Estimation of Carrier Accident Risk Potential

by
Moin El-Herraoui

A thesis
presented to the University of Waterloo
in the fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Civil Engineering

Waterloo, Ontario, Canada 1999

0-612-52026-9

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

The cost of accidents on North America's highways is estimated to be in the billions of dollars annually. Truck safety represents a major safety issue and is of special interest to policy makers. Trucks are involved in a disproportionate number of accidents with respect to their population. Previous studies have analyzed truck safety from a number of perspectives including those of the highway, the driver, the vehicle and the environment. However, few studies have viewed this issue from a carrier perspective. In Ontario, 100% of truck accidents are attributed to 7% of carriers. If we can target carriers who are most at risk of accidents, we may be able to reduce a significant percentage of truck accidents. Carrier attributes and safety management practices can explain large truck accident involvements on a carrier specific basis. Authorities need to understand the accident risk potential of carriers and to identify those who are most at risk of accidents. Those carriers can then be selected for safety audits and other safety interventions.

Safety audits are expensive and, as a result, only a small percentage of carriers are audited in any given year. In Ontario, less than 1% of all carriers are normally audited. This statistic means that the probability of selecting high risk carriers is small. An efficient procedure needs to be established to identify for auditing those carriers with the greatest accident potential. The main objective of this research is to develop a methodology to identify those carriers that have the highest accident potential.

Analysis of Ontario accident data by carrier size underscored a difference between small and large carriers in terms of their respective accident rates per truck. Small carriers within the context of this study are those that have a fleet size of 10 trucks or less. Statistical models which are based on accident history (truck accident involvement in previous years), have proven to be a reasonably good predictor of future accident experience, for large carriers. Large carriers that have experienced a high number of accidents in previous years are likely to have a large number of accidents in the future and vice versa. For small carriers, however, accident history does not provide a good predictor for future accident behaviour. Other attributes of small carriers are needed to provide inference concerning future accident behaviour.

Small carriers comprise over 93% of all carriers registered to operate in Ontario and hence, identifying *high risk* carriers in the small carriers population is more difficult and more important. For the period of 1991 to 1995 in Ontario, small carriers represented 30% of all trucks registered in the province. These same carriers accounted for over 50% of all detentions and convictions for drivers and vehicles infractions. Thus, for small carriers, detentions and convictions could provide inference concerning future accident involvements.

In this thesis, an Empirical Bayes approach was adopted to identify small carriers with the highest risk of having accidents. The model was developed and tested based on the Carrier Vehicle Operator Record (CVOR) database of Ontario for the period of 1991 to 1995. This model makes use of statistical information on carrier convictions and detentions as *prior* and accident history as *likelihood*. The carrier attributes and accident data for the period of 1991 to 1994 were used to develop the model which was then tested and evaluated using the 1995 data.

Different sampling procedures were selected to be compared with the Bayes model. These included a) random selection based on the number of carriers in the population, b) random selection based on the number of trucks in the population, c) selection based on the previous one year of accident history, and d) selection based on the previous 2 years of accident history. Regardless of the number of carriers being sampled, the results of this analysis indicated that the Bayes approach performed significantly better than all other procedures in identifying carriers who experienced accidents in 1995. On average, the Bayes model identified between 3000% to 60% more carriers as compared to methods b) and d), respectively.

We found that for small carriers, detentions and convictions are significant factors in explaining accident involvements. While other variables might further explain accident involvements, many of these variables are unavailable in existing databases. The Bayes model makes an efficient use of the existing data collected in Ontario and does not require any additional data which are usually not available or are too expensive to collect.

This research offers a different approach to highway safety and the potential reduction of truck accidents by investigating the carriers' perspective. The research provides a methodology for identifying high risk carriers that should be targeted for safety interventions which would lead to the improvement of truck safety on highways.

Acknowledgments

I would like to express my thanks and appreciations to my supervisor, Professor Frank Saccomanno for his guidance, advice and encouragement throughout the duration of the research work. I thank Professor Park Reilly for the helpful discussion and advice regarding statistics.

I thank the Ministry of Transportation of Ontario for providing the data used in the thesis. I would like to acknowledge the financial support provided by the Natural Science and Engineering Research Council of Canada through a research assistantship.

I thank my family and my friends for their continuous support. I am most greatful to my mother and my brothers for their love, encouragement and moral support through my many years of school and yet another degree.

To my mother

Contents

List of Figures

List of Tables

Notations

ADS	Accident Data System.
CVOR	Carrier Vehicle Operator Record.
MTO	Ministry of Transportation of Ontario.
Truck	Motor Vehicle Having a Gross Weight More than 4500 kg.
Accident Rate	Accident per Truck per Year.
Carrier	Trucking Firm Having a Legal Entity and Licensed to Operate in Ontario with a Unique CVOR Number.
h	historical accident frequency.
θ	estimate of accident frequency.
$p(h)$	probability of h.
$p(\theta)$	probability of θ.
$p(\theta\|h)$	probability of θ knowing h.
$p(h\|\theta)$	probability of h knowing θ.
$E(\theta)$	Expectation of θ.
$E[\theta\|h]$	Expectation of θ knowing h.

Chapter 1

Background & Literature Review

1.1 Introduction

Road safety is a major concern in the transport sector. The social costs of motor vehicle accidents. or collisions are substantial. In 1990. the costs of accidents that occurred in Ontario were estimated at $9 billion; these costs include human consequences (i.e. injury and death). property damage, time, and material. More than 70% of these costs are the result of injuries or deaths. Using the method of "willingness to pay" by society's members to avoid the collision impact from occurring, the average cost of a motor vehicle collision in Ontario in 1990 was found to be $41,000 and it could be as high as $6.3 million in the case of fatal collisions (Safety Research Office (1994)).

One can see the need to invest efficiently in order to reduce the number of accidents on Ontario's highways. Accidents involving trucks represent a significant part of the total number of accidents, especially fatal accidents.

In 1994 in Ontario, although only 16% of the total registered vehicles were trucks, these vehicles were involved in 25% of the total accidents and 37% of the fatal accidents (Ontario Road Safety Annual Report, 1994). According to Patten et al. (1989), the growing size and number of trucks on highways, and the severity of accidents involving trucks are two good reasons for studying truck safety on highways. In a study carried out by Braver et al. (1997), it was noted that in Virginia in 1995 more than 4800 people died in truck accidents (12% of all motor vehicle fatalities), though trucks accounted for only 3% of registered vehicles in the state (IIHS 1996. FHWA 1995a. 1995b). Virginia is typical of most jurisdictions in North America including Ontario, where truck accidents result in a disproportionate number of persons injured and of fatalities. Reducing truck accidents on highways would result in a significant reduction in the social costs of accidents, and would represent a primary aspect of enhanced highway safety.

There is a need to understand the factors affecting truck accidents so they can be used to predict accidents and to evaluate the effects of the different measures for reducing the number of accidents on highways. Although different approaches have been investigated, most of them have been carried out from the highway, vehicle, driver, or environment perspectives. Few studies have looked at truck safety from the carrier perspective. The carrier perspective is relatively important, since every year a very small percentage of the carrier population (e.g 7% in Ontario) is responsible for *all* truck accidents. In this Chapter, studies that focussed on the highway, vehicle, driver and environment to explain the variations in accidents, are discussed. Also, the few studies that were carried out from the carrier perspective are discussed.

1.2 Highway Safety Analysis

In recent years, different approaches have been used by researchers for establishing relationships between accidents and vehicles, drivers, traffic volumes and environment characteristics. Some studies have explored relationships between accidents and the geometric design of road sections, such as road width, horizontal curvature, gradient (Zegeer et al. (1987, 1990), Okamoto and Koshi (1980), Joshua and Garber (1990), Miaou et al (1992), Miaou and Lum (1993) and Miaou (1994)). For example, Miaou (1994) studied the relationships between truck accidents and horizontal curvature, vertical grade and shoulder width, traffic density, and percentage of trucks in the traffic stream. Many studies have also focused on the relationships between accidents and traffic volumes (Satterthwaite, 1982). Other studies were carried out by Maycock and Hall (1984), Pickering et al. (1986), and Hall (1986), to investigate relationships between truck accidents and highways for different types of road junctions (intersections & interchanges). Nembhard and Young (1995) estimated accident rate on a segment specific basis (i.e. a disaggregate approach) as opposed to an aggregate estimate, using Bayes' approach. Buyco and Saccomanno (1987), studied the relationships between truck accidents and road type, traffic pattern (commuter, non-commuter), traffic volume, truck type (e.g. tractor, tractor-trailer, etc.), load status (empty, loaded), model year, hour of the day (day, night) and driver age.

Bonneson and McCov (1997) studied the effect of a specified median treatment on road safety. They concluded that average daily traffic, median type and adjacent land use were significantly correlated with accident frequency. Hickey (1997) studied the effect of a specific type of shoulder which produces a distinct warning sound and vibration when inattentive drivers drift off the road. They found that "Drift-Off-Road" accidents have been reduced after the installation of such a shoulder. Zhou and Sisiopiku (1997) studied the relationship between hourly accident occurrences and traffic volume to capacity ratios (v/c). It was concluded that, for high v/c values, traffic conflict is considered to be the major factor in explaining high accident rates. However, in low v/c values at night time, driver's inattention is identified as the main explanatory factor for high accident rates.

Other researchers studied the effect of the time of day on accident rates. In Michigan, Blower et al. (1993) concluded that night time is associated with higher casualty rates for heavy truck-tractors, but was also found to have an insignificant effect on property damage accidents. Jovanis and Delleur (1983) found that automobiles had higher accident rates at night time than during the day, while for trucks the accident rates were found to be similar between night time and day time. In studying the effect of different seasons, Evans (1991) found that fatality rates are lower in the winter months. Shankar et al. (1995) studied the effect of rainfall and snowfall on accident rates. They found that these factors had an increased effect on the number of accidents, especially on curves and grade locations. In another study carried out by Chirachavala and Cleveland (1985), it was concluded that truck accident rates are higher on wet/snowy roads.

Vehicle condition has also been investigated in several studies. Hauer et al. (1993) studied different aspects of vehicle conditions (such as brakes and unsafe tires). Evans (1993) found that brakes and friction, improved braking, and brake lights are important factors in determining accident risk. The "model year" of the vehicle was also considered in a study carried out by Chirachavala and Cleveland (1985). El-Herraoui and Saccomanno (1997) studied the relationships between truck accident rates, truck configurations and load status for different highway types in Ontario. Braver et al. (1997) studied the relationships between truck configurations and accident risk in the state of Virginia. They concluded that there is no overall increase in accidents observed among tractors pulling two trailers relative to the ones pulling one trailer. However, they stated that more research is needed to investigate the relationship between truck configuration and accidents under different road surface conditions.

Saccomanno et al. (1993) investigated driver fatigue and its effect on truck accident rates; it was found that higher levels of driver fatigue resulted in higher truck accident rates. Jovanis and Chang (1989) also found that hours of service of the driver had an effect on accident risk. Hauer et al (1991) studied the relationships between accidents and driver traits. They found that using a statistical model with covariates, such as convictions and accidents, is much better than using a "demerit point" system in identifying high risk drivers. In the "demerit point" system, points are assigned to offenses on the basis of their perceived seriousness. In this study, convictions are found to be relevant factors in estimating future accidents. Similar results concerning the "demerit point" system in other jurisdictions were concluded in other studies, by Haight (1964), Burg (1970), Peck et al. (1971), Stewart and Campbell (1972), Harano et al. (1975), Chipman and Morgan (1976), and Peck and Kuan (1983).

The objective of all the previously mentioned studies was to understand the factors that affect accidents and that are related to the highway, environment, vehicle, and driver, in order to improve safety on highways. Few studies have investigated the relationship between carrier characteristics and accident involvements. A study by Corsi et al. (1984) on 1325 carriers found that a disproportionate share of truck accidents occurred in good weather conditions and involved defective vehicles, older trailers, and drivers under 30 years of age. These results indicate that changing the perspective of accident research to the carrier provides a more in-depth understanding of truck accident involvement. Patten et al. (1989) recommended that carriers should be held ultimately responsible for the quality of the vehicles and the drivers that they employ. As a consequence, this recommendation may provide enforcement agencies with an efficient method to induce the trucking industry to increase its overall safety efforts.

To investigate truck accidents from the carrier perspective, the attributes of carriers relevant to accidents, need to be identified. Only a full safety audit could provide that kind of comprehensive data. However, such a comprehensive data for *all* carriers is not available. Safety audits are costly and, as a consequence, a very small percentage of the total carrier population is audited every year. Hence, some studies have been carried out using alternative, less expensive approaches, such as *roadside inspection* and *Roadcheck*, to investigate carriers.

Some researchers such as Patten et al. (1989) and Saccomanno et al. (1997a, 1997b) have developed a framework to estimate accident risk potential for carriers using roadside inspection. Other researchers, such as Savage and Moses (1992, 1994), have developed detailed statistical models based on variables in comprehensive safety audits. A number of concerns associated with existing carrier-based accident analyses is discussed in the following section.

1.3 Previous Carrier-Based Studies and Concerns

Patten et al. (1989) studied the various causes of heavy truck accidents and the efficacy of roadside inspections in reducing accidents. Their study used roadside inspections as a surrogate for estimating carrier accident risk potential; it concluded that driver-related violations are most often reported as the cause of accident, while most of the road inspections focus on the vehicle deficiencies. According to the Bureau of Motor Carrier Safety (BMCS, 1985; U.S. Dept. of Transportation, 1986), more than 80% of out-of-service violations are vehicle related compared to less than 20% of driver out-of-service violations, although more than 80% of the truck-related causal factors are attributed to the truck drivers. To reduce the number of accidents, Patten et al. recommended the adoption of a methodology for identifying significant causes of truck accidents and then removing vehicles and drivers with these defects from the highway. In a similar study, using the 1995 Ontario accident and Roadcheck databases, Saccomanno et al. (1997a, 1997b) studied the link between Roadcheck and carriers' accident risk potentials. The Roadcheck survey is carried out annually throughout North America, over the same 72 hours period. The locations of Roadcheck inspection stations reflect the points of high truck traffic volume on the network. Trucks are chosen randomly at each inspection station. The analysis suggested a link between Roadcheck and carriers' accident risk potential, especially for truck accidents that are caused primarily by mechanical defects. The study showed that roadside inspection could provide a simple way of identifying high risk carriers.

However, there are concerns associated with the Roadcheck approach (Saccomanno et al., 1997). For example, there is a location bias; in Roadcheck, inspection stations are located on major highways (e.g. Highway 401 in Ontario) and therefore do not represent the entire highway network. A fleet size bias also exists; carriers with large number of trucks are most likely to be inspected. There is also a bias towards recording more vehicle violations than driver violations. As a result, roadside inspections represent only a portion of the accident risk analysis of all carriers and all highways.

Moses and Savage (1992, 1994) investigated the effectiveness of motor carrier safety audits in reducing accidents and enhancing safety. The study aimed at estimating carrier accident risk potential from the attributes of those carriers, using a statistical model. A database of 13,000 firms was analyzed in 1992. In 1994, they increased the database to 75,500 firms. A standard questionnaire format, which included 75 questions, was used. The questions varied from driver qualifications to maintenance, reporting of accidents, hours of service, etc. The analysis linked accident rates to carrier attributes using multivariate regression models.

The 1992 analysis by Moses and Savage showed that some aspects of the safety audit (e.g. financial responsibilities, compliance to driving regulations, most maintenance questions) have little to do with the accident performance of firms when they are out on the road. The study also showed that drivers' hours of service and the employment of a safety director were important issues. It states that although the audits could improve the safety of the worst carriers, the probability of inspection is small due to the random selection of the firms to be audited in any given year.

Moses' and Savage's second study, in 1994, confirmed their earlier results from 1992 and also found that the accident rates decline with firm size and not with firm age. The analysis showed that private carriers have lower accident rates (i.e. 20% lower) than those of for-hire carriers, and that general freight carriers have higher accident rates (i.e. 10% higher) than carriers that specialize in particular commodities. They concluded that keeping records of accidents and taking disciplinary or educational actions are effective in reducing accident rates. The findings concerning the weak relationship between maintenance procedure and accident rates imply that the majority of truck accidents are caused by driver error rather than mechanical failure. One major conclusion of the study is that, to have an efficient safety audit process, high risk carriers should be at the top of the list for safety audits, so that scarce resources can be efficiently used to enhance highway safety. For example, in Ontario, carriers are selected randomly for safety audits. A more efficient selection process than the random one needs to be investigated.

In the previous two studies, Moses and Savage used a very detailed set of data from safety audits. Such data are usually *not* available. For example, in Ontario, safety audits are not available for all carriers. Also, the developed model has more than 50 variables, some of which can be collinear. Thus, in that study, there are problems associated with the availability of the data and with the co-linearity among the 50 variables.

There are concerns regarding the few studies carried out from the carriers' perspective. These concerns include limitation of data availability in the case of a comprehensive safety audit and poor representation of the network and carrier population in the case of Roadcheck or roadside inspection. The carriers that are most at risk of accident need to be identified. We believe that the enhancement in carriers' safety will increase the safety of highways, since, as mentioned earlier in this Chapter, trucks are responsible for 37% of fatal accidents even though they represent only 16% of the total number of vehicles registered in Ontario. In the following section, the objectives and scope of the thesis are defined.

1.4 Objectives and Scope

The main objective of this research is to study the carrier-specific accident risk potential. Why a carrier risk perspective? Unsafe trucking firms (carriers) account for a significant share of trucking accidents. In Ontario, for the years of 1991 to 1995, 100% of truck accidents were experienced by only 7% to 8% of the carriers registered to operate in the province (CVOR and ADS). Clearly, there is a need to identify those high risk carriers, so that they can be targeted for appropriate safety interventions (e.g. fines, licensing restrictions, training programs and/or in-depth safety audits).

The objectives of this study are as follows:

1. Develop a model to estimate accident risk potential of carriers knowing their traits and accident history.
2. Evaluate different models using different historical data inputs in order to recommend the best combination of historical data with the statistical model.

3. Compare model results to different selection processes (including the random selection currently used in Ontario to administer safety audits) in identifying high risk carriers.

4. Recommend guidelines for identifying high risk carriers in Ontario.

The scope of this study can be established by the following restrictions:

1. The study is based on data sets made available by MTO, (i.e. CVOR & ADS of Ontario, for the years of 1991 to 1995).

2. Only police reported truck accidents on Ontario highways are considered.

3. Only carriers' exposure on Ontario highways is considered.

1.5 Organization of the Thesis

The thesis consists of seven Chapters and two appendices.

Chapter 2 provides an introduction of the available data.

Chapter 3 provides a definition of the carrier's accident risk potential and a discussion of the relevant factors affecting risk as well as a preliminary analysis of the available data.

Chapter 4 presents different statistical models to explain the variation in accidents among carriers. The results and conclusions from linear and non-linear statistical models are provided.

Chapter 5 introduces the Empirical Bayes approach and a model is developed to account for the extra Poisson variation.

Chapter 6 compares the Bayes model with 4 different approaches of identifying high risk carriers and the results and conclusions of these comparisons are provided.

In Chapter 7, the major findings and conclusions of the thesis are summarized and recommendations for future work are discussed.

Appendix A provides more analysis on detentions and convictions.

Appendix B provides some results of another Bayes model using only the previous 2 years of detentions and convictions.

Appendix C provides some samples of the programs written for the analysis. This include sample of the GLIM program for the linear model, the Poisson model, the Bayes' model ranking.

Chapter 2

Available Data

Any model development relies and depends upon the available data. As a consequence, the available sources of data need to be discussed in order to identify the variables than can be introduced in the model. Models are limited by the number of available variables in the data. From a carrier perspective, certain variables are thought to be relevant to the accident risk potential of carriers. These variables could be characteristics of drivers employed by the carrier, educational programs for drivers, maintenance programs for vehicles, hiring and firing procedures for drivers, detentions, convictions, fleet size, etc.; but a full list of "all" these variables is not available for all carriers.

Two of the major tasks of any statistical study is the availability and cost of the data. In this study, existing Ontario databases are used. This means that there is no major extra cost in collecting more data for the proposed model. In other words, the model represents an efficient way of using the existing databases. A supplementary survey, carried out by the University of Waterloo, is part of the relevant databases as well.

This Chapter introduces the database that is used in this thesis to investigate the accident risk potential of carriers. The following data sets will be used:

1. Carrier Vehicle Operator Record (CVOR) in Ontario.
2. Ontario Accident Data System (ADS).
3. A complementary survey carried out by the University of Waterloo.

In the following sections, each of the data sets mentioned above is discussed in more detail.

2.1 CVOR Database

The CVOR database was made available by the Ministry of Transportation of Ontario (MTO). The database provides registration information (name, address, company officials, nature of business, size, etc.) for all carriers registered to operate in Ontario. The database includes data from the years of 1991 to 1995. A unique *cvor* reference number links each carrier to its fleet size and accident history, as well as its detention and conviction history. The CVOR data sets consists of separate data for each of the attributes (e.g fleet size, convictions, detentions) and for each year (i.e 1991 to 1995). All data was integrated in one single database using the CVOR reference number. The database consists of the traits or attributes of all registered carriers (around 95,000 carriers) in Ontario, including those based outside of the province. The attributes of carriers can be summarized from the Carrier Vehicle Operator Record (CVOR) data set, for the years 1991 to 1995, as follows:

- Fleet Size: indicates the number of trucks operated by the carrier. The fleet size ranged from 1 to more than 3000 trucks for carriers. Most carriers (i.e 93% of carriers), however, have a fleet size smaller than or equal to 10.
- Detention: indicates any roadside inspection of a vehicle where a defect was found and the vehicle was detained until the defect was corrected. There were approximately 10,000 detentions every year for the period of 1991 to 1995.
- Conviction: indicates a conviction for any transportation offense. Convictions may be laid against the driver, operator, or plate/vehicle owner. There were around 33,000 convictions every year for the period of 1991 to 1995.
- Carrier Age: as determined from the first day of registration. This attribute was missing for many carriers. As a consequence, the carrier age is not considered as one of the variables in the model.

