The Relationship of Credit-Based Insurance Scores to Private Passenger Automobile Insurance Loss Propensity

An Actuarial Study
by
EPIC Actuaries, LLC

EPIC

Principal Authors:
Michael J. Miller, *FCAS, MAAA*
Richard A. Smith, *FCAS, MAAA*

Principal Peer Reviewer:
Klayton N. Southwood, *FCAS, MAAA*

June 2003

Foreword

EPIC Actuaries, LLC was retained to conduct an actuarial analysis of the relationship of credit-based insurance scores to the propensity of loss for private passenger automobile insurance. In addition to the correlation study, EPIC was requested to study the extent to which credit-based insurance scores may measure risk that is already being measured by other risk factors and to study the relative importance of credit-based insurance scores to accurate risk assessment.

The study and this report were sponsored by the Alliance of American Insurers, the American Insurance Association, the National Association of Independent Insurers and the National Association of Mutual Insurance Companies.

EPIC had the sole responsibility and the independence to prepare this report and to conduct the study in the way it considered to be actuarially sound. The opinions and conclusions expressed in this report are those of the individuals on EPIC's research team.

About EPIC Actuaries, LLC

EPIC is a privately-held Illinois limited liability corporation, founded by a number of principals and senior consultants previously employed at Miller, Herbers, Lehmann, & Associates. EPIC's professional staff serves clients including insurers, local and state government entities, insurance trade organizations, self-insured businesses and groups, captive insurers and risk retention groups. Many of EPIC's clients have been served continuously by its principals and senior consultants for over 15 years, with some in that group being served continuously since the commencement of practice in 1984.

The authors, Messrs. Miller and Smith, are principals of EPIC, Fellows of the Casualty Actuarial Society and members of the American Academy of Actuaries. Each has been actively involved in ratemaking for personal lines of insurance for over twenty-five years.

The authors are available to answer questions about this report by calling (715) 358-6878, or (309) 828-8351, or writing to EPIC Actuaries, P.O. Box 628, Minocqua, WI 54548.

Table of Contents

Executive Summary

Purpose of the Study

The use of credit information in the risk assessment process for private passenger automobile insurance (as well as use for the various forms of homeowners, mobile homeowners and fire/dwelling coverages issued on private residences) is a relatively new practice. Many questions have been raised regarding the use of credit-related risk factors in the pricing and underwriting of personal lines insurance.

This study addresses the following three questions as they relate to private passenger automobile insurance.

i. Correlation Question: Are credit-based insurance scores related to the propensity of loss?

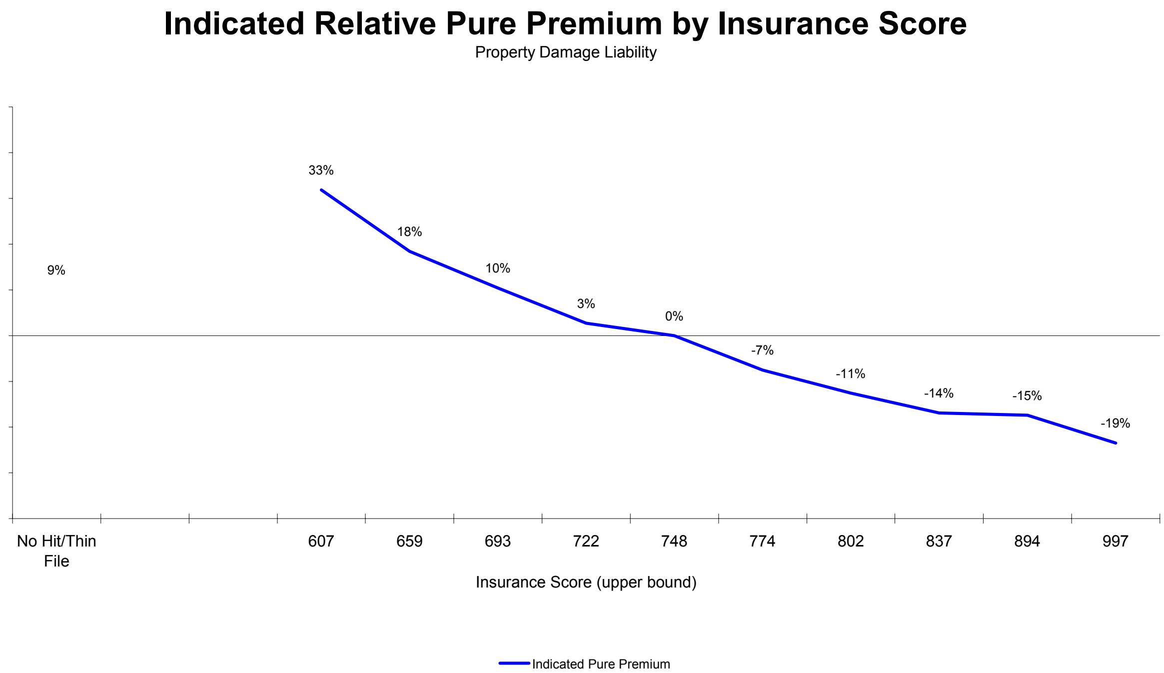
ii. Overlap Question: Do credit-based insurance scores measure risk that is already being measured by other risk factors?

iii. Business-Purpose Question: What is the relative importance to accurate risk assessment of using credit-based insurance scores?

The study and this report were sponsored by the Alliance of American Insurers, the American Insurance Association, the National Association of Independent Insurers and the National Association of Mutual Insurance Companies.

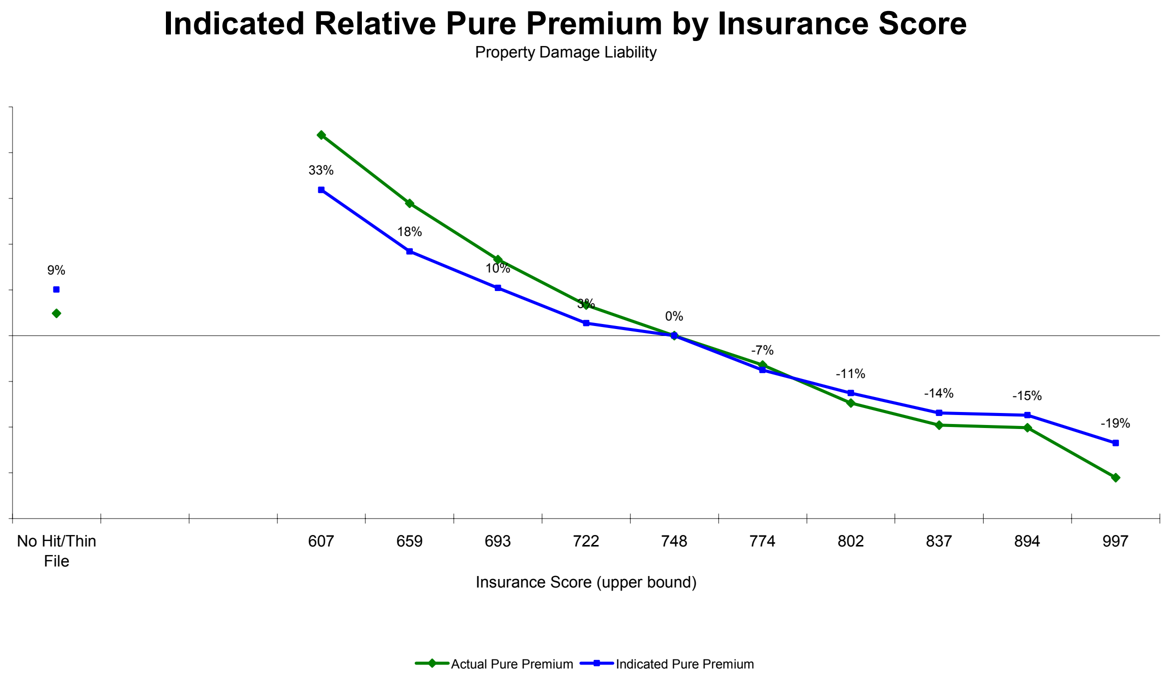
Findings

Finding #1: Using multivariate analysis techniques to adjust the data for interrelationships between risk factors, insurance scores were found to be correlated with the propensity for loss. This correlation is primarily due to a correlation between insurance scores and claim frequency, rather than a correlation between insurance scores and average claim severities.



The relative pure premiums (i.e., propensity of loss) are significantly different from one insurance group to the next and show a clear pattern of decreasing loss propensity as the insurance score increases. For the property damage (PD Liability) coverage shown in the above graph, the lowest range of insurance scores produce indicated pure premiums 33% above average and the highest range of insurance scores produce indicated pure premiums 19% below average. All six of the automobile insurance coverages studied exhibit the same general pattern.

Finding #2: Insurance scores do overlap to some degree with other risk characteristics, but after fully accounting for all interrelationships, insurance scores significantly increase the accuracy of the risk assessment process.



The "Actual" relative pure premiums in the above graph represent the observed values before accounting for any interrelationships between insurance score and the other risk factors. The "Indicated" relative pure premiums represent the differences in risk levels between the various insurance scores, after accounting for all interrelationships between insurance score and the other risk factors. The difference between the "Actual" line and the "Indicated" line is a visual presentation of the extent of overlap. The "Indicated" line is not sloping downward to the right quite as steeply as does the "Actual" line, but the correlation between insurance score after adjustment for overlap is, nevertheless, highly significant.

The steepness of the "Indicated" line in the above graph is a visual presentation of the degree of importance of insurance score in explaining risk that is otherwise not being explained by any other risk factor.

> **Finding #3: Insurance scores are among the three most important risk factors for each of the six automobile coverages studied.**

Coverage	Factor 1	Factor 2	Factor 3
BI Liability	Age/Gender	Ins. Score	Geography
PD Liability	Age/Gender	Ins. Score	Geography
Pers. Inj. Prot.	Ins. Score	Geography	Yrs. Insured
Med Pay	Ins. Score	Limit	Age/Gender
Comprehensive	Model Year	Age/Gender	Ins. Score
Collision	Model Year	Age/Gender	Ins. Score

Finding #4: An analysis of property damage (PD Liability) claim frequencies by insurance score groups for each of the fifty states indicates that the study results apply generally to all states and regions.

