LOSS RATIO MODELING

FOR

BUSINESS OWNER'S POLICY

11TH APRIL, 2012

PRESENTATION OUTLINE

- Introduction
- Goals and Objective
- Exploratory Analysis
- Data Preparation
- Modeling Process
  - Model Design
  - Model Fitting
  - Testing and Adjusting Model
- Key Results
- Conclusions
- Acknowledgements

INTRODUCTION

Background on Project:

- Worcester Polytechnic Institute Final Project
- Past usage of credit scores in Personal Lines
- Recent shift to usage in Commercial Lines

Business Issue:

How can credit scores be used to improve the predictive ability of the current Hanover Business Owner's Policy model?

Proposed Solution:

Model that implements a credit score variable to predict loss ratio for each policy.

PRESENTATION OUTLINE

- Introduction
- Goals and Objective
- Exploratory Analysis
- Data Preparation
- Modeling Process
  - Model Design
  - Model Fitting
  - Testing and Adjusting Model
- Key Results
- Conclusions
- Acknowledgements

PROJECT GOALS AND OBJECTIVES

1. Identify a base set of risk factors for Business Owner's Policy.

2. Use these risk factors to calculate predicted loss ratios for each policy.

3. Bucket predicted loss ratios into categories according to level of risk.

PRESENTATION OUTLINE

- Introduction
- Goals and Objective
- Exploratory Analysis
- Data Preparation
- Modeling Process
  - Model Design
  - Model Fitting
  - Testing and Adjusting Model
- Key Results
- Conclusions
- Acknowledgements

EXPLORATORY ANALYSIS

We outlined <u>three</u> primary goals:

1. **Familiarize ourselves with the data set and composition of various data fields.**
   - Graphical representations of make-up of data variables.
   - Statistical analysis on data variables

2. **Observe relationships between variables different data trends and relationships amongst variables.**
   - Univariate analysis by plotting variables against loss ratios.
   - Correlation among credit variables

3. **Identify changes necessary to improve the data set.**
   - Identification of invalid, missing or inconsistent data
   - Loss ratio analysis to identify outliers.

EXPLORATORY ANALYSIS

There were 299, 441 policies over the period 2006-2011

Numeric Data

- Histograms for numeric variables
- Scatter plots of loss ratio vs. variables.
- Box and Whisker plots

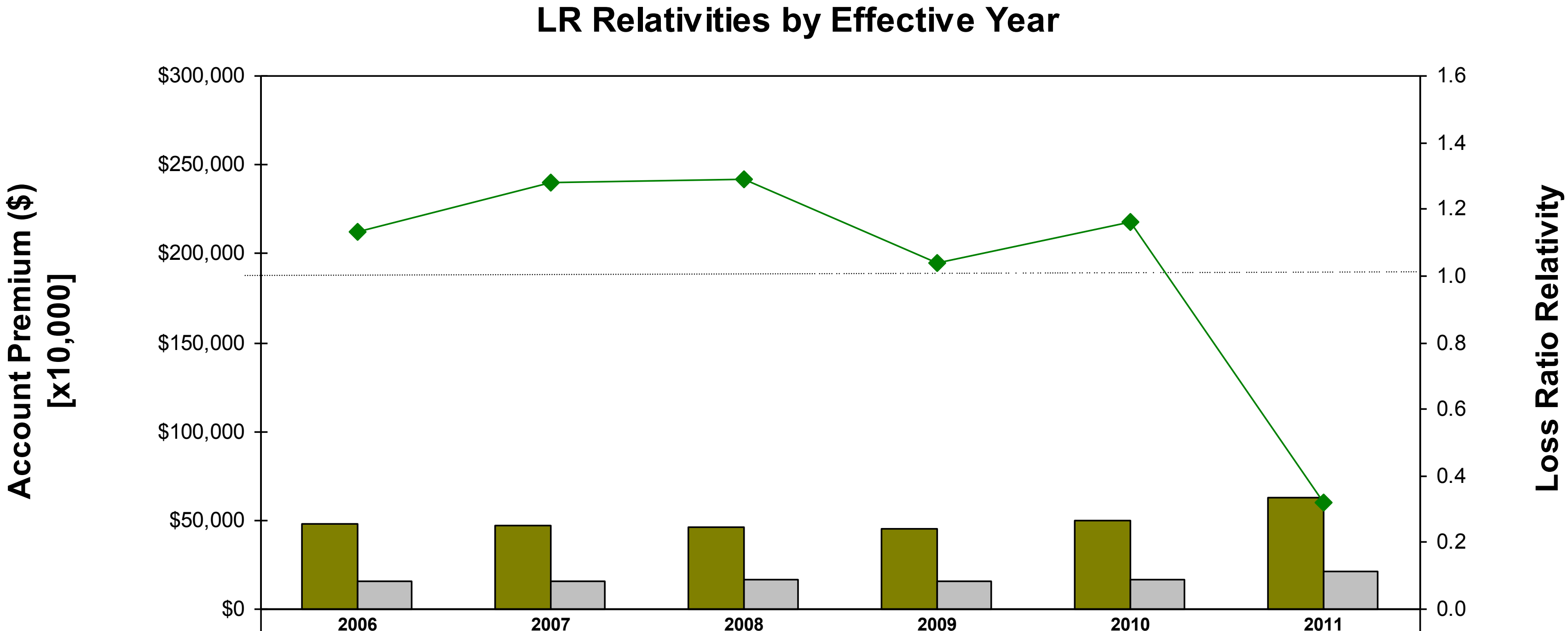
Categorical Data

- Bar Graphs to show make-up of categories
- Frequency tables showing percentage breakup.

Invalid, Missing or Inconsistent Data

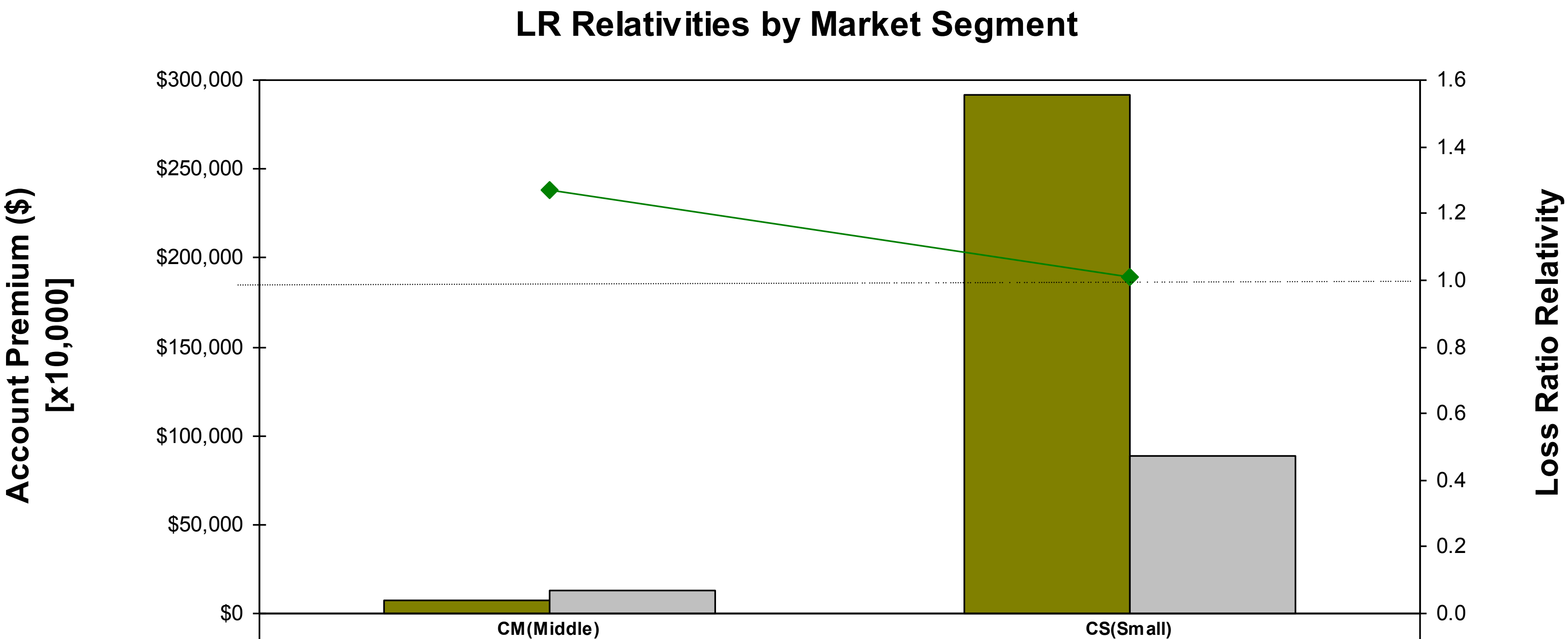
- Policies missing credit information.
- Incurred losses with negative, blank and extreme values.
- Building, Property and Contents limits with values <100.
- Rerated premiums with values <500.
- Business start and control year with years in 18th Century

EXPLORATORY ANALYSIS



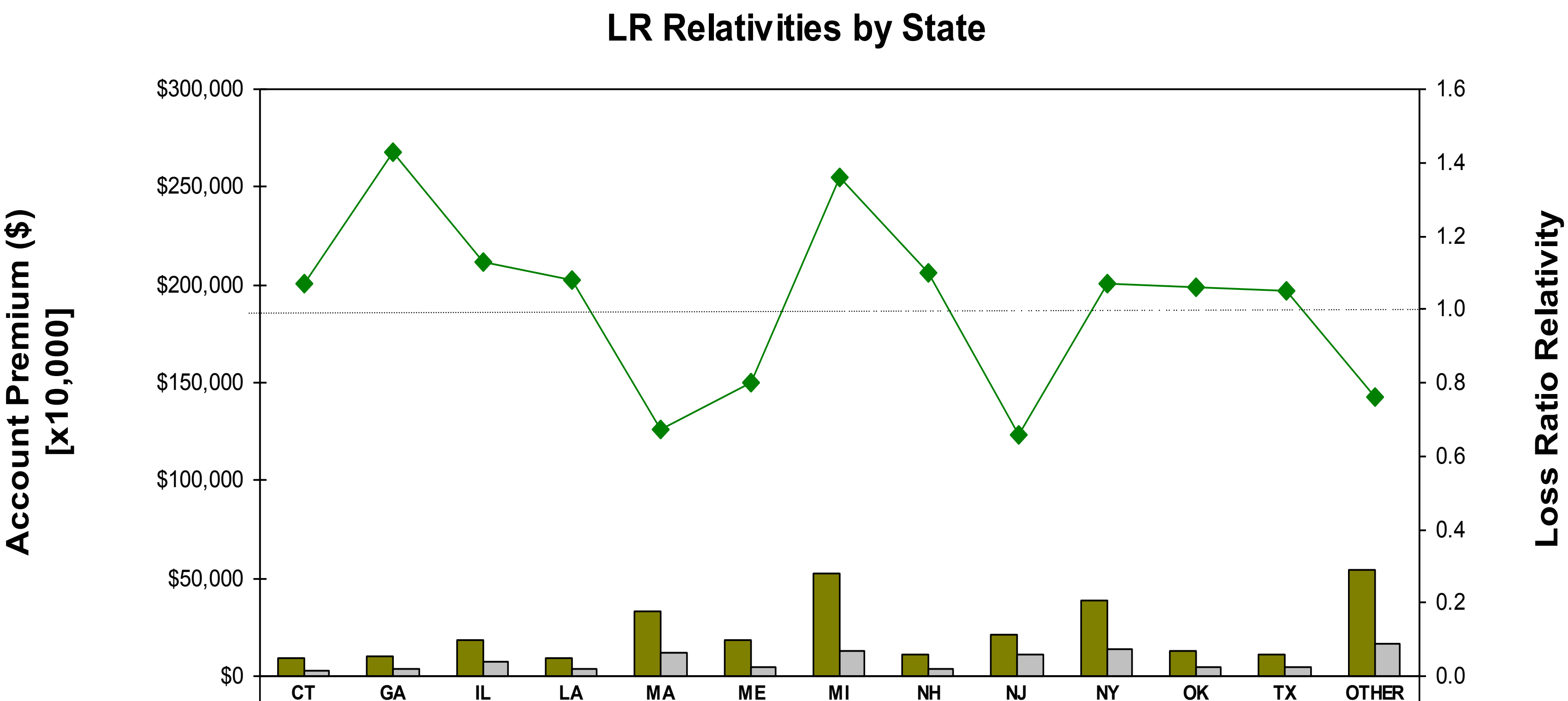
	2006	2007	2008	2009	2010	2011
Policy Count	48,209	46,909	46,465	45,056	49,660	63,142
Premium	$15,770	$15,945	$16,168	$15,392	$16,657	$21,368
Act LR Relativity	1.13	1.28	1.29	1.04	1.16	0.32