2.2 ADS Database

The Ontario Accident Data System (ADS) includes information on vehicle accidents taking place on the provincial highway network assembled by MTO each year. The ADS database consists of all truck accident records reported by police on Ontario provincial highways. The database that was made available for this study includes five years of truck accidents covering the period of 1991 to 1995. Around 15,000 accidents involved trucks every year. A dictionary explaining the information and the variables in the ADS database is available from the Ministry of Transportation (MTO) of Ontario. The database includes information about each accident, such as the highway or location of the accident, the number of vehicles involved, the types of vehicles, the driver's age and driver's gender.

2.3 Complementary Survey

A carrier survey was carried out by the University of Waterloo (Saccomanno et al., 1997) to provide information on carrier safety practices and programs, and to solicit opinions and suggestions on various issues affecting truck safety in Ontario. The survey also provided a check against the accuracy of fleet size and percentage of travel in Ontario as reported in the CVOR databases for the same carriers as in the survey sample. One of the survey's main purposes was to investigate if the fleet size of carriers can be used as a surrogate of their mileage traveled in Ontario, since exposure is not directly available in the CVOR data. The survey consisted of 80 questions, ranging from CVOR number to Ontario exposure. The survey was sent to 208 carriers; only 89 carriers (32 %) responded to all the questions.

2.4 Summary

The available data needs to be used efficiently to obtain an estimate of carrier accident risk potential. In establishing the model framework, the available data must be considered.

Three data sets that are used in this research, were described in this Chapter. The data consisted of the CVOR and ADS databases for the period of 1991 to 1995, as well as a complementary carrier survey carried out by the University of Waterloo.

The data used in the thesis has, like all available data, its limitations in terms of explanatory variables. As a result, the model will be developed in order to efficiently use the existing available data.

Chapter 3

Factors Affecting Risk

3.1 Definition of Carrier Risk

In this thesis, the accident rate (i.e. accident/exposure/year) will be the criteria of ranking carriers and accordingly identifying the high risk ones. In this way, a comparison between carriers of a different fleet size (i.e. exposure) could be made. As mentioned in the previous Chapter, there is no available data on the exposure (i.e. kilometers traveled) of each carrier. But from the supplementary survey carried out by the University of Waterloo, as shown in Figures 3.1 and 3.2, the fleet size was found to be relatively highly correlated (i.e correlation factor of 0.75 and 0.67) with the exposure of the carriers on Ontario highways every year. As a result, the fleet size can act as a surrogate of the exposure of each carrier, and the criteria of ranking carriers (i.e accident rate) is expressed in terms of fleet size as "accident/truck/year". The accident frequency (i.e accident/year) for each carrier is estimated, then the accident rate (i.e. accident/truck/year) is calculated to rank the carriers in order to identify the high risk carriers.

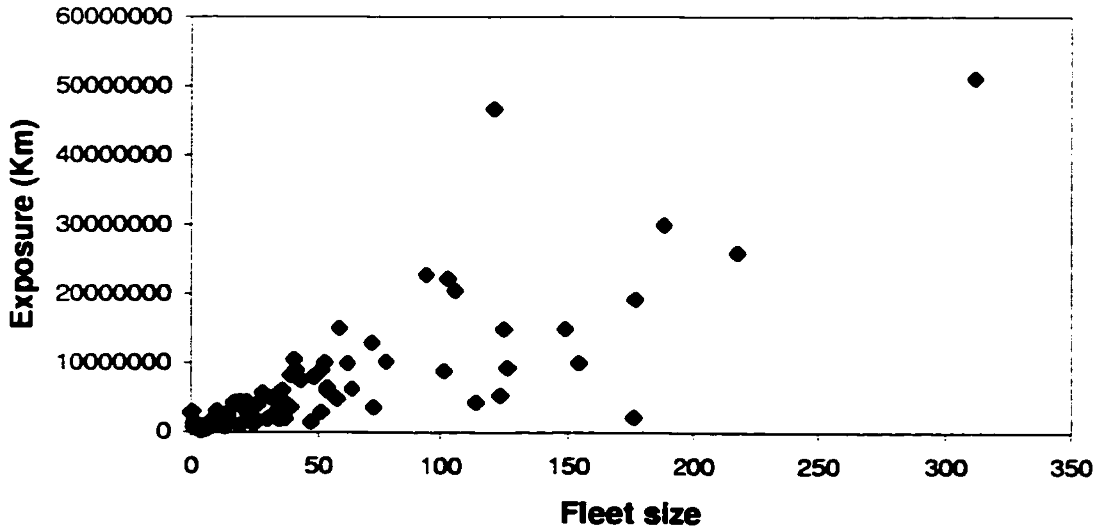

Figure 3.1: Correlation Between Exposure and Fleet Size of Carriers

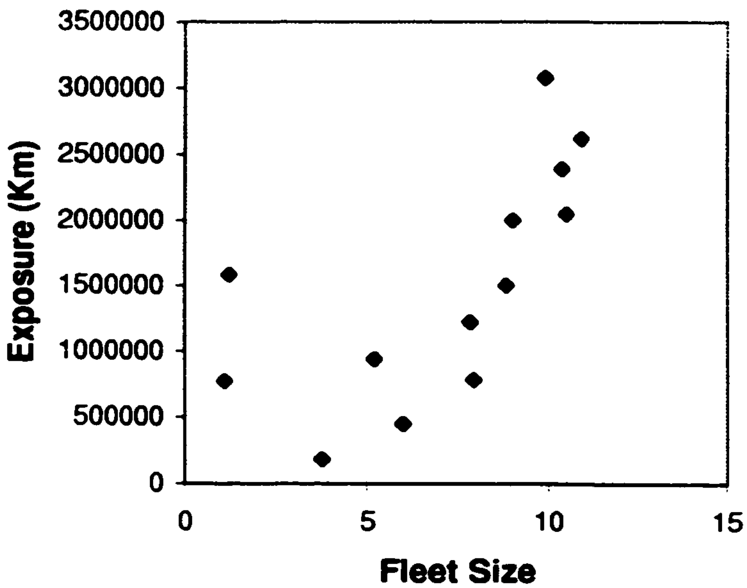

Figure 3.2: Correlation Between Exposure and Fleet Size of Small Carriers

3.2 Relevant Factors

Many factors can be considered relevant when explaining variations in the accident frequency of carriers. These factors could include: age of driver, age of vehicle, detention. conviction, fleet size, accident history, and more. Unfortunately, such data is not available. Thus, analysis can only be carried out using the existing *available* data described in the previous Chapter. Available factors that are used in this study are:

- Accidents for the years of 1991 to 1995
- Detentions for the years of 1991 to 1995
- Convictions for the years of 1991 to 1995
- Fleet size

These factors are considered relevant in explaining variations in the accident frequency of carriers. The accident frequency in previous year(s) could be an indicator of the accident frequency in the following year(s). Some studies suggest (Hauer 1987), though. that what has happened in the past is not necessiraly an indication of what will happen in the future. Also, the number of detentions per year of a carrier could be an indication of the level of maintenance of its fleet or an indication of its drivers' behaviours on the road. The number of convictions per year for a carrier could also be another indication of the level of compliance of its fleet and its drivers to follow regulations in order to avoid accidents.

The fleet size could be a factor in two ways. First, as mentioned in the previous section, fleet size is a surrogate of the exposure. On the other hand, the accident rate (i.e accident/truck/year) could be also a factor of the fleet size. In other words, carriers that have a small fleet size could have a higher or lower accident rate (i.e accident/truck/year) than those of carriers that have a large fleet size. This possibility could indicate that small and large carriers have different maintenance practices (e.g. less maintenance per kilometer), different driver's educational programs, and different ways of operating (e.g. operating fewer trucks). The effect of fleet size on accident rate will be investigated in the next section.

3.3 Preliminary Analysis and Fleet Size Effect

In Ontario, MTO classifies carriers into 8 categories, as shown in Table 3.1. The number of carriers varies from one category to another. The distribution of the number of carriers per each category is shown in Figure 3.3. As shown in Figure 3.3 and Table 3.2, there is a large difference between the first 2 categories and the remaining 6 categories. The number of carriers in the first 2 categories represents the bulk of the population (i.e. 57% & 36%). In this study, the first 2 categories will be combined and dealt with separately from the other 6 categories, in order to investigate if there is a significant difference between these 2 populations.

As shown in Table 3.3, the carriers from categories 1 and 2 (carriers with fleet size of 1 to 10 trucks) represent 93% of the population. In other words, small carriers represent the bulk of the population of carriers. On the other hand, as shown in Table 3.4, small carriers represent 30% of the truck population. Also as shown in Table 3.5, around 30% of the total accidents are experienced by small carriers with a fleet size between 1 and 10 trucks inclusive. This means that if 2 imaginary very large carriers are considered, one having a fleet size of all trucks of small (i.e category 1 and 2) carriers and the other a fleet size of all trucks of large carriers (i.e category 3 to 8), the two imaginary carriers would have the same accident rate. But this is not the case; we do not have 2 very large imaginary carriers.

From the literature of accident analysis, the accident rate is calculated by dividing the yearly number of accidents by the number of kilometers traveled in that year. In other words, the accident rate is expressed in terms of "accident per million truck-kilometer per year" (i.e accident/exposure/year). This accident rate is the one used in the thesis, where fleet size acts as a surrogate of exposure. Thus, the accident rate for each carrier is calculated by dividing the yearly number of accidents by the number of trucks in the fleet. The focus in the thesis is the accident rate for each carrier. In other words, if a carriers has "a" number of accidents and "x" number of trucks, the accident rate would be "a/x". To explain this further, consider 2 carriers , the first has "a" number of accidents and "x" number of trucks, the second carriers has "b" number of accidents and "y" number of trucks. So the accident rate for the first carrier is "a/x" and for the second carrier is "b/y". Then, the mean of the accident rate for both carriers would be $(a/x + b/y)/2$, where 2 is the number of observations. This is totally different from calculating the accident rate for both carriers as $(a+b)/(x+y)$. Statistically, the latter calculation is not the mean of the accident rates of the 2 carriers but it is an overall average of all carriers. It is not carrier specific (i.e the latter rate does not show the variation in accident rate among carriers). Nevertheless, It can be shown mathematically that the first calculation of rates is different from the second way of calculating the rate. To further illustrate that, consider the following example of 2 carriers where:

a: number of accidents in carrier 1

b: number of accidents in carrier 2

x: number of trucks in carrier 1

y: number of trucks in carrier 2

When is the "thesis rate" greater than or equal to the "other rate"?

$$\frac{\frac{a}{x}+\frac{b}{y}}{2} \geq \frac{a+b}{x+y}$$
$$ay^2+bx^2 \geq xy(a+b) \tag{3.1}$$

It can be shown that:

If $a=b$, the "thesis rate" is always greater than or equal to the "other rate"

If $a=0$ and $x>y$, the "thesis rate" is always greater than to the "other rate"

If $a=0$ and $x<y$, the "thesis rate" is always smaller than the "other rate". This case could explain why the mean of the accident rates (i.e thesis rate) of small carriers is smaller than the overall average rate (i.e other rate) because many of the small carriers have experienced zero accidents.

In fact, there are around 95,000 carriers and each has a different accident rates. Furthermore, the mean accident rates will be shown to be different for small (i.e. fleet size of 1 to 10 trucks) and large (i.e. fleet size more than 10 trucks) carriers.

The accident rate per truck is calculated for each individual small carrier and the mean as well as the variance of these accident rates per truck are then calculated. As shown in Table 3.6, the accident rate for small carriers (i.e. category 1 & 2 or carriers with a fleet size of 1 to 10 trucks) is found to be between 0.023 and 0.026 (accident/truck/year), with a standard deviation between 0.15 and 0.17. The same calculations are carried out for large carriers (i.e. category 3 to 8 or carriers with a fleet size more than 10 trucks). As shown in Table 3.7, for the years 1992 to 1995, the accident rate per year per truck for large carriers is found to be between 0.029 and 0.034, with a standard deviation between 0.064 and 0.07.

Looking at Tables 3.6 and 3.7, one can see that the mean as well as the standard deviation are stable over time and that the standard deviation is larger than the mean for both populations (i.e small and large carriers). But if the coefficient of variations (i.e $\frac{std. dev.}{mean}$) of both populations are compared, it can be concluded that the coefficient of variation of small carriers is around 3 times larger than that of large carriers as shown in Table 3.6 and 3.7. This could be an indication that the accident rates per truck for small carriers are more dispersed than those of large carriers.

Table 3.1: Carrier Distribution by Categories Adopted by MTO with Respect to Fleet Size

MTO Category	Fleet Size
1	1
2	2 to 10
3	11 to 25
4	26 to 100
5	101 to 500
6	501 to 1500
7	1501 to 3000
8	3000+

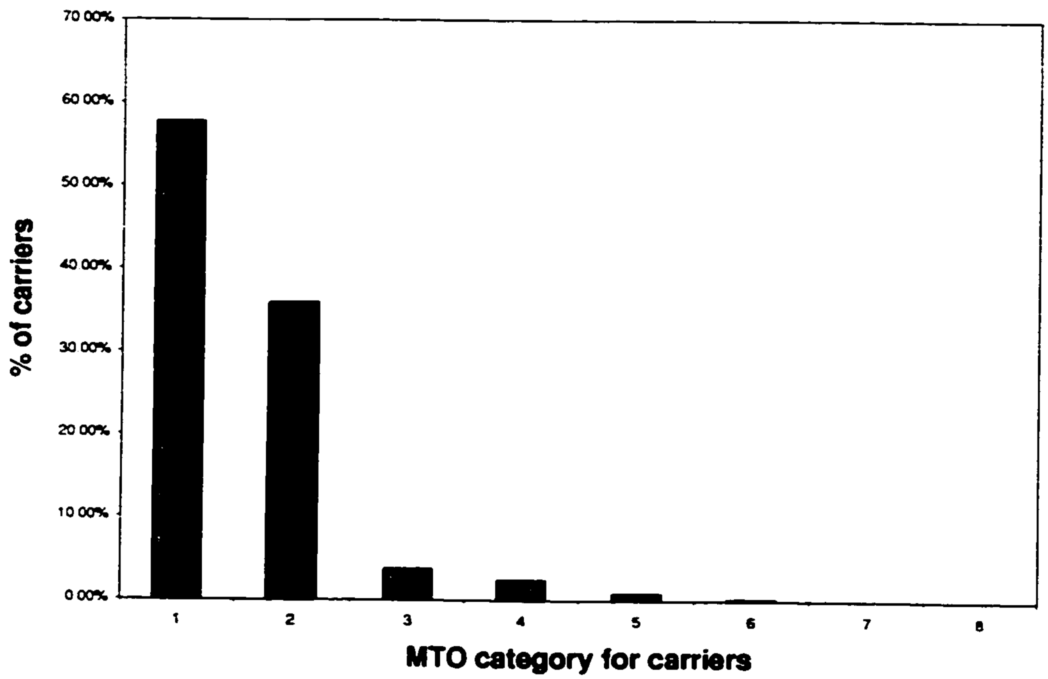

Figure 3.3: Histogram of Carriers Classified by MTO

Table 3.2: Number of Carriers by Categories Adopted by MTO

MTO Category	Fleet Size	% of Total Number of Carriers
1	1	57.60%
2	2 to 10	35.73%
3	11 to 25	3.58%
4	26 to 100	2.26%
5	101 to 500	0.70%
6	501 to 1500	0.10%
7	1501 to 3000	0.02%
8	3000+	0.01%
		100.00%

Table 3.3: Fleet Size and Number of Carriers

Fleet Size	No. of Carriers	% of Carriers	Average No. of Trucks
1-10	88061	93	2
more than 10	6291	7	71
	94354	100	

Table 3.4: Fleet Size and Number of Trucks

Fleet Size	No. of Trucks	% of Trucks
1-10	173935	28
more than 10	447326	72
	621261	100

Table 3.5: Percentage of Accident for Small Carriers (1 to 10 trucks) Compared to All Carriers' Accidents

Year	No of Accident	% of Accidents by Small Carriers
1995	4460	29%
1994	4287	28%
1993	4819	30%
1992	4838	31%

Table 3.6: Mean and Standard Deviation of Accident Rate (accident per truck per year) for All Carriers With 1 to 10 Trucks

Year	92	93	94	95
Mean Accident Rate [a]	0.0258	0.0263	0.0252	0.023
Std. Dev.	0.1598	0.1693	0.1583	0.1468
Std. Dev/Mean	6.19	6.44	6.28	6.38

[a]No. of Observations= 88061

Table 3.7: Mean and Standard Deviation of Accident Rate (accident per truck per year) for All Carriers With More Than 10 Trucks

Year	92	93	94	95
Mean Accident Rate [a]	0.0343	0.0289	0.0328	0.031
Std. Dev.	0.069	0.064	0.073	0.067
Std. Dev/Mean	2.01	2.21	2.23	2.16

[a]No. of Observations= 6291

Testing the difference between the 2 means, using t-statistics as shown in Table 3.8, the null hypothesis that the 2 means are equal is rejected at the 95% as well as at the 99% confidence level. On the other hand, testing the difference between the 2 variances, using F-statistics as shown in Table 3.8, the null hypothesis that the 2 variances are equal, is rejected at the 95% as well as at the 99% confidence level. As a result, the average accident rate per truck and the variance for small carriers are significantly different from those of large carriers. Accident rates are higher for large carriers but with a smaller variance (i.e less dispersal). From the analysis and results, it can be concluded that small carriers (i.e. with a fleet size of 1 to 10 trucks) and large carriers (i.e. with a fleet size more than 10 trucks) belong to different populations with different means and different variances. Also, the significant differences in the variances and coefficients of variation strongly suggest that the accident rates of small carriers are more dispersed.

Also, as shown in Table 3.9, only 4% of the population of small carriers are responsible for all accidents experienced by all small carriers every year. The carriers that experience accidents change significantly from one year to the next. As shown in Table 3.10, for carriers with a fleet size between 1 and 10 trucks (i.e small carriers), 71% to 76% of carriers that had accidents in one year did not have any accidents in the previous year. Also, around 60% of carriers that had accidents in any of the years of 1995, 1994 and 1993 did not have any accidents in the previous 2 years. Also, around 50% of carriers that had accidents in either of the years of 1995 or 1994 did not have any accidents in the previous 3 years.

Thus, identifying high risk carriers is not an easy task, especially since around 75% of those carriers change every year. In other words, only 1% of the carriers' population experience accidents in two consecutive years, but those carriers change every 2 years. This fact adds to the difficulty of identifying high risk carriers in the population of small carriers.

On the other hand, for large carriers (i.e. with a fleet size more than 10 trucks), as shown in Table 3.11. 30% of the total population of large carriers is responsible for 100% of all accidents experienced by large carriers. This 30% compares to only 4% in the case of small carriers. Also, as shown in Table 3.12, every year only 25% of the large carriers that experienced an accident in one year did not have any accidents in the previous years, compared to 75% in the case of small carriers. Again, these results show the difference between small carriers (fleet 1 to 10 trucks) and large carriers (fleet more than 10 trucks) and strongly suggest that they belong to different populations. In the following Chapters. small and large carriers will be investigated separately.

The analysis and results indicate that for small carriers, variables other than accident frequency are needed to explain the variation in accidents. Those variables could be the detentions and convictions of carriers. Preliminary analyses were carried out on the detentions and convictions of small carriers. As shown in Tables 3.13 and 3.14, small carriers are responsible for 45% to 49% of all detentions and 53% to 58% of all convictions. although they represent only 30% of the truck population. These results suggest that. for small carriers. detentions and convictions could be relevant factors in the variation of accidents.

In this thesis, when reference is made to *small* carriers, all carriers that have a fleet size of 1 to 10 trucks are meant. When reference is made to *large* carriers, all carriers that have a fleet size of more than 10 trucks are meant.

