The Impact of Personal Credit History on Loss Performance in Personal Lines

James E. Monaghan, ACAS, MAAA

Introduction

At the time of this writing, a process of both education and debate is occurring with regard to the use of personal credit history in the underwriting or rating of personal lines insurance policies. The insurance industry, the NAIC, and other interested third parties are all involved in educating both themselves and each other on such issues as correlation, multivariate correlation, causality and the social or actuarial appropriateness of using this tool in either underwriting or rating. Although the scope of regulators is more finely focused on rating, the recent trend towards tier rating and the utilization of multiple rating companies by members of the insurance industry has blurred the distinction considerably between the two. The use of personal credit history in personal lines insurance has therefore, through its manifestation in underwriting, gone largely unnoticed until recent years. The rapid increase in its use has brought credit history to the forefront of debate in many jurisdictions, in addition to its use in quasi-rating schema.

The development and use of third-party scoring algorithms for credit evaluation, and the proprietary nature of such models, has made it difficult for regulators, companies, agents and customers to get a firm grasp of the underpinnings of automated risk evaluation based on credit history. Apparently, it is not only actuaries who occasionally take the position that "if I can't touch it, is it actually real?" The key issues under debate are the existence (or non-existence) of a correlation between past credit history and expected loss levels (and which variables are responsible for that correlation) and the establishment of causal links for such correlation. Both will be addressed here, although only the former can be statistically analyzed. Causality will be addressed on an informational (and necessarily subjective) basis. The key questions that will be addressed in this paper are:

1) Is there a correlation between credit history and expected personal lines loss performance?
2) If so, which specific criteria within a credit file are indicative of abnormal loss performance (favorable or unfavorable)?
3) If this correlation exists, is it merely a proxy, i.e., is the correlation actually due to other characteristics (which may already be underwritten for or against, or rated for)?
4) As a corollary to 3), are there dependencies between the impact of credit history on loss performance and other policyholder characteristics or rating variables?
5) What are the ramifications of utilizing such data for underwriting and/or ratemaking?

Research Database Construction

The data utilized in researching the relationships between credit history and private passenger automobile loss experience was assembled from several sources. All policies originally written during calendar year 1993 were first identified. Earned premiums for the calendar/accident years 1993 through 1995 were then appended for all coverages. The longest exposure period for any given policy is therefore 36 months, in the case where the policy was written on January 1st, 1993 and remained inforce through December 31st, 1995. All policies were included in the database, regardless of whether or not they remained inforce through the end of the experience period, making the shortest possible exposure period for any given policy one day. Hence policies are not homogenous in either length of exposure or in coverages afforded. Also of note is the fact that the company did not utilize credit information in underwriting or rating of policies during this time period.

Incurred losses were then added, where incurred loss was defined as the sum of paid losses, case reserves, supplemental reserves on case (which are established to cover adverse development on known losses), loss expenses and salvage and subrogation recoveries. These losses were evaluated as of June 30th, 1996 for the exposure period January 1st, 1993 through December 31st, 1995. Incurred losses during accident year 1993 therefore had 42 months of development, those during accident year 1994 were developed 30 months, and those during accident year 1995 were developed 18 months. All earned premium and incurred loss were determined at the policy level, i.e., accumulated for all vehicles insured on the policy at any time during the experience period and for all coverages afforded on those vehicles.

Data was then appended to each policy record that defined the underwriting and rating characteristics of the policy at the time of initial writing. This dataset contained such information as number of drivers, number of vehicles, prior accident and violation activity, state of residence, residence type and stability and prior insurance carrier information. Some of these variables certainly would have changed value during the experience period for many risks. In order to provide predictive value, information was compiled which related to the conditions in effect at the time of writing.

The dataset was sent to a national credit vendor to append archived credit histories for each match that could be found. These credit histories were retrieved from credit files archived at the time each policy was written (or at the nearest three-month interval). Each record was then stripped of any identifying information (i.e. policy number, name, address) in order to ensure compliance with the Fair Credit Reporting Act. This action permitted analysis of the data without knowledge of the identity of any individual risk. Again, in order to provide predictive value, information gathered was pertinent to the conditions in effect at the time each policy was originally written. The credit information added to the dataset contained all of the information in the insured's credit file. The original listing of policies contained approximately 270,000 records. Matches were obtained on approximately 170,000 of those. This "hit rate" is rather low; recall, however, that many of the policies were no longer actively insured by the company and address and other information could have been outdated.

Queries were then constructed and run against this database, accumulating earned premium and incurred loss during the experience period for various combinations of policy characteristics. In fact, thousands of such queries were run, evaluating the loss ratio and loss ratio relativity of given subsets of data relative to others and to the whole. These subsets each contained one or more variables from the two groups underwriting/rating characteristics and credit characteristics. The database had a grand total of $394 million in earned premiums for all records combined. The results of these queries, and the conclusions that could be drawn from them, shed light on the startling foundations of the credit scoring models: the individual credit characteristics. A data dictionary containing the description of all fields utilized in the results contained herein can be found in the Appendix.

Limitations and Difficulties

The construction of the database caused some inherent difficulties in interpretation and also rendered most traditional ratemaking methodologies unusable. The dataset was not compiled with the intent of applying ratemaking methods and principles. Since the process of risk selection occurs on a policy basis, the data was compiled to be utilized in that setting; loss ratio relativity is the only meaningful measure of performance expected to arise from these data.

The credit file utilized was associated with one individual, although many policies have more than one covered driver. This individual was the named insured. The named insured may or may not have been the individual involved in prior accident or violation events, and may or may not have been involved in subsequent losses during the experience period. This difficulty arises from the use of policy level data. The question remains unanswered as to what kind of loss experience one can expect from, for example, a married couple with significantly different credit histories (as can be expected with policies written on recently married persons).

Another difficulty encountered was determining the appropriate method of binning the data, particularly where the independent variable was of the continuous type (dollars, for example). Any data grouping of a continuous variable will have greater stability when larger bins are employed. Many different bin groupings were used in such cases, although only one will be shown here for each example.

Results of Data Queries

The database contained a large number of variables relating to underwriting characteristics, rating characteristics and credit information. Space limitations preclude presenting information about most of the queries that were run and results obtained. A sampling of this data will be reviewed and discussed. The first section will contain information about individual credit characteristics. All earned premium and incurred loss dollars will be shown in millions unless otherwise specified. The aggregate loss ratio for the entire database is 76.3%; this number is higher than average for the private passenger auto industry but recall this is premium and loss experience during the first (at most) 36 months of experience from a block of newly written policies. New business in general produces higher loss ratios than longer-tenured business.

1. *Amounts Past Due (APD)*

APD is defined as delinquent amounts that are uncollected as of the report date. This amount is the sum of all delinquent amounts on the credit file, regardless of how many accounts are delinquent. A scheduled payment must be at least 30 days late before it appears on the credit file as delinquent. Note that there is a significant amount of premium volume in the categories below $10. This is due to a logistical difficulty with the data: some records contained the value $0, others were blank. In order to run queries, the data must be uniformly formatted, yet there could have been statistically significant differences in results for "blank" versus $0. Therefore, all records with blanks were assigned a value of $1. The premium and loss dollars in the categories below $6-20 should be considered included with $0.

APD	Earned Premium	Incurred Loss	Loss Ratio	Relative Loss Ratio	Fitted Relative LR
$0	$ 257.7	$ 180.9	70.2%	0.92	
1-2	45.8	31.8	69.3%	0.91	1.03
3-5	6.5	4.9	75.9%	1.00	1.07
6-20	4.7	4.4	94.0%	1.23	1.11
21-50	5.5	4.8	87.5%	1.15	1.16
51-99	5.8	5.8	99.7%	1.31	1.19
100-199	7.7	7.3	95.9%	1.26	1.22
200-499	12.0	11.1	92.7%	1.22	1.25
500-999	10.2	10.9	107.2%	1.41	1.28
1K-2K	10.1	9.9	97.2%	1.27	1.31
2K-5K	12.5	12.6	100.5%	1.32	1.35
5K-10K	7.8	8.3	106.1%	1.39	1.38
10K +	7.7	7.6	99.8%	1.31	1.41
Total	$ 394.0	$ 300.4	76.3%	1.00	

A linear regression performed on loss ratio relativity vs. logarithm of APD generated a coefficient of 0.83. The t-statistic for 99.5% significance level with 10 degrees of freedom is 3.17; the t-stat for this dataset is 5.65. Thus the null hypothesis that slope of the regression is 0 is rejected with 99.5% certainty. A less statistical observation would be that loss ratio increases as the APD increases, but the change is very small compared to the large jump in loss ratio from around 70% for $0 to the mid-nineties at almost any value greater than $0. This is somewhat counter-intuitive, as one might speculate that small delinquencies should not have the same impact as large ones. Recall, however, that what is being measured is impact on loss ratio, not credit worthiness or any other characteristic. Since the causal links are not established, preconceived notions should be considered with skepticism.

