresponding observation period matters. We can use extreme value theory to model this event directly. The theoretical background for this approach is the Fisher–Tippett theorem (see for example [6, Theorem 3.2.3]), which identifies all possible limit distributions for properly scaled maxima $M(n) = \max\{X_1, \ldots, X_n\}$ of independent, identically distributed random variables $X_1, \ldots, X_n$ as $n \to \infty$. If the distributions of the properly scaled maxima do converge, then the limiting distribution is either a Fréchet, a Weibull or a Gumbel distribution. In the following we use the Jenkinson–von Mises representation of these extreme value distributions, see [6, Definition 3.4.1]. Let $\mu \in \mathbb{R}$ denote the location parameter, $\tau > 0$ the scale

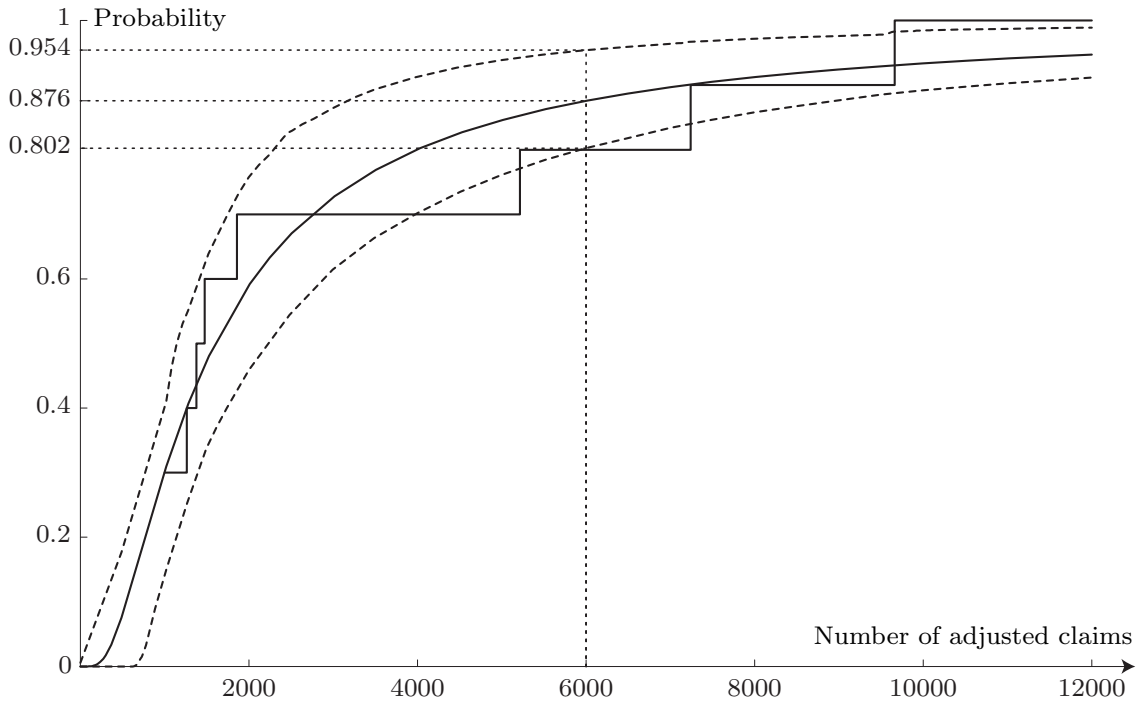

FIGURE 6.1. The censored empirical distribution function of the number of adjusted claims of the most severe event per year (solid step function) and the fitted generalised extreme value distribution (solid curve) with estimated exponent $1/\hat{\xi} \approx 1.316$, scale parameter $\hat{\tau} \approx 1008$ and location parameter $\hat{\mu} \approx 1168$. The estimated probability, that no knock-out event occurs within one year, is approximately 0.876. The two dashed curves, derived from 1000 bootstrap samples, indicate 68%-confidence intervals for the fitted generalised extreme value distribution.

parameter and $\xi \in \mathbb{R}$ the shape parameter. In the case $\xi > 0$, which corresponds to the Fréchet distribution, we define the distribution function $H_{\mu,\xi,\tau}$ by

$$H_{\mu,\xi,\tau}(x) = \begin{cases} \exp\bigl(-(1+\xi(x-\mu)/\tau)^{-1/\xi}\bigr), & \text{if } 1+\xi(x-\mu)/\tau > 0, \\ 0, & \text{otherwise.} \end{cases}$$

In the case $\xi < 0$, which corresponds to the Weibull distribution, we define similarly

$$H_{\mu,\xi,\tau}(x) = \begin{cases} \exp\bigl(-(1+\xi(x-\mu)/\tau)^{-1/\xi}\bigr), & \text{if } 1+\xi(x-\mu)/\tau > 0, \\ 1, & \text{otherwise.} \end{cases}$$

With the above representation, the Gumbel distribution

$$H_{\mu,0,\tau}(x) = \exp\bigl(-\exp(-(x-\mu)/\tau)\bigr), \qquad x \in \mathbb{R},$$

for the case $\xi = 0$ is actually the limit of $H_{\mu,\xi,\tau}$ as $\xi \to 0$.

When fitting the generalised extreme value distribution $H_{\mu,\xi,\tau}$ with $\mu, \xi \in \mathbb{R}$ and $\tau > 0$ to the observed maxima given in Table 1.1, we have to cope censored data. The most severe events of the years 1987–1989 are not given because they caused less than 1 000 adjusted claims (assuming that there were damages caused by storm or hail at all). Next, when we want to use the maximum-likelihood method

to estimate the parameters μ, ξ and τ, we encounter another problem: The density of $H_{\mu,\xi,\tau}$ is unbounded for $\xi < -1$ and $x \nearrow \mu - \tau/\xi$.

Both problems can be solved by discretizing the distribution $H_{\mu,\xi,\tau}$. The censored data for the years 1987–1989 corresponds to three observations in the interval $(0, 1000]$. The most severe events in the years 1990–1996 are adjusted claim numbers which correspond to intervals of the form $(n, n+1]$ with an integer $n \geq 1000$ (at least approximately, ignoring that the vehicles insured index in Table 1.1 is not always exactly one). This suggests the likelihood function

$$L(\mu, \xi, \tau) = \big(H_{\mu,\xi,\tau}(1000) - H_{\mu,\xi,\tau}(0)\big)^3 \times \prod_{y=1990}^{1996} \big(H_{\mu,\xi,\tau}(M_y + 1) - H_{\mu,\xi,\tau}(M_y)\big)$$

with $\mu, \xi \in \mathbb{R}$ and $\tau > 0$, where $M_{1990}, \ldots, M_{1996}$ denote the yearly maxima from Table 1.1. The numerical iteration procedure applied to the log-likelihood function leads to the maximum-likelihood estimates $\hat{\mu} \approx 1168$, $\hat{\xi} \approx 0.760$ and $\hat{\tau} \approx 1008$; the corresponding fit is shown in Figure 6.1. These values lead to an estimated knock-out probability of only $P_{\text{CAT}} = 1 - H_{\hat{\mu},\hat{\xi},\hat{\tau}}(6000) \approx 12.4\,\%$, because the fitted distribution is well above the empirical one at 6 000 in Figure 6.1. A recalculation of Table 3.2 gives a discounted value of CHF 266.62 for the three WINCAT coupons. For further background on parameter estimation for the generalised extreme value distribution, see [6, Section 6.3] and the references given there.

It would be unreasonable to insist on estimates for μ, ξ and τ giving a generalised extreme value distribution with support in $[0, \infty)$, because $H_{\hat{\mu},\hat{\xi},\hat{\tau}}(0) \approx 6.8 \cdot 10^{-8}$ is already a very good approximation of zero, the true distribution is almost certainly not in the family $\{\, H_{\mu,\xi,\tau} \mid \mu, \xi \in \mathbb{R},\ \tau > 0 \,\}$, and a good fit at this end of the distribution, where the data is censored anyway, is not our primary concern.

To estimate the 68.3 %-confidence intervals in Figure 6.1, we use the bootstrap method; see e. g. [5] for an introduction. We take 1 000 bootstrap samples $(M^*_{1987}, \ldots, M^*_{1996})$, where for each component the values $M_{1990}, \ldots, M_{1996}$ have probability $1/10$ of being chosen, and with probability $3/10$ we take a censored observation. For each bootstrap sample we calculate the corresponding maximum-likelihood estimate $(\hat{\mu}^*, \hat{\xi}^*, \hat{\tau}^*)$. This gives 1 000 bootstrap values for $H_{\hat{\mu}^*,\hat{\xi}^*,\hat{\tau}^*}(x)$, we take the 159th and the 841st largest values as boundaries for a 68.3 %-confidence interval for $H_{\hat{\mu},\hat{\xi},\hat{\tau}}(x)$. The estimated 68.3 %-confidence interval for the above knock-out probability is $[0.046, 0.198]$; the conservative estimate $P_{\text{CAT}} = 0.198$ leads to a discounted value of CHF 245.17 for the three WINCAT coupons.

The model of this section has a drawback in our case: While we explicitly use the absence of recorded events in 1987–1989, we are partially discarding 10 of the 17 adjusted claim numbers given in Table 1.1; we only implicitly use that they do not exceed the corresponding maxima.

7. Composite Poisson models with a time-dependent parameter

The constant-parameter composite Poisson models of Section 4 are static ones. They give equal weight to every recorded event and, by construction, do not allow to discover a trend in the data. Every redistribution of the 17 events in Table 1.1 to the ten observation periods would lead to the same result for the coupon values (if we disregard the $15/17$-correction in Table 3.2). However, the tests for overdispersion and time-inhomogeneity from Subsections 5.1 and 5.2 suggest—although

not significantly on the 5%-level but very close—to consider the possibility of a time-dependent distribution. Such a deterministic time-dependence can account for the tendency of over-dispersion considered in Subsection 5.1, it doesn't need to be a randomly varying Poisson parameter as mentioned in Subsection 5.1. In particular, an investor might want to take a possible trend into account when estimating the discounted value of the WINCAT coupons. Even when a constant-parameter model is preferred for pricing the WINCAT coupons, a model capable to accommodate a possible trend can be useful for risk management, because model risk can be an important risk factor. In Section 8, we shall test for the existence of a trend within most of these models we are considering below. There are several reasons why there might be a trend, for example:

- The variability of the weather could change, due to human influence (increased CO_2-part in the atmosphere) or solar activity (11-year cycle of sun spots), for example.
- Winterthur might increase its market share in other regions like the French or Italian speaking parts of Switzerland; this can happen in particular when Winterthur merges with another insurance company (like merging with Neuenburger Schweizerische Allgemeine Versicherungsgesellschaft in 1997, for example). Due to the Swiss Alps, the local climate is in general quite different in different regions of Switzerland, so a change in Winterthur's engagement in a particular region can considerably increase or decrease the company's exposure to storm or hail damages.
- Severe damage caused by hail is a local event. If the density of motor vehicles insured with Winterthur increases (due to more cars per inhabitant, more inhabitants per area or a greater market share of Winterthur Insurance within an area), then more insured motor vehicles are likely to be damaged in every single event.
- The relation of the number of cars to the number of motor-cycles within Winterthur's insurance portfolio of motor vehicles might change.
- The habits of the insured might change. They might buy a second or third car for the family without building or renting an additional garage to protect the car in case of bad weather. Or the insured are better off financially and they can afford the deductable, hence they take chances and don't drive the car to a secure place in case of a storm/hail forecast.
- Motor vehicles might get more or less susceptible of hail damage, because the material changes (different kinds of steel, aluminium, different coats of lacquer, for example) or the thickness of the automobile body sheet changes (a thicker sheet can give more protection in case of an accident, a thinner sheet reduces weight and thereby fuel consumption).

In any case—whatever the particular reason—it is a reasonable idea to consider a model which is flexible enough to take a possible trend in the data into account as long as such a possible trend can not be ruled out by additional information concerning all the points mentioned above (and the ones we have not thought of).

When modelling low-frequency event risks, the scarcity of the available statistical data is a typical problem. If one wants to follow a kind of Bayesian approach, it is desirable to take additional information into account when selecting a model (see [13, Section 6] for such a case study of the correlation of wind storm losses of

the Swedish insurance group Länsförsäkringar with wind speed data provided by the Swedish Meteorological and Hydrological Institute). For a fair and transparent pricing of financial products, such information should either be public or should be published together with the introduction of the financial product. For the pricing of the WINCAT coupons, such additional information besides the historical data of Table 1.1 is contained in the study [21] of Winterthur Insurance. This study, as well as the publicly available report [14], for example, provide information on the variability of the weather; they also describe the development of hail storms, the different frequency of hail storms in the various parts of Switzerland, and the properties of hailstones (size, shape, speed) that cause damage to motor vehicles. The study [21] also points out that damages to agriculture and motor vehicles are mainly caused by different types of hail storms: damage to motor vehicles requires a large momentum of the hailstones (large product of mass and speed), while damage to agriculture can already be caused by small but numerous hailstones. This indicates that the extensive statistical data collected from insured damages to agriculture since 1881 is of limited use when estimating a possible trend in the frequency or severity of damages to motor vehicles caused by hail. Also in the report [14], the severity of hail storms is measured by the number of communities reporting damages to agriculture.

In the following subsections we shall use, for every year $y \in \{1987, \dots, 1996\}$, a random variable N_y describing the number of calendar days within the observation period ending in year y, during which more than 1 000 adjusted claims are caused by storm or hail. We assume that these random variables are independent and that every N_y has a Poisson distribution given by (4.1), but with a parameter $\lambda(y)$ depending on the year $y \in \{1987, \dots, 1996\}$. For the purpose of nicer graphics, we shall treat y as a continuous variable within the figures. We shall discuss five different choices for the dependence $y \mapsto \lambda(y)$.

