

Micro/Macro-Economic and Spatial Individual Risk Models for Catastrophe Insurance Applications

HÉLÈNE COSSETTE, THIERRY DUCHESNE
and ÉTIENNE MARCEAU

July 10, 2002

In this paper we propose two generalized versions of the individual risk model that include the possibility of catastrophes. Even though the models were developed with home insurance in mind, they can be used in other contexts. The first model is a mathematical formalization of the simulation techniques used in the insurance industry. It allows for a micro-economic formulation useful for ratemaking and a macro-economic formulation more appropriate for catastrophe bond or option pricing. The second model is a spatial specification of the individual risk model that includes correlation between risks as a function of geographic proximity. We outline some of the properties of these more general models and propose methods of specifying their various elements and of estimating their parameters using information that can be found in the literature and on the World Wide Web for the first model, or estimated from insurance data for the spatial model.

Key words: Attenuation function, auto-logistic, auto-gaussian, CAT-bond, CAT-option, correlation, damage function, dependence, hurricane, spatial statistics.

1 Introduction

The traditional individual and collective risk models are very useful in casualty insurance. They are simple and convenient and they constitute good approximations to the claim generating mechanism in the usual “daily business” context. Unfortunately, these models assume that the risks are independent, which is not appropriate when single events (catastrophes) can produce several correlated claims at the same time. This is especially true in home in-

insurance where natural disasters such as hurricanes, earthquakes, winter storms, floods, etc. frequently produce thousands of correlated claims, sometimes totalling billions of dollars.

Several authors have pointed out that dependence should be considered in risk theory and have proposed tools to include this dependence in the models. (See Wang, 1998, for a survey of this question.) In the case of home insurance, dependence arises from the geographic proximity between risks. Several papers in the actuarial literature (Clark, 1986, Burger et al., 1996, Walters and Morin, 1996, Huang et al., 2001, Collins and Lowe, 2001) model this spatial dependence by simulating realistic catastrophe scenarios based on historical data and geological, meteorological, hydrological and/or civil engineering models. Unfortunately, the simulation models used are proprietary and very little is said about their use in risk theory. In this paper, we give a mathematical formalization of these simulation models. We first start by generalizing the simple individual risk model so that it mimics more closely the claim generating mechanism. We then derive some of the properties of the new model. We show how this generalized model can be used to perform actuarial calculations and illustrate its use for home insurance calculations. Throughout our development, we illustrate the method using the hurricane risk. We then use a completely different modeling approach and propose a “spatial individual risk” model.

2 Building a realistic catastrophe model

Consider a portfolio with n insured risks, with risk i producing claims of amount X_i in a given period of interest (a calendar year, say). The traditional individual risk model assumes that X_1, \dots, X_n are independent random variables. In a context where natural catastrophes are possible, the independence assumption is not realistic. In this section we take two steps to generalize this model so as to introduce a realistic dependence structure between the claim amounts.

2.1 A simple model with catastrophes

We suppose that the portfolio is partitioned into independent classes. The partitioning is done so that when a catastrophe occurs it affects every single risk within the class. For a given class, suppose that we have n insured risks. Let $S_n = \sum_{i=1}^n X_i$ be the aggregate claim amount for the period under consideration for the class, where X_i is the claim amount for the i th risk. We let $X_i = Y_i + Z_i$, where $Y_i, i = 1, \dots, n$ are n independent random variables with Y_i representing the claim amounts due to “usual business”, and $Z_i, i = 1, \dots, n$, are n random variables that represent the catastrophic losses incurred by risk i during the period

under consideration. If we let $T_n = \sum_{i=1}^n Y_i$ and $W_n = \sum_{i=1}^n Y_i$, then $S_n = T_n + W_n$ can be viewed as the sum of the total claims due to usual business (T_n) and the total claims due to catastrophes (W_n).

We model Y_i and Z_i as follows. First, we put

$$Y_i = \begin{cases} \sum_{k=1}^{M_i} B_{i,k}, & M_i > 0 \\ 0, & M_i = 0 \end{cases}$$

where M_i is the number of claims due to “usual business” for risk i and $B_{i,k}$ is the amount of the k th claim due to usual business for risk i . We suppose that $B_{i,1}, \dots, B_{i,n}$ are iid and independent of M_i and of $B_{j,k}$, $j \neq i$, $k = 1, 2, \dots$. We also suppose that M_1, \dots, M_n are independent. Now we put

$$Z_i = \begin{cases} \sum_{k=1}^{M_0} C_{i,k}, & M_0 > 0 \\ 0, & M_0 = 0 \end{cases}$$

where M_0 is the number of catastrophes to hit the class in the period under consideration and where the $C_{i,k}$'s share the same properties as the $B_{i,k}$'s above. Obviously, the Y_i 's are independent while the Z_i 's are not.

For a more “macro-economic” approach, if $M_0 > 0$ let $D_k = C_{1,k} + \dots + C_{n,k}$, $k = 1, \dots, M_0$; D_k represents the total cost of the k th catastrophe. Then W_n can be rewritten as

$$W_n = \begin{cases} \sum_{k=1}^{M_0} D_k, & M_0 > 0 \\ 0, & M_0 = 0. \end{cases}$$

This representation is closer to the approach used in extreme value modeling (see Beirlant and Teugels, 1992 or Embrechts et al., 1998, for example) and catastrophe option pricing (see Christensen and Schmidli, 2000, for example), where we usually model the distribution of D_k or W_n directly.

This simple model with dependence has been proposed and studied by Cossette et al. (2000), among others. Unfortunately, this model is somewhat unrealistic. It does not correlate risks from different classes together (which requires the classes to be large to be appropriate) and it assumes that if a catastrophe hits a class, every risk in the class is hit in the same way (which now requires the classes to be small to be appropriate). Moreover, it assumes that all the M_0 catastrophes that happen have the same effect on every risk.

2.2 A general catastrophe model

We now generalize the catastrophe model so as to add correlations between classes/areas and varying intensities from catastrophe to catastrophe, and from risk to risk for a given

catastrophe. We base our model on the catastrophe simulation algorithms used in the insurance industry (Clark, 1986, Burger et al., 1996, Walters and Morin, 1996, Huang et al., 2001, Collins and Lowe, 2001).

DEFINITION 2.1 *Our general catastrophe model (GCM) is given by*

$$S = \sum_{r=1}^R \sum_{i=1}^{n_r} X_{ri} = \sum_{r=1}^R \sum_{i=1}^{n_r} Y_{ri} + \sum_{r=1}^R \sum_{i=1}^{n_r} \sum_{k=1}^{M_0} C_{rik}(I_{rk}), \quad (1)$$

where

- S = aggregate claim amount for portfolio
- X_{ri} = total claims for risk i in area r
- R = number of areas
- n_r = number of insured risks in area r
- Y_{ri} = claims due to “usual business” for risk i in area r
- M_0 = number of catastrophes to hit the portfolio in period
- I_{rk} = “intensity” of k th catastrophe felt in area r
- $C_{rik}(I_{rk})$ = claim due to the k th catastrophe for risk i in area r ,
whose distribution depends on I_{rk}

We assume that the Y_{ri} ’s are as in Section 2.1 and that **given** M_0 **and the** I_{rk} ’s, the $C_{rik}(I_{rk})$ ’s are independent. We assume that M_0 is independent of the Y_{ri} ’s and of the $C_{rik}(I_{rk})$ ’s and that the Y_{ri} ’s are independent of the $C_{rik}(I_{rk})$ ’s.

