A Discussion of "Loss Estimates Using S-Curves: Environmental and Mass Tort Liabilities" by Bruce E. Ollodart

Kirk G. Fleming, FCAS, MAAA

A Discussion of "Loss Estimates Using S-Curves: Environmental and Mass Tort Liabilities"

by Kirk Fleming

Abstract: This paper is a discussion of Bruce Ollodart's 1997 Winter *Forum* paper on using S-Curves to model environmental and mass tort liabilities. To start, there is a brief summary of Ollodart's paper. Then I introduce a type of S-Curve known as the logistic curve. The logistic curve assumes a maximum number of claims so it eliminates at least one of the problems Ollodart mentions with the curves he discusses. Finally, I finish with some comments on the modeling process.

Bruce Ollodart in his paper "Loss Estimates using S-Curves: Environmental and Mass Tort Liabilities" proposes using S-Curves for the analysis of mass tort liabilities (e.g. environmental and asbestos). The appeal of these curves is that their shape matches how we have seen environmental claims emerge and be paid. The paid losses associated with these claims start out slowly, increase rapidly for a period of time and then finally slow down again. The cumulative payment pattern or cumulative reporting patterns follow this S pattern. Ollodart discusses potential candidates for S-Curves to use for this analysis as well as the strengths and weaknesses of these individual candidates.

In this paper, I discuss another type of S-curve called the logistic curve along with a justification for its use. This form of S-curve has a long history of use in economics and the social sciences. It can eliminate some of the shortcomings that Ollodart points out with some of the curves he uses.

A Quick Summary of Bruce Ollodart's Paper

Ollodart says that he has tested a number of curves as appropriate S-Curve models. One successful candidate is the power curve that he discusses in his paper. Another curve he discusses is the gamma curve.

The power curve has the following form:

$$Y = s\,(x-b)^P + c.$$

The variable Y represents the cumulative paid or reported losses, s is a scalar coefficient greater than zero; x is the year of projection (or year corresponding to the historical data), b represents the time at which the curve's inflection point occurs, P is an odd power between zero and one, and c is a constant representing the projected cumulative paid losses at time b. The power P is typically chosen from among the family of fractional powers 1/3, 1/5, 3/5, 1/7, 3/7, 5/7, 1/9, etc.

Ollodart mentions several problems with power functions. The curve increases without bound so the actuary must select a maximum runoff period. This selection has to be made possibly with little information to justify the selection. In practice the rate of change for the curve might be very low once you get out many years on the curve. It may not make a significant difference if you pick a 20-year runoff period or a 60-year runoff period. But it opens up an area for others to disagree with the methods used and the projections.

For example, the 1997 paper uses the power curve to project asbestos losses and the selection is a 20-year runoff period. Does it make a difference if we use 20 years or 60 years? Well depending upon the situation and the data, perhaps not. But if I am talking to the outside auditors, they might be reluctant to buy into the 20-year runoff period and if I am talking to my CFO he or she might not readily buy into a 60-years runoff. Either way, it's a conversation that we would rather not have after presenting results.

Another potential problem with the power curve is that it is very sensitive to the selection of the factors. To get around this problem Ollodart suggests restricting the parameter P to ten possible values. Given those ten different values, fit the curves to the data to get the remaining parameters in the curves. Then select the best curve from among those ten choices.

Logistic Curves

An excellent book to study, if people have the time, is Martin Braun's "Differential Equations and Their Applications". The book is an introduction to differential equations. Braun describes more and more complex and successful applications of differential equations in economics and the social sciences. We can use some of the techniques that Braun describes to construct a model of the claim payment process for mass torts.

Suppose the following story describes an environmental claim payment process. We are dealing with a type of pollution claim that was once covered under an occurrence policy but is no longer covered. Let us say the change in coverage took place in 1986. So there is the possibility of these pollution claims being reported from old policies but there will be no new claims of this type from policy years after 1986.

