© *The Journal of Risk and Insurance*, 2004, Vol. 71, No. 2, 309-322

A Data-Analytic Method for Forecasting Next Record Catastrophe Loss

Ping-Hung Hsieh

Abstract

We develop in this article a data-analytic method to forecast the severity of next record insured loss to property caused by natural catastrophic events. The method requires and employs the knowledge of an expert and accounts for uncertainty in parameter estimation. Both considerations are essential for the task at hand because the available data are typically scarce in extreme value analysis. In addition, we consider three-parameter Gamma priors for the parameter in the model and thus provide simple analytical solutions to several key elements of interest, such as the predictive moments of record value. As a result, the model enables practitioners to gain insights into the behavior of such predictive moments without concerning themselves with the computational issues that are often associated with a complex Bayesian analysis. A data set consisting of catastrophe losses occurring in the United States between 1990 and 1999 is analyzed, and the forecasts of next record loss are made under various prior assumptions. We demonstrate that the proposed method provides more reliable and theoretically sound forecasts, whereas the conditional mean approach, which does not account for either prior information or uncertainty in parameter estimation, may provide inadmissible forecasts.

Introduction

Forecasting extreme losses caused by catastrophic events such as hurricanes and earthquakes has been a major concern for market participants in the property/casualty (P/C) insurance industry. In particular, the recent series of catastrophic events has raised concerns about the financial stability of the industry. Because the industry's ability to model extreme losses has a direct impact on its decisions about managing catastrophe risk, and consequently on its capacity to provide adequate financial protection, modeling extreme losses has become of primary importance in the P/C

Ping-Hung Hsieh is Associate Professor of Quantitative Methods, College of Business, Oregon State University, USA, e-mail: hsiehph@bus.orst.edu; tel: 541-737-4297; fax 541-737-4890. The author would like to thank two anonymous referees for the comments and Gary Kerney of the Insurance Services Office for his assistance in collecting data. This study was supported by Summer Research Fellowship 2001, College of Business, Oregon State University.

insurance industry. Thus, developing statistical techniques to forecast extreme losses is certainly a major task for market participants.

One of the many problems encountered in forecasting extreme losses is the availability of extreme loss data. By definition, catastrophic events occur infrequently, and thus, any statistical analysis related to extreme events must deal with tail probability or extreme quantiles of the underlying loss distribution, using only the scarcity of historical data. Traditional statistical methods based on the estimation of the entire density are inappropriate for such tasks because these methods typically produce a good fit in those regions in which most of the data reside but at the expense of good fit in the tails. In recent years, extreme value theory (EVT), which includes wide-ranging topics such as record values and times, tail index estimation, point processes, excess over threshold, weak convergence of distributions, long memory, and self-similarity, has received great attention in the field of risk management. One of the main features of EVT is that it is tailored to the object of interest, namely, the tail rather than the center of the loss distribution, and thus offers sensible methods to deal with extreme losses. It has become one of the most important tools in modeling extreme losses and has motivated a great deal of applied work pertaining to various social, economic, and environmental issues; see, for example, Smith (1989), Koedijk, Schafgans, and de Vries (1990), Coles and Tawn (1990, 1991), Hols and de Vries (1991), Danielsson and de Vries (1997), Walshaw (2000), and Embrechts (2000).

In addition to the problem of data availability, another problem of representative data arises in forecasting extreme losses: consistent conditions. The underlying conditions of the time period over which the data are collected may be quite different from those of current and future periods. In the context of hurricane or earthquake losses, such change in conditions could be due to changes in land use, population densities, construction techniques, or building codes. For example, a 1994 report by Insurance Services Office, Inc. (ISO, 1994) indicates that the population density on the southeastern Atlantic coast of the United States, a storm-prone area, increased nearly 75 percent from 1970 to 1990, far greater than the countrywide population increase of more than 20 percent. It is therefore logical to include such information in the analysis, along with other sources of knowledge not already contained in the form of data. Considering the use of prior information naturally links to the Bayesian approach of extreme value analysis in forecasting extreme losses.

Increasing effort has been made to incorporate Bayesian methodology into extreme value analysis. Smith and Naylor (1987) systematically study the effect of selected prior distributions on the posterior distributions of the parameters of a Weibull distribution. Pickands (1994) discusses Bayesian estimation of extreme quantiles assuming independent noninformative priors among parameters. Hill (1994) proposes a Bayesian forecasting model of next record value and examines the issue of (in)finite predictive moments, which arises often in the context of long-tailed distributions. Coles and Tawn (1996) develop a Bayesian model to analyze extreme rainfall data. They illustrate how the careful elicitation of prior expert information can supplement data and lead to improved estimates of extremal behavior. Recently, a mixture model for extreme wind speeds was proposed by Walshaw (2000) to characterize two distinct extremal processes. A review article by Coles and Powell (1996) provides a comprehensive literature review that links the themes of Bayesian and extreme value analysis.