Since the “intensities felt” I_{rk} are assumed constant over areas, the partition of the portfolio into areas must be fine. Computer simulations of hurricane and earthquake events used in actuarial practice tend to generate constant intensities at the zip/post code level, which suggests that defining areas as zip/post codes to be a sensible approach. If the portfolio is partitioned into big areas where the intensity felt is unlikely to be the same for each risk within the area, one can use the same model but replace $C_{rik}(I_{rk})$ in (1) by $C_{rik}(I_{rik})$ and now I_{rik} represents the intensity of the k th catastrophe felt by risk i in area r .

For CAT-bond or CAT-option pricing, it is more convenient to rewrite the catastrophic part of S in the GCM, say S_{CAT} , in a more “macro-economic” fashion:

$$S_{CAT} = \sum_{k=1}^{M_0} D_k,$$

where $D_k = \sum_{r=1}^R \sum_{i=1}^{n_r} C_{rik}(I_{rk})$ is the total cost of the k th catastrophe. Christensen and Schmidli (2000) discuss how CAT-futures can be priced when models for M_0 and D_k are available. In the remainder of this paper, we will discuss modeling options for $C_{rik}(I_{rk})$ and I_{rk} (which in turn will lead to models for D_k) and for M_0 .

2.2.1 Models for $C_{rik}(I_{rk})$

Portion of insured amount

Let b_{ri} be the insured amount for risk i of area r . Let \mathcal{I} be the set of all possible values that can be taken by the intensity random variables I_{rk} . We suppose that there exist random variables $\psi_{ri}(I_{rk})$ that, given an intensity I_{rk} , give the ‘‘proportion of damage’’ caused to each risk. Thus, we define

$$C_{rik}(I_{rk}) = b_{ri}\psi_{ri}(I_{rk}). \quad (2)$$

We let the cumulative distribution function of ψ_{ri} given an intensity I_{rk} be given by

$$P[\psi_{ri}(I_{rk}) \leq x | I = I_{rk}] = F_{\psi|I}(x|I_{rk}), \quad x \in [0, 1].$$

For convenience we let

$$\begin{aligned} \mu_{\psi}(I_{rk}) &= \int_0^1 x dF_{\psi|I}(x|I_{rk}) \\ \sigma_{\psi}^2(I_{rk}) &= \int_0^1 (x - \mu_{\psi}(I_{rk}))^2 dF_{\psi|I}(x|I_{rk}) \\ \text{and } \varphi_{\psi}(t; I_{rk}) &= \int_0^1 e^{tx} dF_{\psi|I}(x|I_{rk}). \end{aligned}$$

If the specification of the distribution of the random variable ψ_{ri} is close to what happens in real life, then this model should be appropriate for most insurance applications. If, however, ψ_{ri} is only a rough approximation of the (distribution of the) proportion of damage caused, then one might prefer to use the accelerated failure time or proportional hazards models below.

Depending on the application, \mathcal{I} may be continuous (e.g., maximum 10-minute average wind speed of a hurricane) or discrete (e.g., intensity of an earthquake on the Modified Mercali scale).

Accelerated failure time

Here we assume that the claim amount due to a catastrophe is a random variable. The effect of the catastrophe intensity is to shift the scale of the distribution. Let $S_{ri}(x) = P[C_{rik}(I_0) >$

$x]$, where I_0 is a “baseline intensity” that makes $h[\psi_{ri}(I_0)] = 1$, where $h : [0, 1] \rightarrow [0, \infty)$ is an increasing function that maps the proportion of destruction to $[0, \infty)$. Then we propose the model

$$P[C_{rik}(I_{rk}) > x | I_{rk}] = S_{ri} [xh[\psi_{ri}(I_{rk})]]. \quad (3)$$

Equation (4) is the *accelerated failure time model* that is commonly used in survival analysis (see Lawless, 1982, Chapter 6 for instance). This model is equivalent to letting $C_{rik}(I_{rk}) = \{1/h[\psi(I_{rk})]\}Y_{ri}$, where Y_{ri} has survivor function S_{ri} . If we let $\mu_{ri} = \int_0^\infty S_{ri}(x)dx$, $\sigma_{ri}^2 = -\int_0^\infty (x - \mu_{ri})^2 dS_{ri}(x)$, and $\varphi_{ri}(t) = -\int_0^\infty e^{tx} dS_{ri}(x)$, we get

$$\begin{aligned} E[C_{rik}(I_{rk}) | I_{rk}] &= \frac{1}{h[\psi_{ri}(I_{rk})]} \mu_{ri} \\ \text{Var}[C_{rik}(I_{rk}) | I_{rk}] &= \left(\frac{1}{h[\psi_{ri}(I_{rk})]} \right)^2 \sigma_{ri}^2 \\ E[e^{tC_{rik}(I_{rk})} | I_{rk}] &= \varphi_{ri} \left(\frac{t}{h[\psi(I_{rk})]} \right). \end{aligned}$$

Proportional hazards

This model is similar in concept to the accelerated failure time model. The difference is that the catastrophe intensity will not change the scale of the distribution but rather distort the survivor function (see Wang, 1998). The model is given by

$$P[C_{rik}(I_{rk}) > x | I_{rk}] = [S_{ri}(x)]^{h[\psi_{ri}(I_{rk})]}, \quad (4)$$

where S_{ri} , ψ_{ri} , and h keep the same definitions as in the accelerated failure time model. One can obtain the mean, variance and characteristic function of $C_{rik}(I_{rk})$ with this model directly from equation (5). Unfortunately, the expressions do not simplify as nicely as they did with the accelerated failure time model.

2.2.2 Model for I_{rk}

Suppose that for each catastrophe that happens, d of its characteristics are needed in order to be able to compute the intensity felt at each area of the portfolio. For example, if we are considering hurricanes, the important characteristics are central pressure or maximum winds, radius of maximum winds, forward speed, track angle and point of entry (see Clark, 1986, Kaplan and DeMaria 1995, 2001, Walters and Morin, 1996, or Huang et al, 2001). Let $\boldsymbol{\theta}_k$, with $\dim(\boldsymbol{\theta}_k) = d$, be the vector containing the important characteristics of the k th

catastrophe. Then we suppose that there exist functions $\phi_r : \mathbb{R}^d \rightarrow \mathcal{I}$ such that $I_{rk} = \phi_r(\boldsymbol{\theta}_k)$, $r = 1, \dots, R$. The functions ϕ_r are sometimes referred to as *attenuation functions*.

2.2.3 Attenuation functions $\phi_r(\cdot)$

The attenuation function is the function that maps the characteristics of the k th catastrophe, $\boldsymbol{\theta}_k$, onto a measure of the intensity felt in area r for that catastrophe, I_{rk} . There is a rich literature on attenuation functions for many natural phenomena (hurricanes, earthquakes, floods, etc.). In this section, we give an example of attenuation function for hurricanes. Information for other types of catastrophe can be found in the geology, hydrology, etc. literature.

For hurricanes: For insurance purposes, the intensity of hurricanes is most conveniently expressed in terms of maximum one-minute sustained wind speed or maximum 10 minute average wind speed. The National Hurricane Center and Casson and Coles (2000) suggest using the simple model of Kaplan and DeMaria (1995, 2001) in order to compute how hurricanes lose their intensity once they make landfall.

