

# **LOSS RATIO MODELING FOR COMMERCIAL AUTO LINE OF BUSINESS**

---



15<sup>TH</sup> FEBRUARY, 2012

# PRESENTATION OUTLINE

---

- Introduction
- Goals and Objectives
- Exploratory Analysis
- Data Preparation
- Modeling Process
  - Model Design
  - Model Fitting
  - Testing and Adjusting Model
- Key Results
- Conclusions
- Future Considerations
- 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 Commercial Auto pricing model?

## Proposed Solution:

Model that implements a credit score variable to predict loss ratio for each policy and provide a recommendation on appropriate credit factor.

# PRESENTATION OUTLINE

---

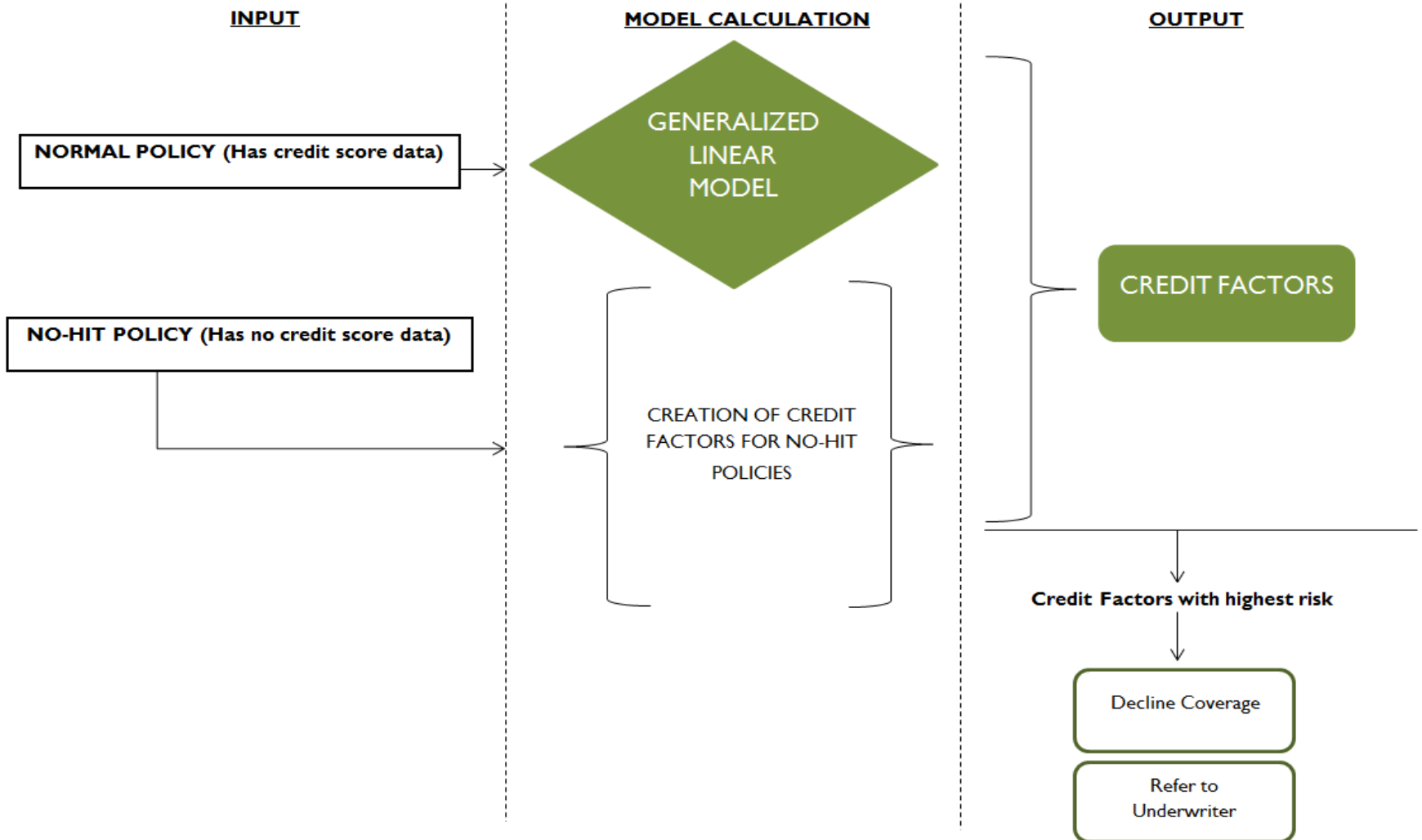
- Introduction
- **Goals and Objectives**
- Exploratory Analysis
- Data Preparation
- Modeling Process
  - Model Design
  - Model Fitting
  - Testing and Adjusting Model
- Key Results
- Conclusions
- Future Considerations
- Acknowledgements

# PROJECT GOALS AND OBJECTIVES

---

1. Identify a base set of risk factors for commercial auto.
2. Use these risk factors to calculate predicted loss ratios for each policy.
3. Use the predicted loss ratios to develop credit factors for each policy.
4. Assess current pricing and underwriting techniques to determine how best to incorporate credit factors into current calculations.

# PROPOSED PROCESS OVERVIEW



# PRESENTATION OUTLINE

---

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

# EXPLORATORY ANALYSIS

---

We outlined three 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 various data fields and loss ratio as well as relationships amongst variables.**
  - Univariate analysis by plotting variables against loss ratios.
  - Correlation between C points and F points
  
- 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 165,142 policies over the period 2005-2011