This article proposes a data-analytic method for forecasting the severity of next record-breaking catastrophe loss. The underlying theoretical model, extended from Hill's (1994) forecasting model, incorporates both Bayesian and EVT methods and allows for the use of prior knowledge and information. The forecasting procedure, including data normalization, assumption validation, hyperparameter assessment, and forecast computation, is discussed. In particular, we apply the model to examine natural catastrophes that caused US $25 million or more in direct insured losses to property in the United States between 1990 and 1999. By considering three-parameter Gamma priors for the parameter in the model, we extend Hill's (1994) model and derive analytical solutions to the predictive distribution and moments of next record value. The model thereby enables us to explore their behaviors with respect to prior information. Although recent advances in numerical techniques on parameter estimation, such as Markov chain Monte Carlo (MCMC) methodology, have made Bayesian evaluation of complex models straightforward, our closed-form formulas still have a significant role to play, especially in the context of forecasting catastrophe losses. We believe that the impact of prior information not contained in the form of data can be substantial in this case, and thus, market participants, who have limited training in, say, MCMC, should be allowed to experiment with different sets of information without fearing computational issues such as convergence of the chain, global versus local optima, and starting values of algorithms that are often related to the practice of MCMC. Note that the proposed method is close in spirit to the estimation of expected shortfall in finance; see Embrechts (2000) for more details.

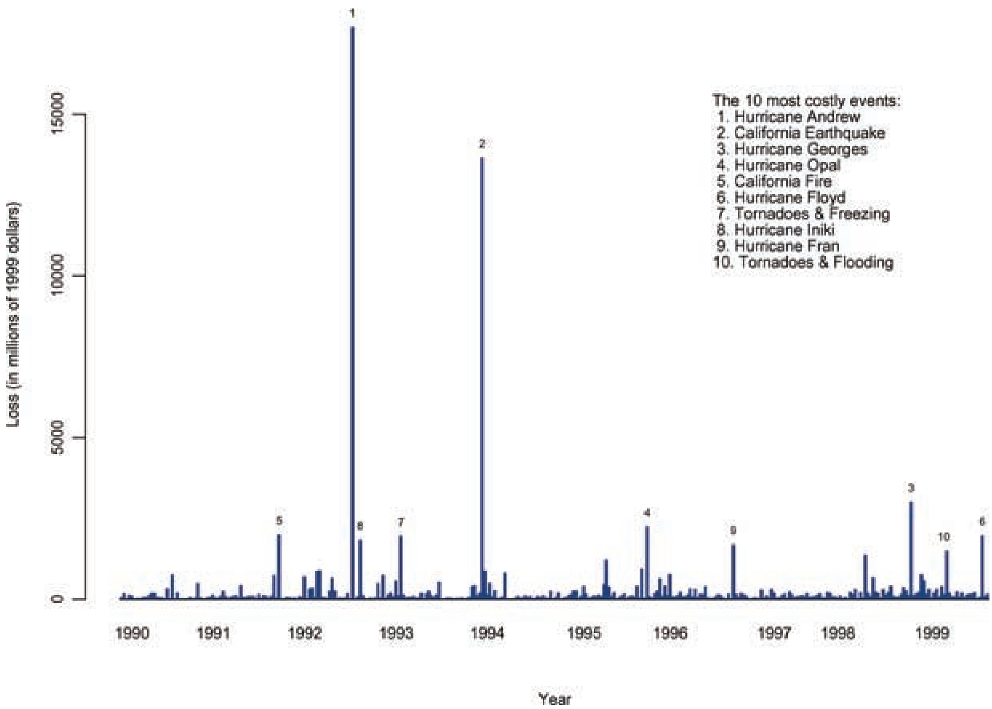
The remainder of this article is organized as follows. In the following section, the catastrophe data used in this study are described in detail. Also outlined is a data normalization process that normalizes the losses to values that are more representative of the impacts of past catastrophic events in today's context. In "A Bayesian Forecasting Model of Next Record Value," the theoretical model is formulated, and analytical solutions to the posterior distribution and moments of the parameter, as well as the predictive distribution and moment of next record value, are derived. The subsequent section discusses issues related to prior elicitation and presents the analysis results, followed by a summary in the final section.

NATURAL CATASTROPHE LOSSES BETWEEN 1990 AND 1999

Description of the Data

The data, provided by ISO, consist of information on natural catastrophes that occurred in the United States between 1990 and 1999, each of which resulted in US $25 million or more in property damage. For each catastrophic event, the data contain the following information: catastrophe serial number (CAT#), date(s) of occurrence, affected geographic areas, types of perils, and final insured property loss in current dollars. The catastrophe serial number assigned by the Property Claim Services permits all insurance companies to track reserves and losses related to a single, discrete event and is used throughout the P/C industry. Among the types of perils recorded, tropical storms are the prime cause of extreme losses. For example, among 33 natural catastrophes occurred in 1995, three losses caused by hurricanes (Hurricanes Erin,

FIGURE 1
Estimated Property Losses (in Millions of 1999 Dollars) Caused by Natural Catastrophes in the United States Between 1990 and 1999. The 10 Most Costly Events Are Numbered



Marilyn, and Opal) contributed to 40 percent of the total annual loss of US $8.3 billion in 1995 prices. Other types of perils include earthquakes, tornadoes, and flooding, all of which also contribute to great losses. Figure 1 presents the inflation-adjusted insured losses of these events in chronological order. As shown in the figure, among the 10 most costly catastrophic events, 6 of them are resulted from hurricanes.

Data Normalization

The problem that prevents us from using the data directly in modeling is the infrequent nature of catastrophic events, necessitating a long data-collection period. Even after adjusting for inflation, the comparison of two losses estimated at different points in time is still meaningless. This is because underlying societal conditions have changed over the years, and thus, a more effective adjustment should be applied to the data to account for such changes.

