Credit-Based Insurance Score –
Homeowners Insurance – PO44804

Comments Regarding Proposed Order to File
A Special Report of Homeowners Insurance Data

by Michael J. Miller, FCAS, MAAA
EPIC Consulting, LLC

On Behalf of the National Association of Mutual Insurance Companies

June 18, 2008

Comments Regarding
the Federal Trade Commission's ("FTC") Proposed Special Report
of Data for the Study of Credit-Based Insurance Scores
as Applied to Homeowners Insurance

FTC Matter No. PO44804

Introduction

My name is Michael J. Miller. I am a consulting actuary with EPIC Consulting, LLC. My business address is Carlock, Illinois.

I am a Fellow of the Casualty Actuarial Society (CAS) and a member of the American Academy of Actuaries. I have served the CAS as a director (two terms), as Vice-President – Research, and as chairperson of several professional committees. My work as a consulting actuary has included the creation of many large research databases. I have consulted with two statistical agencies for the insurance industry on better ways to edit data received from insurers prior to its consolidation. I am a co-author with Richard Smith of a June 2003 study entitled "The Relationship of Credit-Based Insurance Scores to Private Passenger Automobile Insurance Loss Propensity". The Miller/Smith study was based on a large database which we created by consolidating data from multiple data providers. The comments offered in this report are based on my experience in creating and editing large research databases.

A more complete description of my background and professional experience is provided in the attached curriculum vitae.

Standards for Judging Proposed Data Call

The FTC has created an impossible task for itself. The sheer volume of data being demanded will make it impossible to understand, edit, reconcile and consolidate the data into a meaningful research database. If the FTC would simplify the data demand and provide precise definitions as to each data element to be provided, there is a good chance that a usable research database could be constructed.

The creation of an accurate database is the most important aspect of any statistical analysis. It is meaningless to apply sophisticated statistical analysis techniques to incomplete, inaccurate, or irrelevant data.

The cost of the data is also an important consideration in the design of a study's database. It is costly for the data providers to extract the requested data from their computer systems. It is also costly for the data recipient to properly edit the data for reasonable accuracy. Without an effective editing protocol, the entire study will be unreliable. The cost and difficulty of editing increases exponentially as the number of requested data elements increases.

The costs and difficulties in creating a reliable study database also increase substantially when the database is consolidated data from multiple providers, rather than data produced by a single provider. Data from multiple sources requires very precise definitions so as to guarantee that the data elements received are identical. Data from multiple sources also requires that the editing protocol be applied separately to the data received from each provider before the data are consolidated.

Because of the significant cost and data accuracy considerations, a researcher should limit the data request to only those data elements which are necessary to complete the predetermined statistical analysis. It would be fiscally irresponsible to do otherwise.

Based on these cost and data accuracy considerations, the following standards have been applied in our review of the FTC's data subpoena.

1. Has each data element been defined with sufficient precision so that the data received from each provider can be properly edited and consolidated?

2. Is each data element essential to the statistical analysis to be undertaken?

Purpose of the Study

The first step in the creation of any research database is the design of the statistical analysis. Only after the calculations to be undertaken have been decided upon, can the researcher then identify precisely which data elements will be necessary in order to complete the calculations.

The FTC has requested public comments on the data elements to be included in the Special Reports, but has not provided a description of the statistical analyses it intends to undertake. Without a clear description of the statistical analyses which the FTC intends to undertake, it is impossible to fully judge the necessity for each of the data elements being demanded.

We acknowledge that the FTC does say that it intends to "prepare a study regarding the use and effect of credit-based insurance scores on homeowners' insurance". Unfortunately, the FTC's description of its study is vague and lacks the precision necessary to properly determine which data elements being demanded are essential to the calculations.

In our review of the demanded data elements we have assumed that the FTC's statistical analyses of the homeowners data will be limited to, and exactly replicate, the statistical analyses undertaken by the FTC for its July 2007 study of automobile insurance. If the FTC intends to undertake new and/or additional statistical analyses of the homeowners data, as compared to its automobile study, our conclusions and recommendations in this report could be materially impacted and subject to change.

Specific Comments

1. *Volume of Data Demanded* – We understand that the subpeoenae for data are being directed to nine insurer groups. Because some of the individual insurers within these groups will be individually incorporated and may have their own computer systems and records, data will be produced by more than nine data providers. We believe the FTC's

implied suggestion that it will be receiving data from "only" nine data providers is misleading. Consolidating and editing data from nine or more data providers will be a daunting and costly task for the FTC. The difficulties and costs associated with creating a reliable, consolidated database increases exponentially as the number of data providers increases.