## Numeric Data

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

## Categorical Data

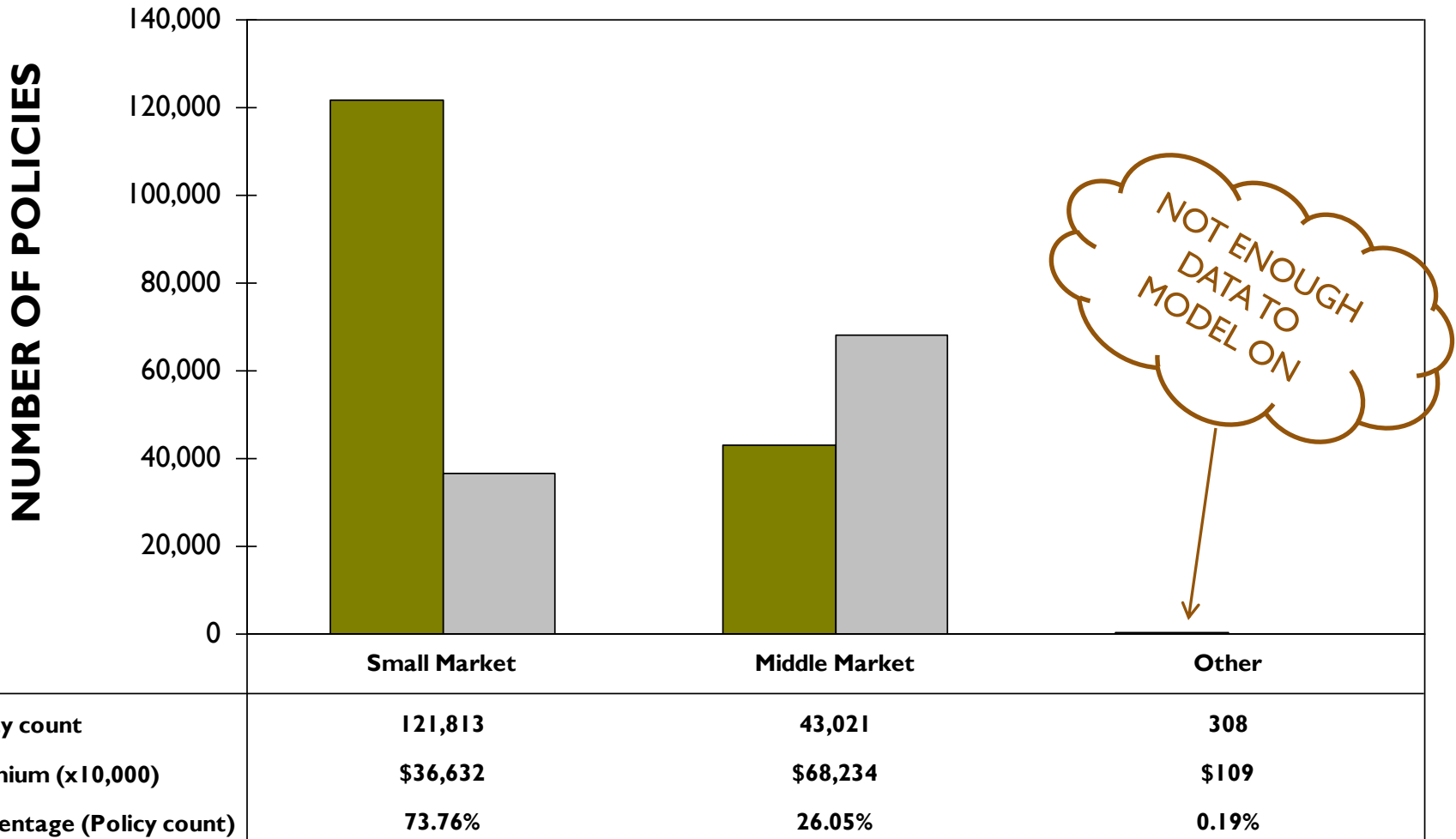
- Bar graphs to show make-up of categories
- Frequency tables showing percentage breakup.

## Invalid, Missing or Inconsistent Data

- 35.3% policies have no credit information.
- 3.5% policies have no premiums.
- 3.4% manually written policies
- 2.7% policies have premium below \$500
- <0.1% policies have negative incurred losses.

