Journal of Risk Finance, Vol. 7, No. 5, 2006.

EDITORIAL

Catastrophe Forecasting: Seeing "Gray" among the "Black Boxes"

Michael R. Powers*

A catastrophe is, by its very nature, *rare*. In insurance parlance, it can be described as an event whose severity is so far out on the loss distribution that its frequency is necessarily low.[1] One inevitable result of the rarity of catastrophes is the sparseness of relevant data for estimating loss frequencies and severities. For example, over the past half-century, the rate of hurricanes making landfall in the U.S. has been only about two storms per year.[2]

The paucity of catastrophe data leaves insurers, reinsurers, and their risk-assessment experts with a difficult statistical problem: how to make reasonable forecasts of insured catastrophe losses based upon few historical observations. Most catastrophe-risk analysts[3] generate "black-box" forecasts of both loss frequency and loss severity. In other words, they employ highly complex mathematical models – often with subjective components reflecting the judgmental forecasts of catastrophe experts – which, ostensibly because of "proprietary" concerns, remain unpublished and therefore untested by impartial outsiders.

Faithful readers of this journal already know (see Powers, 2005) that I am somewhat skeptical of black-box forecasts. No matter how experienced or intelligent the individuals generating such a forecast, they cannot escape the possibility of gross errors and oversights, some of which may have a significant impact on forecast magnitudes.

This does not mean that all black boxes are bad. I am quite comfortable relying on a vast array of high-tech gadgets – everything from CAT[4] scanners to television sets – about whose inner workings I have little detailed understanding. However, my willingness to rely on such

---

* Editor, *Journal of Risk Finance*; Professor and Director, Advanta Center for Financial Services Studies, The Fox School, Temple University; e-mail: michael.powers@temple.edu.

devices is directly related to the amount of empirical evidence supporting their efficacy. For example, I find the images of current events relayed by television useful because of the medium's long successful track record. But before purchasing a similar device purporting to provide images from the spirit world (such as one of the computer displays in the 2005 motion picture *White Noise*), I would require *extensive* corroboration by *impartial* sources of the seller's claims.

The problem with black-box catastrophe forecasts is the absence of both extensive validation data and impartial peer review. Fortunately, both of these issues may be addressed; the former simply by comparing the black-box forecasts to a naïve alternative, to ensure a basic "reality check," and the latter by insisting that this comparison be done by an independent party. To see how this approach would work, we will now consider some historical forecast data from one of the most publicly discussed (and therefore virtually "transparent") black boxes in the catastrophe field: the hurricane-prediction methodology of Dr. William M. Gray, professor at Colorado State University.

Since 1984, Dr. Gray has employed a complex meteorological and statistical model of global weather patterns to make forecasts of the numbers of (named) tropical storms and hurricanes of various categories originating in the Atlantic Basin.[[5]] Prior to each year's hurricane season, he makes several forecasts, typically in early December, early April, and late May/early June. For the purposes at hand, we will focus on Dr. Gray's May/June predictions of the total number of hurricanes for the 23 years from 1984 through 2006. These forecasts are presented in the third column of Table I below.

Using a term familiar to financial investors, Dr. Gray's forecasts may be said to derive from a rigorous *fundamental* analysis of numerous prevailing meteorological conditions (comparable to the firm characteristics and market conditions considered in a fundamental analysis of stock prices). To offer a simple reality check of the value of these forecasts, we will provide a naïve *technical* analysis of the complete set of historical hurricane frequency data compiled by the National Hurricane Center (see Charts 1a-1c), and use this technical analysis to generate a set of competing forecasts for 1984 through 2006.

Chart 1a. Atlantic Basin Hurricane Frequency (1851-1900)

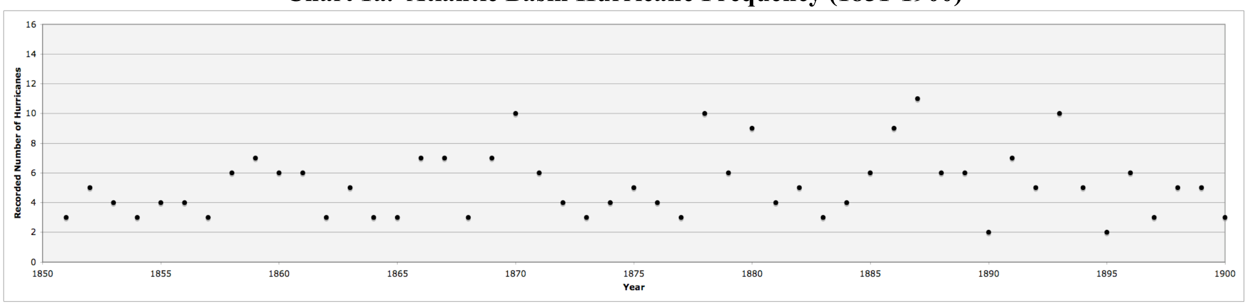


Chart 1b. Atlantic Basin Hurricane Frequency (1901-1950)