EXPLORATORY ANALYSIS



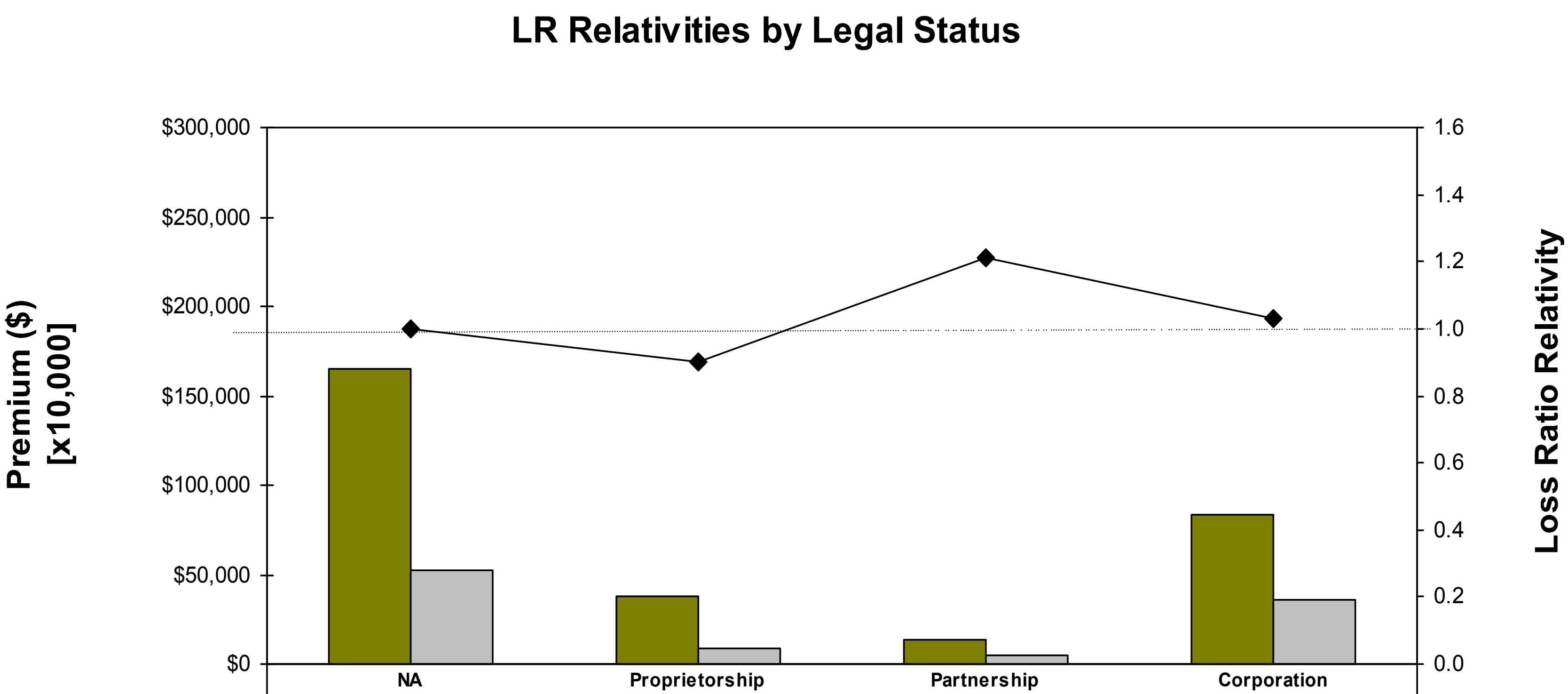
	CM(Middle)	CS(Small)
Policy Count	7,706	291,735
Premium	$12,853	$88,447
Act LR Relativity	1.27	1.01

EXPLORATORY ANALYSIS



	CT	GA	IL	LA	MA	ME	MI	NH	NJ	NY	OK	TX	OTHER
Policy Count	9,505	9,912	18,228	9,448	33,268	18,123	52,391	10,613	21,578	38,297	12,553	11,251	54,294
Premium	$3,014	$3,225	$7,276	$3,664	$12,297	$5,042	$12,889	$3,802	$10,771	$13,819	$4,505	$4,640	$16,352
Act LR Relativity	1.07	1.43	1.13	1.08	0.67	0.8	1.36	1.1	0.66	1.07	1.06	1.05	0.76

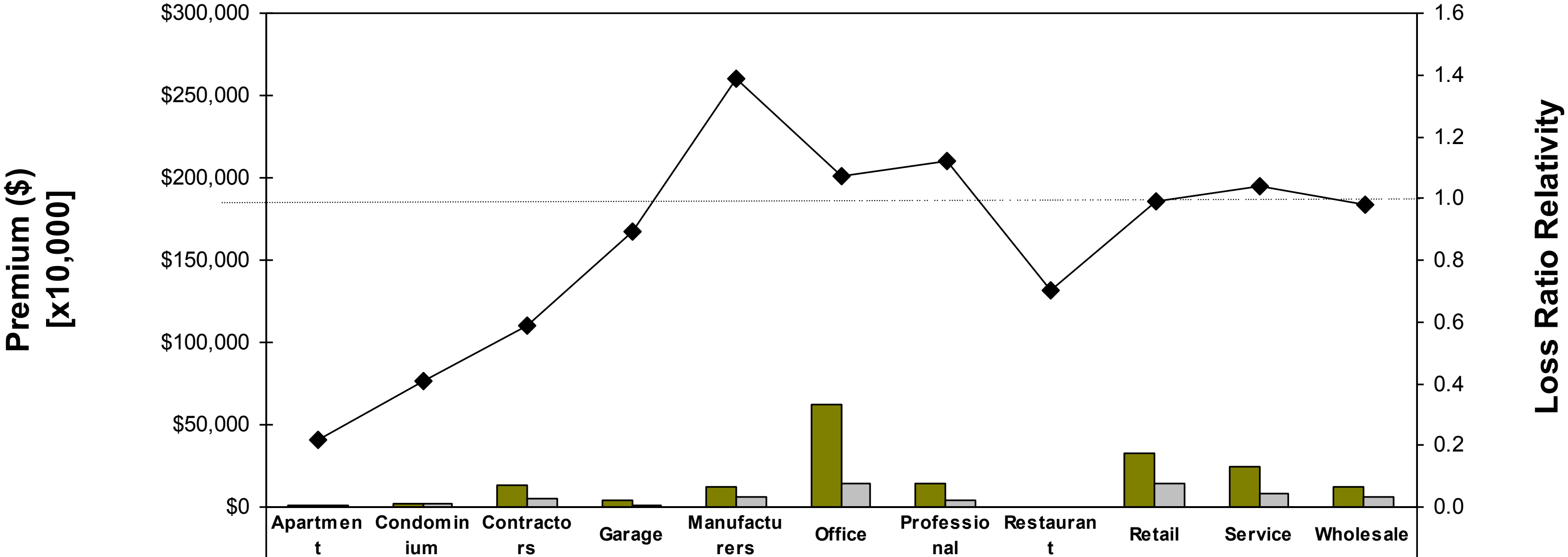
EXPLORATORY ANALYSIS



	NA	Proprietorship	Partnership	Corporation
Policy Count	164,803	37,798	13,750	83,090
Premium (x 10,000)	$51,943	$8,742	$4,742	$35,872
LR Relativity	1	0.9	1.21	1.03

EXPLORATORY ANALYSIS

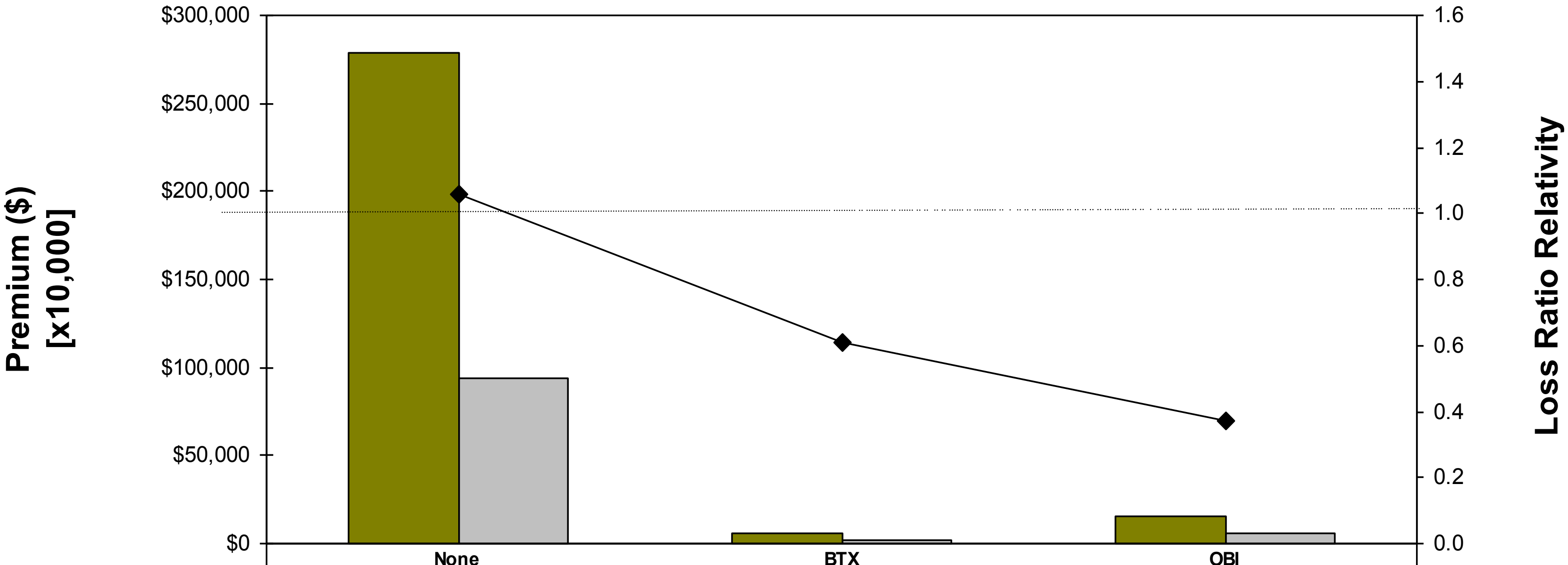
LR Relativities by Program



	Apartment	Condominium	Contractors	Garage	Manufacturers	Office	Professional	Restaurant	Retail	Service	Wholesale
Policy Count	985	2,055	13,288	3,793	12,628	61,850	14,706	332	32,366	24,670	12,542
Premium (x 10,000)	$1,105	$1,660	$4,639	$1,497	$6,059	$14,134	$3,645	$333	$14,475	$7,674	$6,079
LR Relativity	0.22	0.41	0.59	0.89	1.39	1.07	1.12	0.7	0.99	1.04	0.98