# EXPLORATORY ANALYSIS

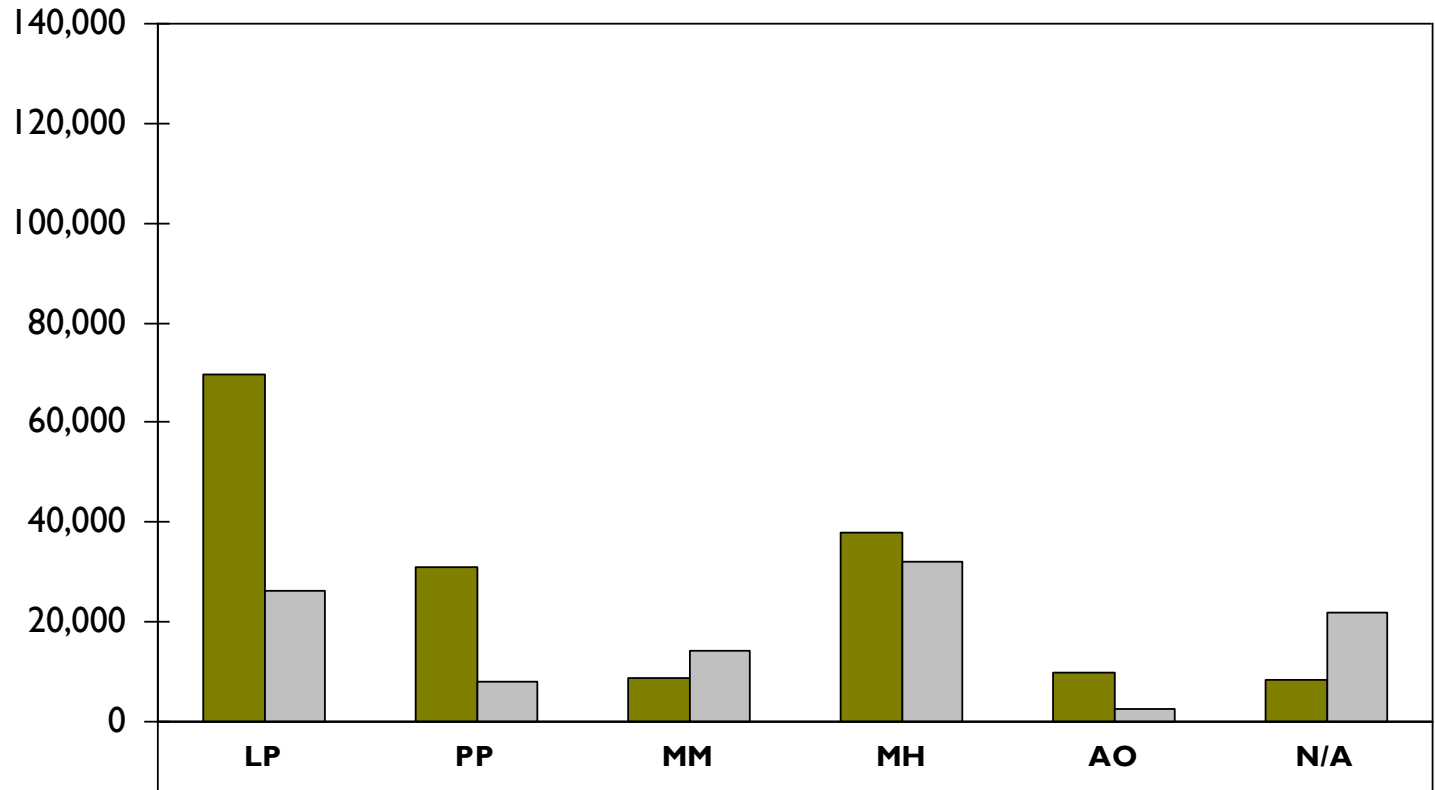
## DISTRIBUTION BY MARKET SEGMENT





# EXPLORATORY ANALYSIS

## DISTRIBUTION BY POLICY TYPE

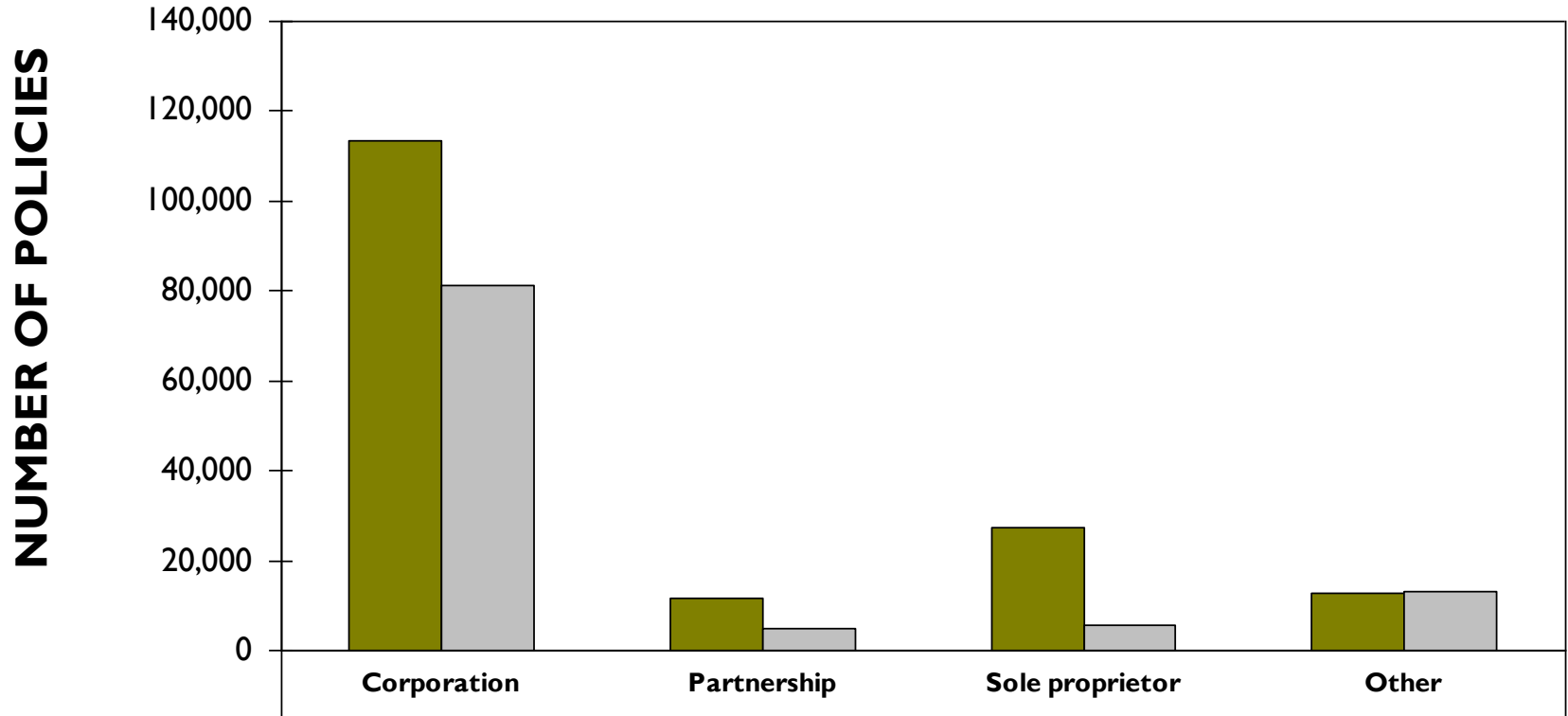
**NUMBER OF POLICIES**



 Policy count	69,528	31,057	8,734	37,735	9,738	8,350
 Premium (x10,000)	\$26,281	\$7,958	\$14,228	\$32,085	\$2,563	\$21,860
Percentage (Policy Count)	42.09%	18.81%	5.29%	22.85%	5.90%	5.06%