7.1. **Linear trend of the parameter.** To start with the apparently simplest dependence, we assume that the Poisson parameter for the number of events with more than 1 000 adjusted claims depends linearly on the year y, namely

$$\lambda_{\alpha,\beta}(y) = \alpha + \beta(y - 1987), \tag{7.1}$$

where we subtract 1987 to get reasonable numbers for α. When using (7.1), we have to make sure that $\lambda_{\alpha,\beta}(y) \geq 0$ for all years under consideration. This will certainly be the case when $\alpha, \beta \geq 0$. The corresponding likelihood function arising from the ten observations $N_{1987}, \dots, N_{1996}$ is

$$L(\alpha, \beta) = \prod_{y=1987}^{1996} \text{Poisson}_{\lambda_{\alpha,\beta}(y)}(N_y) \tag{7.2}$$

with the Poisson distribution given by (4.1) and the parameter $\lambda_{\alpha,\beta}(y)$ as in (7.1). When trying to calculate the maximum-likelihood estimators for α and β numerically, it turns out that for the given data there is no simultaneous solution of

$$\frac{\partial}{\partial\alpha} L(\alpha, \beta) = 0 \qquad \text{and} \qquad \frac{\partial}{\partial\beta} L(\alpha, \beta) = 0 \tag{7.3}$$

satisfying $\alpha \geq 0$. As a pragmatic approach, let us set $\alpha = 0$. This means we consider the special case where the Poisson parameter $\lambda_\beta(y)$ depends on β in the

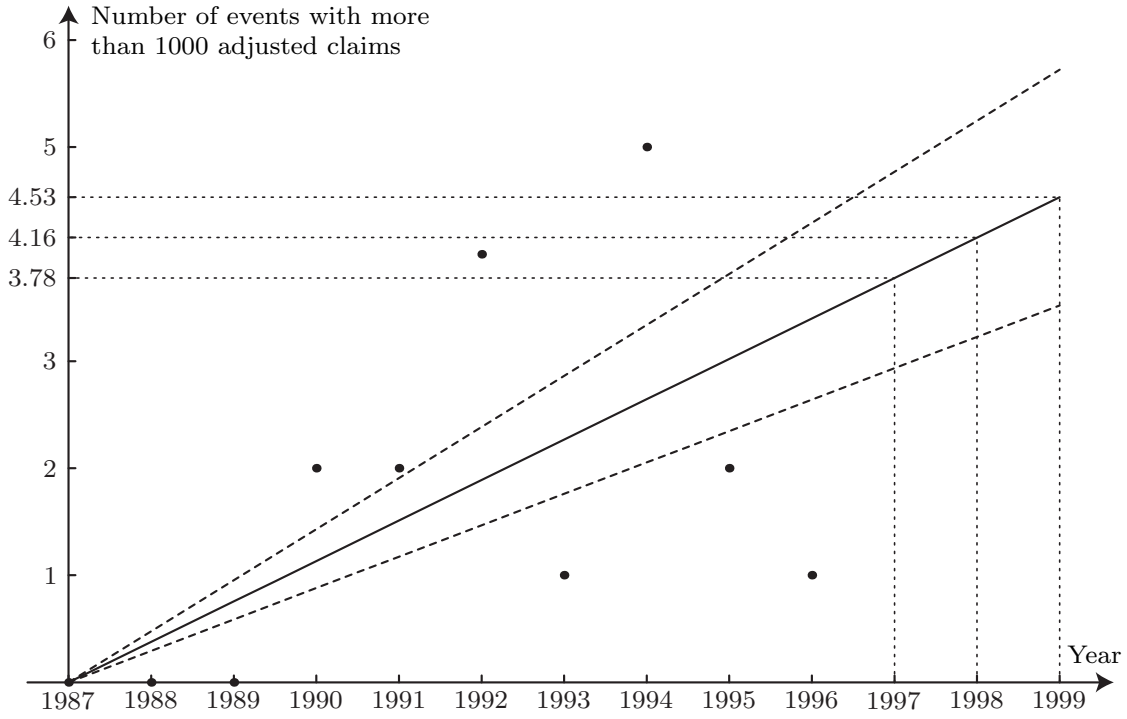

FIGURE 7.1. Observed number of events causing more than 1 000 adjusted claims. A linear fit of the intensity $\lambda_\beta(y) = \beta(y-1987)$, using the maximum-likelihood method, leads to $\hat{\beta} = 17/45 \approx 0.378$. The increasing dashed lines indicate this estimated standard deviation of $\lambda_{\hat{\beta}}(y)$. This model has a problem with the years up to 1987 and it certainly underestimates the standard deviation in the first years.

form $\lambda_\beta(y) = \beta(y - 1987)$. In this case the equation $\frac{\partial}{\partial\beta} \log L(0, \hat{\beta}) = 0$ for the maximum-likelihood estimator $\hat{\beta}$ can be solved explicitly, leading to

$$\hat{\beta} = \frac{\sum_{y=1988}^{1996} N_y}{\sum_{y=1988}^{1996} (y - 1987)} = \frac{1}{45} \sum_{y=1988}^{1996} N_y = \frac{17}{45} \approx 0.378\,. \tag{7.4}$$

The corresponding straight line is shown in Figure 7.1. Extrapolation to the years 1997–1999 gives the estimated values for $\lambda_{\hat{\beta}}(y)$ contained in Figure 7.1 and Table 7.1. Using these extrapolated Poisson parameters and the conditional probability $\hat{p}_{6000} \approx 0.0757$ from (4.32), which was estimated by a generalised Pareto distribution, the knock-out probabilities can be calculated as in (4.26) by the formula

$$P_{\text{CAT}}(y) = 1 - \exp\bigl(-\hat{p}_{6000} \cdot \lambda_{\hat{\beta}}(y)\bigr) \tag{7.5}$$

for $y \in \{1997, 1998, 1999\}$. The results are given in the fourth column of Table 7.1. Applying the $^{15}/_{17}$-correction of (2.1) to $P_{\text{CAT}}(1997)$ and inserting the resulting coupon-dependent knock-out probabilities into Table 3.2, a recalculation of this table leads to the discounted values of the three WINCAT coupons. These values are given in the last column of Table 7.1. The sum of these discounted values of the three WINCAT coupons is CHF 223.88.

To estimate the standard deviation of the Poisson parameter $\lambda_{\hat{\beta}}(y)$, note that $\text{Var}(N_y) = \lambda_\beta(y) = \beta(y - 1987)$ for every one of the independent random variables

Year y	$\lambda_{\hat\beta}(y)$	$\hat p_{6000}$	$P_{\text{CAT}}(y)$	Coupon value
1997	3.78	0.0757	24.9%	CHF 80.64
1998	4.16	0.0757	27%	CHF 73.72
1999	4.53	0.0757	29%	CHF 69.52
Discounted value of the three WINCAT coupons:				CHF 223.88

TABLE 7.1. Calculation of the discounted value of the three WINCAT coupons in the case of a linear dependence $\lambda_\beta(y) = \beta(y - 1987)$ of the Poisson parameter. The Poisson parameters $\lambda_{\hat\beta}(y)$ are the extrapolated values from Figure 7.1. The conditional probability $\hat p_{6000}$ for a knock-out event, given that an event occurs, is taken from (4.32). The forth column contains $P_{\text{CAT}}(y) = 1-\exp(-\hat p_{6000}\,\lambda_{\hat\beta}(y))$. The discounted coupon values are then calculated according to Table 3.2 taking into account the $^{15}/_{17}$-correction from (2.1) for the shorter first observation period.

$N_{1987}, \dots, N_{1996}$ in this model. By (7.4),

$$\operatorname{Var}(\hat\beta) = \frac{1}{45^2}\sum_{y=1988}^{1996}\operatorname{Var}(N_y) = \frac{\beta}{45^2}\sum_{y=1988}^{1996}(y-1987) = \frac{\beta}{45}.$$

Using (7.4), this leads to $\hat\sigma(\lambda_{\hat\beta}(y)) = \hat\sigma(\hat\beta)(y-1987)$ with $\hat\sigma(\hat\beta) = \sqrt{17}/45 \approx 0.0916$.

This model with a linear trend in the Poisson parameter $y \mapsto \lambda_\beta(y)$ has a severe problem with the year 1987, because the estimate $\lambda_{\hat\beta}(1987) = 0$ is certainly wrong. The estimated standard deviations for the years 1987–1989, as shown in Figure 7.1, are quite unrealistic, too. Model predictions for the years before 1987 are impossible, because negative values for $\lambda_{\hat\beta}(y)$ are unacceptable. In the following subsections we shall discuss models which do not have these deficiencies.

7.2. **Log-linear trend of the parameter.** To avoid the problem of negative Poisson parameters, let us consider the prime example of a model where this cannot occur, namely a generalised linear model with the log-linear dependence

$$\lambda_{\alpha,\beta}(y) = \exp(\alpha + \beta(y - 1991.5)), \qquad \alpha, \beta, y \in \mathbb{R}. \tag{7.6}$$

We subtract 1991.5 from y in order to get approximately orthogonal parameters, meaning that the maximum-likelihood estimators for α and β have only a small correlation (for the notion of orthogonal parameters, see e. g. [2, p. 182–185]). This parametrisation is also useful in Section 8, because it introduces a symmetry which reduces the computational effort for the permutation test. The log-likelihood function for this model arising from (7.2) with $\lambda_{\alpha,\beta}(y)$ as in (7.6) is given by

$$l(\alpha,\beta) = \sum_{y=1987}^{1996}\big(N_y(\alpha + \beta(y - 1991.5)) - e^{\alpha+\beta(y-1991.5)} - \log N_y!\big). \tag{7.7}$$

The maximum-likelihood estimates for α and β, calculated numerically, are

$$\hat\alpha \approx 0.406 \qquad \text{and} \qquad \hat\beta \approx 0.176\,. \tag{7.8}$$

The corresponding curve $y \mapsto \lambda_{\hat\alpha,\hat\beta}(y)$ is shown in Figure 7.2. It approximates quite well within the time span 1987–1996. The extrapolated values of the Poisson parameter $\lambda_{\hat\alpha,\hat\beta}(y)$ for the years 1997–1999 can be read off from Figure 7.2. A

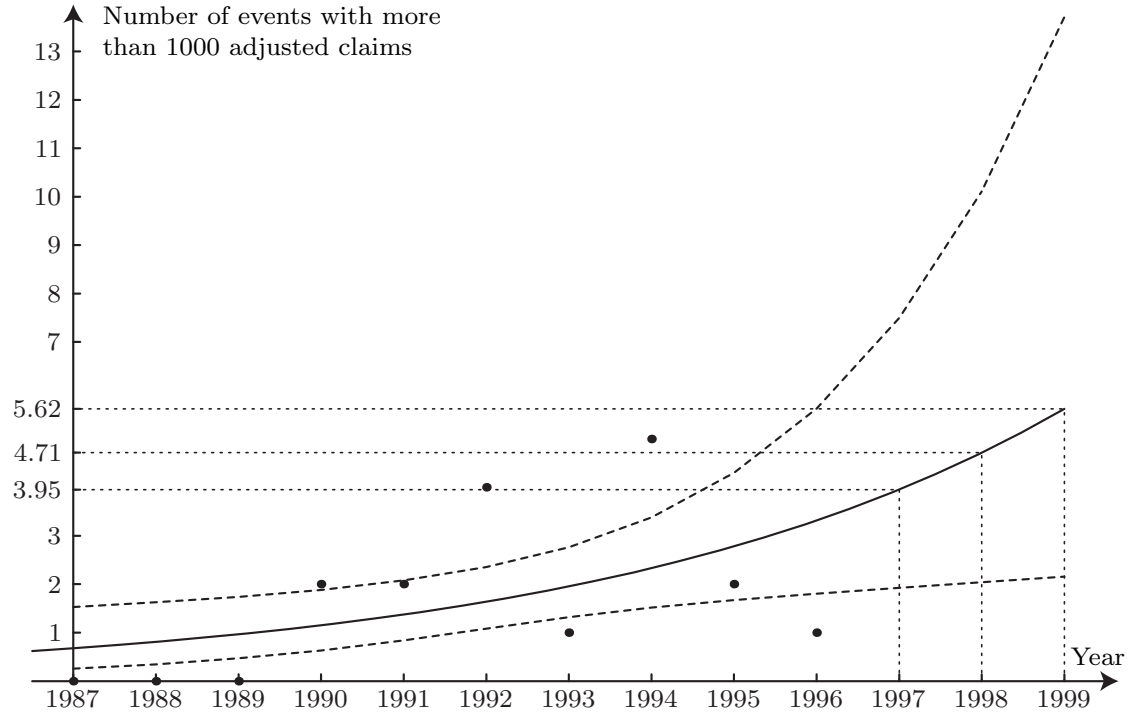

FIGURE 7.2. Log-linear dependence $\lambda_{\alpha,\beta}(y) = \exp(\alpha + \beta(y - 1991.5))$ of the Poisson parameter. The maximum-likelihood method leads to $\hat{\alpha} \approx 0.406$ and $\hat{\beta} \approx 0.176$, the result is shown as a solid curve together with the extrapolation. The dashed piecewise linear curves indicate for every year y the estimated standard deviation of $\lambda_{\hat{\alpha},\hat{\beta}}(y)$ as derived from the log-likelihood ratio statistic. The uncertainty of the extrapolated values is very large for the years 1997–1999.

recalculation of Table 7.1 with these numbers leads to the knock-out probabilities 25.9% (without $^{15}/_{17}$-correction), 30% and 34.7% for the three WINCAT coupons and a discounted value of CHF 214.37.

For calculating the maximum-likelihood estimates (7.8), we needed appropriate starting values for the numerical iteration procedure. We took $\alpha_0 = \log \lambda_{1000}^{\text{const}} = \log {}^{17}/_{10} \approx 0.53$ with $\lambda_{1000}^{\text{const}}$ from (4.2). This value for α is the correct choice in the constant-parameter case $\beta = 0$. For β we used the following heuristic: The approximating tangent of $y \mapsto \lambda_{\alpha_0,\beta}(y)$ at the middle $y = 1991.5$ of the interval $[1987, 1996]$ is $\lambda_{\alpha_0,\beta}^{\text{tangent}}(y) = e^{\alpha_0} + \beta e^{\alpha_0}(y - 1991.5)$ with $e^{\alpha_0} = {}^{17}/_{10}$. The optimal β for the least-squares fit of this tangent to the data $N_{1987}, \ldots, N_{1996}$ has to satisfy

$$\frac{\partial}{\partial \beta} \sum_{y=1987}^{1996} \left(N_y - \lambda_{\alpha_0,\beta}^{\text{tangent}}(y)\right)^2 = 0.$$

This linear equation in β is solved by

$$\beta_0 = \frac{\sum_{y=1987}^{1996} N_y (y - 1991.5)}{e^{\alpha_0} \sum_{y=1987}^{1996} (y - 1991.5)^2} = \frac{94}{561} \approx 0.168 \,.$$

To obtain an estimate for the standard deviation of the estimated Poisson parameter $\lambda_{\hat{\alpha},\hat{\beta}}(y)$ for every year $y \in \{1987, \ldots, 1999\}$, we use the log-likelihood ratio

statistic, which we already applied in Subsection 4.3. Similarly to (4.33), we define the log-likelihood ratio statistic or deviance by $D(\alpha,\beta) = 2l(\hat\alpha,\hat\beta) - 2l(\alpha,\beta)$ for $\alpha,\beta\in\mathbb{R}$. Corresponding to (4.34), for every year $\tilde y \in \{1987,\dots,1999\}$, we want to determine the smallest 68%-confidence interval $[\lambda^-_{\tilde y},\lambda^+_{\tilde y}]\subset(0,\infty)$ such that

$$\big\{\,(\alpha,\beta)\in\mathbb{R}^2 \;\big|\; D(\alpha,\beta)\le \chi^2_{2,0.32}\,\big\} \subset \big\{\,(\alpha,\beta)\in\mathbb{R}^2 \;\big|\; \lambda_{\alpha,\beta}(\tilde y)\in[\lambda^-_{\tilde y},\lambda^+_{\tilde y}]\,\big\}. \quad (7.9)$$

Solving equation (7.6) for α yields $\alpha_{\tilde y}(\beta,\lambda) = \log\lambda - \beta(\tilde y - 1991.5)$. With this reparametrisation, the inclusion (7.9) reduces to

$$\big\{\,(\beta,\lambda)\in\mathbb{R}\times(0,\infty) \;\big|\; D(\alpha_{\tilde y}(\beta,\lambda),\beta)\le \chi^2_{2,0.32}\,\big\} \subset \mathbb{R}\times[\lambda^-_{\tilde y},\lambda^+_{\tilde y}]$$

The numerical results are shown in Figure 7.2 as dashed piecewise linear curves.

