A Comparison of Data-Driven and Model-Based Approaches to Quantifying Railway Risk

Tim Bedford, John Quigley
Department of Management Science, Strathclyde University,
40, George St, Glasgow, UK,

Simon French
Manchester Business School, University of Manchester,
Booth Street West, Manchester, M15 6PB, UK

Abstract

This paper presents some of the results of a project sponsored by the UK Railway Safety and Standards Board (RSSB). An earlier statistical evaluation of a previous version of the RSSB Safety Risk Model (SRM), a combined Fault/Event Tree, conducted by Prof Andrew Evans had concluded that the model was unduly pessimistic. We have constructed a hypothesis test based on the relative likelihood techniques using the most recent version of the SRM as the null hypothesis. The results support the SRM being consistent with the historical data. Two significant differences between these two studies are the statistical methods employed to support the analysis and the removal of certain significant conservative assumptions from updating the versions of the SRM.

The paper discusses the demands that different model purposes place on these models, and explores the question of whether or not it is meaningful to compare their outputs. The use of expected fatalities as a metric for expressing risk in both models is questioned because of the heavy-tailed form of the distribution for fatality numbers given a fatal accident.

1 Introduction

A quantitative risk model for the UK railways called the Safety Risk Model (SRM) is maintained by the Rail Safety and Standards Board (RSSB). This model (RSSB 2003) quantifies risk for a number of different categories – for example in terms of expected numbers of minor injuries, major injuries and fatalities for passengers, staff and members of the public. The SRM contributes to the safety strategy of RSSB and allows it to appraise potential benefits of safety improvements, at least at a system wide level (although more detailed models are used in actual safety cases). The model is a large combined Fault and Event tree model with 122 hazardous event trees and more than 4000 end states.

Use of the SRM to quantify risk levels on the UK railways has been criticised by Prof Andrew Evans (2000, 2002). He has taken a data driven approach to assess passenger fatality risk. Based largely on historical UK data from 1967 onwards, he has used non-homogeneous Poisson processes with an exponentially decreasing intensity function to model the changing (decreasing) rate of fatal accidents and has estimated the mean numbers of fatalities given a fatal train accident. Using this information he has been able to build a time dependent model of the expected number of fatalities per year. This model is rather more optimistic than the SRM (version 2) predictions. Hence there has been some debate between Prof Evans and RSSB about whose model gave the better picture of railway risk. Version 3 of the SRM removed some known conservatisms and gave a considerably closer assessment of expected fatalities, but the need for an understanding of the sources of the differences remained.

On the one hand we have a black-box statistical model of fatal accidents and their consequences which is quantified by – fairly sparse – historical data. On the other hand we have the SRM, a more detailed Fault and event tree model, which is quantified by a combination of expert judgement and historical data. The SRM gives information about non-fatal as well as about fatal accidents.

A study has been performed by a team from the Universities of Strathclyde and Manchester, with the objects of reassessing SRM structure, input and outputs from a statistical point of view, taking into account the work of Andrew Evans.

This paper discusses the results of that study. The opinions given here are those of the authors and not necessarily of the Railway Safety and Standards Board. In Section 2, we describe the Safety Risk Model in relation to its purpose. Section 3 is a brief summary of the work of Professor Evans, identifying his underlying assumptions and use of data. Section 4 provides a summary of the quantitative assessment we conducted of the latest version of the SRM and the available data; we make the same parametric assumptions for the rate of occurrence of fatal incidents as Professor Evans, but construct statistic inference following Likelihood Ratio procedures rather than appealing Normal distribution approximations. The results are presented in terms of confidence intervals for comparison. Finally, we reflect on the analysis conducted and provide recommendations for further improvements of the SRM.

2 The Safety Risk Model

The Safety Risk Model (SRM) is currently in version 3, and version 4 is in preparation. Its top-level aims are to provide:
- An understanding of the nature of the current risk on the mainline railway
- Risk information and risk profiles relating to the mainline railway.

The SRM has been developed as a joint Fault and Event Tree model, with 122 different event trees leading to more than 4000 endpoint scenarios. Event trees allow the specification of scenarios involving different time periods to take account of varying passenger loadings, distinctions between freight and passenger lines, and accidents involving various numbers of trains, amongst many other factors.

Accident sequence endpoints have associated data about passenger injury probabilities for up to three trains involved in the incident and specify also expected numbers of staff and member of the public injuries.

Consequences are measured in terms of expected numbers of casualties (minor, major and fatal) in the three categories of passengers, members of the public, and staff. The expected numbers in these different categories can be expressed in terms of "equivalent fatalities" using the ratio's 1:0.1:0.005 in order to give an overall measure of risk. An FN-curve is also generated, but does not play a significant role in the decision making process.