In the case of hurricanes, $\boldsymbol{\theta}^T = (x, y, a, \rho, V_0, s)$, where x and y represent the latitude and longitude of the point of landfall, a is the direction of travel of the hurricane, ρ is the radius of maximum winds, V_0 is the speed of maximum one-minute sustained wind at landfall and s is the forward speed of the hurricane. From the information given in $\boldsymbol{\theta}$, one can compute the speed of the wind felt in area r as a function of time.

Since V_0 is the maximum 1-minute sustained wind speed found somewhere within the entire hurricane wind field, we need a formula to compute the wind speed at a given point in the wind field. If r is the distance in km from the point of interest to the center of the storm, θ is the angle measured clockwise from a perpendicular to the right of the storm's trajectory to the line joining the center of the hurricane and the point of interest (see Figure 1), and $\alpha = 0.4$ is some constant that varies from storm to storm (lower values mean stronger winds, but $V_0(r, \theta)$ is not much sensitive to lowering the value of α), the wind speed in knots at the point of interest is

$$V_0(r, \theta) = s \cos \theta + (V_0 - s) \left(\frac{r}{\rho} \right) \exp \left\{ \frac{1}{\alpha} \left[1 - \left(\frac{r}{\rho} \right)^\alpha \right] \right\}, \quad r > \rho. \quad (5)$$

If $r \leq \rho$, we use (8) with $r = \rho$. To compute how V_0 evolves as a function of time since landfall, Kaplan and DeMaria (1995, 2001) propose the use of

$$\phi_r(\boldsymbol{\theta}) = \begin{cases} 26.7 + (0.9V_0 - 26.7)e^{-0.095t_r}, & \text{south of 40 degrees of latitude} \\ 29.6 + (0.9V_0 - 29.6)e^{-0.187t_r}, & \text{north of 40 degrees of latitude,} \end{cases} \quad (6)$$

where V_0 and ϕ_r are wind speeds in knots and t_r is the time since landfall in hours. The idea now consists in using (8) in conjunction with (??) in order to calculate the wind speed at the point of interest at several times during the storm, recording all the wind speed values, moving a 10-minute moving average over all these values, and keeping the highest 10-minute average wind speed as our I_{rk} .

Place Figure 1 about here.

2.2.4 Damage function $\psi_{ri}(\cdot)$

The damage function ψ_{ri} gives the proportion of destruction to building i in region r given the intensity felt in the region. This function may be either deterministic or stochastic.

For hurricanes: Unanwa et al. (2000) propose a general algorithm to compute the amount of destruction to a building given wind speed. Note that this damage function can be used for other types of wind storms as well. In their Figures 11-14, they present upper and lower confidence curves for the proportion of damage as a function of wind speed for each of 4 types of constructions. These curves represent quantiles of the distribution of what we term ψ_{ri} in this paper.

Huang et al (2001) give a formula for the average proportion of damage caused by a maximum average 10 minutes wind speed of x in m/s (1 knot = 0.5148 m/s):

$$\mu_\psi[x] = \begin{cases} 0.01 \exp(0.252x - 5.823), & x \leq 41.4 \\ 1, & x > 41.4. \end{cases} \quad (7)$$

We can also obtain a rough estimate of the conditional variance of the proportion of destruction given wind speed x by looking at their Figure 3:

$$\sigma_{\psi}^2[x] = \begin{cases} 0, & x < 25 \\ 0.000464, & 25 \leq x < 35 \\ 0.02567, & 35 \leq x < 41.4 \\ 0, & x \geq 41.4. \end{cases} \quad (8)$$

Note, however, that (9) and (10) were obtained from insurance data collected in South Carolina and Florida and are not a function of building type. Hence, these damage functions may have to be adapted if one wishes to compute hurricane premiums for houses in northern states.

2.2.5 Models for Θ and M_0

We now study models to generate natural catastrophes and their properties. We consider two approaches. In the first approach, we propose models for the distribution of M_0 and of Θ . These models are mostly based on catalogues of events available on the World Wide Web.

Following the assumptions, we assume that given M_0 , the catastrophes $\Theta_1, \dots, \Theta_{M_0}$ are independent and identically distributed. This assumption is not quite realistic for earthquakes, as several small aftershocks tend to follow a large event, but these assumptions are not unreasonable if we only consider major earthquakes that cause important losses to a specific region. For hurricanes, these assumptions seem to be reasonable.

For hurricanes: The characteristics of past hurricanes (location, time, intensity, etc.) have all been recorded and are freely available from the Internet (for example at

<http://weather.unisys.com/hurricane/>).

These data can be used to model the distribution of M_0 and Θ .

In the insurance literature about hurricanes (e.g., Clark, 1986), it is argued that the distribution of M_0 , the number of hurricanes to hit the US. main land in a year, is well approximated by a negative binomial distribution. A simpler Bernoulli distribution may be appropriate if M_0 represents the number of hurricanes to make landfall in a given state (except perhaps for a few states such as Florida or North Carolina, where getting more than one hurricane hit per year has non-negligible probability).

The distribution of the vector of hurricane characteristics, Θ , is not as easy to describe, given the complex dependence structure of all of its elements. Huang et al. (2001) give a table summarizing the distribution of the important hurricane characteristics for four Southeastern

US states. Darling (1991) proposes a very elaborate empirical model to estimate probabilities of hurricane winds at virtually any location in the Atlantic basin..

A simpler approach often convenient to get moment estimates is to sample hurricanes from the historical database provided by the NHC. If we let each of the H hurricanes to make landfall on record be described by $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_H$, then the distribution of Θ is simply given by

$$P[\Theta = \boldsymbol{x}] = \begin{cases} \frac{1}{H}, & \boldsymbol{x} = \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_H \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

For instance, if we use the available data up to 2000, $H = 164$. It is easily seen that (11) yields

$$E_{\Theta_1}[g(\Theta_1)] = \frac{1}{H} \sum_{h=1}^H g(\boldsymbol{\theta}_h) \equiv \bar{g}; \quad Var_{\Theta_1}[g(\Theta_1)] = \frac{1}{H} \sum_{h=1}^H (g(\boldsymbol{\theta}_h) - \bar{g})^2$$

for any function $g(\cdot)$. To simplify some of the analytical calculations, we can consider only a subset of the hurricanes on record. In the example of Section 2.4, we suppose that four “typical” hurricanes can hit Long Island, NY, even though the historical records contain at least 7 hurricane hits.

2.3 Properties of the general catastrophe model

We now give a list of distributional properties for S and X_{ri} of the GCM under model (2); properties under models (4) and (5) can be obtained in a similar fashion. All our results in this section are given in terms of (properties of) the distribution of the catastrophe characteristics Θ . We prove these results in the Appendix.