What is clear from the Specifications is that the data being demanded are 100% of each data provider's policyholder records for the Relevant Time Period. The number of data lines received by the FTC from this 100% sample will likely amount to 400 million records or more. The question is not whether a computer exists that is big enough and fast enough to handle 400 million data records. A critical question is whether or not the FTC has the resources to properly edit and consolidate 400 million data records. A sample significantly smaller than the 100% sample of three years of policy records could be constructed so as to provide a more workable, and likely more accurate, consolidated research database.

Further, the 100% sample proposed by the FTC will most likely produce a research database that does not represent the distribution of insured homes throughout the 50 States. For instance, the largest insurer of homes in Florida will apparently not be a data provider. This means that Florida will likely be under-represented in the research database, especially with regard to homes on or near the coasts. We recommend that the sample be constructed so as to produce a research database which represents a cross-section of the insured homes throughout the United States.

2. *Specification 3(d), Policy Data* – This data element is unnecessary. Whether or not the policy is new or a renewal can be determined by comparing the inception date (3i) to the start date (3b).

3. *Specifications 3(e) and 3(f), Cause of Cancellation or Non-Renewal* – It is unlikely that an insurer's computerized premium records will retain information regarding the cause for non-renewal or cancellation. To the extent such information is in the premium records, the information will be captured with a code that is unique to the insurer. We doubt that much, if any, of the data produced in response to 3(e) and 3(f) can be

consolidated into a research database. None of these data are essential to the statistical analysis.

4. *Specification 3(h), Name of Entity* – The data demand appears to be constructed for a specific insurance company within a company group. There is no need to re-identify the insurer of the policy. Perhaps the problem with 3(h) is one of definition. Why is the term "company or entity" used in 3(h), whereas the capitalized and defined term of "Company" is used elsewhere in the data call? Is the FTC seeking information in 3(h) about some non-insurance company entity, such as an agency? If so, there would be no need for such data for statistical research.

5. *Specification 4(a) and 4(b), Premium Data* – Homeowners insurance is a package coverage and the premium is usually indivisible by component coverage. There is no need to obtain premium data by "coverage type". If the FTC were to seek premium data by coverage type, there would need to be a precise and uniform definition of each "type" of coverage and a description as to how the premium is to be estimated. As it stands, it is unlikely that any premium data received by coverage type could be consolidated into the research database because the data providers will apply unique definitions as to what the term "coverage type" means.

   Further, 4(a) and 4(b) refer to premiums "net of any rebates, refunds, dividends, etc." No definition is provided so that the data providers know what the FTC intends to be reported as a rebate or refund. Also, a research data call should never use the term "etc". The lack of definition of these terms will likely produce data that are dissimilar and cannot be consolidated into a research database.

   It is also unclear as to the difference, if any, between the premiums requested in 4(a) and the premiums requested in 4(b). The data providers will be insurance professionals accustomed to insurance technology. The terminology describing the premium in 4(b) is vague.

6. *Specification 4(d), Miscellaneous Fees* – This data demand is vague and will result in data that cannot be combined into a consolidated research database. Further, this

information is just as irrelevant and unnecessary for the statistical study of homeowners insurance as it would have been for the July 2007 FTC study of auto insurance.

7. *Specification 5(a), Policy Form* – The subpoena specifically excludes data from homeowners policies which cover renters, condominiums, and mobile/ manufactured homes. With these exclusions in place there is no analytical purpose for obtaining a further breakdown of the data by policy form.

   It is unlikely that the multiple data providers use a uniform set of policy form codes or identifiers. As such, it is likely that the FTC will receive policy form data that it cannot accurately consolidate into the research database. If the FTC were to decide that it needed policy form data, it needs to specifically define the categories of policy forms it seeks and then rely on the data providers to accurately slot the data into the prescribed policy form categories. Terminology such as "including but not limited to" is open-ended and an unworkable approach for creating a research database.

   If the FTC desires to statistically test the hypothesis that the type of coverage selected by the policyholder is a predictor of risk, we recommend that 5(d) be revised to simply identify the policy as either a "replacement cost policy" or "all other".

8. *Specifications 5(d), 5(e), and 5(f), Coverage Detail* – None of these data elements are necessary for the statistical analysis being contemplated. Even if these data were relevant to the statistical analysis, the data being demanded could not be consolidated in a meaningful way into a combined research database.