Table 3.8: Testing Significant Difference Between Small and Large Carriers' means and variances

Year	92	93	94	95
t-statistics for 2 Means	-8.07	-2.35	-7.25	-8.23
F-statistics for 2 Variances	13.79	17.99	12.09	12.34

Table 3.9: No. of Carriers (With a Fleet Size of 1 to 10 Trucks) That Have Accidents Each Year

Year	No of Carriers Having Accidents	% to Total No. of Small Carriers
1995	3513	4.0%
1994	3382	3.8%
1993	3869	4.4%
1992	3935	4.5%

Table 3.10: Change in Small Carriers (1 to 10 trucks) That Had Accidents From One Year to Next

Year	In 1995	In 95 and Not in 94	In 95 and Not in 94 Nor 93	In 95 and Not in 94 Nor 93 Nor 92
No. Of carriers	3513	2581	2072	1816
% of carriers		73%	59%	52%
Year	In 1994	In 94 and Not in 93	In 94 and Not in 93 Nor 92	In 94 and Not in 93 Nor 92 Nor 91
No. Of carriers	3382	2398	1924	1624
% of carriers		71%	57%	48%
Year	In 1993	In 93 and Not in 92	In 93 and Not in 92 Nor 91	
No. Of carriers	3869	2861	2353	
% of carriers		74%	61%	
Year	In 1992	In 92 and Not in 91		
No. Of carriers	3935	2991		
% of carriers		76%		

Table 3.11: No. of Large Carriers (With Fleet Size More Than 10 Trucks) That Have Accidents Each Year

Year	No of Carriers Having Accidents	% to Total No. of Large Carriers
1995	1922	30.6%
1994	1914	30.4%
1993	1911	30.4%
1992	1855	29.5%

Table 3.12: Change in Large Carriers (more than 10 trucks) That Had Accidents From One Year To Next

Year	In 1995	In 95 and Not in 94	In 95 and Not in 94 Nor 93	In 95 and Not in 94 Nor 93 Nor 92
No. Of carriers	1922	496	288	209
% of carriers		26%	15%	11%
Year	In 1994	In 94 and Not in 93	In 94 and Not in 93 Nor 92	In 94 and Not in 93 Nor 92 Nor 91
No. Of carriers	1914	487	300	219
% of carriers		25%	16%	11%
Year	In 1993	In 93 and Not in 92	In 93 and Not in 92 Nor 91	
No. Of carriers	1911	511	302	
% of carriers		27%	16%	
Year	In 1992	In 92 and Not in 91		
No. Of carriers	1855	512		
% of carriers		27%		

Table 3.13: Percentage of Detentions for Small Carriers

Year	% of All Detentions
1995	49%
1994	45%
1993	47%
1992	48%
1991	48%

Table 3.14: Percentage of Convictions for Small Carriers

Year	% of All Convictions
1995	58%
1994	55%
1993	53%
1992	53%
1991	54%

3.4 Summary

The analyses and results in this Chapter can be summarized as follows:

1. Carriers with a fleet size of 1 to 10 trucks (i.e. small carriers) represent the bulk of the total population of carriers (i.e 93%).
2. There is a significant difference between the mean as well as the variance of the accident rates per truck for small and large carriers. This suggests that small and large carriers are from different populations with different means and variances. As a consequence, in the work presented in this thesis, small and large carriers will be dealt with separately.
3. Every year, 4% of small carriers are responsible for 100% of all accidents experienced by small carriers.
4. Every year, 75% of the small carriers that experienced accidents in one year did not have any accidents in the previous year.
5. Small carriers are responsible for more than 50% of all carriers' detentions and convictions, although they represent only 30% of the truck population. This fact suggests that detentions and convictions could be relevant factors in explaining the variation of accidents.

Chapter 4

Statistical Model

4.1 Rationale of Statistical Model

The aim of a statistical model is to replace the data with a set of fitted values derived from the model. While a perfect model is not possible, any model must have a certain goodness-of-fit criteria. Theoretically, there is "one" model (sometimes called *full* model) which is a perfect fit with as many parameters as observations. This model is just a replication of the data and does not produce a simple theoretical pattern for the data. However, this perfect or *full* model can be used as a benchmark to compare between all other models. According to McCullagh and Nedler (1989), *"all models are wrong; some, though, are more useful than others and we should seek those... Data will often point with almost equal emphasis at several possible models and it is important that statisticians recognize and accept this."* The simplest model is the "null" model which is a mean of the observed data. However, the "null" model is not usually the best one. Therefore, the aim is to seek a model that has the best fit measured using the goodness-of-fit criteria. In this thesis, we will start with a simple linear regression model. This model will be investigated in the following section.

4.2 Linear Regression Models and Results

One would assume previous years' accident frequency (i.e accident history) as a good indicator of the future accident frequency of a carrier. A simple linear model could be a good model in such a case. Thus, some linear regression models will be developed to explain the variations in the accident frequency for large as well as for small carriers. In these models, as only 5 years of data are available, the independent variables will be the accidents for the period of 1991 to 1994, while the dependent variable will be the accidents in the year of 1995. First, a linear regression model is developed for large carriers that have a fleet size of more than 10 trucks. The linear model uses 4 years of accident history to estimate the fifth year. In other words, the model can be expressed as follows:

$$a5 = \beta_1 a1 + \beta_2 a2 + \beta_3 a3 + \beta_4 a4 \tag{4.1}$$

where

$a1, a2, a3, a4$: Independent variables or observed accidents per year for carriers in 1991, 1992, 1993 and 1994.

$\beta_1, \beta_2, \beta_3, \beta_4$: are the paramters.

$a5$: Dependent variable or estimated accidents frequency in 1995.

The parameter estimates and the t-statistics test are shown in Table 4.1. All parameters were found to be statistically significant at the 95% confidence level. The result of this model shows an R-square value of 0.97. This statistic suggests that accident history in the years of 1991 to 1994 may possibly explain the number of accidents in 1995. The value of "t-statistics" for the coefficient of "a1" (i.e. accidents in 1991) was found to be 1.69 (slightly larger than 1.645 from the t-statistics table at the 95% confidence level). Another linear model was developed excluding the variable "a1". The parameter estimates and the t-statistics test of this new model are shown in Table 4.2. Again, all parameters were found to be statistically significant at the 95% confidence level and the model has an R-square value of 0.97. This fact suggests that for large carriers (i.e. more than 10 trucks) the accident history is a good indicator of the accident frequency, which could be used as a ranking criterion for those carriers. In other words, the fluctuation and variation could be explained to a large extent using the previous 3 years of accident history.

Table 4.1: Parameter Estimates of Linear Regression for Large Carriers (More Than 10 Trucks) Using Previous 4 Years Accident History

Parameter	Coefficient	t-statistics [a]
1	0.1401	5.15
a1	0.02272	1.69
a2	-0.1038	-6.50
a3	0.3092	21.52
a4	0.6915	58.11

[a] t-critical=1.645

Table 4.2: Parameter Estimates of Linear Regression for Large Carriers (More Than 10 Trucks) Using Previous 3 Years of Accident History

Parameter	Coefficient	t-statistics
1	0.1406	5.17
a2	-0.08528	-7.33
a3	0.3133	22.11
a4	0.692	58.15

A similar linear model, using the accident history for the period of 1991 to 1994 as dependend variables and the accidents in 1995 as independent variables, is developed for small carriers (1 to 10 trucks fleet size). The model in this case produced an R-square of 0.16, which suggests that for small carriers the accident history is not sufficient in estimating the accident frequency nor could it be used as ranking criterion for small carriers.

Two more linear models were developed for small carriers of 1 to 10 truck fleet size using the accident frequency in 1995 as the dependent variable and fleet size, detentions, and convictions of the previous 4 years as independent variables. The results show that, for small carriers, the R-square was found to be 0.11, which may possibly indicate that fleet size, convictions and detentions are as important a factor as the accident history. As a consequence, the two clues can be used to produce a better model for small carriers. When using the two clues in a third linear model (i.e. accident history and detentions, convictions and fleet size), as shown below:

$$a5 \quad = \quad f(a1, a2, a3, a4, d1, d2, d3, d4, c1, c2, c3, c4, fleet) \qquad (4.2)$$

where

$a1, a2, a3, a4$: observed accidents per year for carriers in 1991, 1992, 1993 and 1994.

$d1, d2, d3, d4$: detentions per year for carriers in 1991, 1992, 1993 and 1994.

$c1, c2, c3, c4$: convictions per year for carriers in 1991, 1992, 1993 and 1994.

$fleet$: fleet size for carriers.

$a5$: Dependent variable or estimated accidents frequency in 1995.

R-square was found to be 0.18.

The analysis in this section demonstrates the difference between carriers with small fleet size (i.e. 1 to 10 trucks) and ones with large fleet size (i.e. more than 10 trucks). For large carriers, a linear model using the previous 3 years of accident history seems to be a good model for estimating accident frequency in the future. The accident rate (i.e. accident/year/truck) can then be calculated to rank large carriers. On the other hand, the above results lead to the conclusion that small carriers, which represent 93% of the carrier population, cannot be ranked based only on the accident history. Also, the linear model does not adequately explain the variation in accident frequency even when including the other attributes (i.e conviction, detention, fleet size). Since small carriers represent 93% of the total population of the carriers, it is extremely important to develop a better model to identify high risk carriers for the population of small carriers. These results seem to be consistent with those in Chapter 3, where 75% of carriers that have accidents in one year did not have any accidents in the previous year. The results imply that for small carriers, accident history might not be the primary attribute to explain the variation. Therefore, other attributes, such as detention and convictions history, might provide inference for future accident behaviour. In the next section, non-linear models for small carriers will be developed.

4.3 Non-Linear Model Development

4.3.1 Model Selection

The proposed non-linear model is selected to be in the following form:

$$\begin{aligned} \theta &= e^{(x^T\beta)} \times fleet \\ &= AR \times fleet \end{aligned} \tag{4.3}$$

where

θ : is the mean of a Poisson (or other) distribution describing the accident frequency (i.e accident per year) for carriers.

β : are the parameters.

x : are the carrier attributes or the independent variables.

x^T : is the transpose of the vector $\boldsymbol{x}$.

AR: is the accident rate (i.e. accident per truck per year).

$fleet$: acts as a surrogate for exposure

This model form ensures that accident-involvement is always non-negative. This type of rate function has been widely used in the statistical literature and found to be very flexible in fitting different types of data (e.g. Cox and Lewis 1966; Cameron and Trivedi 1986: Frome et al. 1990).

It is necessary to mention here that the data from the year of 1995 will not be part of the model development. Instead, it will be used at the end to check and test the model by comparing the observed accident frequency in 1995 with the estimated one from the model. As a result, the dependent variable will be the number of accidents in 1994 (i.e. a4) and the independent variables will be :

- detentions in the previous three years (91,92,93) or d1,d2,d3
- convictions in the previous three years (91,92,93) or c1,c2,c3
- fleet size

A probability distribution that represents the accident events needs to be used. According to many researchers, the number of accidents in a single year is assumed to follow a Poisson distribution. A Poisson regression model has been used in many studies; by Hauer et al. (1987). Joshua and Garber (1990), and Miaou and Lum (1993). The Poisson model was found to be better than the conventional linear regression model, which was used in different studies (e.g. Zeeger et al. (1987,1990), Okamoto and Koshi (1989), Miaou and Lum (1993)). According to Jovanis and Chang (1986), Saccomanno and Buyco (1988), and Miaou and Lum (1993), the linear regression model cannot adequately describe the random and discrete vehicle accident events. Joshua and Garber (1990) and Miaou and Lum (1993) have shown that the Poisson regression can better explain the variation in the accidents than the conventional linear regression. As a result, Poisson distribution is suggested for use in estimating the parameters. Nevertheless, after developing the model, the Poisson distribution hypothesis will be tested.

4.3.2 Estimation of Parameters

After selecting the model, the parameters of the statistical model could then be estimated using different methods. The moment estimate was used by Breslow (1984), Lawless (1987), and Hauer (1992) and the quasi-likelihood method was used by McCullagh and Nedler (1989), while the generalized least squares method was used by Carrol and Ruppert (1989). An iterative regression-based estimation method was suggested by Cameron and Trivedi (1986, 1990), and the maximum likelihood method was used by Cramer (1986). The estimated parameters from the quasi-likelihood and the generalized least squares methods converge to those of the maximum likelihood method as more iterations are used (Miaou, 1994). Thus, the parameters (β) will be estimated using the maximum likelihood technique (McGullan and Nedler, 1989). The maximum likelihood technique was applied and the coefficients were estimated using the statistical software package GLIM. GLIM is short for Generalized Linear Interactive Modeling. GLIM is a program designed to facilitate the fitting of Generalized Linear Models (i.e. GLM).

Many useful statistical models prove to be special cases of Generalized Linear Models. GLIM has a simple but powerful interpretive language suitable for complex analytical techniques. The estimates of the parameters are shown in Table 4.3 and all coefficients of the parameters are found to be significant at the 95% level of confidence as shown in Table 4.3.

Table 4.3: Testing Statistical Significance of Coefficients Using t-statistics

Parameter	Estimate	t-statistics
1	-3.914	-150.19
FLEET	0.03906	7.35
D1	-0.1818	-6.81
D2	0.1346	5.72
D3	0.1463	8.72
C1	0.05795	4.75
C2	-0.01295	-1.99
C3	0.0425	8.39

Looking at the parameter estimates in Table 4.3, one can see that the accident rate increases with the fleet size as well as all detentions (except d1 or detentions in 1991) and convictions (except c2 or convictions in 1992), as expected. But the coefficients of d1 (i.e detention in 91) and c2 (i.e. conviction in 92) have a negative sign, which implies that the accident rate decreases when d1 and c2 increase. This could be due to the fact that d1 and d2 as well as c1 and c2 were found to be highly correlated. So another model excluding the variables d1 and c1 was developed (i.e independent variables are fleet,d2,d3,c2,c3). The parameter estimates of this new model are shown in Table 4.4. As shown in this table, all coefficients have a positive sign, which is expected as detentions and convictions increase with the accident frequency. Also, as shown in Table 4.4, the coefficient of d2 was found to be not significant at the 95% confidence level, so a third model excluding d2 (i.e. independent variables are fleet, d3,c2,c3) was developed to determine the significance of the improvement of the model when including d2. Another subset of the model using the fleet, d3 (i.e detentions in 1993) and c3 (i.e. convictions in 1993), was developed to determine if there is a significant improvement to the model when including more variables. For each developed model, the parameters were estimated in order to maximize the likelihood (or the log-likelihood) of the data observed. In the following section, all the developed models will be compared, using the *likelihood ratio* (i.e. the goodness-of-fit criteria used here), to find the *best* model.

4.4 Comparison of Models Using the Likelihood Ratio (i.e. Scaled Deviance)

Each of the developed models in the previous section has a maximum log-likelihood value. One can consider a maximum log-likelihood achievable for an exact fit, where the fitted values are equal to the observed data. Then, any model can be compared with the exact or perfect model by comparing their log-likelihood values. The smaller the difference of log-likelihoods (i.e. ratio of likelihoods), the closer the model is from the "perfect" model. This difference between log-likelihoods is called the "scaled deviance". The scaled deviance is χ^2 distributed so the models can be compared to investigate the statistical significance of adding or omitting any parameter using the difference in the scaled deviance.

As shown in Table 4.5, the difference between the scaled deviances of the model that contained all 7 variables (i.e. fleet,d1,d2,d3,c1,c2,c3) and all other subset models was found to be significant at the 95% confidence level. The results, as shown in Table 4.5, demonstrate that the model (i.e. fleet,d1,d2,d3,c1,c2,c3) is significantly better (i.e. closer to the *perfect* model) than all other models, although d1 and c1 are highly correlated with d2 and c2, respectively.

4.5 Extra Poisson Variation

After the development of the statistical model, a test was carried out to determine if the data observed was not significantly different from the assumed Poisson distribution. The test of Fisher (1954) was used to evaluate whether or not there is a significant difference from Poisson distribution. As shown in Table 4.6, the observed data is statistically significant from Poisson (critical value is 1.645) at the 95% confidence level (according to Fisher (1954), the term $\sqrt{(2\chi^2)} - \sqrt{(2n-1)}$ is normally distributed for higher value on "n") . Also, when conducting the same test on other developed models that include the accident history as independent variables, the results indicate that, as shown in Table 4.6, there is still a significant difference from the Poisson distribution. Thus, even when including the accident history in the statistical model, there is an indication of an extra-Poisson variation.

Many studies by Dean and Lawless (1989), Collings and Barry (1985), Breslow (1984), Cox (1983), and Hauer and Persaud (1991) have shown some form of the variance to account for the extra-Poisson variation. The variance shown in Equation 4.4 is the one many of these studies have used.

$$Var(h) = E(\theta) + k \times (E(\theta))^2 \tag{4.4}$$

where

$E(\theta)$ is the expected value of θ.

The Poisson variance is a special case of the variance shown in Equation 4.4 (when k=0).

An Empirical Bayes approach was used in different studies by Hogg and Craig (1978), Robbins (1980), Hauer (1992) nd Nembhard and Young (1995) to account for this variance with an extra-Poisson component. Empirical Bayes methodology will be introduced in the next Chapter in order to enhance the Poisson statistical model.

Table 4.4: Testing the Nested Model for Statistical Significance of Coefficients Using t-statistics

Parameter	Estimate	t-statistics
1	-3.913	-150.27
FLEET	0.04021	7.59
D2	0.01387	1.04
D3	0.1171	6.33
C2	0.006537	1.34
C3	0.05096	10.64

Table 4.5: Comparison of Developed Models Using Scaled Deviance

Model	Scaled Deviance	Delta Scaled Deviance [a]	Delta D.o.f
"fleet,d1,d2,d3,c1,c2,c3"	24281		
"fleet,d2,d3,c2,c3"	24331	50	2
"fleet,d3,c2,c3"	24332	51	3
"fleet,d3,c3"	24336	55	4

[a] $\chi^2 - critical$ is between 5.99 and 9.48

Table 4.6: Test of Poisson Using Fisher Test

Model	$\sqrt{(2\chi^2)} - \sqrt{(2n-1)}$ [a]
"d1,d2,d3,c1,c2,c3"	228.37
"d1,d2,d3,c1,c2,c3,a3"	170.51
"d1,d2,d3,c1,c2,c3,a2,a3"	129.18
"d1,d2,d3,c1,c2,c3,a1.a2,a3"	101.30

[a] critical value = 1.645

4.6 Summary

The analysis and results in this Chapter can be summarized as follows:

1. A linear model is a good fit and indicator of high risk carriers for large carriers but not for small carriers.

2. Since small carriers represent 93% of the population. there is a need to develop a better model for identifying high risk carriers.

3. Non-linear Poisson models were developed for small carriers and the models were compared using the scaled deviance to find the best one.

4. Extra-Poisson variation was found in the data, leading to the introduction of the Bayes model in the following chapter.

Chapter 5

An Empirical Bayes Approach

5.1 Rationale for Bayes Approach

The Bayes approach is based on prior knowledge of the probabilities of a variable. The concept is to combine the prior knowledge with the current data observations to get a posterior that is considered a better estimate than using either the *prior* or the *likelihood* by itself. Subjective probability is the basis of the ordinary use of Bayesian method. Subjectivity, according to Miller et al. (1990), plays an important role in statistical methods or formulas that are used in a given situation. For example, subjectivity is used to decide the size of a sample in an experiment, or to specify the probabilities with which we are willing to risk errors, or even to decide how to formulate a hypothesis and the alternative against which it is to be tested. Bayes' estimates consider parameters to be random variables which have a prior distribution that reflects the strength of one's belief about the possible values they can take on, or which can reflect other indirect information. In Bayes' estimates, prior knowledge about the possible values of a variable is combined with direct sample evidence (i.e. likelihood) to get a posterior value. In this work, the expression "Empirical" means the use of Bayes theorem along with an informative prior derived from mathematical models involving the traits.

Bayes estimates of accident frequency for a certain carrier can be calculated by combining accident frequency from a mean of a certain population or a statistical model (i.e. prior knowledge) with accident rates from historical accident records (i.e. likelihood). The Bayes method was used in different studies by Hogg & Craig (1978), Robbins (1980), Hauer & Persaud (1987,1992) and Nembhard & Young (1995). The Bayes method is found to be an efficient method in the following two cases:

The first case occurs when the available traits or variables used in a statistical model are not sufficient to explain the variation in the data. In other words, there are other traits or variables which could explain the variation in the data, but they are missing from the statistical model due to the fact that these traits are not available in the data. This is shown in the work carried out by Hauer (1987,1991). Hauer showed that the use of the Bayes approach resulted in a better estimate of accident rates for railway crossings as well as for drivers in Ontario. In a study carried out by Hauer and Persaud (1987) on the safety of railway crossing, it was demonstrated that the use of Bayes estimates was a better estimate than using the traits of the crossing or the accident history of the crossing alone. It was demonstrated that, in the following 2 years, the actual observed rate was 7 times higher when only the traits were considered, and that the actual observed rate was overestimated by a factor of 5.5 when only the accident history was considered. On the other hand, the model estimate was very close to the observed accident rate.

The second case occurs when the available data on the disaggregate level is very small. Accident rates are estimated using the aggregate approach. For example, Nembhard et al. (1995) used the Bayes procedure to obtain reliable accident rate estimates using an optimal compromise between the aggregate estimate (or prior) and the historical accident record of the segment-specific (or likelihood).

5.1.1 Model Framework

There are two kinds of clues to the lack of safety of a carrier: its traits (such as fleet, conviction, detention) and its historical accident record. As shown in Figure 5.1, from the traits (first clue) of the carriers, a statistical model is developed, which represents the *prior* knowledge of carrier risk potential. Then, using an Empirical Bayes approach, the latter statistical model is combined with the historical accident record (second clue) which is considered as the *likelihood*, to estimate the *posterior* accident risk potential of carriers.

In this thesis, different models are developed using the statistical model with different combinations of historical record of accidents "h" for carriers. These models are compared to find the "best" combination. For example, in one model, "h" is a simple last year accident record. In other models, "h" is the average of the last 2 or 3 years. Thus, the model that has the "best" goodness of fit criteria (using *Pearson* χ^2) will be recommended. The proposed approach is to use the Bayes estimates as follows:

- A statistical model (i.e. prior) was developed in the previous Chapter to explain as much as possible of the variation in the number of accidents in terms of the attributes of the carrier.
- A historical accident record (i.e. likelihood) is used to provide experimental inference on carrier accident risk as observed for a certain number of years.
- The statistical model is linked with the observed accident history to get an estimate

of the accident risk potential of carriers (i.e. posterior).