RESULT 2.1 (Expectation)

$$E[X_{ri}] = \mu_{Y_{ri}} + E[M_0] E_{\Theta_1}[\mu_{ri}(\Theta_1)],$$

where $\mu_{Y_{ri}} = E[Y_{ri}]$ and

$$\mu_{ri}(\boldsymbol{\theta}_1) = E[C_{ri1}(I_{r1})|\Theta_1 = \boldsymbol{\theta}_1] = b_{ri}\mu_{\psi}(\phi_r(\boldsymbol{\theta}_1)). \quad (10)$$

$$E[S] = \sum_{r=1}^R \sum_{i=1}^{n_r} E[X_{ri}] = \sum_{r=1}^R \sum_{i=1}^{n_r} \mu_{Y_{ri}} + E[M_0] \sum_{r=1}^R \sum_{i=1}^{n_r} E_{\Theta_1}[\mu_{ri}(\Theta_1)].$$

RESULT 2.2 (Variance and covariances)

$$\begin{aligned} \text{Var}[X_{ri}] &= \sigma_{Y_{ri}}^2 + E[M_0] \left\{ E_{\Theta_1} [\sigma_{ri}^2(\Theta_1)] + \text{Var}_{\Theta_1} [\mu_{ri}(\Theta_1)] \right\} \\ &\quad + \text{Var}[M_0] \left\{ E_{\Theta_1} [\mu_{ri}(\Theta_1)] \right\}^2, \end{aligned} \quad (11)$$

where $\sigma_{Y_{ri}}^2 = \text{Var}[Y_{ri}]$, $\mu_{ri}(\Theta_1)$ is as in (12) and

$$\sigma_{ri}^2(\theta_1) = \text{Var} [C_{ri1}(I_{r1}) | \Theta_1 = \theta_1] = b_{ri}^2 \sigma_{\psi}^2(\phi_r(\theta_1)).$$

$$\begin{aligned} \text{Cov}(X_{ri}, X_{sj}) &= E[M_0^2] E_{\Theta_1} [\mu_{ri,sj}(\Theta_1)] \\ &\quad - (E[M_0])^2 E_{\Theta_1} [\mu_{ri}(\Theta_1)] E_{\Theta_1} [\mu_{sj}(\Theta_1)], \end{aligned} \quad (12)$$

where

$$\begin{aligned} \mu_{ri,sj}(\theta_1) &= E [C_{ri1}(I_{r1}) C_{sj1}(I_{s1}) | \Theta_1 = \theta_1] \\ &= b_{ri} b_{sj} E[\psi_{ri}(\phi_r(\theta_1)) \psi_{sj}(\phi_s(\theta_1)) | \Theta_1 = \theta_1]. \end{aligned}$$

$$\text{Var}[S] = \sum_{r=1}^R \sum_{i=1}^{n_r} \text{Var}[X_{ri}] + \sum_{r=1}^R \sum_{s=1}^R \sum_{i=1, i \neq sj}^{n_r} \sum_{j=1}^{n_s} \text{Cov}(X_{ri}, X_{sj}),$$

with $\text{Var}[X_{ri}]$ given by (14) and $\text{Cov}(X_{ri}, X_{sj})$ given by (16).

RESULT 2.3 (Characteristic function) Let $\iota = \sqrt{-1}$. Then

$$\varphi_{X_{ri}}(t) = E [\exp(itX_{ri})] = \varphi_{Y_{ri}}(t) \mathcal{M}_{M_0} \left(\ln E_{\Theta_1} [\varphi_{ri}(t; \Theta_1)] \right),$$

where $\varphi_{Y_{ri}}(t) = E[\exp(itY_{ri})]$, $\mathcal{M}_{M_0}(t) = E[\exp(tM_0)]$ and

$$\varphi_{ri}(t; \theta_1) = E [\exp(itC_{ri1}(I_{r1})) | \Theta_1 = \theta_1] = \varphi_{\psi}(tb_{ri}; \phi_r(\theta_1)).$$

$$\begin{aligned} \varphi_S(t) &= E[\exp(itS)] \\ &= \left\{ \prod_{r=1}^R \prod_{i=1}^{n_r} \varphi_{Y_{ri}}(t) \right\} \mathcal{M}_{M_0} \left(\ln E_{\Theta_1} \left[\prod_{r=1}^R \prod_{i=1}^{n_r} \varphi_{ri}(t; \Theta_1) \right] \right). \end{aligned}$$

Result 2.3 is probably the most important of this section, since it is usually straightforward to compute probabilities, stop-loss premiums, etc. from the characteristic function using numerical techniques such as the FFT (Klugman et al, 1998), for example.

Note that the expressions in Results 2.1-2.3 are relatively easy to evaluate numerically when the distribution of Θ is of the form (11). We now illustrate this fact and how actuarial calculations can be performed with the GCM.

2.4 An example

We consider home insurance on Long Island, NY, particularly in the two eastern counties of Nassau and Suffolk. We chose this region because of its relatively quiet, yet non-negligible, hurricane history (7 landfalls since 1851) that allows us to illustrate the methods without too much of a computational burden. Our first risk is in the county of Nassau, in zip code 11554, at latitude and longitude (40.7176, -73.564), and our second risk is in the county of Suffolk, near the shore line in zip code 11951, at latitude and longitude (40.7661, -72.84). Figure 2 shows the counties of the state of New York, with Nassau and Suffolk being the two easternmost counties on the island. The 7 straight lines represent the paths of the hurricanes that have made landfall on Long Island since 1850, and the two large dots are the locations of the two risks in our portfolio. We segmented the south shore of Long Islands into four segments (see Fig. 2), and our four “typical” storms make landfall in the center of each segment with probabilities $2/7$, $1/7$, $3/7$, $2/7$, respectively from west to east. For each storm, we suppose that $s = 40$ kt, $\rho = 70$ km, and $V_0 = 85$ kt (these numbers are representative of the 7 hurricanes that have made landfall on Long Island). We use the attenuation formulas (??) and (8) and the damage function (9). Note that we will obtain pretty conservative number, as the houses on Long Island are probably of a more solid construction than the homes that entered the calculations of Huang et al. (2001) to yield (9).

Place Figure 2 about here.

Place Figure 3 about here.

Now the idea is to get the maximal 10-minute average of maximum sustained wind speed for each risk given each of the four typical storms. To do so, we have to compute the storm center location and the one-minute wind speed at the risk location over a fine grid of times (see Figure 3). Then we use a ten-minute moving window and find the maximum 10-minute average wind speed. This is easily done when formulas for spherical geometry are available. These formulas are implemented in the S-Plus code available freely on the web site given in the Appendix. Applying all these formulas, we get the results of Table 1.

Put Table 1 about here.

One striking result from Table 1 is that for the quickly moving storms of the Northeast, points on the right hand side of the storms' trajectories get hit a lot harder. Using a Bernoulli q , with $q = 7/151 \approx 0.045$ as the probability that a hurricane will hit Long Island in a given year, and Results 2.1 and 2.2, we get that the expected value and variance of the losses due to hurricanes for risk NY 11554 (with insured value b_1) are $0.00012b_1$ and $0.000001b_1^2$,

respectively. For a home worth \$200,000, this means a hurricane premium of about \$24 and a standard deviation in annual losses due to hurricanes of about \$200. For risk NY 11951 (of insured value b_2), we get $0.0069b_2$ and $0.0178b_2^2$ as our mean and variance, respectively. For a home worth \$200,000, this means a hurricane premium of \$1,377 with an annual standard deviation of yearly hurricane losses of about \$26,700. Again, we emphasize that these numbers are somewhat conservative, given that the entries of columns μ_ψ and σ_ψ^2 in Table 1 were computed from insurance data obtained in Florida and South Carolina. The covariance between the two risks is $0.00003b_1b_2$, for a correlation coefficient of 0.226. As expected, there is a non-negligible positive correlation between the two risks.