9. *Specification 5(g), Deductibles* – Deductible information is necessary for the statistical analysis being contemplated. However, the deductible information being demanded in 5(g) is not well-defined and will produce data that cannot be accurately consolidated into a combined research database. We recommend that this data demand be revised to reflect a list of the most popular deductibles and that the list include an "all other" category.

10. *Specifications 5(i) and 5(j), Market Value* – These data are irrelevant to any statistical study concerning the prediction of risk. Further, at what point in time should the market

value of the home be determined? How are insurers to determine market value if the house has not recently been sold? Insurers have no need to determine the market value of the home at the time of renewing the policy. It is likely that any market value in the records of an insurer is out of date as of the renewal date of the policy. It is highly unlikely that any insurer captures the value of the building lot. Such information is neither relevant to the rating of homeowners insurance, nor could the value be determined accurately.

11. *Specification (6), Insurance Score* – None of the data demanded in Specification 6 is necessary for a statistical study regarding the predictive power of credit-based insurance scores or the degree to which the scores are a proxy for protected minority groups. None of the data can be consolidated in a meaningful way into a research database. Also, there will be no data which could be used for comparison purposes, thereby making the data in Specification (6) useless for statistical analysis purposes. For example, assume that a credit-based insurance score was used in the determination of a rate tier. Of what value is that information unless it is also known what the tier would have been had an insurance score not been used?

    Further, request 6(e) is meaningless. The only way to determine the rate impact of the use of a credit-based insurance score is to compare the rate charged when the score is used to what the rate would have been if no score was used to underwrite or rate homeowners insurance. In order to respond to 6(e), a data provider would need to carry two, very different rate schedules in its computer system, or it would need to calculate what the alternate rate would have been for every policy at every renewal date. Accurate information to respond to 6(e) is simply not available and cannot be estimated by the FTC with the data it is demanding.

12. *Specifications 7(a), 7(b), 7(d), and 7(f), Risk Data* – There will be little, if any, uniformity in territory definitions, tier definitions, or protection class between the various data providers. There also is no uniformity in the class codes used by the data providers. The information in 7(a), 7(b), 7(d) and 7(f) will be impossible to consolidate into a combined research database. Even if the FTC were to receive a definition of the class codes used by each data provider, it could not accurately consolidate this data into a research database.

Further, the rate classes (i.e., territory, tier, and protection class) are not combinable for any data provider across state lines. Territories, tiers, and protection classes vary by state for each insurer.

There is a critical need for the FTC to be able to control its statistical analysis for the portion of the homeowners risk that varies geographically. Control for geographic risk was also necessary for the FTC's 2007 study of auto insurance. The FTC's inability to adequately control for geographic risk on all of the auto coverages, except possibly for property damage liability, was the single largest weakness of the FTC's 2007 auto study. Control of geographic risk for homeowners insurance will be even more difficult than it was for the auto study because the geographic risk varies differently for each peril covered by homeowners insurance. The data demanded in 7(a), 7(b), 7(d), and 7(f) afford absolutely no usable information with respect to the geographic risk of homeowners insurance.

13. *Specifications 7(c) and 7(e), Premium Adjustments* – These two data demands lack sufficient definition to even venture a guess as to what data the FTC is demanding.

14. *Specification 7(g) and 7(l), Construction and Renovation* – These demands are examples of how a call for research data should not be constructed. A data call should never be based on open-ended "examples". Precise categories of construction type and renovation type should be provided. Otherwise, the data received from multiple data providers runs a good chance of not being combinable.

15. *Specification 7(m), Condition of Dwelling* – This data demand is so vague as to be meaningless. If any data is reported for this item, it will be impossible to edit for reasonableness and it will be impossible to combine in any meaningful way with data from another data provider.

16. *Specifications 7(r) through 7(bb), Risk Factors* – These demands for data are worded in such a way that the FTC will receive no useful information. Requests for "values" or "codes" will not produce information that can be combined into a consolidated research database. How could a researcher combine income information from just one person in

the family with total family income reported on another policy? How will a researcher know whether the income being reported is for one person or a family total? How will a researcher handle a policy for which no break-in-coverage information, or no income information, or no home equity line of credit information is available to be reported? How will those policy records be combined with policy records where such information is available?

Precisely defined categories of breaks in coverage, prior claims experience, person and household characteristics, income, multi-line discounts, group discounts, security systems, and mortgage information must be provided if the FTC has any hope of receiving usable research data. All of the open-ended, vague data demands will most assuredly produce information that the FTC will ultimately be forced to scrap as being unusable.