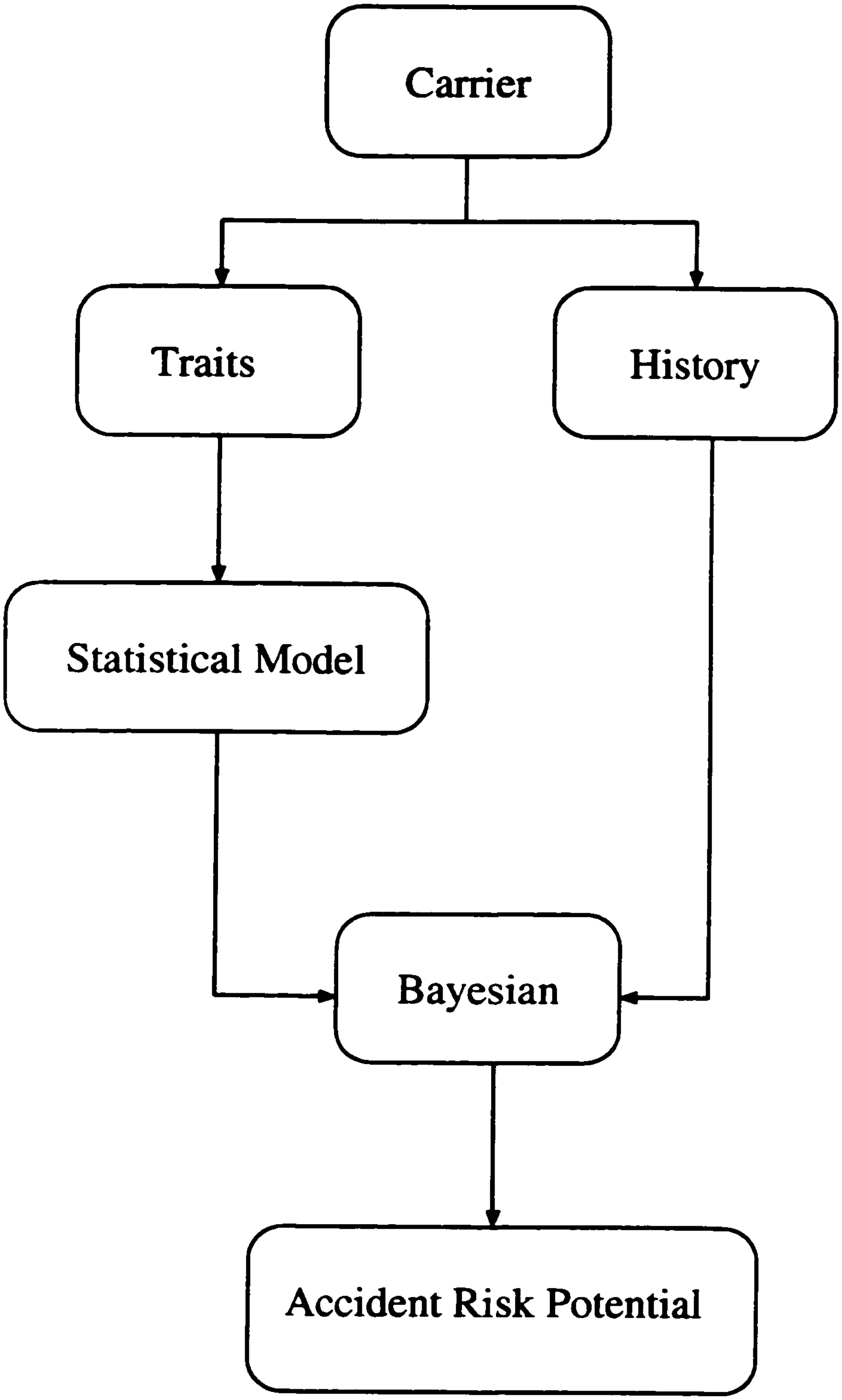

Figure 5.1: Model Structure of Accident Risk Potential for Carriers

In other words, in this thesis, the statistical model developed in the previous Chapter acts as a mean "θ" for the reference population that the carrier belong to. Then an empirical assumption is made that this mean "θ" follows a certain probability distribution for the reference population, and, using Bayes' theorem, the posterior is calculated.

5.2 Model Derivation

Using the estimated mean accident frequency "θ" from the previous Chapter as the "prior", the "likelihood" estimate from the accident history "h" is combined with it in order to produce the "posterior". This process is explained in this section. The conditional probability of "θ" knowing "h" can be expressed as follows:

$$p(\theta|h) \tag{5.1}$$

Using Bayes' Theorem

$$p(\theta|h) = \frac{p(\theta)\, p(h|\theta)}{p(h)} \tag{5.2}$$

where $p(h|\theta)$ is the probability of "h" knowing "θ" and $p(\theta)$ is the probability of "θ". Using a Gamma distribution for $p(\theta)$ (i.e probability of "θ") and a Poisson distribution for $p(h|\theta)$ (i.e probability of "h" knowing "θ") as many researchers have recommended (Robbins (1980), Hauer and Persaud (1987,1992), Nembhard and Young (1995)), the solution of $E[\theta|h]$ or the "posterior" is in the form of:

$$\begin{aligned} E[\theta|h] &= \int \theta\, p(\theta|h) \\ &= \alpha\, E(\theta) + (1-\alpha)\, h \end{aligned} \tag{5.3}$$

where

$$\alpha = \frac{1}{1 + \frac{Var(\theta)}{E(\theta)}} \tag{5.4}$$

α and $(1-\alpha)$ are the weights associated with the statistical estimate "θ" and the observed historical accident record "h" respectively. To explain the model, the following two extreme situations are considered:

- If $Var(\theta) \approx 0$, which means that the model is "perfect", α would equal 1.0 and $(1-\alpha)$ would equal zero, meaning that accident history is not needed. Equation 5.3 would be $E[\theta|h] = E(\theta)$.
- On the other hand, if $Var(\theta) \longrightarrow \infty$, which means that the model does not explain any variation at all, α would equal zero and $(1-\alpha)$ would equal 1.0, meaning that the accident history is the only factor considered in estimating $E[\theta|h]$. Equation 5.3 would be $E[\theta|h] = h$,

Having the variance of the observed accident frequency "$Var(h)$" expressed as follows:

$$Var(h) \quad = \quad E(\theta) + Var(\theta) \tag{5.5}$$

An estimate of the variance of "θ" (i.e $Var(\theta)$) is developed using $E(\theta)$, assuming a linear relationship between $Var(\theta)$ and the square of $E(\theta)$ as follows:

$$Var(\theta) = k \times [E(\theta)]^2 \tag{5.6}$$

An estimate of "k" is found to be around 0.6. Then α is calculated as shown below:

$$\alpha \quad = \quad \frac{1}{1 + \frac{Var(\theta)}{E(\theta)}} \tag{5.7}$$

$$= \frac{1}{1 + \frac{k \times (E(\theta))^2}{E(\theta)}}$$
$$= \frac{1}{1 + k \times E(\theta)}$$

Three Bayes models are examined using the *likelihood* component as follows:

1. the previous year's accident history.
2. the average of the previous 2 years' accident history.
3. the average of the previous 3 years' accident history.

The three models were compared to the statistical model developed in Chapter 4, using χ^2-test. As shown in Table 5.1, the difference in χ^2 between the statistical model developed in the previous Chapter and each of the three Bayes models was found to be significant at the 95% confidence level. Also, as in Table 5.1, when comparing the three models among themselves using the incremental differences in χ^2, it was found that the model using only the previous year's accident history is the best model.

Table 5.1: Comparison of Models Using Chi-square

Model	Chi-square	Delta chi-square [a]	D.o.f
"fleet.d1,d2,d3,c1,c2,c3"	107005	4727	7
"fleet.d1.d2.d3,c1,c2,c3,a3"	102278		8
"fleet.d1,d2,d3,c1,c2,c3,a3,a2"	102362	93	9
"fleet.d1,d2,d3,c1,c2,c3,a3,a2,a1"	102456	178	10

[a] $\chi^2 - critical$ is between 3.84 and 7.81

5.3 Preparing Bayes Model for Validation

To validate the model, it was applied to estimate the accident frequency in 1995 for each small carrier in the population (i.e fleet size of 1 to 10 trucks). It must be noted, as mentioned earlier in Chapter 4, that the data from the year of 1995 was not used in the development of the model. The estimated accidents in 1995 was calculated from the Bayes model as shown in Equation 5.8 as follows:

$$a5_{est} = \alpha \times f(fleet, d2, d3, d4, c2, c3, c4) + (1-\alpha) \times a4_{obs} \tag{5.8}$$

where

$a5_{est}$: estimated accidents in 1995

$d2, d3, d4$: the previous 3 years of detentions (92,93,94)

$c2, c3, c4$: the previous 3 years of convictions (92,93,94)

$a4_{obs}$: accidents observed in the previous year (1994)

Although, as shown in Table 5.1, the best of the 3 models was found to be the one using only the previous year's accident history, the Bayes model using the average of the previous 2 years accident history will also be used to test the model against the observed accidents in 1995. When using 2 years of accident history, the estimated accident frequency in 1995 is calculated as shown in Equation 5.9 as follows:

$$a5_{est} = \alpha \times f(fleet, d2, d3, d4, c2, c3, c4) + (1-\alpha) \times \frac{a4_{obs} + a3_{obs}}{2} \tag{5.9}$$

where

$a4_{obs}$: accidents observed in the previous year (1994)

$a3_{obs}$: accidents observed in the year of 1993

Then, the accident frequency estimates are divided by the corresponding fleet size, as shown in Equation 5.10, to produce the "accident rate per truck per year" according to which the carriers are ranked. At the top of the list are the carriers with the highest accident rate per truck per year (i.e. high risk carriers). The ranking of carriers using the proposed developed model will be compared to other ranking procedures in the following chapter.

$$Ranking\,Criteria = \frac{a5_{est}}{Fleet} \tag{5.10}$$

5.4 Summary

The analysis and results in this Chapter are summarized as follows:

1. The Bayes approach, which is used in different studies to deal with extra-Poisson variation, is explained and demonstrated.

2. An Empirical Bayes model is developed and it is shown that the Bayes model is significantly better than the statistical model developed in the previous Chapter because of the significant difference in Chi-square.

3. It is demonstrated, by the difference in the value of Chi-square, that using the accident history of one previous year in the Bayes model is better than using the averages of the previous 2 years' and 3 years' accident history.

Chapter 6

Comparison of Model with Data and Other Approaches

6.1 Test Criteria and Rationale

In this chapter, the proposed model for small carriers is compared to other approaches of identifying high risk carriers to investigate whether or not there is a significant difference between the model and these approaches. The following four alternative approaches are considered:

1. random selection based on the number of carriers in the population
2. random selection based on the number of trucks in the population
3. selection based on the previous one year of accident history
4. selection based on the previous 2 years of accident history

The first approach (i.e random based on the number of carriers in the population) is the audit approach used in many jurisdictions in North America, including Ontario (MTO in Ontario is trying to develop another approach). This approach will be investigated here because one would like to know how the Bayes model is significantly different from the existing approach.

One could argue that the more trucks in the fleet of a carrier, the more chance that this carrier should be randomly selected for a safety audit. In other words, the carrier that has a fleet size of 10 trucks would have 10 times more chances of having an accident than one with a fleet size of 1 truck. As a result, it should have 10 times as many chances of being audited than a one-truck carrier. Thus, the second approach (i.e. random based on the number of trucks in the population) is investigated as well.

The question remains: why not use accident history to rank carriers? The hypothesis that the previous year accident rate could be a good indicator of high risk carriers led to the introduction of the third approach (selection based on previous one year of accident history). In this third approach, the ranking of carriers is based on the previous one year accident rate (i.e. accident/year/truck) and the high risk carriers to be audited are identified.

The fourth approach is similar to the third, except that one would think an average of two years' accident history could be a better representation of high risk carriers than one single year of accident history. As a result, the accident rate in the fourth approach is based on the previous *two* years' accident history and the high risk carriers are identified accordingly for safety audit.

For the above selection approaches, the number of carriers selected include different cutoffs. The top 100, 200, 400, 600, 800, 1000, 1600 and 3382 carriers are selected. The rationale behind these cutoffs is explained in the following manner: when ranking the carriers according to their accident history in 1994, only 3382 carriers can be ranked because only this number of carriers had accidents in 1994. In other words, all the remaining carriers experienced zero accidents, which means that these carriers have the same ranking according to their accident history. As a result, the maximum number of ranked carriers that can be compared to Bayes approach is 3382. In Ontario, 400 to 800 carriers are audited every year. In the process of validating the Bayes model, these cutoffs are considered because they represent the actual number of carriers audited in Ontario. This is a reflection of the available resources for safety audit in the province. But the available resources for safety audit in Ontario can be increased or reduced in the future. The number of carriers audited in Ontario can then change as well. The cutoffs of 1000 and 1600 carriers audited represent the potential increase (by twice as much) of the number of carriers in Ontario. On the other hand, the cutoffs at 100 and 200 carriers audited represent the potential decrease (by half or more) of the number of carriers in Ontario.

Safety audits are not directly reducing the number of accidents, but the objective of safety audits is to identify those safety practices that make some carriers less safe than others. The variance of the accident rates of small carriers is around 5 times more than that of large carriers as shown in Chapter 3. As a consequence, the higher the variance, the easier would be to identify the difference in safety practices among carriers with low and high accident rates. Then, a certain safety program (e.g. educational programs for drivers) could be identified and introduced in order to make the high risk carriers safer. This would lead to the main goal which is reducing the number of accidents. Conducting safety audits on large carriers only would not necessarily reduce the number of accidents. In other words, many of those large carriers might be relatively safe carriers. Also, the variance of the accident rates of large carriers is not as large as those of small carriers.

Thus, the difference in safety practices through a safety audit of large carriers might not be as easily identified as in the case of small carriers.

Also, in order to test the model, we need to identify what constitutes "high risk" carrier. As mentioned in Chapter 4, the data from 1995, was not used in developing the model so that it can be used in comparing the model with other approaches of identifying high risk carriers. In this analysis, in comparing the model with other approaches, correctly identified high risk carriers means those carriers having at least one accident in 1995.

6.2 Proposed Model Versus Random Approach Based on Number of Carriers

First, the random selection method must be explained. Random selection in the first approach is based on the number of carriers in each fleet size category. Using the ratio of the number of carriers in a particular category to the total number of carriers in all categories, as shown in Table 6.1, the number of carriers are randomly generated for the different cutoffs as shown in Table 6.2.

A comparison of the number of carriers that experienced accidents in 1995 based on the random selection versus the proposed model is demonstrated in Table 6.3. It has to be noted that the random selection was carried out 4 times and all results were consistent. The proposed model identifies 31 to 66 times more carriers than the random process. The increase of efficiency and the robust nature of the proposed model over this random selection process is very clear. For example, the proposed model identified 197 carriers out of 400 while the random process identified only 3. In other words, using the same scarce resources to audit 400 carriers, the proposed model is 66 times more efficient in identifying high risk carriers than the random approach. This is a very significant improvement with no extra cost in increasing the number of carriers audited. One can therefore conclude that the Bayes model is significantly better in identifying high risk carriers than the random approach which is used in different jurisdictions in North America.

Table 6.1: Ratio Based on Number of carriers in Population

Fleet Size	Number of Carriers	Ratio
1	54346	0.617
2	15084	0.171
3	6796	0.077
4	3983	0.045
5	2569	0.029
6	1710	0.019
7	1072	0.012
8	962	0.011
9	594	0.007
10	945	0.011
	88061	1

Table 6.2: Different Cutoffs of Carrier Audited Based on Number of Carriers in Population

Fleet Size	Ratio	100	200	400	600	800	1000	1600	3382
1	0.62	62	123	247	370	494	617	987	2087
2	0.17	17	34	69	103	137	171	274	579
3	0.08	8	15	31	46	62	77	123	261
4	0.05	5	9	18	27	36	45	72	153
5	0.03	3	6	12	18	23	29	47	99
6	0.02	2	4	8	12	16	19	31	66
7	0.01	1	2	5	7	10	12	19	41
8	0.01	1	2	4	7	9	11	17	37
9	0.01	1	1	3	4	5	7	11	23
10	0.01	1	2	4	6	9	11	17	36
	1	100	200	400	600	800	1000	1600	3382

Table 6.3: Comparison Between Number of Carriers Identified Using Proposed Bayes Model and Model Using the Random Selection Based on Number of Carriers

Number of Carriers	Random	Bayes	Bayes/Random
100	0	57	infinity
200	0	102	infinity
400	3	197	66
600	7	276	39
800	7	347	50
1000	8	403	50
1600	14	596	43
3382	30	943	31

6.3 Proposed Model Versus Random Approach Based on Number of Trucks

As mentioned earlier in this Chapter, some people would argue that the random selection should be based on the number of trucks and not on the number of carriers. Another comparison will be carried out in this section between the proposed model and a random selection process, based on the number of trucks in each fleet size category. This means that a carrier that has a fleet size of 10 trucks has 10 times more chances to be randomly selected over the carrier that has a fleet size of one truck.

As shown in the third column of Table 6.4, the number of trucks in each category of fleet size is calculated by multiplying the fleet size by the corresponding number of carriers. Then, the ratio between the number of trucks in each category to the total number of trucks in all categories is calculated as shown in Table 6.4. Using this ratio, the number of carriers are randomly generated for the different cutoffs as shown in Table 6.5.

The number of carriers identified using the random selection process explained above that had experienced accidents in 1995 was compared with the number of carriers identified using the proposed model that had accidents in 1995. As shown in Table 6.6, the number of carriers identified in the proposed model is 20 to 35 times more than the number of carriers identified using the random process. For example, the proposed model identified 197 carriers out of 400 while the random process identified only 7. This is again a significant increase in the effectiveness and efficiency of the proposed model in identifying high risk carriers without any increase in the number of carriers audited (i.e. no increase in cost).

The previous two sections illustrate how the proposed model significantly increases the efficiency of identifying high risk carriers over the random selection process, a process that many jurisdictions in North America, including Ontario, have adopted.

Table 6.4: Ratio Based on Number of trucks in Population

Fleet Size	Number of Carriers	Number of Trucks	Ratio
1	54346	54346	0.312
2	15084	30168	0.173
3	6796	20388	0.117
4	3983	15932	0.092
5	2569	12845	0.074
6	1710	10260	0.059
7	1072	7504	0.043
8	962	7696	0.044
9	594	5346	0.031
10	945	9450	0.054
	88061	173935	1

Table 6.5: Different Cutoffs of carrier Audited Based on Number of Trucks in Population

Fleet Size	Ratio	100	200	400	600	800	1000	1600	3382
1	0.31	31	62	125	187	250	312	500	1057
2	0.17	17	35	69	104	139	173	278	587
3	0.12	12	23	47	70	94	117	188	396
4	0.09	9	18	37	55	73	92	147	310
5	0.07	7	15	30	44	59	74	118	250
6	0.06	6	12	24	35	47	59	94	199
7	0.04	4	9	17	26	35	43	69	146
8	0.04	4	9	18	27	35	44	71	150
9	0.03	3	6	12	18	25	31	49	104
10	0.05	5	11	22	33	43	54	87	184
	1	100	200	400	600	800	1000	1600	3382

Table 6.6: Comparison Between Number of Carriers Identified Using Proposed Bayes Model and One Using Random Selection Based on Number of Trucks

Number of Carriers	Random	Bayes	Bayes/Random
100	2	57	29
200	5	102	20
400	7	197	28
600	8	276	35
800	11	347	32
1000	13	403	31
1600	18	596	33
3382	47	943	20

The following two sections further test the robustness of the proposed model by comparing it to other methods of identifying high risk carriers. These other methods are based on the accident history of the carriers. The Bayes Model is compared to 2 different approaches of ranking carriers using accident history. In one approach, the selection or ranking process is based on the previous year's accident history. In the other approach, the ranking process is based on the average of the previous *two* years' accident history. These approaches will be discussed in the next 2 sections.

6.4 Proposed Model Versus Previous Year Accident History Approach

First. the procedure of how the ranking of carriers using the previous year's accident history needs to be explained. The accidents of the previous year of each carrier was divided by its fleet size to produce the accident rate based on the previous year's accident history. The carriers were then ranked according to their accident rates where the highest rate was at the top of the list.

The accident rates based on the accidents in 1994 were calculated and the carriers were ranked according to those calculated accident rate. The number of carriers that experienced accidents in 1995 and were identified using the previous year's accident history (i.e. 1994). was compared with the number of carriers that experienced accidents in 1995 and were identified using the Bayes model ranking. It was found that, as shown in Table 6.7, the number of carriers identified using the proposed model was 6% to 163% higher than the number of carriers identified using the previous year's accident history approach. The number of carriers that is usually audited every year in Ontario is between 400 and 800 carriers. As shown in Table 6.7, when 400 to 800 carriers are audited, the efficiency of using the proposed model over the previous year's accident history approach is increased by 140% to 163%. This demonstrates a significant increase in the number of carriers identified as having had accidents in 1995 (or high risk carriers in 1995) in Ontario with no extra carriers to audit and no extra cost of collecting additional data. In this way, scarce resources are efficiently used.

In Table 6.8. it is demonstrated that 12% to 98% of the number of carriers identified using the proposed model did not experience any accidents in the previous year but experienced accidents in the following year. This demonstrates that the carriers' attributes (i.e detentions, convictions and fleet size) used in the proposed model provide more information and help identify more carriers, including the ones that did not experience any accidents in the previous year. When the range of the number of carriers audited (i.e. high risk) is between 100 to 1600, 71% to 93% of these carriers identified using the proposed model did not have any accidents in the previous year. This result is consistent with the findings in Chapter 3, where 75% of the carriers that had accidents in one year did not have any accidents in the previous year. for the period of 1991 to 1995.