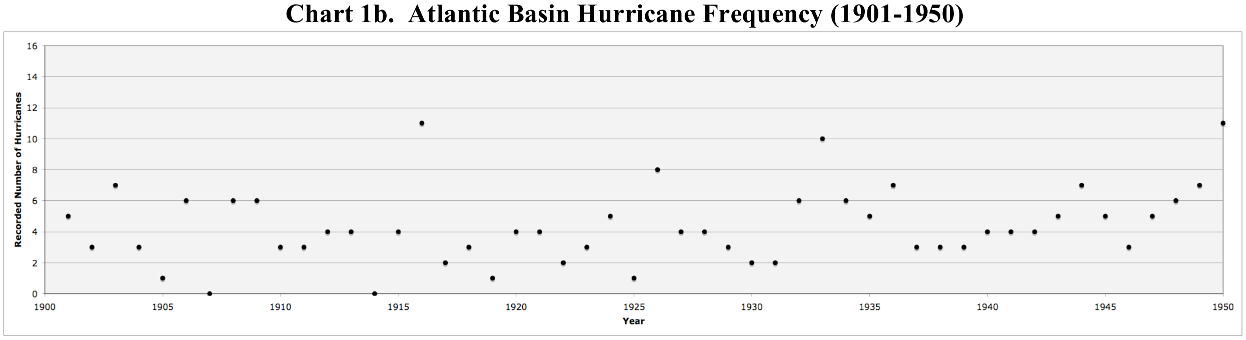
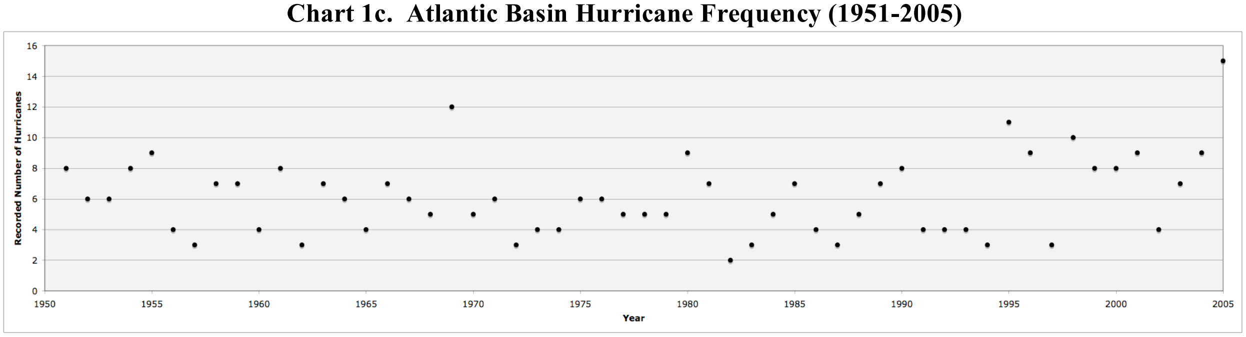


Chart 1c. Atlantic Basin Hurricane Frequency (1951-2005)



Source: National Hurricane Center.

Although it is somewhat difficult to see from Charts 1a-1c, the hurricane frequency series is nonstationary, with a mean (average number of hurricanes per year) that increases slightly, but significantly, over time. A simple linear regression of the data against time provides an estimated increase of 0.01 hurricane per year.[6] Another characteristic of the time series is that it appears to manifest serial correlation, as the plot of the points moves up and down more smoothly than might be expected of sequentially independent observations.

In conducting our competing technical analysis, it is crucial that we place ourselves in the first half of 1984, at about the same time Dr. Gray made his first forecast, and select our model based upon only considerations possible at that time. To this end, we use the Atlantic Basin hurricane data from 1851 to 1983 (i.e., all but the last 22 points in the above charts).