Graphs for each state, as provided in Appendix Q of the report, exhibit strikingly similar patterns of decreasing claim frequencies with increasing insurance scores to the pattern observed in the countrywide data.

About the Study

The study was based on a countrywide, random sample of private passenger automobile policy and claim records. Records from all fifty states were included in approximately the same proportion as each state's registered motor vehicles bear to the total registered vehicles in the United States. After elimination of the incomplete records, there were records for analysis equivalent to nearly 2.7 million earned car years (i.e., the equivalent of one car insured for twelve months).

The random sample of records was drawn from all policies in effect at any time during the twelve-month period ending June 30, 2001. This included policies that were in effect on July 1, 2000 and continued in effect during at least a part of the following year, as well as new policies first written in the year ending June 30, 2001. Premiums included in the study were those earned during the year ending June 30, 2001.

The claim record for each policy included accidents which occurred in the accident-year ending June 30, 2001. Claim counts, paid claim amounts and reserves on known outstanding claims were reported as they had developed as of June 30, 2002.

Data extracted for the study included information about the policy, about each vehicle insured on the policy, about each driver insured on the policy and about each claim on the policy.

The insurance score used in the study was the ChoicePoint Attract ™ score provided by ChoicePoint Services, Inc. ChoicePoint is a commercial vendor of proprietary insurance scores to insurers throughout the United States.

ChoicePoint was unable to either match the policy record with a credit record (i.e., no-hits) or had insufficient credit information to develop an insurance score (i.e., thin-files) for approximately 10% of the database. The no-hits and thin-files were included in the study as a separate category of risks.

Data were not shared among the providers of the sample database, nor did EPIC provide the database to the sponsors of the study.

The study was primarily based on relative pure premiums by insurance score groupings, rather than reliance on relative loss ratios. Relative pure premiums allow users of the study to more readily generalize the results without the need to consider the specific rating plan and specific rate factors being used by any specific insurer. The study was performed separately on each of six automobile insurance coverages.

A multivariate analysis technique (i.e., generalized linear modeling) was used to determine indicated risk factors. Multivariate analysis involves analyzing all risk factors simultaneously so as to adjust for any interaction between insurance scores and other risk factors. The software used for the analysis was Pretium®, which is owned by Watson Wyatt Pretium Limited. The statistical models used in the fitting of the curves to the raw data were the Poisson distribution for claim frequencies and the Gamma distribution for the average claim costs.

Introduction

This report begins with definitions and discussion of basic terms and concepts which pertain to ratemaking and risk assessment. While these basic concepts may be a review for many readers, this foundational material may help some non-actuaries understand the actuarial principles which apply to risk assessment in general and to the use of credit-based insurance scores in particular.

The report concludes with a description of the study methodology and the major findings and conclusions. To minimize the volume of the report's narrative, many important graphs and exhibits have been relegated to a separately bound Appendix. The Appendix is an integral part of this report and is necessary for a full understanding of EPIC's analyses and findings.

Definition of Important Terms and Concepts

Private Passenger Automobiles

This study has been limited to private passenger automobile insurance losses. Private passenger automobiles, as the term is used in this report, means the same as the term typically means in the insurance industry. In other words, four-wheeled passenger type vehicles (including sport utility, sport van and station wagon types) that are used for personal pleasure and family purposes. Utility-type vehicles (i.e., pickup, panel truck, or utility van) are also included in the definition if used for personal pleasure, family, or business purposes and the gross vehicle weight is not more than 10,000 pounds or the load capacity is not more than one ton.

Claim Frequency

Claim frequency is the ratio of the number of insurance claims to the number of autos insured. For example, a claim frequency of .150 means there are 150 claims for every 1,000 autos insured. A claim frequency of .150 can also be interpreted as a 15% chance or likelihood that a particular insured will incur a claim.

Average Cost Per Claim

The average cost of a claim is calculated as the total dollars of claim losses incurred divided by the total number of claims. This value is often referred to as Claim Severity.

Pure Premium

The pure premium is the average cost of claims per insured auto. It is calculated as the total dollars of incurred claim losses divided by the total number of autos insured. As shown in the following algebraic formula, the pure premium is the product of the claim frequency times the average cost per claim.

Let:

N = number of insured autos

C = number of claims

D = dollars of claim losses

C/N = claim frequency

D/C = average cost per claim (i.e., severity)

D/N = pure premium

Then:

(C/N) x (D/C) = D/N = Pure Premium

Since the pure premium is a combination of the probability of a claim occurring (i.e., claim frequency) and the average cost of the claim once it occurs (i.e., claim severity), it is considered as the best measure of risk for a group of insureds or for an individual insured. An insured with an expected pure premium of $450 would be considered a "higher risk" than an insured with an expected pure premium of $300.

Relative Pure Premium

When a pure premium is expressed as a ratio to a selected base, it is referred to as a relative pure premium. If from the previous example the $450 pure premium were expressed as a ratio to the $300 pure premium, the relative pure premium would be 1.50. In the example, we would say that the insured with the $450 pure premium possessed a 50% higher total risk than the insured with the $300 pure premium.

Loss Ratio

Often, claim losses are expressed as a ratio to premiums. The total dollars of claim losses divided by the total dollars of premiums is a loss ratio. For example, a loss ratio of .70 (i.e., 70%) means that 70% of the premium dollars went toward the payment of claim losses.

Relative Loss Ratio

A loss ratio expressed as a ratio to another loss ratio, is referred to as a relative loss ratio. For example, if one group of insureds has a loss ratio of 70% and a second group has a loss ratio of 50%, the first group is said to have a relative loss ratio of 1.40 compared to the second group.

Loss Ratio versus Pure Premium

Both loss ratios and pure premiums play important roles in actuarial ratemaking calculations. Loss ratios and pure premiums are algebraically related.

Let:

P =	total premium dollars
N =	number of insured autos
D =	total dollar of claim losses
D/N =	pure premium or average loss per insured
P/N =	average premium

Then:

i. Loss Ratio = D/P, or

ii. Loss Ratio = (D/N) / (P/N)

The second algebraic expression simply says that a loss ratio may be calculated as the pure premium divided by the average premium per insured.

Loss ratios are commonly used by actuaries to determine the <u>needed adjustment</u> to current rates and/or current rate factors. If the loss ratios are identical between two risk groups, that means the current rate factors are in relative proportion to the losses. If the loss ratios are not identical between two risk groups, that means the current rate factors needed to be adjusted to bring the rates into proportion with the losses.

If loss ratios are used in an analysis of risk, then the loss ratio must be coupled with additional knowledge about the rates and rate factors. Loss ratios <u>alone</u> reveal nothing about the actual value of the underlying rates or rate factors, nor do they <u>alone</u> reveal anything about the level of risk. An insured with a high degree of risk (assume a $450 pure premium from the previous example) may have a low loss ratio of 50%, if that insured is being charged a premium of $900. On the other hand, an insured with a low degree of risk (assume a $300 pure premium) may have a very high loss ratio of 100%, if that insured is being charged a premium of $300. Insureds with the highest propensity for loss may have the lowest loss ratios and vice versa.

Expressing loss ratios as relative loss ratios does not correct for the loss ratio shortcomings cited above. In the example above, the insured with the higher degree of risk produced a 50% loss ratio and the insured with the lower degree of risk produced a 100% loss ratio. The relative loss ratios are .50 to 1.00, with the higher-risk insured possessing the lower relative loss ratio.

Pure premiums are used to directly measure the level of risk. Pure premiums directly reveal the level of risk, independent from the rate that is being charged to the insured. High-risk insureds have high expected pure premiums. Low-risk insureds have low expected pure premiums.

Risk Factors

There are several risk characteristics, or risk factors, that have been found to measure and predict at least a portion of the total risk associated with each insured. For private passenger auto insurance, where the car is garaged and principally operated has been found to affect the average cost of claims (i.e., severity) and also affect the frequency of claims. Other risk factors found to be related to the risk associated with each insured include driver characteristics (i.e., age, gender, or marital status), driving record, how the auto is used (i.e., pleasure, commuting, or business), and the make and model of the auto. There are many other risk factors not listed here.

No single risk factor has been found that measures or predicts the total risk. All risk factors work in combination to measure and predict risk. One of the questions being addressed in this report is whether a new risk factor using credit-based insurance scores can add significant accuracy to the risk assessment process without overlapping, or duplicating, the risk factors already in use.

As previously stated, the level of risk is measured directly by the pure premium because the pure premium accounts for both the likelihood of a claim occurring and the cost of the claim once it does occur. Throughout this report, references to either the level of risk or to relative risk are based on pure premiums or relative pure premiums.

Rate Factors

Rates for all property/casualty insurance coverages reflect four broad categories of costs:

i. the anticipated claim losses,
ii. the anticipated expenses associated with settling the claims,
iii. the anticipated operational/administrative expenses, and
iv. the cost of capital necessary to support the insurance process.

Rates vary between insureds because the combination of the above four cost categories varies between insureds.

The first of the four cost categories (Item i above), the anticipated claim losses, is measured by the pure premium. Throughout this report references to “risk factors” are based on an analysis of pure premiums. Since the pure premium is only one of the four cost categories that make up a rate, it follows that “rate factors” (which reflect a combination of all four cost categories) are often different than “risk factors,” which reflect only the claim loss portion of the rate.

This study was not intended to determine indicated rates, rate factors, or relative rate factors between any insureds or groups of insureds, including groupings by credit-based insurance scores. All references to risk, or relative risk, are based on the pure premium portion of the rate, and exclude consideration of the various expenses and cost of capital components of the rate.

Risk Classification Plans

The actual claims history of an individual insured is unreliable as the sole basis for determining that individual’s propensity for loss. It would be a simpler world if we could solely look to an individual’s past driving history to reliably predict future losses. Unfortunately our world is not that simple. An insured’s propensity for loss is the result of a complex combination of several risk factors. A risk classification plan is a schedule of all the applicable risk factors.

The purpose of any risk classification plan is to group insureds with substantially similar risk characteristics so that claims data for each risk grouping can be accumulated and analyzed to determine an accurate value for each risk factor. Once the value of each risk factor is determined the insurance rate for an individual can be accurately calculated using the risk factors that apply to a specific individual. If credit-based insurance scores are used in rating, then these factors become part of the risk classification plan so that claims data can be gathered to accurately determine the contribution that insurance scores make to the overall risk assessment process.