2. *Derogatory Public Records (DPR)*

DPRs include such items as bankruptcies, federal, state or municipal tax liens, civil judgments and foreclosures. The presence of a DPR on a credit file also has significant impact on future loss performance. This should come as no surprise, as this variable is the one that has been utilized in the personal lines industry for the longest time and is the most widely accepted.

DPR	Earned Premium	Incurred Loss	Loss Ratio	Relative Loss Ratio	Fitted Relative LR
None	$ 358.6	$ 264.7	73.8%	0.97	1.04
1	22.4	21.6	96.5%	1.27	1.18
2	7.1	7.4	104.2%	1.37	1.33
3 or more	5.9	6.7	114.1%	1.50	1.54

Linear regression on number of DPR vs. relative loss ratio generated an R^2 value of 0.95. The loss ratio for all DPR that had an outstanding liability on the file of greater than $0 is 102.2%, (relativity = 1.34) and premium volume of $31.1. Although many will not be surprised that there is a correlation with this variable, the size of the difference in loss ratio may confirm the underlying reason for its historic use.

3. *Collection Records*

A collection record is generated when responsibility for collecting a delinquent account (or trade line as they are generally referred) is transferred to a collection agency In general, this occurs when a delinquency is more than 120 days past due. Collection records can, however, occur for delinquencies that are not associated with a trade line, i.e., in the case of a utility bill.

Collections	Earned Premium	Incurred Loss	Loss Ratio	Relative Loss Ratio	Fitted Relative LR
0	$ 364.6	$ 270.1	74.1%	0.97	1.05
1	19.0	18.5	97.5%	1.28	1.21
2	5.5	6.0	108.4%	1.42	1 37
3 or more	5.0	5.9	118.6%	1.56	1.61

R^2 value for the regression of number of collections vs. relative loss ratio is 0.96. The loss ratio for any collections with outstanding liability greater than $0 is 107.6% with a premium volume of $22.3. The results for this variable are very similar to those for DPR. Although there is increasing loss ratio for increasing number of collections, the largest jump in loss ratio occurs between 0 and 1.

4. *Status of Trade Lines*

Each trade line is given a rating based on its current status. A rating of 0 indicates no information is available, while a rating of 1 indicates that the most recent payment made was as agreed, or no more than 30 days past the payment due date. Status codes 2-5 are used to indicated trade lines where the most recent payment made was 30-59, 60-89, 90-119, or over 120 days past due, respectively. Codes 7-9 are used to denote such situations as accounts which are being paid under a wage earner plan, are in repossession, have been written off as bad debt, and others.

Condition	Earned Premium	Incurred Loss	Loss Ratio	Relative Loss Ratio
All trade lines not rated 2-5	$ 314.8	$ 227.3	72.2%	0.95
At least 1 trade line rated 2-5	79.2	73.1	92.3%	1.21

Condition	Earned Premium	Incurred Loss	Loss Ratio	Relative Loss Ratio
All trade lines not rated 7, 8 or 9	$ 334.1	$ 240.8	72.1%	0.95
1 or more trade line rated 7, 8, or 9	59.8	59.6	99.6%	1.31

If these two types of ratings are viewed exclusively, the following results are obtained:

All trade lines rated 1	$ 286.7	$ 198.8	69.3%	0.91
1 or more rated 2-5, none 7-9	47.5	42.1	88.6%	1.16
1 or more rated 7-9, none 2-5	28.1	28.5	101.5%	1.33
1 or more of each type	31.7	31.0	97.8%	1.28

When combining both types of trade line status, Note the difference between this variable and APD: APD refers to amounts that are currently delinquent, whereas status refers to the account evaluation based on the most recent payment made.

5. *Age of Oldest Trade Line*

This variable measures the time between the report date and the oldest date that any trade line was opened. Trade lines include more than just revolving-type accounts; home improvement loans, installment loans, car loans and mortgages are also considered trade lines. The years listed in the following table reflect the fact that the database involved policies written in 1993.

Year of Opening/ Age of Oldest Line	Earned Premium	Incurred Loss	Loss Ratio	Relative Loss Ratio	Fitted Loss Ratio Relativity
1963 & Prior (30+ yrs)	$ 9.6	$ 6.4	66.4%	0.87	0.79
1964-1968 (25-29 yrs)	24.4	14.7	60.2%	0.79	0.85
1969-1973 (20-24 yrs)	41.0	29.4	71.8%	0.94	0.91
1974-1978 (15-19 yrs)	68.3	48.9	71.5%	0.94	0.97
1979-1983 (10-14 yrs)	82.9	60.5	73.0%	0.96	1.03
1984 (9 years)	26.5	20.2	76.2%	1.00	1.07
1985 (8 years)	26.4	20.6	78.2%	1.03	1.08
1986 (7 years)	23.2	19.3	82.9%	1.09	1.09
1987 (6 years)	21.2	19.8	93.3%	1.22	1.10
1988 (5 years)	18.9	15.9	84.2%	1.10	1.11
1989 (4 years)	16.5	12.8	77.6%	1.02	1.13
1990 (3 years)	14.0	12.2	87.2%	1.14	1.14
1991 (2 years)	10.4	9.6	92.5%	1.21	1.15
1992 (1 year)	10.7	10.2	95.0%	1.25	1.16

The t-statistic for the dataset is (5.86); the t-stat for the 99.5% significance level for 12 degrees of freedom is (3.06), thus the null hypothesis that the slope of the regression is zero is rejected at the 99.5% confidence level. The linear regression on years since opening and relative loss ratio generated an R^2 value of 0.86. Here is a correlation that has drawn skepticism: are these results arising merely from the age of the insured, rather than the age of the oldest trade line? This question will be answered in the multivariate section using driver age data, but one can nevertheless deduce that if younger drivers are responsible for the poorer loss results in the lower section of this table, then the same results should be found in the class experience for those ages. This is not true for policies in this dataset, nor is it true for the insurance industry as a whole.

6. *Non-Promotional Inquiry Count*

A strong relationship was also found between loss ratio and non-promotional inquiry count. An inquiry is posted to an individual's credit history file any time that file is reviewed. Many such inquiries are made for direct mail marketing campaigns, which are not requested by the insured. These inquiries are excluded from consideration, and only those that arise from the activities and requests of the insured are included. Federal law prohibits the maintenance of inquiry records for longer than 24 months, at which point they are purged by the credit bureaus.

Number of Inquiries	Earned Premium	Incurred Loss	Loss Ratio	Relative Loss Ratio	Fitted Loss Ratio Relativity
0	$ 130.9	$ 92.9	71.0%	0.93	0.92
1	82.7	58.4	70.6%	0.93	0.96
2	55.1	40.9	74.2%	0.97	0.99
3	37.4	28.8	77.0%	1.01	1.03
4	24.9	20.8	83.4%	1.09	1.07
5	17.5	15.2	87.0%	1.14	1.11
6	12.0	9.7	80.6%	1.06	1.15
7	8.7	7.9	90.8%	1.19	1.18
8	6.0	5.3	87.7%	1.15	1.22
9	4.4	4.8	110.0%	1.44	1.26
10	3.2	3.2	100.1%	1.31	1.30
11-15	7.6	8.2	108.6%	1.42	1.41
16 or more	3.7	4.4	117.5%	1.54	1.60

The t-statistic is 9.51; the t-statistic for 11 degrees of freedom for the 99.5% significance level is 3.11. The correlation coefficient for the regression is 0.94. Once again, a single characteristic from an individual's financial management history has a surprisingly large and consistent impact on loss ratio, even in the smaller premium volume cells.

7. *Leverage Ratio on Revolving-Type Accounts*

This variable is calculated as the ratio of the sum of all revolving debt to the sum of all revolving account limits. Trade lines such as mortgages and installment loans are excluded due to the difference in the nature of such accounts. Since leverage ratio is a continuous-type variable, it was difficult to determine how to define data bins.

When the data was initially reviewed, it was found that the loss ratio relativity for leverage ratio = 0% was 1.04, while the relativities for leverage ratios below 10% were in the 0.75-0.90 range, and subsequently rose as leverage ratio increased. This anomaly occurred due to the fact that records with limits of $0 caused a zero divide, and were given a default leverage value of 0%. Therefore, the table displays a more detailed breakdown of records with 0% leverage, due to the marked difference that was evident in loss ratio impact where limits were low or zero.