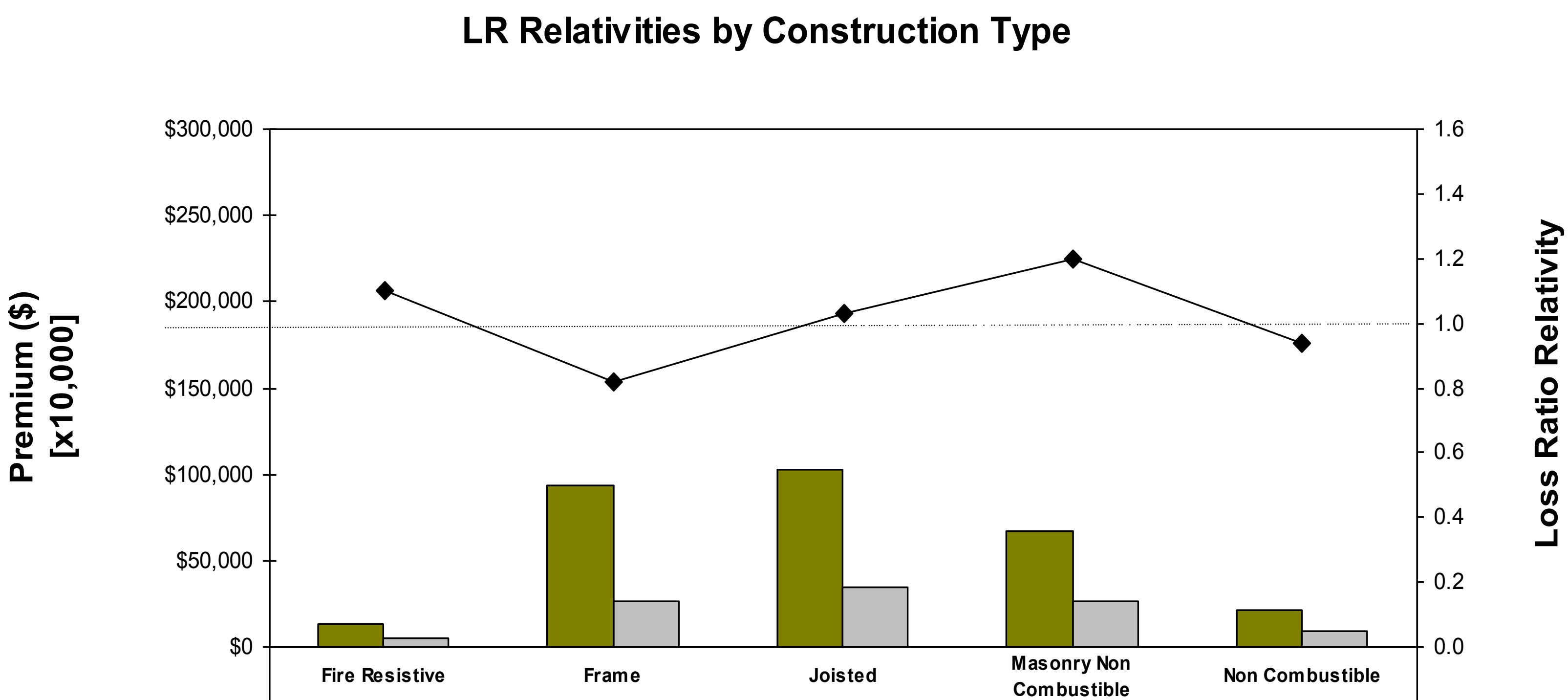
EXPLORATORY ANALYSIS

LR Relativities by Book Transfer Code



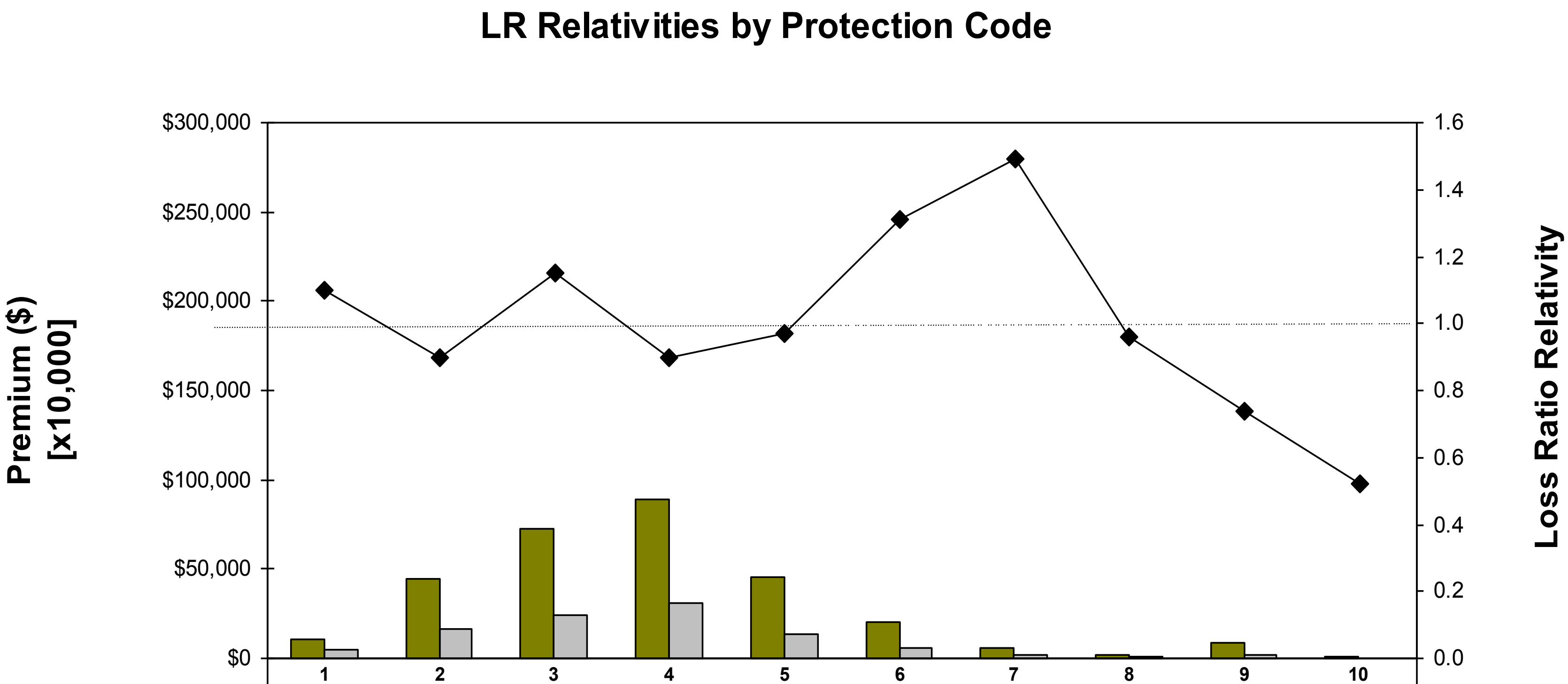
	None	BTX	OBI
Policy Count	278,270	6,005	15,165
Premium (x 10,000)	$93,614	$1,529	$6,157
LR Relativity	1.06	0.61	0.37

EXPLORATORY ANALYSIS



	Fire Resistive	Frame	Joisted	Masonry Non Combustible	Non Combustible
Policy Count	13,432	94,040	102,869	67,541	21,559
Premium (x 10,000)	$4,896	$26,723	$34,120	$26,180	$9,380
LR Relativity	1.1	0.82	1.03	1.2	0.94

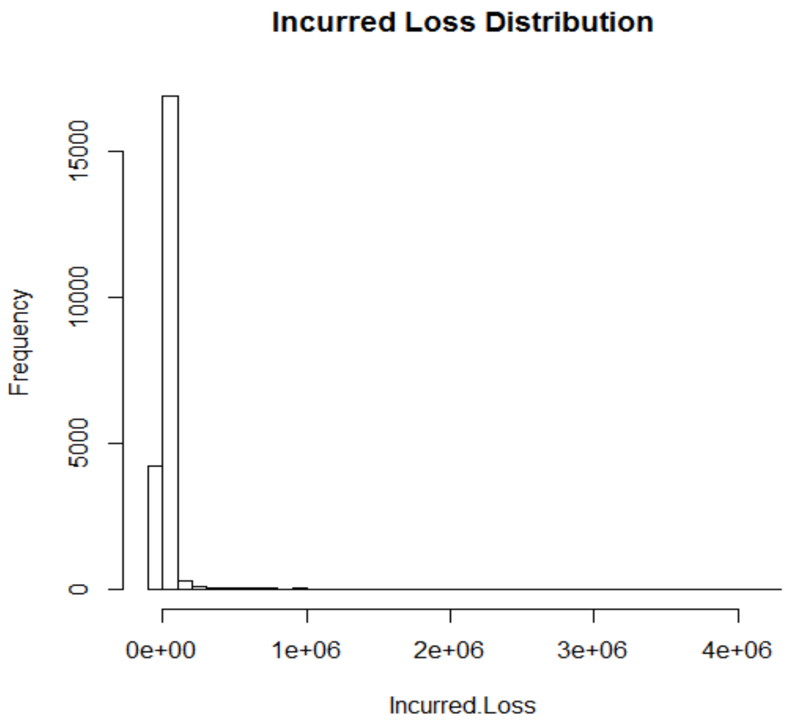
EXPLORATORY ANALYSIS

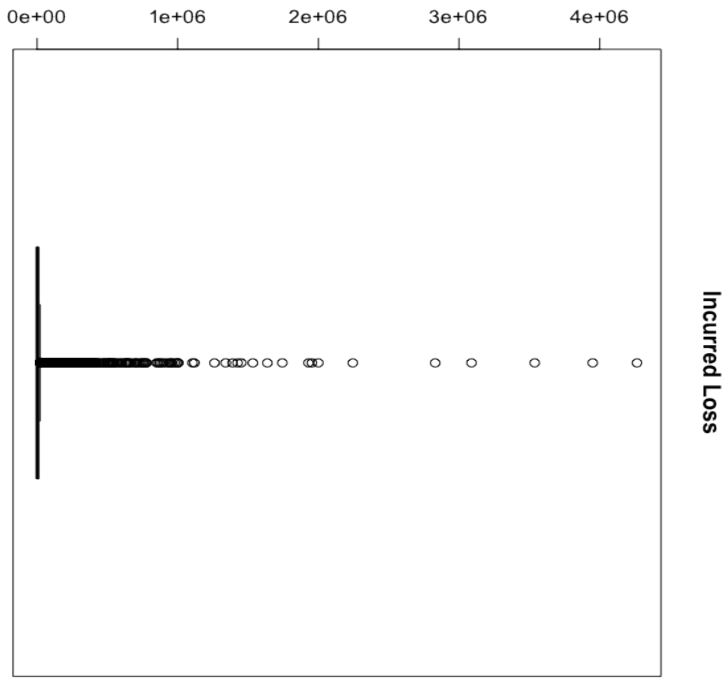


	1	2	3	4	5	6	7	8	9	10
Policy Count	10,892	44,624	72,494	89,079	45,629	20,529	5,378	1,816	8,516	484
Premium (x 10,000)	$4,821	$16,663	$24,653	$31,343	$13,857	$5,510	$1,464	$520	$2,336	$131
LR Relativity	1.1	0.9	1.15	0.9	0.97	1.31	1.49	0.96	0.74	0.52