Univariate versus Multivariate Analysis

Univariate analysis refers to analyzing insurance claim loss data, one risk factor at a time. Multivariate analysis refers to simultaneously analyzing the data for two or more risk factors. Perhaps the easiest way to understand the concept is to consider a hypothetical example.

Assume an insurer subdivided its total claims data by gender of driver as shown in Table 1. This univariate analysis indicates that males are of higher risk than females by approximately 15% (i.e., $196/$170).

Table 1: Univariate by Gender

Gender	# of Insureds	Pure Premium	Relative Pure Premium
Male	100	$196	1.15
Female	100	$170	1.00
Total	200	$183	

Assume the insurer then subdivides the same total claims data by mileage as shown in Table 2. This univariate analysis indicates that long-mileage drivers are of higher risk than short-mileage drivers by approximately 16% (i.e., $192/$166).

Table 2: Univariate by Mileage

Mileage	# of Insureds	Pure Premium	Relative Pure Premium
Long	130	$192	1.16
Short	70	$166	1.00
Total	200	$183	

The univariate analyses in Tables 1 and 2 do not tell us whether the apparent difference in risk due to gender is real or whether we are merely observing the results of females who may tend to drive fewer miles than males. Similarly the two univariate analyses do not tell us whether the apparent difference in risk due to mileage is real or whether we are merely observing the result of having high-risk males dominate the long-mileage category and the low-risk females dominate the short-mileage category. It is possible that gender is a real risk factor and mileage is not. It is possible that mileage is a real risk factor and gender is not. It is also possible that both are valid risk factors.

The univariate analysis technique does not allow an analyst to observe any overlap, or interrelationship, between the risk factors. Univariate analysis does not reveal whether any interaction exists between two or more risk factors. Univariate analysis can produce reasonable results if the analyst knows from other research that there is no significant overlap between the risk factors being analyzed.

A multivariate analysis would allow us to analyze both gender and mileage simultaneously to determine if one or both factors are related to risk and to eliminate the possibility that the two factors are interrelated in a way that is distorting the data.

Table 3: Multivariate by Gender and Mileage

Pure Premium (# Insureds)

Mileage	Male	Female	Total
Long	$200 (80)	$180 (50)	$192 (130)
Short	$180 (20)	$160 (50)	$166 (70)
Total	$196 (100)	$170 (100)	$183 (200)

The multivariate analysis in Table 3 allows us to determine that both mileage and gender are risk-related. There is a measurable difference in loss propensity between the genders within both of the mileage categories. Males driving long-mileage are approximately 11% ($200/$180) higher risk than females driving long-mileage. Males driving short-mileage are approximately 13% ($180/$160) higher risk than short-mileage female drivers. Obviously, miles driven does not fully explain the apparent differences in gender observed in Table 1.

The data in Table 3 also show there is a discernible difference in loss propensity by mileage within each gender category. This means that gender does not entirely explain the observed pure premium differences between the mileage categories observed in Table 2.

While we have considered a simple example with only two risk factors, there are many more complex relationships that occur when an insurer uses dozens of rating and underwriting variables to assess risk. If univariate analysis techniques are used, inaccuracies in risk assessment could occur for each risk factor, and these inaccuracies could be compounded and lead to larger inaccuracies in the measurement of the total risk. If a multivariate analysis technique is used, interrelationships between the risk factors are taken into account and more accurate risk relationships can be determined.

Credit Scores versus Insurance Scores

Credit reports contain a wide variety of credit information concerning an individual consumer. In addition to information that identifies a specific individual, the report contains data on credit card and loan balances, types of credit, status of each account, judgments, liens, collections, bankruptcies, and requests for credit information. Each data element in the credit report is commonly referred to as an "attribute".

Credit-score modelers combine and weight selected attributes to develop a single "credit score". These "credit scores" have long been used by lending institutions to predict the risk associated with the repayment of a loan or satisfaction of some other financial responsibility.

Insurance-score modelers have begun to combine and weight selected credit attributes to develop a single "insurance score". These "insurance scores" are being added as a risk factor to create risk classification plans with the intent to more accurately assess the risk associated with the propensity for an insurance loss.

Even though both a "credit score" and an "insurance score" are derived from an individual's credit report, the two scores are different. There is no reason to believe that a credit score measuring the likelihood of loan repayment will be based on the same credit attributes (or that each attribute will be assigned the same weight) as are used to derive an insurance score, and vice versa. Unfortunately, some in the insurance business have come to refer to credit-based insurance scores as "credit scores". This misuse of the language may have led some to conclude that the advantages and disadvantages of using credit scores in the lending industry have direct application to the insurance industry. It may have also led some to attempt to apply the results of "credit score" studies by lending institutions to the use of "insurance scores" by insurers.

In this paper we refer to "insurance scores" or "credit-based insurance scores" to avoid confusion with the term "credit score" commonly used by lending institutions.

While there is a difference between "credit scores" and "insurance scores," there may also be an important difference between insurance scores used for rating one type of insurance versus another type. For instance, the credit attributes and the weighting of those attributes to develop an insurance score for private passenger automobile insurance may be different than a score used for commercial automobile insurance, or for homeowners insurance. In this study, all references to insurance scores are meant to apply specifically to insurance scores for private passenger automobile insurance.

Actuarial Principles and Fairness

The Casualty Actuarial Society (CAS) has adopted a "Statement of Principles Regarding Property and Casualty Insurance Ratemaking" which has direct application to the use of insurance scores in the risk assessment of personal lines insureds (see Appendix A). The Statement of Principles says that equity among insureds is maintained if the ratemaking process provides for all the costs associated with the risk transfer. It goes on to say that rates are reasonable and neither excessive, inadequate, nor unfairly discriminatory if all costs are provided for in the rate.

The Actuarial Standards Board (ASB) has adopted Actuarial Standard of Practice No. 12, entitled "Concerning Risk Classification", which has direct application to the use of insurance scores in the risk assessment of personal lines insureds (see Appendix B). ASOP No. 12 states as its first basic principle that a sound risk classification system "should reflect cost and experience differences on the basis of relevant risk characteristics." It goes on to say that it is equitable when "material differences in costs for risk characteristics are appropriately reflected in the rate" and that "a relationship between a risk characteristic and cost is demonstrated if it can be shown that experience is different when the characteristic is present."

The foundations of actuarial science clearly establish the following:

i. the use of a risk factor is fair and equitable if the risk factor reflects the differences in the expected value of anticipated insurance costs, and

ii. in a voluntary and competitive market system where the insurance buyer has the freedom to choose among several insurers, accurate risk assessment is vital "to ensure the equity and financial soundness of the system".

The actuarial definition of equity and fairness based on the underlying insurance costs and the degree of risk is not a definition first invented by actuaries. It is a definition that has been adopted by actuaries because of its long history of acceptance in insurance ratemaking and rate regulation. Rather than applying subjective judgments as to which insurance consumers should pay more and which should pay less, the rate can be objectively determined so that insureds representing the greatest insurable risk pay the highest premium, and vice versa.

In addition to the concepts of fairness and equity, ASOP No. 12 also establishes the principle that rates based on underlying costs and the degree of risk "permit economic incentives to operate, and thereby encourage widespread availability of coverage" in the marketplace. When insurers are not permitted to charge a rate commensurate with the known propensity for loss, there is a strong economic incentive to reduce the availability of coverage for those underpriced risks. This is not an economic phenomenon which applies only to insurance. Rational producers of goods or services are not usually willing to make products available to the market unless the price is adequate.

Some mistakenly argue that insurance is really a system of one insured subsidizing another and that the insurer can overcharge some insureds in order to undercharge others. But that doesn't work in a market where there are multiple providers and the buyers are free to move from one insurer to another, as demonstrated by the following example.

There are two insurers. Insurer A charges one rate that is an average for high-risk insureds and low-risk insureds. Insurer B charges a separate rate for each risk group. All else being equal, low-risk insureds will leave Insurer A and gravitate to Insurer B which has the lowest rate available for low-risk insureds. Insurer A will be left with only high-risk insureds and Insurer A's rate will increase to reflect the higher degree of risk of its insureds. The only way that Insurer A can continue to serve the full spectrum of insureds in a competitive market is to accurately price each insured.

The economic forces of any competitive market drive prices toward the underlying costs and the degree of risk represented by each insured, consistent with the commonly held definitions of fair and equitable pricing.

Causation, Correlation, and Predictive Value

Actuarial Standard of Practice No. 12 establishes that a risk factor is appropriate for use if there is a demonstrated relationship between the risk factor and the insurance losses. ASOP No. 12 states that the relationship may be demonstrated by the statistical analysis of data, but that the relationship need not be a cause-and-effect relationship.

Causation

A risk factor need not be the "cause" of the insurance losses. While understanding the cause of the losses is of interest in attempting to reduce losses, non-causal factors may be powerful predictors of insurance losses. Indeed, most risk factors are not the direct cause of a loss.

The classical example of a relationship that is not a cause-and-effect relationship is a home built in a river valley. Living in a river valley does not "cause" a flood. But there is a predictive relationship between the risk of a flood loss and the construction of a home in the flood plain. It would be foolish to presume there is no risk of a flood loss merely because the location of the home does not "cause" the flood.

Many other examples of non-causal relationships can be cited. Past traffic violations do not "cause" future insurance losses, but there is a predictive relationship between past driving records and future losses. Past fires in a home do not "cause" future fires, but past claim records are predictive of future losses.

Just as is the case of all other risk factors, causality should not be the basis for allowing or disallowing the use of credit-based insurance scores. The basis for allowing the use should be the ability of the insurance scores to measure the propensity for insurance losses.

It has long been a tenet of risk assessment that financial stability/responsibility was a risk predictor for private passenger automobile insurance. However, the concepts of financial stability and responsibility have been heretofore difficult to translate into objective, measurable risk factors. Credit-based insurance scores may be the means of objectively measuring financial responsibility.

While it would be inconsistent with sound actuarial principles to require credit-based insurance scores to demonstrate a causal relationship, we could reasonably speculate that there are psychological factors that likely affect how we manage our personal lives. We could reasonably speculate that the results of these psychological tendencies can be observed in many aspects of our personal lives, including our credit history and insurance losses. Insurance scores seem to provide an objective means of measuring personal responsibility and its effect on insurance losses, even though we may never fully understand the psychology involved.