The results in Figure 7.2 for the years 1987 up to 1996 look quite satisfactory, although the 68%-confidence intervals are larger than the estimated standard deviation $\hat\sigma(\lambda^{\text{const}}_{1000})$ in the constant-parameter model, see Figure 4.1. (However, since the inclusion (7.9) is strict in general, we may have slightly overestimated the size of the confidence intervals here.) In Figure 7.2, the upper 68%-confidence bounds λ^+_{1997}, λ^+_{1998} and λ^+_{1999} for the future observation periods show a large uncertainty of the estimates. Of course, the small size of the historical data set is partially responsible for this uncertainty. The main contribution, however, comes from the log-linear model itself, because it blows up the unavoidable uncertainty of the maximum-likelihood estimators $\hat\alpha$ and $\hat\beta$ in an exponential way. This log-linear model is already a very pessimistic one with respect to the future development of the event frequency. Due to the exponential amplification of the estimator uncertainty, the log-linear model is certainly not the favourite one for calculating a conservative estimate for the value of the WINCAT coupons.

7.3. **Square-root linear trend of the parameter.** To avoid the possibly negative Poisson parameters of the linear model from Subsection 7.1 and the very pessimistic perspective of the future event frequency in the log-linear model from Subsection 7.2, we want to consider the usual root-linear model

$$\lambda_{\alpha,\beta}(y) = (\alpha+\beta(y-1991.5))^2, \qquad \alpha,\beta,y\in\mathbb{R}.$$

Along the lines of the previous subsection, we obtain Figure 7.3. A recalculation of Table 7.1 leads to the knock-out probabilities 27.8% (without $^{15}/_{17}$-correction), 31.4% and 35% for the WINCAT coupons and a discounted value of CHF 210.86.

Incidentally, note that the maximum-likelihood estimators $\hat\alpha$ and $\hat\beta$ are not uniquely defined in this square-root linear model: If $(\hat\alpha,\hat\beta)$ maximises the likelihood function, then so does $(-\hat\alpha,-\hat\beta)$. The starting values for α and β determine which of these solutions is found by the numerical iteration procedure.

In this square-root linear model, the estimated Poisson parameter drops to zero between 1983 and 1984 and increases in the more distant past. This model deficiency, however, is not of great importance for the extrapolation into the future. A more important problem is, as in the log-linear model of Subsection 7.2, the built-in pessimistic perspective of a future quadratic growth of the event frequency. In the next subsection we shall present a model which is better tailored for the extrapolation of the estimated Poisson parameter in our case.

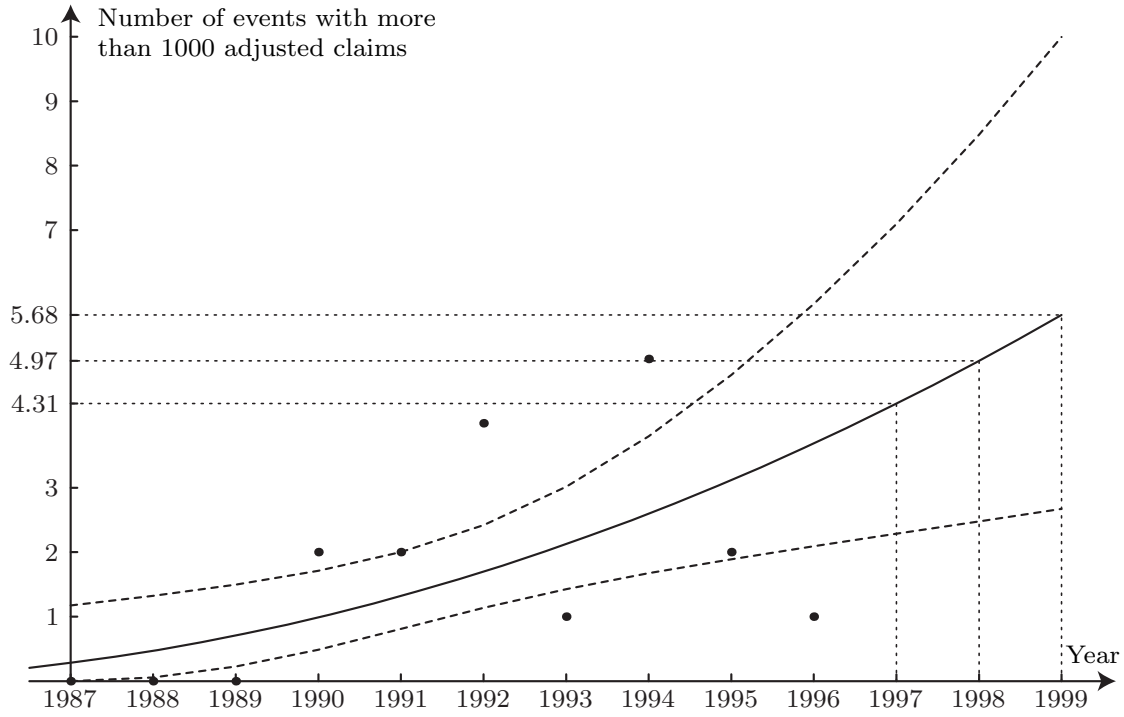

FIGURE 7.3. Root-linear dependence $\lambda_{\alpha,\beta}(y) = (\alpha + \beta(y - 1991.5))^2$ of the Poisson parameter. The maximum-likelihood method leads to $\hat{\alpha} \approx 1.23$ and $\hat{\beta} \approx 0.154$. These estimates give the solid curve and the extrapolated values. The 68%-confidence bounds are indicated by the dashed piecewise linear curves.

7.4. **Modified-linear trend of the parameter.** To avoid the problems of negative Poisson parameters and too pessimistic perspectives of the future event frequency, we want to consider the dependence

$$\lambda_{\alpha,\beta}(y) = \log\bigl(1 + \exp(\alpha + \beta(y - 1991.5))\bigr), \qquad \alpha, \beta, y \in \mathbb{R}. \tag{7.10}$$

We shall call it *modified linear,* because $y \mapsto \lambda_{\alpha,\beta}(y)$ is approximately linear for $\alpha + \beta(y - 1991.5) \gg 0$ and the graph bends smoothly for $\alpha + \beta(y - 1991.5) \approx 0$ to avoid negative values. The number one in (7.10) arises from this restriction; another value, let us call it γ, would lead to the asymptotic value $\log \gamma$ when $\alpha + \beta(y - 1991.5) \ll 0$. Maximum-likelihood estimators and 68%-confidence intervals are calculated with the procedures outlined in Subsections 7.1 and 7.2. The results are shown in Figure 7.4, the maximum-likelihood estimates are $\hat{\alpha} \approx 1.36$ and $\hat{\beta} \approx 0.521$. Figure 7.4 also shows that the estimator uncertainty is amplified only in a linear way. A recalculation of Table 7.1 with the extrapolated values of the Poisson parameter from Figure 7.4 lead to the knock-out probabilities 27.4% (without $^{15}/_{17}$-correction), 30.2% and 32.9% for the three WINCAT coupons and a discounted value of CHF 214.44.

7.5. **Smooth transition of the parameter.** The models from the previous subsections can be criticised in the sense that an extrapolation far into the future gives unreasonable results. Of course, such a far-reaching extrapolation should not be done in the current case with just ten observations, and the models of the previous subsections were not chosen for this purpose. If future observations would confirm

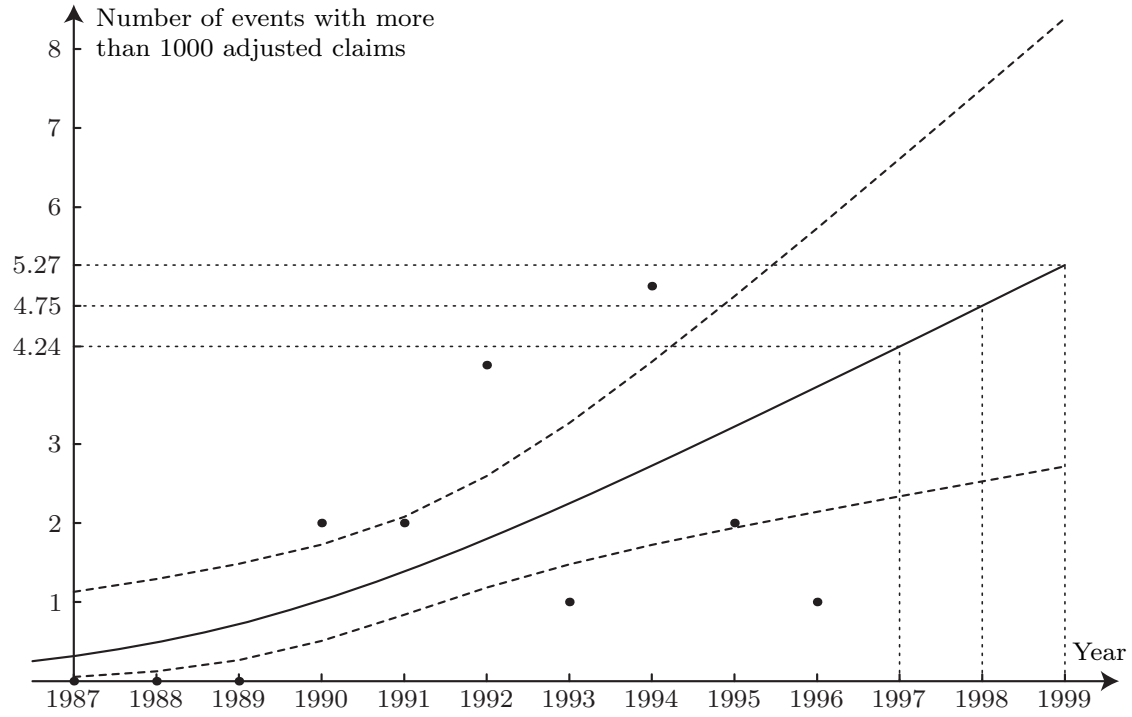

FIGURE 7.4. Observed number of events with more than 1 000 adjusted claims together with a modified-linear fit of the Poisson parameter for such events using $\lambda_{\alpha,\beta}(y) = \log(1 + \exp(\alpha + \beta(y - 1991.5)))$. The maximum-likelihood method leads to $\hat{\alpha} \approx 1.36$ and $\hat{\beta} \approx 0.521$. The dashed piecewise linear curves indicate the estimated 68%-confidence interval for $\lambda_{\hat{\alpha},\hat{\beta}}(y)$ as derived from the log-likelihood ratio statistic. The uncertainty of the estimates grows only linearly.

an increasing trend of the strength estimated in the previous subsections, then Winterthur Insurance would have a strong incentive to introduce preventive measures. Notice that the WINCAT coupons provide a reinsurance on the financial market for only half the amount necessary for the adjustment of damages arising from just *one* event per observation period with at least 6 000 adjusted claims. Hence, let us also consider a model for a smooth transition between two limiting intensities, where the speed and the direction of the transition is measured by a parameter β. While zero is certainly a natural lower limit for the intensities, there is some ambiguity for the upper bound. A plausible ad hoc choice based on the data is the largest observation. To avoid an exponentially fast convergence to the limiting values, we do not propose a scaled version of the normal or the logistic distribution function but take a scaled version of the Cauchy distribution function instead:

$$\lambda_{\alpha,\beta}(y) = 5\Big(\frac{1}{2} + \frac{1}{\pi}\arctan(\alpha + \beta(y - 1991.5))\Big), \qquad \alpha, \beta, y \in \mathbb{R}. \tag{7.11}$$

Maximum-likelihood estimators and 68%-confidence intervals are calculated with the procedures outlined in Subsections 7.1 and 7.2. The results are shown in Figure 7.5, the values of the estimators are $\hat{\alpha} \approx -0.466$ and $\hat{\beta} \approx 0.442$. A recalculation of Table 7.1 with the extrapolated values of the Poisson parameter from Figure 7.5 lead to the knock-out probabilities 27.5% (without $^{15}/_{17}$-correction), 28.2% and 28.7% for the three WINCAT coupons and a discounted value of CHF 220.53.