EXPLORATORY ANALYSIS

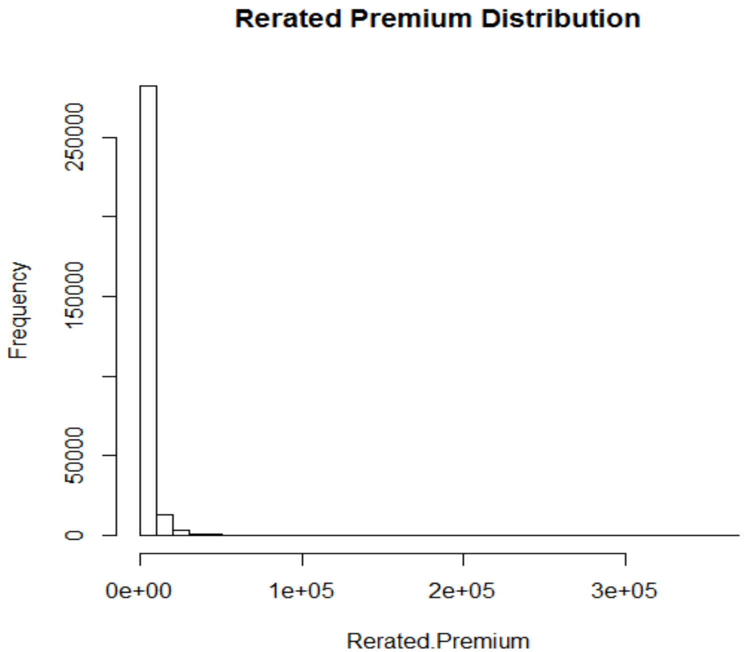
Incurred Loss

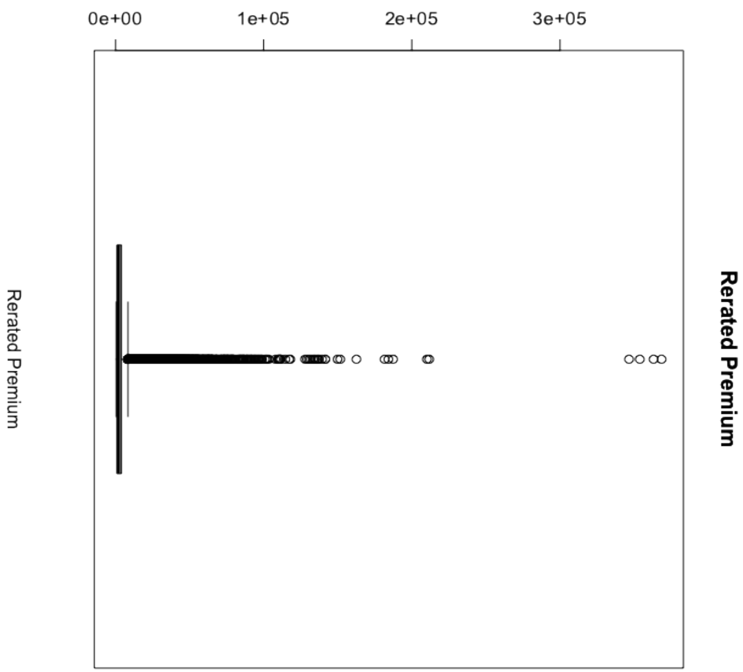




EXPLORATORY ANALYSIS

Rerated Premium





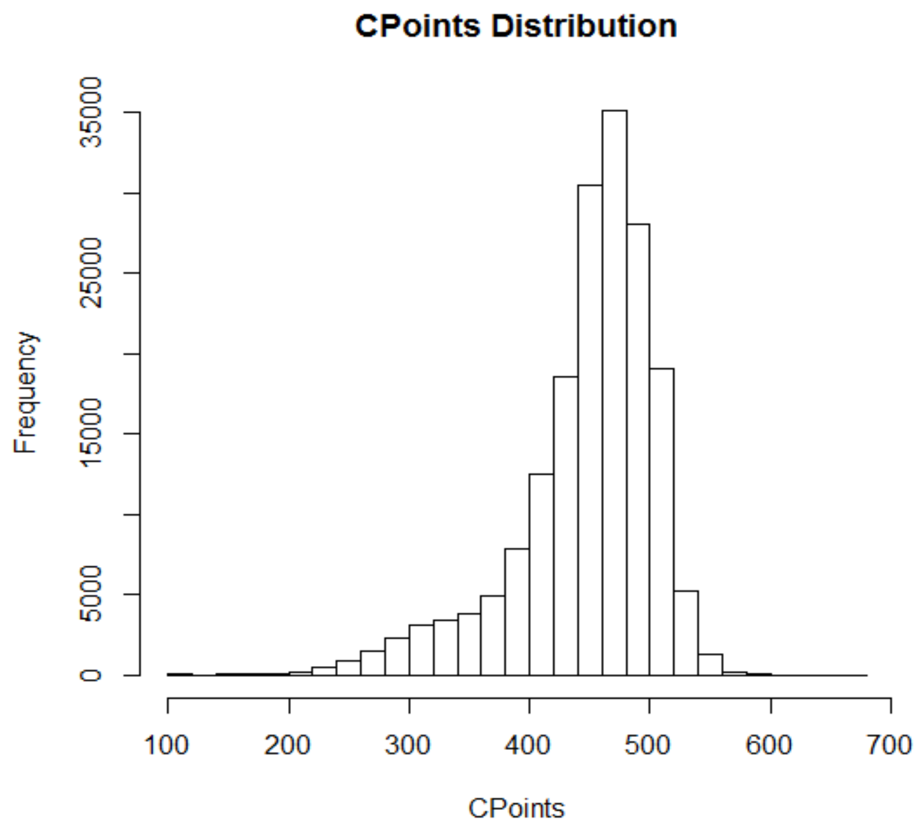
EXPLORATORY ANALYSIS

Population of Credit Variables.

Variable	Number of Zeros	Number of non-zero policies	Number of NA's	Total
C Points	5,822	179,316	114,303	299,441
F Points	5,822	179,216	114,403	299,441
Liens	187,374	195	111,872	299,441
Suits	187,421	148	111,872	299,441
Judgments	187515	54	111,872	299,441
Legal Status	0	134, 638	164,803	299,441

EXPLORATORY ANALYSIS

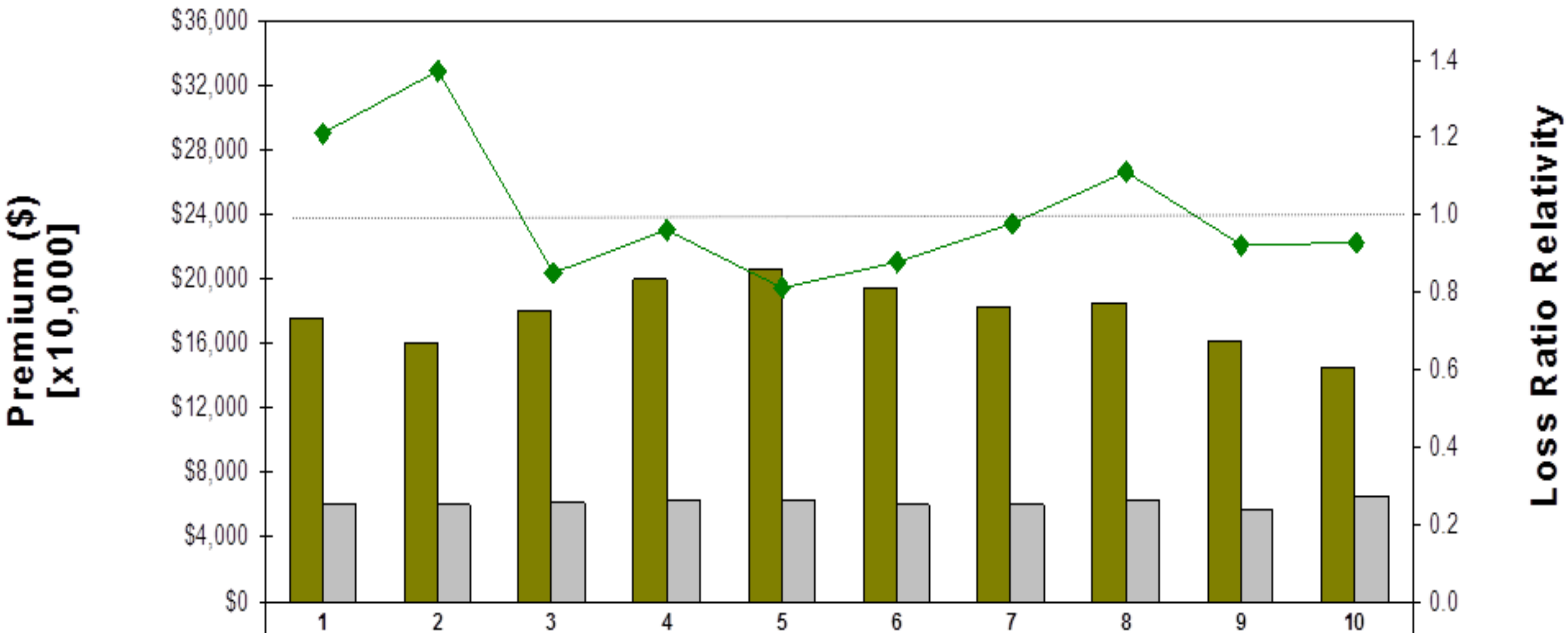
C Points (Risk of credit default)



EXPLORATORY ANALYSIS

C Points (Risk of credit default)

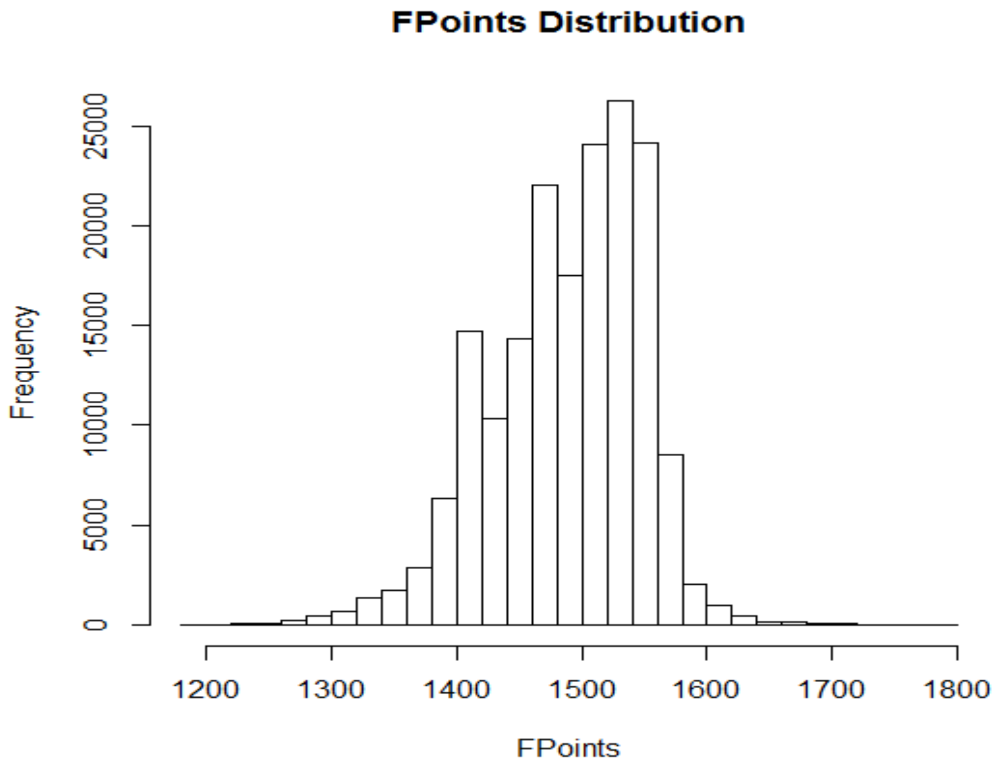
LR Relativities by C Points



	1	2	3	4	5	6	7	8	9	10
Policy Count	17,600	16,018	18,050	20,044	20,629	19,409	18,212	18,538	16,199	14,516
Premium (x 10,000)	$6,016	$6,040	$6,086	$6,270	$6,224	$6,017	$6,043	$6,274	$5,800	$6,530
LR Relativity	1.21	1.37	0.85	0.96	0.81	0.88	0.98	1.11	0.92	0.93
Lower Limit	101	369	410	433	450	462	473	484	497	511
Upper Limit	369	410	433	450	462	473	484	497	511	670

EXPLORATORY ANALYSIS

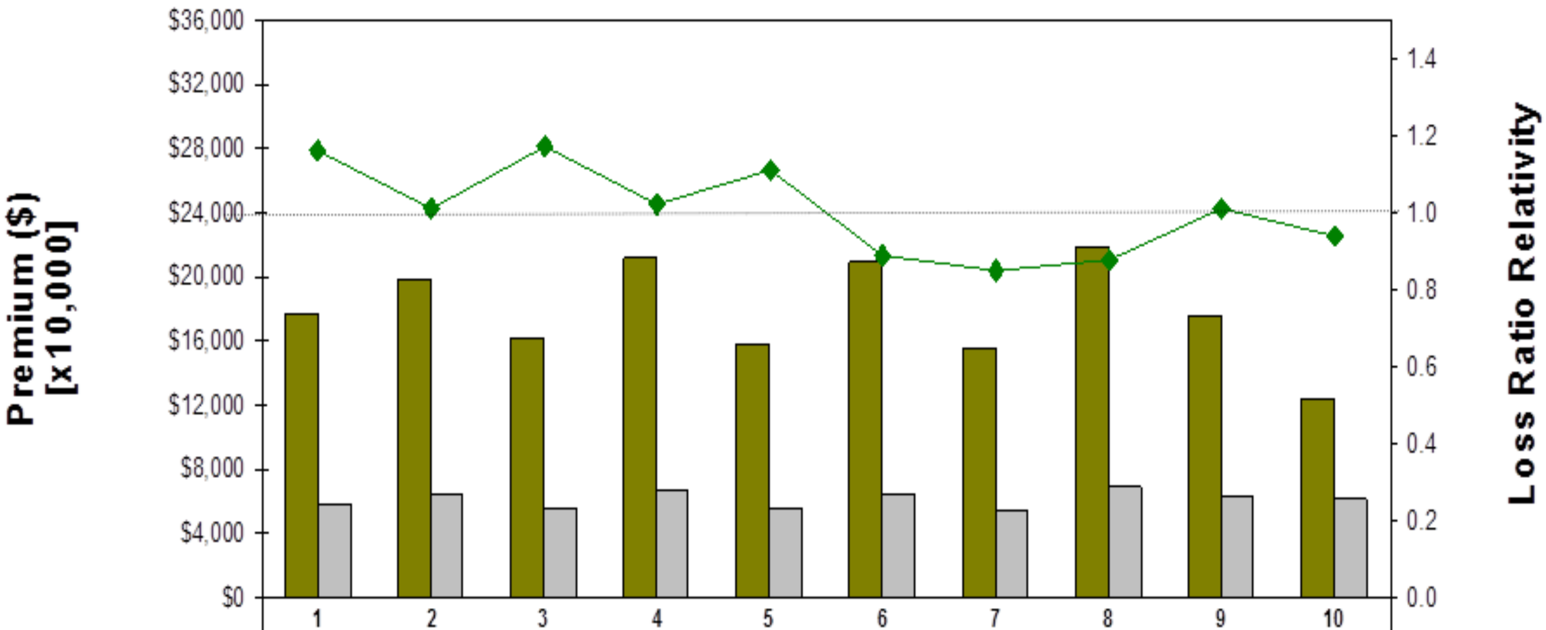
F Points (Risk of financial stress)