# EXPLORATORY ANALYSIS

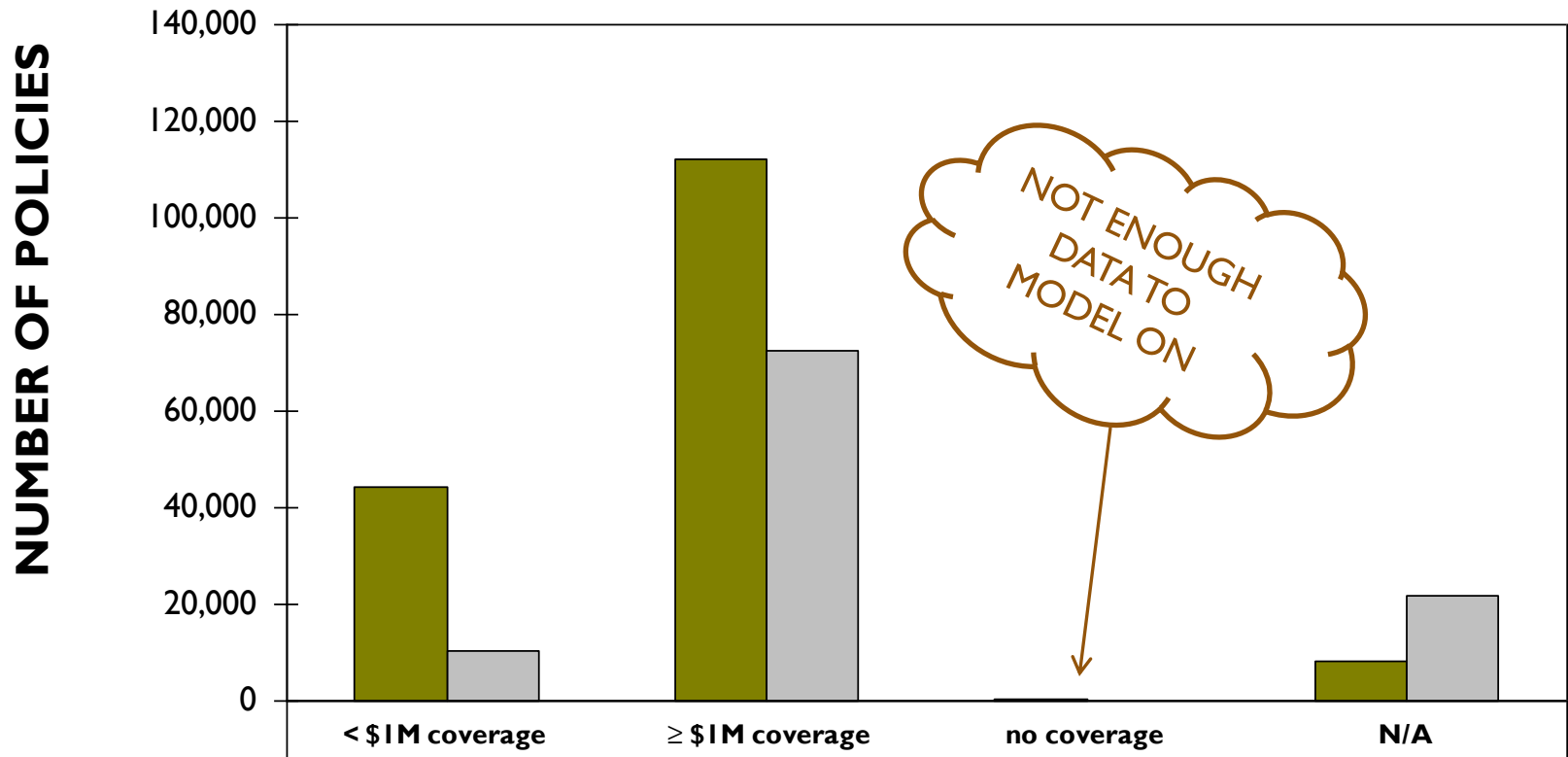
## DISTRIBUTION BY BUSINESS TYPE



<b>Policy count</b>	<b>113,302</b>	<b>11,695</b>	<b>27,333</b>	<b>12,812</b>
<b>Premium(x10,000)</b>	<b>\$81,155</b>	<b>\$4,966</b>	<b>\$5,628</b>	<b>\$13,226</b>
<b>Percentage (Policy Count)</b>	<b>68.61%</b>	<b>7.08%</b>	<b>16.55%</b>	<b>7.76%</b>