For states that are hit very often by hurricanes, such as Florida or North Carolina, the calculations above become quite tedious. However for these states, companies with a few decades of insurance data may prefer to use the spatial approach described in the next section.

3 Spatial individual risk model

We now take a different approach in modeling the spatial dependence between claim amounts. Rather than modeling the “mechanics” of catastrophes and their effect, we model directly the joint distribution of the claim amounts. Since in home insurance a major source of dependence between claim amounts is geographic proximity, we propose a version of the GCM based on models for spatial (irregular lattice) data found in the literature and summarized in Cressie (1993, Chapters 6 and 7). An attractive feature of this approach is that parameters can be estimated directly from catastrophe insurance data, without having to recourse to meteorological or geological models. However only larger insurance companies doing business in areas where catastrophes are frequent (e.g. Southeastern US. for hurricanes) are likely to have access to enough data; smaller companies may therefore find the original GCM from Section 2.2 more useful.

For the sake of clarity, let us consider a simplified version of our GCM where we can only observe at most one catastrophe per period (we make this assumption for clarity and mathematical convenience, but this is not a strong assumption because the probability of two catastrophes hitting the same region (zip code) twice in the same year is usually small):

$$S = \sum_{r=1}^R \sum_{i=1}^{n_r} Y_{ri} + \sum_{r=1}^R \sum_{i=1}^{n_r} B_{r,i},$$

where

$$B_{r,i} = \begin{cases} b_{r,i}\psi_{ri}, & J_r = 1 \\ 0, & J_r = 0, \end{cases}$$

where this time J_r is the random variable indicating whether there was a catastrophe affecting region r and $\psi_{r,i}$ is the proportion of destruction of risk i in region r due to a catastrophe. We still assume that the $\psi_{r,i}$'s are independent of the J_r 's and that the $Y_{r,i}$'s are the "usual business" claim amounts and are independent of each other and of any other random variable.

Our main interest here is to obtain a joint distribution for the $\psi_{r,i}$'s and one for the J_r 's. Ideally, we would like that the parameters of these joint distributions be estimable from insurance data. The two models that we consider here are an auto-logistic model for the joint distribution of J_1, \dots, J_R and an auto-gaussian model for $Z_{r,i} = g(\psi_{r,i})$, where $g(\cdot)$ is an $\mathbb{R} \rightarrow [0, 1]$ differentiable and strictly increasing link function. A common choice for g is $g(x) = e^x/(1 + e^x)$. Before we give a description of these two models, we need some notation. Let $(x_{r,i}, y_{r,i})$ represent the latitude and longitude of risk i in region r , and let (x_r, y_r) be the latitude and longitude of the center of region r . Let $d((r, i), (s, j)) = ((x_{r,i} - x_{s,j})^2 + (y_{r,i} - y_{s,j})^2)^{1/2}$ be the Euclidean distance between risk i in region r and risk j in region s , and let $d^*(r, s) = ((x_r - x_s)^2 + (y_r - y_s)^2)^{1/2}$ be the distance between the centers of regions r and s . Let $N_{r,i} = \{(s, j) : d((r, i), (s, j)) \leq d_0\}$ and $N_r^* = \{s : d^*(r, s) \leq d_0^*\}$ be the neighborhoods of risk i in region r and region r , respectively, i.e. we define the neighborhood of a given risk (resp. region) as the set of risks (resp. regions) that are within a certain fixed distance d_0 (resp. d_0^*). Cressie (1993) suggests a few statistical methods to choose appropriate values for d_0 and d_0^* ; these values may or may not be the same. Based on the studies of Kaplan and DeMaria (1995) and Huang et al. (2001), using $d_0 = d_0^* = 70$ km seems to be a reasonable choice.

3.1 Auto-logistic model for the J_r 's

The auto-logistic model for J_1, \dots, J_R is obtained by specifying the following conditional probability:

$$P[J_r = u_r | \{J_j, j \neq r\}] = \frac{\exp\{\alpha_r u_r + \sum_{j=1}^R \theta_{r,j} J_j\}}{1 + \exp\{\alpha_r + \sum_{j=1}^R \theta_{r,j} J_j\}}, \quad u_r = 0, 1, \quad r = 1, \dots, R, \quad (13)$$

where α_r represents the large-scale variation (trend) and $\theta_{r,j}$ the small-scale variation (spatial dependence). Cressie (1993) proposes various ways of choosing and specifying α_r and $\theta_{r,j}$ from the data. Possible simple forms sensible for catastrophe insurance applications are

$\alpha_r = \mu + \beta_1 x_r + \beta_2 y_r + \beta_3 x_r y_r$ (a trend in latitude, longitude, and a possible interaction between latitude and longitude), and

$$\theta_{r,j} = \begin{cases} 0, & j \notin N_r^* \text{ or } j = r \\ \gamma \frac{d^*(r,j)^{-k}}{\max\{d^*(r,j)^{-k}, j \in N_r^*\}}, & j \in N_r^* \text{ and } j \neq r, \end{cases}$$

with $k = 0, 1$ or 2 (1 is a sensible choice for our purposes). The parameters γ, β_1, β_2 and β_3 are estimated from data using a pseudolikelihood method based on a simplification of the joint probability mass function of J_1, \dots, J_R :

$$P[J_1 = u_1, \dots, J_R = u_r] = \exp \left\{ \sum_{r=1}^R \alpha_r u_r + \sum_{1 \leq r < j \leq R} \theta_{r,j} u_r u_j \right\} \\ \times \left[\sum_{(v_1, \dots, v_R) \in \{0,1\}^R} \exp \left\{ \sum_{r=1}^R \alpha_r v_r + \sum_{1 \leq r < j \leq R} \theta_{r,j} v_r v_j \right\} \right]^{-1}. \quad (14)$$

The pseudolikelihood method is described in Cressie (1993, Section 7.2). From (25), we can easily compute probabilities of the form $q_r = P[J_r = 1]$ and $q_{rs} = P[J_r = 1, J_s = 1]$; these probabilities enter the calculations of the moments of the $B_{r,i}$'s.

Statistical packages such as S-Plus make this model quite tractable. Code for the pseudolikelihood estimation of the parameters is available from various sources, and the matrix algebra capabilities of such software make calculations involving the joint distribution (25) straightforward.

To simulate data from this auto-logistic model, one can use the Gibbs Sampler algorithm based on (23) given by Cressie (1993, Section 7.7.1).