Following Pielke and Landsea (1998), we adjust the loss data to values that are more representative in today's context. In other words, we "normalize" the losses to produce the estimated loss of a catastrophic event as if it had happened this year. In this article, we consider three factors for data normalization: inflation, wealth, and population. Consideration of the inflation factor is a standard practice for data collected over a long period of time. A penny today is worth more than a penny tomorrow. The wealth factor accounts for the fact that people today have more to lose in terms of their possessions than did people years ago. Finally, the rapid growth in population in recent years has resulted in more people exposed to catastrophic risks. A disaster that occurred many years ago could cause much greater damage today, simply because more people reside in vulnerable locations. Therefore, by assuming the estimated loss is proportional to all three factors, we use the following formula to normalize each loss to its 1999 value:

$$NL_{99} = L_t \times I_t \times W_t \times P_t,$$

where NL_{99} is a normalized loss in 1999 value, L_t is a loss in year t in current dollars, I_t is the inflation factor determined by the ratio of the implicit price deflator for gross domestic product in 1999 to that of year t, W_t is the wealth factor determined by the ratio of the inflation-adjusted national income per capita in 1999 to that of year t; and P_t is the population factor determined by the ratio of the population in the United States in 1999 to that of year t.

Implicit price deflator and national income are provided by the U.S. Bureau of Economic Analysis, whereas estimates of the United States population can be found on the U.S. Census Bureau's Web site.

The top 10 normalized catastrophe losses between 1990 and 1999 are listed in Table 1. Although the list of the top 10 normalized losses roughly corresponds to those of losses in current and 1999 real dollars, the impacts of several events are perceived differently. For example, Hurricane Floyd (CAT# 97) is considered the sixth largest

TABLE 1
Top Catastrophe Losses Based on Current, Real (1999), and Normalized (to 1999) Dollars. All Losses Are in Millions of Dollars

		Loss in Millions of			Rank by		
Year	Event (CAT#)	Current Dollars	1999 $	Normalized $	Current	1999 $	Normalized $
1992	Hurricane Andrew (27)	15,500	17,682	30,166	1	1	1
1994	California earthquake (78)	12,500	13,641	20,009	2	2	2
1991	California fire (87)	1,700	1,986	3,645	7	5	3
1998	Hurricane Georges (69)	2,955	2,999	3,231	3	3	4
1992	Hurricane Iniki (30)	1,600	1,825	3,114	8	8	5
1993	Tornadoes/freezing (46)	1,750	1,949	3,089	6	7	6
1995	Hurricane Opal (54)	2,100	2,243	3,046	4	4	7
1996	Hurricane Fran (88)	1,600	1,676	2,112	9	9	8
1999	Hurricane Floyd (97)	1,960	1,960	1,960	5	6	9
1995	Wind/hail/flooding (38)	1,135	1,212	1,646	12	12	10

loss in real dollar terms but drops to ninth in normalized dollar term. The event (CAT# 38) of wind, hail, and flooding that hit Texas and New Mexico in 1995 is not on the top 10 list when measured in current and real dollars but is the 10th largest normalized loss. Table 1 also shows that damages caused by these catastrophic events are much more severe than what is generally believed. For example, the reported property loss caused by Hurricane Andrew is typically in the neighborhood of US $18 billion in 1999 value. However, when taking into account the increase in wealth and the growth of the population, the (normalized) loss could exceed US $30 billion, a value that is more representative of the impact of Hurricane Andrew if it had made landfall in 1999.

A BAYESIAN FORECASTING MODEL OF NEXT RECORD VALUE

Assume that a random sample of $X_1, X_2, \ldots, X_n$ is drawn from a distribution F with a density f defined on the positive domain. Further assume that the distribution F belongs to a class of distributions whose upper tail above a high threshold D can be approximated by an algebraic functional form. That is, $1 - F(x \mid \alpha) \approx Cx^{-\alpha}$ for $x \geq D$, where α is a positive real-valued parameter and C is a normalizing constant. Note that many important distributions commonly used in the insurance literature belong to this class. They include Pareto, generalized Pareto, Burr, transformed beta, and many more. See, for example, Stuart (1983), Philbrick (1985), Hogg and Klugman (1983, 1984), and Zajdenweber (1994, 1996). However, in contrast to these studies, we do not assume any specific distribution in our analysis.

Define $X_{(1)} \geq X_{(1)} \geq \cdots \geq X_{(n)}$ as the descending-order statistics. By conditioning on the parameter α and the largest $r + 1$-order statistics being greater than D, the conditional likelihood function for the parameter α is

$$L(\alpha) \propto \alpha^r \times \exp(-\alpha \cdot t), \tag{1}$$

where $t = \sum_{i=1}^{r} i \times \ln(x_{(i)}/x_{(i+1)})$, and $r + 1 \leq n$. Let $\pi(\cdot)$ be the prior distribution of α. For the purpose of forecasting the next record value (i.e., given $X > x_{(1)}$), we are interested in deriving the predictive distribution of next record value $F^*(X \mid X > x_{(1)})$ and its corresponding kth moment $E^*[X^k \mid X > x_{(1)}]$, both of which can be evaluated by integrating out the parameter α with respect to the posterior distribution $\pi^*(\alpha) \propto L(\alpha) \times \pi(\alpha)$.