# EXPLORATORY ANALYSIS

## DISTRIBUTION BY LIABILITY GROUP

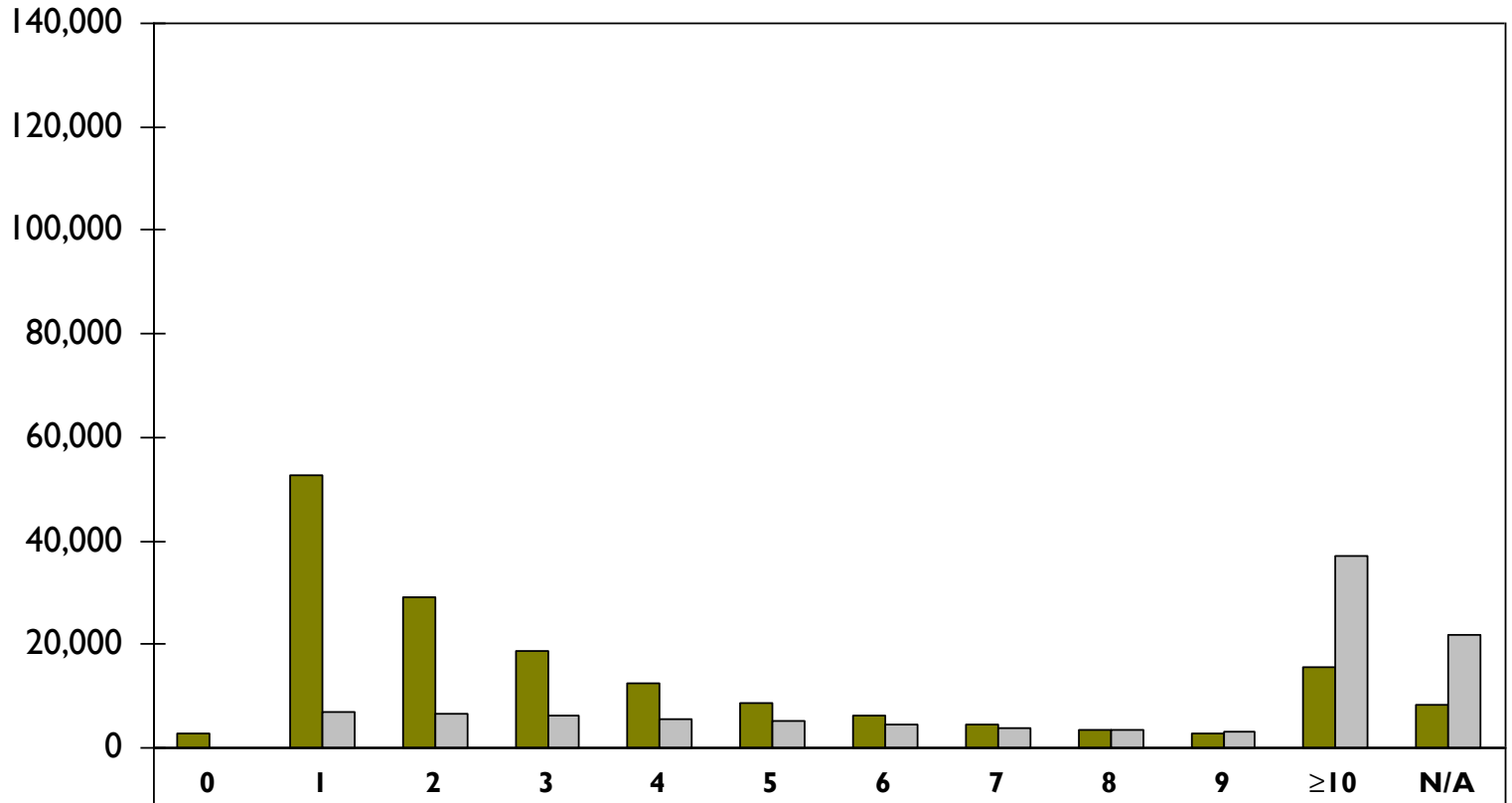


■ Policy count	44,414	112,048	322	8,358
■ Premium(x10,000)	\$10,462	\$72,631	\$22	\$21,860
Percentage (Policy Count)	26.89%	67.85%	0.20%	5.06%

# EXPLORATORY ANALYSIS

## DISTRIBUTION BY FLEET SIZE

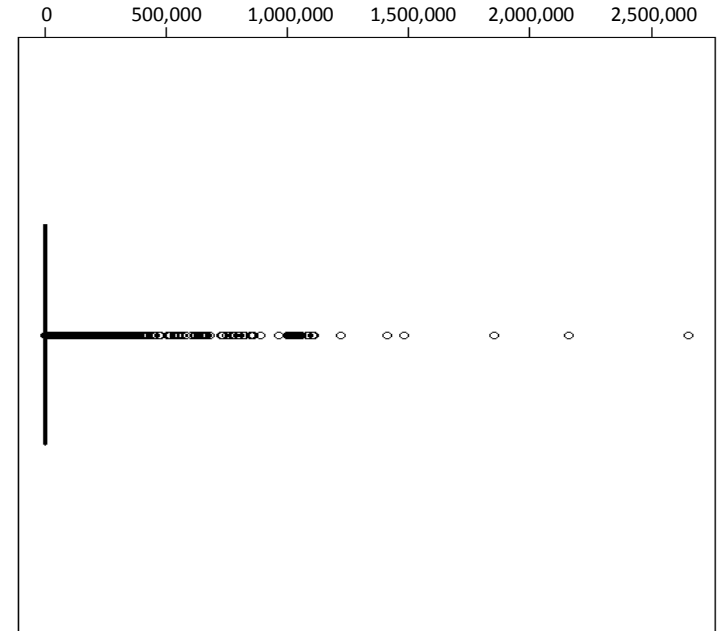
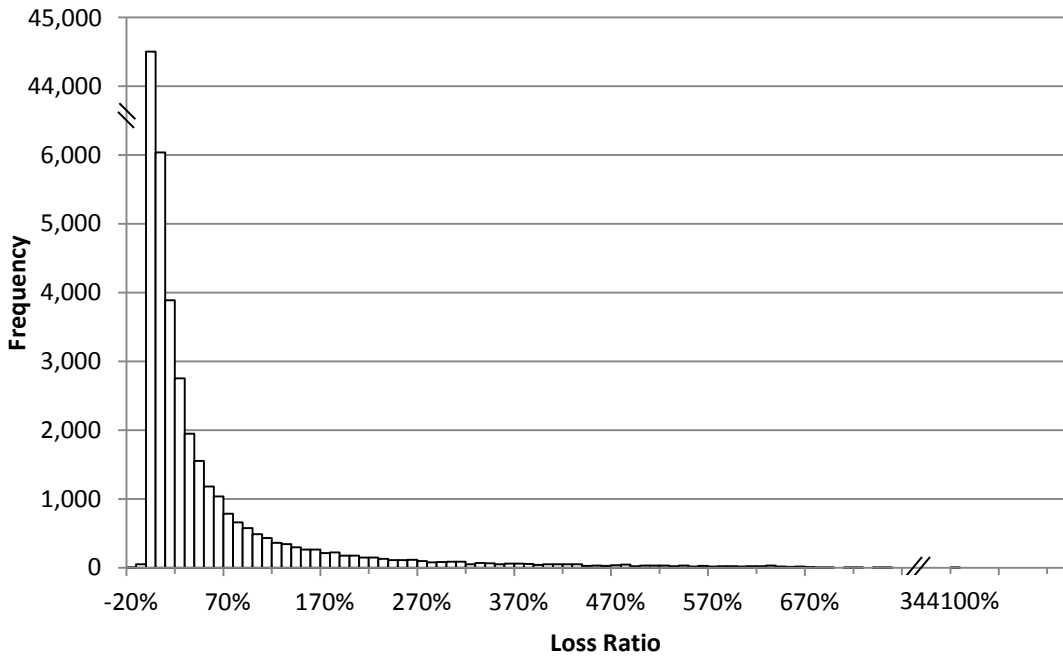
**NUMBER OF POLICIES**