3.2 Auto-gaussian model for the transformed $\psi_{r,i}$'s

Now for the proportions of damage, let $\boldsymbol{\psi} = (\psi_{1,1}, \psi_{1,2}, \dots, \psi_{1,n_1}, \psi_{2,1}, \dots, \psi_{R,n_R})^T$ and let $K = \sum_{r=1}^R n_r$. The auto-gaussian model specifies a multivariate normal distribution for $\mathbf{Z} = g(\boldsymbol{\psi})$, i.e.

$$\mathbf{Z} \sim \text{MVN}(\boldsymbol{\mu}, \Sigma),$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} 1 & x_{1,1} & y_{1,1} & x_{1,1}y_{1,1} \\ \vdots & & & \\ 1 & x_{R,n_R} & y_{R,n_R} & x_{R,n_R}y_{R,n_R} \end{pmatrix} \begin{pmatrix} \mu \\ \eta_1 \\ \eta_2 \\ \eta_3 \end{pmatrix}$$

and $\Sigma = (I_{K \times K} - C)^{-1}M$, where $I_{K \times K}$ is the $K \times K$ identity matrix, $M = \tau^2 I_{K \times K}$ with $\tau^2 = \text{Var}[Z_{r,i} | \{Z_{(j,s)}, (j,s) \neq (r,i)\}]$ and C is a $K \times K$ matrix whose element in row k and column l is $c_{k,l}$ with $c_{k,k} = 0$ and $c_{k,l} = c_{l,k}$. Using the fact that the correlation between close neighbors should be positive and higher than between distant neighbors, we can use the following specification of $c_{k,l}$. Let k represent risk i in region r and l represent risk j in region s . Then

$$C_{k,l} = \begin{cases} 0, & (j,s) = (r,i) \text{ or } (j,s) \notin N_{r,i} \\ \phi \frac{d((r,i),(j,s))^{-k}}{\max\{d((r,i),(j,s))^{-k}, (j,s) \in N_{r,i}\}}, & (j,s) \in N_{r,i} \text{ and } (j,s) \neq (r,i), \end{cases}$$

As before, it is sensible to choose $k = 1$ and Cressie (1993) shows how to estimate $\mu, \eta_1, \eta_2, \eta_3, \phi$ and τ by maximum likelihood. For convenience, we let

$$\begin{aligned} \mu_{r,i} &= E[Z_{r,i}] = \mu + \eta_1 x_{r,i} + \eta_2 y_{r,i} + \eta_3 x_{r,i} y_{r,i} \\ \sigma_{r,i}^2 &= \text{Var}[Z_{r,i}] = \Sigma_{(r,i),(r,i)} \\ \sigma_{(r,i),(s,j)} &= \text{Cov}(Z_{r,i}, Z_{s,j}) = \Sigma_{(r,i),(s,j)}. \end{aligned}$$

If we let $h = g^{-1}$ and let h' be its derivative, then we get that the joint density function of $\psi_{1,1}, \dots, \psi_{R,n_R}$ is given by

$$f_{\boldsymbol{\psi}}(\mathbf{w}) = (2\pi)^{K/2} \left| \Sigma^{-1} \right|^{1/2} \exp \left\{ \frac{1}{2} (\mathbf{h}(\mathbf{w}) - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{h}(\mathbf{w}) - \boldsymbol{\mu}) \right\} \left| \prod_{r=1}^R \prod_{i=1}^{n_r} h'(w_{r,i}) \right|, \quad (15)$$

where $\|A\|$ denotes the determinant of a matrix A . To compute the moments of the $\psi_{r,i}$'s, we can use two approaches. One is to use numerical techniques to derive these moments from the joint density (27). Another is to use Taylor series and/or delta-rule approximations. Using this latter approach we get

$$\mu_{r,i}^{\psi} = E[\psi_{r,i}] \approx g(\mu_{r,i}) + g''(\mu_{r,i}) \sigma_{r,i}^2 / 2$$

If we let $g(x) = e^x / (1 + e^x)$, we obtain

$$\mu_{r,i}^{\psi} = \frac{e^{\mu_{r,i}}}{1 + e^{\mu_{r,i}}} + \frac{(e^{\mu_{r,i}} - e^{2\mu_{r,i}}) \sigma_{r,i}^2}{(1 + e^{\mu_{r,i}})^3} \frac{1}{2}.$$

To get the variance-covariance matrix of $\boldsymbol{\psi}$, let $\nabla g = (g'(\mu_{1,1}), \dots, g'(\mu_{R,n_R}))^T$. Then the delta rule yields

$$\Sigma^{\psi} = \text{Var}(\boldsymbol{\psi}) \approx \nabla g^T \Sigma \nabla g.$$

For convenience, we let

$$\begin{aligned}\sigma_{r,i}^{\psi^2} &= \text{Var}[\psi_{r,i}] = \Sigma_{(r,i),(r,i)}^{\psi} = [g'(\mu_{r,i})]^2 \sigma_{r,i}^2 \\ \sigma_{(r,i),(s,j)}^{\psi} &= \text{Cov}(\psi_{r,i}, \psi_{s,j}) = \Sigma_{(r,i),(s,j)}^{\psi}.\end{aligned}$$

There is a simple algorithm to simulate from the distribution of $\boldsymbol{\psi}$:

1. Simulate K independent standard normal variables $\mathbf{Z}^* = (Z_1^*, \dots, Z_{R,n_R}^*)^T$;
2. Using Choleski decomposition, find a matrix A such that $A^T A = \Sigma$;
3. Let $\mathbf{Z} = A\mathbf{Z}^* + \boldsymbol{\mu}$;
4. Let $\boldsymbol{\psi} = g(\mathbf{Z})$.

The characteristic function $\varphi_{\boldsymbol{\psi}}$ can be approximated using the simulation algorithm for \mathbf{Z} given above since

$$\varphi_{\boldsymbol{\psi}}(t_1, \dots, t_K) = E \left[e^{t \sum_{i=1}^K t_i g(Z_i)} \right]. \quad (16)$$

3.3 Loss distribution properties

If we forget about the “usual business” portion of the model, we have that the probability distributions of each risk (r, i) and S are completely specified. We now explore some of the properties of these distributions.

By conditioning on J_r , the net catastrophe premium for risk (r, i) is easily shown to be

$$E[B_{r,i}] = b_{r,i} P[J_r = 1] E[\psi_{r,i}] = b_{r,i} q_r \mu_{r,i}^{\psi}.$$

Using similar arguments, we also have that

$$\text{Var}[B_{r,i}] = b_{r,i}^2 \left\{ \mu_{r,i}^{\psi^2} q_r (1 - q_r) + \sigma_{r,i}^{\psi^2} q_r \right\}$$

and

$$\text{Cov}(B_{r,i}, B_{s,j}) = b_{r,i} b_{s,j} \left\{ \sigma_{(r,i),(s,j)}^{\psi} q_{rs} + \mu_{r,i}^{\psi} \mu_{s,j}^{\psi} (q_{rs} - q_r q_s) \right\}. \quad (17)$$

When $r = s$, (31) simplifies to

$$\text{Cov}(B_{r,i}, B_{r,j}) = b_{r,i} b_{r,j} \left\{ \sigma_{(r,i),(r,j)}^{\psi} q_r + \mu_{r,i}^{\psi} \mu_{r,j}^{\psi} q_r (1 - q_r) \right\}.$$

If we let $\varphi_{\boldsymbol{\psi}}(t_1, \dots, t_K)$ denote the characteristic function of $\boldsymbol{\psi}$, $\varphi_{Y_{r,i}}(t)$ denote the characteristic function of $Y_{r,i}$ and p_{i_1, \dots, i_R} denote $P[J_1 = i_1, \dots, J_R = i_R]$, then we get that the characteristic function of S is given by

$$E[e^{tS}] = \left\{ \prod_{r=1}^R \prod_{i=1}^{n_R} \varphi_{Y_{r,i}}(t) \right\} \times \left\{ \sum_{(i_1, \dots, i_R) \in \{0,1\}^R} \varphi_{\boldsymbol{\psi}}(tb_{1,1}i_1, tb_{1,2}i_2, \dots, tb_{R,n_R}i_R) p_{i_1, \dots, i_R} \right\}.$$

We can also compute the characteristic function of S at t via the following algorithm:

1. Set $J_1 = i_1, \dots, J_R = i_R$.
2. Compute $t_1 = tb_{1,1}i_1, \dots, t_K = tb_{R,n_R}i_R$.
3. Simulate several realizations of \boldsymbol{Z} .
4. For each simulated realization of \boldsymbol{Z} , use (28) to compute $\varphi_{\boldsymbol{\psi}}(t_1, \dots, t_K)$ and average out the results over all realizations.
5. Multiply this average by p_{i_1, \dots, i_R} .
6. Repeat these steps for every possible values of (i_1, \dots, i_R) and sum up the values obtained in step 5 to obtain the desired result.