Insurance scores are the output of scoring models. Scoring models provide a disciplined, objective, and consistent way to combine a multitude of credit attributes. The insurance scores produced by the scoring models can be tested against actual insurance losses and the correlation to loss propensity objectively determined.

Correlation

Correlation is an objective, statistical means of establishing the relationship between a risk factor and the propensity for an insurance loss. If credit-based insurance scores increase as insurance losses increase, the statistics will indicate a positive linear correlation. If credit-based insurance scores increase as insurance losses decrease, the statistics will indicate a negative linear correlation. Either a positive or negative correlation between the risk factor and insurance losses is helpful in measuring the propensity for loss.

Predictive Value

Unfortunately, determining the relationship between a risk factor and the propensity of loss is not as simple as just determining the linear correlation. It turns out that linear correlation is part of the statistical evidence, but it is neither a necessary nor a sufficient condition for establishing predictability. There are instances when a risk factor demonstrates little or no linear correlation, but is still a powerful predictor of insurance losses.

It is possible that a study of the relationship between credit-based insurance scores and insurance losses may identify situations where little or no linear correlation exists. In those possible situations, credit-based insurance scores may still possess strong predictive value.

A hypothetical example may be the best means of comparing the concepts of linear correlation and predictive value.

Example 1:

Driver Age	Pure Premium	Stand. Dev. of Pure Premium	Actual Observations
20	$850	$245	$500, $600, $700, $800, $900, $1000, $1100, $1200
30	$650	$245	$300, $400, $500, $600, $700, $800, $900, $1000
55	$450	$245	$100, $200, $300, $400, $500, $600, $700, $800

Example 1 shows a strong linear correlation between the driver age and the pure premiums for each class of risk. But the observations (i.e., actual losses) within each driver age class are highly dispersed so that there is very little predictive value in this risk classification plan. The data in Example 1 suggest so much overlap of pure premiums by driver age class that one cannot conclude that driver age is reliably predictive of insurance loses.

Example 2:

Driver Age	Pure Premium	Stand. Dev. of Pure Premium	Actual Observations
20	$103.50	$2.45	$100, $101, $102, $103, $104, $105, $106, $107
30	$128.50	$2.45	$125, $126, $127, $128, $129, $130, $131, $132
55	$78.50	$2.45	$75, $76, $77, $78, $79, $80, $81, $82

Example 2 is another hypothetical, this time constructed with little correlation between driver age and pure premiums. However, the predictive value of these risk factors is far stronger than the risk factors in Example 1 because there is less dispersion within each risk class.

As Example 2 shows, analysis of risk characteristics need not be restricted to linear relationships. Predictive value is a variance concept which refers to the variation of losses within each risk class. If there is little variance of loss within each risk class, then the risk factor has strong statistical link to expected insurance losses.

Study Methodology

State Representation

This study was conducted on a random sample of individual policy records from throughout the United States. Since each insurer's book of automobile business is distributed differently across the fifty states, it was necessary to assign a different sampling percentage to each data provider. Each provider drew its countrywide random sample of policies using its assigned sampling percentage. The result was a total database for this study which is distributed across the fifty states approximately the same as registered vehicles are distributed by state. In other words, each state is represented in this study in approximately the same proportion as its registered motor vehicles bear to the total registered vehicles in the United States (see Appendix C).

Type of Policies and Size of Sample

The samples of policy records were drawn from entire books of private passenger automobile insureds. This means that the sample includes random representation across all tiers of insureds, including those insured through each state's residual market insurance mechanism.

The total number of earned car years (i.e., equivalent of one car insured for 12 months) produced for study by the random sampling process approached 2.7 million. Less than one percent of the records were incomplete and unusable for study. After the elimination of the incomplete records, there were records equivalent to approximately 2,690,000 earned car years available for analysis.

Policy Data and Claims Records

The random sample was drawn from all policies that were in effect at any time during the year beginning July 1, 2000 and ending June 30, 2001. This included policies that were in effect on July 1, 2000 and continued in effect during at least a part of the following twelve months, as well as new policies written in the twelve-month period ending June 30, 2001. Premiums included in the study were those earned during the year ending June 30, 2001.

The claim record for each policy included accidents which occurred in the accident-year beginning July 1, 2000 and ending June 30, 2001. Claim counts, paid claim amounts and reserves on known outstanding claims were reported as they had developed as of June 30, 2002.

Insurance Coverages

Premiums and claims data for the following coverages were included in the database: bodily injury liability, property damage liability, medical payments, personal injury protection, comprehensive and collision.

Data Detail

Data extracted for the study included information about the policy, about each vehicle insured on the policy, about each driver insured on the policy and about each claim on the policy. Details of the data included in the study are set forth in Appendix D. At no time did EPIC share the sample data with any providers or any of the study sponsors.

Insurance Scores

After receiving and consolidating the policy records from each of the participating insurers, EPIC submitted the consolidated database to ChoicePoint, a commercial firm that sells proprietary insurance scores to automobile insurers throughout the United States.

ChoicePoint attached its insurance scores to the policy records. Before returning the database to EPIC, ChoicePoint eliminated from the database the policy number, the name of the insured, the street address of the insured, the social security number of the insured, and the vehicle identification number. Upon the return of the database from ChoicePoint, EPIC possessed individual policy and vehicle records with insurance scores attached, but no way to tie the insurance scores either to a specific person or household. This procedure was followed to ensure the confidentiality of each person's insurance score.

ChoicePoint markets several different insurance scores. EPIC conducted this study using the Attract ™ score because the records in the database were drawn from a broad spectrum of the market and because this score has been most commonly used by ChoicePoint customers.

Territory Rating

The treatment of geographic rating territories presented a special challenge for this study because different territory definitions exist in each state. If the study had been focused only on a state such as

Texas or North Carolina where insurers tend to use the same territory definitions, then we would have used the state's "benchmark" territory definitions. But with a countrywide study, there are no "countrywide benchmark" territory definitions to use.

Eliminating the territory or geographical risk factor from consideration was not an option for this study because of the importance of the geographic risk in rating private passenger auto insurance. No study of the potential overlap of credit-based insurance scores with other risk factors, or the study of relative importance of various risk factors, could be complete without consideration of the geographic risk factor.

To recognize territory rating in this countrywide study, EPIC first ranked all U.S. zip codes by population density and then grouped the zip codes into twenty population/density groupings so that the total population was distributed among twenty groups of equal size. This means that for purposes of this study, each of the fifty states has potentially twenty rating territories, the definitions of which are based entirely on population density.

No-Hits/Thin-Files

Of the policy records submitted, ChoicePoint was unable to match the record to a credit history on approximately 7% of the total database for the study. The "no-hit" files constituted 183,183 earned car years.

ChoicePoint did match a credit history on some records, but the credit information was insufficient for the development of an insurance score. These "thin-files" represented slightly over 3% of the total database for the study and constituted 90,932 earned car years.

EPIC included both the no-hit records and the thin-files in the study as separate categories of insureds.

Multivariate Analysis

The research into the questions of: relationship of insurance scores to loss propensity, potential overlap of insurance scores with other rate factors, and the relative importance of all risk factors requires the application of multivariate analysis techniques whereby all risk factors are analyzed simultaneously. Those portions of this study requiring multivariate analyses were conducted with the use of Pretium®.

Pretium® is software which is owned by, and has been authored by, Watson Wyatt Pretium Limited of London, U.K. EPIC has been properly licensed to utilize Pretium® for this study and has used the model with the full knowledge of Watson Wyatt.

There exist a number of multivariate analyses techniques. One is commonly referred to in the literature as generalized linear modeling (GLM). Pretium® is software built on the GLM concept.

GLM analysis techniques involve the fitting of curves (i.e., statistical distributions) to the raw data to determine the indicated risk factors. For this study, EPIC utilized the Poisson distribution when analyzing indicated claim frequencies and the Gamma distribution when analyzing indicated average claim costs. These components were then combined to develop the indicated pure premiums.

Analyses and Findings

Relationship to Loss Propensity

The study was conducted separately for each of the six major automobile insurance coverages. So as not to overburden the narrative part of this report, it was decided to include only the Property Damage Liability graphs and charts in the narrative. PD Liability was selected because it is unaffected by deductibles, by wide divergence in limits of coverage, and is commonly considered to be the best "barometer" for accident and/or claim frequencies of all the coverages.

We have included the study results for all six coverages in the Appendices and often throughout the narrative we refer the reader to the Appendices.

Exhibit I summarizes the following data elements for the PD Liability coverage:

a) Number of records in the study by twenty-one ranges of insurance scores, plus the no-hit and thin-file categories.

b) Average claim frequency by insurance score range.

c) Average claim cost by insurance score range.

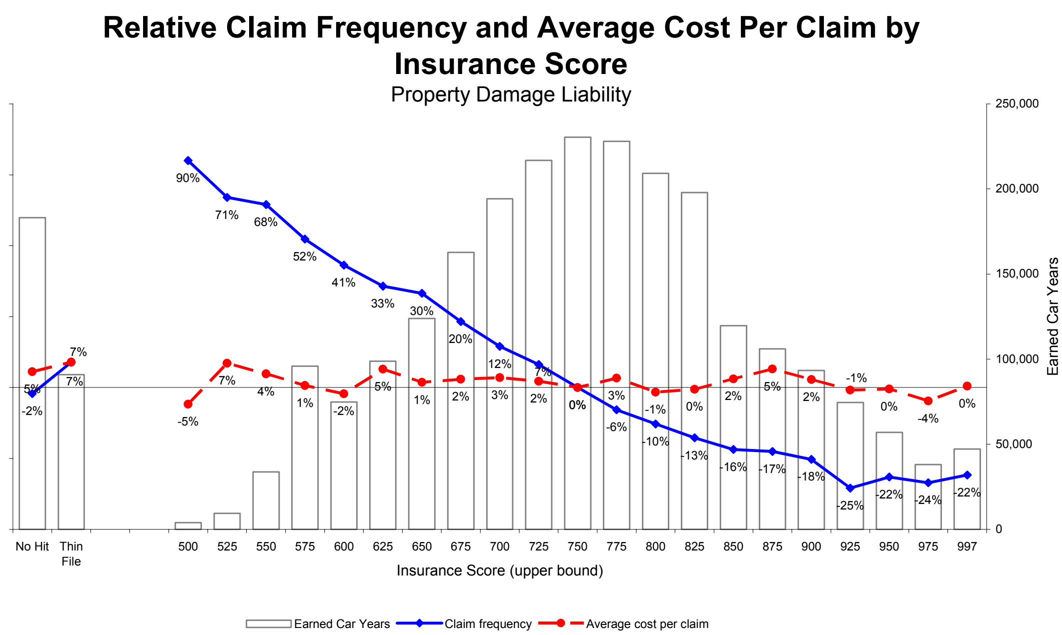
The claim frequencies and average claim severities in Exhibit I are raw data from the sample database and have not been adjusted for any potential biases arising from different risk demographics within each insurance score range or any overlap with other risk factors.