Leverage Ratio	Revolving Limits	Earned Premium	Incurred Loss	Loss Ratio	Relative Loss Ratio	Fitted Loss Ratio Relativity
0%	$0	$ 20.3	$ 20.0	98.4%	1.29	
0%	$ 1 – 499	8.6	8.0	93.0%	1.22	
0%	$500 or more	35.8	23.2	64.9%	0.85	0.84
1-10%		91.6	58.9	64.3%	0.84	0.85
11-39%		91.6	65.0	70.9%	0.93	0.92
40-60%		41.8	31.5	75.2%	0.99	1.01
61-80%		30.5	24.8	81.2%	1.07	1.08
81-100%		24.6	21.7	88.1%	1.16	1.14
101% or more		49.0	47.3	96.6%	1.27	1.26

T-statistic for this dataset (excluding the low-limit, 0% leverage group) is 26.3, using weighted means of the leverage ratio ranges. The 99.5% confidence t-stat is 4.03. The R^2 value is 0.996. The practice of some insurance companies of utilizing the characteristic 'possession of a major credit card' as an underwriting criteria for company placement seems justifiable when the top segment of this table is considered. This depends of course on the average rate level of the writing company.

8. Revolving Account Limits

This variable is the denominator in the calculation of leverage ratio discussed previously. It is the sum of credit limits for all revolving-type trade lines on the report for a given individual.

Revolving Limits	Earned Premium	Incurred Loss	Loss Ratio	Relative Loss Ratio
$0	$ 41.5	$ 39.4	95%	1.25
$1 - $500 ·	9.8	8.6	88	1.15
501-1000	13.0	12.5	96	1.26
1001-1500	12.0	10.3	86	1.13
1501-2000	11.2	10.8	96	1.26
2001-2500	10.0	8.1	81	1.06
2501-3500	18.8	15.3	81	1.07
3501-5000	26.0	20.6	79	1.04
5001-7500	36.2	28.2	78	1.02
7501-10 K	31.4	24.5	78	1.02
10 – 15 K	50.8	34.8	69	0.90
15 - 20 K	37.7	24.0	64	0.83
20 – 25 K	27.6	19.0	69	0.91
25 – 30 K	18.7	12.9	69	0.91
30 – 40 K	22.0	13.5	61	0.80
40 – 50 K	10.9	7.3	67	0.88
50 K +	16.4	10.7	65	0.85

Correlation coefficient for this regression is (0.78), using midpoints of the limit ranges. The first conclusion that could be drawn is that this correlation only duplicates the one already discussed in the leverage ratio section. This will be addressed in the multivariate section. Another conclusion that has been drawn is that this variable is directly correlated to personal income, and use of revolving limits in any underwriting or rating program is discriminatory towards lower income individuals (disparate impact). This may or may not be true; the data does not contain income information. It would be erroneous however, to assume that all people with low revolving limits are also low-income. Many people choose not to use credit; others may have substantial income but low revolving limits due to the fact that they cannot obtain such credit lines based on their past bill payment performance.

Many other individual variables were reviewed from the credit file. Some exhibited correlation to loss ratio at various significance levels, others had no such correlation. Those displayed thus far, however, show a systematic predictive power that requires explanation and understanding.

Causality

Explanation of these correlations, for the most part, cannot be found in the data assembled for this research. I would be remiss, however, if I did not at least attempt to set down those arguments which could be made suggesting reasonable causal links between an individual's bill paying history and expected loss experience for insured losses under a private passenger auto insurance policy.

Before listing such arguments, it is first appropriate to review the Actuarial Standards of Practice #12, entitled "Concerning Risk Classification". The relevant section is 5.2, which states the following:

5.2 Causality – Risk classification systems provide a framework of information which can be used to understand and project future costs. If a cause-and-effect relationship can be

established, this tends to boost confidence that such information is useful in projecting future costs, and may produce some stability of results.

However, in financial security systems, it is often impossible or impractical to prove statistically any postulated cause-and-effect relationship. Causality cannot, therefore, be made a requirement for risk classification systems.

Often, the term "causality" is not used in a rigorous sense of cause and effect, but in a general sense, implying the existence of a plausible relationship between the characteristics of a class and the hazard for which financial security is provided. For example, living in a river valley would not by itself cause a flood insurance claim, but it does bear a reasonable relationship to the hazard insured against, and thus would be a reasonable basis for classification.

Risk classification characteristics should be neither obscure nor irrelevant to the protection provided, but they need not exhibit a cause-and-effect relationship.

Clearly, the operative word in this Standard of Practice is irrelevant, as the historical data in question is not obscure. Therefore, arguments must be put forth which, despite being speculative, are reasonable statements that a reasonable person would find relevant.

Why would an individual who has current or past difficulties with meeting financial obligations be expected to have above-average costs to an auto insurer? Since there is an administrative expense associated with the processing of insurance premiums and related transactions, it can be argued that subsequent lapses in the individual's payment history is a direct cost to the insurer. This cost would fall under the category of expenses, however. The focus here is loss costs.

Maintenance

The argument has already been made, and often, that auto insurers' underwriting practices are created for risk selection, and one characteristic that is viewed as favorable for selection is described in various quarters as "stability" or "responsibility". Few, however, could give an objective definition of how one could measure such a characteristic, but historically many customer characteristics have been utilized as an assumed proxy for this nebulous attribute, such as home ownership, marital status, number of vehicles, coverage and limits selected, etc. It is entirely possible that a person's current and historical management of debt is another indicator that could be utilized to identify this quality. If a person manages their financial affairs responsibly such that debts are paid on time, they may also take the same approach to the maintenance of other aspects of their lives, including their automobile. A vehicle kept in good working order and condition is less likely to be involved in an accident than one that is not, all other things being equal. Such an individual may also take greater care in operating that vehicle.

Morale Hazard

The CPCU textbook "Personal Insurance" defines morale hazard in the following way:

> **Morale hazard is a condition that exists when a person is less careful because of the existence of insurance. Morale hazard does not involve an intent to cause or exaggerate a loss. Instead, the insured becomes careless about potential losses because insurance is available. Leaving the keys in an unlocked car or allowing fire hazards to remain uncorrected are examples of morale hazard. Morale hazard results in additional losses that drive up the cost of insurance because of injuries and damage that could have been prevented."**

The previous discussion of responsibility could lead to the argument that individuals who are careless in the management of finances also present a morale hazard in the area of automobile insurance.

Claims Consciousness

An insurer's loss experience measures dollars of loss which are paid on claims that are filed. The number of claims filed is less than the number of accidents that actually occur. Consider two risks that are identical in all ways (from an insurer's perspective) except for the fact that one manages their financial affairs much better than the other does. The risk who has a troubled financial history and condition is much more likely to be in debt and to a larger degree; the need for capital to satisfy financial obligations has a bearing on decisions made in many areas of his/her life. Suppose for example, that these two risks are both involved in an auto accident, involving no injuries, but causing property damage to their own vehicles which is some nominal amount (say, $100) more than the deductible. The risk whose financial condition is more sound has a disincentive to file the claim. It may impact his/her rates at the next renewal; the time and effort involved may not be even worth the compensation obtained. The risk with the poorer record of financial management has a greater incentive to file the claim and obtain the compensation, as it has greater value to that individual.

Fraud: Increased Severities

Continuing with these same two risks, consider now the situation in which the damage to property was much greater than the deductible; the vehicles each sustained damage measuring in the thousands of dollars. If an auto repair technician suggested a relatively easy way of recouping the deductible for the insured, or the benefits of padding the repair costs, the individual under the greater financial pressure would be more susceptible to acquiesce. This does *not*, however, imply that risks with poor bill-paying histories have any less integrity than other risks. Some people would never commit fraud on any level; others would do so with no need for provocation or encouragement; still others could be convinced to do so only under the proper conditions. This argument only implies that any individual who *could* be induced to participate in this level of fraud would be more likely to do so if they were under financial pressure from other sources.

Fraud. Increased Frequencies

The presence of severe financial pressure could also produce claims that would not have existed otherwise. There is some segment of the population that either does or could view the insurance mechanism as a financial opportunity. Fraudulent claims in the form of staged accidents, phantom claimants, phantom vehicles or arson are a way that an individual can extract funds from the insurance mechanism. Once again, this argument does not imply anything about the integrity of a risk with poor bill-paying history. What it does assert is that an individual with severe financial pressure could look to all possible sources of funds to alleviate that pressure. Therefore, any individual who was capable of committing this type of fraud is more likely to do so given the existence of that financial pressure compared to the absence of it.