The benefit of identifying more carriers in terms of the extra number of accidents that can be targeted, further demonstrates the efficiency of the proposed model over the previous year's accident history approach. Table 6.9 shows the extra number of accidents targeted using the proposed model over the accident history approach. The extra number of accidents that can be targeted (potentially avoided) ranges from 105 to 614. This number can be translated into a dollar figure. According to an MTO study of accident cost in Ontario (Safety Research Office (1994)), an average accident costs society around \$41,000. Savings generated from using the proposed model could range from \$4.3 million to \$25 million. The figures are probably higher than the ones mentioned above due to the fact that trucks are involved in a larger number of *fatal* accidents with respect to their population, as mentioned in Chapter 1. In Ontario, an average *fatal* accident costs $6.3 million (Safety Research Office (1994)). These dollar figures illustrate the direct benefit to society of adopting the proposed model with an increased rate of return on the resources allocated for programs to improve highway safety.

Table 6.7: Comparison Between Number of Carriers Identified Using Proposed Bayes Model and One Using Previous Year Accident History

No. of Carriers Audited	One-Year Accident History	Bayes Model	% Increase
100	24	57	138
400	82	197	140
600	105	276	163
800	139	347	150
1000	173	403	133
1600	309	596	93
3382	886	943	6

Table 6.8: Number of Carriers Identified Using Proposed Bayes Model Only (i.e. carriers that had no accidents in the previous year) Versus One Using Previous Year's Accident History

No. of Carriers Audited	One-year Accident History	Bayes Model	In Bayes Only	% In Bayes Only
100	24	57	53	93
400	82	197	172	87
600	105	276	243	88
800	139	347	305	88
1000	173	403	344	85
1600	309	596	422	71
3382	886	943	111	12

Table 6.9: Extra Number of Potential Accidents Targeted Using Proposed Bayes Model Versus Using Previous Year Accident History

No. of Carriers Audited	One-Year Accident History	Bayes Model	No. of Extra Accident targeted
100	24	57	105
400	82	197	312
600	105	276	440
800	139	347	505
1000	173	403	555
1600	309	596	614
3382	886	943	106

6.5 Proposed Model Versus Previous Two Years Accident History Approach

In this section, the robust nature of the proposed model is tested one more time. Here, the proposed model is compared to a ranking process using the previous two years' accident history not only the previous year's accidents, as demonstrated in the previous section. The ranking in this case is based on the accident rate calculated from the previous two years. The accident rate is calculated by taking the average of the number of accidents over the previous two years (i.e 1993 & 1994) for each carrier and dividing it by the fleet size.

As shown in Table 6.10, 58% to 133% more carriers were identified using the proposed model than were identified using the previous two years' accident history approach. This demonstrates, one more time, the significant increase in the number of carriers identified using the proposed model. Table 6.11 shows the number of carriers identified using the proposed model that had no accidents in the previous 2 years. The percentage of those carriers is between 50% to 86%. Again, this demonstrates that the attributes of the carriers used in the proposed model provide more information that help in identifying high risk carriers. It has to be noted that these results are consistent with the findings in Chapter 3, where 60% of carriers that had accidents in one year did not have any accidents in the previous 2 years.

Also, the extra number of accidents that can be targeted (potentially avoided) by using the proposed model versus the previous two years' accident history approach is a good indicator of the efficiency of the model. Table 6.12 shows the extra number of accidents for different cutoffs of carriers audited/targeted; the extra number of accidents range between 94 to 734 when using the proposed model.

As in the previous section, this number can be translated into a dollar figure. Again, using the MTO study carried out in 1994, the savings generated from the use of the proposed model could range from \$4 million to \$30 million. In fact, due to the disproportionate number of *fatal* accidents that trucks are involved in, as previously mentioned, the dollar figure of savings is potentially much higher. Therefore, applying the proposed model in identifying the high risk carriers that should be targeted for safety audit can be of direct benefit to society.

Table 6.10: Comparison Between Number of Carriers Identified Using Proposed Bayes Model and One Using Average of Previous 2 Years Accident History

No. of Carriers Audited	Two-Year Accident History	Bayes Model	% Increase
100	28	56	100
400	108	181	68
600	131	266	103
800	131	338	121
1000	153	401	123
1600	261	609	133
3382	627	990	58

Table 6.11: Number of Carriers Identified Using Proposed Bayes Model Only (i.e carriers that had no accident in the previous 2 years) Versus One Using Average of Previous 2 Years Accidents History

No. of Carriers Audited	Two-Year Accident History	Bayes Model	In Bayes Only	% In Bayes Only
100	28	56	48	86
400	108	181	146	81
600	131	266	214	80
800	131	338	278	82
1000	153	401	332	83
1600	261	609	498	82
3382	627	990	583	59

Table 6.12: Extra Number of Potential Accident Targeted Using Proposed Bayes Model Versus Using Last Year's Accident History

No. of Carriers Audited	Two-Year Accident History	Bayes Model	No. of Extra Accident Targeted
100	28	56	94
400	108	181	226
600	131	266	354
800	131	338	463
1000	153	401	533
1600	261	609	734
3382	627	990	701

Another analysis was carried out to investigate the distribution of carriers, by fleet size, ranked by the Bayes model. As shown in Table 6.13 and Figure 6.1, the distribution is consistent with the expected distribution of carriers having accidents every year, except for carriers having a fleet size of 1 truck. In the case of carriers of 1 truck, the Bayes model seems to be underestimating the number of carriers identified as high risk ones. An interpretation of that statistic could be that carriers with a fleet size of 1 truck who experience an accident in one year do not necessarily fall into the *high risk* category. In other words, the proposed model does not consider these carriers as "high risk" because it combines their accident history with their conviction and detention history.

Most importantly, carriers identified using the proposed model are not necessarily the same as those identified using the previous year's accident history. For example, as shown in Table 6.13, the percentages of high risk carriers identified for carriers having a fleet size of 2 trucks are 16% and 15% when using the Bayes model and the previous accident history approach respectively. But the carriers identified by the Bayes model are *different* from those identified using accident history. This was demonstrated earlier in this Chapter, where 12% to 93% of the carriers identified using the Bayes model were not identified using the accident history approach.

Table 6.13: Distribution of Carriers by Fleet Size

Fleet Size	Accidents in 1995	Bayes Model	Accidents in 1994
1	32%	19%	30%
2	17%	16%	15%
3	12%	14%	12%
4	10%	11%	10%
5	8%	10%	8%
6	7%	7%	6%
7	5%	7%	6%
8	5%	6%	5%
9	2%	4%	3%
10	2%	5%	4%
	100%	100%	100%

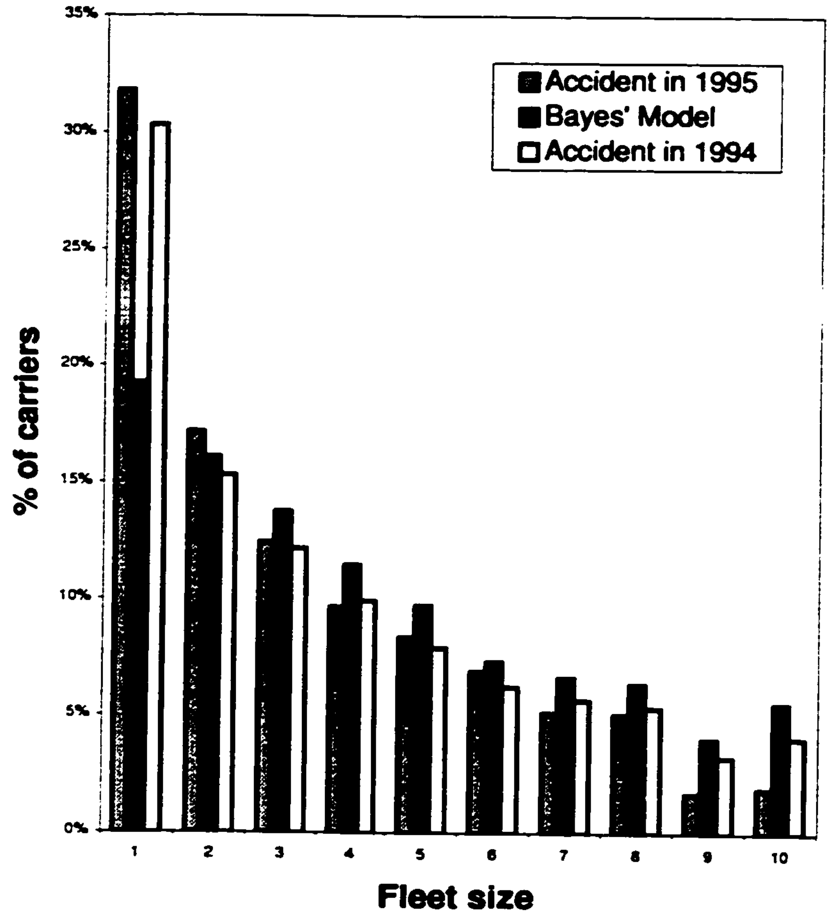

Figure 6.1: Distribution of High Risk Carriers By Fleet Size Using Bayes Model Compared to Distribution of Carriers That Had Accidents in 1994 and 1995

6.6 Summary

The analysis and results in this Chapter can be summarized as follows:

1. Four alternative approaches in ranking carriers and identifying high risk ones are defined and the rationale behind each is explained.

2. The proposed Bayes model is compared to the four alternative approaches in ranking carriers and identifying high risk ones.

3. The number of carriers identified by the Bayes model is significantly larger (i.e. 2000% to 6600%) than that identified by either of the 2 random approaches.

4. Using the Bayes model in identifying the 400 to 800 carriers audited every year in Ontario. the number of carriers correctly identified as high risk is more than 100% of the number identified using the accident history of one or 2 years. This is a significant improvement of the efficiency of the safety audit program in the province at no extra cost.

5. The proposed Bayes model is significantly more efficient in identifying high risk carriers than any of the 4 alternative approaches. This is substantiated by the larger number of accidents associated with the carriers identified by the Bayes model than the number associated with the carriers identified by the accident history approach.

6. Most importantly, most of the carriers identified by the Bayes model (around 60% to 90%) did not have any accidents in the previous year. This is consistent with the findings in Chapter 3 where 75% of the carriers that had accidents in one year did not have any accidents in the previous year.

Chapter 7

Conclusions and Recommendations

The resources available to enhance highway safety are scarce. As a consequence, these scarce resources need to be used efficiently and effectively. In this thesis, highway safety issues were studied from the carrier perspective because in many jurisdictions, a relatively small percentage of carriers are responsible for most accidents. The objective of the study was to estimate accident risk potential of carriers so that the "*high risk*" ones can be targeted for different safety interventions, such as full safety audits. Presumably, if high risk carriers are targeted, safety interventions can be made more cost-effective and efficient. This will enhance carriers' safety, which will in turn lead to the overall improvement of safety on highways.

In this thesis, an Empirical Bayes model was developed that addresses the problem of the limited number of carrier attributes in available data. The Bayes model has been proven to be robust when compared to various alternative approaches for sampling carriers. In this Chapter, conclusions and recommendations derived from the research will be summarized.

7.1 Conclusions

The following conclusions are drawn from the analysis and results detailed in the thesis.

1. There is a difference in the mean accident rate of small carriers and that of large carriers when calculated for each individual carrier. The number of accidents of large carriers tend to be more stable, with less fluctuation from year to year.
2. For large carriers, accident history is a good estimator of future accident frequency. A linear regression model provides a simple and good model in estimating accident frequency using the previous three years of accident frequencies.
3. The bulk of the carrier population consists of small carriers; the fleet size of 93% of all carriers is smaller than or equal to 10 trucks.
4. Only 4% of small carriers are responsible of 100% of accidents experienced by all small carriers in any one year. These 4% change from one year to the next. In other words, 75% of the small carriers that had accidents in one year, did not have any accidents in the previous year.
5. Detentions, convictions, and the fleet size of carriers are relevant and significant factors in explaining the variations in accidents for small carriers, which represent 93% of the total population of carriers.
6. Accident frequency follows usually a Poisson distribution. In this thesis, Poisson distribution was used but an extra-Poisson variation was found in the accident frequency of small carriers. An Empirical Bayes approach was presented to account for this extra-Poisson variation.

7. The Empirical Bayes Model explains the variations in accidents better than the pure Poisson regression model. The Bayes Model combines information from detentions, convictions, and fleet size with the accident history of carriers.

8. Comparing the Empirical Bayes Model with a random process (the existing safety audit approach) in ranking small carriers and identifying high risk ones, the number of high risk carriers identified using the Bayes model is found to be 20 to 60 times larger than the number of carriers identified using the random approach. In this thesis, high risk carriers means those carriers having at least one accident in 1995 (the data in 1995 was not used in developing the model). The comparison between the model and the other approaches was carried out for small carriers only.

9. Comparing the Empirical Bayes Model with the accident history approach in ranking carriers and identifying high risk ones, the number of high risk carriers identified using the Bayes model is found to be significantly larger than the one using the accident history approach. For example, in Ontario, 400 to 800 carriers are audited every year. In this range, the number of carriers that experienced accidents in 1995 and are identified using the Bayes model, was found to be 70% to 120% larger than the number of carriers that experienced accidents in 1995 and were identified using the ranking approach based on the accident history. Also, using the Bayes model, the number of accidents that could potentially be avoided is found to be around 450.

10. When the number of carriers audited varies from 100 to 1600 carriers (1600 is twice the maximum number of carriers audited in Ontario every year), 71% to 88% of carriers identified by the Bayes Model as high risk carriers did not have any accidents in the previous year. This is found to be consistent with the results from the observed data where, every year, 75% of the carriers that had accidents in one year did not have any in the previous year.

7.2 Recommendations

1. Two different models are recommended in ranking small and large carriers. In the case of large carriers, accident history is a good indicator of ranking carriers using a simple linear regression model. Most importantly, an Empirical Bayes model is recommended in ranking small carriers, which represent 93% of the total population of carriers.

2. It is recommended that *safety audits* be based on the proposed model so that more high risk carriers are targeted. As a consequence, scarce resources are effectively used and more accidents will potentially be avoided, which will lead to the overall improvement of highway safety.

3. Future research should explore other classifications of carriers, with respect to the fleet size, than the one used in this thesis. In other words, classifications other than *small* and *large* carriers should be investigated to determine if the model can be improved.

4. After identifying the accident rates for small carriers as well as for large carriers, they can be combined in one single list where the top of the list are the carriers (small and large) having high accident rates. Then, those carriers could be targeted for safety audits.

Appendix A

More Analysis of Detentions and Convictions

Table A.1: Change In Small Carriers (1 to 10 trucks) That Had Detentions From One Year To Next

Year	1995	95 and not in 94	95 and not in 94 nor 93	95 and not in 94 nor 93 nor 92
No. Of carriers	4260	3197	2722	2232
of carriers		75%	64%	52%
Year	1994	94 and not in 93	94 and not in 93 nor 92	94 and not in 93 nor 92 nor 91
No. Of carriers	2568	1876	1340	1053
% of carriers		73%	52%	41%
Year	1993	93 and not in 92	93 and not in 92 nor 91	
No. Of carriers	2508	1677	1278	
% of carriers		67%	51%	
Year	1992	92 and not in 91		
No. Of carriers	4127	2992		
% of carriers		72%		

Table A.2: Change In Large Carriers (more than 10 trucks) That Had Detentions From One Year To Next

Year	1995	95 and not in 94	95 and not in 94 nor 93	95 and not in 94 nor 93 nor 92
No. Of carriers	1647	651	438	293
% of carriers		40%	27%	18%
Year	1994	94 and not in 93	94 and not in 93 nor 92	94 and not in 93 nor 92 nor 91
No. Of carriers	1299	501	254	167
% of carriers		39%	20%	13%
Year	1993	93 and not in 92	93 and not in 92 nor 91	
No. Of carriers	1253	338	192	
% of carriers		27%	15%	
Year	1992	92 and not in 91		
No. Of carriers	1665	575		
% of carriers		35%		

Table A.3: Change In Small Carriers (1 to 10 trucks) That Had Convictions From One Year To Next

Year	1995	95 and not in 94	95 and not in 94 nor 93	95 and not in 94 nor 93 nor 92
No. Of carriers	10399	6414	5078	4289
% of carriers		62%	49%	41%
Year	1994	94 and not in 93	94 and not in 93 nor 92	94 and not in 93 nor 92 nor 91
No. Of carriers	9737	6037	4682	4164
% of carriers		62%	48%	43%
Year	1993	93 and not in 92	93 and not in 92 nor 91	
No. Of carriers	9241	5576	4593	
% of carriers		60%	50%	
Year	1992	92 and not in 91		
No. Of carriers	9726	6434		
% of carriers		66%		

Table A.4: Change In Large Carriers (more than 10 trucks) That Had Convictions From One Year To Next

Year	1995	95 and not in 94	95 and not in 94 nor 93	95 and not in 94 nor 93 nor 92
No. Of carriers	2413	602	356	262
% of carriers		25%	15%	11%
Year	1994	94 and not in 93	94 and not in 93 nor 92	94 and not in 93 nor 92 nor 91
No. Of carriers	2405	573	346	287
% of carriers		24%	14%	12%
Year	1993	93 and not in 92	93 and not in 92 nor 91	
No. Of carriers	2432	578	417	
% of carriers		24%	17%	
Year	1992	92 and not in 91		
No. Of carriers	2474	748		
% of carriers		30%		

Appendix B

Results of Bayes' Model Using 2 Years of Detentions and Convictions

Table B.1: Comparison Between Number of Carriers Identified Using Second Best Bayes Model and One Using Previous 2 Years Accident History

No audited	bayes count	a4 count	% increase
100	57	28	104%
200	98	46	113%
400	192	108	78%
600	281	131	115%
800	350	153	129%
1000	420	180	133%
1600	614	261	135%
3382	1019	627	63%

Table B.2: Number of Carriers Identified Using Second Best Bayes Model Only (i.e carriers that had no accidents in the previous 2 years) versus the one using the previous 2 years' accident history

No audited	bayes count	a4 count	only Bayes	% of Bayes only
100	57	28	50	88%
200	98	46	88	90%
400	192	108	154	80%
600	281	131	230	82%
800	350	153	288	82%
1000	420	180	346	82%
1600	614	261	505	82%
3382	1019	627	621	61%

Table B.3: Extra Number of Potential Accidents Targeted Using Second Best Bayes Model Versus Using Previous 2 Years Accident History

No audited	bayes count	a4 count	No. of potential accident avoided
100	57	28	101
200	98	46	180
400	192	108	255
600	281	131	393
800	350	153	474
1000	420	180	561
1600	614	261	750
3382	1019	627	750

Table B.4: Distribution of Carriers Having Accident in 1995 and 1994 Versus Distrubition of Carriers Using Second Best Bayes' Model for Ranked 3382 carriers

Fleet	% in 1995	% in Bayes	% in 1994
1	32%	22%	30%
2	17%	16%	15%
3	12%	13%	12%
4	10%	11%	10%
5	8%	9%	8%
6	7%	7%	6%
7	5%	6%	6%
8	5%	6%	5%
9	2%	4%	3%
10	2%	5%	4%