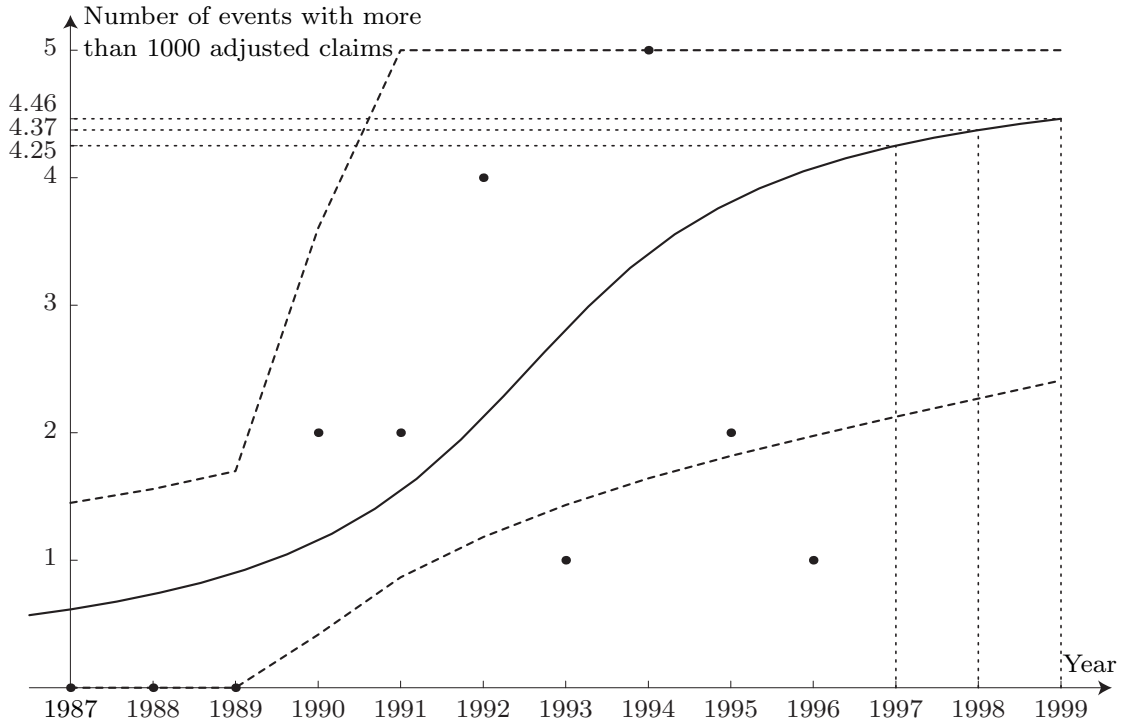

FIGURE 7.5. A smooth transition of the Poisson parameter according to (7.11). The possible range of the parameter is determined by the smallest and largest observed value, respectively. The maximum-likelihood method leads to $\hat{\alpha} \approx -0.466$ and $\hat{\beta} \approx 0.442$. The dashed piecewise linear curves indicate the estimated 68%-confidence intervals for $\lambda_{\hat{\alpha},\hat{\beta}}(y)$ as derived from the log-likelihood ratio statistic. These intervals clearly show the artificial restriction of the range of the Poisson parameter.

The drawback of this model is illustrated by the 68%-confidence intervals given in Figure 7.5. For the years 1990–1994, they are considerably larger than the ones in Figure 7.4, and starting from 1991, one clearly sees the influence of the artificial model assumption about the upper bound.

8. Testing for a positive trend in the Poisson parameter

For the various models with a time-dependent Poisson parameter discussed in Section 7, it is of interest to check whether the additional parameter β for modelling a trend is significantly different from zero. For this purpose we shall assume that there is no trend in the data, and try to determine the probability that the estimated trend parameter $\hat{\beta}$ takes values which are at least as large as the observed one.

The model with a linear trend in the Poisson parameter given by (7.1) has a serious deficiency with respect to the positivity of the Poisson parameter as pointed out in Subsection 7.1. However, we can test whether the observations $N_{1987}, \ldots, N_{1996}$ exhibit a linear dependence on the year of the observation period. For this test, as for the one in Subsection 5.2, we assume that $N_{1987}, \ldots, N_{1996}$ are exchangeable so that every permutation of the ten observations has the same probability. We do not assume that $N_{1987}, \ldots, N_{1996}$ are independent or that they have a Poisson distribution. For every permutation π of the ten years 1987, ..., 1996, which gives

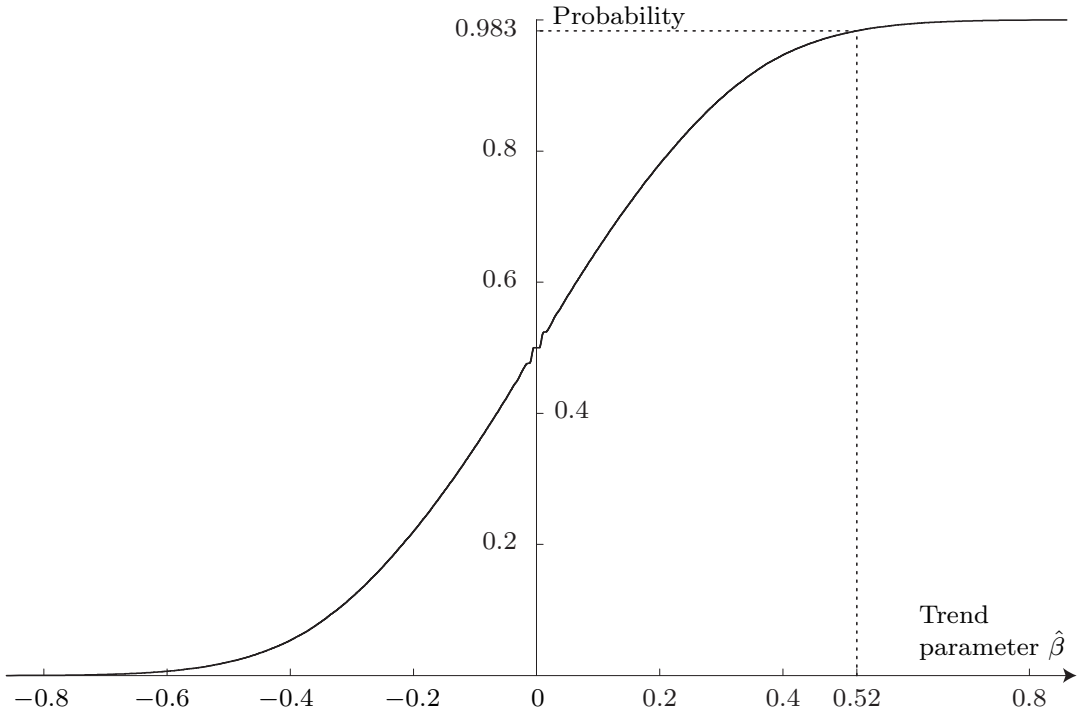

FIGURE 8.1. Distribution function of the estimated trend parameter $\hat{\beta}$ in the modified-linear model under the assumption that there is no trend, i. e., all permutations of the ten observations have the same probability. Under this assumption the probability for $\hat{\beta} \geq 0.52$ is about 1.66%.

a different sequence of observations, we calculate the value of

$$\varphi(\pi) = \sum_{y=1987}^{1996} N_{\pi(y)}(y - 1991.5). \tag{8.1}$$

This test statistic has a symmetric distribution under the above null hypothesis and attains all 88 possible values of the form $k + \frac{1}{2}$, $k \in \mathbb{Z}$, between -43.5 and 43.5. When the permutation π is the identity id, then $N_{\pi(1987)}, \ldots, N_{\pi(1996)}$ are given by the observed values $(0, 0, 0, 2, 2, 4, 1, 5, 2, 1)$ and we obtain $\varphi(\mathrm{id}) = 23.5$. Working through all 50 400 different permutations from (5.5), it turns out that for 3 726 of them, about 7.39%, the value of $\varphi(\pi)$ is larger or equal to $\varphi(\mathrm{id}) = 23.5$. (Due to symmetry, it suffices to consider only one half of these permutations in our case. If there were a substantially higher number of permutations, then a suitable number of random permutations would have to be generated.)

For the remaining tests of this section we assume that $N_{1987}, \ldots, N_{1996}$ are independent and identically distributed according to a Poisson distribution, meaning that the trend models are correct with $\beta = 0$. This no-trend assumption implies in particular that every permutation π of $N_{1987}, \ldots, N_{1996}$ has the same probability. The maximum-likelihood estimator $\hat{\beta}_\pi$ corresponding to $N_{\pi(1987)}, \ldots, N_{\pi(1996)}$ via one of the trend models is then a function of the random permutation π, and we can determine the percentage of all permutations which lead to a value of $\hat{\beta}_\pi$ larger or equal to the observed value for $\hat{\beta}_{\mathrm{id}}$.

The model with the log-linear trend (7.6) of the Poisson parameter is not more sensitive to permutations than the test statistic (8.1): For a permutation π of the years 1987–1996 we can define the log-likelihood function l_π similarly to (7.7) by

$$l_\pi(\alpha,\beta) = \sum_{y=1987}^{1996} \big(N_{\pi(y)}(\alpha+\beta(y-1991.5)) - e^{\alpha+\beta(y-1991.5)} - \log N_{\pi(y)}!\big)$$

for $\alpha,\beta\in\mathbb{R}$. Using (8.1) and the associativity of summation to rewrite $l_\pi(\alpha,\beta)$, we see that $l_\pi(\alpha,\beta) = \beta\varphi(\pi) + \sum_{y=1987}^{1996}\big(\alpha N_y - e^{\alpha+\beta(y-1991.5)} - \log N_y!\big)$, which means that all the dependence of the log-likelihood function l_π on the permutation π is contained in the test statistic $\varphi(\pi)$ and no new information can be obtained by calculating the maximum-likelihood estimator $\hat\beta_\pi$ corresponding to l_π.

A similar permutation test for the model from Subsection 7.3 with a square-root linear trend of the Poisson parameter is meaningless. As already noted in Subsection 7.3, if $(\hat\alpha,\hat\beta)$ maximises the likelihood function, then so does $(-\hat\alpha,-\hat\beta)$. Hence we should only consider the absolute value of the estimator $\hat\beta$. Furthermore, a large value such as $|\hat\beta|\approx 0.2486$ for the maximum-likelihood estimator arises from an increasing tuple like $(0,0,0,1,1,2,2,2,4,5)$ and also from a decreasing tuple like $(5,4,2,2,2,1,1,0,0,0)$. An even larger value of $|\hat\beta|\approx 0.4535$ arises from a "$\cup$-shaped" tuple like $(5,2,1,1,0,0,0,2,2,4)$. A corresponding "$\cap$-shaped" tuple like $(0,1,2,2,5,4,2,1,0,0)$ gives $|\hat\beta|\approx 0.047$ however. Therefore, large values of $|\hat\beta|$ are not equivalent to a large "rate of change" in the tuple. Nonetheless, for the curious reader: About 24.7% of all permutations result in $|\hat\beta|\geq 0.154$.

We can use the permutation test to challenge the no-trend hypothesis within the modified-linear trend model. For every permutation π of the 50 400 ones, which lead to different sequences of the ten observations, we determine the estimator $\hat\beta_\pi$ which maximises the corresponding log-likelihood function

$$l_\pi(\alpha,\beta) = \sum_{y=1987}^{1996} \log\big(\mathrm{Poisson}_{\lambda_{\alpha,\beta}(y)}(N_{\pi(y)})\big) \tag{8.2}$$

with $\lambda_{\alpha,\beta}(y)$ given by (7.10). The corresponding distribution function for $\hat\beta$, viewed as a random variable depending on π, is given in Figure 8.1. Note that $\hat\beta$ is quite sensitive to the permutations, because—except close to zero—the distribution function in Figure 8.1 looks "smooth" compared to the one in Figure 5.1 or the one with 88 jumps which would arise from the test statistic (8.1). The probability under the no-trend assumption for $\hat\beta\geq 0.52$ is about 1.66%, which is quite significant.

For the permutation test in the model with a smooth transition of the Poisson parameter, we also determine, for every permutation π, the estimator $\hat\beta_\pi$ which maximises the log-likelihood function (8.2) with $\lambda_{\alpha,\beta}(y)$ given by (7.11). The corresponding distribution function looks similar to the one in Figure 8.1 in the sense that it is "smooth" and therefore sensitive to the permutations. There are 1 504 permutations of the 50 400 different ones, about 2.98% of them, which lead to a maximum-likelihood estimate $\hat\beta\geq 0.442$. Again, the result is quite significant for the small data set.

Note that the various trend models "measure trend" by the parameter β in different ways, therefore it is no surprise that the test results depend on the used model. More explicitly, every model defines a map $\pi\mapsto\hat\beta_\pi$ on the set of all

permutations of the ten observations, and if $\hat{\beta}_\pi \leq \hat{\beta}_{\pi'}$ for two permutations π and π' in one trend model, then the corresponding inequality does not need to hold for the maximum-likelihood estimators in another trend model.

9. Composite Poisson model with a change-point

The models in Section 7 allowed to take a continuously changing Poisson parameter into account by choosing a non-vanishing trend parameter β. By extrapolating far enough into the future, these models—with the exception of the smooth-transition model of Subsection 7.5—lead to extrapolated values of the Poisson parameters which are larger than any value observed so far. Since the Poisson parameter is equal to the mean of the corresponding Poisson distribution, such large values might not be desirable or might be too pessimistic. While it is desirable that more recent observations have a greater influence than older observations for estimating the value of the WinCat coupons, new observations can have a considerable impact on these estimated values when one of the models of Section 7 is chosen; see Section 12 for a more detailed discussion. If one is willing to accept a discontinuously changing Poisson parameter, then a compromise between the constant parameter models of Section 4 and the trend models of Section 7 can be considered, namely a model with a change-point in the Poisson parameter.