■ Policy count	2,812	52,672	29,174	18,575	12,554	8,621	6,144	4,566	3,456	2,761	15,467	8,350
■ Premium(x10,000)	\$91	\$6,839	\$6,752	\$6,175	\$5,655	\$5,227	\$4,507	\$3,793	\$3,602	\$3,274	\$37,199	\$21,860
Percentage (Policy count)	1.70%	31.89%	17.67%	11.25%	7.60%	5.22%	3.72%	2.76%	2.09%	1.67%	9.37%	5.06%

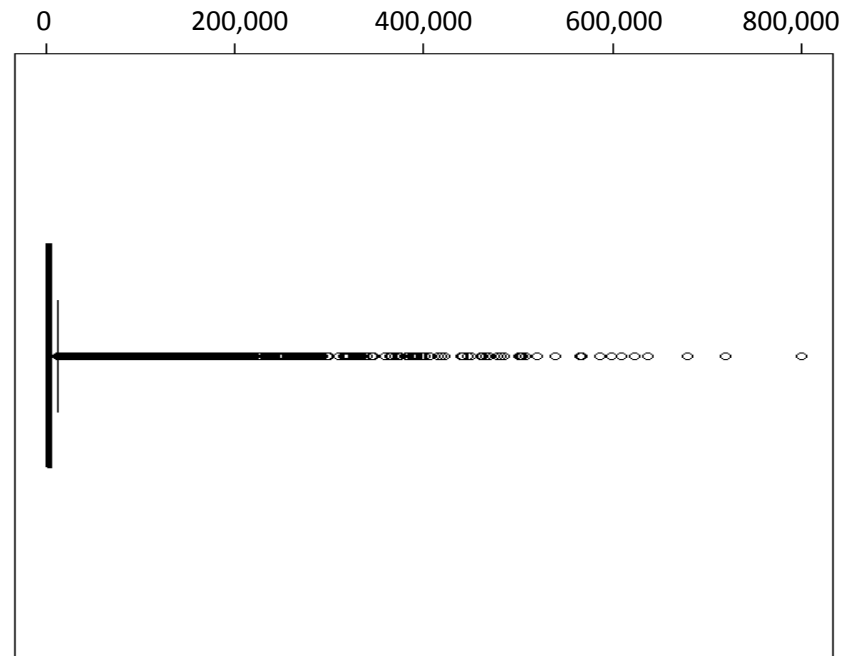
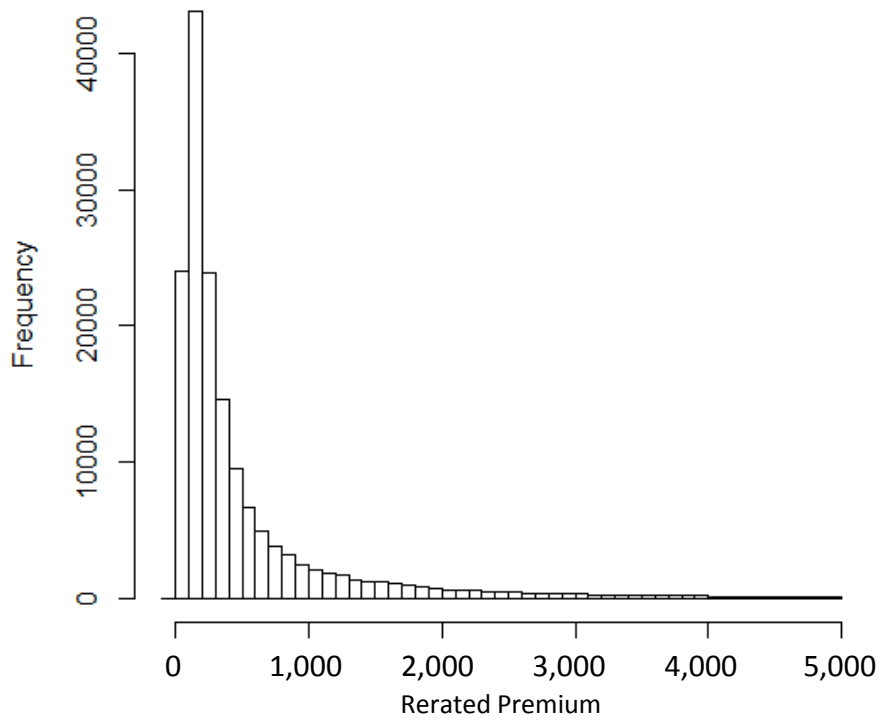
# EXPLORATORY ANALYSIS