Appendix C

Sample of Programs Used in Analysis

C.1 Samples of Linear Model Programs

```
%########################################################################
file name : /u2/melherra/CVOR/regre_1_10_Normal_Acc.glm
%########################################################################
$data 142828 cvor fleet a1 a2 a3 a4 a5 d1 d2 d3 d4 d5 c1 c2 c3 c4 c5
$dinput 'cvor_fleet_accid_det_conv.out'
$output 'regre_1_10_Normal_Acc.glm.out'
$print 'Analysis regre_1_10_Normal_Acc.glm';;;
$c ---------------------------------------------
$c calc I=(fleet==1)
$calc I=(fleet>=1 & fleet<=10)
:       %n=%CU(I)
$var %n fleet_1 a1_1 a2_1 a3_1 a4_1 a5_1
$var %n         d1_1 d2_1 d3_1 d4_1 d5_1
$var %n         c1_1 c2_1 c3_1 c4_1 c5_1

$calc J=I*%CU(I)
$calc       fleet_1(J)=fleet
$calc       a1_1(J)=a1
$delete a1 $calc      a2_1(J)=a2
$delete a2 $calc      a3_1(J)=a3
```

```
$delete a3 $calc      a4_1(J)=a4
$delete a4 $calc      a5_1(J)=a5
$delete a5 $calc      d1_1(J)=d1
$delete d1 $calc      d2_1(J)=d2
$delete d2 $calc      d3_1(J)=d3
$delete d3 $calc      d4_1(J)=d4
$delete d4 $calc      d5_1(J)=d5
$delete d5 $calc      c1_1(J)=c1
$delete c1 $calc      c2_1(J)=c2
$delete c2 $calc      c3_1(J)=c3
$delete c3 $calc      c4_1(J)=c4
$delete c4 $calc      c5_1(J)=c5
$delete c5


$c link L
$c error Poisson
$link I
$error Normal

$yvariate a5_1

$cycle  50 1
$c fit d1_1+d2_1+d3_1+d4_1+c1_1+c2_1+c3_1+c4_1
$fit a1_1+a2_1+a3_1+a4_1
$display e m t v c

$print 'Chi Square = Deviance /d.o.f    '  $look %x2$

$c the histotical part of Beyesian
$var %n history
$calc history=(a3_1+a2_1+a1_1)/3


$c calculation of un-explained  variance

$var 1 unexp_var sum_unexp_var r_square
$var 1 exp_var sum_exp_var
$var %n one_avg

$c calculation of average
$calc avg=%cu(%fv)/%n
$calc one_avg=avg(%n)

$c calculation of un-explained  variance

$calc vec_unexp_var=(%fv - %yv)**2
```

```
$calc %x=%cu(vec_unexp_var)
$calc unexp_var=%x/%n
$calc sum_unexp_var=%x
$print 'un_explained variance    ' $look unexp_var sum_unexp_var

$c calculation of -explained  variance

$calc vec_exp_var=(%fv - one_avg)**2
$calc %x=%cu(vec_exp_var)/%n
$calc exp_var=%x
$calc sum_exp_var=%x*%n
$print 'explained variance    ' $look exp_var sum_exp_var

$print 'Linear Regression Calculations'
$var 1 r_square
$calc r_square=1-(unexp_var/(unexp_var+exp_var))
$look r_square

$c calculation of Model Pearson Chi_square

$var 1 mod_unexp_var mod_sum_unexp_var
$calc mod_vec_unexp_var=((%fv - %yv)**2) / %fv
$calc %x=%cu(mod_vec_unexp_var)
$calc mod_unexp_var=%x/(%n)
$calc mod_sum_unexp_var=%x
$print 'Model Pearson Chi_square' mod_sum_unexp_var

$c calculations of the Bayesian Estimate for the specific group

$var %n v_alpha
$calc alpha=1/(1+(unexp_var/one_avg(%n)))
$calc v_alpha=alpha(1)
$calc bayes=v_alpha*(%fv)+(1-v_alpha)*(history)
$print 'alpha =' $look alpha

$c calculation of Bayesian Pearson Chi_square

$var 1 bay_unexp_var bay_sum_unexp_var
$calc bay_vec_unexp_var=((bayes - %yv)**2) /bayes
$calc %x=%cu(bay_vec_unexp_var)
$calc bay_unexp_var=%x/(%n)
$calc bay_sum_unexp_var=%x
$print 'Bayesian Pearson Chi_square' bay_sum_unexp_var

$return
```

```
%############################################################################
file name : /u2/melherra/CVOR/regre_more_10_Normal_Acc.glm
%############################################################################
$data 142828 cvor fleet a1 a2 a3 a4 a5 d1 d2 d3 d4 d5 c1 c2 c3 c4 c5
$dinput 'cvor_fleet_accid_det_conv.out'
$output 'regre_more_10_Normal_Acc.out'
$print 'Analysis for regre_more_10_Normal_Acc.glm';;;
$c ---------------------------------------------
$c calc I=(fleet==1)
$c calc I=(fleet>=11 & fleet<=50)
$calc I=(fleet>10)
:       %n=%CU(I)
$var %n fleet_1 a1_1 a2_1 a3_1 a4_1 a5_1
$var %n         d1_1 d2_1 d3_1 d4_1 d5_1
$var %n         c1_1 c2_1 c3_1 c4_1 c5_1

$calc J=I*%CU(I)
$calc       fleet_1(J)=fleet
$calc       a1_1(J)=a1
$delete a1 $calc       a2_1(J)=a2
$delete a2 $calc       a3_1(J)=a3
$delete a3 $calc       a4_1(J)=a4
$delete a4 $calc       a5_1(J)=a5
$delete a5 $calc       d1_1(J)=d1
$delete d1 $calc       d2_1(J)=d2
$delete d2 $calc       d3_1(J)=d3
$delete d3 $calc       d4_1(J)=d4
$delete d4 $calc       d5_1(J)=d5
$delete d5 $calc       c1_1(J)=c1
$delete c1 $calc       c2_1(J)=c2
$delete c2 $calc       c3_1(J)=c3
$delete c3 $calc       c4_1(J)=c4
$delete c4 $calc       c5_1(J)=c5
$delete c5


$c link L
$c error Poisson
$link I
$error Normal

$yvariate a5_1

$cycle  50 1
$c fit d1_1+d2_1+d3_1+d4_1+c1_1+c2_1+c3_1+c4_1
$fit a1_1+a2_1+a3_1+a4_1
$display e m t v c
```

```
$print 'Chi Square = Deviance /d.o.f   '  $look %x2$

$c the histotical part of Beyesian
$var %n history
$calc history=(a3_1+a2_1+a1_1)/3


$c calculation of un-explained  variance

$var 1 unexp_var sum_unexp_var r_square
$var 1 exp_var sum_exp_var
$var %n one_avg

$c calculation of average
$calc avg=%cu(%fv)/%n
$calc one_avg=avg(%n)

$c calculation of un-explained  variance

$calc vec_unexp_var=(%fv - %yv)**2
$calc %x=%cu(vec_unexp_var)
$calc unexp_var=%x/%n
$calc sum_unexp_var=%x
$print 'un_explained variance   ' $look unexp_var sum_unexp_var

$c calculation of -explained  variance

$calc vec_exp_var=(%fv - one_avg)**2
$calc %x=%cu(vec_exp_var)/%n
$calc exp_var=%x
$calc sum_exp_var=%x*%n
$print 'explained variance   ' $look exp_var sum_exp_var

$print 'Linear Regression Calculations'
$var 1 r_square
$calc r_square=1-(unexp_var/(unexp_var+exp_var))
$look r_square

$c calculations of the Bayesian Estimate for the specific group

$var %n v_alpha
$calc alpha=1/(1+(unexp_var/one_avg(%n)))
$calc v_alpha=alpha(1)
$calc bayes=v_alpha*(%fv)+(1-v_alpha)*(history)
$print 'alpha =' $look alpha
```

```
$c calculation of un-explained  variance

$var 1 bay_unexp_var bay_sum_unexp_var
$calc bay_vec_unexp_var=(bayes - %yv)**2
$calc %x=%cu(bay_vec_unexp_var)/%n
$calc bay_unexp_var=%x
$calc bay_sum_unexp_var=%x*%n
$print 'un_explained variance   ' $look bay_unexp_var bay_sum_unexp_var


$return
```

C.2 Samples of Poisson Model Programs

```
%###########################################################################
file name : /u2/melherra/CVOR/CHI_model_3years_bayes_1history_m.glm
%###########################################################################
$data 142828 cvor fleet a1 a2 a3 a4 a5 d1 d2 d3 d4 d5 c1 c2 c3 c4 c5
$dinput 'cvor_fleet_accid_det_conv.out'
$output 'CHI_model_3years_bayes_1history_m.out'
$print 'Analysis for Rating_1_10_a4_a5_1year_fv_Exclude_Zero';;;
$c ------------------------------------------------
$c calc I=(fleet>=1 & fleet<=5)
$calc I=(fleet>=1 & fleet<=10)
$c calc I=(fleet>=11 & fleet<=99)
$c calc I=(fleet>=11)
$c  calc I=(fleet==1)
$calc %n=%CU(I)
$var %n cvor_1 fleet_1 a1_1 a2_1 a3_1 a4_1 a5_1
$var %n        d1_1 d2_1 d3_1 d4_1 d5_1
$var %n        c1_1 c2_1 c3_1 c4_1 c5_1

$calc J=I*%CU(I)
$calc     cvor_1(J)=cvor
$calc       fleet_1(J)=fleet
$delete cvor fleet
$calc       a1_1(J)=a1
$delete a1 $calc       a2_1(J)=a2
$delete a2 $calc       a3_1(J)=a3
$delete a3 $calc       a4_1(J)=a4
$delete a4 $calc       a5_1(J)=a5
$delete a5 $calc       d1_1(J)=d1
$delete d1 $calc       d2_1(J)=d2
$delete d2 $calc       d3_1(J)=d3
$delete d3 $calc       d4_1(J)=d4
$delete d4 $calc       d5_1(J)=d5
$delete d5 $calc       c1_1(J)=c1
$delete c1 $calc       c2_1(J)=c2
$delete c2 $calc       c3_1(J)=c3
$delete c3 $calc       c4_1(J)=c4
$delete c4 $calc       c5_1(J)=c5
$delete c5


$link L
$error Poisson
$c link I
$c error Normal
```

```
$c yvariate a5_1
$yvariate a4_1
$calc log_fleet_1=%log(fleet_1)
$offset log_fleet_1

$cycle  50 1
$fit fleet_1+d1_1+d2_1+d3_1+c1_1+c2_1+c3_1
$c fit fleet_1+d2_1+d3_1+c2_1+c3_1
$c fit fleet_1+d3_1+c3_1
$c fit a1_1+a2_1+a3_1+a4_1
$display e m t v c

$print 'Chi Square = Deviance /d.o.f   '  $look %x2$

$c the histotical part of Beyesian
$var %n history
$c calc history=(a4_1+a3_1+a2_1+a1_1)/4
$c  calc history=(a3_1+a2_1+a1_1)/3
$c calc history=(a4_1+a3_1)/2
$c calc history=a4_1
$calc history=a3_1
$c calculation of un-explained  variance

$var 1 unexp_var sum_unexp_var r_square
$var 1 exp_var sum_exp_var
$var %n one_avg

$c calculation of average
$calc avg=%cu(%fv)/%n
$calc one_avg=avg(%n)

$c calculation of un-explained  variance

$calc vec_unexp_var=(%fv - %yv)**2
$calc %x=%cu(vec_unexp_var)
$calc unexp_var=%x/%n
$calc sum_unexp_var=%x
$print 'un_explained variance   ' $look unexp_var sum_unexp_var

$c calculation of -explained  variance

$calc vec_exp_var=(%fv - one_avg)**2
$calc %x=%cu(vec_exp_var)/%n
$calc exp_var=%x
$calc sum_exp_var=%x*%n
$print 'explained variance   ' $look exp_var sum_exp_var
```

```
$print 'Linear Regression Calculations'
$var 1 r_square
$calc r_square=1-(unexp_var/(unexp_var+exp_var))
$look r_square

$c calculation of Model Pearson Chi_square

$var 1 mod_unexp_var mod_sum_unexp_var
$calc mod_vec_unexp_var=((%fv - %yv)**2) / %fv
$calc %x=%cu(mod_vec_unexp_var)
$calc mod_unexp_var=%x/(%n)
$calc mod_sum_unexp_var=%x
$print 'Model Pearson Chi_square' mod_sum_unexp_var

$c calculations of the Bayesian Estimate for the specific group

$var %n v_alpha

$calc var_m=(%fv-%yv)**2 - %fv
$c calc var_m_1=%if(var_m<0,0,var_m)
$plot var_m %fv

$extract %pe
$var %pl coeff
$c calc para=1,fleet,d2_1,d3_1,d4_1,c2_1,c3_1,c4_1
$c calc para=1,fleet,d1_1,d2_1,d3_1,c1_1,c2_1,c3_1
$calc coeff=%pe
$calc yv_1=%yv
$calc estimate=%fv
$calc square_est= %fv**2

$link I
$error Normal
$yvariate var_m
$calc my_os=0
$offset my_os
$calc %os=0
$fit square_est-1
$display e m
$plot my_os %os
$c ------------------
$delete my_os
$c ------------------
$calc v_alpha=1/(1+(%fv/estimate))
$c calc v_alpha=1/(1+(var_m_1/%fv))
$c calc v_alpha=alpha(1)
```

```
$var %n fv_1,temp1,temp2,temp
$calc temp1=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d1_1+coeff(4)*d2_1$
$calc temp2=coeff(6)*c1_1+coeff(7)*c2_1+coeff(8)*c3_1+coeff(5)*d3_1$
$c
$c calc temp1=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d2_1+coeff(4)*d3_1$
$c calc temp2=coeff(6)*c2_1+coeff(7)*c3_1+coeff(8)*c4_1+coeff(5)*d4_1$
$calc temp=temp1+temp2$
$c ------------------
$c calc temp=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d2_1+coeff(4)*d3_1+coeff(5)*d4_1+coeff(6)*c2_1+coeff(7)*c3_1+coeff(8)*
$c calc temp=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d1_1+coeff(4)*d2_1+coeff(5)*d3_1+coeff(6)*c1_1+coeff(7)*c2_1+coeff(8)*
$c----------
$delete temp1 temp2 one_avg
$c----------
$calc fv_1=%exp(temp)*fleet_1
$calc bayes=v_alpha*(fv_1)+(1-v_alpha)*(history)
$c calc bayes=v_alpha*(estimate)+(1-v_alpha)*(history)
$c print 'alpha =' $c look alpha
$plot v_alpha %fv
$plot estimate fv_1
$c ---------------------------------------------
$c delete temp estimate fv_1
$delete temp  fv_1
$c ---------------------------------------------
$c calculation of Bayesian Pearson Chi_square

$var 1 bay_unexp_var bay_sum_unexp_var
$calc bay_vec_unexp_var=((bayes - yv_1)**2) /(estimate+%fv)
$c calc bay_vec_unexp_var=((bayes - yv_1)**2) /bayes
$c calc bay_vec_unexp_var=((bayes - a4_1)**2) /bayes
$calc %x=%cu(bay_vec_unexp_var)
$calc bay_unexp_var=%x/(%n)
$calc bay_sum_unexp_var=%x
$print 'Bayesian Pearson Chi_square' bay_sum_unexp_var


$c calculation of un-explained  variance for Bayesian Solution

$var 1 b_unexp_var sum_b_unexp_var

$calc vec_b_unexp_var=(bayes - yv_1)**2
$c calc vec_b_unexp_var=(bayes - a4_1)**2
$calc %x=%cu(vec_b_unexp_var)
$calc b_unexp_var=%x/%n
$calc sum_b_unexp_var=%x
$print 'Bayesian un_explained variance'   $look sum_b_unexp_var
$c ---------------------------------------------
$look coeff
```

```
$return
```

```
%###########################################################################
file name : /u2/melherra/CVOR/CHI_model_2years_bayes_2history_m.glm
%###########################################################################
$data 142828 cvor fleet a1 a2 a3 a4 a5 d1 d2 d3 d4 d5 c1 c2 c3 c4 c5
$dinput 'cvor_fleet_accid_det_conv.out'
$output 'CHI_model_2years_bayes_2history_m.out'
$print 'Analysis for Rating_1_10_a4_a5_1year_fv_Exclude_Zero';;;
$c ------------------------------------------------
$c calc I=(fleet>=1 & fleet<=5)
$calc I=(fleet>=1 & fleet<=10)
$c calc I=(fleet>=11 & fleet<=99)
$c calc I=(fleet>=11)
$c  calc I=(fleet==1)
$calc %n=%CU(I)
$var %n cvor_1 fleet_1 a1_1 a2_1 a3_1 a4_1 a5_1
$var %n         d1_1 d2_1 d3_1 d4_1 d5_1
$var %n         c1_1 c2_1 c3_1 c4_1 c5_1

$calc J=I*%CU(I)
$calc     cvor_1(J)=cvor
$calc       fleet_1(J)=fleet
$delete cvor fleet
$calc       a1_1(J)=a1
$delete a1 $calc       a2_1(J)=a2
$delete a2 $calc       a3_1(J)=a3
$delete a3 $calc       a4_1(J)=a4
$delete a4 $calc       a5_1(J)=a5
$delete a5 $calc       d1_1(J)=d1
$delete d1 $calc       d2_1(J)=d2
$delete d2 $calc       d3_1(J)=d3
$delete d3 $calc       d4_1(J)=d4
$delete d4 $calc       d5_1(J)=d5
$delete d5 $calc       c1_1(J)=c1
$delete c1 $calc       c2_1(J)=c2
$delete c2 $calc       c3_1(J)=c3
$delete c3 $calc       c4_1(J)=c4
$delete c4 $calc       c5_1(J)=c5
$delete c5


$link L
$error Poisson
$c link I
$c error Normal

$c yvariate a5_1
$yvariate a4_1
```

```
$calc log_fleet_1=%log(fleet_1)
$offset log_fleet_1

$cycle  50 1
$c fit fleet_1+d1_1+d2_1+d3_1+c1_1+c2_1+c3_1
$fit fleet_1+d2_1+d3_1+c2_1+c3_1
$c fit fleet_1+d3_1+c3_1
$c fit a1_1+a2_1+a3_1+a4_1
$display e m t v c

$print 'Chi Square = Deviance /d.o.f   '  $look %x2$

$c the histotical part of Beyesian
$var %n history
$c calc history=(a4_1+a3_1+a2_1+a1_1)/4
$c  calc history=(a3_1+a2_1+a1_1)/3
$c calc history=a4_1
$calc history=(a3_1+a2_1)/2
$c calc history=a3_1
$c calculation of un-explained  variance

$var 1 unexp_var sum_unexp_var r_square
$var 1 exp_var sum_exp_var
$var %n one_avg

$c calculation of average
$calc avg=%cu(%fv)/%n
$calc one_avg=avg(%n)

$c calculation of un-explained  variance

$calc vec_unexp_var=(%fv - %yv)**2
$calc %x=%cu(vec_unexp_var)
$calc unexp_var=%x/%n
$calc sum_unexp_var=%x
$print 'un_explained variance   ' $look unexp_var sum_unexp_var

$c calculation of -explained  variance

$calc vec_exp_var=(%fv - one_avg)**2
$calc %x=%cu(vec_exp_var)/%n
$calc exp_var=%x
$calc sum_exp_var=%x*%n
$print 'explained variance   ' $look exp_var sum_exp_var

$print 'Linear Regression Calculations'
$var 1 r_square
```

```
$calc r_square=1-(unexp_var/(unexp_var+exp_var))
$look r_square

$c calculation of Model Pearson Chi_square

$var 1 mod_unexp_var mod_sum_unexp_var
$calc mod_vec_unexp_var=((%fv - %yv)**2) / %fv
$calc %x=%cu(mod_vec_unexp_var)
$calc mod_unexp_var=%x/(%n)
$calc mod_sum_unexp_var=%x
$print 'Model Pearson Chi_square' mod_sum_unexp_var

$c calculations of the Bayesian Estimate for the specific group

$var %n v_alpha

$calc var_m=(%fv-%yv)**2 - %fv
$c calc var_m_1=%if(var_m<0,0,var_m)
$plot var_m %fv

$extract %pe
$var %pl coeff
$c calc para=1,fleet,d2_1,d3_1,d4_1,c2_1,c3_1,c4_1
$c calc para=1,fleet,d1_1,d2_1,d3_1,c1_1,c2_1,c3_1
$calc coeff=%pe
$calc yv_1=%yv
$calc estimate=%fv
$calc square_est= %fv**2

$link I
$error Normal
$yvariate var_m
$calc my_os=0
$offset my_os
$calc %os=0
$fit square_est-1
$display e m
$plot my_os %os
$c ------------------
$delete my_os
$c ------------------
$calc v_alpha=1/(1+(%fv/estimate))
$c calc v_alpha=1/(1+(var_m_1/%fv))
$c calc v_alpha=alpha(1)
$var %n fv_1,temp1,temp2,temp
$calc temp1=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d2_1+coeff(4)*d3_1$
$calc temp2=coeff(5)*c2_1+coeff(6)*c3_1$
```

```
$c calc temp1=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d2_1+coeff(4)*d3_1$
$c calc temp2=coeff(6)*c2_1+coeff(7)*c3_1+coeff(8)*c4_1+coeff(5)*d4_1$
$calc temp=temp1+temp2$
$c ------------------
$c calc temp=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d2_1+coeff(4)*d3_1+coeff(5)*d4_1+coeff(6)*c2_1+coeff(7)*c3_1+coeff(8)*
$c calc temp=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d1_1+coeff(4)*d2_1+coeff(5)*d3_1+coeff(6)*c1_1+coeff(7)*c2_1+coeff(8)*
$c----------
$delete temp1 temp2 one_avg
$c----------
$calc fv_1=%exp(temp)*fleet_1
$calc bayes=v_alpha*(fv_1)+(1-v_alpha)*(history)
$c calc bayes=v_alpha*(estimate)+(1-v_alpha)*(history)
$c print 'alpha =' $c look alpha
$plot v_alpha %fv
$plot estimate fv_1
$c ---------------------------------------------
$c delete temp estimate fv_1
$delete temp  fv_1
$c ---------------------------------------------
$c calculation of Bayesian Pearson Chi_square

$var 1 bay_unexp_var bay_sum_unexp_var
$calc bay_vec_unexp_var=((bayes - yv_1)**2) /(estimate+%fv)
$c calc bay_vec_unexp_var=((bayes - yv_1)**2) /bayes
$c calc bay_vec_unexp_var=((bayes - a4_1)**2) /bayes
$calc %x=%cu(bay_vec_unexp_var)
$calc bay_unexp_var=%x/(%n)
$calc bay_sum_unexp_var=%x
$print 'Bayesian Pearson Chi_square' bay_sum_unexp_var


$c calculation of un-explained  variance for Bayesian Solution

$var 1 b_unexp_var sum_b_unexp_var

$calc vec_b_unexp_var=(bayes - yv_1)**2
$c calc vec_b_unexp_var=(bayes - a4_1)**2
$calc %x=%cu(vec_b_unexp_var)
$calc b_unexp_var=%x/%n
$calc sum_b_unexp_var=%x
$print 'Bayesian un_explained variance'  $look sum_b_unexp_var
$c ---------------------------------------------
$look coeff
$return
```