Data for the model comes from a number of different sources. Historical data from the UK railways is collected within the railway system and made available to the SRM project. With so many different event tree branches and endpoints though, historical data does not provide a detailed enough basis for quantifying the full model. Hence there is a large reliance on the use of – currently fairly unstructured – expert opinion to make judgements about failure frequencies, probabilities and consequences.

One of the most important modelling aspects is that the SRM is a system-level model. It does not attempt to model individual features of individual lines within the rail network. Instead, risk is determined for particular units (for example, per passenger-train mile) using data that is assumed to be appropriate for the whole network.

As part of the study we did make various recommendations for improvements to the probabilistic modelling in the SRM. Some of these will be discussed briefly below.

3 Historical data-driven approaches to quantifying railway risk

The work of Andrew Evans (refs) is the most significant recent work performed in the UK using the UK national data for fatal accidents. It should immediately be noted that his work is based on fatal accident data, and seeks to assess the frequency and consequences of fatal accidents, in contrast to the SRM that looks at all accidents including non-fatal ones.

In broad terms, Evans has looked at a data set of all fatal collisions, derailments and overruns in the UK national railway system from 1967 to the present. A 'fatal accident' is one in which there was at least one fatality. He has selected a coherent dataset of 78 fatal accidents which for the last full analysis (Evans, 2002c) led to 311 fatalities over the period. His estimation and forecasting process has essentially the following steps.

- Estimate the number of fatal accidents per annum, λ, and its (declining) trend over the years.
- Estimate the number of fatalities per fatal accident, μ.
- Combine the estimates to give an estimate of the number of fatalities per annum: $\lambda \times \mu$.

Estimates are provided for three mutually exclusive classes of fatal incidents (i.e. SPAD, ATP Preventable and Non-ATP Preventable) and adjustments are made for changes in rolling stock. While it is assumed that accident rates are decreasing, there is no similar assumption for fatalities per accident (see Figure 1).
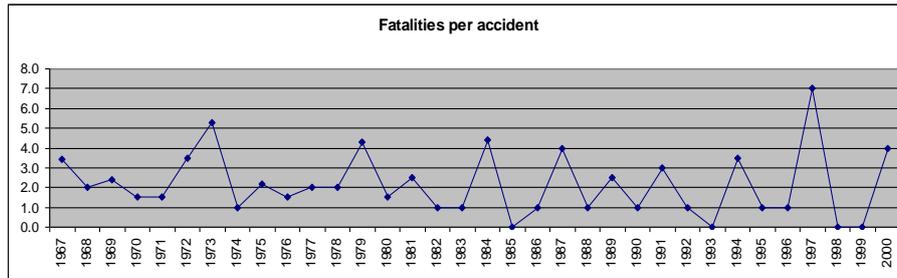

Figure 1   Average number of fatalities per accident by year with the atypical Hither Green (1967), Clapham Junction (1988) and Ladbrooke Grove (1999) accidents eliminated.

Evans makes no claim of any distributional form for the number of fatalities given a fatal accident. The data for the number of fatalities in an accident (Figure 2) gives a high probability to single deaths but also shows a "thick tail". There is no reason to fit a particular parametric model to the data. The SRM could in fact could provide a model if extended to distributions of fatalities rather than just expectations.
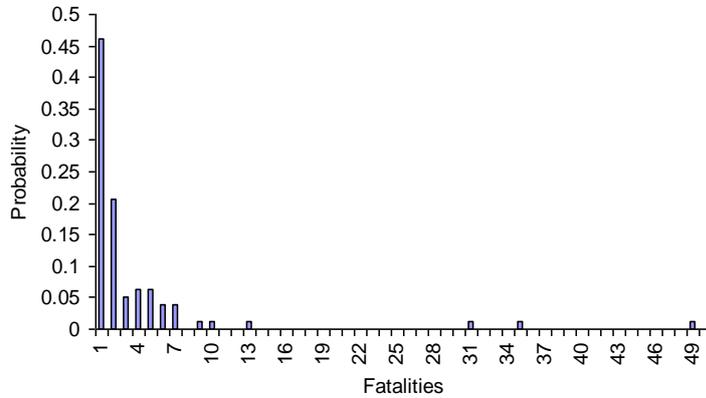

Figure 2 Empirical distribution of the number of fatalities given a fatal incident

The overall results in 2002 were that Evans' model was suggesting a mean number of 4.3 fatalities per year (Evans 2002c, p13) while SRM version 2 was suggesting 11.3. In the HSE Report 066 p75 (HSE, 2003) and in Evans (2002a) p11, he refers to the development of confidence intervals. These are based on the normal distribution approximation for the sample mean. The confidence intervals are important, because the confidence interval for the expected number of fatalities per

year does not contain the overall version 2 SRM estimate. This led Evans to conclude that the SRM was overestimating the overall risk.