## DISTRIBUTION OF INCURRED LOSSES



# EXPLORATORY ANALYSIS

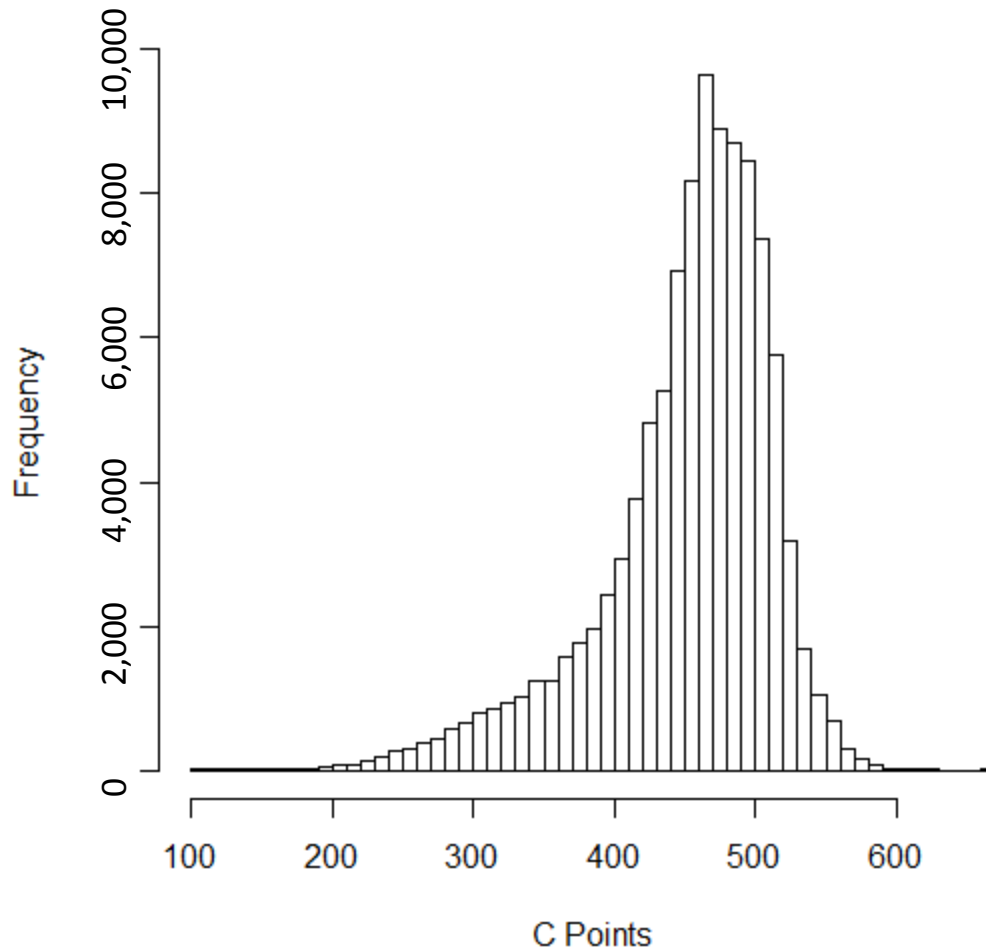
## DISTRIBUTION OF RERATED PREMIUM





# EXPLORATORY ANALYSIS

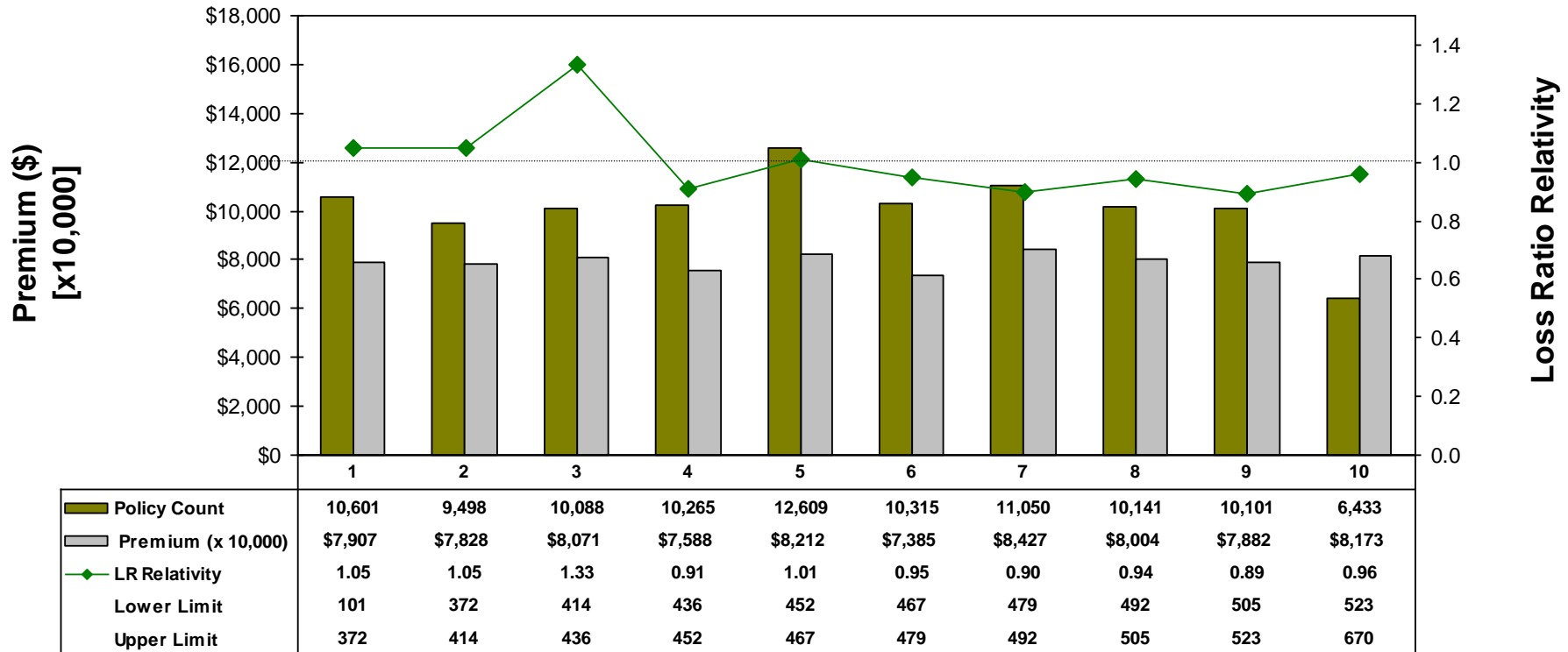
## DISTRIBUTION OF C POINTS (risk of credit default)