Stress

The assumption is made here that individuals who are under financial pressure from debt exist under a greater level of stress than average. This stress could exist from the associated worries over future impact of financial condition. Individuals under such stress may be less focused on proper operation of a motor vehicle and make them more susceptible to accidents resulting from chance occurrences or distraction. It would be useful if there were some other condition which could produce this same level of stress, for which loss data was available, to strengthen the argument. A few currently coded customer characteristics could be considered candidates. One such variable is number of children under the age of 16. One must first make the assumption that risks with three or more children under the age of 16 have a higher level of stress than average. Whether or not one agrees with that probably depends on whether or not they are a parent! In any case, the loss ratio for such risks reviewed in a 1993 research study was over 20 points higher than average. Another possible variable candidate could be self-employed risks. The added responsibilities and worries of a small business owner could imply that their level of stress is higher than average. From that same 1993 study, self-employed risks had a loss ratio which was roughly 15% higher than average.

It is important to make note that this list is not suggested as a menu from which to select the one correct answer. It is likely that the impact on losses of financial management history is a cumulative impact of some or all of these situations, as well as others not listed here.

Multivariate Analysis: Underwriting Characteristics

There have been many assertions made, in the absence of data, about this relationship between loss experience and credit history. The following comes from the NAIC's "Credit Reports and Insurance Underwriting", dated December 14, 1996:

> **"There still is insufficient data to prove to all regulators' satisfaction whether credit history ... are or are not valid indicators ... independent multivariate analysis, a statistical method some regulators view as necessary, has not been performed." (p. 15) "Some regulators suggest that an unbiased and reasonably precise multivariate analysis is necessary to determine the actual rating factor.... They ask whether a person's credit history is truly correlated with future loss experience or whether it is a spurious correlation?" (p. 17)**

It is beyond the scope of this paper to determine whether or not the loss ratio method is appropriate to analyze this particular database. This method is questioned in the aforementioned NAIC report; the assertion is made that small errors in pricing for a number of rating factors could add up to a fairly significant overall pricing error, making loss ratios a biased measure. For purposes here, it is assumed that differences in relative loss ratio are due to differences in expected average loss costs after adjustments for individual premiums, and that this method is a reasonable way of measuring such differences when reviewing more than one variable simultaneously.

The utilization of the factors discussed earlier when performing multivariate queries tended to produce premium volumes in the individual cells which were smaller than desired for credible results. Strict credibility adjustments could not be performed, due to the fact that a) claim counts were not contained in the data and b) the premium and loss on each record arose from all coverages combined. In order to generate larger premium volumes, the credit variables were combined into four mutually exclusive profiles. These profiles were designed to achieve significant loss ratio differences and significant premium volumes described by each. Group A is defined by those characteristics producing the highest loss ratio, i.e., derogatory public records, collection records and large amounts past due. Group D is defined by those characteristics producing the lowest loss ratio, i.e. low leverage ratio, high age of oldest trade line, good account ratings, etc. The precise definitions of the four groups are contained in the appendix. These profiles will be used in this multivariate section for the sake of simplicity and brevity. Each individual credit characteristic was reviewed in conjunction with the underwriting and rating variables described herein. The variables discussed here are a sampling of all those reviewed; they were selected based on assumed relevance. The overall performance of these four profiles is as follows:

Group	Earned Premium	Incurred Loss	Loss Ratio	Loss Ratio Relativity
A	$ 74,279	75,333	101.4%	1.33
B	158,922	124,723	78.5%	1.03
C	69,043	47,681	69.1%	0.91
D	91,746	52,688	57.4%	0.75

Prior Driving Record

The loss performance of various prior driving record combinations is influenced by two significant factors: the underwriting practices of a given company and the experience modification system utilized in rating. Earned premium and incurred loss were aggregated for risks based on their prior accident and violation activity (in the three year period before they were originally written) and based on credit category (A-D):

Prior Driving Record	Group A Prem	Group A LR	Group B Prem	Group B LR	Group C Prem	Group C LR	Group D Prem	Group D LR	All Groups Prem	All Groups LR
No incidents	28.4	93%	66.0	71%	30.7	64%	45.8	53%	170.9	68.6%
1minor*	8.0	94%	17.3	68%	7.5	68%	8.4	50%	41.2	69.4%
1 at-fault accident	3.7	101%	7.7	74%	4.1	68%	5.9	65%	21.4	75.2%
1 non-fault acc.	6.6	109%	14.8	81%	7.3	70%	9.9	70%	38.7	80.7%
2 minors*	2.5	86%	6.0	59%	1.9	41%	2.4	43%	12.8	58.7%
2 incidents (any)	6.5	108%	13.5	96%	6.6	82%	7.9	64%	34.4	88.2%
All other (more Than 2 incidents)	18.6	114%	33.7	95%	10.8	83%	11.5	66%	74.6	93.1%

* *minor refers to a minor moving violation*

The favorable overall performance of the category '2 minor moving violations' can be attributed to both underwriting practice and experience modification surcharge system of the company from which this data was obtained. Of note here is the marked consistency of the loss ratio relationships across credit groups, regardless of prior driving record. Loss ratio relativities, calculated relative to each driving record sub-group, display this consistency:

Group	A	B	C	D	All Groups
No incidents	1.36	1.04	0.93	0.77	1.00
1minor moving violation	1.36	0.98	0.98	0.72	1.00
1 at-fault accident	1.35	0.99	0.90	0.87	1.00
1 non-fault accident	1.35	1.00	0.87	0.86	1.00
2 minor moving violations	1.47	1.01	0.69	0.74	1.00
2 incidents of any kind	1.23	1.08	0.93	0.73	1.00
All other (> 2 incidents)	1.22	1.01	0.89	0.70	1.00
Total	1.33	1.03	0.91	0.75	

Of particular note in this table is the wide difference in performance between clean driving record/poor credit history risks (93%) vs. poor driving record/good credit history risks (66%).

Age of Driver

It could be argued that the loss experience for poorer credit history risks is influenced by driver age distribution. If a disproportionate percentage of young drivers are contained in Group A, then credit history is merely substituting for age. However, as stated earlier, this would only be true if loss experience for younger drivers was adverse, which is not the case. There is a distributional difference in the four groups by age, but the loss experience relationships across credit groups is again robust:

Age of Driver 1	A Prem	A LR	B Prem	B LR	C Prem	C LR	D Prem	D LR	Total Prem	Total LR
< 25	$ 3.8	121%	$23.6	75%	$ 1.4	51%	$ 1.9	53%	$ 30.8	78%
25-34	21.1	103%	55.8	79%	22.6	66%	8 9	63%	108.4	80%
35-39	13.0	100%	21 8	81%	12.9	65%	13.0	54%	60.7	76%
40-44	12.4	109%	18.5	82%	10.4	76%	15.6	52%	57.0	79%
45-49	9.8	93%	14.6	83%	8.2	·76%	14.8	58%	47.4	76%
50-59	9.2	97%	14.4	78%	7.9	68%	16.5	53%	48.0	71%
60+	3.8	110%	8.3	75%	4.9	81%	20.0	67%	37.1	75%

Some of the individual cells in this table have significantly lower premium volumes than prior tables; they are shown nonetheless for completeness. Clearly, age of driver is not the cause of the poor loss experience in Group A.