4 Concluding remarks

We have proposed two general versions of the individual risk model useful for actuarial modeling when catastrophes such as hurricanes may be present. The first model is a mathematical formalization of the simulation algorithms described in the actuarial literature. This model is most useful to actuaries who do not have access to several decades of catastrophe insurance data and is particularly tractable when the number of possible “typical catastrophes” that can hit the portfolio is small.

The second model is a spatial formulation of the individual risk model that takes into account the correlation induced by geographic proximity between the risks. The attractive feature of this model is that if enough data are available, there is no need to specify the attenuation and damage functions that, in general, require some meteorological or geological and engineering knowledge.

Acknowledgments

The three authors are grateful to the Actuarial Education and Research Fund and the Casualty Actuarial Society for their financial support of this project.

Appendix A: Derivation of Results 2.1-2.3

This Appendix contains outlines of the proofs of Results 2.1, 2.2, and 2.3.

Result 2.1

$$\begin{aligned}
 E[S] &= E \left[\sum_{r=1}^R \sum_{i=1}^{n_r} \{Y_{ri} + \sum_{k=1}^{M_0} C_{rik}(I_{rk})\} \right] \\
 &= \sum_{r=1}^R \sum_{i=1}^{n_r} E[Y_{ri}] + E_{M_0} \left[E \left[\sum_{r=1}^R \sum_{i=1}^{n_r} \sum_{k=1}^{M_0} C_{rik}(I_{rk}) \middle| M_0 \right] \right] \\
 &= \sum_{r=1}^R \sum_{i=1}^{n_r} \mu_{Y_{ri}} + E[M_0] \sum_{r=1}^R \sum_{i=1}^{n_r} E[C_{ri1}(I_{r1})] \\
 &= \sum_{r=1}^R \sum_{i=1}^{n_r} \mu_{Y_{ri}} + E[M_0] \sum_{r=1}^R \sum_{i=1}^{n_r} E_{\Theta}[\mu_{ri}(\Theta)].
 \end{aligned}$$

Result 2.2

$$Var[S] = Var \left[\sum_{r=1}^R \sum_{i=1}^{n_r} X_{ri} \right] = \sum_{r=1}^R \sum_{s=1}^R \sum_{i=1}^{n_r} \sum_{j=1}^{n_s} Cov(X_{ri}, X_{sj}),$$

where $Cov(X_{ri}, X_{ri}) \equiv Var[X_{ri}]$, and

$$\begin{aligned}
 Var[X_{ri}] &= Var[Y_{ri}] + Var \left[\sum_{k=1}^{M_0} C_{rik}(I_{rk}) \right] \\
 &= \sigma_{Y_{ri}}^2 + E_{M_0} \left[Var \left[\sum_{k=1}^{M_0} C_{rik}(I_{rk}) \middle| M_0 \right] \right]
 \end{aligned}$$

$$\begin{aligned}
& +Var_{M_0} \left[E \left[\sum_{k=1}^{M_0} C_{rik}(I_{rk}) \middle| M_0 \right] \right] \\
= & \sigma_{Y_{ri}}^2 + E[M_0]Var[C_{ri1}(I_{r1})] \\
& +Var[M_0] \{E[C_{ri1}(I_{r1})]\}^2 \\
= & \sigma_{Y_{ri}}^2 + E[M_0] \{E_{\Theta}[\sigma_{ri}^2(\Theta)] + Var_{\Theta}[\mu_{ri}(\Theta)]\} \\
& +Var[M_0] \{E_{\Theta}[\mu_{ri}(\Theta)]\}^2
\end{aligned}$$

and

$$\begin{aligned}
Cov(X_{ri}, X_{sj}) & = E \left[\sum_{k=1}^{M_0} C_{rik}(I_{rk}) \sum_{l=1}^{M_0} C_{sjl}(I_{jl}) \right] \\
& \quad - E \left[\sum_{k=1}^{M_0} C_{rik}(I_{rk}) \right] E \left[\sum_{k=1}^{M_0} C_{sjk}(I_{sk}) \right] \\
= & E_{M_0} \left[E \left[\sum_{k=1}^{M_0} C_{rik}(I_{rk}) \sum_{l=1}^{M_0} C_{sjl}(I_{jl}) \middle| M_0 \right] \right] \\
& \quad - E_{M_0} \left[E \left[\sum_{k=1}^{M_0} C_{rik}(I_{rk}) \middle| M_0 \right] \right] E_{M_0} \left[E \left[\sum_{k=1}^{M_0} C_{sjk}(I_{sk}) \middle| M_0 \right] \right] \\
= & E[M_0^2]E[C_{ri1}(I_{r1})C_{sj1}(I_{s1})] \\
& \quad - (E[M_0])^2 E[C_{ri1}(I_{r1})] E[C_{sj1}(I_{s1})] \\
= & E[M_0^2]E_{\Theta}[\mu_{ri,sj}(\Theta)] - (E[M_0])^2 E_{\Theta}[\mu_{ri}(\Theta)]E_{\Theta}[\mu_{sj}(\Theta)].
\end{aligned}$$

Result 2.3

$$\begin{aligned}
\varphi_S(t) & = E[\exp(itS)] = E \left[\exp \left(\sum_{r=1}^R \sum_{i=1}^{n_r} Y_{ri} + \sum_{k=1}^{M_0} \sum_{r=1}^R \sum_{i=1}^{n_r} C_{rik}(I_{rk}) \right) \right] \\
= & E \left[\exp \left(it \sum_{r=1}^R \sum_{i=1}^{n_r} Y_{ri} \right) \right] E \left[\exp \left(it \sum_{k=1}^{M_0} \sum_{r=1}^R \sum_{i=1}^{n_r} C_{rik}(I_{rk}) \right) \right] \\
= & \left\{ \prod_{r=1}^R \prod_{i=1}^{n_r} \varphi_{Y_{ri}}(t) \right\} E_{M_0} \left[E \left[\exp \left(it \sum_{k=1}^{M_0} \sum_{r=1}^R \sum_{i=1}^{n_r} C_{rik}(I_{rk}) \right) \middle| M_0 \right] \right]
\end{aligned}$$

$$\begin{aligned}
&= \left\{ \dots \right\} E_{M_0} \left[\left(E \left[\exp \left(ut \sum_{r=1}^R \sum_{i=1}^{n_r} C_{ri1}(I_{r1}) \right) \right] \right)^{M_0} \right] \\
&= \left\{ \dots \right\} E_{M_0} \left[\left(E_{\Theta_1} \left[E \left[\exp \left(ut \sum_{r=1}^R \sum_{i=1}^{n_r} C_{ri1}(\psi_{ri}(\phi_r(\Theta_1))) \right) \middle| \Theta_1 \right] \right] \right)^{M_0} \right] \\
&= \left\{ \dots \right\} E_{M_0} \left[\left(E_{\Theta_1} \left[\prod_{r=1}^R \prod_{i=1}^{n_r} \varphi_{ri}(t; \Theta_1) \right] \right)^{M_0} \right] \\
&= \left\{ \dots \right\} E_{M_0} \left[\exp \left(M_0 \left\{ \ln E_{\Theta_1} \left[\prod_{r=1}^R \prod_{i=1}^{n_r} \varphi_{ri}(t; \Theta_1) \right] \right\} \right) \right] \\
&= \left\{ \dots \right\} \mathcal{M}_{M_0} \left(\ln E_{\Theta_1} \left[\prod_{r=1}^R \prod_{i=1}^{n_r} \varphi_{ri}(t; \Theta_1) \right] \right),
\end{aligned}$$

where

$$\left\{ \dots \right\} = \left\{ \prod_{r=1}^R \prod_{i=1}^{n_r} \varphi_{Y_{ri}}(t) \right\}.$$