C.3 Samples of Bayes Model Programs and Comparison with Accident History Approach

```
%###########################################################################
file name : /u2/melherra/CVOR/Rating_1_10_a4_a5_2year_fv_Exclude_Zero_Mod2.glm
%###########################################################################
$data 142828 cvor fleet a1 a2 a3 a4 a5 d1 d2 d3 d4 d5 c1 c2 c3 c4 c5
$dinput 'cvor_fleet_accid_det_conv.out'
$output 'Rating_1_10_a4_a5_2year_fv_Exclude_Zero_Mod2.out'
$print 'Analysis for Rating_1_10_a4_a5_2year_fv_Exclude_Zero_Mod2';;;
$c -----------------------------------------------
$c calc I=(fleet>=1 & fleet<=5)
$calc I=(fleet>=1 & fleet<=10)
$c calc I=(fleet>=11 & fleet<=99)
$c calc I=(fleet>=11)
$c  calc I=(fleet==1)
$calc %n=%CU(I)
$var %n cvor_1 fleet_1 a1_1 a2_1 a3_1 a4_1 a5_1
$var %n         d1_1 d2_1 d3_1 d4_1 d5_1
$var %n         c1_1 c2_1 c3_1 c4_1 c5_1

$calc J=I*%CU(I)
$calc     cvor_1(J)=cvor
$calc       fleet_1(J)=fleet
$delete cvor fleet
$calc       a1_1(J)=a1
$delete a1 $calc       a2_1(J)=a2
$delete a2 $calc       a3_1(J)=a3
$delete a3 $calc       a4_1(J)=a4
$delete a4 $calc       a5_1(J)=a5
$delete a5 $calc       d1_1(J)=d1
$delete d1 $calc       d2_1(J)=d2
$delete d2 $calc       d3_1(J)=d3
$delete d3 $calc       d4_1(J)=d4
$delete d4 $calc       d5_1(J)=d5
$delete d5 $calc       c1_1(J)=c1
$delete c1 $calc       c2_1(J)=c2
$delete c2 $calc       c3_1(J)=c3
$delete c3 $calc       c4_1(J)=c4
$delete c4 $calc       c5_1(J)=c5
$delete c5
$c to calculate the number of carriers that have non-zero accident
$calc a5_1_no=(a5_1 ==0)
$calc %v=%CU(a5_1_no)
```

```
$print 'number of carriers with zero accident in a5 ='
$look %w

$c to calculate the number of carriers that have non-zero accident in year 4
$calc a4_1_no=(a4_1 ==0)
$calc %w=%CU(a4_1_no)
$print 'number of carriers with zero accident in a4 ='
$look %w

$c to calculate the number of carriers that have non-zero accident in years 3 & 4
$calc a_2years_no=(a4_1 ==0 & a3_1==0)
$calc %w=%CU(a_2years_no)
$print 'number of carriers with zero accident in a4 and a3  (2 years)='
$look %w

$c to calculate the total number of accident in year 5 for this fleet range
$calc sum_a5_1=%cu(a5_1)
$c look sum_a5_1
$print 'Number of accident in a5'
$print sum_a5_1(%n)

$c to calculate the total number of accident in year 4 for this fleet range
$calc sum_a4_1=%cu(a4_1)
$c look sum_a4_1
$print 'Number of accident in a4'
$print sum_a4_1(%n)

$c to calculate the total number of accident in year 3 for this fleet range
$calc sum_a3_1=%cu(a3_1)
$c look sum_a3_1
$print 'Number of accident in a3'
$print sum_a3_1(%n)

$c calc I=(fleet>=11 & fleet<=99)
$c calc bay_zero=(d1_1==0 & d2_1==0 & d3_1==0 & c1_1==0 & c2_1==0 & c3_1==0 & a4_1==0)
$calc bay_zero=(d1_1==0 & d2_1==0 & d3_1==0 & c1_1==0 & c2_1==0 & c3_1==0)
$c--- bay_zero_1=(bay_zero==1 & a4_1==0)
$c calc   %n=%CU(I)
$calc %q=%CU(bay_zero)
$c--- calc %v=%CU(bay_zero_1)
$look %q
$c--- look %v
$c---------------------------------------------
$delete a5_1_no a4_1_no a_2years_no bay_zero
$delete sum_a5_1 sum_a4_1 sum_a3_1
$c---------------------------------------------
```

```
$link L
$error Poisson
$c link I
$c error Normal

$c yvariate a5_1
$yvariate a4_1
$calc log_fleet_1=%log(fleet_1)
$offset log_fleet_1

$cycle  50 1
$c fit fleet_1+d2_1+d3_1+d4_1+c2_1+c3_1+c4_1
$fit fleet_1+d1_1+d2_1+d3_1+c1_1+c2_1+c3_1

$c fit a1_1+a2_1+a3_1+a4_1
$display e m t v c

$print 'Chi Square = Deviance /d.o.f    '  $look %x2$

$c the histotical part of Beyesian
$var %n history
$c calc history=(a4_1+a3_1+a2_1+a1_1)/4
$c  calc history=(a3_1+a2_1+a1_1)/3
$calc history=(a4_1+a3_1)/2
$c ---***calc history=a4_1
$c calc history=a3_1
$c calculation of un-explained  variance

$var 1 unexp_var sum_unexp_var r_square
$var 1 exp_var sum_exp_var
$var %n one_avg

$c calculation of average
$calc avg=%cu(%fv)/%n
$calc one_avg=avg(%n)

$c calculation of un-explained  variance

$calc vec_unexp_var=(%fv - %yv)**2
$calc %x=%cu(vec_unexp_var)
$calc unexp_var=%x/%n
$calc sum_unexp_var=%x
$print 'un_explained variance   ' $look unexp_var sum_unexp_var

$c calculation of -explained  variance

$calc vec_exp_var=(%fv - one_avg)**2
```

```
$calc %x=%cu(vec_exp_var)/%n
$calc exp_var=%x
$calc sum_exp_var=%x*%n
$print 'explained variance    ' $look exp_var sum_exp_var

$print 'Linear Regression Calculations'
$var 1 r_square
$calc r_square=1-(unexp_var/(unexp_var+exp_var))
$look r_square

$c calculation of Model Pearson Chi_square

$var 1 mod_unexp_var mod_sum_unexp_var
$calc mod_vec_unexp_var=((%fv - %yv)**2) / %fv
$calc %x=%cu(mod_vec_unexp_var)
$calc mod_unexp_var=%x/(%n)
$calc mod_sum_unexp_var=%x
$print 'Model Pearson Chi_square' mod_sum_unexp_var

$c calculations of the Bayesian Estimate for the specific group

$var %n v_alpha

$calc var_m=(%fv-%yv)**2 - %fv
$c calc var_m_1=%if(var_m<0,0,var_m)
$plot var_m %fv

$extract %pe
$var %pl coeff
$c calc para=1,fleet,d2_1,d3_1,d4_1,c2_1,c3_1,c4_1
$c calc para=1,fleet,d1_1,d2_1,d3_1,c1_1,c2_1,c3_1
$calc coeff=%pe
$calc yv_1=%yv
$calc estimate=%fv
$calc square_est= %fv**2

$link I
$error Normal
$yvariate var_m
$calc my_os=0
$offset my_os
$calc %os=0
$fit square_est-1
$display e m
$plot my_os %os
$c ------------------
$delete my_os
```

```
$c ------------------
$calc v_alpha=1/(1+(%fv/estimate))
$c calc v_alpha=1/(1+(var_m_1/%fv))
$c calc v_alpha=alpha(1)
$var %n fv_1,temp1,temp2,temp
$calc temp1=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d2_1+coeff(4)*d3_1$
$calc temp2=coeff(6)*c2_1+coeff(7)*c3_1+coeff(8)*c4_1+coeff(5)*d4_1$
$calc temp=temp1+temp2$
$c ------------------
$c calc temp=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d2_1+coeff(4)*d3_1+coeff(5)*d4_1+coeff(6)*c2_1+coeff(7)*c3_1+coeff(8)*
$c calc temp=coeff(1)*1+coeff(2)*fleet_1+coeff(3)*d1_1+coeff(4)*d2_1+coeff(5)*d3_1+coeff(6)*c1_1+coeff(7)*c2_1+coeff(8)*
$c----------
$delete temp1 temp2 one_avg
$c----------
$calc fv_1=%exp(temp)*fleet_1
$calc bayes=v_alpha*(fv_1)+(1-v_alpha)*(history)
$c calc bayes=v_alpha*(estimate)+(1-v_alpha)*(history)
$c print 'alpha =' $c look alpha
$plot v_alpha %fv
$plot estimate fv_1
$c ---------------------------------------------
$delete temp estimate fv_1
$c ---------------------------------------------
$c calculation of Bayesian Pearson Chi_square

$var 1 bay_unexp_var bay_sum_unexp_var
$calc bay_vec_unexp_var=((bayes - yv_1)**2) /bayes
$c calc bay_vec_unexp_var=((bayes - a4_1)**2) /bayes
$calc %x=%cu(bay_vec_unexp_var)
$calc bay_unexp_var=%x/(%n)
$calc bay_sum_unexp_var=%x
$print 'Bayesian Pearson Chi_square' bay_sum_unexp_var


$c calculation of un-explained  variance for Bayesian Solution

$var 1 b_unexp_var sum_b_unexp_var

$calc vec_b_unexp_var=(bayes - yv_1)**2
$c calc vec_b_unexp_var=(bayes - a4_1)**2
$calc %x=%cu(vec_b_unexp_var)
$calc b_unexp_var=%x/%n
$calc sum_b_unexp_var=%x
$print 'Bayesian un_explained variance'   $look sum_b_unexp_var
$c ---------------------------------------------
$delete one_avg vec_unexp_var vec_exp_var mod_vec_unexp_var
$delete bay_vec_unexp_var vec_b_unexp_var yv_1
```

```
$c ---------------------------------------------
$c sorting the data according to the Bayesian Rate
$var %n xx new_xx new_cvor_1
$calc xx=a5_1
$calc b_rate=bayes/fleet_1
$c sort  z 1 bayes
$sort  z 1 b_rate

$calc %i=%n
$c calc %i=%tr(0.02 * %n)
$calc %l=1

$c var %i new_bayes new_xx new_yv sum_yv new_1_bayes
$macro test
$c print %i
$c calc new_1_bayes(%i)=bayes(z(%i))

$c calc new_bayes(%l)=bayes(z(%i))
$c calc new_b_rate(%l)=b_rate(z(%i))
$calc new_xx(%l)=xx(z(%i))
$calc  sum_xx=%cu(new_xx)
$calc new_cvor_1(%l)=cvor_1(z(%i))

$c calc sort_alpha(%l)=v_alpha(z(%i))
$c calc sort_var_m(%l)=var_m(z(%i))

 $calc %i=%i-1$
$calc %l=%l+1
$endmac

$while %i test$
$c ---------------------------------------------
$delete bayes b_rate v_alpha var_m
$c ---------------------------------------------
$c sorting the data according to the Lat Year Accident_Rate
$var %n x_axix  a4_xx a4_1_m
$calc a4_rate=(a4_1+a3_1)/(2*fleet_1)
$c calc a3_rate=a3_1/fleet_1
$c sort  z 1 bayes
$sort  z 1 a4_rate

$calc %i=%n
$c calc %i=%tr(0.02 * %n)
$calc %l=1

$c var %i new_bayes new_xx sum_xx new_1_bayes
$macro test1
```

```
$c print %i
$c calc new_1_bayes(%i)=bayes(z(%i))

$calc a4_xx(%l)=xx(z(%i))
$calc a4_1_m(%l)=a4_1(z(%i))
$calc  sum_a4_xx=%cu(a4_xx)
$c calc a4_cvor_1(%l)=%if(a4_1_m(%l)==0,0,cvor_1(z(%i)))
$calc a4_cvor_1(%l)=cvor_1(z(%i))

 $calc %i=%i-1$
$calc %l=%l+1
$calc x_axis(%l-1)=%l-1
$endmac

$while %i test1$
$c ----------------------------------------------
$calc diff=sum_xx-sum_a4_xx
$plot diff x_axis
$look diff
$c ----------------------------------------------
$c ------------a4_xx----------------------------------
$delete x_axix a4_rate
$c ----------------------------------------------
$c sorting the data according to the  Year 1995 Accident_Rate
$c ----var %n x_axix
$calc a5_rate=a5_1/fleet_1
$c calc a3_rate=a3_1/fleet_1
$c sort  z 1 bayes
$sort  z 1 a5_rate

$calc %i=%n
$c calc %i=%tr(0.02 * %n)
$calc %l=1

$c var %i new_bayes new_xx sum_xx new_1_bayes

$var %n  a5_xx
$macro test2
$c print %i
$c calc new_1_bayes(%i)=bayes(z(%i))

$calc a5_xx(%l)=xx(z(%i))
$c ----calc  sum_a5_xx=%cu(a5_xx)
$calc a5_cvor_1(%l)=cvor_1(z(%i))

 $calc %i=%i-1$
$calc %l=%l+1
```

```
$c ----calc x_axis(%l-1)=%l-1
$endmac

$while %i test2$
$c ---------------------------------------------

$c look sum_xx
$c look new_b_rate b_rate new_bayes new_xx sum_xx
$c look sum_xx sum_a4_xx
$c plot (style=1 ylimit=0,500 xlimit=0,1000) sum_xx sum_a4_xx  'ba' x_axis
$c plot sum_xx sum_a4_xx  'ba' x_axis

$c -----look diff
$c ----look estimate fv_1
$c look a5_cvor_1 new_cvor_1 a4_cvor_1 a4_1_m a4_xx  a5_xx
$look a5_cvor_1 new_cvor_1 a4_cvor_1
$look coeff
$c look diff sort_var_m sort_alpha new_bayes new_xx a4_xx sum_xx sum_a4_xx

$c ---------------------------------------------
$c ----------------Addition-----------------------------
$c ---------------------------------------------
$c ---------------------------------------------
$c-- Initializing %i (top 2%) in outer loop (a5)

$return
```

```
%###########################################################################
file name : /u2/melherra/CVOR/Model_develop/ranking_1_10_truck.c
%###########################################################################
#define TheYear
#define MAX 7000
#define MaxLine 1000
#include <stdio.h>
#include <math.h>
#include <string.h>
FILE *stOut, *stIn_1, *stIn_2;
main(int argc,char *argv[])
{
int i,j,k,l,m,n,nn,exist,bayes_count,a4_count,exist1,both_count;
/* int a5_CVOR[MAX],bayes_CVOR[MAX],a4_CVOR[MAX],junk_number,temp_a5;*/
char *a5_CVOR[MAX][30],*bayes_CVOR[MAX][30],*a4_CVOR[MAX][30],*junk_number[30],*p_temp_a5[30],temp_a5[30];
/* char *a5_CVOR[MAX],*bayes_CVOR[MAX],*a4_CVOR[MAX],*junk_number,*temp_a5;*/
char input_file_name_1[50],input_file_name_2[50],output_file_name[50];
char basic_1[200],basic_2[200],blaa[10];
strcpy(input_file_name_1,argv[1]);
strcpy(output_file_name,argv[2]);
/*
strcpy(input_file_name_1,"ranking_1_10_truck_1600.inp");
strcpy(output_file_name,"ranking_1_10_truck_1600.out");
*/
stIn_1=fopen(input_file_name_1, "r");
stOut=fopen(output_file_name, "w");

printf("Starting ...");
sscanf(argv[3],"%d",&n);
/*n=1600;*/
nn=3382;
bayes_count=0;
a4_count=0;
both_count=0;

fscanf(stIn_1,"%s %s %s ",blaa,blaa,blaa);
printf("Starting ...");
for (k=0;k<nn;k++){
fscanf(stIn_1,"%s",junk_number);
fscanf(stIn_1,"%s",a5_CVOR[k]) ;
fscanf(stIn_1,"%s",bayes_CVOR[k]) ;
fscanf(stIn_1,"%s",a4_CVOR[k]) ;
/*fscanf(stIn_1,"%s %s %s",blaa,blaa,blaa);*/
/*printf("%s\t %s\t %s\t %s\n",junk_number,a5_CVOR[k],bayes_CVOR[k],a4_CVOR[k]);
fprintf(stOut,"%s\t %s\t %s\t %s\n",junk_number,a5_CVOR[k],bayes_CVOR[k],a4_CVOR[k]);*/
}
```

```
for (k=0;k<nn;k++){
/*strcpy(p_temp_a5,a5_CVOR[k]);*/

/*strcpy(temp_a5,"*a5_CVOR[k]");   OK but take literally the string*/
/*strcpy(temp_a5,a5_CVOR[k]);   compiled but warning:passing arg 2 of 'strcpy' from incompatible pointer type*/
/*strcpy(temp_a5,*a5_CVOR[k]);  compiled but segmentation fault */
strcpy(temp_a5,a5_CVOR[k]);
/*
printf("%d\t %s\n",k+1,temp_a5);
fprintf(stOut,"%d\t %s\n",k+1,temp_a5);
*/
i =0;
j=0;
exist=0;
exist1=0;

/*
while((exist==0) && (i<n)){
if (strcmp("*a5_CVOR[k]","*bayes_CVOR[i]")){
bayes_count=bayes_count+1;
exist=1;
}
i=i+1;
}
*/
while(!((exist==1) || (i>=n))){
if (!strcmp(temp_a5,bayes_CVOR[i])){
bayes_count=bayes_count+1;
        exist=1;
/*fprintf(stOut,"bayes %d\t %s\t %s\n",i,temp_a5,bayes_CVOR[i]);
printf("%d\t %s\t %s\n",i,temp_a5,bayes_CVOR[i]);*/
}
i=i+1;
}



/*while((exist1==0) && (j<n)){*/
while(!((exist1==1) || (j>=n))){
if (!strcmp(temp_a5,a4_CVOR[j])){
/*if (strcmp("*a5_CVOR[k]","*a4_CVOR[j]")){*/
/*if (a5_CVOR[k]==a4_CVOR[j]){*/
        a4_count=a4_count+1;
        exist1=1;
/*fprintf(stOut,"a4a4 %d\t %s\t %s\n",j,temp_a5,a4_CVOR[j]);
printf("%d\t %s\t %s\n",j,temp_a5,a4_CVOR[j]);*/
}
```

```
j=j+1;
}

if ((exist==1) && (exist1!=1)){
/*fprintf(stOut,"In bayesButNOTina4 %s\n",temp_a5);*/
}

if ((exist==1) && (exist1==1)){
both_count=both_count+1;
/*fprintf(stOut,"both_count %d\n",both_count);*/
}

}
/*
fprintf(stOut,"%s\n",argv[2]);
fprintf(stOut,"bayes_count %d\n",bayes_count);
fprintf(stOut,"a4_count %d\n",a4_count);
fprintf(stOut,"both_count %d\n",both_count);
*/

/*fprintf(stOut,"%s\n\n",argv[2]);*/
fprintf(stOut,"No_audited\t bayes_count\t a4_count\t only_Bayes\t  both_count\n");
fprintf(stOut,"%d\t %d\t %d\t %d\t %d\n\n",n,bayes_count,a4_count,bayes_count-both_count,both_count);
/*
fprintf(stOut,"==============================================================\n");
*/

/*for (k=0;k<n;k++){
fprintf(stOut,"%s\t %s\t %s\t",a5_CVOR[k],bayes_CVOR[k],a4_CVOR[k]);
}*/

/*printf("%d\t %d\t,bayes_count,a4_count);*/
fclose(stIn_1);

fclose(stOut);


}
```