Age of driver was also reviewed in conjunction with many of the individual credit variables. For example, the following is the cross-hatching of relative loss ratios for age of driver and non-promotional inquiry count:

Inquiry Count	Age of Driver 1 ⇒ Under 30	30-39	40-49	50-59	60+	Total
0-3	1.01	0.95	0.95	0.87	0.92	0.95
4-7	1.09	1.07	1.18	1.06	1.38	1.12
8-15	1.22	1.34	1.32	1.43	1.69	1.33
16+	1.48	1.88	1.25			1.56

(values are not shown for cells with premium volume less than $ 0.5 M)

The variable age of oldest trade line, reviewed earlier, could have a relationship to losses that is dependent upon age of operator. When these two variables were combined, the impact exhibited independence:

Age of Oldest Trade Line	Age of Driver 1 ⇒ 26-30	31-35	36-40	41-45	46-50	51-55	56-60	60+	Total
< 7 years	1.15	1.23	1.19	1.43	1.25	1.19	1.44	1.15	1.15
7-9 years	1.02	1.03	1.01	1.20	0.96	1.07	0.87	0.92	1.05
10+ years	0.90	0.93	0.94	0.95	0.94	0.87	0.89	0.98	0.93

Classical Underwriting Profile

Historically, the underwriting function has identified and selected for various combinations of characteristics. The risk groups exhibited lower than average frequency of loss, which in the absence of premium adjustments, produced more profitable results. One such profile is the married, multicar, homeowner risk with clean driving record. In an effort to produce a favorable loss ratio within Group A, this characteristic was evaluated:

	Married multicar homeowner				All risks NOT married multicar homeowner			
Group	Clean Driving Record		All other		Clean Driving Record		All other	
A	$ 10.2	97%	10.6	102%	$ 27.8	92%	$ 25.6	113%
B	22.3	77%	20.2	85%	62.9	69%	53.4	88%
C	14.5	76%	13.5	76%	24.4	58%	16.7	74%
D	20.2	57%	16.0	58%	34.4	50%	21.2	70%
Total	67.3	74%	60.3	79%	149.5	67%	116.9	88%

Again, it is important to keep in mind that these results are heavily influenced by underwriting practice at the time of writing by a given company; this can influence column totals. The underwriting function, however, had no knowledge of the information that defines credit groups A-D, and the relationships across these groups are again consistent.

Rating Territory

A key concern voiced by regulators in at least a handful of states is the potentially disparate impact that the utilization of credit history in underwriting or rating could have on lower income urban risks. This paper will not address whether or not income levels in urban areas are in fact lower than suburban or rural areas. The issue of rating territory, however, was analyzed. Although rating territory was not a variable in the original database, subsequent state profiles were developed for inforce policies in order to determine distribution of risks by credit characteristics (again using the Groups A through D) in a sampling of states. The exposure distribution shown below exhibited no clear-cut disparate impact on urban territories when compared to non-urban territories:

State	Exposure Distribution Type	A	Group B	C	D
Connecticut	Urban	14%	32%	12%	42%
	All Other	13	29	13	46
	Total CT	13	30	12	45
New York	New York City	10	26	8	55
	Other urban	14	23	11	52
	All other	13	25	13	49
	Total NY	13	25	12	50
Ohio	Urban	14	20	12	54
	All other	10	19	16	54
	Total OH	11	20	15	54

Data is also available for many other underwriting characteristics, including number of vehicles, number of drivers, residence type, residence stability, job stability, prior insurance type, gender, marital status and many others. These characteristics were also queried against the individual credit variables, in addition to queries run against the four groups utilized above. The results were very similar. There were no variables that produced even roughly uniform results across the credit characteristics.

Multivariate Analysis: Credit Characteristics

Another group of variables that was analyzed is credit characteristics in combination with other credit characteristics. This is necessary to ensure that no dependencies or cross-correlations exist within these characteristics. As with the other analyses, this group contains many cross combinations that were reviewed; only a sampling will be discussed here.

Leverage and Revolving Limits

It was noted in single variable section that leverage ratio could be duplicating the impact of revolving account limits. When reviewing the numerator of leverage, revolving balances, it was found that there was virtually no relationship between that variable and loss ratio (R^2 value of 0.04). The array of loss ratio relativities (for all cells with premium greater than $ 0.5 M) for leverage ratio versus revolving limits shows the independence of their impacts:

		Leverage Ratios ⇒						
Revolv. Limits	Selected Midpoint	0% 0.00	0-50% 0.25	50-75% 0.625	75-100 0.875	100% + 1.20	All	Correl. Coefficient
$ 0	0	1.27	1.25			1.18	1.25	
1-999	500	1.02	1.01	1.35	1.38	1.34	1.21	0.87
1K-3K	2000	0.96	1.11	1.15	1.23	1.33	1.16	0.97
3K-5K	4000	0.78	0.99	1.04	1.19	1.34	1.05	0.98
5K-10K	7500	0.77	0.95	1.11	1.13	1.25	1.01	0.97
10K-25K	17500	0.78	0.83	1.09	1.07	1.07	0.88	0.88
25K +	35000	0.65	0.85	0.87	0.95	0.98	0.86	0.92
Total All		1.08	0.89	1.07	1.16	1.24	1.00	0.74
Correl Coefficient		-0.72	-0.74	-0.80	-0.90	-0.86	-0.87	

Note the consistency of the coefficients in both directions. This would not exist if one variable simply proxied for the other. In more general terms, risks with high leverage ratios have poorer loss performance than those with lower leverage ratios, regardless of limits; risks with low revolving limits have poorer loss performance than those with higher limits, regardless of leverage ratio.

Derogatory Public Records and Collections

Given the similarity of the distribution and loss results of these two characteristics, it might be expected that there is overlap between the two, i.e., individuals that exhibit one type of record commonly exhibit the other. This did not turn out to be the case:

DPR	Collections	Earned Premim	Loss Ratio	Loss Ratio Relativity
0	0	$ 339.2	72%	0.95
0	1	17.2	94%	1.23
1	0	13.7	96%	1.25
	Total 1any	30.9	95%	1.24
0	2	4.8	93%	1.22
1	1	3.1	88%	1.15
2	0	3.4	107%	1.41
	Total 2 any	11.2	96%	1.26
	Total 3 or more	12.6	117%	1.53

Each variable produced poor loss results regardless of whether or not the other variable was present. Both variables also had significant distributional volume.

Leverage Ratio and Inquiry Count

If the basis for the relationship between credit history and loss performance can be attributed to a more general characteristic, one might refer to that characteristic as financial stress, distress or duress. Since leverage ratio and high inquiry count can be expected to occur under such situations, it is reasonable to assume that there may be some overlap between these two variables also. As with the other multivariate combinations that are reviewed, it is important to keep in mind the distinction between distributional imbalance and loss ratio imbalance. In the driver age vs. credit group (A-D) table, there is a clear distributional imbalance, with older drivers being disproportionately represented in the best performing credit group. The loss ratio impact, however, remains consistent across credit groups and is not offset by the inclusion of age. This is also true to a lesser degree in the table of loss ratio relativities below: risks with higher leverage ratios are disproportionately represented in the higher inquiry count groupings, but the two-way impact on loss ratio remains:

Limits:	<500	>500			Leverage Ratio		
Inquiries	0%	0%	1-50%	50-75%	75-100%	100%+	Total
0	1.25	0.74	0.86	0.94	1.01	1.04	0.93
1-3	1.27	0.87	0.86	1.03	1.05	1.26	0.96
4-6	1.23	1.12	0.95	1.21	1.57	1.30	1.10
7-10	1.24		1.20	1.36	1.22	1.35	1.25
11+			1.18	1.28	1.54	1.99	1.46
Total	1.26	0.85	0.89	1.07	1.16	1.24	1.00

Trade Line Counts and Status

In addition to searching for variables that duplicated loss ratio impact within the credit characteristics, bivariate tables were reviewed to determine if some variables partially mitigated those impacts. For example, trade line status showed a strong impact earlier. One could argue that the impact of any trade line not rated 1 would diminish as the total number of trade lines increases. That is, if just one trade line is not in good standing, should that not have less significance for a risk with many trade lines, compared to one with only a few? The following table reveals that this appears not to be true generally:

Total Trade Lines	Total Rated 2 through 9	Earned Premium	Loss Ratio	Loss Ratio Relativity
1	0	$ 12.9	78%	1.02
	>0	3.3	116%	1.52
2	0	9.7	88%	1.15
	>0	6.2	103%	1.34
3	0	9.0	72%	0.94
	>0	8.1	93%	1.22
4	0	13.2	68%	0.89
	>0	5.1	90%	1.18
5	0	13.8	69%	0.91
	1	2.3	101%	1.32
	2 or more	3.0	104%	1.37
6	0	14.3	72%	0.94
	1	2.3	94%	1.23
	2 or more	3.1	117%	1.54
7-8	0	31.4	67%	0.88
	1	4.7	96%	1.26
	2	2.4	103%	1.36
	3 or more	4.3	105%	1.38
9-10	0	31.4	66%	0.87
	1	4.6	101%	1.33
	2-3	3.7	95%	1.25
	4-6	2.5	88%	1.16
	7 or more	0.5	134%	1.76
11-15	0	67.7	66%	0.86
	1	9.9	76%	0.99
	2-3	7.2	91%	1.19
	4-6	5.2	88%	1.15
	7 or more	2.3	106%	1.39
16 or more	0	75.6	69%	0.91
	1	13.5	89%	1.16
	2-3	8.5	82%	1.07
	4-6	5.5	99%	1.29
	7 or more	6.3	97%	1.27