Initially, reporting and payments from these types of pollution claim were light. There was no incentive for insureds to report the claims and the regulatory agencies were not really pushing to get things cleaned up. Then some policyholders actually received some large settlements to clean up their pollution sites. As news of these settlements began to spread, other policyholders began to take notice. Regulators also began to take notice of the money available for clean up costs. As more and more insureds won settlements, it generated more and more reports of claims and demands from insureds for money for clean up. Eventually, the pool of reported and closed claims began to reach the maximum number of potential claims. As there were less and less claims that potentially could be reported, the rate of new reports slowed down. The same would be true with payments. As there were less and less claims that were yet to be settled, the rate of claims payments slowed down.

We can start to model this story for claim payments by saying that the population of cumulative claim payments is a function of time $Y(t)$. Suppose that N is the ultimate loss for all claims and suppose that c is a constant. As claims start to be settled successfully for large amounts, it causes more and more claims to be reported and settled. As we get closer to the ultimate loss dollars for all claims, the rate of new payments slows down. An initial value differential equation that describes this process is

$$dY/dt = c\,Y\,(N - Y) \quad \text{with} \quad Y(0) = 0. \qquad (1)$$

The solution to this problem is

$$Y(t) = N\exp(cNt) \,/\, (N - 1 + \exp(cNt)). \qquad (2)$$

There are a few desirable features about this solution. The graph of this solution is an S-curve. Also, it has a maximum value since that was one of the assumptions we started with. And finally, it is the result of modeling a process.

The equation (2) describes the logistics law of population growth. Braun points out that it was first introduced in 1837 by the Dutch mathematical-biologist Verhulst. It was an enhancement to the Malthusian law of population growth that had the unrealistic implication that populations grow to an infinite size. The logistic law assumes a maximum point for the population.

The logistic equation has been used to model many different growth patterns. One such pattern was the spread of technological innovations. Braun points out that an implication of the logistic equation is that growth speeds up to a point where the modeled population has reached its half way point. After that it slows down. At least for the spread of technological innovations, it seems that the actual data follows a pattern where the rate of growth slows down beyond the half way point. When describing the model for the spread of technological innovations, this could be explained by an enhancement to the story that allows for the impact of advertising in addition to word of mouth and an extra term in the differential equation. Let c' be a constant.

$$dY/dt = c\,Y(N-Y) + c'(N-Y) \quad Y(0) = 0. \qquad (3)$$

This extra term in the model says that when the number of people who have not heard of the technological innovation is large, there is a definite influence due to advertising.

The solution to this problem is

$$Y(t) = Nc'[\exp((c' + cN)\,t) - 1] \,/\, (c\,N + c'\exp((c' + cN)\,t)). \qquad (4)$$

The graph of this enhanced logistic equation has a maximum and will also be an S curve for appropriate choices of c and c'.

Using the Curve with the Original Data

The source of the original industry data in Ollodart's paper was confidential. Because the details behind the data are not available, I do not know what the claims process is and I was not able to construct an appropriate model to represent the claims process.

However, just to show that logistic curves can be fit to data and to give people a set of numbers they can use to check their work if they reproduce these equations, I refit the original data in Ollodart's paper with the enhanced logistics curve. I want to strongly emphasize that these calculations are not alternative projections of the results. One of my main points in this paper is that it is important to model a process rather than just fit a curve to data.

In practice when using a logistic curve, I have found the parameters N, c and c' using the Solver feature in Excel. I have the Solver minimized the sum of the squares of the actual points and the fitted points. Whether minimizing the sum of the squares of the actual points and the fitted points is the best function to minimize is up for discussion. Ollodart points out that it might be more appropriate to minimize a function that gives more weight to later data. That sounds like a good approach since later data is presumably more relevant.

The original asbestos data is shown on Exhibit 1 along with the power curve results and the results of the enhanced logistic curve. A graph is shown on Exhibit 2. Exhibit 3 has the original pollution data along the modeled power curve results and the enhanced logistic curve results. The first point in the pollution data looks too big to be the initial point. I assumed the first available point was actually year 6 as opposed to year 1 based on the annual change in losses. Exhibit 4 shows a graph of the results.