Let $y_c \in \{1988, \dots, 1996\}$ be the year of the parameter change. The corresponding likelihood function arising from the ten observations $N_{1987}, \dots, N_{1996}$ is

$$L(y_c) = \prod_{y=1987}^{y_c-1} \text{Poisson}_{\lambda_0(y_c)}(N_y) \times \prod_{y=y_c}^{1996} \text{Poisson}_{\lambda_1(y_c)}(N_y), \tag{9.1}$$

where

$$\lambda_0(y_c) = \frac{1}{y_c - 1987} \sum_{y=1987}^{y_c-1} N_y \qquad \text{and} \qquad \lambda_1(y_c) = \frac{1}{1997 - y_c} \sum_{y=y_c}^{1996} N_y$$

are the (maximum-likelihood) Poisson parameter estimators arising from the observations before and after the change-point. For $y_c = 1987$, which means that there is no change-point in the observed data, we omit the first product in (9.1). For the log-likelihood function $l(y_c) = \log L(y_c)$, we get

y_c	$l(y_c)$	y_c	$l(y_c)$	y_c	$l(y_c)$
1988	−16.23	1991	−14.69	1994	−16.94
1989	−14.23	1992	−15.52	1995	−17.99
1990	−11.96	1993	−17.45	1996	−17.84

and $l(1987) = -18.02$, if there is no change-point. Clearly, the choice $y_c = 1990$ by far maximises the log-likelihood, leading to $\lambda_0(1990) = 0$ and $\lambda_1(1990) = {}^{17}\!/_{7} \approx 2.43$, see Figure 9.1. Of course, the estimate $\lambda_0(1990) = 0$ cannot be the true value and this might be considered a model deficiency. However, only the Poisson parameter $\lambda_1(1990)$ for the observations after the jump is of interest for the extrapolation and the estimate of the knock-out probability P_{CAT}. This also means that all observations before the change-point are ignored for the extrapolation. Using $\hat{p}_{6000} \approx 0.0757$ from (4.32), we obtain similarly to (4.7) and (4.8) that

$$P_{\text{CAT}} = 1 - \exp\bigl(-\hat{p}_{6000}\, \lambda_1(1990)\bigr) \approx 0.1680.$$

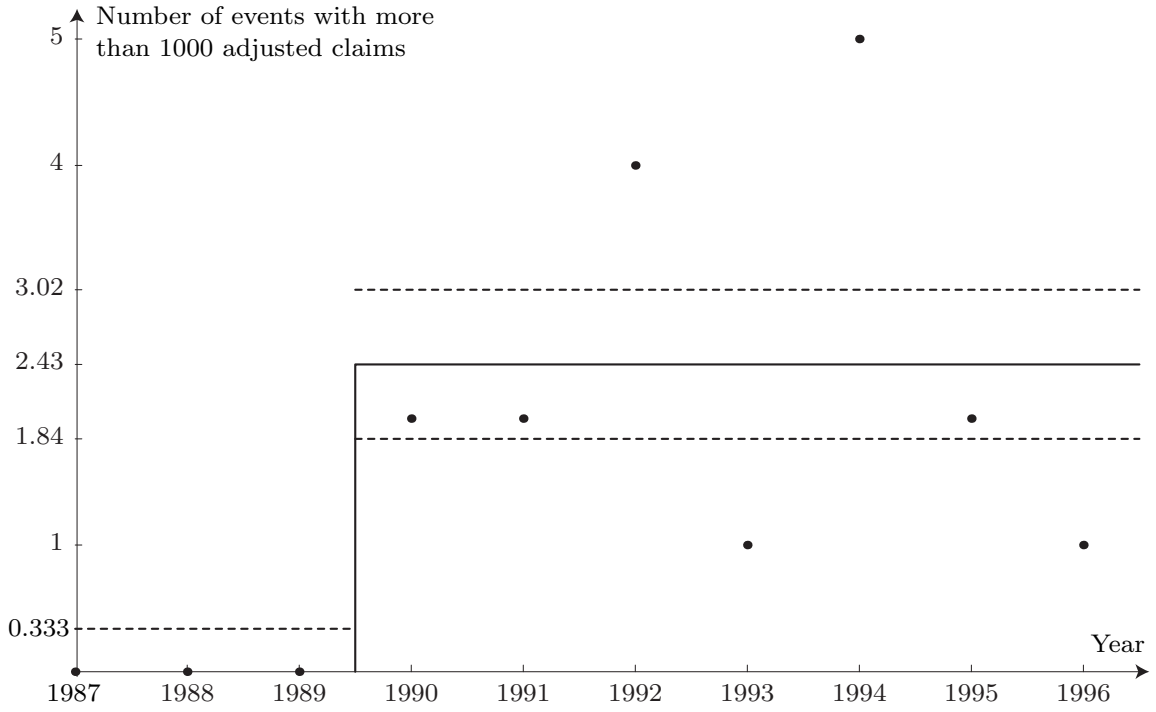

FIGURE 9.1. A model with a change-point in the Poisson parameter. A change from $\lambda_0 = 0$ to $\lambda_1 = {}^{17}\!/_{7} \approx 2.43$ between 1989 and 1990 maximises the likelihood. The dashed lines indicate the estimated standard deviations of the Poisson parameters disregarding the uncertainty arising from the estimate of the location of the change-point.

A recalculation of Table 3.2 gives a discounted value of CHF 253.80 for the WINCAT coupons. Using an analogue of (4.28) leads to $P_{\text{CAT}} \approx 0.1689$ and CHF 253.56.

For the standard deviation of $\lambda_1(1990)$ indicated by the dashed lines in Figure 9.1, we take an estimate similar to (4.4), namely $\hat{\sigma}(\lambda_1(1990)) = \sqrt{\lambda_1(1990)/7} = \sqrt{17}/7 \approx 0.59$. A similar estimate for $\lambda_0(1990)$ does not lead to a meaningful result, because $\lambda_0(1990) = 0$. The log-likelihood ratio statistic does not seem to be suitable here, because the application of asymptotic results to just three observations is questionable. Looking for the largest Poisson parameter λ such that the likelihood for the joint occurrence of the three independent events $N_{1987} = 0$, $N_{1988} = 0$ and $N_{1989} = 0$ is 68%, we get $\exp(-3\hat{\sigma}(\lambda_0(1990))) = 0.68$, which gives $\hat{\sigma}(\lambda_0(1990)) \approx 0.13$. However, the smallest possible observation for $\lambda_0(1990)$ besides zero is ${}^{1}\!/_{3}$, hence we should take at least $\hat{\sigma}(\lambda_0(1990)) = {}^{1}\!/_{3}$. This is also the smallest value besides zero, which we can get by the formula $\sqrt{\lambda_0(1990)/3}$ similarly to (4.4).

Note that the two estimates $\hat{\sigma}(\lambda_0(1990))$ and $\hat{\sigma}(\lambda_1(1990))$ do not incorporate any uncertainty about the location of the change-point. While asymptotic results for confidence regions of change-point estimates based on likelihood ratio tests are available, see e. g. [9] and [22], it is problematical to apply these to the present short sequence with a priori unknown λ_0 and λ_1, in particular since the log-likelihood function gives such a clear-cut answer for the location of the change-point here.

10. Peaks-over-threshold method

The composite Poisson models discussed in Sections 4, 5 and 7–9 make an a priori distinction between the event frequency and the event severity, which is the adjusted number of claims arising from these events. This distinction allows to choose from a selection of constant-parameter and time-dependent parameter models for the event frequency and, independently, to choose a distribution for the event severity: the Bernoulli distribution, the Pareto distribution and the generalised Pareto distribution have been discussed in Section 4.

So far we fitted several Poisson models with a time-dependent parameter to the observed event frequencies. It is an obvious idea to consider also a time-dependent distribution for the event severity; most of the arguments given at the beginning of Section 7 can be used to support this idea. However, there is also the aim of parameter parsimony and the danger of overfitting. With the present small historical data set, this danger probably becomes real when we try to estimate trend parameters for the event frequency and the event severity separately. If we tear down the wall between event frequency and severity, then there is a three-parameter model available, which incorporates the for theoretical reasons desirable generalised Pareto distribution for the event severity and the Poisson distribution for the event frequency. The Poisson distribution is also backed up by extreme value theory, because the point process of exceedances over thresholds, in an appropriate set-up, converges weakly to a time-homogeneous Poisson point process as the threshold increases and the time is rescaled accordingly to keep the expected number of exceedances constant, see [6, Theorem 5.3.2]. By allowing a trend in one of the three parameters, which requires a fourth parameter, we can model a possible time-inhomogeneity in both distributions. Note that such a joint model for event frequency and severity might better account for the two events in 1994 which just pierced the threshold of 1 000 adjusted claims.

The original continuous-time model is called *peaks-over-threshold model* and we shall give a brief outline adapted to the present problem below. For a more detailed discussion, see [6, Chapter 6.5] for example. This peaks-over-threshold model is also used in a case study of wind storm losses encountered by a Swedish insurance group, see H. Rootzén and N. Tajvidi [13]; their fifth section is devoted to trend detection.

Consider a Poisson point process on the set $\{1987,\dots,1996\}\times\mathbb{R}$ with an intensity measure $\Lambda_{\mu,\xi,\tau}$, which is uniquely determined by

$$\Lambda_{\mu,\xi,\tau}(\{y\}\times(x,\infty)) = \Big(1+\xi\frac{x-\mu}{\tau}\Big)_+^{-1/\xi}, \quad x\in\mathbb{R},\ y\in\{1987,\dots,1996\}; \quad (10.1)$$

where τ and ξ are strictly positive model parameters and $\mu\in\mathbb{R}$ denotes a location parameter. If $\xi(x-\mu)\le-\tau$, then the right-hand side of (10.1) has to be interpreted as infinity. The value

$$\Lambda_{\mu,\xi,\tau}(\{y\}\times(a,\infty)) = \Big(1+\xi\frac{a-\mu}{\tau}\Big)_+^{-1/\xi} \quad (10.2)$$

is the Poisson parameter for the number of events exceeding the threshold a in the observation period ending in year y. All events for which the number of adjusted claims exceeds the threshold of $a = 1\,000$ are recorded in Table 1.1. Of course, the model parameters should be chosen such that the Poisson parameter in (10.2) is finite, which is equivalent to $\xi(a-\mu) > -\tau$. For $x \ge a$, the ratio of the Poisson

parameters from (10.1) and (10.2) is

$$\frac{\Lambda_{\mu,\xi,\tau}(\{y\}\times(x,\infty))}{\Lambda_{\mu,\xi,\tau}(\{y\}\times(a,\infty))} = \Big(1+\xi\frac{x-a}{\tau_a}\Big)^{-1/\xi} \quad \text{with} \quad \tau_a = \tau+\xi(a-\mu). \tag{10.3}$$

This ratio is the conditional probability for a Poisson point to lie in the subset $\{y\}\times(x,\infty)$ given that it lies in $\{y\}\times(a,\infty)$. It is also the expected number of points in $\{y\}\times(x,\infty)$ divided by the expected number of points in $\{y\}\times(a,\infty)$. As the right-hand side of (10.3) shows, the conditional probability is the tail $1-G_{a,\xi,\tau_a}(x)$ of a shifted generalised Pareto distribution of the form (4.29).

The above calculation shows that the points-over-threshold model is structurally stable with respect to an increase in the threshold a; according to (10.2) and (10.3) we just get different Poisson and scale parameters. This stability can be used to determine an appropriate threshold by exploratory data analysis; due to our small historical data set, we do not attempt such an analysis here. Note that we have chosen the set $\{1987,\dots,1996\}$ instead of a ten-year interval because we want to refrain from modelling a seasonal dependence of the storm and hail damages.

As an abbreviation for the Poisson parameter from (10.2), let $V(x;\mu,\xi,\tau)$ denote the right-hand side of (10.1) for all $\xi,\tau>0$ and $\mu,x\in\mathbb{R}$ satisfying $\xi(x-\mu)>-\tau$, and let $v(x;\mu,\xi,\tau)=-(\partial/\partial x)V(x;\mu,\xi,\tau)$. The function

$$[a,\infty)\ni x\mapsto \frac{v(x;\mu,\xi,\tau)}{V(a;\mu,\xi,\tau)} = \frac{1}{\tau_a}\Big(1+\xi\frac{x-a}{\tau_a}\Big)^{-1-1/\xi}$$

is the $(-\partial/\partial x)$-derivative of (10.3) and therefore the density of the shifted generalised Pareto distribution G_{a,ξ,τ_a} on $[a,\infty)$, see (4.29).

To take a time-inhomogeneity into account, we first consider a linear dependence of the location parameter given by

$$\mu_{\alpha,\beta}(y)=\alpha+\beta(y-1991.5), \qquad \alpha,\beta,y\in\mathbb{R}. \tag{10.4}$$

The location parameter has to satisfy $\xi(a-\mu_{\alpha,\beta}(y))>-\tau$ for all years y under consideration. In this way we get a family of Poisson parameters and shifted generalised Pareto distributions via (10.2) and (10.3), respectively. Motivated by the simulation study in [13, Section 3], we use only the maximum-likelihood method for estimating the parameters and do neither consider the method of moments nor probability weighted moments. The maximum-likelihood method is also more flexible with respect to model extensions.

The likelihood function is given by

$$L(\alpha,\beta,\xi,\tau)=\prod_{y=1987}^{1996}\Bigg(\mathrm{Poisson}_{V(a;\mu_{\alpha,\beta}(y),\xi,\tau)}(N_y)\prod_{i=1}^{N_y}\frac{v(X_{i,y};\mu_{\alpha,\beta}(y),\xi,\tau)}{V(a;\mu_{\alpha,\beta}(y),\xi,\tau)}\Bigg) \tag{10.5}$$

with the Poisson distribution as in (4.1), where N_{1987}, …, N_{1996} are the numbers of observed events exceeding the threshold $a=1000$ in the years 1987, …, 1996 and $X_{1,y}$, …, $X_{N_y,y}$ are the numbers of adjusted claims of the events in the year y given by Table 1.1. Using the abbreviation $c=-\sum_{y=1987}^{1996}\log N_y!$, we can write the log-likelihood function corresponding to (10.5) as

$$l(\alpha,\beta,\xi,\tau)=c-\sum_{y=1987}^{1996}V(a;\mu_{\alpha,\beta}(y),\xi,\tau)+\sum_{i=1}^{17}\log v(X_i;\mu_{\alpha,\beta}(y_i),\xi,\tau), \tag{10.6}$$

	Linear dependence of the location parameter				Log-linear dependence of the scale and location parameters			
Year	$\hat{\lambda}_{1000}$	$\hat{\lambda}_{6000}$	$\hat{p}_{6000}$	P_{CAT}	$\hat{\lambda}_{1000}$	$\hat{\lambda}_{6000}$	$\hat{p}_{6000}$	P_{CAT}
1987	0.957	0.110	0.115	10.4%	0.616	0.036	0.058	3.5%
1988	1.049	0.113	0.107	10.6%	0.756	0.045	0.059	4.4%
1989	1.157	0.115	0.099	10.9%	0.925	0.056	0.061	5.5%
1990	1.285	0.118	0.092	11.1%	1.128	0.070	0.062	6.8%
1991	1.437	0.120	0.084	11.3%	1.369	0.088	0.064	8.4%
1992	1.623	0.123	0.076	11.6%	1.655	0.110	0.066	10.4%
1993	1.851	0.126	0.068	11.9%	1.990	0.137	0.069	12.8%
1994	2.139	0.129	0.060	12.1%	2.381	0.171	0.071	15.7%
1995	2.508	0.132	0.053	12.4%	2.831	0.213	0.075	19.2%
1996	2.995	0.136	0.045	12.7%	3.347	0.265	0.079	23.3%
1997	3.663	0.139	0.038	13.0%	3.930	0.328	0.084	28.0%
1998	4.618	0.143	0.031	13.3%	4.583	0.406	0.089	33.4%
1999	6.067	0.147	0.024	13.6%	5.307	0.501	0.094	39.4%

TABLE 10.1. Results for the peaks-over-threshold model with a linear dependence $\mu_{\alpha,\beta}(y) = \alpha + \beta(y - 1991.5)$ of the location parameter or a log-linear dependence (10.7) of the scale and the location parameters. For both models the fitted Poisson parameters $\hat{\lambda}_{1000}$ and $\hat{\lambda}_{6000}$ for the number of events with more than 1 000 and 6 000 claims, respectively, are listed. In addition, the conditional probability $\hat{p}_{6000} = \hat{\lambda}_{6000}/\hat{\lambda}_{1000}$ that an event causes more than 6 000 claims and the estimated knock-out probability P_{CAT} are given for the years 1987–1999.

where $X_1, \ldots, X_{17}$ are the numbers of adjusted claims of the 17 events recorded in Table 1.1 and $y_1, \ldots, y_{17}$ are the corresponding years these events happened. Incidentally, note that the log-likelihood function in (10.6) would be flexible enough to accommodate a changing threshold for recording historical events; the constant $a = 1000$ simply has to be replaced by a function $y \mapsto a_y$ for this purpose.