Derogatory Public Records and Collections: Age and Amount

Another area of concern for both regulators and the insurance industry is the severity of a given event and its age. It is common practice for other variables, such as prior claims, to be evaluated differently based on their severity or amount paid. Thresholds are established to determine whether or not experience modification surcharges should apply in such cases. The age of a claim is also an important consideration in making underwriting decisions for private passenger auto applications. This concept is being applied to credit characteristics as well, as insurance companies apply different criteria to both age and amount when it comes to such items as DPRs and collections. The most commonly used vendor scoring algorithm also applies lesser weights to older events. This research database unfortunately was not large enough to have sufficient premium volumes in all the sub-groups, but those that have substantial weight indicate that severity and age may not be nearly as relevant factors as the existance of the record itself:

Age of Event	*Event=Collection* Premium	Loss Ratio	*Event=Derog. Public Record* Premium	Loss Ratio
Within 12 months	$ 5.8	110%	$ 7.6	103%
12-24 months	7.3	108%	7.5	93%
24-36 months	5.7	102%	6.1	107%
36-48 months	3.7	100%	4.7	106%
48-60 months	2.9	90%	3.6	111%
60-84 months	3.8	99%	5.9	92%
No collection records	364.7	74%	No DPR 358.9	74%

Amounts	*Event=Collection* Premium	Loss Ratio	*Event=Derog. Public Record* Premium	Loss Ratio
$0	$371.7	74%	$362.9	74%
$1 - $49	3.6	98%	6.9	95%
$50 - $99	3.7	102%	0.2	-
$100 - $499	9.6	106%	4.4	99%
$500 or more	5.4	120%	19.6	106%

Again, there were hundreds of other combinations of variables reviewed and analyzed; these have been provided as a sample. What has arisen is a significant number of variables within the credit history of an individual each of which has independent influence on private passenger auto loss experience. Such an environment lends itself most readily to a scoring-type mechanism, as the variables can be assigned independent weights that can be accumulated for an overall impact estimate for a given potential applicant. But the social and regulatory acceptability (or lack thereof) of these relationships has made it such that univariate scoring models are not viewed as the most favorable way of treating this particular set of data.

Other Impacts: Retention

One of the variables that was included in the research database was an indicator which designated whether or not a policy was still inforce at the end of the experience period, December 31st, 1995 (anywhere from 24 to 36 months since policy inception). The length of time that an auto policy remains inforce has a direct relationship to overall profitability, both from a loss and an expense standpoint. Characteristics that indicate better policy retention therefore indicate better expected experience over the lifetime of the policy.

The credit characteristics reviewed showed that in general, risks with better bill payment histories were retained at a higher rate than those with poorer bill paying histories. The reason for non-renewal was not available, therefore policies could have been no longer active due to a variety of reasons such as price shopping, underwriting cancellation, non-payment of premium, or any other reason for which a policy can normally cease to be inforce The following table shows percentages of policies still inforce at the end of the experience period for various categories:

All policies	48%	Number of Inquiries = 0	51%
		1-3 inquiries	48%
Policies with no collection records	49%	4-6 inquiries	44%
One collection record	36%	7-10 inquiries	41%
2 or more collections	30%	11 or more inquiries	33%
No derogatory public records	49%		
One DPR	38%	Leverage = 0 ($0 limits)	33%
Two or more DPR	33%	=0 ($1-$500 limits)	39%
		= 0 (limits > $500)	51%
Amounts Past Due = $0	52%		
$1 - $20	52%	0% - 50%	53%
$21 - $100	40%	50% - 75%	47%
$101 - $499	36%	75% - 100%	44%
$500 or more	33%	100% or more	38%

It could appear as though the increase in losses and the deterioration of retention are two effects of the same cause. This is not the case, however, as the loss ratio variation by, for example, number of collections still exists within both subsets of policies: those that remained inforce at the end of the experience period and those that did not. The loss ratios for policies still inforce are 72%, 101% and 114% for risks with none, one, or two or more collections, respectively. The same values for policies that did not remain inforce throughout the experience period are 80%, 93% and 113% for risks with none, one, or two or more collections. This pattern is true for other variables as well. This is a second way in which credit history can impact loss experience.

Homeowners Line of Business

A database was constructed to analyze the impact of credit history on loss experience for the homeowners line of business. The procedure was nearly identical to that described above for the auto line of business, with the exception that the policies included were those originally written in policy years 1993 and 1994. In addition to obtaining the credit data at the time the policy was written, similar data was obtained on those same policies at later dates. This was done in an effort to determine what percentage of risks experience significant changes in their bill-paying profiles over time. Policies were not included in the study from other miscellaneous property lines such as renter, condominium, dwelling fire and landlord policies.

There are some differences in the two datasets. This homeowners database contains $120 million in earned premium and has an overall loss ratio of 64.1%, excluding catastrophe losses. The loss ratio is 79.2% with those catastrophe losses included. The experience period was extended to December 31, 1996 for the policies originally written in 1994, making the experience period 36 months for both policy years. For the majority of the writing period, 1/1/93 through 12/31/94, the company that wrote the policies did not use credit as an underwriting or rating tool. Approximately 10% of the policies were written after such a program was implemented in the underwriting area. During the experience period, all policies inforce were re-underwritten using credit score. While no action was taken directly due to the score, some policies received condition and maintenance reviews and had inspection reports ordered, if such reports were not ordered upon first issuance of the policy. Also, rating territory was included in this database from the outset.

There were striking similarities between the auto and home databases with regard to credit impact on loss experience. The most significant difference seemed to be that derogatory information on a credit report for a homeowners policy had a more severe impact on loss performance (Group A below). If premium and loss are aggregated according to the same Groups A through D as was done with the auto line of business, the results are as follows, with the auto experience displayed again for comparison (premiums are in millions and loss ratios exclude catastrophes for homeowners):

	Homeowners			***Auto***		
Group	Earned Premium	Loss Ratio	Loss Ratio Relativity	Earned Premium	Loss Ratio	Loss Ratio Relativity
A	$ 17.6	111.7%	1.74	$ 74.3	101.4%	1.33
B	41.4	66.5%	1.04	158.9	78.5%	1.03
C	11.9	54.5%	0.85	69.0	69.1%	0.91
D	49.1	47.4%	0.74	91.7	57.4%	0.75
Total	120.0	64.1%		394.0	76.3%	

The similarities between the loss ratio relativities for these profiles lends credence to the assertion that the impact of bill paying history on insured losses transcends line of business, and is not a characteristic attributable only to property policies and claims associated with them. Note that there is a much larger premium distribution in group D for homeowners, the best performing group. This could arise due to a variety of reasons. The same derogatory characteristics that make up Group A are considered in a loan or mortgage application, so a homeowners policy applicant has already (at some point) undergone a screening process based on credit history. The company's underwriting program during the experience period likely decreased the volume of group A policies in the cohort, increasing the proportional amount of Group D.

Individual Credit Variables

The review of individual variables will not be discussed in depth here, as many of the results were parallel with those obtained from the auto study. A handful of examples will be displayed. Compare these with the tables for auto on pages 3 through 5.

Amounts Past Due

APD	Earned Premium	Loss Ratio	Relative Loss Ratio
$0	$ 106.7	58.9%	0.92
$1 - $20	0.9	67.8%	1.06
$21 - $100	2.1	69.2%	1.08
$101-$500	3.5	100.0%	1.56
$501 +	6.8	124.9%	1.95

Collection Records

Number of Collections	Earned Premium	Loss Ratio	Relative Loss Ratio
0	$ 112.0	59.7%	0.93
1	5.2	125.3%	1.95
2+	2.9	124.9%	1.97

Derogatory Public Records

Number of DPRs	Earned Premium	Loss Ratio	Relative Loss Ratio
0	$ 105.4	57.7%	0.90
1	8.0	99.3%	1.55
2	3.0	122.5%	1.91
3+	3.6	125.1%	1.95

Age of Oldest Trade Line

Age in Years	Earned Premium	Loss Ratio	Relative Loss Ratio
< 1	$ 2.3	115.8%	1.81
2 – 3	3.0	68.7%	1.07
4 – 5	5.1	70.9%	1.11
6 – 7	8.3	77.6%	1.21
8-10	19.6	73.8%	1.15
11-15	26.6	60.5%	0.94
16-20	23.6	65.3%	1.02
21+	30.2	48.9%	0.76

Non-Promotional Inquiry Count

Number of Inquiries	Earned Premium	Loss Ratio	Relative Loss Ratio
0	$ 82.2	60.4%	0.94
1	19.5	59.5%	0.93
2	8.1	65.9%	1.03
3	4.1	84.2%	1.31
4-6	4.3	96.8%	1.51
7-10	1.3	106.7%	1.66
11+	0.5	261.2%	4.07

In nearly all characteristics reviewed, it was found that the range of the variable that was correlated with poorer loss experience produced more severe values for the homeowners line than for auto. The linear correlation coefficients for the above tables for loss ratio relativity were 0.95 for APD (0.78 for logarithm of APD versus loss ratio relativity), 0.81 for collection records, -0.74 for age of oldest trade line and 0.93 for non-promotional inquiry count.