Curves With and Without Stories

In the spirit of provoking discussion, I will throw out the following thought – in some circles, data mining is considered a bad thing. Now I am hesitant to say that because I have friends who think data mining is a really good thing. And I suppose the explanation must be found in the way we each think about data mining.

There is a paper on the Chartered Financial Analyst syllabus called "Using Economic Models" written by Avery B. Shenfeld, the Senior Economist for CIBC Wood Gundy. In the paper he discusses different forecasting approaches and he discusses problems in the use of models. One of the potential problems that he discusses with the modeling process is data mining. By data mining he is referring to a process where the researcher will do multiple calculations with the data in order to get something that works. So in a sense, the modeler just lucks out in finding something that works on the past data but has no explanation for why it should work going forward.

When I first read the Ollodart paper, the process of continually fitting different S-curves to the data with the only justification being that the data looked like an S-curve struck me as being open to this type of criticism. The process described was modeling the data as opposed to modeling a process.

Some actuaries would argue that our field is threatened by other professionals who are just as qualified to do the same type of analysis as we do. We have to be careful to construct models with their appropriate inputs and solve for the implications of those models. Then we have to accept or reject the results of those models and assumptions based on our best judgment along with the input and insights of other experts. For those people who defend data mining, my guess is that they argue data mining produces a model that had previously gone unnoticed. Once the model is uncovered, the modeler would only use it if they understood how the model should work going forward.

As I wrote, it's a point for discussion.

Other Ideas

One of the reviewers of this paper asked, "I am curious, is there a reason that the S shaped curves do not 'work' for relatively shorter tail lines such as medical malpractice and workers compensation?" My answer to that is, "Who says they don't work?" I have not used S-Curves for development work because there are other accepted loss development models. There are papers that discuss using mathematical curve models for the development process such as Richard Sherman's useful and practical paper, "Extrapolating, Smoothing and Interpolating Development Factors". I would say the work is still to be done to see if there is a model of the claims process that justifies the use of an S-Curve on lines other than environmental.

Closing Comments

In corresponding with Bruce Ollodart about his paper, Bruce pointed out that it is important for others to take the basic ideas proposed by some and work to develop them. I certainly agree with that. All of the work that we do is building on things that others have done.

Given that, I would have to give thanks to my all my teachers. That would include Martin Braun for writing the book so that I could lift material

directly from it. I also have to give thanks to all the people that I have worked with over the years and all the people that I work with now. Finally, thanks go to Bruce Ollodart and all casualty actuaries who have built and expanded the Casualty Actuarial Society so that we all have a profession to share.

Exhibit 1

Asbestos Indemnity and Expense
Cumulative Paid Loss
Original Data from Ollodart Paper
(000's)

CY	(1) Actual Calendar Yr Cumulative Paid Loss	(2) Annual Change In Losses	(3) Fitted Calendar Yr Cumulative Paid Loss Power Curve	(4) Fitted Calendar Yr Cumulative Paid Loss Logistics Curve	CY	(5) Fitted Calendar Yr Cumulative Paid Loss Power Curve	(6) Fitted Calendar Yr Cumulative Paid Loss Logistics Curve
1978	362	362		18,455	2003	3,624,338	8,777,269
1979	17,918	17,556	57,426	41,236	2004	3,693,242	9,596,461
1980	33,987	16,069	117,252	69,342	2005	3,758,660	10,377,967
1981	84,014	50,027	179,775	103,989	2006	3,821,037	11,108,040
1982	193,596	109,582	245,358	146,660	2007	3,880,734	11,776,818
1983	258,994	65,398	314,454	199,154	2008	3,938,045	12,378,529
1984	284,030	25,036	387,632	263,638	2009	3,993,215	12,911,201
1985	324,534	40,504	465,635	342,713	2010	4,046,451	13,376,037
1986	374,068	49,534	549,452	439,475	2011	4,097,929	13,776,626
1987	612,636	238,568	640,459	557,567	2012	4,147,800	14,118,133
1988	752,146	139,510	740,659	701,232	2013	4,196,194	14,406,601
1989	898,011	145,865	853,169	875,330	2014	4,243,225	14,648,374
1990	1,026,623	128,612	983,338	1,085,315	2015	4,288,996	14,849,691
1991	1,259,167	232,544	1,141,855	1,337,149	2016	4,333,593	15,016,410
1992	1,585,463	326,296	1,357,474	1,637,124	2017	4,377,097	15,153,856
1993	2,078,939	493,476	2,095,513	1,991,557	2018	4,419,578	15,266,747
1994	2,470,635	391,696	2,591,167	2,406,346	2019	4,461,100	15,359,187
1995	2,835,848	365,213	2,802,383	2,886,366	2020	4,501,720	15,434,691
1996			2,959,027	3,434,733	2021	4,541,489	15,496,237
1997			3,088,106	4,051,996	2022	4,580,456	15,546,320
1998			3,199,887	4,735,365	2023	4,618,663	15,587,021
1999			3,299,556	5,478,147	2024	4,656,149	15,620,061
2000			3,390,156	6,269,559	2025	4,692,951	15,646,857
2001			3,473,648	7,095,057	2026	4,729,102	15,668,574
2002			3,551,384	7,937,246	2027	4,764,633	15,686,165