If we take $\beta = 0$, then we actually have a three-parameter model consisting of the constant-parameter Poisson model for the event frequencies and the generalised Pareto distribution for the event severity. This model coincides with the one discussed in Subsection 4.3. The maximum-likelihood method using (10.6) leads to the estimates $\hat{\mu}_0 = \hat{\alpha}_0 \approx 1427.5$, $\hat{\xi}_0 \approx 0.7243$ and $\hat{\tau}_0 \approx 970.3$, which give via (10.2) and (10.3) the Poisson parameter $\Lambda_{\hat{\mu}_0,\hat{\xi}_0,\hat{\tau}_0}(\{y\} \times (1000, \infty)) \approx 1.7$ for every year y and the scale parameter $\hat{\tau}_{a,0} \approx 660.7$, where the index zero refers to $\beta = 0$. These estimated values coincide with the ones in (4.3) and (4.31), which is reassuring. On the other hand, we can use this fact to determine the maximum-likelihood estimates arising from (10.6) with $\beta = 0$ in an easier way. The scale parameter $\hat{\tau}_{a,0}$ and the shape parameter $\hat{\xi}_0$ for the generalised Pareto distribution can be calculated as in Subsection 4.3; the Poisson parameter $\lambda_{1000}^{\text{const}}$ is given by (4.2). Solving (10.2) and (10.3) leads to

$$\hat{\tau}_0 = \hat{\tau}_{a,0}\big(\lambda_{1000}^{\text{const}}\big)^{\hat{\xi}_0} \qquad \text{and} \qquad \hat{\mu}_0 = a + (\hat{\tau}_0 - \hat{\tau}_{a,0})/\hat{\xi}_0.$$

In this way we can determine starting values for the numerical iteration procedure in the time-inhomogeneous cases discussed below.

Maximising the log-likelihood function (10.6) without the restriction $\beta = 0$, we obtain the estimates $\hat{\alpha} \approx 1376.9$, $\hat{\beta} \approx 93.99$, $\hat{\xi} \approx 0.6534$ and $\hat{\tau} \approx 1021.3$. The positive value of $\hat{\beta}$ leads to an increasing value of the location parameter, which leads to an increasing Poisson parameter via (10.2) and to a decreasing scale parameter via (10.3). This means that the event frequency increases, but the event severity decreases. The expected number of events above the thresholds 1 000 and 6 000 for the years under consideration, calculated according to (10.2), are given in the second and third column of Table 10.1. A recalculation of Table 7.1 using the estimated parameters from the last three lines of Table 10.1 leads to a discounted value of CHF 264.00 for the three WINCAT coupons.

Note that $\hat{\beta}$ is so large that the restriction $\hat{\xi}(a - \mu_{\hat{\alpha},\hat{\beta}}(y)) > -\hat{\tau}$ is violated for the years $y \geq 2005$, which shows again that extrapolation has to be done with great care. We should test whether the trend parameter β is significantly different from zero. Under the null hypothesis $\beta = 0$, every redistribution of the 17 observations to the ten observation periods has the same probability; the log-likelihood function in (10.6) would not depend on this redistribution. In principle, we could use this observation for a permutation test. In practice, it is numerically demanding to determine the four maximum-likelihood estimators for several thousand random redistributions (there are 10^{17} redistributions in total, hence it is impossible to use all of them). We therefore resort to asymptotic results and use the log-likelihood ratio statistic. Assuming the model with $\beta = 0$ and the above estimates $\hat{\alpha}_0$, $\hat{\xi}_0$, and $\hat{\tau}_0$ to be the correct one, the deviance $2l(\hat{\alpha}, \hat{\beta}, \hat{\xi}, \hat{\tau}) - 2l(\hat{\alpha}_0, 0, \hat{\xi}_0, \hat{\tau}_0)$ has approximately a χ^2-distribution with one degree of freedom [15, Section 5.2.3]. With the above maximum-likelihood estimates $\hat{\alpha} \approx 1376.9$, $\hat{\beta} \approx 93.99$, $\hat{\xi} \approx 0.6534$ and $\hat{\tau} \approx 1021.3$, we obtain $2l(\hat{\alpha}, \hat{\beta}, \hat{\xi}, \hat{\tau}) - 2l(\hat{\alpha}_0, 0, \hat{\xi}_0, \hat{\tau}_0) \approx 3.166$, which corresponds to the 7.52%-quantil of the χ^2_1-distribution.

Another possibility to allow for a trend in the model is to assume that the location parameter μ and the strictly positive scale parameter τ depend in the same way log-linearly on the year, meaning that

$$\begin{aligned} \mu_{\alpha,\beta}(y) &= \exp(\alpha + \beta(y - 1991.5)), \\ \tau_{\beta,\gamma}(y) &= \exp(\gamma + \beta(y - 1991.5)), \end{aligned} \qquad \alpha, \beta, \gamma, y \in \mathbb{R}. \tag{10.7}$$

Such a model extension is appropriate when the event severity is measured in currency units subject to a yearly inflation rate of $e^{\beta} - 1$, for example. The parameters have to satisfy $\xi(a - \mu_{\alpha,\beta}(y)) > -\tau_{\beta,\gamma}(y)$ for all years y under consideration. Defining the corresponding log-likelihood function similarly to (10.6) by

$$\tilde{l}(\alpha, \beta, \gamma, \xi) = c - \sum_{y=1987}^{1996} V(a; \mu_{\alpha,\beta}(y), \xi, \tau_{\beta,\gamma}(y)) + \sum_{i=1}^{17} \log v(X_i; \mu_{\alpha,\beta}(y_i), \xi, \tau_{\beta,\gamma}(y_i)), \tag{10.8}$$

and maximising it numerically, we obtain the estimates $\hat{\alpha} \approx 7.199$, $\hat{\beta} \approx 0.1379$, $\hat{\gamma} \approx 6.835$ and $\hat{\xi} \approx 0.5972$. Since $\hat{\xi} \exp(\hat{\alpha}) \approx 799.1 < 929.8 \approx \exp(\hat{\gamma})$, the above inequality $\hat{\xi}(a - \mu_{\hat{\alpha},\hat{\beta}}(y)) > -\tau_{\hat{\beta},\hat{\gamma}}(y)$ is satisfied for all $y \in \mathbb{R}$. The results are given in Table 10.1, the discounted value of the three WINCAT coupons is CHF 204.34. With the values estimated above, the Poisson parameters $\hat{\lambda}_{1000}$ and $\hat{\lambda}_{6000}$ for the frequency of events with more than 1 000 or 6 000 claims, respectively, increase with time. This is illustrated by the sixth and seventh column of Table 10.1. Note that the two extensions of the peaks-over-threshold model come to opposite conclusions

concerning the trend of the event severity. The fourth column of Table 10.1 exhibits a decreasing trend of the probability $\hat{p}_{6000}$, that an event causes at least 6 000 claims, while the eighth column shows an increasing trend of $\hat{p}_{6000}$. This partial cancellation or superposition of trends leads to a difference of CHF 59.66 between the two estimated values of the three WINCAT coupons.

To test the null hypotheses $\beta = 0$ in the model extension specified by (10.7), we assume that this model with $\beta = 0$ and the above estimates $\hat{\mu}_0 \approx 1427.5$, $\hat{\xi}_0 \approx 0.7243$ and $\hat{\tau}_0 \approx 970.3$ is the correct one (which via (10.7) gives $\hat{\alpha}'_0 = \log \hat{\mu}_0 \approx 7.264$ and $\hat{\gamma}_0 = \log \hat{\tau}_0 \approx 6.88$ here) and determine the value of the log-likelihood ratio statistic. We get $2\tilde{l}(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\xi}) - 2\tilde{l}(\hat{\alpha}'_0, 0, \hat{\gamma}_0, \hat{\xi}_0) \approx 4.088$, which corresponds to the 4.32%-quantil of the χ^2_1-distribution.

Various other extensions of the peaks-over-threshold model for incorporating a trend are possible. We could consider a time-dependent shape parameter ξ, for example. Furthermore, instead of a linear or log-linear dependency as in (10.4) or (10.7), we could consider a greater selection of possible dependencies as we have done for the Poisson parameter in Sections 7 and 9. If the historical data set were bigger, also other suitable selections of two or even all three parameters μ, ξ, and τ could be made time-dependent. We refrain from justifying, discussing, fitting and testing such models here, but we hope that the worked-out cases in this section can serve as a guideline when the need for one of the above-mentioned extensions arises.

11. Comparison of the estimated values

Table 11.1 contains the estimated discounted values of all three WINCAT coupons for the models considered in the previous sections. We have added several additional models, which arise by combining the Bernoulli or Pareto distribution for the event severity with the various models for the event frequency. The entries of the table are grouped according to the used model for the event frequency. There is an additional partial order according to the value of the coupons.

The following remarks should be kept in mind when comparing the "conservative" value of CHF 229.78 obtained in [3] with the results of this paper.

- No explicit risk premium is included in the discounted values of the WINCAT coupons given in Table 11.1.
- The extrapolated estimated Poisson parameters are in the region from 3.78 in Figure 7.1 up to 5.68 in Figure 7.3. In the historical data set, two observation periods with four and five events are recorded, hence the extrapolated parameters are not unreasonable if one accepts the possibility of a trend.
- Time homogeneity is possible with all the models considered in this paper by choosing either $\beta = 0$ or no change-point. It is the historical data set that leads to the positive estimates for β or the clear-cut location of the change-point, respectively.

A higher event frequency magnifies the differences between the estimate $\hat{p}_{6000} \approx 0.0857$ from (4.24) obtained by the Pareto fit of the adjusted claim numbers, the estimate $\hat{p}_{6000} \approx 0.0757$ from (4.32) obtained by the generalised-Pareto fit, and the estimate $\hat{p}_{6000} = 2/17 \approx 0.118$ from the Bernoulli distribution of Subsection 4.1. As Table 11.1 shows, these differences between the empirical and the fitted distribution functions in Figures 4.2 and 4.3 give rise to quite different values of the WINCAT coupons. The specific form of a possible trend is of minor importance.

No.	Coupon value	Corresponding model
1	CHF 244.44	Binomial model of Section 3
		Constant-parameter Poisson model of Section 4 and
2	CHF 267.48	— *generalised Pareto distribution of Subsection 4.3*
3	CHF 263.29	— *Pareto distribution of Subsection 4.2*
4	CHF 247.37	— *Bernoulli distribution of Subsection 4.1*
5	CHF 266.62	Generalised extreme value distribution of Section 6
		Peaks-over-threshold model of Section 10 and
6	CHF 264.00	— *linear trend of the location parameter*
7	CHF 204.34	— *log-linear trend of the scale and location parameter*
		Change-point model of Section 9 and
8	CHF 253.80	— *generalised Pareto distribution of Subsection 4.3*
9	CHF 247.99	— *Pareto distribution of Subsection 4.2*
10	CHF 225.28	— *Bernoulli distribution of Subsection 4.1*
		Generalised Pareto distribution (4.29) and
		time-dependent Poisson parameter with the
11	CHF 223.88	— *linear trend of Subsection 7.1*
12	CHF 220.53	— *smooth transition of Subsection 7.5*
13	CHF 214.44	— *modified-linear trend of Subsection 7.4*
14	CHF 214.37	— *log-linear trend of Subsection 7.2*
15	CHF 210.86	— *square-root linear trend of Subsection 7.3*
		Pareto distribution (4.15) and a
		time-dependent Poisson parameter with the
16	CHF 215.19	— *linear trend of Subsection 7.1*
17	CHF 211.54	— *smooth transition of Subsection 7.5*
18	CHF 204.96	— *modified-linear trend of Subsection 7.4*
19	CHF 204.93	— *log-linear trend of Subsection 7.2*
20	CHF 201.12	— *square-root linear trend of Subsection 7.3*
		Bernoulli distribution and a
		time-dependent Poisson parameter with the
21	CHF 189.56	— *linear trend of Subsection 7.1*
22	CHF 185.11	— *smooth transition of Subsection 7.5*
23	CHF 177.36	— *modified-linear trend of Subsection 7.4*
24	CHF 177.44	— *log-linear trend of Subsection 7.2*
25	CHF 172.87	— *square-root linear trend of Subsection 7.3*

TABLE 11.1. Comparison of the value of the three discounted WINCAT coupons arising from the binomial model, the composite Poisson models, the generalised extreme value model and the two extensions of the peaks-over-threshold model. No explicit risk premium is included. The various estimated coupon values clearly illustrate the model risk.

In the linear-trend model of Subsection 7.1, the slope β is restricted by the positivity requirement of the Poisson parameter. Similarly, in the smooth-transition model of Subsection 7.5, there is the imposed upper bound for the Poisson parameter. Both restrictions lead to higher estimated coupon values.

Model No.	Coupon value in CHF	Value change in scenario with no event	Value change in scenario with 4 events	Coupon value distribution (in CHF) arising from simulated model data: Mean	Median	St. dev.	15.9%-quantil
1	244.44	5.33	−21.35	244.74	244.44	37.45	215.07
2	267.48	3.04	−8.90	272.26	274.20	22.51	249.68
3	263.29	3.36	−7.83	258.46	260.43	22.95	234.81
4	247.37	4.81	−17.21	249.41	247.37	33.69	223.60
5	266.62	3.20	−12.51	271.18	273.96	23.83	246.69
6	264.00	4.78	−9.98	266.55	268.70	26.52	241.89
7	204.34	43.24	−21.19	199.05	206.37	70.18	121.53
8	253.80	5.66	−9.88	259.95	262.69	31.04	226.60
9	247.99	6.27	−8.29	240.95	243.39	31.32	208.54
10	225.28	9.09	−19.07	230.90	225.28	46.00	194.34
11	223.88	7.36	−16.21	235.17	236.03	47.37	185.22
12	220.53	28.89	−10.14	232.08	231.12	43.26	185.78
13	214.44	24.41	−16.51	225.32	230.68	54.55	165.14
14	214.37	32.83	−24.93	223.21	233.39	62.27	154.02
15	210.86	28.59	−18.60	223.55	227.69	56.13	159.11
16	215.19	8.02	−13.76	208.39	210.39	41.98	162.12
17	211.54	31.61	−7.26	207.96	209.10	39.16	166.62
18	204.96	26.56	−13.71	195.08	198.98	49.90	140.85
19	204.93	35.79	−22.48	196.78	205.29	59.55	132.81
20	201.12	31.07	−15.76	189.15	193.58	51.33	134.90
21	189.56	9.74	−27.72	196.58	189.56	60.75	150.18
22	185.11	38.98	−19.84	195.82	195.63	55.54	136.24
23	177.36	32.19	−27.85	185.41	182.04	65.20	115.16
24	177.44	43.60	−38.05	187.14	189.51	73.88	106.55
25	172.87	37.51	−30.28	183.51	182.14	70.87	108.60

TABLE 12.1. Assume that a similar bond with three WINCAT coupons is issued one year later and assume that the interest rate structure is the same. If there is no event in 1997 (favourable scenario), then the models listed in Table 11.1 give higher values for the corresponding three WINCAT coupons. If there are four events in 1997 (stress scenario) with adjusted claims as in 1992, see Table 1.1, then the models lead to lower values. The above changes in value show the robustness/sensitivity of the models with respect to new data. The last four columns characterise the distributions of the coupon values based on 1 000 simulated data sets for every fitted model, see Section 13.