Although the advent of MCMC methodology has made the evaluation of F^* and E^* straightforward for a wide range of π, we consider in this article a fairly flexible prior family, such that the analytical solutions to F^* and E^* are available and can be conveniently used by experts to gain insights into the behavior of next record values. Specifically, we consider the use of a three-parameter Gamma prior of α, denoted by GAM$[\lambda, \beta, \tau]$, with density defined as

$$\frac{\lambda^\beta}{\Gamma(\beta)}(\alpha - \tau)^{\beta-1} \exp[-\lambda \cdot (\alpha - \tau)], \tag{2}$$

where $\lambda > 0, \beta > 0, \alpha > \tau > 0$, and $\Gamma(\cdot)$ is the Gamma function. Relying on the repeated use of the binomial theorem $(1 + y)^n = \sum_{j=1}^{n} \binom{n}{j} y^j$ and the fact that the kth moment of a GAM[a, b, 0] is $a^{-k}\Gamma(b+k)/\Gamma(b)$, Hsieh (2001a) derives the analytical solutions to the following key quantities of interests:

1. The kth posterior moment of α is

$$E^{\pi^*}[\alpha^k] = \tau^k \frac{E\left[W_1^{r+k}\right]}{E\left[W_1^r\right]}. \tag{3}$$

2. The predictive distribution of X given $X > x_{(1)}$ is

$$F^*(x \mid X > x_{(1)}) = 1 - \left(\frac{x}{x_{(1)}}\right)^{-\tau} \left(\frac{t+\lambda}{t+\lambda+\ln\left(\frac{x}{x_{(1)}}\right)}\right)^{\beta} \frac{E[W_2^r]}{E[W_1^r]}. \tag{4}$$

3. The kth predictive moment of X given $X > x_{(1)}$ for $k \le \tau$ is

$$E^*\left[X^k \mid X > x_{(1)}\right] = x_{(1)}^k \frac{-\sum_{j=0}^{\infty} \left(\frac{\tau}{k}\right)^{j+1} E\left[W_1^{j+r+1}\right]}{E\left[W_1^r\right]}. \tag{5}$$

4. In particular, if $k = \tau$, then Equation (5) reduces to

$$E^*[X^\tau \mid X > x_{(1)}] = x_{(1)}^\tau \frac{\tau(t+\lambda)}{\beta - 1} \frac{E\left[W_3^{r+1}\right]}{E\left[W_1^r\right]}, \tag{6}$$

where W_1, W_2, and W_3 are random variables with distributions GAM[$\tau(t+\lambda)$, β, 1], GAM[$\tau(t+\lambda+\ln(x/x_{(1)}))$, β, 1], and GAM[$\tau(t+\lambda)$, $\beta - 1$, 1], respectively, and $t = \sum_{r=1}^{r} i \times \ln(x_{(i)}/x_{(i+1)})$.

FORECASTING NEXT RECORD CATASTROPHE LOSS

Before we can apply Equations (3)–(6) to forecast next record loss, we must first determine the values of the parameters r, τ, λ, and β. Methods and issues pertaining to this task are now discussed, followed by the analysis results.

Validation of the Pareto Assumption

The likelihood function $L(\alpha)$ defined in Equation (1) is derived on the basis of the assumption that the largest $r+1$ upper-order statistics reside in the region where the underlying distribution function is of algebraic form. If a large value of r is selected, then the algebraic form assumption may not be valid, and the resulting forecasts can be biased. However, a small value of r leads to a smaller effective sample size, and the forecasts may have relatively large variance. This variance-bias trade-off problem has become a research topic of great interest in extreme value analysis, and various studies, mostly based on asymptotic argument, have been conducted along this line (see, for example, Hall, 1982, 1990; Dekkers and de Haan, 1993; Drees and Kaufmann, 1998; Beirlant et al., 1999; Feuerverger and Hall, 1999; Matthys and Beirlant, 2000). As pointed out by Resnick (1997), however, some of these selection procedures require rather explicit knowledge of the underlying distribution and thus are not practical solutions to the problem of choosing an r. Furthermore, because of the asymptotic nature of these procedures, they have to be applied with some care in finite samples.

Selection of an optimal r can be approached by using data-analytic methods (Dupuis, 1999; Hsieh, 1999). These methods are more relevant to our current analysis because only a finite sample is available. The main idea in Hsieh (1999) is to measure the discrepancy between the empirical distribution and the assumed algebraic functional form using a test statistic. Starting from $r = 1$, if the discrepancy is within a predetermined tolerance level s, then the next largest upper-order statistic is included and the discrepancy is recalculated. The process then repeats until, eventually, the discrepancy is greater than s, at which time the assumption of the algebraic form is no longer considered valid and the desired r is obtained.