4 A comparison of the SRM with the historical data record

A quantitative comparison between the SRM and the model developed by Andrew Evans (referred henceforth as AEM) was conducted with the aim of evaluating the extent to which these approaches differed. AEM considers incidents that result in fatalities only, while the SRM considers incidents that result in fatalities, major injuries and minor injuries. Moreover, AEM was developed firstly through modelling the rate at which fatal incidents occur and secondly through modelling the number of fatalities in an incident given at least one fatality. As such, an immediate comparison between these modelling approaches was not possible. Seven different SRM event trees were identified as corresponding to the categories considered in the historical analysis of Andrew Evans. The different endpoints were classified according to how they correspond to each of Evan's categories.

A graver problem is that the SRM-predicted accident rates include non-fatal accidents. Therefore, for each event sequence the rate of occurrence must be reduced to measure the rate at which fatal incidents occur. We assumed that the probability of death of individuals at risk are independent and identically distributed, resulting in a Binomial distribution to describe the number of fatalities in a given fatal incident with a known number at risk. We approximated the Binomial distribution with a Poisson distribution for computational convenience. This enabled a comparison of accident rates.

In order to generate confidence intervals for the historical data we adopted the likelihood ratio method (Kalbfleisch and Prentice 2002). In order to generate confidence intervals for the mean number of fatalities given a fatal accident we used Bootstrapping (Davison and Hinkley 1997).

The results of this comparison were that the average number of fatalities given a fatal accident is 3.99 and the SRM (version 3) estimates 4.9. This was found to be within the 95% Confidence Intervals for the mean fatalities per fatal accident of (2.5, 5.8). The number of fatal incidents estimated by AEM in 2002 due to SPAD, ATP Preventable and ATP Non Preventable are 0.45, 0.011 and 0.45 respectively. The SRM's estimates were 0.38, 0.09 and 0.92 respectively, all of which are within the 95% confidence intervals for 2002. These are presented in Table 1.

Type	Lower 95% CI 2002	Upper 95% CI 2002	Point Estimate AEM	SRM
SPAD (i=1)	0.17	1.04	0.45	0.38
ATP Preventable	0.0002	0.12	0.011	0.09
ATP Non Preventable	0.19	0.93	0.45	0.92
Total	0.31	2.03	0.911	1.39

Table 1   95% Confidence Interval of Mean Fatal Accidents (AEM) 2002 and SRM

5 Discussion and Conclusions

That two distinct models for fatality numbers should give different results will come as no surprise. The basis for comparison is difficult however. The use of confidence intervals to give an estimate of the uncertainty around the mean number of fatalities per year is fraught with difficulties. The main difficulty is that such confidence intervals are model based and hence do not reflect uncertainty outside the model class. Even in this case where an appeal is made to the Central Limit Theorem, it is not clear whether the number of observations is large enough, which is particularly worrying with such "thick tail" as seen in Figure 2. We found no statistically significant difference between the SRM version 3 and the data.

Although the overall results of the SRM are broadly consistent with the fatality data and the model is well structured and documented, there remain a number of points where the statistical modelling can be improved. In particular, the parameter estimates depend on very sparse data sets. This arises partly because of the engineering imperative to give a finely structured model to gain insights. In other work we have looked at the use of credibility theory (a technique that has been much applied in the insurance industry) to adjust estimates where there is little data. This is a good alternative to the chi-squared estimator often used in risk assessment.

The SRM overall expression of risk is made through the use of "expected fatalities". The estimator of this metric is not robust when the underlying distribution is heavy tailed. More importantly, the metric does not capture public perception about risk on the railways, which is driven largely by the rare multi-fatality accidents. From both an estimation point of view, and a risk perception point of view it would be better to consider the median fatalities or other quantile based metrics. As far as the SRM is concerned, there is a need to model the ET endpoint conditional distributions fully rather than through conditional expected value. The scenarios developed currently have a large amount of variance in the consequences. Again, in further work for the RSSB we have explored this in more detail.

References

1. RSSB 2003 *Q3 2002/03 Safety Performance Report*, downloadable from http://www.railwaysafety.org.uk/q3spr0203.asp
2. Evans 'Fatal train accidents on Britain's mainline railways' *J. Royal Statistical Society* A **163**, 99 –119, 2000.
3. Evans 'Fatal train accidents on Britain's main line railways: end of 2001 analysis' Centre for Transport Studies, University College London, 2002.
4. Mehlhorn K. Data structures and algorithms, vol 1, Sorting and searching. Springer Berlin Heidelberg New York, 1984
5. de Kleer J. An assumption-based TMS. Artificial Intelligence 1986; 28:127-162
6. Kalbfleisch J and Prentice R *The Statistical Analysis of Failure Time Data*, Wiley 2002
7. Davison A and Hinkley D *Bootstrap Methods and Their Applications*, Cambridge University Press, 1997.