17. *Specification 8*, *Endorsements* – There is typically a wide array of endorsements available for use with a homeowners insurance policy. There is little uniformity between insurers with regard to endorsements. Some insurers may include a particular coverage in its policy, another insurer may provide the coverage as a mandatory endorsement, and yet another insurer may provide the coverage as an optional endorsement.

    Most of the information demanded in Specification 8 will not be combinable into a consolidated research database. Little, if any, of the information in Specification 8 is essential to the central issue being researched by the FTC.

    To simplify the database, reduce costs, and eliminate the likelihood of the FTC receiving a high volume of unusable information, we recommend the entire elimination of Specification 8.

18. *Specification 9 (c)(iii), Payment Data* – Partial claim payments may occur on several dates. Is it the intent of this request to obtain multiple payment dates?

19. *Specification 9(d), Claim Status* – This request is too vague. Open or closed should not be provided as an "example" of claim status. If it desired to know whether the claim is open or closed, then the request should so specify. Is there any other claim status in

which the FTC is interested? If so, the data demand should precisely specify the categories of claim status the FTC is seeking.

20. *Specification 9(j), Type of Loss* – Well-defined categories by type of loss should be provided. It is likely that the information received from this data demand, as it is currently written will vary among the data providers, making it impossible to combine the data.

21. *Specification 9(k), Catastrophe* – There is no uniformity among insurers as to the identification of "catastrophe" losses. As currently written, this data demand will provide no useful information that can be combined into a consolidated research database. Catastrophe losses are typically weather-related. If the data demand for type of loss is properly constructed, 9(k) becomes completely unnecessary.

22. *Specification 9(l), Coverage* – The term "coverage" is not defined and as such there will be no uniformity in the data produced by multiple data providers. If the type of loss category is defined properly in 9(j), Specification 9(l) becomes unnecessary.

23. *Specifications 10, 11, 12, 13 and 14* – None of the information in these five Specifications can be edited for reasonableness or combined into a consolidated research database. None of this information is necessary for the conduct of a statistical analysis similar to that which the FTC conducted for auto insurance in 2007.

    Further, Specification 11 is an admission by the FTC that it fully expects to receive non-uniform data from the multiple data providers. What will happen is that each data provider will do its best to respond to the poorly worded and vague data demands. Then the FTC will be forced to use the documents demanded in Specification 11 in a futile attempt to understand and manipulate the information it has received so that the data can be combined with the data from other data providers. The likelihood that the FTC would be able to accurately combine all of this data into a single database is zero.

    Receiving dissimilar data from multiple providers, and then trying to interpret rate manuals and other documents in hopes of understanding the data that has been provided, is an impossible task for the FTC, as it would be for any researcher. The

sheer volume of data that will be provided will make it impossible for the FTC to decipher, edit, reconcile and consolidate the data into a meaningful research database. The data demands need to be reduced in size and simplified with precise definitions for each data element.

Recommendation and Conclusions

We find that the volume of data records being demanded is more than what is necessary for the conduct of a reliable statistical analysis and that the data sample has not been constructed so as to produce a representative cross-section of homeowners insurance policies across all 50 States.

We find that the data elements being demanded are generally not well-defined and are likely to result in data from multiple providers that cannot be edited and checked for reasonableness and cannot be combined into a reliable, consolidated research database.

We find that most of the data and information being demanded are unnecessary for the conduct of a statistical analysis that would be similar to the analysis conducted in 2007 by the FTC for auto insurance.

We recommend that the FTC first approach the study by first developing a study protocol which defines the type of statistical analysis and the mathematical calculations which it intends to undertake. Once the statistical analysis and calculations are defined, a revised data demand should be restricted solely to the data elements necessary to complete the calculations. The revised data demand should include precise definitions for each data element so that the data received from multiple data providers will be similar; can be subjected to tests of reasonableness; and can be combined into a reliable, consolidated research database. If the FTC adopts this recommendation there is a good chance that a usable research database can be constructed within the bounds of a reasonable FTC budget and the study can be completed in a reasonable time period.

As currently written, the data subpoena will not produce data which the FTDC can use to satisfy its legislative charge. We recommend that the FTC withdraw the proposed data demand.