Without getting into the gory details, a fairly straightforward and conventional analysis of autocorrelation functions suggests a first-order moving average model (MA(1)) in the first differences of the original hurricane series. Checking residuals further suggests including a first-order autoregressive component (AR(1)), which ultimately yields the three-parameter ARIMA(1,1,1) model:

$$\left(\Delta_t - \mu\right) - \varphi \cdot \left(\Delta_{t-1} - \mu\right) = \varepsilon_t - \theta \cdot \varepsilon_{t-1},$$

where Δ_t denotes the difference between the number of hurricanes in year t and the number in year $t-1$, and the parameters μ, φ, and θ must be re-estimated as each successive year's number of hurricanes becomes available.

Table I presents a comparison of Dr. Gray's forecasts with those of the author's ARIMA model. For each year, the forecast that is closest to the actual recorded number of hurricanes is indicated by a shaded box; if both forecasts are equally distant from the true value, then neither box is shaded. Setting aside the latter forecasts, we see that out of the 17 years in which one of the forecasts did better than the other, Dr. Gray's forecast is better 12 times and the author's is better only 5 times. Intuitively, this seems to suggest that Dr. Gray's methodology adds some value beyond the simple technical analysis, especially in time periods (e.g., 1998 to 2005) in which the number of hurricanes deviates substantially from its recent average. Interestingly, however, various statistical tests fail to show that Dr. Gray's forecasts are significantly better than the alternative (at the 5 percent level).[7] This is because of the relatively short historical period for comparison – a problem that only the passage of time can remedy.

Table I. Atlantic Basin Hurricane Forecasts

Year	Actual Number	Dr. Gray's Forecast	Dr. Powers' Forecast
1984	5	7	5
1985	7	8	5
1986	4	4	6
1987	3	5	5
1988	5	7	5
1989	7	4	5
1990	8	7	5
1991	4	4	6
1992	4	4	5
1993	4	7	5
1994	3	5	5
1995	11	8	5
1996	9	6	6
1997	3	7	6
1998	10	6	5
1999	8	9	6
2000	8	8	6
2001	9	7	6
2002	4	6	6
2003	7	8	6
2004	9	8	6
2005	15	8	7
2006	**?**	**9**	**8**
% Time Closer to Actual		55%	23%

Sources: National Hurricane Center; Dr. William M. Gray <http://typhoon.atmos.colostate.edu/forecasts/>; and the author's calculations.

Although unable to demonstrate the clear superiority of Dr. Gray's forecasts over those of a simple technical model, the above exercise is clearly useful. Most importantly, it reveals that Dr. Gray's complex fundamental analysis does in fact provide reasonable forecasts, and may indeed offer value beyond a naïve alternative model. As a result, I would argue that, whenever confronted with black-box catastrophe forecasts, all parties of interest (insurers, reinsurers, rating agencies, regulators, and the public at large) should demand that such forecasts be accompanied by an impartial peer review.

Reference

Powers, M. R. (2005), "The terror of the 'black box,'" *Journal of Risk Finance*, Vol. 6, No. 4, pp. 289-291.

---

[1] In fact, Property Claim Services – a division of ISO (the Insurance Services Office) – currently defines a "catastrophe" in the United States as a single event generating at least $25 million in insured property losses.
[2] National Hurricane Center data.
[3] The largest commercial risk analysis firms include Risk Management Solutions (RMS), Applied Insurance Research (AIR), and Eqecat (EQE).
[4] As in "computer assisted tomography," not "catastrophe".
[5] See <http://typhoon.atmos.colostate.edu/forecasts/>.
[6] This is true whether or not the 2005 spike is included. The small estimated increase is clearly relevant in assessing whether or not the *global warming* phenomenon is responsible for increases in hurricane frequency, especially when one observes that the number of recorded hurricanes in earlier years (i.e., prior to the weather satellite era) may have been underreported.
[7] The simplest such test is the nonparametric sign test using only the 17 years for which one forecast is better than the other. Under the null hypothesis, each year's pair of forecasts yields an i.i.d. Bernoulli trial, with probability 1/2 that Dr. Gray's forecast is better (or worse) than the author's. For this test, the recorded p-value is approximately 7.3 percent. More sophisticated tests accounting for the 5 ties and the magnitudes of the forecast errors yield even larger p-values.