Appendix B: S-Plus code

The S-Plus functions used for the Long Island hurricane insurance example are available at <http://www.utstat.toronto.edu/~duchesne/hurricane/>.

There are a few software packages available on the Internet that can fit the auto-logistic model and the auto-gaussian. For the auto-logistic, free software for Bayesian inference is available at

<http://www.stat.colostate.edu/~jah/software/jahsoftware.html>.

For the auto-gaussian, a package that can be used with the free software R is available at

<http://www.geo.uni-bayreuth.de/~martin>.

References

Beirlant, J. and Teugels, J.L. (1992) "Modeling large claims in non-life insurance", *Insurance: Mathematics and Economics* 11, 17-29.

- Burger, G., Fitzgerald, B.E., White, J. and Woods, P.B. (1996), "Incorporating a hurricane model into property ratemaking", CAS Forums, Winter 1996.
- Campbell, K.W. (1989), "The dependence of peak horizontal acceleration on magnitude, distance, and site effects for small-magnitude earthquakes in California and Eastern North-America", *Bulletin of the Seismological Society of America*, **79**, 1311-1339.
- Casson, E. and Coles, S. (2000), "Simulation of extremal analysis of hurricane events", *Journal of the Royal Statistical Society, Series C*, **49**, 227-245.
- Christensen, C.V. , Schmidli, H. (2000), "Pricing catastrophe insurance products based on actually reported claims", *Insurance: Mathematics and Economics*, **27**, 189-200.
- Clark, K.M. (1986), "A formal approach to catastrophe risk assessment and management", PCAS LXXIII.
- Collins, D.J. and Lowe, S.P. (2001), "A macro validation dataset for US hurricane models", CAS Forums, Winter 2001.
- Cossette, H., Duchesne, T. and Marceau, E. (2000), "A dependent individual risk model based on catastrophes, with estimation". Technical Report 2000-05, École d'Actuariat, Université Laval.
- Cressie, N.A.C. (1993), *Statistics for Spatial Data*, New York: Wiley.
- Darling, R.W.R. (1991), "Estimating probabilities of hurricane wind speeds using a large scale empirical model", *Journal of Climate*, **4**, 1035-1046.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer-Verlag.
- Huang, Z., Rosowsky, D.V. and Sparks, P.R. (2001), "Hurricane simulation techniques for the evaluation of wind-speeds and expected insurance losses", *Journal of Wind Engineering and Industrial Aerodynamics*, **89**, 605-617.
- Kaplan, J. and DeMaria, M. (1995), "A simple empirical model for predicting the decay of tropical cyclone winds after landfall", *Journal of Applied Meteorology*, **34**, 2499-2512.
- Kaplan, J. and DeMaria, M. (2001), "On the decay of tropical cyclone winds after landfall in the New England area", *Journal of Applied Meteorology*, **40**, 280-286.
- Klugman, S.A., Panjer, H.H. and Willmot, G.E. (1998), *Loss Models: From Data to Decision*. New York: Wiley.

Lamarre, M., Townshend, B. and Shah, H. (1992), “Application of the bootstrap method to quantify uncertainty in seismic hazard estimates”, *Bulletin of the Seismological Society of America*, **82**, 104-119.

Unanwa, C.O., McDonald, R.J., Mehta, K.C. and Smith, D.A. (2000), “The development of wind damage bands for buildings”, *Journal of Wind Engineering and Industrial Aerodynamics*, **84**, 119-149.

Walters, M.A. and Morin, F. (1996), “Catastrophe ratemaking revisited (Use of computer models to estimate loss costs)”, CAS Forums, Winter 1996.

Wang, S. (1998) ”Aggregation of correlated risk portfolios: Models and Algorithms”, *Proceedings of the Casualty Actuarial Society*, **vol. LXXXV**, 848-939.

Addresses for correspondence:

Hélène Cossette
École d’Actuariat, Pavillon Vachon
Université Laval
Ste-Foy, QC, G1K 7P4, Canada
hcossett@act.ulaval.ca

Thierry Duchesne
Department of Statistics
University of Toronto
100 St. George, Toronto, ON, M5S 3G3, Canada
duchesne@utstat.toronto.edu

Étienne Marceau
École d’Actuariat, Pavillon Vachon
Université Laval
Ste-Foy, QC, G1K 7P4, Canada
emarceau@act.ulaval.ca

Table 1: Maximum 10-minute average wind speed (in m/s) at risk location $[\phi(\boldsymbol{\theta})]$, expectation and variance of the proportion of destruction $[\mu_\psi(\phi(\boldsymbol{\theta}))]$ and $\sigma_\psi^2(\phi(\boldsymbol{\theta}))$, respectively], and probability of the storm having characteristics $\boldsymbol{\theta}$, given there is a storm.

Risk 1, NY 11554				
$\boldsymbol{\theta}$	$\phi(\boldsymbol{\theta})$	$\mu_\psi(\phi(\boldsymbol{\theta}))$	$\sigma_\psi^2(\phi(\boldsymbol{\theta}))$	$P[\boldsymbol{\Theta} = \boldsymbol{\theta}]$
$\boldsymbol{\theta}_1$	21.74	0.0071	0	2/7
$\boldsymbol{\theta}_2$	14.57	0.0012	0	1/7
$\boldsymbol{\theta}_3$	13.47	0.0009	0	3/7
$\boldsymbol{\theta}_4$	13.13	0.0008	0	1/7
$E_{\boldsymbol{\Theta}}[\cdot]$		0.0027	0	
$Var_{\boldsymbol{\Theta}}[\cdot]$		0.000015		
Risk 2, NY 11951				
$\boldsymbol{\theta}$	$\phi(\boldsymbol{\theta})$	$\mu_\psi(\phi(\boldsymbol{\theta}))$	$\sigma_\psi^2(\phi(\boldsymbol{\theta}))$	$P[\boldsymbol{\Theta} = \boldsymbol{\theta}]$
$\boldsymbol{\theta}_1$	36.60	0.3000	0.02567	2/7
$\boldsymbol{\theta}_2$	38.24	0.4530	0.02567	1/7
$\boldsymbol{\theta}_3$	20.63	0.0054	0	3/7
$\boldsymbol{\theta}_4$	17.24	0.0023	0	1/7
$E_{\boldsymbol{\Theta}}[\cdot]$		0.1530	0.0110	
$Var_{\boldsymbol{\Theta}}[\cdot]$		0.3619		