Specifically, let $\hat{\alpha}_r$ be the maximum likelihood estimator of α derived from Equation (1) (Hill, 1975). Define a test statistic

$$H^{(r)} = \hat{\alpha}_r^2 \sum_{i=1}^{r} \left(i V_i - \hat{\alpha}_r^{-1}\right)^2, \tag{7}$$

where $V_i = \ln(x_{(i)}/x_{(i+1)})$ for $i = 1, \ldots, r$. If the largest $r + 1$ observations fall in the upper tail region where the algebraic form assumption holds, then the variables iV_i should in all respects behave like a random sample from an exponential distribution with parameter α for $i = 1, \ldots, r$. Hence, the test statistic $H^{(r)}$ can be viewed as a measure of discrepancy between the underlying (transformed) distribution and exponential distribution. A decision parameter s is predetermined such that the largest observations will be included sequentially in calculation of $\hat{\alpha}_r$ as long as the (standardized) $H^{(r)}$ is below s, that is, the discrepancy is below the tolerance level. A decision rule for choosing an optimal s^* under the squared error loss $(1/\alpha - 1/\hat{\alpha}_r)^2$, as studied in Hsieh (1999), is

$$s^* = \begin{cases} 0.34 \log_{10} n & \text{for } n \leq 1{,}000 \\ 1.02 & \text{for } n \geq 1{,}000, \end{cases} \tag{8}$$

where n is the sample size, and the decision rule s^* is shown to hold for a large class of distributions including the inverted gamma, T, generalized Pareto, Pareto, Burr, and log-logistic distributions. In our present analysis, we rely on this approach to determine the optimal r value. Note that a similar sequential procedure, albeit asymptotic in nature, can be found in a recent article by Guillou and Hall (2001).

A Guideline for Choosing the Hyperparameters τ, λ, and β

In an ideal situation, if an expert's prior knowledge about the parameter α is rather extensive (e.g., he or she can specify the quartiles of the prior $\pi(\alpha)$), then the values of the hyperparameters τ, λ, and β that are consistent with the specified quartiles can be solved numerically using Equation (2). In most cases, however, such extensive elicitation is not realistic. An expert may, at best, have some opinions regarding the average size of historical losses and possibly about a particular value of α. In this section, we offer some guidelines for choosing initial values for these parameters.

The parameter τ reflects the belief of a finite kth moment but an infinite $(k+1)$th moment of the loss X. Note that conditional on the parameter α and the new record value $X > x_{(1)}$, the kth conditional moment of X is finite if and only if $\alpha > k$. Consequently, the kth predictive moment $E^*[X^k \mid X > x_{(1)}]$ is finite if and only if the *a priori* distribution $\pi(\alpha)$ sufficiently downweights values of α below k. Thus, the kth predictive moment of $X > x_{(1)}$ is finite for all $k \le \tau$. Hence, if an expert believes, for example, that the first moment of a loss distribution is finite but the second moment may be infinite, then he or she can assume the domain of the prior $\pi(\alpha)$ to be $[1.0, \infty)$, that is $\tau = 1$, to reflect such a belief.

The parameters λ and β together can be used to represent the degree of certainty of the average value (m) or the most likely value (m_0) of α perceived by an expert. This is because the mean m and the mode m_0 of a GAM$[\lambda, \beta, \tau]$ are $\tau + \beta/\lambda$ and $\tau + (\beta - 1)/\lambda$, respectively, with variance $(\beta/\lambda) \times (1/\lambda)$. With a specified m or m_0, and thus, the ratio β/λ or $(\beta - 1)/\lambda$ is fixed, the distribution becomes more concentrated about the mean m, and the mode m_0 as λ or β increases. Hence, if prior knowledge on m or m_0 is rather vague, then a small value of λ or β should be chosen. As certainty about the m or m_0 value grows, a larger value of λ or β should be considered.

In our current analysis of catastrophe losses, we set $\tau = 1$ and examine both the mean and mode values of α between 1.01 and 2.50. We arbitrarily choose $\beta = 2$, 5, and 10 to reflect the degree of certainty of the mode and mean values of α, ranging from the least to the most certain.

Empirical Results

Given the loss distribution F where $1 - F(x \mid \alpha) \approx Cx^{-\alpha}$ for $x \ge D$, the conditional mean $E[X|X > x_{(1)}, \alpha] = x_{(1)} \times \alpha/(\alpha - 1)$ may be used as a forecast of next record value, where α may be replaced with the maximum likelihood estimate $\hat{\alpha}_r$. By applying the $H^{(r)}$ statistic on our data set of size $n = 342$, we find that the discrepancy between the assumed algebraic form and the empirical distribution exceeds the optimal $s^* = 0.86$ at $r + 1 = 4$, which leads to an estimate of $\hat{\alpha}_r = 0.72$. This result presents a problem to decision makers because the conditional mean $E[X|X > x_{(1)}, \hat{\alpha}_r]$ does not exist for $\hat{\alpha}_r \le 1$ and, thus, the forecast is not available. To demonstrate that the conditional mean approach is unreliable, we randomly permute the data and generate a new data set using the first half of the data. For each permutation, we apply the $H^{(r)}$ statistic to determine the optimal r value and then calculate both $\hat{\alpha}_r$ and $E[X|X > x_{(1)}, \hat{\alpha}_r]$. More than 50 percent of the time, $\hat{\alpha}_r$ is less than or equal to 1, and thus, the conditional mean approach cannot provide reliable forecasts. We also try fixing r at 2, 5, 10, and 20 for each permutation, and the results provide a similar conclusion.