12. Model robustness and sensitivity analysis

When choosing a model, in particular for low frequency event risks, it is of interest to know how sensitive the model reacts to changes of the data. We refrain from manipulating the available historical data of Table 1.1 for this purpose. Instead, we employ a scenario technique by adding fictitious data for 1997 to the historical data set of Table 1.1. For a favourable scenario, we assume the best possible case,

namely that no event with more than 1 000 claims is recorded in 1997. Such an event history happened three times already during the recorded 10-year history. For a stress scenario, we want to add a bad event record for 1997. To remain realistic, we prefer to pick a bad year from the available historical data set. While the year 1994 is certainly the worst case with respect to the event frequency, it would not lead to a knock-out of the coupon and therefore counts as a favourable year for the binomial model of Section 3. Hence we choose the data of the year 1992 as a common stress scenario for all models listed in Table 11.1.

For an easy comparison of the previous results with the coupon values arising from these scenarios, we assume that an identical three-year bond is issued in February 1998, that the observation period for the first coupon is shorter for applying the $^{15}/_{17}$-correction from (2.1), and that the interest-rate structure for the coupons is again given by Table 3.1. Based on the extended 11-year data set, we estimate the discounted value of the corresponding three WINCAT coupons using all models discussed so far. The model-dependent changes of the value are given in the third and forth column of Table 12.1.

The robustness and sensitivity of the models should not be judged exclusively on the numbers in Table 12.1, because the two scenarios do not illustrate every possible behaviour of the models. In the change-point model of Section 9, the change-point illustrated by Figure 9.1 remains between 1989 and 1990 in both scenarios. Similarly, in the linear-trend model of Subsection 7.1, there is no solution of the likelihood equations (7.3) in either scenario, hence we use the pragmatic approach again and set $\alpha = 0$. Three trend models, namely the root-linear model of Subsection 7.3, the modified-linear model of Subsection 7.4, and the smooth-transition model of Subsection 7.5 lead to a slightly smaller estimate for the trend parameter β even in the stress scenario. These are exactly those trend models of Section 7 that predicted a Poisson parameter larger than four for the year 1997. For the favourable scenario, the biggest relative change shows up in the smooth-transition model of Subsection 7.5; the trend parameter β drops from 0.442 to 0.114. Due to our choice of the stress scenario, the imposed upper bound in the smooth-transition model remained at 5.

The trend models of Section 7 combined with the generalised Pareto distribution of Subsection 4.3 have four parameters, similarly as the two extensions of the peaks-over-threshold model of Section 10. However, as Table 12.1 for models no. 6 and 7 shows, an extension of the peaks-over-threshold model may be more robust, because the increasing trend in the event frequency is partially compensated by the decreasing trend in the event severity (see the left part of Table 10.1 for model no. 6), or the extension can be more sensitive if both trends go in the same direction (see the right part of Table 10.1 for model no. 7).

13. Model consistency and dispersion analysis

For every fitted model listed in Table 11.1—under the assumption that the model describes reality correctly—we generate 1 000 new random data sets according to the distribution specified by the fitted model. These data sets replace the actual observations recorded in Table 1.1, and we use the model to estimate the discounted coupon values based on the random data set. This gives a list of 1 000 (not necessarily different) values for the coupons; for some selected models the corresponding

Model variant no.	Coupon value in CHF	Values from 1 000 simulated data sets			
		Mean	Median	St. dev.	15.9%-quantil
$8'$	253.80	256.91	259.99	34.00	223.36
$9'$	247.99	238.05	240.69	31.67	206.93
$10'$	225.28	229.77	225.28	47.84	179.22
$12'$	220.53	226.20	229.06	49.56	175.94
$17'$	211.54	196.89	199.26	45.22	147.65
$22'$	185.11	185.85	186.34	58.50	126.56

TABLE 13.1. Mean, median, standard deviation and 15.9%-quantil in CHF of the simulated distribution for variants of the models 8–10 (variable location of the change-point of the Poisson parameter) and for the variants of models 12, 17, and 22 (variable upper bound for the smooth transition of the Poisson parameter). Compare the results with the corresponding entries in Table 12.1.

histograms are shown in Figures 13.1 and 13.2. Using these simulated distributions, we can calculate the mean value of the estimated coupon values, the median, and the standard deviation; these results are listed in Table 12.1. For models no. 1, 4, 10, and 21, the exact distribution could have been used. For some model variants, the results are in Table 13.1. For comparison with conservative estimates calculated in previous sections, we also give the 15.9%-quantiles of these simulated distributions. This 15.9%-quantil corresponds to one standard deviation in the case of a normal distribution. If the standard deviation or the desired quantil cannot be derived by analytical means, these simulated values can give an indication for a conservative estimate *within the used model;* see the end of the introduction for a further discussion of this topic.

We want to explain the model variants first. For the models no. 8–10, the change-point of the Poisson parameter was kept between 1989 and 1990. Therefore, these models correspond to the constant-parameter Poisson models no. 2–4 with a shorter historical data set of seven years. Table 13.1 contains the simulation results for the model variants $8'$–$10'$, where the change-point is taken to be the earliest year such that the likelihood function (9.1) corresponding to the simulated data set is maximised. Since $\lambda_0(y) = 0$ for $y \in \{1987, 1988, 1989\}$, the change-point y_c can only move to later years; it has done so for 133 of the 1 000 simulated data sets.

To obtain the simulation results for the models with a smooth transition of the Poisson parameter (models no. 12, 17, and 22), the imposed upper bound for the Poisson parameter was kept at 5. However, in Subsection 7.5, we took the largest observation as the upper bound. Doing the same for the simulated data sets leads to the model variants $12'$, $17'$, and $22'$.

Using the same model for generating the data and for fitting this simulated data creates a bias towards the model, hence the model should exhibit a good performance. Let us comment on some peculiarities of the models.

The binomial model no. 1 can lead to at most 11 different estimated values for the WINCAT coupons; seven of them are shown in the upper left histogram of Figure 13.1. Since we used $\hat{p} = 0.2$ from (3.1), the other four values have a probability of $\sum_{k=7}^{10} \binom{10}{k} \hat{p}^k (1-\hat{p})^{10-k} \approx 0.000864$ together and did not show up in the simulation.

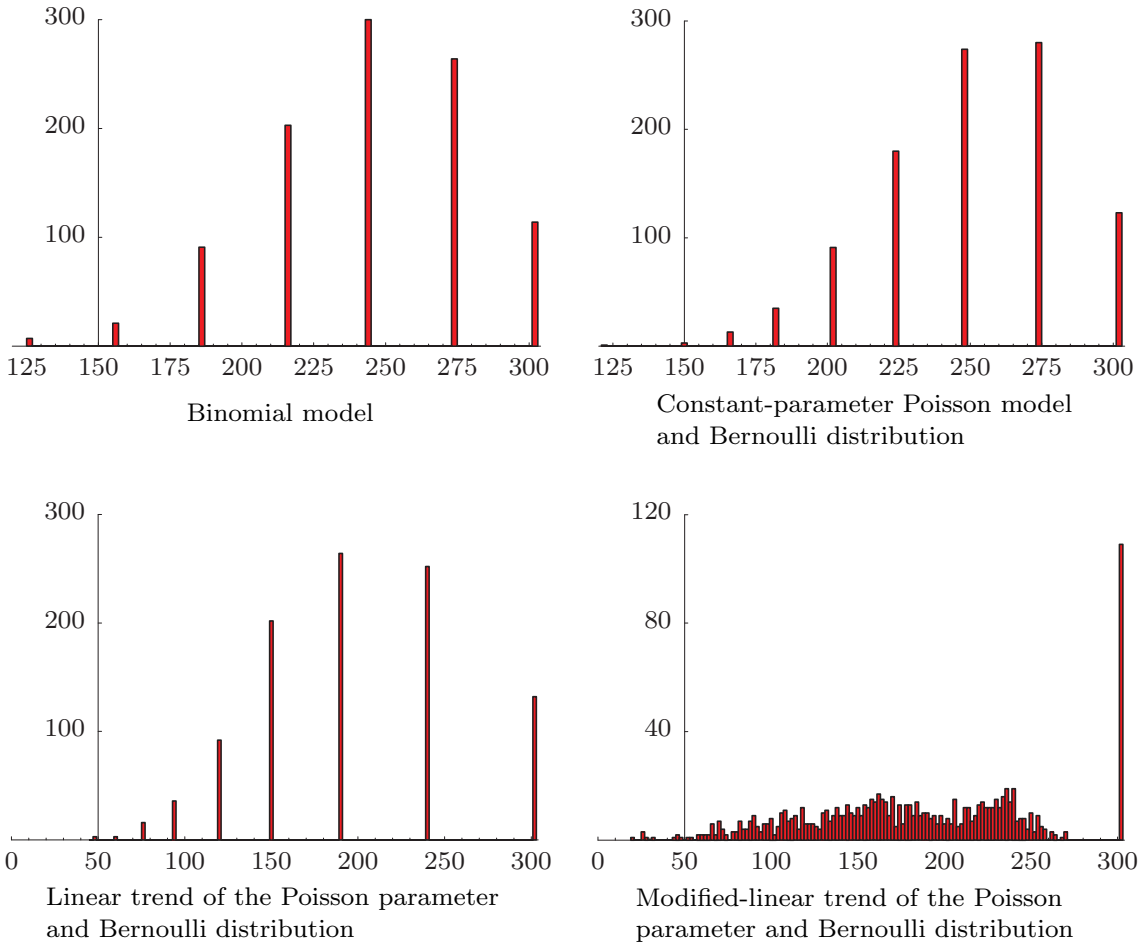

FIGURE 13.1. Distribution of the estimated coupon values for models no. 1, 4, 21, and 23 of Table 11.1 resulting from 1 000 simulated data sets. The mean, median, standard deviation and 15.9%-quantil are given in Table 12.1. Note the different scales, the discreteness of the two upper and the lower left distribution, the different sizes of the gaps, and the persistence of the peak at the highest possible coupon value representing the number of simulated data sets without a knock-out event.

For the constant-parameter Poisson model with a Bernoulli distribution for the event severity (model no. 4), we generate the number $N_{6000,10}$ of knock-out events during the ten years using a Poisson distribution with parameter 2, see Subsection 4.1. The simulated knock-out probability P_{CAT} is calculated from $N_{6000,10}$ using the unbiased estimator (4.12), the resulting discrete distribution is the upper right histogram in Figure 13.1. Similarly for model no. 10: we generate $N_{6000,7}$ from a Poisson distribution with parameter 2 and use (4.12) with $n = 7$. The resulting histogram has bigger gaps than the upper right one in Figure 13.1. When the change-point can move (model no. 10′), then small bars appear between the large ones of the histogram, representing about 13% of the simulated data sets with a later change-point.

The model with a linear trend $\lambda_\beta(y) = \beta(y - 1987)$ of the Poisson parameter and a Bernoulli distribution for the event severity (model no. 21) also results in a discrete distribution, see the lower left histogram in Figure 13.1. Notice the

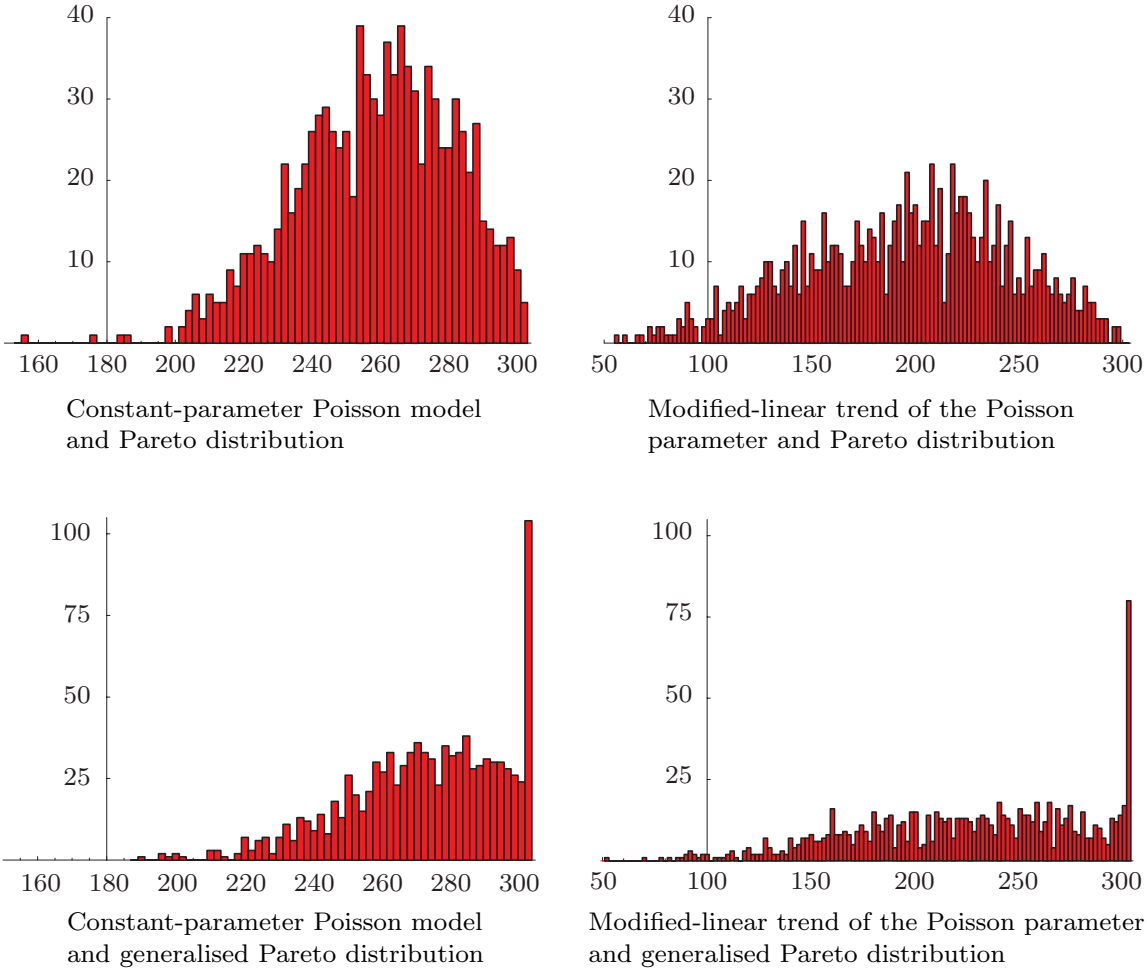

FIGURE 13.2. Distribution of the estimated coupon values for models no. 3, 18, 2, and 13 of Table 11.1 resulting from 1 000 simulated data sets. Note the different scales and the peak at the highest possible coupon value for the two models using a generalised Pareto distribution. The mean, median, standard deviation and 15.9 %-quantil of these distributions are given in Table 12.1.