Power Curve 7
Fulcrum Year 1993

Logistic Curve Paremeters

N	15,760,600
c	1.35E-08
c prime	1.05E-03

Asbestos Indemnity and Expense Cumulative Paid Loss Data from Ollodart's Paper

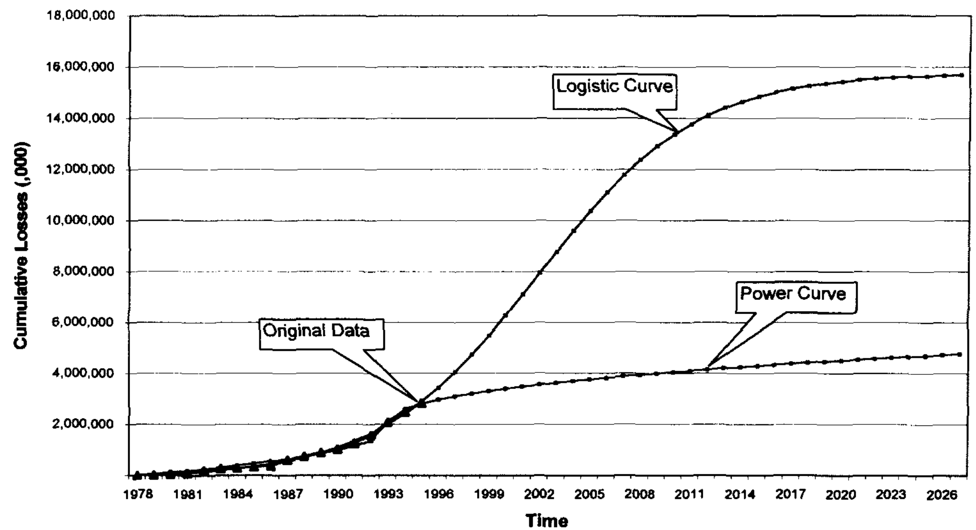

Exhibit 2

Exhibit 3

Pollution Indemnity and Expense
Cumulative Paid Loss
Original Data from Ollodart Paper
(000's)