Multivariate: Underwriting and Credit Combinations

As with the auto line of business, queries were run to produce premium and loss data for various combinations of risk characteristic and credit characteristic. For purposes of credibility, the credit characteristics were grouped into the same profiles shown above, Groups A through D. A sampling of those results are shown here.

Prior Loss History

At the time of application, an effort is made to determine if there were prior losses filed on the residence. This information arose either from a property CLUE (Comprehensive Loss Underwriting Exchange) report or from the interview with the applicant. Note that the loss ratio across credit levels is not that much different for risks with prior losses compared to those risks with no such prior losses. This is due to a) underwriting practice of the company writing the business and b) relatively less complete information in property CLUE than is present in the auto CLUE system and the state motor vehicle record histories combined.

	Risks with no prior losses			**Risks with at least 1 prior loss**		
Credit Group	Earned Premium	Loss Ratio	Relative Loss Ratio	Earned Premium	Loss Ratio	Relative Loss Ratio
A	$ 15.6	111.2%	1.73	$ 1.9	115.5%	1.80
B	37.7	66.7%	1.04	3.8	64.4%	1.00
C	11.0	56.2%	0.88	1.0	35.3%	0.55
D	43.6	45.7%	0.71	5.5	61.0%	0.95
Total	$ 107.9	63.6%	0.99	$ 12.2	68.6%	1.07

Town Class or Protection Class

Loss experience in the form of loss ratio relativities for credit groups A through D are evaluated within the various protection class designations and is shown below. Values are not shown for cells that possess a premium volume below $500,000.

Protection Class	Credit Profile Group A	B	C	D	Total
1	1.30	0.68		0.65	0.77
2	1.63	1.06	0.84	0.66	1.00
3	2.15	1.20	0.92	0.77	1.14
4	1.61	1.03	0.93	0.71	0.97
5	1.95	0.92	0.72	0.83	1.00
6	1.48	0.88	0.55	0.79	0.90
7		0.63	0.42	0.42	0.79
8		0.67		1.26	1.31
9		1.72		0.48	0.97
10					
Total	1.74	1.04	0.85	0.74	1.00

There is much more fluctuation for individual cells for this dataset compared to the auto line due to both the overall smaller premium volume and the greater volatility of homeowners losses. The consistency across the profile groups is still quite evident for various protection classes, and the relativities decrease monotonically wherever there is significant premium volume in the cells.

Liability Limits

During the two-year period of policy writing, the company wrote an approximately equal proportion of $100,000 and $300,000 liability limits on homeowners policies. A much smaller volume of premium was written with other limits of liability. The base premium was set based on the former limit, and the latter was offered as additional optional coverage.

	Liability Limit = $100,000			**Liability Limit = $300,000**		
Credit Profile Group	Earned Premium	Loss Ratio	Relative Loss Ratio	Earned Premium	Loss Ratio	Relative Loss Ratio
A	$ 9.7	115.5%	1.80	$ 6.4	100.3%	1.56
B	20.4	63.3%	0.99	17.5	70.4%	1.10
C	5.7	59.4%	0.93	5.2	48.4%	0.75
D	21.1	50.9%	0.79	23.2	43.7%	0.68
Total	$ 56.9	67.2%	1.05	$ 52.2	60.1%	0.94

Note the steady shift in distribution of premium between the two limits by group. The premium distribution of the $100,000 limit for the four groups (A through D) is 60%, 54%, 52% and 48%, respectively. Risks with poorer bill paying histories are more likely to choose the lower liability limit, even though the cost of this additional coverage was less than $10 in most cases.

Bill Mode

The two most common forms of payment of homeowners insurance premiums are direct bill, in which the policyholder pays the premium directly, or mortgagee bill, where the financial institution which holds the note on the property pays the premium.

	Direct Bill			**Mortgagee Bill**		
Credit Profile Group	Earned Premium	Loss Ratio	Relative Loss Ratio	Earned Premium	Loss Ratio	Relative Loss Ratio
A	$ 7.7	117.2%	1.83	$ 7.8	103.9%	1.62
B	19.9	69.5%	1.08	17.4	62.6%	0.98
C	5.3	56.7%	0.88	5.5	54.3%	0.85
D	27.7	46.8%	0.73	15.2	48.2%	0.75
Total	$ 60.6	64.0%	1.00	$ 46.0	63.9%	1.00

Rating Territory

As with the auto line, premiums and losses were aggregated by rating territory by assigning characteristic definitions to each rating territory, designating each territory as urban, suburban or rural. This designation was done by eye, without any objective definition of urban (such as population density); major urban areas were designated as such, satellite territories around urban areas and smaller population centers were referred to as suburban, and the remaining regions were called rural. Although there was little credibility when this data was reviewed at the state level, there was sufficient volume when premiums were accumulated by territory type across states. The credit-defined groups showed consistent impact on losses within each group, and there were only slight distributional differences. Only the largest 12 states were included in this query; these states made up roughly two-thirds of the premium volume of the entire sample.

	Urban		**Suburban**		**Rural**	
Credit Profile Group	Earned Premium	Relative Loss Ratio	Earned Premium	Relative Loss Ratio	Earned Premium	Relative Loss Ratio
A	$ 2.8	1.23	$ 7.3	1.99	$ 2.3	1.31
B	7.0	1.07	18.6	1.02	5.1	1.14
C	1.6	0.96	5.7	0.91	1.3	0.57
D	5.5	0.64	23.4	0.80	6.8	0.66
Total	$ 16.9	0.95	$ 54.9	1.04	$ 15.5	0.91

Motility

In order to understand the migration of risks from one credit profile to another over time, additional data was added to the homeowners database. Credit files from future dates were included, which were taken from archived records approximately 12 months after original writing date, and again at 48 months after the original writing date. For this discussion, the same four credit profiles will be used as in the above exhibits.

Group A, the poorest performing profile, was populated with 10,737 policies written in 1993. Of these, 84% still had Group A characteristics 12 months later, and 66% of those risks were still categorized as Group A 48 months later. 20% had migrated to Group B,and the remaining 14% to C and D. This is not surprising, given that 2 of the 3 criteria for Group A are maintained for many years on the credit file (derogatory public records and collections).

Group B was not as stable over time, significant portions of the population migrated in both directions. Of the original Group B in 1993, 67% were still in the group 12 months later, and 36% 48 months later. At that time, 31% had moved to D, 12% to group C, and 21% to A.

Group C was the least stable. Since this group is defined by better than average characteristics, it is not surprising that as those characteristics continue to improve, much of the distribution migrates to Group D. Only 50% of the group still had the Group C characteristics 12 months later, and only 11% at 48 months. 65% of the entire group migrated to Group D in four years. This is not surprising due to the fact that one of the differences between C and D is age of oldest trade line; for those risks that did not qualify as D, time can be the only factor necessary to cause a migration over the subsequent 3 year period. (Again, refer to the Appendix for exact Group definitions.)

Group D, the best performing group, showed the most stability. Risks with the best credit profiles are more likely to maintain those profiles over time. Of the 23,248 policies in this group, 87% still met the criteria for D 12 months later, and 78% met those criteria 48 months later.

This data was not collected on the original auto cohort, so the above data is for homeowners only. It does provide some indication about the necessity of updating the review of credit profile for the purpose of rating and/or underwriting.

Implications and Other Related Issues

The impact of credit history on expected loss performance is a major factor influencing whether or not this variable should be utilized in the rating of personal lines insurance premiums. There are, however, many other relevant issues that must be considered.