Date: June 18, 2008

Michael J. Miller, FCAS, MAAA

CURRICULUM VITAE

NAME: Michael J. Miller

BUSINESS ADDRESS: 21253 N 825 East Road
Carlock, IL 61725
E-Mail: mmiller@ask-epic.com

EDUCATION:

ILLINOIS STATE UNIVERSITY
Bachelor of Science – 1968
Major – Mathematics
Minor – Accounting

CONTINUING EDUCATION:

Estimated study time exceeding 3,000 hours necessary for completion of 10 qualifying exams for membership in Casualty Actuarial Society (CAS).

Participation as an attendee and on the faculty of the CAS Loss Reserve Seminar, the CAS Ratemaking Seminar, and other CAS educational seminars on special topics, such as rate of return and underwriting practices.

Meet all continuing education requirements of the American Academy of Actuaries necessary to sign a public actuarial opinion.

MEMBERSHIP IN PROFESSIONAL ORGANIZATIONS:

Casualty Actuarial Society (CAS)	
Associate Member	1971
Fellow	1981
American Academy of Actuaries (AAA)	1975
Conference of Consulting Actuaries	2002-2004
Fellow	
International Actuarial Association	
Midwestern Actuarial Forum	
Chartered Life Underwriter (CLU)	

PROFESSIONAL ACTIVITIES:

CAS Committee on Risk Classification,	
Member	1982-1984
Chairman	1983-1984
CAS Committee on Principles of Ratemaking	
Member	1985-1987
Chairman	1991-1992
CAS Examination Consultant	1987-1990
CAS Long-Range Planning Committee	1993-1994 1997-2000
CAS Board of Directors	1992-1993 2001-2003
CAS Officer,	
Vice President – Research and Development	1993-1996
CAS Task Force on Non-Traditional Practice Areas	
Chairman	1998-2000
CAS/SOA Joint Task Force on Financial Engineers	1998-2001
AAA, Liaison Committee to the National	
Association of Insurance Commissioners	1985-1988
Actuarial Education and Research Fund	
Board of Directors	1994-1996
AAA, Casualty Practice Council	1990-1993
Property Casualty Committee of Actuarial	
Standards Board, Member	1987-1993
Chairman of Ratemaking Subcommittee	1987-1988
Chairman of Property/Casualty Committee	1989-1993
Midwestern Actuarial Forum	
Education Officer	1986-1987
President	1988

EMPLOYMENT HISTORY:

State Farm Insurance	1967-1984
M. J. Miller and Company	1984
Tillinghast	1984-1993
Miller, Herbers, Lehmann, & Associates, Inc.	1994-2002
EPIC Consulting, LLC	2003-Present

PROFESSIONAL PUBLICATIONS:

"Private Passenger Automobile Insurance Ratemaking", Proceedings of CAS, Volume LXVI.

"Review – Risk Classification Standards by Walters", Proceedings of CAS, Volume LXVIII.

"A History of the Rating and Regulation of Personal Car Insurance in the United States", The Institute of Actuaries of Australia, February, 1990.

"An Evaluation of Surplus Allocation Methods Underlying Risk Based Capital Applications", CAS Discussion Paper Program, Volume I, 1992.

"How to Successfully Manage the Pricing Decision Process", CAS Discussion Paper Program, 1993.

"Building a Public Access PC-Based DFA Model", CAS Forum, Summer 1997, Volume 2.

"Auto Choice: Whose Fault Is It Anyway", Contingencies, January/February 1998

"Actuarial Implications of Texas Tort Reform", CAS Forum, Spring 1998.

"The Relationship of Credit-Based Insurance Scores to Private Passenger Automobile Insurance Loss Propensity", June 2003.

PRESENTATIONS:

Faculty member on National Association of Insurance Commissioners' orientation program for new insurance commissioners, 1987-1994.

Faculty member on National Association of Independent Insurers' seminars on ratemaking and loss reserving.

"Key Provision in Rate Filings", Society of State Filers.

Numerous presentations at educational seminars and meetings conducted by the Casualty Actuarial Society on topics including ratemaking, loss reserving, underwriting, risk classification and rate of return.

EXPERT TESTIMONY: Rate Regulatory Hearings in Alberta, California, Florida, Georgia, Louisiana, Maryland, Massachusetts, Michigan, Mississippi, New Brunswick, New Jersey, New York, North Carolina, Ohio, Oklahoma, Ontario, Pennsylvania, Texas, Vermont, West Virginia, and Wyoming.

Courts in Alabama, California, Florida, Minnesota, Mississippi, New Hampshire, Pennsylvania.