C.4 Samples of Programs Comparing Random Aprroach to Bayes Model

```
%##########################################################################
file name : /u2/melherra/CVOR/Shortreed_Analysis/GENERIC/my_random_all.c
%##########################################################################
#define TheYear
#define MAX 60000
#define MaxLine 1000
#include <stdio.h>
#include <math.h>
#include <string.h>
#include <stdlib.h>
FILE *stOut_1,*stOut_2,*stOut_3,*stOut_4,*stOut_5,*stIn_1, *stIn_2;
FILE *stOut[5];
main(int argc,char *argv[])
{
int i,j,k,l,m,mm,n,exist,bayes_count,a4_count,exist1,fleet[MAX];
/* int a5_CVOR[MAX],bayes_CVOR[MAX],a4_CVOR[MAX],junk_number,temp_a5;
char *a5_CVOR[MAX][30],*bayes_CVOR[MAX][30],*a4_CVOR[MAX][30],*junk_number[30],*temp_a5[30];*/
char input_file_name_1[50],input_file_name_2[50],output_file_name[10][50],output_file_name_1[50];
char basic_1[200],basic_2[200],blaa[10],cvor[MAX][50];
unsigned int seed=153;
/*unsigned int seed=system("date +%H%M%S");*/
srand(seed);

strcpy(input_file_name_1,argv[1]);
strcpy(output_file_name_1,argv[2]);
/*
strcpy(input_file_name_1,"cvor_fleet_10.out");
strcpy(output_file_name_1,"random_10.out");
*/
stIn_1=fopen(input_file_name_1, "r");
stOut_1=fopen(output_file_name_1, "w");
i=0;
while(!feof(stIn_1)){
i=i+1;
fscanf(stIn_1,"%s %d",cvor[i],&fleet[i]);
}

/*m=system('expr argv[3] + 0');to have the number of the Carriers randomly selected)*/
//m=argv[3];
//printf("%s",argv[3]);
sscanf(argv[3],"%d",&mm);
```

```
sscanf(argv[4],"%d",&m);
//m=10;
for (i=0;i<m;i++){
k=random(1,mm);
l=rand();
printf("%d %d\n",k,l);

fprintf(stOut_1,"%s %d\n",cvor[k],fleet[k]);
}

fclose(stIn_1);
fclose(stOut_1);
}

int random(int first,int last)
{
/* this function returens a random integer from first to last,inclusive*/
int offset;
offset=rand()/(RAND_MAX+1.0)*(last-first+1);
return(first+offset);
}
```

```
%############################################################################
file name : /u2/melherra/CVOR/Shortreed_Analysis/GENERIC/modified.c
%############################################################################
/* SYNTAX : modified.run output_file number_n */
#define TheYear
#define MAX 7000
#define MaxLine 1000
#include <stdio.h>
#include <math.h>
#include <string.h>
FILE *stOut, *stIn_1, *stIn_2;
main(int argc,char *argv[])
{
int i,j,k,l,m,n,nn,exist,bayes_count,a4_count,exist1,both_count,random_count;
/* int a5_CVOR[MAX],bayes_CVOR[MAX],a4_CVOR[MAX],junk_number,temp_a5;*/
char *a5_CVOR[MAX][30],*random_CVOR[MAX][30],*bayes_CVOR[MAX][30],*a4_CVOR[MAX][30],*junk_number[30],*p_temp_a5[30],temp
/* char *a5_CVOR[MAX],*bayes_CVOR[MAX],*a4_CVOR[MAX],*junk_number,*temp_a5;*/
char input_file_name_1[50],input_file_name_2[50],output_file_name[50];
char basic_1[200],basic_2[200],blaa[10];
strcpy(input_file_name_1,"test_1_10truck.inp");
strcpy(input_file_name_2,"random.inp");
strcpy(output_file_name,argv[1]);

stIn_1=fopen(input_file_name_1, "r");
stIn_2=fopen(input_file_name_2, "r");
stOut=fopen(output_file_name, "w");

printf("Starting ...");
//n=3382;
sscanf(argv[2],"%d",&n);
nn=3382;
bayes_count=0;
a4_count=0;
random_count=0;;
both_count=0;

printf("Starting ...");
for (k=0;k<nn;k++){
fscanf(stIn_2,"%s",random_CVOR[k]) ;
fscanf(stIn_2,"%s",junk_number);
}
for (k=0;k<nn;k++){
fscanf(stIn_1,"%s",junk_number);
fscanf(stIn_1,"%s",a5_CVOR[k]) ;
fscanf(stIn_1,"%s",bayes_CVOR[k]) ;
fscanf(stIn_1,"%s",a4_CVOR[k]) ;
printf("%s\t %s\t %s\t %s\n",junk_number,a5_CVOR[k],bayes_CVOR[k],a4_CVOR[k]);
```

```
fprintf(stOut,"%s\t %s\t %s\t %s\n",junk_number,a5_CVOR[k],bayes_CVOR[k],a4_CVOR[k]);
}

for (k=0;k<nn;k++){
/*strcpy(p_temp_a5,a5_CVOR[k]);*/

/*strcpy(temp_a5,"*a5_CVOR[k]");   OK but take literally the string*/
/*strcpy(temp_a5,a5_CVOR[k]);   compiled but warning:passing arg 2 of 'strcpy' from incompatible pointer type*/
/*strcpy(temp_a5,*a5_CVOR[k]);  compiled but segmentation fault */
strcpy(temp_a5,a5_CVOR[k]);

printf("%d\t %s\n",k+1,temp_a5);
fprintf(stOut,"%d\t %s\n",k+1,temp_a5);
i =0;
j=0;
exist=0;
exist1=0;

/*
while((exist==0) && (i<n)){
if (strcmp("*a5_CVOR[k]","*bayes_CVOR[i]")){
bayes_count=bayes_count+1;
exist=1;
}
i=i+1;
}
*/
while(!((exist1==1) || (i>=n))){
if (!strcmp(temp_a5,bayes_CVOR[i])){
bayes_count=bayes_count+1;
        exist=1;
fprintf(stOut,"bayes %d\t %s\t %s\n",i,temp_a5,bayes_CVOR[i]);
printf("%d\t %s\t %s\n",i,temp_a5,bayes_CVOR[i]);
}
i=i+1;
}



/*while((exist1==0) && (j<n)){*/
while(!((exist1==1) || (j>=n))){
if (!strcmp(temp_a5,random_CVOR[j])){
/*if (strcmp("*a5_CVOR[k]","*random_CVOR[j]")){*/
/*if (a5_CVOR[k]==random_CVOR[j]){*/
        random_count=random_count+1;
        exist1=1;
fprintf(stOut,"randomrandom %d\t %s\t %s\n",j,temp_a5,random_CVOR[j]);
```

```
printf("%d\t %s\t %s\n",j,temp_a5,random_CVOR[j]);
}
j=j+1;
}

if ((exist==1) && (exist1!=1)){
fprintf(stOut,"In bayesButNOTinrandom %s\n",temp_a5);
}

if ((exist==1) && (exist1==1)){
both_count=both_count+1;
fprintf(stOut,"both_count %d\n",both_count);
}

fprintf(stOut,"bayes_count %d\n",bayes_count);
fprintf(stOut,"random_count %d\n",random_count);
}

fprintf(stOut,"========================================================\n");

/*for (k=0;k<n;k++){
fprintf(stOut,"%s\t %s\t %s\t",a5_CVOR[k],bayes_CVOR[k],random_CVOR[k]);
}*/

/*printf("%d\t %d\t,bayes_count,random_count);*/
fclose(stIn_1);
fclose(stIn_2);
fclose(stOut);


}
```

```
%#######################################################################
file name : /u2/melherra/CVOR/Shortreed_Analysis/GENERIC/back_glim_1_10_truck_random.c
%#######################################################################
#define TheYear
#define MAX 7000
#define MaxLine 1000
#include <stdio.h>
#include <math.h>
#include <string.h>
FILE *stOut, *stIn_1, *stIn_2;
main(int argc,char *argv[])
{
int i,j,k,l,m,n,nn,exist,bayes_count,a4_count,exist1,both_count,random_count;
/* int a5_CVOR[MAX],bayes_CVOR[MAX],a4_CVOR[MAX],junk_number,temp_a5;*/
char *a5_CVOR[MAX][30],*random_CVOR[MAX][30],*bayes_CVOR[MAX][30],*a4_CVOR[MAX][30],*junk_number[30],*p_temp_a5[30],temp
/* char *a5_CVOR[MAX],*bayes_CVOR[MAX],*a4_CVOR[MAX],*junk_number,*temp_a5;*/
char input_file_name_1[50],input_file_name_2[50],output_file_name[50];
char basic_1[200],basic_2[200],blaa[10];
strcpy(input_file_name_1,"test_1_10truck_random.inp");
strcpy(output_file_name,argv[1]);
stIn_1=fopen(input_file_name_1, "r");
stOut=fopen(output_file_name, "w");

printf("Starting ...");
n=3382;
nn=3382;
bayes_count=0;
a4_count=0;
random_count=0;;
both_count=0;

printf("Starting ...");
for (k=0;k<nn;k++){
fscanf(stIn_1,"%s",random_CVOR[k]) ;
fscanf(stIn_1,"%s",junk_number);
fscanf(stIn_1,"%s",junk_number);
fscanf(stIn_1,"%s",a5_CVOR[k]) ;
fscanf(stIn_1,"%s",bayes_CVOR[k]) ;
fscanf(stIn_1,"%s",a4_CVOR[k]) ;
printf("%s\t %s\t %s\t %s\n",junk_number,a5_CVOR[k],bayes_CVOR[k],a4_CVOR[k]);
fprintf(stOut,"%s\t %s\t %s\t %s\n",junk_number,a5_CVOR[k],bayes_CVOR[k],a4_CVOR[k]);
}

for (k=0;k<nn;k++){
/*strcpy(p_temp_a5,a5_CVOR[k]);*/

/*strcpy(temp_a5,"*a5_CVOR[k]");   OK but take literally the string*/
```

```
/*strcpy(temp_a5,a5_CVOR[k]);   compiled but warning:passing arg 2 of 'strcpy' from incompatible pointer type*/
/*strcpy(temp_a5,*a5_CVOR[k]);  compiled but segmentation fault */
strcpy(temp_a5,a5_CVOR[k]);

printf("%d\t %s\n",k+1,temp_a5);
fprintf(stOut,"%d\t %s\n",k+1,temp_a5);
i =0;
j=0;
exist=0;
exist1=0;

/*
while((exist==0) && (i<n)){
if (strcmp("*a5_CVOR[k]","*bayes_CVOR[i]")){
bayes_count=bayes_count+1;
exist=1;
}
i=i+1;
}
*/
while(!((exist1==1) || (i>=n))){
if (!strcmp(temp_a5,bayes_CVOR[i])){
bayes_count=bayes_count+1;
        exist=1;
fprintf(stOut,"bayes %d\t %s\t %s\n",i,temp_a5,bayes_CVOR[i]);
printf("%d\t %s\t %s\n",i,temp_a5,bayes_CVOR[i]);
}
i=i+1;
}



/*while((exist1==0) && (j<n)){*/
while(!((exist1==1) || (j>=n))){
if (!strcmp(temp_a5,random_CVOR[j])){
/*if (strcmp("*a5_CVOR[k]","*random_CVOR[j]")){*/
/*if (a5_CVOR[k]==random_CVOR[j]){*/
        random_count=random_count+1;
        exist1=1;
fprintf(stOut,"randomrandom %d\t %s\t %s\n",j,temp_a5,random_CVOR[j]);
printf("%d\t %s\t %s\n",j,temp_a5,random_CVOR[j]);
}
j=j+1;
}

if ((exist==1) && (exist1!=1)){
fprintf(stOut,"In bayesButNOTinrandom %s\n",temp_a5);
```

```
}

if ((exist==1) && (exist1==1)){
both_count=both_count+1;
fprintf(stOut,"both_count %d\n",both_count);
}

fprintf(stOut,"bayes_count %d\n",bayes_count);
fprintf(stOut,"random_count %d\n",random_count);
}

fprintf(stOut,"==================================================================\n");

/*for (k=0;k<n;k++){
fprintf(stOut,"%s\t %s\t %s\t",a5_CVOR[k],bayes_CVOR[k],random_CVOR[k]);
}*/

/*printf("%d\t %d\t,bayes_count,random_count);*/
fclose(stIn_1);

fclose(stOut);


}
```

References

Accident Data Systems (ADS) for the years 1991 to 1995 of The Province of Ontario. Ministry of Transportation (MTO), Ontario, Canada.

Abbess, C., Jarrett, D., Wright, C.C. (1981). *Accidents at blackspots: Estimating the effectiveness of remedial treatment, with special reference to the " regression-to-the-mean" effect.* Traffic Eng. Control 22, pp 535-542.

Anscombe, F.J. (1950). *Sampling Theory of the Negative Binomial and Logarithmic Series Distributions.* Biometrika, 37, pp 358-382.

Baker, R.J. and Nedler, J. A. (1978). *The GLIM System Manual.* The Numerical Algorithms Group, England.

Bonneson, J.A. and McCoy, P.T. (1997). *Effect of Median Treatment on Urban Arterial Safety: An Accident Prediction Model.* Transportation Research Board, 76th annual meeting, Jan.12-16,1997, Washington, D.C.

Braver, E.R., Zador P.L., Thum D., Mitter E.L., Baum H.M. and Vilardo F.J. (1997). *Tractor-Trailer Crashes in Indiana: A Case-Control Study of The Role of Truck Configuration.* Accid. Anal. & Prev. Vol. 29, No. 1, pp. 79-96.

Breslow, N. E. *Extra-Poisson Variation in Log-Linear Models.* Appl. Statist., 1984, 33, No. 1, pp 38-44.

Burg, A. (1970). *The Stability of Driving Record Over Time.* Accid. Anal. & Prev. 2, pp. 57-65.

Buyco, C. and Saccomanno F.F. (1988).*Analysis of Truck Accident Rates Using Log-linear Models.* Can. J. Civ. Eng. 15, pp397-408.

Cameron. A. C. and Trivedi, P. K. (1986). *Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators and Tests.* Journal of Applied Econometrics. 1. pp 29-53.

Carrier Vehicle Operator Record (CVOR) for the years 1991 to 1995 of The Province of Ontario. Ministry of Transportation (MTO). Ontario, Canada.

Chipman. M.L. and Morgan. P.P. (1976). *The Driver Demerit Point in Ontario as a Long-term predictor of Collisions.* Proceedings of the 20th Conference. Association of the Advancement of Automotive Medicine. Morton Grove, IL: AAAM, pp 45-51.

Chirachavala. T. and Cleveland, D. (1985). *Causal Analysis of Accident involvements for the nation's largest truck and combination vehicles.* Transportation Research Record 847. pp37-42.

Collings. B. J. and Margolin, B. H. (1985). *Testing of Goodness of Fit for the Poisson Assumption when Observations are not identically Distributed.* Journal of the American Statistical Association, June 1985, Vol. 80. No. 390, pp 411-418.

Corsi. T. M. and Fanara, P. (1988). *Driver Management Policies and Motor Carrier Safety.* Logistics and Transportation Review. Vol. 24, N0. 2, pp 153-163.

Cox, D. R. (1983). *Some Remarks on Over-Dispersion.* Biometrika, 70. pp 269-274.

Dean. C. and Lawless, J. F. (1989). *Tests for Detecting Overdispersion in Poisson Regression Models.* Journal of the American Statistical Association, June 1989, Vol. 84, No. 406. Theory and Methods.

El-Herraoui, M. and Saccomanno F. F. (1997). *Estimation of Large Truck Accident Rates Using Commercial Vehicle Survey (CVS) Station Counts.* Presented at the Canadian Transportation Research Forum in June 1997.

Federal Highway Administration (FHWA) (1995a). Highway Statistics 1994. U.S. Department of Transportation, Washington, D.C.

Federal Highway Administration (FHWA) (1995b). Truck and Bus Accident Factbook 1993. Washington, DC. U.S. Department of Transportation, Washington, D.C.

Fisher. R. A. (1954). *Statistical Methods for Research Workers.* Oliver and Boyd.

Haight. F. (1964). *Accident Proneness. The History of an Idea.* Automobilismo 12. pp 534-546.

Hall, R. D. (1986). *Accidents at Four-Arm Single Carriageway Urban Traffic Signals.* /TRRL Contractor Report 65. U.K. Transport and Road Research Laboratory. Crowthorne. Berkshire. England, 1986.

Harano. R.M.. Peck. R.C. and McBride, R. S. (1975). *The Prediction of Accident Liability Through Biographical Data and pshycometric tests.* J Saf. Res. 7, pp 16-52.

Hauer. E. (1992). *Empirical Bayes Approach to The Estimation of "Unsafety": The Multivariate Regression Method.* Accid. Anal. & Prev. Vol. 24, No. 5, pp. 457-477.

Hauer. E.. Persaud B.N., Smiley A. and Duncan D. (1991).*Estimation The Accident Potential of An Ontario Driver.* Accid. Anal. & Prev. Vol. 23, Nos. 2/3, pp. 133-152.

Hauer. E. and Persaud B.N. (1987).*How to Estimate The Safety of Rail-Highway Grade Crossings and The Safety Effects of Warning Devices.* Transportation Research Record 1114. pp. 131-139.

Hauer, E. (1986). *On The Estimation of The Expected Number of Accidents.* Accid.

Anal. & Prev. Vol. 18, No. 1, pp. 1-12.

Hickey. J.J. (1997). Shoulder Rumble Strip Effectiveness: Drift-Off-Road Accident Reductions on the Pennsylvania Turnpike. Transportation Research Board, 76th annual meeting, Jan.12-16,1997, Washington, D.C.

Hogg R.V. and Craig, A.T. (1978).*Introduction to Mathematical Statistics.* Macmillan Publishing Co., Inc.

Insurance Institute for Highway Safety (IIHS) (1996). Large Trucks. Fatality Facts. Insurance Institute for Highway Safety, Arlington, VA.

Joshua S. C. and Garber, N. J. (1990). *Estimating Truck Accident Rate and Involvements Using Linear and Poisson Regression models.* Transport. Planning Technol. 15: pp 41-58.

Lawless. J. F. *Negative Binomial and mixed Poisson Regression.* The Canadian Journal of Statistics, Vol. 15, No. 3, 1987, pp 209-225.

Maher. M.J. and Summersgill, I. (1996). *A Comprehensive Methodology for The Fitting of Predictive Accident Models.* Accid. Anal. and Prev., Vol. 28, No. 3. pp. 281-296.

Maycock, G. and Hall, R. D. (1984). *Accident at 4-Arm Round-abouts.* TRRL Laboratory Report 1120. U.K. Transport and Road Research Laboratory. Crowthorne. Berkshire. England, 1984.

McGullan. P., Nedler, J.A., (1983).*Generalized Linear Models.* Chapman and Hall, London.

Miller, I.R., Freud, J.E. and Johnson, R. (1990). Probability and Statistics For Engineers. Prentice-Hall, Inc.

Miaou. S.P. (1994). *The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson versus Negative Binomial Regressions.* 73rd Annual Meeting Transportation Research Board, Washington, D.C.

Miaou, S.P. and Lum H. (1993). *Modeling Vehicle Accidents and Highway Geometric Design Relationships.* Accid. Anal. & Prev. Vol. 25, pp. 689-709.

Ministry of Transportation (MTO). *The Social Cost of Motor Vehicle Crashes in Ontario.* Safety Research Office. Safety Policy Branch in cooperation with the Research and Development Branch. March 1994, SRO-94-101, ISBN 0-7778-2616-X.

Miaou. S.P.. Hu. P.S., Wright, T., Rathi, A.K. and Davis, S.C. (1992). *Relationships between Truck Accidents and Highway Geometric Design: A Poisson Regression Approach.* Tranp. Res. record 1376. pp 10-18.

Moses. L.N. and Savage. I. (1992).*The Effectiveness of Motor Carrier Safety Audits.* Accid. Anal. & Prev. Vol. 24, No. 5. pp. 479-496.

Moses. L.N. and Savage. I. (1994).*The Effect of Firm Characteristics on Truck Accidents.* Accid. Anal. & Prev. Vol. 26, No. 2. pp. 173-179.

Nembhard D.A. and Young M.R. (1995).*Parametric Empirical Bayes Estimates of Truck Accident Rates.* Journal of Transportation Engineers, July/August 1995, pp. 359-363.

Okamoto. H. and Koshi, M. (1989). *A Method to Cope with The Random Errors of Observed Accident Rates in Regression Analysis.* Accid. Anal. & Prev. Vol. 21, pp. 317-332.

Ontario Road Safety Annual Report (1994). Safety Research Office, Ministry of Transportation, Ontario, Canada.

Patten, M.L., Carroll, J.L. and Thomchick, E.A. (1989). *The Efficacy of roadside Inspections in Reducing Heavy Truck Accidents.* Journal of Transportation Research Forum, 29, pp. 269-276.

Peck, R.C., McBride, R.S. and Coppin, R.S. (1971). *The Distribution and Prediction of Driver Accident Frequencies.* Accid. Anal. & Prev. 2, pp. 243-299.

Peck, R.C. and Kuan, J.A. (1983). *A Statistical Model of Individual Accident Risk prediction Using Driver Record, Territory and Other Biographical Factors.* Accid. Anal. & Prev. 15, pp. 371-393.

Pickering, D., Hall R. D., and Grimmer, M. (1986). *Accidents at Rural T-Junctions.* TRRL Research Report 65. U.K. Transport and Road Research Laboratory. Crowthorne, Berkshire, England, 1986.

Robbins, H. (1956). *An Empirical Bayes Approach to Statistics.* 3rd Berkeley Symposium on Math. Statist. Prob. 1, 157-163.

Robbins, H. (1980). *An Empirical Bayes Estimation Problem.* Proc. Natl. Acad. Sci. USA, Vol. 77, No. 12, pp. 6988-6989.

Saccomanno, F.,Yu, M. and Shortreed, J. (1993). *Effect of Driver Fatigue on Truck Accident Rates.* Urban Transport and The Environment Mechanics Publication, Southampton, U.K.

Saccomanno, F., Kormendi S. and Lamble R. (1997a). *Establishing Trucking Firm Accident Risk Potential From Roadcheck Vehicle Inspections.* Tennessee Conference.

Saccomanno, F., Kormendi S., El-Herraoui M. and Lamble R. (1997b). *Identifying High Accident Risk Trucking Firms Using Roadcheck Vehicle Inspections.* Submitted to Can.

J. Civ. Eng. for review.

Satterthwaite, S.P. (1981). *A Survey of Research Into Relationships Between Traffic Accidents and Traffic Volumes.* Supplementary Report 692. Crowthorne, Berks, U.K.: Transport Research Laboratory.

Stewart R.J. and Campbell, B.J. (1972). *The Statistical Association Between Past and Future Accidents and Violations.* Chapel Hill, NC: Highway Safety Research center, University of North Carolina.

Tomberlin, T.J. (1988). *Predicting Accident Frequencies for drivers classified by two factors.* J. Am. Stat. Assoc. 83, pp 309-321.

U.S. Department of Transportation. *Accident of Motor Carriers of Property*, Washington, DC: Federal Highway Administration, May 1986.

U.S. Department of Transportation. *BMCS Annual roadside Inspections 1983 and 1984*, Washington, DC: Federal Highway Administration, September 1985.

Zeeger, C. V., Hummer, J. Reinfurt, D., Herf, L. and Hunter, W. (1987). *Safety Effects of Cross-section Design for Two-lane Roads, Vols I and II.* Prepared for the Federal Highway Administration and Transportation Research Board by the University of North Carolina.

Zeeger, C. V., Stewart, R., Reinfurt, D., Herf, Council, F., Neumann, T., Hamilton E., Miller T. and Hunter, W. (1990). *Cost Effective geometric improvements for Safety Upgrading of Horizontal curves.* Prepared for the Federal Highway Administration by the University of North Carolina.

Zhou, M. and Sisiopiku, V. (1997). *On The Relationship Between Volume to Capacity Ratios and Accident Rates.* Transportation Research Board, 76th annual meeting,