GOOD AND BAD DRIVERS—A MARKOV MODEL OF ACCIDENT PRONENESS

EMILIO VENEZIAN

VOLUME LXVIII

DISCUSSION BY SANFORD R. SQUIRES

A series of studies by the Department of Transportation in the 1960's and early 1970's first popularized analysis of automobile insurance experience through accident proneness models. Unlike many actuarial approaches which tend to be empirical and emphasize practical interpretations, accident proneness models tend to be heavily theoretical and often prove difficult to interpret on a non-technical level. If one is not already familiar with the various Poisson process models found in actuarial literature, Dr. Venezian's paper might be extremely slow going.

Dr. Venezian's summary of other authors' models in the introduction is remarkable in its terseness without omitting the salient points. In the same style, the author points out many strengths and implications of his own proposed model. This is a paper which requires careful reading, working out details derived by the author, for full appreciation and understanding.

Dr. Venezian analyzes some of the weaknesses in his own model. In particular, the crucial role of mileage in determining the fitting parameters is noted. According to his data, males have almost twice the accident frequency of females in the younger age groups. The paper attributes all this difference to mileage alone, since mileage differences conveniently follow the same pattern of accident likelihood differences between the sexes. The author notes that detailed data of mileage by age and sex would be helpful. Such data was analyzed in the early seventies[1]—but the mileage patterns would seem to disrupt rather than enhance Dr. Venezian's model. In particular, average annual mileage for male drivers rises steeply in the younger ages. This is directly counter to the exponential decay of accidents with age shown in the model. Female mileages show much less variation by age than male mileages. Mileage does contribute to explaining

[1] The *National Transportation Survey,* 1969–1970, conducted by the Bureau of Census for the Federal Highway Administration developed average annual mileage data for the general driver population by age and sex of drivers. This data has recently been published and used in analysis of automobile classification plans by the Rate Regulation Division of the California Department of Insurance in Phase II of their *Study of California Driving Performance* (November 1979), p. 17.

the variation in accident frequencies by age and sex, but not in as simple a way as used in the author's paper.

Dr. Venezian has created a "Markov process" model. That is, drivers bounce between "good" and "bad" states with a predetermined probability of transition back and forth. To someone familiar with Markov processes, it is obvious that the model will produce an exponential decay of accident likelihood over time. If driver data shows accidents decreasing with age in a somewhat exponential fashion, this model has to fit. It is disappointing that the way it fits is so philosophically unappealing—at least from this reviewer's viewpoint. It is difficult to give a layman's explanation for the Markov process other than "it fits the data."

Only three drivers in a hundred go from the good to the bad state every year while seventeen in a hundred go the other direction. Intuitively one would like to see more frequent transitions of some sort reflecting heavy use of autos during the day, but not at night, occasional bad driving conditions and occasional non-use of the automobile for weeks at a time while on vacation—or any of the hundreds of other everyday real life situations where accident likelihood for an individual varies considerably over a short period of time. The long time between transitions in the model stems from the transition probability parameters fitted to the data. These parameters are a direct result of the overall rate of exponential decay in accident likelihood with age. These low transition probabilities fit best, but seem to defy intuitive interpretation other than to produce the steady, long term decrease in accident likelihood with age. High frequency transitions are more intuitive but are practically useless to produce the desired data fit from the model.

The ad-hoc assumption made by the author that males and females start their transitions at different ages seems non-intuitive. In the author's model, female transitions start before male transitions. This seems intuitively counter to the fact that females tend to start driving at later ages than males in our society. It would seem more plausible to either allow different accident likelihood ratios between "bad" and "good" states for men and women or allow different transition probabilities or starting ages. These possibilities are somewhat ad-hoc, too, but could fit the data. They are more in keeping with some particular data published by the state of California in Phase II of its recent study.[2] A mileage-adjusted surrogate for accident likelihood shown graphically

[2] *Study of California Driving Performance Phase II*, Rate Regulation Division—California Department of Insurance (November 1979), p. 22.

in that study suggests different "bad" to "good" ratios for males and females. Better data from strictly insurance sources with loss costs and mileage would be helpful, but is not generally available.

The data fit in the author's paper does not strike this reviewer as particularly great. However, the calculated statistical significance test on the excess variance test shows that other models, like a simple Poisson, produce much worse fits to the data! This is the best fit offered so far.

Dr. Venezian observes that one might be able to distinguish different models by looking at the distribution of time between accidents. For his own model, he implies that one would most likely see two clustered distributions for the number of accidents for an individual over subintervals of time. Both statements would seem to be only of theoretical interest and difficult to measure in the real world because of the low frequencies of accidents and the likelihood of changing driving environments during the period of years between accidents. Although data in this regard might support assertions about excess variance in total, it seems unlikely that one could use this test to distinguish between different competing models.

Models on the order of complexity of the author's will quickly approach the point where their mathematical robustness will necessitate more detailed data and more intricately structured statistical tests in order that further progress be made. It is this reviewer's opinion that intuitive interpretation of models will become important in selecting the best among competing models.

Several extensions of the author's model are possible.

One could possibly add more discrete states. In addition to "good" and "bad," I'm sure someone will suggest a third "ugly" state. Surely, this would produce much improved data fits—but new insights are unlikely to result.

One could create a continuous spectrum of states and a corresponding set of transition probabilities—a task to be tackled by only the purest of pure actuaries, but not an impossible task.

Finally, one might follow up on the author's speculations on the further utilization of mileage data. One might create a model using cumulative lifetime mileage and its first derivative (or first difference), annual mileage, as the entire basis for accident likelihood. (The mathematical treatment of accident likelihoods as separate from mileage in the author's paper is cosmetic, resulting from a two variable notation, rather than any intrinsic functional independence in the equations.)

Dr. Venezian has written a very professional and accurate paper proposing a new model for accident proneness. Although this review has made general criticisms on non-intuitive aspects of the author's model, the paper is of professional calibre and few technical flaws can be found. Dr. Venezian has taken us a step in complexity beyond previous models. The long journey yet to go should provide significant challenge for future contributors to the *Proceedings*.