The data are shown for twenty-one ranges of insurance scores, plus the no-hits and thin-files. While our actuarial analyses were based on ten ranges of insurance scores, we chose to show data in a more refined breakdown so that the reader could appreciate the relatively few insureds at both extremes of the range of insurance scores. The distribution of automobile insureds follows the pattern of a normal distribution with the greatest concentration of insureds in the 650 to 825 range of insurance scores.

The highest claim frequencies are found in the lowest insurance score categories where there are relatively fewer insureds.

The pattern of variation of average claim severities across the insurance score categories differs by coverage. For PD Liability there is little variation of claim severities by insurance score. However, as shown in Appendix E, the Bodily Injury Liability coverage shows higher average claim severities as the insurance score increases up to a score of 900. As we will see later, this phenomenon is largely explained by other risk characteristics. Appendix E shows generally decreasing claim severities as the insurance score increases for the comprehensive and collision coverages.

Exhibit I:



Exhibits II, III and IV continue to present raw statistics, but with only ten ranges of insurance scores, plus one group with no score (i.e., no-hits and thin-files combined). Ten equal size groupings of insurance scores were chosen for this study because other similar studies have used ten groupings and because using the more refined, twenty-one groupings might raise data credibility issues with the groups that contain relatively few insureds. The choice to present the analysis based on ten equal-sized groupings, rather than the twenty-one score ranges shown in Exhibit I, has no impact on the general findings of this report.

Exhibit II shows that the pure premium (i.e., average dollars of loss per insured) tends to decrease as the insurance score increases. This phenomenon is observable for each automobile insurance coverage as shown in Appendix F. As evidenced by the data in Exhibit III and IV, the primary reason that the PD Liability pure premium tends to decrease as the insurance score increases is because the claim frequency tends to decrease as the insurance score increases. This phenomenon is observable by comparing Exhibits II, III and IV. The relative pure premiums in Exhibit II range from 48% above average to 24% below average. The range in relative pure premiums is nearly identical to the range of relative claim frequencies in Exhibit III. However, the range of relative average claim costs in Exhibit IV is much smaller with values between +3% to -1% of the average.

Exhibit II:

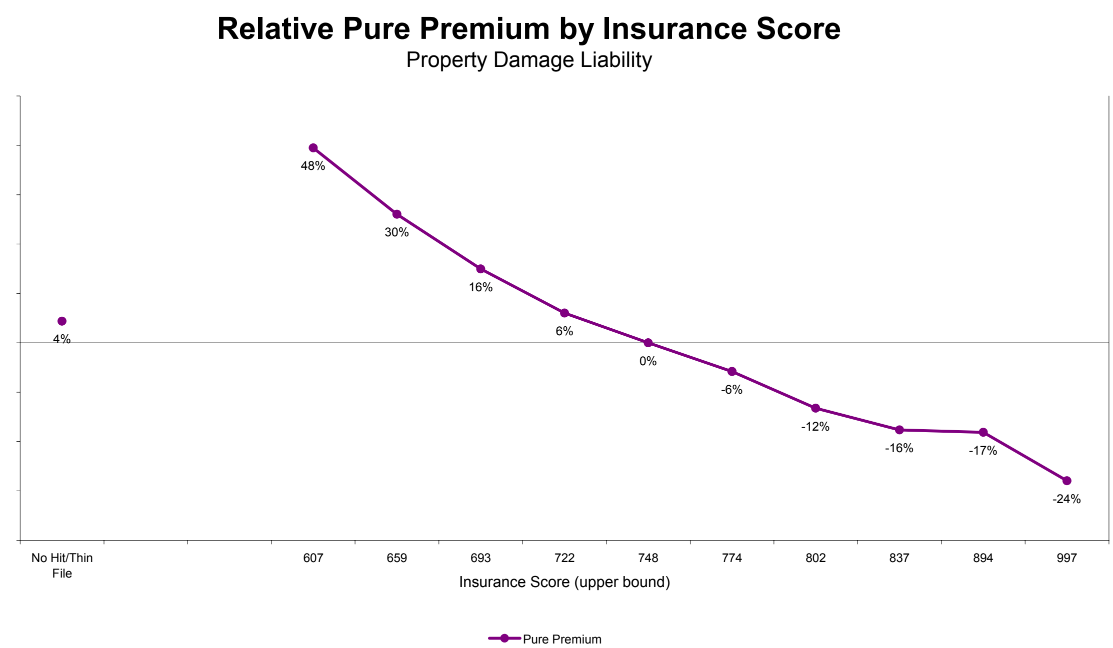


Exhibit III:

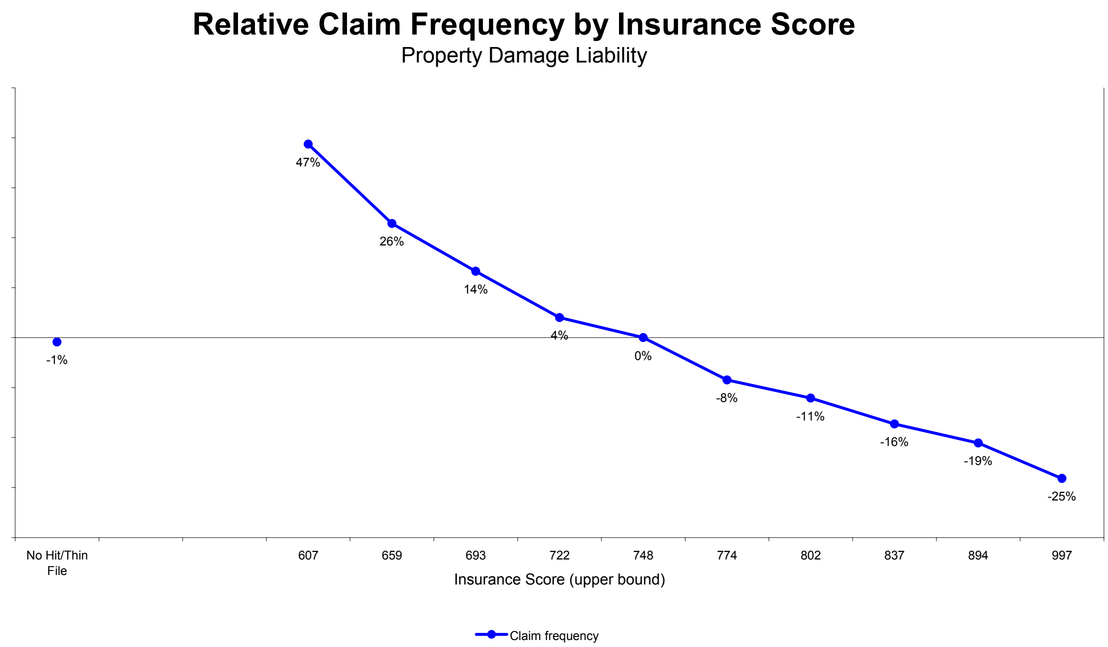
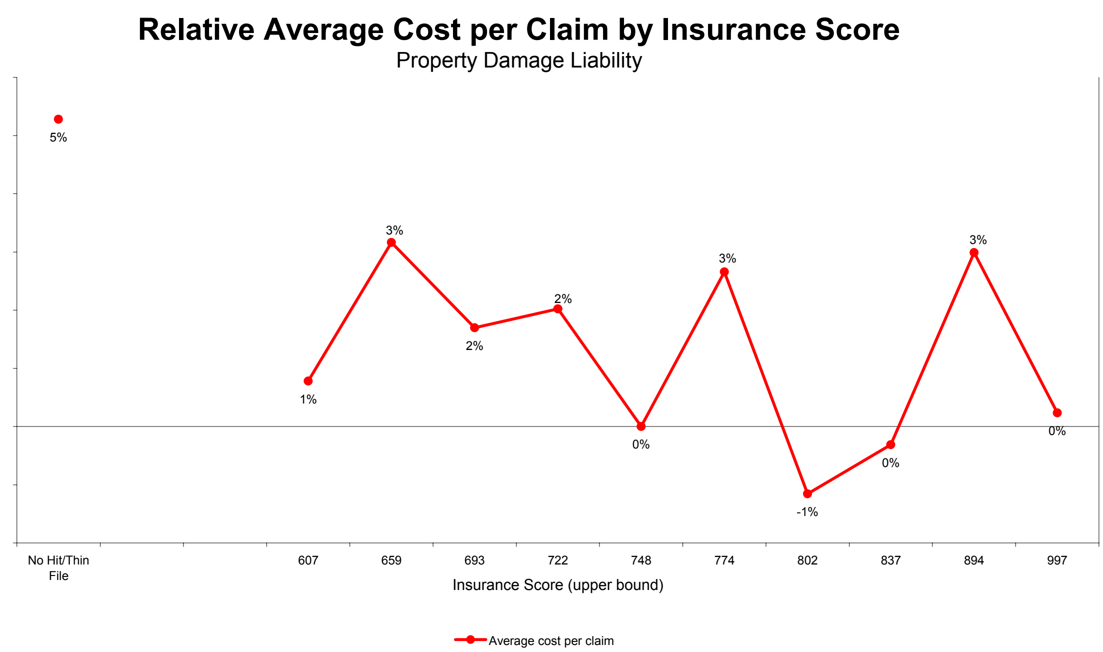


Exhibit IV:



In terms of actuarial analysis, the primary limitation with the data in Exhibits I through IV (also Appendices E, F, G, and H) is that the data are potentially distorted by distributions of insureds which differ by the various risk characteristics within each insurance score range. To understand this potential problem, consider Exhibit II. A cursory review of the graph suggests a strong correlation between insurance score and loss propensity (i.e., pure premium). However, it would be a mistake to draw such a conclusion without a more rigorous analysis of the data.