EXPLORATORY ANALYSIS

F Points (Risk of financial stress)

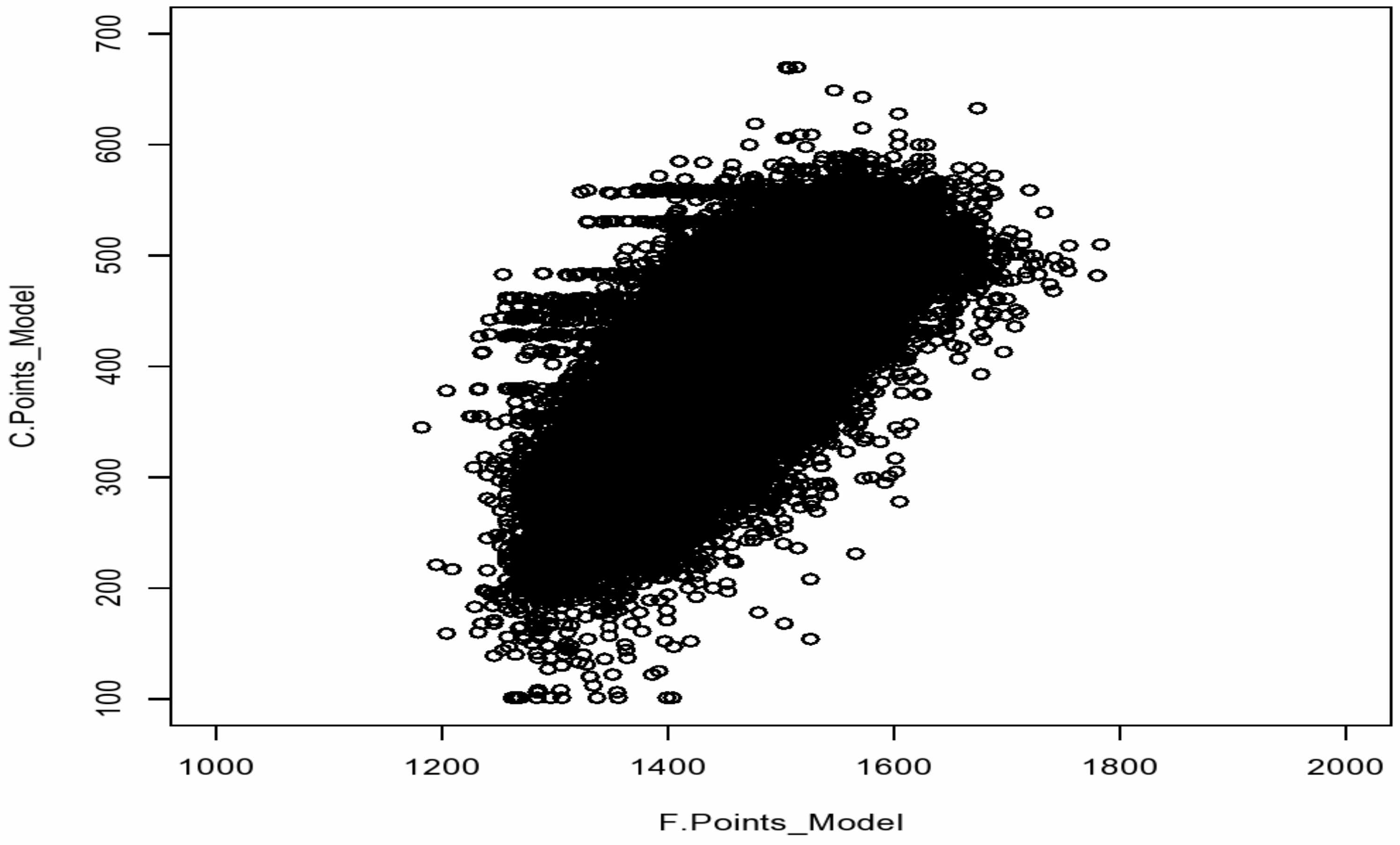
LR Relativities by F Points



	1	2	3	4	5	6	7	8	9	10
Policy Count	17,766	19,836	16,186	21,225	15,730	20,942	15,586	21,885	17,649	12,411
Premium (x 10,000)	$5,813	$6,415	$5,592	$6,700	$5,621	$6,398	$5,480	$6,892	$6,250	$6,140
LR Relativity	1.16	1.01	1.17	1.02	1.11	0.89	0.85	0.88	1.01	0.94
Lower Limit	1,182	1,408	1,439	1,462	1,481	1,500	1,516	1,530	1,546	1,560
Upper Limit	1,408	1,439	1,462	1,481	1,500	1,516	1,530	1,546	1,560	1,783

EXPLORATORY ANALYSIS

C Points vs. F Points Scatter plot



EXPLORATORY ANALYSIS (SUMMARY)

- Roughly 40% of policies had no credit data.
- Noticed unexpected values for some variables:
  - Rerated premiums (below $500, the minimum premium amount)
  - Incurred losses (below $0)
  - Building year (polices with years of zero and starting in 17th Century)
  - Building, property and contents limit (values less than 100)
  - Building Start and Control years (values given as 0)
- Many policies had data given as N/A: incurred loss, legal status, business start and control year.
- We noticed a number of extreme values in rerated premiums, incurred losses and incurred loss ratios.

PRESENTATION OUTLINE

- Introduction
- Goals and Objective
- Exploratory Analysis
- Data Preparation
- Modeling Process
  - Model Design
  - Model Fitting
  - Testing and Adjusting Model
- Key Results
- Conclusions
- Acknowledgements

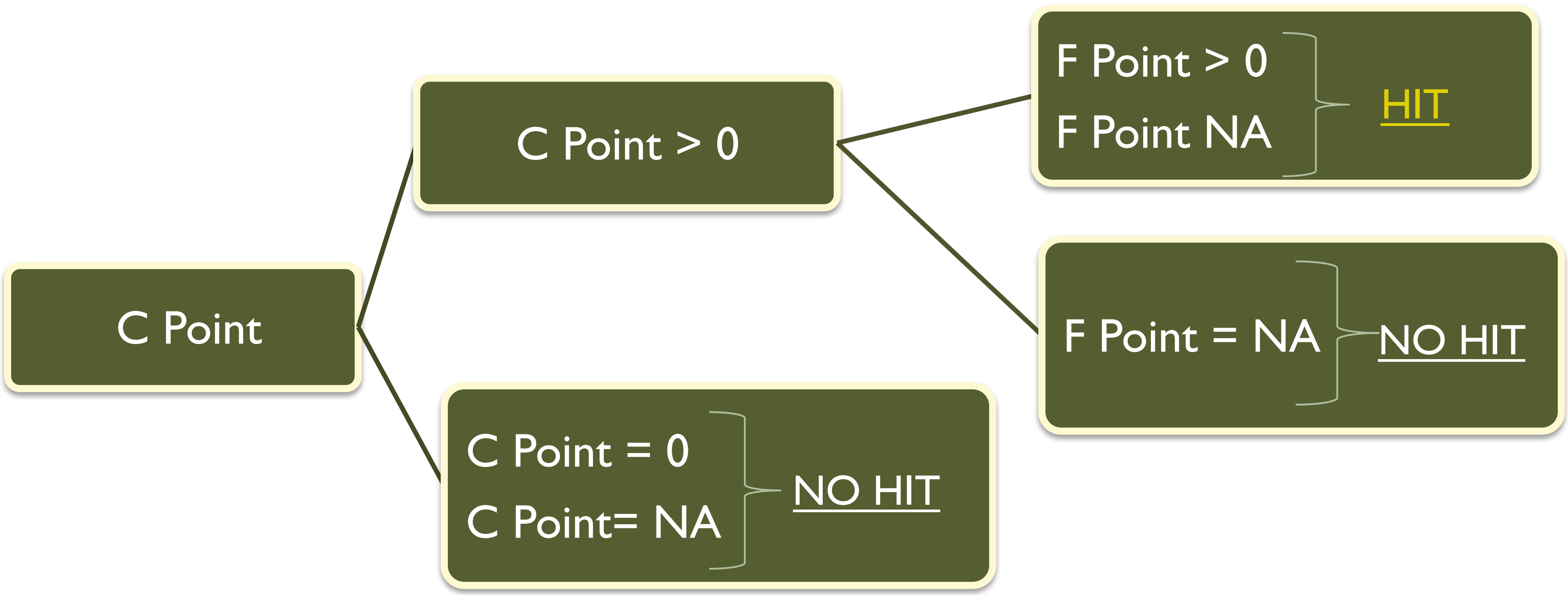
DATA PREPARATION (INITIAL STEPS)

1. We began with 299,441 policies
2. We removed the following
   - 120,226 polices with no credit information ("no-hits")
   - 332 restaurant policies were removed
3. We were left with 178,883 policies for use in modeling

	Original data	Remove No Hits	Remove Restaurant	After data Deletions
Incurred Loss Ratio	33.26%	32.94%	32.99%	32.99%
Number of Policies	299,441	179,215	178,883	178,883

DATA PREPARATION (INITIAL STEPS)

Definition of Hits/No-Hits



Total Policies	Total No-Hits	Total Hits
299,441	120,226	179,215

DATA PREPARATION

Once we reduced the data to 178,883 policies, we made additional adjustments:

1. Rerated premium capped at 500, impacting 423 policies.
2. Building, property and contents limit between 0 and 100 were set to 0, impacting 113 policies.
3. Building years before 1631 were set to NA, impacting 22 policies.
4. Incurred losses given as negative or NA were set to zero, impacting 165,809 policies (only 19 policies were negative).
5. Incurred loss ratios capped at 95th percentile of the positive loss ratios (2071%), impacting 501 policies.

	LR before adjustments	Capping at 95th percentile
Incurred Loss Ratio	**32.99%**	**21.76%**

DATA PREPARATION

- After cleaning the data we had to determine the appropriate variables to use in the model.

- However we had to check for multi-collinearity between variables. As expected our results showed that C points and F points were correlated which had to be corrected.

- Factor Analysis:
  - enabled analysis of multi-collinearity among continuous variables.
  - an uncorrelated factor was found by weighting the sum of the standardized variables (C Points and F Points).
  - this factor was named Financial Stability and used as an input variable.

DATA PREPARATION

Creation of Data Variables

- Control Age= 2012 – max (business start year, control year)
- Business Age= Effective year – Building year
- Effective Age= 2012 – effective year

- Created an indicator for:
  - polices with contents limit only, building limits only and both building and contents limits.

- Miscellaneous modifications:
  - grouped protection codes 1-2, 3, 4-6, and 7-10.
  - grouped location count into those =1 and those > 1.
  - combined apartment and condominium program names
  - included states with less than 1,000 policies in Southeast, Midwest or West regions.
  - grouped by GL Limit: less than 1 million, 1 million, or 2 million.
  - New Transfer?

DATA PREPARATION (SUMMARY)

1. Deleted policies with missing, invalid or inconsistent data (i.e., no-hits and restaurant policies)

2. Data Adjustments:
   - Adjusted negative incurred losses.
   - Capped:
     - Rerated premiums at $500.
     - Positive incurred loss-ratios at 95th percentile.
   - Building, property and contents limit below 100.
   - Building years before 1631.

3. Modified Data Variables
   - Created control age, effective age and financial stability.
   - Created indicators for certain variables.
   - Grouped specific variables into major sections for better analysis.