The problem of infeasible forecasts resulting from the conditional mean approach cannot be avoided even if a decision maker is certain that $\alpha > 1$, and some arbitrary rules might be implemented to supply a forecast. The proposed predictive mean approach as defined in Equations (5) and (6), in contrast, allows information such as $\alpha > 1$ to be incorporated into the formal analysis and thus provides a sensible and theoretically sound forecast. Table 2 summarizes the forecasts made by the predictive

TABLE 2
Forecasts of Next Record Catastrophe Loss. For $\beta = 2$, 5, and 10, Forecasts Are Made Using Equation (6) With $\tau = 1$. For $\beta \to \infty$, the Forecasts Approach the Conditional Mean Defined as $x_{(1)} \times m_0/(m_0 - 1) \approx x_{(1)} \times m/(m - 1)$. All Forecasts Are in Billions of 1999 Normalized Dollars

	$\beta = 2$ Least Informative	$\beta = 5$ Informative	$\beta = 10$ Most Informative	$\beta \to \infty$
		Prior mode m_0 of $\pi(\alpha)$ is specified		
1.01	3,096.45	3,058.98	3,052.18	3,046.77
1.1	395.18	346.82	338.39	331.83
1.2	251.78	198.19	188.57	181.00
1.3	205.66	149.52	139.08	130.72
1.4	183.15	125.60	114.57	105.58
1.5	169.89	111.45	100.01	90.50
1.6	161.16	102.15	90.38	80.44
1.7	154.99	95.58	83.57	73.26
1.8	150.39	90.70	78.49	67.87
1.9	146.84	86.93	74.57	63.68
2.0	144.02	83.94	71.45	60.33
2.1	141.72	81.52	68.91	57.59
2.2	139.81	79.50	66.80	55.30
2.3	138.20	77.81	65.03	53.37
2.4	136.83	76.37	63.52	51.71
2.5	135.64	75.12	62.21	50.28
		Prior mean m of $\pi(\alpha)$ is specified		
1.01	6,111.93	3,813.06	3,387.34	3,046.77
1.1	690.55	421.69	371.80	331.83
1.2	395.18	235.11	205.14	181.00
1.3	298.94	173.76	150.03	130.72
1.4	251.78	143.51	122.72	105.58
1.5	223.95	125.60	106.47	90.50
1.6	205.66	113.80	95.72	80.44
1.7	192.74	105.46	88.11	73.26
1.8	183.15	99.26	82.43	67.87
1.9	175.76	94.49	78.05	63.68
2.0	169.89	90.70	74.57	60.33
2.1	165.11	87.61	71.73	57.59
2.2	161.16	85.06	69.38	55.30
2.3	157.83	82.92	67.40	53.37
2.4	154.99	81.08	65.71	51.71
2.5	152.53	79.50	64.25	50.28

mean with respect to a decision maker's beliefs on different values of the modes (m_0) and the means (m) of prior $\pi(\alpha)$, as well as the degrees ($\beta = 2, 5$, and 10) of uncertainty of these beliefs. For example, suppose an expert is confident that the conditional mean exists (i.e., $\alpha > 1$ and thus $\tau = 1$) but believes with a small degree of certainty (i.e., $\beta = 2$) that the mode m_0 of the prior $\pi(\alpha)$ is 2. Given the current record (normalized) loss of US \$30 billion, the forecast of next record loss is thus US \$144 billion, a value that accounts for a decision maker's uncertainty on the parameter α. In this scenario, should the belief about $m_0 = 2$ become stronger ($\beta \to \infty$) as new knowledge or experience is acquired, the forecast of next record loss will be revised toward US \$60 billion. In other words, as the posterior distribution of α degenerates to a single value $\alpha = m_0$, the resulting predictive mean is essentially equivalent to the conditional mean of $x_{(1)} \times m_0/(m_0 - 1)$.

We stress that Table 2 is not meant to be used retrospectively after historical data have been observed. It is presented to demonstrate that the proposed method is able to provide admissible forecasts even when $m_0(m) \to 1$ and to offer more conservative forecasts when the prior is less informative. It also serves for a decision maker to gain insights into the impact of a misspecified $m_0(m)$ value. If the loss distribution is relatively short-tailed (i.e., the value of $m_0(m)$ is large), then a small discrepancy between a specified $m_0(m)$ and its true value has a minimal impact on the resulting forecasts, regardless of the amount of information (i.e., the β value). For $m_0(m)$ close to 1, extra effort is required to obtain more information on α because a small discrepancy between a specified $m_0(m)$ and its true value may lead to a mega difference in the resulting forecasts (compare, e.g., the forecasts made with $m_0(m) = 1.01$ versus those with $m_0(m) = 1.1$).

One final question of interest is whether the forecasts presented in Table 2 using only loss data between 1990 and 1999 are "in the ball park." By examining a much larger and more comprehensive data set, Pielke and Landsea (1998) report a record (normalized) loss of US \$74.4 billion in 1995 value (approximately US \$108 billion in 1999 value) set by a hurricane in 1926. The forecasts in Table 2 seem to be in line with this value.

DISCUSSION

The $H^{(r)}$ statistic plays an important role in the proposed data-analytic method for forecasting next record catastrophe loss. The statistic is constructed based on the so-called Hill estimator $\hat{\alpha}_r$ (Hill, 1975). It has been noted that the Hill estimator can be severely biased. Embrechts, Klüppelberg, and Mikosch (1997) even created and termed "Hill horror plots" to illustrate the poor performance of the estimator. Hence, one might be concerned about how the biased problem of the Hill estimator will affect the performance of the $H^{(r)}$ statistic, and consequently, the forecasts of the proposed data-analytic method.