To properly interpret the data, each insurance score group should contain the same distribution of insureds by territory, age of driver, limit of coverage, model of car, and etc. In that way we can compare apples-to-apples and get a true picture of the differences in loss propensity across the insurance score groupings. We can approximate a "normalization" of the data through a multivariate analysis technique that considers all risk factors simultaneously.

Exhibit V presents the PD Liability pure premiums for each insurance score category after application of the multivariate analysis technique. The pure premiums in Exhibit V can be considered as the relative pure premiums from Exhibit II after "correction" for any distributional biases. More accurately, the pure premiums in Exhibit V are the indicated relative pure premiums for the various insurance score ranges after accounting for all the overlap, or interrelationships, with all other risk factors.

Table 4:

(1) Insurance Score	(2) Relative Pure Premium Univariate*	(3) Multivariate*
Less than 607	1.48	1.26
607 – 659	1.30	1.14
660 – 693	1.16	1.07
694 – 722	1.06	1.01
723 – 748	1.00	1.00
749 – 774	.94	.94
775 – 802	.88	.91
803 – 837	.84	.88
838 – 894	.83	.88
895 – 997	.76	.84
No-Hit/Thin-File	1.04	1.10

* Source for Column 2 is Exhibit II and for Column 3 is Exhibit V

A comparison of the relative pure premiums in Columns 2 and 3 above shows that after accounting for the interaction with all other risk factors, the propensity for loss decreases as insurance score increases. The univariate analysis in Column 2 correctly indicated the same pattern, but the relationship in Column 3 is "flatter" after adjustment for the interrelationship between all risk factors.

The PD Liability claims frequencies and average claim severities, after adjustment for the interrelationship with all other risk factors, are presented in Exhibits VI and VII. These exhibits show that claim frequencies are the primary reason that loss propensity varies by insurance score.

Pure premiums, claim frequencies, and average claim severities which have been adjusted for the interrelationship with all other risk factors are shown for the six major coverages in Appendices I, J, and K.

Exhibit V:

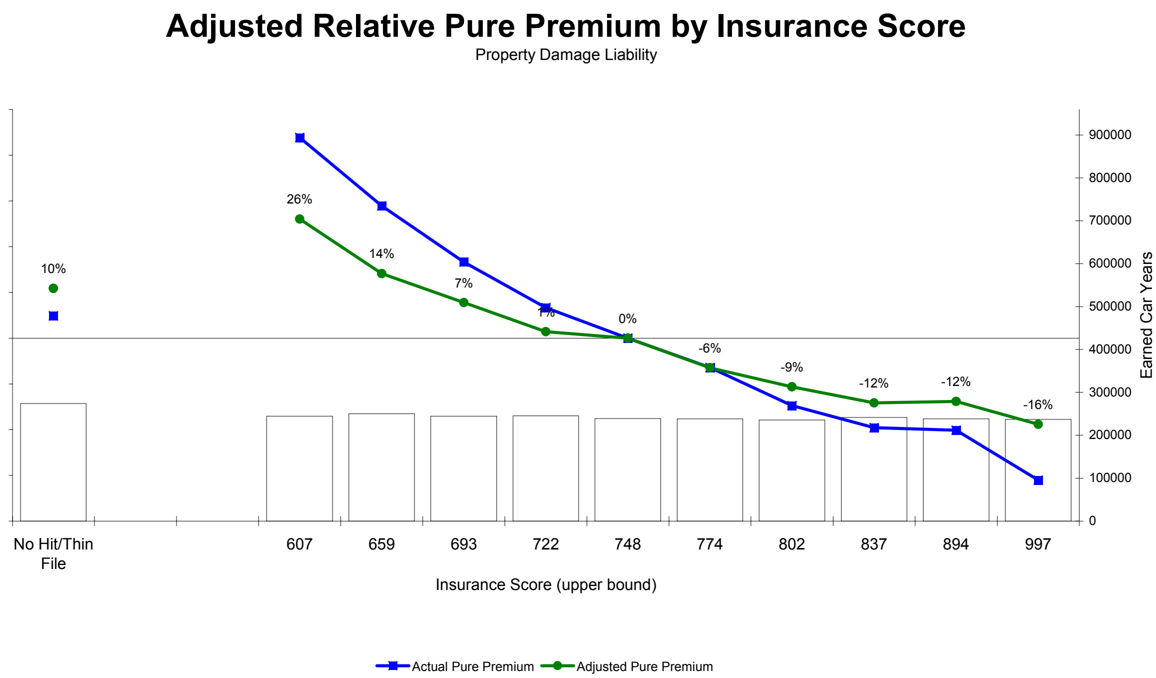


Exhibit VI:

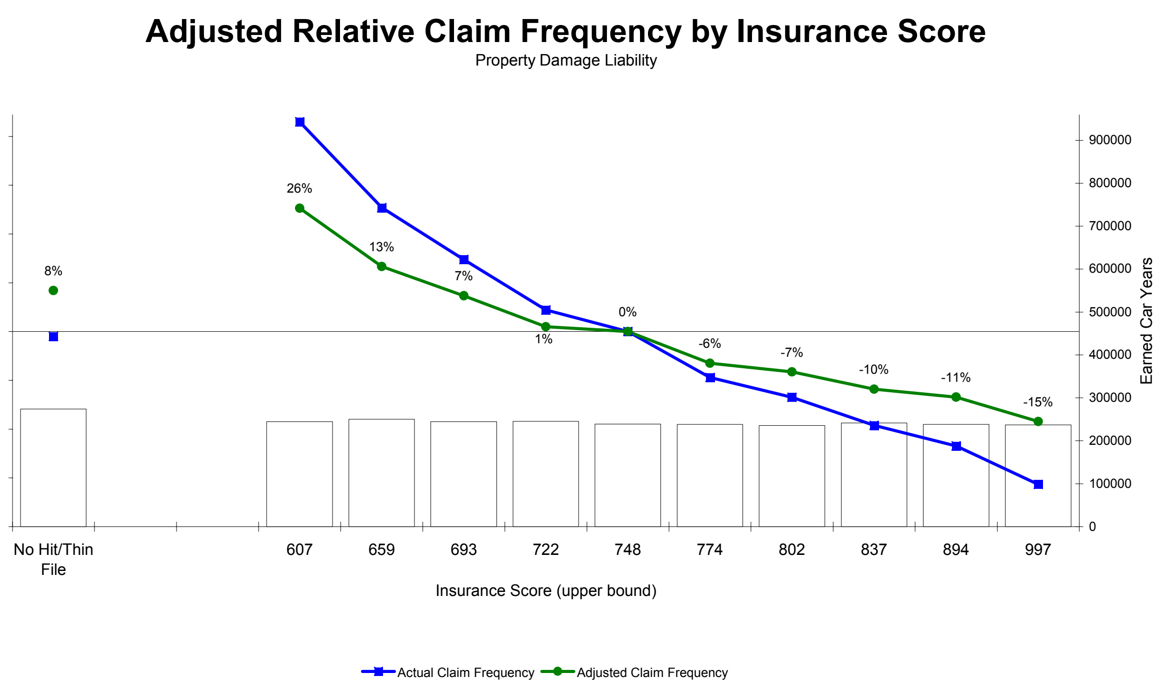
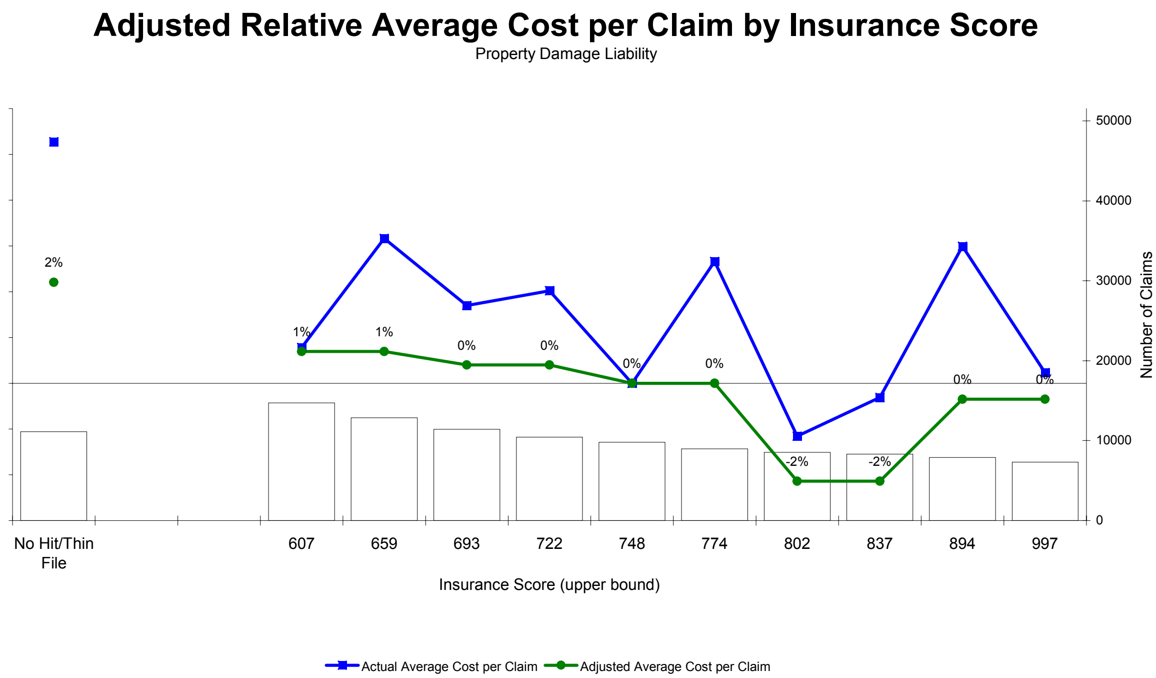


Exhibit VII:



Overlap/Interaction

We begin the discussion of overlap by presenting loss ratios calculated from the raw data. Several previously published studies on the subject have relied on relative loss ratios as a measure of loss propensity. For comparison to other studies, it is important to present the loss ratios and relative loss ratios derived from this database.