CY	(1) Actual Calendar Yr Cumulative Paid Loss	(2) Annual Change In Losses	(3) Fitted Calendar Yr Cumulative Paid Loss Power Curve	(4) Fitted Calendar Yr Cumulative Paid Loss Logistics Curve	CY	(5) Fitted Calendar Yr Cumulative Paid Loss Power Curve	(6) Fitted Calendar Yr Cumulative Paid Loss Logistics Curve
1983	135,953	135,953		184,894	2006	5,922,432	6,020,415
1984	172,946	36,993	160,048	255,951	2007	6,072,103	6,077,768
1985	222,134	49,188	326,616	347,749	2008	6,217,219	6,121,693
1986	407,273	185,139	500,739	465,358	2009	6,358,232	6,155,202
1987	579,370	172,097	683,772	614,438	2010	6,495,522	6,180,686
1988	914,273	334,903	877,553	800,879	2011	6,629,411	6,200,022
1989	1,150,537	236,264	1,084,678	1,030,149	2012	6,760,177	6,214,668
1990	1,410,354	259,817	1,309,057	1,306,319	2013	6,888,061	6,225,746
1991	1,613,107	202,753	1,557,103	1,630,820	2014	7,013,275	6,234,117
1992	1,951,047	337,940	1,840,889	2,001,161	2015	7,136,002	6,240,438
1993	2,334,475	383,428	2,189,868	2,410,024	2016	7,256,409	6,245,208
1994	2,779,049	444,574	2,791,813	2,845,207	2017	7,374,642	6,248,806
1995	3,373,188	594,139	3,557,863	3,290,741	2018	7,490,832	6,251,519
1996			3,914,670	3,729,083	2019	7,605,098	6,253,565
1997			4,202,122	4,143,787	2020	7,717,546	6,255,107
1998			4,452,431	4,521,847	2021	7,828,275	6,256,268
1999			4,678,395	4,855,025	2022	7,937,373	6,257,144
2000			4,886,715	5,140,000	2023	8,044,921	6,257,804
2001			5,081,439	5,377,579	2024	8,150,993	6,258,301
2002			5,265,244	5,571,447	2025	8,255,658	6,258,675
2003			5,440,014	5,726,903	2026	8,358,979	6,258,957
2004			5,607,135	5,849,815	2027	8,461,015	6,259,170
2005			5,767,665	5,945,921			

Power Curve 9
Fulcrum Year 1995

Logistic Curve Paremeters

N	6,259,819
c	4.49E-08
c prime	1.93E-03

Pollution Indemnity and Expense Cumulative Paid loss Data from Ollodart's Paper

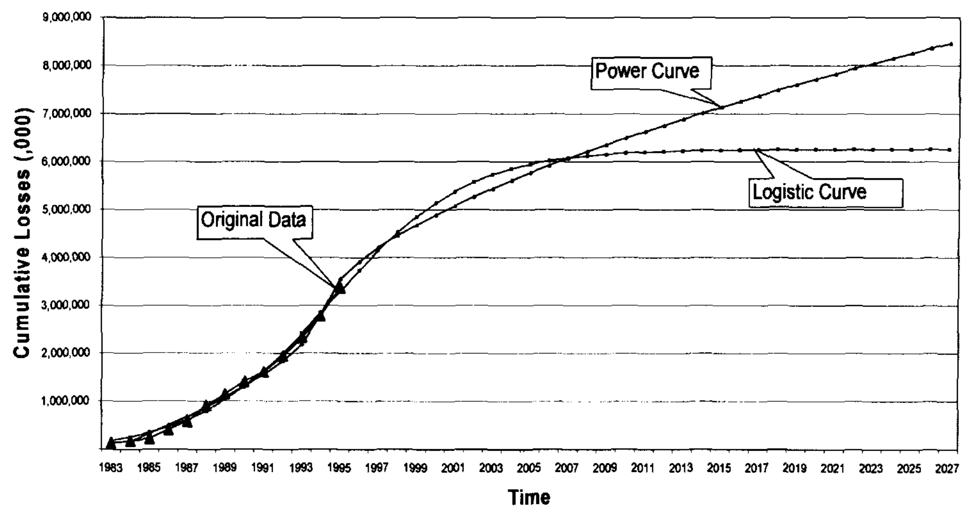

Exhibit 4

Bibliography

Braun, Martin 1993. Differential Equations and Their Applications: an Introduction to Applied Mathematics. New York: Springer-Verlag

Ollodart, Bruce C. 1997. "Loss Estimates Using S-Curves: Environmental and Mass Tort Liabilites," Casualty Actuarial Society Forum Winter 1997: 111-132.

Sherman, Richard E. 1984. "Extrapolating, Smoothing and Interpolating Development Factors," Proceedings of the Casualty Actuarial Society LXXI: 122-155.

Shenfeld, Avery B. 2001. "Using Economic Models," 2002 CFA Level III Candidate Readings Volume 1: 7-11.