large gap between the two highest values, it is about twice as large as the gaps in the binomial model. With $\hat{\beta}$ from (7.4), the number $N_{1000,10} = \sum_{1988}^{1996} N_y$ of events from 1987 until 1996 in the linear-trend model has a Poisson distribution with parameter 17. Remember that in this model there can be no events during 1987. With $\hat{p}_{6000} = 2/17$ from (4.5), the number $N_{6000,10}$ of knock-out events in the 10-year period has again a Poisson distribution with parameter 2. Corresponding to (7.5), the knock-out probability is calculated from $N_{6000,10}$ by

$$P_{\text{CAT}}(y) = 1 - \exp(-N_{6000,10}(y - 1987)/45)$$

for $y \in \{1997, 1998, 1999\}$. If desired, an unbiased estimator for $P_{\text{CAT}}(y)$ could be used, see the discussion in Subsection 4.1.

For the discrete distributions of models no. 1, 4, 10, 10′, and 21, the median of the simulated distribution coincides with the estimated value of the coupons; the other possible case of a substantial deviation did not occur in our simulation runs.

The other models with a time-dependent Poisson parameter and a Bernoulli distribution for the event severity (models no. 22–25 and 22′) give a histogram similar to the lower right one in Figure 13.1. The peak in the histogram at the highest possible value of the WINCAT coupons reflects the fact that, depending on the estimated event frequency, about 10% of the simulated data sets do not contain a knock-out event. Table 12.1 shows that the mean of these simulated distributions is about CHF 10 above the estimated values of the WINCAT coupons. This indicates a bias towards higher coupon values. Moving the imposed upper bound of the smooth-transition model according to the data seems to compensate this effect, see the results for model no. 22′ in Table 13.1.

All models using a Pareto distribution for the event severity (models no. 3, 9, 9′, 17′, and 16–20) lead to a histogram similar to the two upper ones in Figure 13.2. Depending on the estimated event frequency for the years 1997–1999 in the used models, the simulated distribution is more or less concentrated. Since the Pareto distribution for the event severity reflects the model assumption of a heavy-tailed distribution, none of the histograms contains a peak at the highest possible value. As Tables 12.1 and 13.1 show, the mean of the simulated distributions for these models is always below the estimated value of the coupons. With the exception of model no. 19, the same observation applies to the median. This indicates a bias towards lower, hence more conservative coupon values.

The constant-parameter composite Poisson model with a Pareto distribution for the event severity (model no. 3) that was used in [3], exhibits one of the smallest standard deviations in this simulation study; only the standard deviation of model no. 2 is slightly smaller, but the difference is within the variation arising from repeated simulation runs. Note that model no. 5, which uses a generalised extreme value distribution, exhibits a standard deviation of similar small size as models no. 2 and 3. The histogram for model no. 5 looks similar to the lower left one in Figure 13.2.

The models no. 12, 17, and 22 use a smooth transition of the Poisson parameter with a fixed upper bound at 5. This clearly restricts the possible extrapolated values of the coupons and leads to the smallest standard deviation of the simulated distributions within the corresponding model groups 11–15, 16–20, and 21–25. If the upper bound is allowed to move to the highest simulated event frequency (models no. 12′, 17′, and 22′ in Table 13.1), then the simulated mean coupon values and the 15.9%-quantiles drop considerably and the standard deviations increase.

When estimating the shape parameter ξ and the scale parameter τ of the generalised Pareto distribution (4.29) for the models no. 2, 8, 8′, 12′, and 11–15, then in about 10% of all simulated data sets there is no pair (ξ, τ) of *strictly positive* numbers maximising the log-likelihood function. An examination of the corresponding data sets reveals that this tends to happen when the simulated data set is smaller than the average one and does not contain events with a large number of claims. For this reason we consider the full family of (shifted) generalised Pareto distributions, which are defined, for all location parameters $a \in \mathbb{R}$ and scale parameters $\tau > 0$, by (4.29) in the case $\xi > 0$, by the shifted exponential distribution

$$G_{a,0,\tau}(x) = \begin{cases} 1 - \exp(-(x-a)/\tau), & \text{if } x \geq a, \\ 0, & \text{if } x < a, \end{cases} \tag{13.1}$$

in the case $\xi = 0$, and by

$$G_{a,\xi,\tau}(x) = \begin{cases} 1 - (1 + \xi(x-a)/\tau)^{-1/\xi}, & \text{if } x \in [a, a - \tau/\xi), \\ 0, & \text{if } x < a, \\ 1, & \text{if } x \geq a - \tau/\xi, \end{cases} \tag{13.2}$$

in the case $\xi < 0$. The shifted exponential distribution $G_{a,0,\tau}$ is the pointwise limit of $G_{a,\xi,\tau}$ as $\xi \to 0$. Note that the distributions $G_{a,\xi,\tau}$ with $\xi < 0$ have a bounded support. For $\xi < -1$, the densities of $G_{a,\xi,\tau}$ are unbounded at the left-hand side of $a-\tau/\xi$. Hence, maximum-likelihood estimators $(\hat{\tau}, \hat{\xi})$ are not well-defined within the full family consisting of (4.29), (13.1), and (13.2), because the likelihood function is unbounded for every sequence $\{(\tau_n, \xi_n)\}_{n\in\mathbb{N}}$ in $(0,\infty) \times (-\infty, -1)$ such that $a - \tau_n/\xi_n$ approaches the largest observation from the right-hand side as $n \to \infty$. Therefore, maximum-likelihood estimators[7] are only meaningful within the range $\tau > 0$ and $\xi \geq -1$. However, in about 2% of all simulated data sets, there is no maximum of the log-likelihood function in the range $\tau > 0$ and $\xi > -1$. An example is the set of claim numbers 1 021, 1 256, 1 420, 1 450, 1 493, 1 839, 2 251, 2 326, 3 109, 3 167, 3 403, 4 857, 4 865, and 4 877, which showed up in one of the simulation runs. A similar problem for obtaining maximum-likelihood estimates for small sample sizes in a simulation study is reported in [8, Section 4].

A possible solution well adopted to the present case is to maximise a grouped-data likelihood function; this method is investigated in [7]. For this method we have to remind ourselves that the spacing of adjusted claim numbers in the last column of Table 1.1 is given by the corresponding vehicles insured index of that table; for simplicity we set this spacing to one in the following considerations. Every simulated value from the generalised Pareto distribution with parameters given by (4.31) then represents an interval of length one. If $X_1, \ldots, X_n$ represent the simulated values, then we use the grouped-data log-likelihood function

$$l(\xi, \tau) = \sum_{k=1}^{n} \log\big(G_{a,\xi,\tau}(\lfloor X_k \rfloor + 1) - G_{a,\xi,\tau}(\lfloor X_k \rfloor)\big), \tag{13.3}$$

where $\xi \in \mathbb{R}$ and $\tau > 0$ have to satisfy $\max\{X_1, \ldots, X_n\} < a - \tau/\xi$ in the case $\xi < 0$. This log-likelihood function avoids the singularity of the density of $G_{a,\xi,\tau}$ for $\xi < -1$. All simulation results of this section for the models using a generalised Pareto distribution rely on the grouped-data log-likelihood function (13.3).

We conclude from these observations, that in about 10% of all simulated data sets the models using a generalised Pareto distribution do not reproduce the original heavy-tailed behaviour because there is no such evidence in the simulated data sets. The data sets with a negative maximum-likelihood estimate for the shape parameter ξ are mainly responsible for the high peaks in the lower two histograms of Figure 13.2: If $\hat{\xi} < 0$, then the support of the distribution is bounded above. If this bound is below 6 000, then the probability of events with more than 6 000 claims is zero, making a knock-out of the coupons impossible and thereby leading to the highest possible value for the coupons. As Tables 12.1 and 13.1 show, the mean and the median of the simulated distributions are well above the coupon value

[7] Remember that the data sets are generated from a generalised Pareto distribution with parameters given by (4.31). According to [16, Section 7], the classical asymptotic theory of maximum-likelihood estimators is applicable, because $\hat{\xi}$ from (4.31) satisfies $\hat{\xi} > -1/2$.

for all the models using a generalised Pareto distribution. This indicates a bias of these models towards higher coupon values.

When we use the grouped-data log-likelihood function (13.3) instead of (4.30) for the original data set of Table 1.1, then the maximum-likelihood estimates are $\hat{\xi} \approx 0.7229$ and $\hat{\tau} \approx 662.1$ instead of $\hat{\xi} \approx 0.7243$ and $\hat{\tau} \approx 660.7$ from (4.31). This leads to $\hat{p}_{6000} \approx 0.07573$ instead of $\hat{p}_{6000} \approx 0.07575$ from (4.32). The change in the coupon value stays below CHF 0.02. Therefore, our use of different likelihood functions does not explain the observed bias.

To perform the simulation study also for the two extensions of the peaks-over-threshold model from Section 10, we extended the definition of the Poisson point-process intensity (10.1) to the cases $\xi = 0$ and $\xi < 0$ as we have done for the generalised Pareto distribution (4.29) by adding (13.1) and (13.2). We modified the log-likelihood functions (10.6) and (10.8) by using grouped-data log-likelihood functions similar to (13.3). We refrain from spelling out all the details here. Note that in the case $\xi < 0$ and the extensions specified by (10.4) and (10.7), the range of the possible number of claims depends on the year. The histograms for these extensions look like the lower left and the lower right ones in Figure 13.2, respectively. The same log-linear trend in the scale and location parameters leads to a more spread-out distribution of the simulated coupon values.

Acknowledgements

It's a great pleasure for me to thank Prof. Dr. F. Delbaen (Mathematical Finance, Department of Mathematics, ETH Zürich), Dr. A. Gisler (Winterthur Insurance) and in particular Dr. A. McNeil (Swiss Re Research Fellow at the ETH Zürich) for many fruitful discussions. In addition, want to express my gratitude to the anonymous referees for their careful reading of the paper, their detailed comments and the very helpful suggestions.

References

[1] A. D. Barbour, L. Holst and S. Janson, *Poisson Approximation,* Oxford University Press, Oxford, 1992.

[2] O. Barndorff-Nielsen, *Information and Exponential Families in Statistical Theory,* John Wiley & Sons, Chichester, 1978.

[3] Crédit Suisse First Boston, Fixed Income Research, *Convertible bond Winterthur Insurance with WinCAT coupons "Hail"*, Zürich, January 1997.

[4] M. Crouhy, D. Galai and R. Mark, *Model risk,* J. of Financial Engineering **7**, no. 3/4 (1998) 267–288.

[5] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap,* Chapman & Hall, New York, 1993.

[6] P. Embrechts, C. Klüppelberg and T. Mikosch, *Modelling Extremal Events for Insurance and Finance,* Springer-Verlag, Berlin, 1997.

[7] F. Giesbrecht and O. Kempthorne, *Maximum likelihood estimation in the three-parameter lognormal distribution,* J. Roy. Statist. Soc. Ser. B **38** (1976) 257–264.

[8] J. R. M. Hosking and J. R. Wallis, *Parameter and quantile estimation for the generalized Pareto distribution,* Technometrics **29** (1987) 339–349.

[9] D. V. Hinkley, *Inference about the change-point in a sequence of random variables,* Biometrika **57** (1970) 1–17.

[10] E. L. Lehmann, *Testing Statistical Hypotheses*, 2. ed., John Wiley & Sons, New York, 1986.

[11] H. Linhart and W. Zucchini, *Model Selection,* John Wiley & Sons, New York, 1986.

[12] P. McCullagh and J. P. Nelder, *Generalized Linear Models,* 2. ed., Chapman & Hall, London, 1989.

[13] H. Rootzén and N. Tajvidi, *Extreme value statistics and wind storm losses: A case study,* Scand. Actuarial J. (1997) 70–94.
[14] H.-H. Schiesser et al., *Klimatologie der Stürme und Sturmsysteme anhand von Radar- und Schadendaten,* Vdf, Hochschulverlag AG an der ETH, Zürich, 1997.
[15] G. A. F. Seber and C. J. Wild, *Nonlinear Regression,* John Wiley & Sons, New York, 1989.
[16] R. L. Smith, *Maximum likelihood estimation in a class of nonregular cases,* Biometrika **72** (1985) 67–90.
[17] G. G. Venter, *Premium calculation implications of reinsurance without arbitrage,* Astin Bulletin **21**, no. 2 (1991) 223–230.
[18] Winterthur Insurance, *Nachrangige Wandelanleihe 1997–2000 von CHF 399'500'000 mit* $2\,1/4\,\%$ *WinCAT-Coupons "Hagel",* Emissions- und Kotierungsprospekt.
[19] Winterthur Insurance, $2\,1/4\,\%$ *convertible bond with WinCAT coupon "Hail"*, Online information at URL `http://www.winterthur.com/prod/wincat/index-e.html.`
[20] Winterthur Insurance, *The Winterthur Share,* Online information at URL `http://www.winterthur.com/prod/aktien/index-e.html.`
[21] Winterthur Insurance, *Hagel und Schäden an Fahrzeugen, Ein Kurzbericht über die Bedrohung in der Schweiz,* internal report, January 1997.
[22] K. J. Worsley, *Confidence regions and tests for a change-point in a sequence of exponential family random variables,* Biometrika **73** (1986) 91–104.