PRESENTATION OUTLINE

- Introduction
- Goals and Objective
- Exploratory Analysis
- Data Preparation
- Modeling Process
  - Model Design
  - Model Fitting
  - Testing and Adjusting Model
- Key Results
- Conclusions
- Acknowledgements

MODEL DESIGN

Response variable: Incurred loss ratio

Model distribution: Tweedie (member of exponential family)

Model type: Generalized linear model (GLM)

Input variables:

- Program Name, Region, and Construction Type
- Property Limit, Legal Status, Property Deductible, and New/Renew
- Financial stability, Control Age, and Effective age
- Limit Indicator, Location count and Protection Group
- New transfer

MODEL DESIGN

$\log(E[loss\ ratio]) = B0 + \beta 1 \times Program\ name\downarrow\ + \beta 2 \times Legal\ Status\downarrow\ + \beta 3 \times Financial\ Stability\downarrow\ + \ldots + \mathcal{E}$

Components of GLM:

– Random component- a group of n independent observations with a distribution from the exponential family.

– Systematic component- a linear predictor $\eta = \beta \mathbf{X}$ is used to weight the predictor variables for each individual observation.

– Link function- A logarithmic link function was used which set the predictor above ($\eta = \beta \mathbf{X}$) equal to log (μ) where μ represents the predicted incurred loss ratio.

MODEL FITTING

1. **Create random sample**: 80% build sample, 20% for validation sample.
   - Assigned random number to each policy
   - Ordered the policies from lowest to highest
   - Chose top 80% for build sample

2. **Model using software:** Used GLM method in R software to model Loss Ratios with Tweedie family.

3. **Rerun on Full Sample:**
   - Model created using the build sample, then run on validation sample.
   - After optimal model was achieved, model was run on full sample.

TESTING AND ADJUSTING MODEL

Before we were able to analyze results we performed tests to determine goodness of model fit:

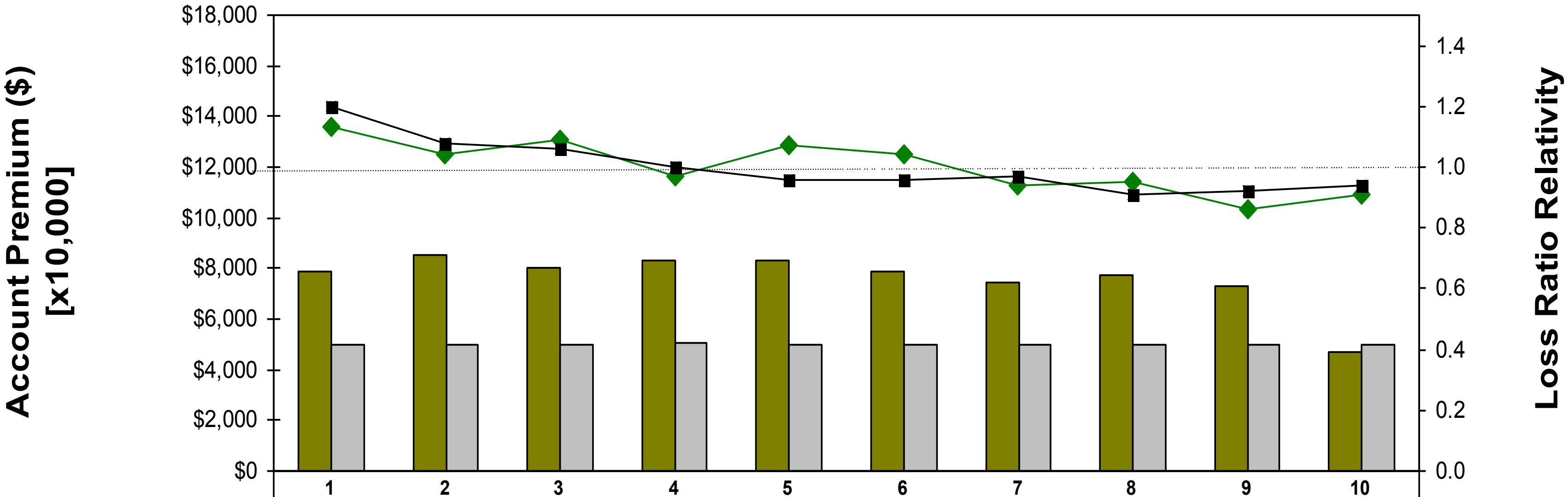
- Analysis of Variance (ANOVA) with Chi-Squared test P-value
  - Measures the significance of each input variable to the overall model fit
- Akaike Information Criteria (AIC)
  - Measures the relative fit of a candidate model fit compared to other candidates
- Lift Charts
  - Breaks data into groups based on variable being measured.
  - Calculates and plots actual and predicted incurred loss ratio relativities.

PRESENTATION OUTLINE

- Introduction
- Goals and Objective
- Exploratory Analysis
- Data Preparation
- Modeling Process
  - Model Design
  - Model Fitting
  - Testing and Adjusting Model
- Key Results
- Conclusions
- Acknowledgements

CREDIT MODEL LIFT BY FINANCIAL STABILITY (BUILD SAMPLE)

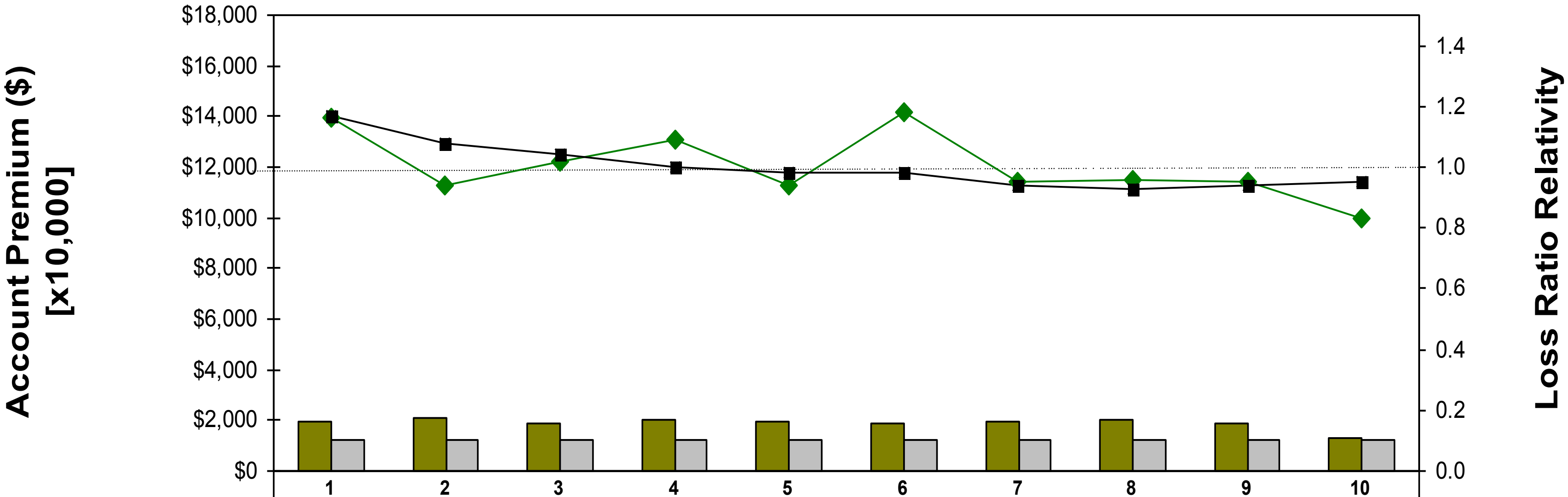
Actual versus Predicted LR Relativities by Financial Stability



	1	2	3	4	5	6	7	8	9	10
Account count	7,880	8,518	7,993	8,348	8,287	7,900	7,452	7,710	7,322	4,727
Premium	$5,007.77	$5,009.78	$4,998.51	$5,027.92	$4,993.79	$5,014.25	$5,009.04	$5,007.46	$5,007.91	$5,010.24
Act LR Relativity	1.13	1.04	1.09	0.97	1.07	1.04	0.94	0.95	0.86	0.91
Pred LR Relativity	1.2	1.08	1.06	1	0.96	0.96	0.97	0.91	0.92	0.94
Lower Limit	-4.84	-1.3	-0.72	-0.29	0	0.27	0.5	0.72	0.94	1.21
Upper Limit	-1.3	-0.72	-0.29	0	0.27	0.5	0.72	0.94	1.21	3.17

CREDIT MODEL LIFT BY FINANCIAL STABILITY (VALIDATION SAMPLE)

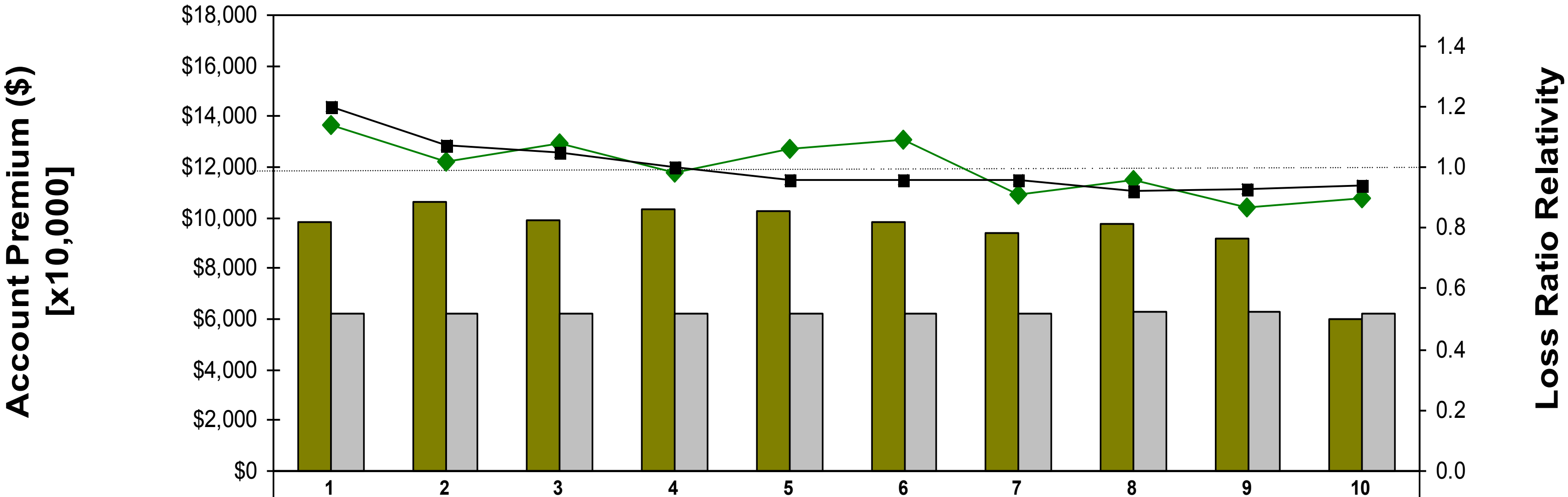
Actual versus Predicted LR Relativities by Financial Stability



	1	2	3	4	5	6	7	8	9	10
Account count	1,975.00	2,115.00	1,906.00	2,041.00	1,973.00	1,880.00	1,951.00	2,037.00	1,886.00	1,272.00
Premium	$1,231.17	$1,231.32	$1,248.05	$1,236.48	$1,237.40	$1,236.56	$1,235.29	$1,230.72	$1,235.52	$1,246.01
Act LR Relativity	1.16	0.94	1.02	1.09	0.94	1.18	0.95	0.96	0.95	0.83
Pred LR Relativity	1.17	1.08	1.04	1	0.98	0.98	0.94	0.93	0.94	0.95
Lower Limit	-4.58	-1.32	-0.73	-0.31	-0.03	0.22	0.45	0.69	0.91	1.18
Upper Limit	-1.32	-0.73	-0.31	-0.03	0.22	0.45	0.69	0.91	1.18	3.02

CREDIT MODEL LIFT BY FINANCIAL STABILITY (FULL SAMPLE)