Exhibit VIII and IX present the PD Liability relative loss ratios, unadjusted for overlap with other risk factors, for both the twenty-one insurance score groupings and the ten insurance score groupings. Relative loss ratios for all coverages may be found in Appendices L and M.

Exhibit VIII:

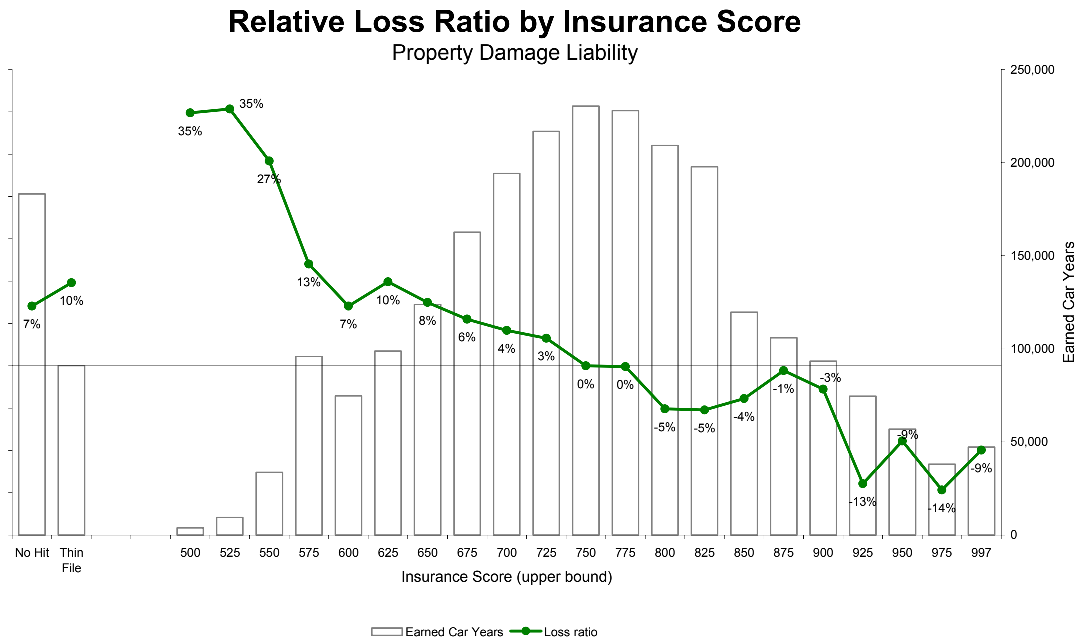


Exhibit IX:

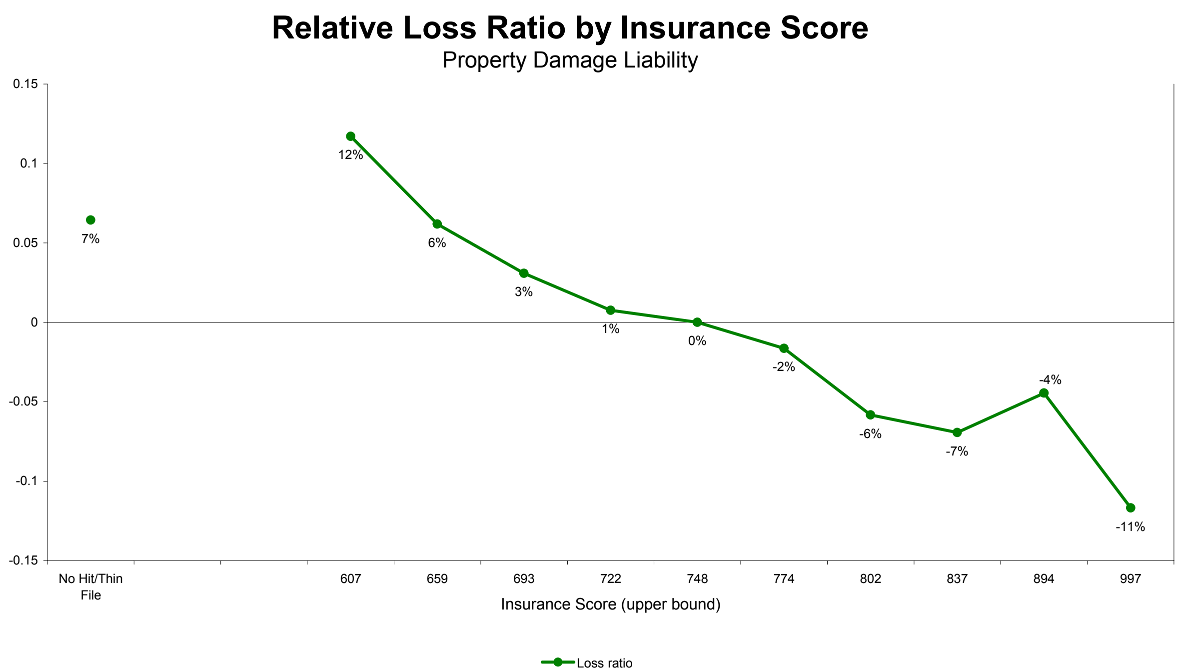


Table 5:

(1) Insurance Score	(2) Relative Loss Ratio*	(3) Relative Pure Premium*
Less than 607	1.12	1.48
607 – 659	1.06	1.30
660 – 693	1.03	1.16
694 – 722	1.01	1.06
723 – 748	1.00	1.00
749 – 774	.98	.94
775 – 802	.94	.88
803 – 837	.93	.84
838 – 894	.96	.83
895 – 997	.89	.76
No-Hit/Thin-File	1.07	1.04

* Source: Column (2) from Exhibit IX and Column 3 from Exhibit V

Both the relative loss ratios and the relative pure premiums in Columns 2 and 3 of Table 5 seem to show a significant difference in loss propensity by insurance score. But as the reader can see, the relativities between insurance scores are significantly different if based on pure premiums rather than on loss ratios. The differences in these two sets of relativities are not due to any differences in loss propensity. The losses underlying the calculation of both Columns 2 and 3 of Table 5 are identical. The differences in the relative values are entirely due to the effect of the premiums which are in the denominator of the loss ratios. Because the premiums (i.e., the rates and rate factors) can have such a significant impact on loss ratios, and thereby obfuscate the pure premiums in the numerator of the ratio, we have avoided analysis of loss ratios and concentrated directly on the pure premiums.

The reader will recall a previous discussion of relative loss ratios in which it was stated that the relative loss ratio provides the <u>indicated adjustment</u> which needs to be made to the underlying premium charges while the pure premium provides a direct measure of loss propensity.

Loss ratios and relative loss ratios are dependent upon the underlying premiums that are being charged. That means a study of loss propensity based on relative loss ratios cannot be easily generalized to all insurers because most insurers charge significantly different rates and use significantly different rating plans.

There may be exceptions to our concern with using relative loss ratios. If the study were confined to a state such as Texas or North Carolina, where insurers usually follow the state's "benchmark" rating plan, then a study of relative loss ratios could be safely generalized to other insurers in each of those states. However, this is a countrywide study and the many differences in rates and rating plans between insurers dictate the analysis be performed on relative pure premiums.

Exhibits X, XI, and XII present the PD Liability pure premiums, claim frequencies and average claim severities, after adjustment for overlap with all other rating factors. The line on the graph labeled "Actual" is based on raw data before any overlap adjustments and repeats the data from Exhibits II, III, and IV. The line on the graphs labeled "Adjusted" was calculated by using a multivariate analysis technique on all risk factors, except insurance score, and then leaving the remainder of risk to be explained by insurance score. These "Adjusted" data are a repeat of the data from Exhibits V, VI, and VII.

The line on the graphs labeled "Indicated" is calculated using the multivariate analysis technique on all risk factors simultaneously, including insurance score. The "Indicated" relative pure premiums, relative claim frequencies and relative average claim severities are the best statistical indication of the relationship between insurance scores and loss propensity.

"Indicated" pure premiums, claim frequencies and claim severities for all coverages are presented in Appendices N, O, and P.

Exhibit X:

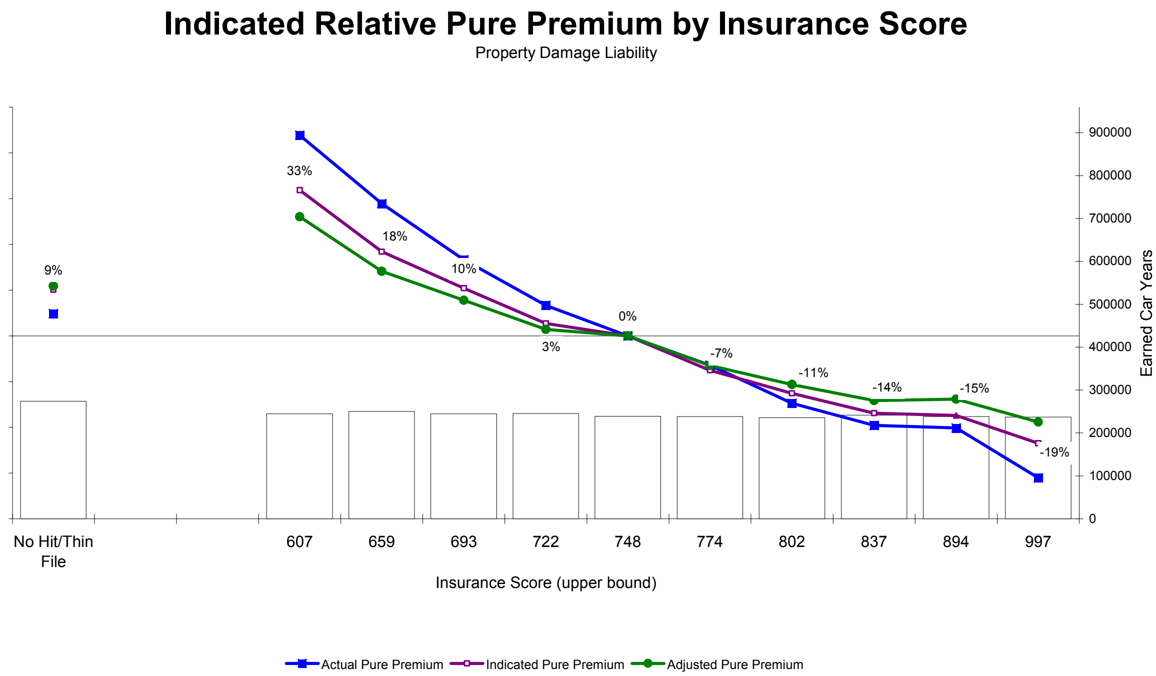


Exhibit XI:

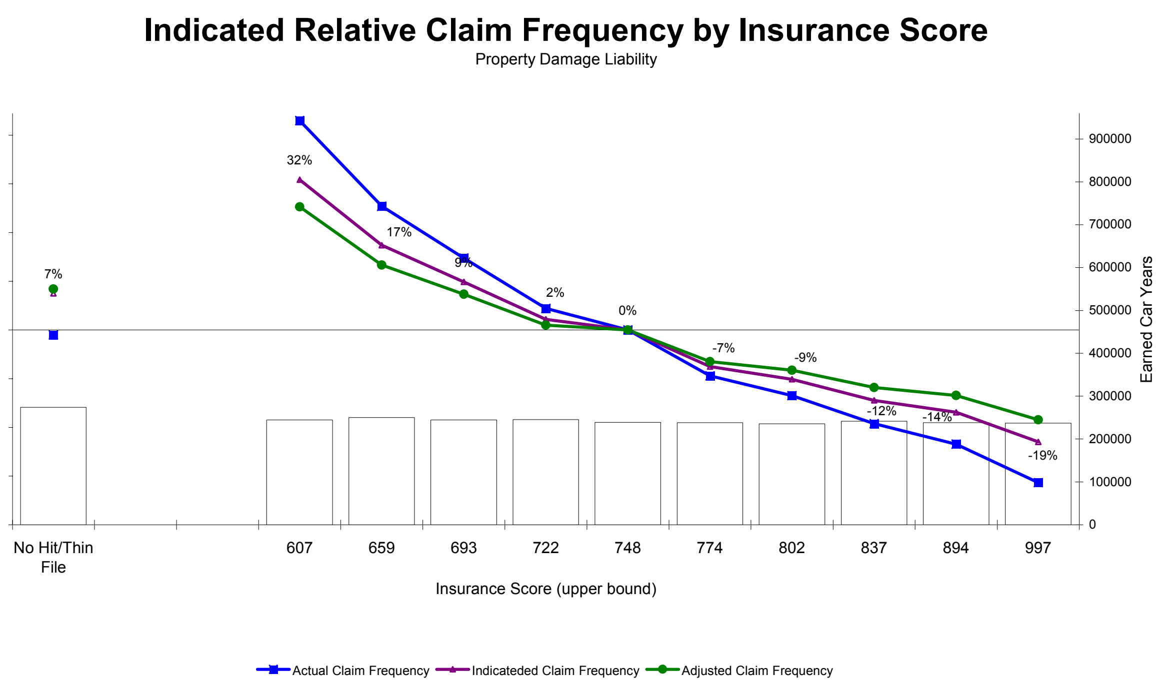
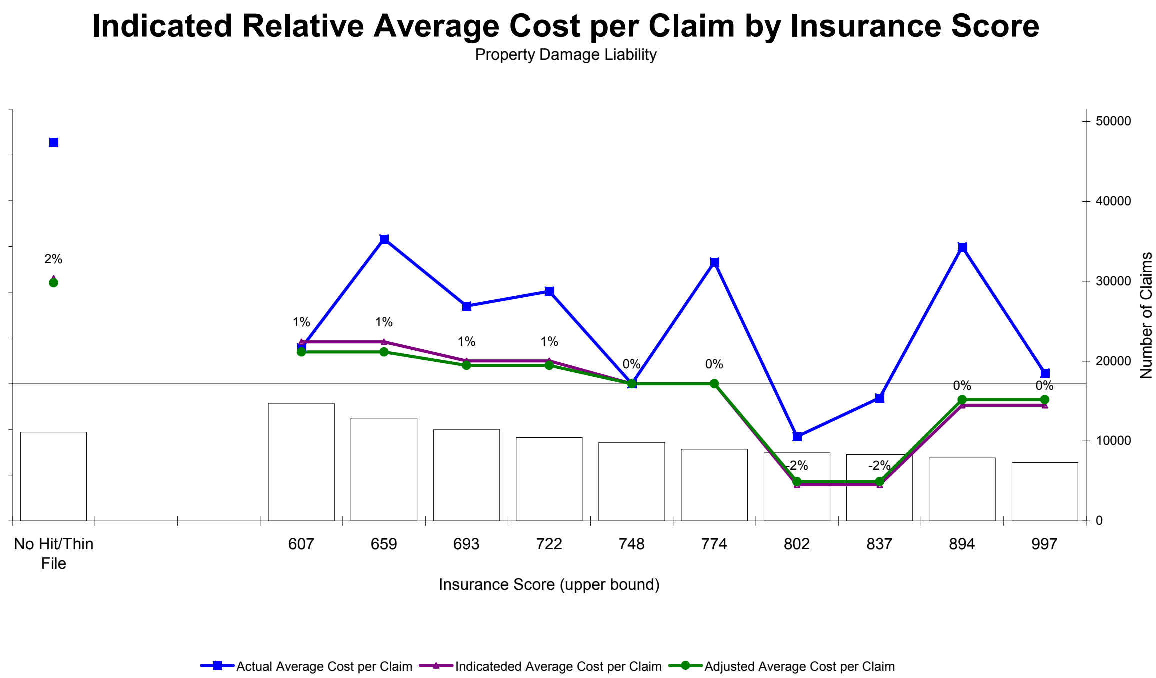


Exhibit XII:



We reviewed the statistics from the models that test the significance of the factors that were included. In statistical terms, a Type III test estimates a "P value" – the probability that the differences explained by a given risk factor are due to chance. In other words, the smaller the statistic, the more likely it is that the risk factor is identifying meaningful differences in risk.

For claim frequencies, the P values were all less than .002, suggesting that the findings are significant. Claim severities, as mentioned above, can tend to be more difficult to model, especially for coverages that have low claim frequencies. For all coverages except medical payments, the P values for claim severities were all less than .002, again suggesting significant differences in risk. The P value for medical payments claim severity was approximately .300, which would normally suggest dropping the factor when modeling claim severities. Since insurance score is the focus of the study, and is relatively "flat" for claim severities, we chose to leave the factor in place. This result does not change our opinion about the importance of insurance scores as a risk factor, and is consistent with our observation that claim frequency tends to be the main component explaining differences in pure premiums for insurance scores.

Importance

In the discussions about insurance scores that have taken place in the last decade, a question has sometimes been raised – "How important are insurance scores in predicting expected losses?" There are statistical measures that are available to measure importance, but there is not always agreement on exactly what to measure and how to do so.

The measure of importance used for this study is a measure of the relative impact of the various risk factors on the pure premiums.

To understand the measure of importance, consider a hypothetical risk class that has three subgroups: A, B, and C. The hypothesized risk class could be violation history with A equal to no violations, B equal to one violation, and C equal to two or more violations. The hypothesized risk class could be car usage with A equal to pleasure use, B equal to commuter use, and C equal to business use.

In our hypothetical subgroups A and C each contain one percent of the data being analyzed, and subgroup B contains the remaining ninety-eight percent.

Table 6:

Risk Class	Distribution Earned Car Years	Indicated Risk Factor
A	0.010	1.01
B	0.980	1.00
C	0.010	0.99
Total	1.000	

There is a very small difference in the risk factors between the three subgroups in Table 6. A number of other risk factors could have been hypothesized with much larger spreads between subgroups. In our measure of importance, a given rating factor is considered more meaningful as the spread in the risk factors increases between the subgroups and as the distribution of earned car years is more dispersed between the subgroups.

In this study, the actual calculation of importance was made by first rescaling the subgroup factors for each risk class so that the average for that class equaled one (1.00). We then subtracted 1.00 from the result and calculated the absolute values (i.e., made the signs of the values all positive). Finally, the earned car years were used to calculate a weighted-average, absolute value. These resulting statistics were then ranked by risk factor. Of all the risk factors included in the database for this study, insurance score ranked among the top three for each coverage, as shown in the following Table 7.

Table 7:

Coverage	Factor 1	Statistic	Factor 2	Statistic	Factor 3	Statistic
BI Liability	Age/Gender	.1808	Ins. Score	.1766	Geography	.1517
PD Liability	Age/Gender	.1835	Ins. Score	.1247	Geography	.1178
Pers. Inj. Prot.	Ins. Score	.2982	Geography	.2183	Yrs. Insured	.1959
Med Pay	Ins. Score	.2737	Limit	.2129	Age/Gender	.1594
Comprehensive	Model Year	.1619	Age/Gender	.1493	Ins. Score	.1254
Collision	Model Year	.2163	Age/Gender	.1572	Ins. Score	.1470

The analysis indicates that insurance score is an important risk factor and that it significantly explains risk that is otherwise not being explained by any other risk factor. Other methods of measuring "importance" could have been selected for analysis. Perhaps insurance score would not be in the top three for all measures, for all coverages. However, insurance score does not need to be in the top three to be considered important.

By-State Analysis

The graphs of Appendix Q present relative PD Liability claim frequencies for each of the fifty states. These frequency data are directly comparable to the countrywide claim frequency data presented in Exhibit III. While the by-state data exhibit greater fluctuation than the larger, countrywide database, the underlying pattern of decreasing claim frequencies with rising insurance scores is unmistakable in each of the fifty states. These data increase our confidence that an analysis of an individual state's data would produce the same general conclusions as has been drawn from the countrywide data.

Limitations of the Study

The study was limited to private passenger automobile insurance and as such the study results cannot be generalized and applied to other lines of insurance.

Actuarial principles do not require that risk factors demonstrate a causal relationship. The study was limited to the ability of insurance scores to predict the propensity for claim losses. No attempt was made to explain why insurance scores predict claim losses.

The study addresses the relative risk, or propensity for loss, between various levels of insurance scores. The relative risk factors in the study are not relative rate factors. Rate factors take into account the risk of loss plus various expenses and the cost of capital. It is highly unlikely that the relative risk factors in this report would be appropriate for use as rate factors by any specific insurance company.

The study tested a particular insurance score. It is reasonable to assume that other similarly constructed credit-based insurance scores will produce similar study results. However, not all insurance scores are identical in construction. There may exist credit-based insurance scores which do not show the same strong relationship to loss propensity as the score tested in this study.