The Hill estimator is indeed an unbiased estimator provided that the underlying Pareto assumption is valid. Its bias problem often arises from the user overlooking such an assumption. Typically, the Pareto assumption is only valid in the very far tail (i.e., a very high threshold D), but one who applies the peaks over threshold method may choose the threshold D simply for the reason of convenience (e.g., use the retention limit of an insurance policy as the threshold). But there is no guaranty that such threshold is actually located in the tail area where the Pareto assumption is valid, and thus, one may still take observations residing outside the Pareto tail area in calculating the Hill estimate and the bias results. The $H^{(r)}$ statistic (Hsieh, 1999) takes into account this bias problem. It measures the discrepancy between the underlying distribution and the assumed Pareto functional form, and provides a threshold D above which the underlying distribution can be reasonably approximated by a Pareto functional form. Thus, by using the $H^{(r)}$ statistic in the proposed forecasting model, we offer a simple and statistically defensible way to validate the Pareto assumption.

We also acknowledge that a more complex tail model such as $1 - F(x) \propto Cx^{-\alpha}(1 + \delta(x))$ (e.g., Feuerverger and Hall, 1999) may be appropriate in some applications. But for our current application of forecasting next record value, we believe that the assumed tail model $1 - F(x) \propto Cx^{-\alpha}$ is sufficient for the following reasons: First, the proposed data-analytic method calls for the use of the $H^{(r)}$ statistic to identify the threshold $D < x_{(1)}$, above which the underlying distribution can be reasonably approximated by $Cx^{-\alpha}$. The research problem of interest is to predict the next record value exceeding the current record $x_{(1)}$. Hence, it is reasonable to assume that the tail of the loss distribution beyond $x_{(1)}$ can be approximated by $Cx^{-\alpha}$. That is, $\delta(x) \approx 0$. Second, the loss data under consideration were caused by natural disasters with a loss of US \$25 million or more. The effective sample size is relatively small. Given the amount of the data, it is difficult to assess an appropriate functional form for $\delta(x)$. Even if the form of $\delta(x)$ is assumed known (a big *if*), the additional parameters in $\delta(x)$ and their corresponding hyperparameters in the prior distributions might lead us to wonder if it is worth the trouble. After all, we are interested in the tail of the distribution exceeding the current record value caused by a catastrophic event and presumably located far in the tail.

SUMMARY AND FURTHER RESEARCH

In this article, we proposed a data-analytical method to study the behavior of next catastrophe loss. There are four steps involved in the forecasting process: (1) data normalization, (2) validation of the Pareto assumption, (3) selection of hyperparameters, and (4) computation of the predictive moments. A real data set is used to demonstrate the use of this method. The study offers a theoretically sound and defensible approach to extrapolating beyond the range of data in the context of forecasting record catastrophe losses.

The data set examined consists of natural catastrophic losses occurring in the United States between 1995 and 1999. The proposed model recognizes the importance of incorporating an expert's prior experience and knowledge and accounts for uncertainty in parameter estimation. Both considerations are essential in extreme value analysis in which data are typically scarce. We derived simple analytical solutions to several key elements of interest, such as the posterior moments of the parameter α and the predictive distribution and moments of record value. These analytical solutions thereby enable practitioners to gain insights into the behavior of the predictive moments without concerning themselves with issues related only to computationally intensive approaches. Finally, we analyzed the data set and observed that, in contrast with the traditional conditional mean approach, the proposed model provides more sensible and reliable forecasts for decision makers.

Although we did not discuss the problem of infinite moments that is often associated with extreme value analysis, we consider it an important extension of this work. The maximum likelihood estimate, as well as the posterior means and standard deviations of α derived under various values of m_0 and m, seem to suggest that the second moment of the underlying loss distribution is possibly infinite because the evidence of $\alpha \leq 2$ is strong. Therefore, we were not able to provide the standard deviation of the predictive distribution or perform further statistical inference. To derive a sensible approximation to the standard deviation or any higher moment in our analysis, we could select an upper bound M on the extreme right tail of the loss distribution, such that the probability mass beyond M is negligible. Should we be able to justify the selection of M and understand its impact on the resulting forecasts, the problem of infinite moment will become irrelevant. A more detailed discussion on the issues pertaining to the use of the finite versus infinite idealized models can be found in Hill (1994) and Hsieh (2001b).

Finally, if the issues associated with the computationally intensive Bayesian approaches are not of primary concern, the exploration of other families of prior distributions is also recommended.

REFERENCES

Beirlant, J., G. Dierckx, Y. Goegebeur, and G. Matthys, 1999, Tail Index Estimation and an Exponential Regression Model, *Extremes*, 2: 177-200.

Coles, S. G., and E. A. Powell, 1996, Bayesian Methods in Extreme Value Modelling: A Review and New Developments, *International Statistical Review*, 64: 119-136.

Coles, S. G., and J. A. Tawn, 1990, Statistics of Coastal Flood Prevention, *Philosophical Transactions of the Royal Society of London, Series A*, 332: 457-476.

Coles, S. G., and J. A. Tawn, 1991, Modelling Extreme Multivariate Events, *Journal of the Royal Statistical Society, Series B*, 53: 377-392.

Coles, S. G., and J. A. Tawn, 1996, A Bayesian Analysis of Extreme Rainfall Data, *Applied Statistics*, 45: 463-478.