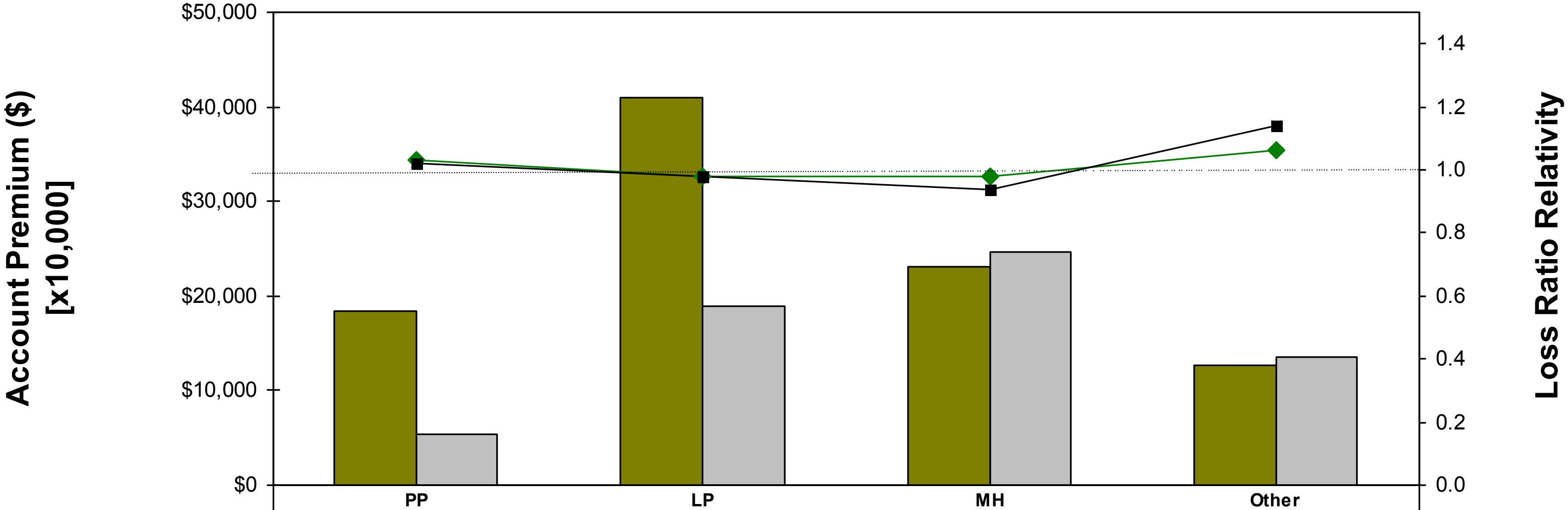
Actual versus Predicted LR Relativities by Financial Stability



	1	2	3	4	5	6	7	8	9	10
Account count	9,839	10,648	9,931	10,334	10,278	9,837	9,372	9,733	9,169	6,032
Premium	$6,246.43	$6,239.23	$6,247.46	$6,246.28	$6,244.76	$6,241.51	$6,227.36	$6,260.65	$6,255.70	$6,245.82
Act LR Relativity	1.14	1.02	1.08	0.98	1.06	1.09	0.91	0.96	0.87	0.9
Pred LR Relativity	1.2	1.07	1.05	1	0.96	0.96	0.96	0.92	0.93	0.94
Lower Limit	-5.03	-1.31	-0.73	-0.29	0	0.26	0.49	0.72	0.93	1.2
Upper Limit	-1.31	-0.73	-0.29	0	0.26	0.49	0.72	0.93	1.2	3.17

CREDIT MODEL LIFT BY POLICY TYPE (FULL SAMPLE)

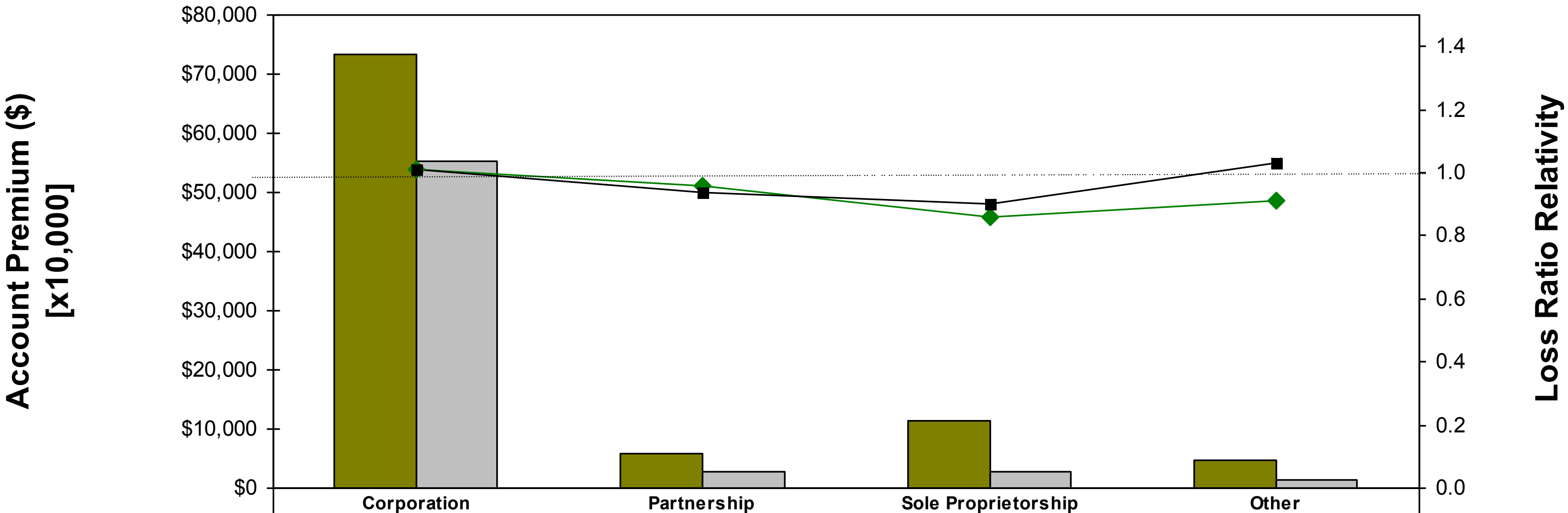
Actual versus Predicted LR Relativities by Policy Type



	PP	LP	MH	Other
Account count	18,401	41,043	23,075	12,654
Premium	$5,464.66	$18,924.83	$24,593.18	$13,472.53
Act LR Relativity	1.03	0.98	0.98	1.06
Pred LR Relativity	1.02	0.98	0.94	1.14

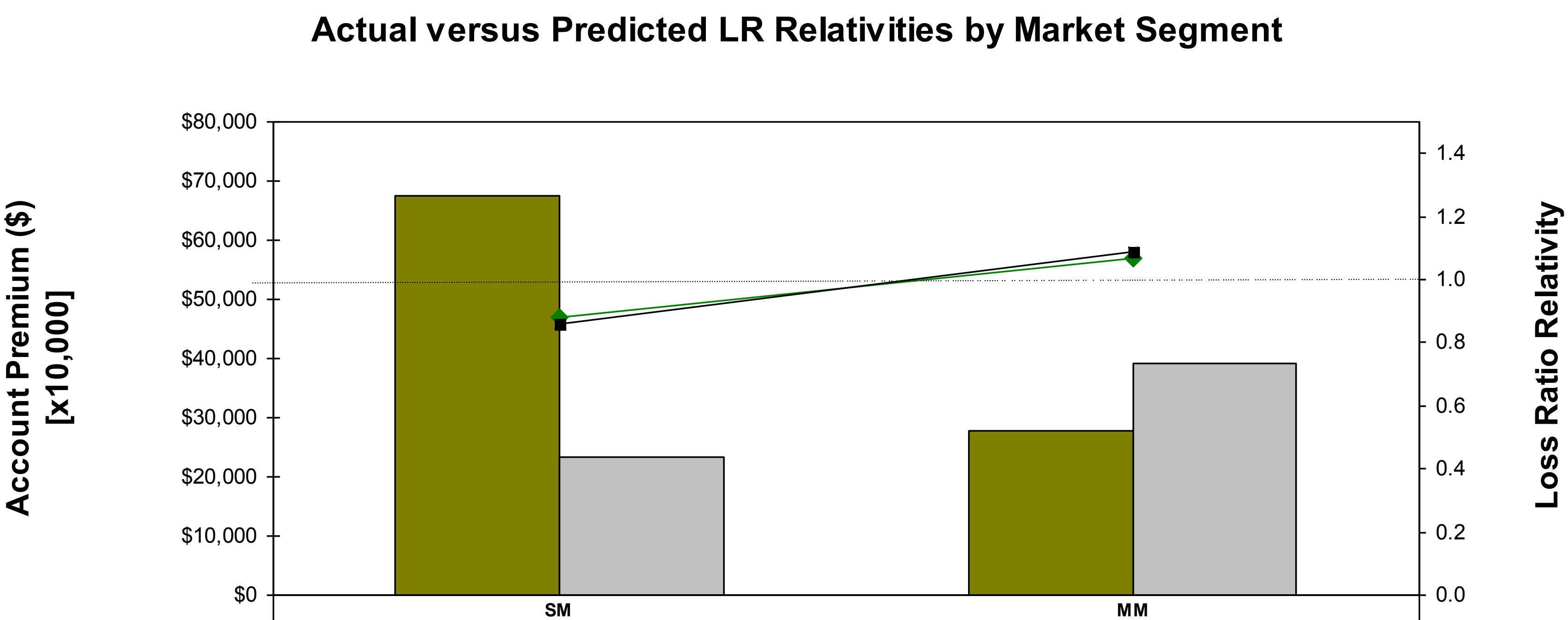
CREDIT MODEL LIFT BY BUSINESS TYPE (FULL SAMPLE)

Actual versus Predicted LR Relativities by Business Type



	Corporation	Partnership	Sole Proprietorship	Other
Account count	73,435	5,859	11,274	4,605
Premium	$55,383.04	$2,846.00	$2,747.89	$1,478.26
Act LR Relativity	1.01	0.96	0.86	0.91
Pred LR Relativity	1.01	0.94	0.9	1.03

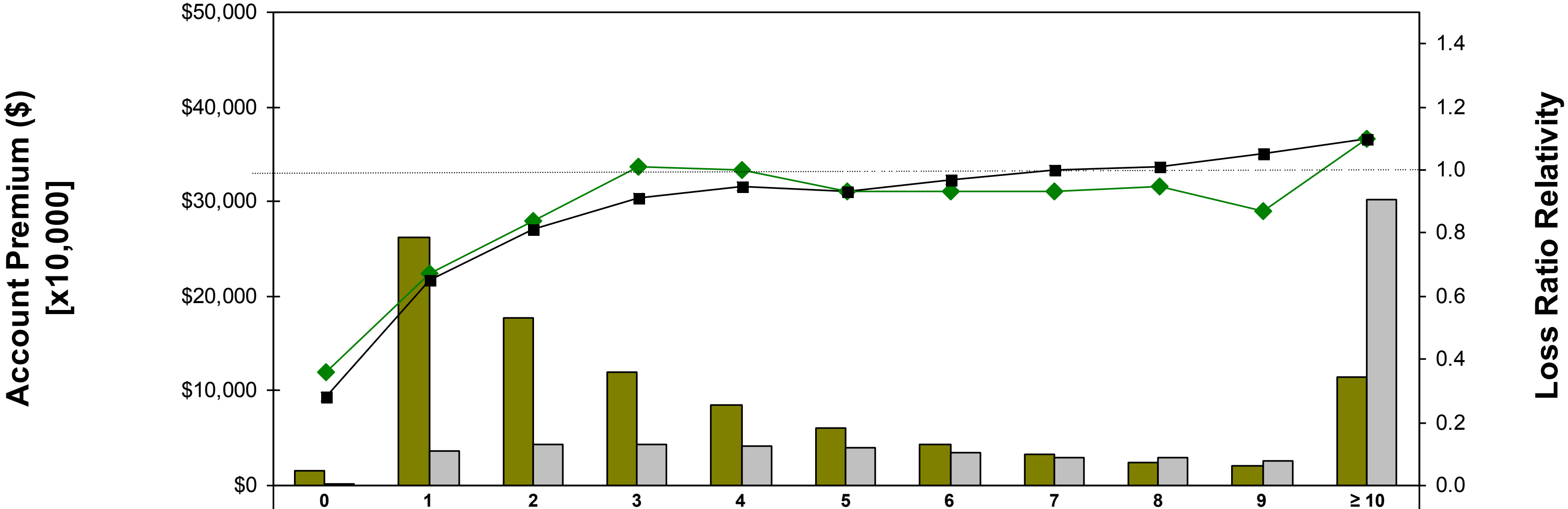
CREDIT MODEL LIFT BY MARKET SEGMENT (FULL SAMPLE)



	SM	MM
Account count	67,476	27,697
Premium	$23,374.82	$39,080.38
Act LR Relativity	0.88	1.07
Pred LR Relativity	0.86	1.09

CREDIT MODEL LIFT BY FLEET SIZE (FULL SAMPLE)