The credit history contains a large amount of data. The impact on loss performance has been measured in this study as if arising from a single variable, which is one particular accumulation of the credit data. There is of course an enormously large number of ways in which the data can be combined for this purpose of measurement. When the variables are inspected, individually, one finds that there are some that are historic, and cannot change until they are purged from the record (i.e., derogatory public records, collection records, inquiries and delinquent payments). Others contain information about current conditions, such as account status, current balances and limits, and overdue amounts. The method of combination of these variables will determine where the model falls in the responsiveness versus stability spectrum. This study has shown that both types have strong influences on loss performance. How they are combined is currently an open field for individual insurers' discretion. This study utilized a mutually exclusive profiling technique; scoring models can and do utilize a large number of variables, giving numeric weights to each individual characteristic which are then added to obtain a total. Either method can be accomplished using a wide range of variable counts.

An important gap in this study is the impact of credit history on loss performance for customers who have been insured with the same company for a number of years. Recall that the data was assembled from new policies written in a give policy year, and the subsequent three-year loss experience. This data cannot show if long-term customers who have similar credit characteristics are expected to have the same differences in loss performance. The creation of a rating factor based on credit history can affect renewal customers as well as new customers, yet there is currently no data publicly available to my knowledge that shows such relationships. Without such data, it would be speculative at best to assume that the relationships hold true regardless of tenure. Studies have shown that long-tenured customers produce far better loss experience than new customers. Opinions vary as to whether this is due more or less to two (or more) dominant factors which can cause such improvement: 1) the fact that longer term customers have more experience in operating a motor vehicle or maintaining a home, and 2) that the underwriting function of a given company will selectively non-renew poor performing risks, which could not be identified accurately in the underwriting process when the policies were originally written. The research done with this data has shown that longer-tenured customers tend to have better credit profiles than newer customers. This is one variable, policy tenure, that could be both distributionally and loss performance-linked to credit history.

The question as to how often the credit history needs to be reevaluated is also of concern. Although the motility information above indicates that there is a fair amount of stability over time for credit conditions, there is still significant change that occurs within such distributions. Each reevaluation will cause the creation of an additional inquiry record on the file. Although such inquiries should not be utilized for evaluation, there is no guarantee that all financial institutions and other users of credit data will ignore their existence. When such a reevaluation occurs, there is also the question as to which risks should experience premium adjustment. Is there reasonable justification for an individual risk to experience an increase in premium solely due to a change in a variable within the credit file? A different type of database construction technique would be required to answer such a question.

From an actuarial standpoint, questions arise concerning the nature of the variable. The literature is replete with admonitions concerning the use of variables that are, or can be, under the control of the insured. Although the historic variables are not under the control of the insured, certainly those that measure current conditions are. Worth considering, however, is the argument that such control is not nearly as relevant as other rating factors that are not utilized for this reason. An individual who has a poor history of timely bill payment, and is under a considerable debt load is already experiencing detrimental effects from these conditions. Such conditions are causing economic penalties in the form of monthly interest payment, or debt service, and can also result in higher interest rates charged for credit lines, installment loans and mortgage loans. There already exists a financial disincentive to maintain financial management habits that produce these conditions. Will a difference in auto or homeowners insurance premiums cause a change in such habits, where these other economic disincentives have not? It is likely, in my opinion, that the magnitude of the premium difference would not be as large as the sum of all other financial consequences of such a credit profile in most cases. This may mitigate the concern over the control the risk appears to have over the data contained in the credit file.

Another area of concern that is related to variable control is data accuracy. Reports as to the accuracy of credit history data vary widely depending upon the source. Credit bureau sources quote data accuracy values in the 99% to 100% range. Some consumer groups have quoted this number to be as low as 30% to 40%. This discrepancy is due to the way in which errors are measured. One could obtain the first result if errors were considered to exist only in cases where a) an adverse decision was made for a financial transaction, b) the customer inquired as to the credit data, c) discovered an error, d) contacted the creditor to correct the error, and e) the financial institution reversed the decision based on that correction. Dividing the number of such events by the entire credit warehouse would produce a very high level of accuracy. To produce the second, much lower values, one could simply count every possible error within the file, including seemingly irrelevant errors such as street name misspellings, and divide this count by the total number of records. Neither is a very good measure of data accuracy. For all parties concerned to get a true understanding of accuracy, a good method of measurement must be established. In any case, the utilization of credit history for rating requires the insurance industry to assist its customers by informing them of the method for resolving true inaccuracies on record, and taking those corrections into account through reevaluation.

An outstanding issue that will likely remain outstanding is causality. Although arguments were put forward earlier in this paper which attempted to link financial management responsibility and future expected loss levels, such arguments are unsupported, even if reasonable, speculation. The arguments of causality are generalized; in fact the difference between one rate level and another charged to a given individual could be different due to only one particular variable within the credit file. That individual may ask for an argument of causality pertaining only to the one characteristic that separates him or her from the next lower rate. Such questions may never be answered with statistical causality, even if the entire credit file (however that is aggregated) can be demonstrated to be causal in a way that goes beyond the mathematical correlations.

The issue of acceptance of credit history data in personal lines insurance has more obstacles than mere causality. The social and regulatory acceptance of such data in the rating of personal lines insurance may be restricted for other reasons. Arguments have already been made that indicate that some groups consider its use invasive, and that credit-based rating is a breach of privacy, regardless of its strength as a tool to reduce rate subsidies between risks. The auto line of business has considered past driving record to be a key factor in underwriting and rating. One key characteristic of prior accidents is negligence, i.e., whether the accident was the fault of the insured or not. It is natural for some people to immediately apply this concept to credit history as well. Credit files contain information about derogatory events that an individual may feel are perfectly explainable. Such explanations are commonplace in the area of mortgage financing, where an event is not considered if there is a suitable explanation for its existence in some cases. The key difference, however, is that the use of this data for rating or underwriting is not done for the purpose of credit worthiness. It is not done for the purposes of judging character, lifestyle, integrity or financial soundness. The purpose is to segregate risks by different levels of expected losses only, a point which may be difficult to communicate.

It may be easier to obtain regulatory acceptance comparted to social acceptance with regard to the use of credit history as a rating tool. The NAIC White Paper on the use of credit in underwriting, referred to earlier, makes several specific statements which indicate their deference to rating, rather than underwriting. The use of credit in rating requires the filing of a rating plan with supporting documentation. It permits inspection of content by both regulators and consumers. Such filing gives a regulatory body the evidence required to give valid statistical response to constituents who may call to inquire or register a complaint.

The data reviewed in this study produced clear evidence of a strong correlation between credit history and future loss performance. The understanding of this relationship, and its acceptance, have grown rapidly over the last few years. This understanding has come primarily in the form of scoring model results. Hopefully, this paper will serve as a starting point in an effort to place more detailed information from credit history, other than scoring models, and the relationship such data has to personal lines losses, in a public forum. This effort is necessary in order to promote greater understanding of the driving forces behind this relationship, and can only serve to improve the quality of discussion during future debates on the ways in which it will be utilized.

APPENDIX

1. Data Fields

Policy Variables included and reviewed:
State transfer indicator
Policy Tier
Original policy written month, day and year
Acive status indicator
Months of coverage
Writing company
Original producer code
Risk state
Vehicle type
Non-standard indicator
Number of vehicles
Number of operators
Number of potential operators
Payment plan
Residence stability
Residence code
Residence type
Number of years employed
Prior insurance code
Number of vehicles financed
For each driver:
  Age
  Gender
  Marital status
  Occupation code
  Number of years licensed
  Driving record: fault losses, non-fault losses, moving violations
  Comprehensive losses
Earned premium
Incurred losses

Variables included from National Credit File:
Trade Record: Subscriber code, date opened, high credit, date verified, date reported, date closed, date paid out, associated code, payment pattern, current balance, amount past due, account type, current manner of payment (status), credit limit, terms, maximum delinquency date, maximum delinquency amount, number of months 30-59 days past due, 60-89 days past due, 90+ days past due, loan type, dispute code, collateral field, duplicate indicator, account number, short subscriber name.
Inquiry Record: Subscriber code, inquiry date, type, loan type, loan amount.
Public Record: Date reported, amount, public record type, date paid, assets, liabilities, attorney, plaintiff, docket number.
Collection Record: Date reported, subscriber code, amount owed, status, date paid, creditor name.
Summary Record: Number of inquiries, trades, collections, public records, manner of payment totals for each status code.

2. Definitions of Credit Profiles Used in Exhibits

Group A: Existance of any of the following: Derogatory public record with liability amount >$0, collection record, or amount past due of $500 or more.

Group B: Does not meet any other group criteria.

Group C: No DPR or collection records, no APD; no trade lines with status codes other than 0 or 1, leverage ratio on revolving accounts less than 60%, age of oldest trade line at least 7 years.

Group D: Same as group C, plus nonpromotional inquiry count less than 4 and age of oldest trade line at least 10 years.