Danielsson, J., and C. G. de Vries, 1997, Tail Index and Quantile Estimation with Very High Frequency Data, *Journal of Empirical Finance*, 4: 241-257.

Dekkers, A., and L. de Haan, 1993, Optimal Choice of Sample Fraction in Extreme-Value Estimation, *Journal of Multivariate Analysis*, 47: 173-195.

Drees, H., and E. Kaufmann, 1998, Selecting the Optimal Sample Fraction in Univariate Extreme Value Estimation, *Stochastic Processes and their Applications*, 75: 149-172.

Dupuis, D. J., 1999, Exceedances over High Thresholds: A Guide to Threshold Selection, *Extremes*, 1: 251-261.

Embrechts, P., ed., 2000, *Extremes and Integrated Risk Management* (London, UK: Risk Books).

Embrechts, P., C. Klüppelberg, and T. Mikosch, 1997, *Modelling Extremal Events* (Berlin: Springer-Verlag).

Feuerverger, A., and P. Hall, 1999, Estimating a Tail Exponent by Modelling Departure from a Pareto Distribution, *The Annals of Statistics*, 27: 760-781.

Guillou, A., and P. Hall, 2001, A Diagnostic for Selecting the Threshold in Extreme Value Analysis, *Journal of the Royal Statistical Society, Series B*, 63: 293-305.

Hall, P., 1982, On Simple Estimates of an Exponent of Regular Variation, *Journal of the Royal Statistical Society, Series B*, 44: 37-42.

Hall, P., 1990, Using the Bootstrap to Estimate Mean Squared Error and Select Smoothing Parameter in Nonparametric Problems, *Journal of Multivariate Analysis*, 32: 177-203.

Hill, B. M., 1975, A Simple General Approach to Inference about the Tail of a Distribution, *The Annals of Statistics*, 3: 1163-1174.

Hill, B. M., 1994, Bayesian Forecasting of Extreme Values in an Exchangeable Sequence, *Journal of Research of the National Institute of Standards and Technology*, 99: 521-538.

Hogg, R.V., and S. A. Klugman, 1983, On the Estimation of Long Tailed Skewed Distributions with Actuarial Applications, *Journal of Econometrics*, 23: 91-102.

Hogg, R.V., and S. A. Klugman, 1984, *Loss Distributions* (New York: Wiley).

Hols, M. C. A. B., and C. G. de Vries, 1991, The Limiting Distribution of Extremal Exchange Rate Returns, *Journal of Applied Econometrics*, 6: 287-302.

Hsieh, P.-H., 1999, Robustness of Tail Index Estimation, *Journal of Computational and Graphical Statistics*, 8: 318-332.

Hsieh, P.-H., 2001a, On Bayesian Predictive Moments of Next Record Value Using Three-Parameter Gamma Priors, *Communications in Statistics: Theory and Methods*, 30: 729-738.

Hsieh, P.-H., 2001b, Robustness of Conditional Moments: An Application to Premium Calculation of Reinsurance Treaties, *Risk Analysis*, 21, 225-234.

Insurance Services Office, Inc., 1994, The Impact of Catastrophes on Property Insurance. ISO Insurance Issues Series.

Koedijk, K. G., M. M. A. Schafgans, and C. G. de Vries, 1990, The Tail Index of Exchange Rate Returns, *Journal of International Economics*, 29: 93-108.

Matthys, G., and J. Beirlant, 2000, Adaptive Threshold Selection in Tail Index Estimation, in: P. Embrechts, ed., *Extremes and Integrated Risk Management* (London, UK: Risk Books).

Philbrick, S. W., 1985, A Practical Guide to the Single Parameter Pareto Distribution, *Proceedings of the Casualty Actuarial Society*, LXXII: 44-123.

Pickands, J., III, 1994, Bayes Quantile Estimation and Threshold Selection for the Generalized Pareto Family, in: J. Galambos et al., eds., *Extreme Value Theory and Applications* (Dordrecht: Kluwer Academic).

Pielke, R. A., Jr., and C. W. Landsea, 1998, Normalized Hurricane Damages in the United States: 1925–95, *Weather Forecasting*, 13: 621-631.

Resnick, S. I., 1997, Heavy Tail Modeling and Teletraffic Data, *The Annals of Statistics*, 25: 1805-1869.

Smith, R. L., 1989, Extreme Value Analysis of Environmental Time Series: An Application to Trend Detection in Ground-Level Ozone, *Statistical Science*, 4: 367-377.

Smith, R. L., and J. C. Naylor, 1987, A Comparison of Maximum Likelihood and Bayesian Estimators for the Three-Parameter Weibull Distribution, *Applied Statistics*, 36: 358-369.

Stuart, C., 1983, Pareto-Optimal Deductibles in Property-Liability Insurance: The Case of Homeowner Insurance in Sweden, *Scandinavian Actuarial Journal*, 4: 227-238.

Walshaw, D., 2000, Modelling Extreme Wind Speeds in Regions Prone to Hurricanes, *Applied Statistics*, 49, 51-62.

Zajdenweber, D., 1994, Extreme Values in Business Interruption Insurance, *Journal of Research of the National Institute of Standards and Technology*, 99, 231-235.

Zajdenweber, D., 1996, Extreme Values in Business Interruption Insurance, *Journal of Risk and Insurance*, 63, 95-110.