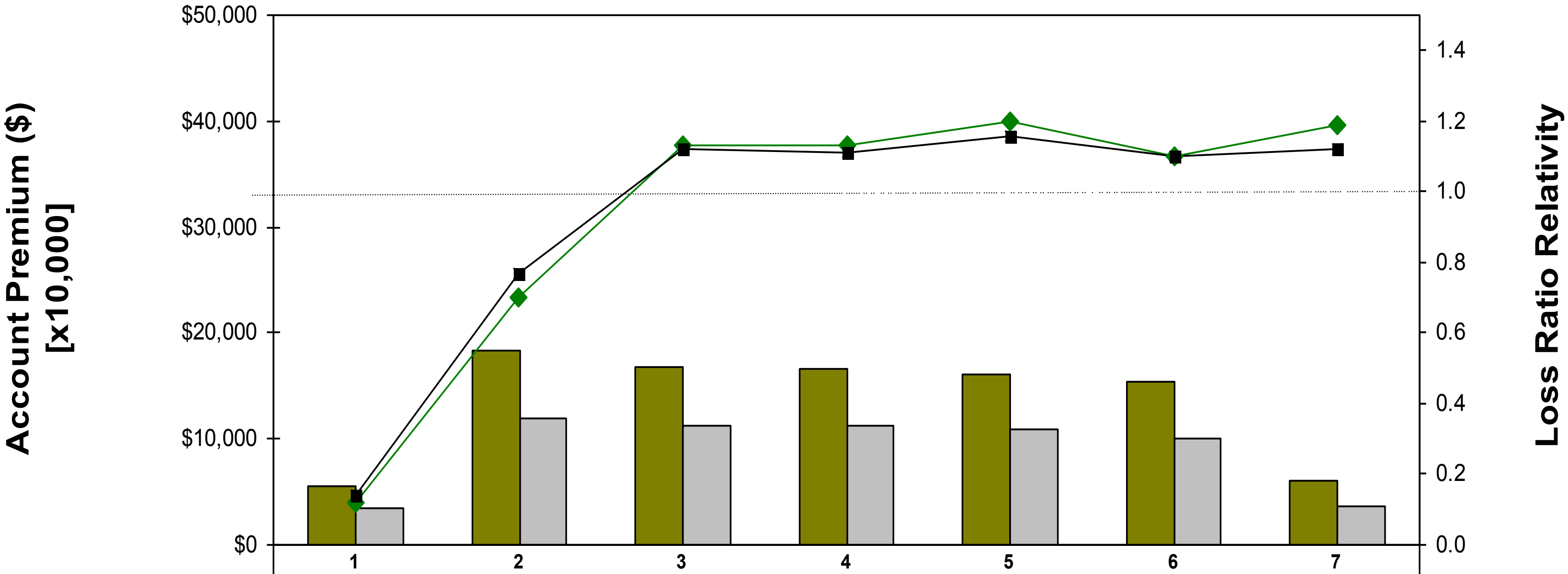
Actual versus Predicted LR Relativities by Fleet Size



	0	1	2	3	4	5	6	7	8	9	≥ 10
Account count	1,485	26,157	17,626	11,900	8,547	6,007	4,313	3,224	2,517	2,000	11,397
Premium	$89.33	$3,658.74	$4,341.88	$4,273.81	$4,156.83	$3,929.52	$3,423.57	$2,919.19	$2,891.27	$2,605.45	$30,165.60
Act LR Relativity	0.36	0.67	0.84	1.01	1	0.93	0.93	0.93	0.95	0.87	1.1
Pred LR Relativity	0.28	0.65	0.81	0.91	0.95	0.93	0.97	1	1.01	1.05	1.1

CREDIT MODEL LIFT BY POLICY EFFECTIVE AGE (FULL SAMPLE)

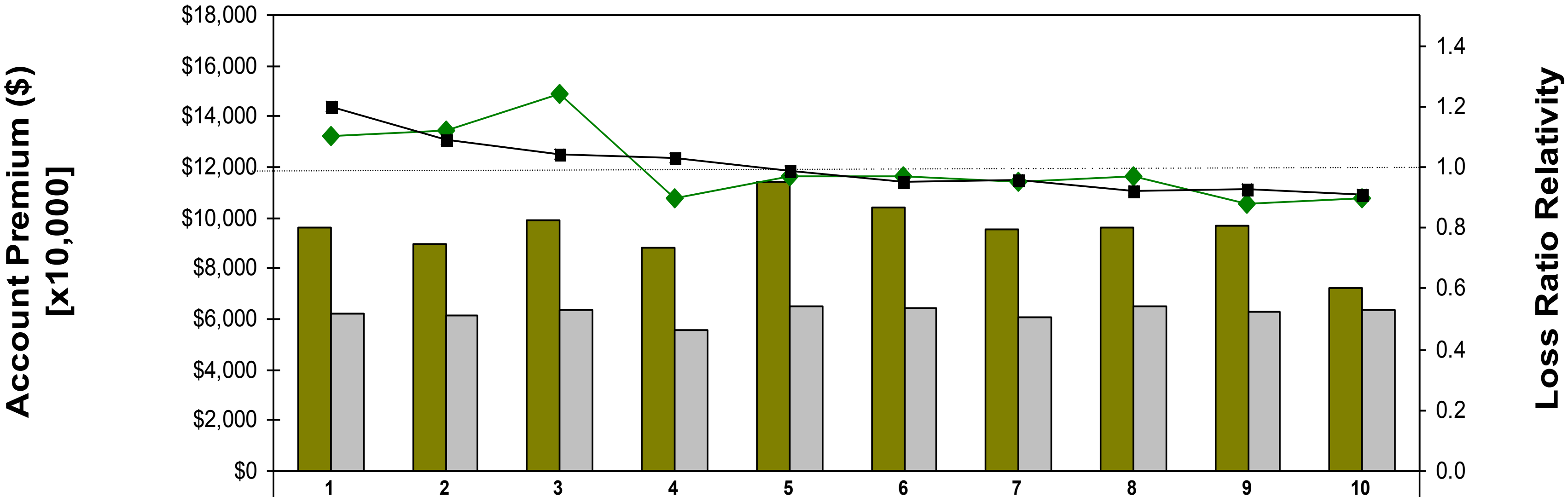
Actual versus Predicted LR Relativities by Policy Effective Age



	1	2	3	4	5	6	7
Account count	5,575	18,403	16,830	16,694	16,118	15,416	6,137
Premium	$3,444.20	$12,007.22	$11,209.58	$11,209.22	$10,923.88	$10,065.32	$3,595.78
Act LR Relativity	0.12	0.7	1.13	1.13	1.2	1.1	1.19
Pred LR Relativity	0.14	0.77	1.12	1.11	1.16	1.1	1.12

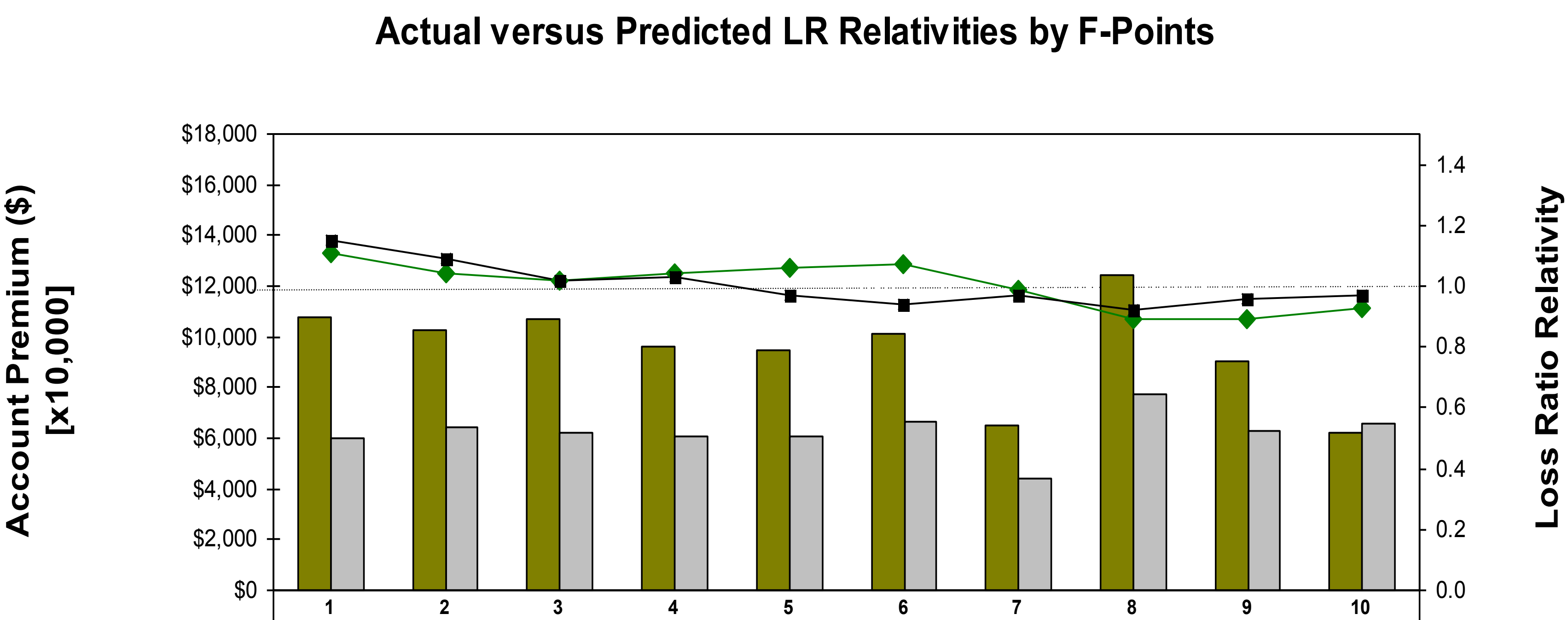
CREDIT MODEL LIFT BY C-POINTS (FULL SAMPLE)

Actual versus Predicted LR Relativities by C-Points



	1	2	3	4	5	6	7	8	9	10
Account count	9,594	8,939	9,869	8,793	11,434	10,437	9,551	9,642	9,716	7,198
Premium	$6,213.14	$6,141.37	$6,376.25	$5,577.17	$6,475.39	$6,446.94	$6,075.53	$6,474.84	$6,294.70	$6,379.87
Act LR Relativity	1.1	1.12	1.24	0.9	0.97	0.97	0.95	0.97	0.88	0.9
Pred LR Relativity	1.2	1.09	1.04	1.03	0.99	0.95	0.96	0.92	0.93	0.91
Lower Limit	101	369	412	435	451	465	478	490	503	519
Upper Limit	369	412	435	451	465	478	490	503	519	655

CREDIT MODEL LIFT BY F-POINTS (FULL SAMPLE)



	1	2	3	4	5	6	7	8	9	10
Account count	10,784	10,265	10,665	9,631	9,442	10,139	6,510	12,467	9,038	6,232
Premium	$6,026.29	$6,403.25	$6,188.18	$6,095.60	$6,058.55	$6,644.52	$4,436.31	$7,732.18	$6,298.98	$6,571.32
Act LR Relativity	1.11	1.04	1.02	1.04	1.06	1.07	0.99	0.89	0.89	0.93
Pred LR Relativity	1.15	1.09	1.02	1.03	0.97	0.94	0.97	0.92	0.96	0.97
Lower Limit	1,198	1,410	1,450	1,473	1,493	1,510	1,526	1,538	1,554	1,579
Upper Limit	1,410	1,450	1,473	1,493	1,510	1,526	1,538	1,554	1,579	1,827

PRESENTATION OUTLINE

CONCLUSIONS

- The credit variable Financial Stability is an powerful predictor of future loss ratio of a policy.
- Implementation of a credit factor will allow for better differentiation of risk, ultimately improving underwriting profit.
- Usage will ensure that Hanover's underwriting techniques are competitive and more advanced.

PRESENTATION OUTLINE

- Introduction
- Goals and Objective
- Exploratory Analysis
- Data Preparation
- Modeling Process
  - Model Design
  - Model Fitting
  - Testing and Adjusting Model
- Key Results
- Conclusions
- Acknowledgements

ACKNOWLEDGEMENTS

The WPI Team would like to thank all the persons that invested time to help shape the direction and outcome of our project:

Isin Ozaksoy

Marc Cournoyer

Jonathan Blake

Chen Li

Alyssa Potter

Professor Jon Abraham

And everyone else at Hanover and Worcester Polytechnic Institute that made this project possible!

THANK YOU!

QUESTIONS?