# EXPLORATORY ANALYSIS

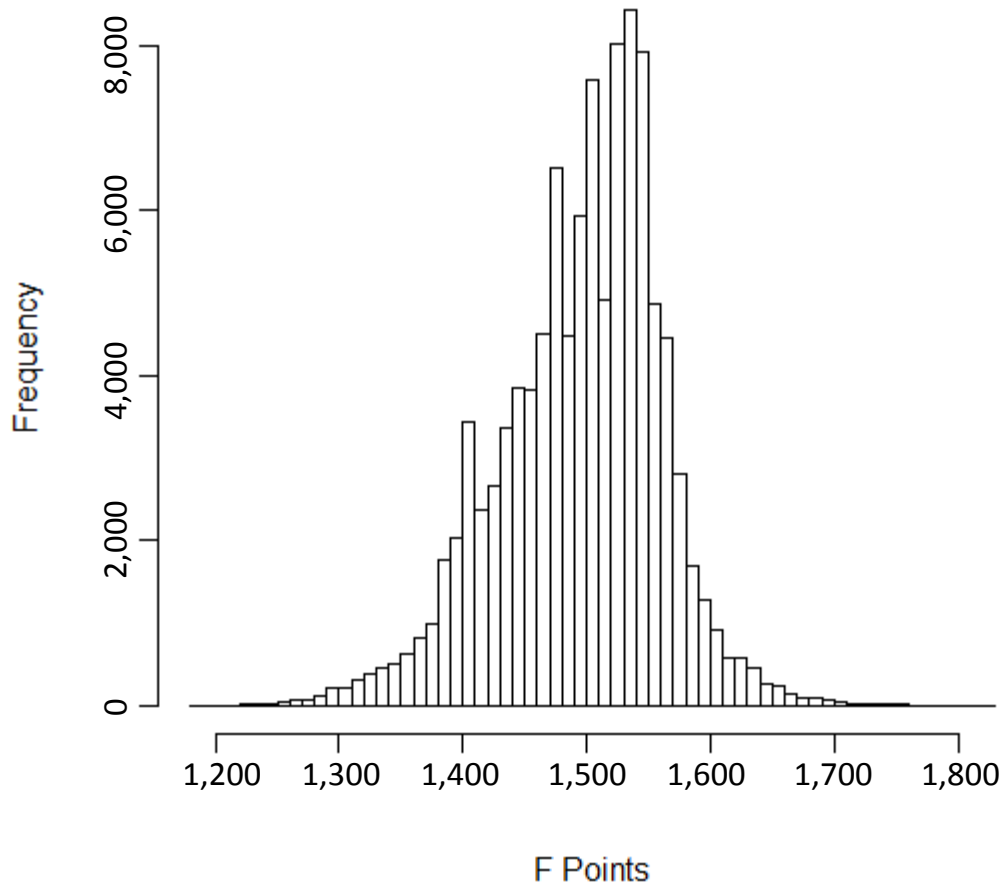
## C POINTS (risk of credit default)

### LR Relativities by C Points



# EXPLORATORY ANALYSIS

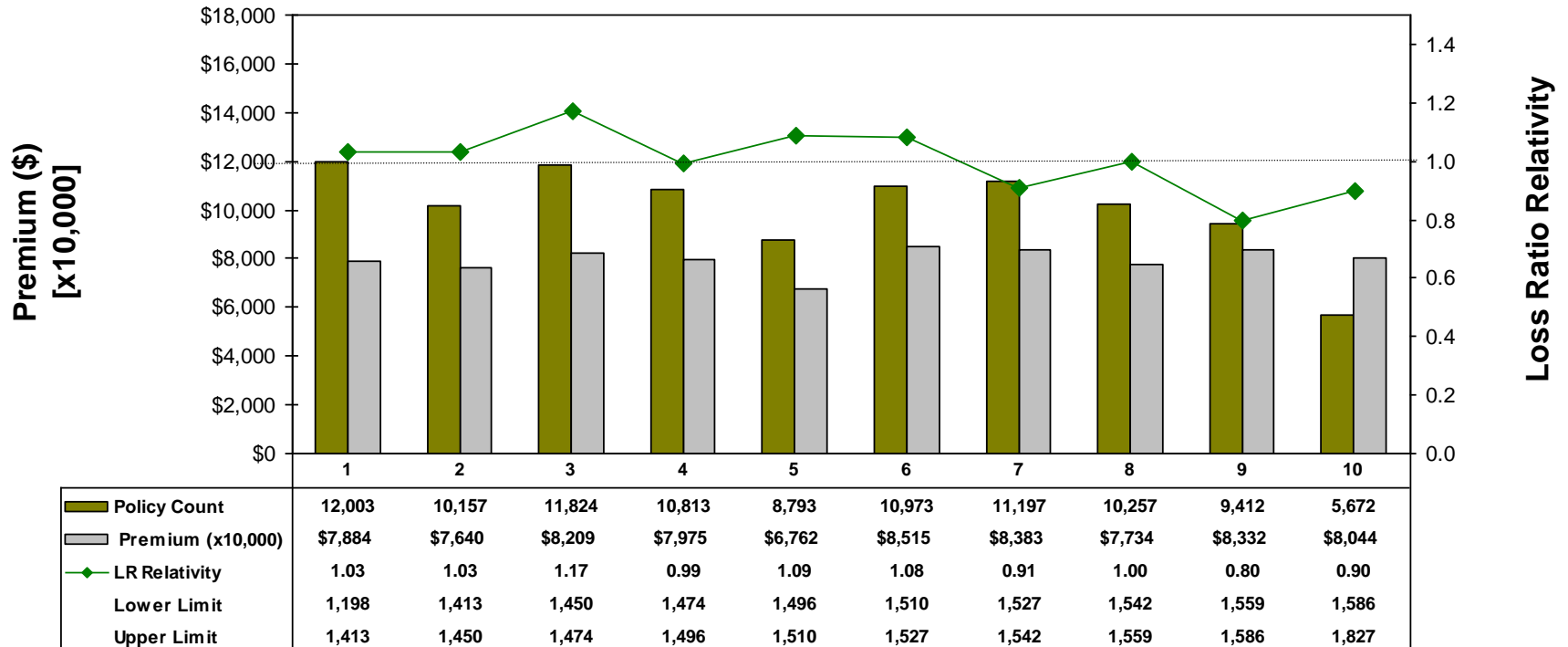
## DISTRIBUTION OF F POINTS (risk of financial stress)



# EXPLORATORY ANALYSIS

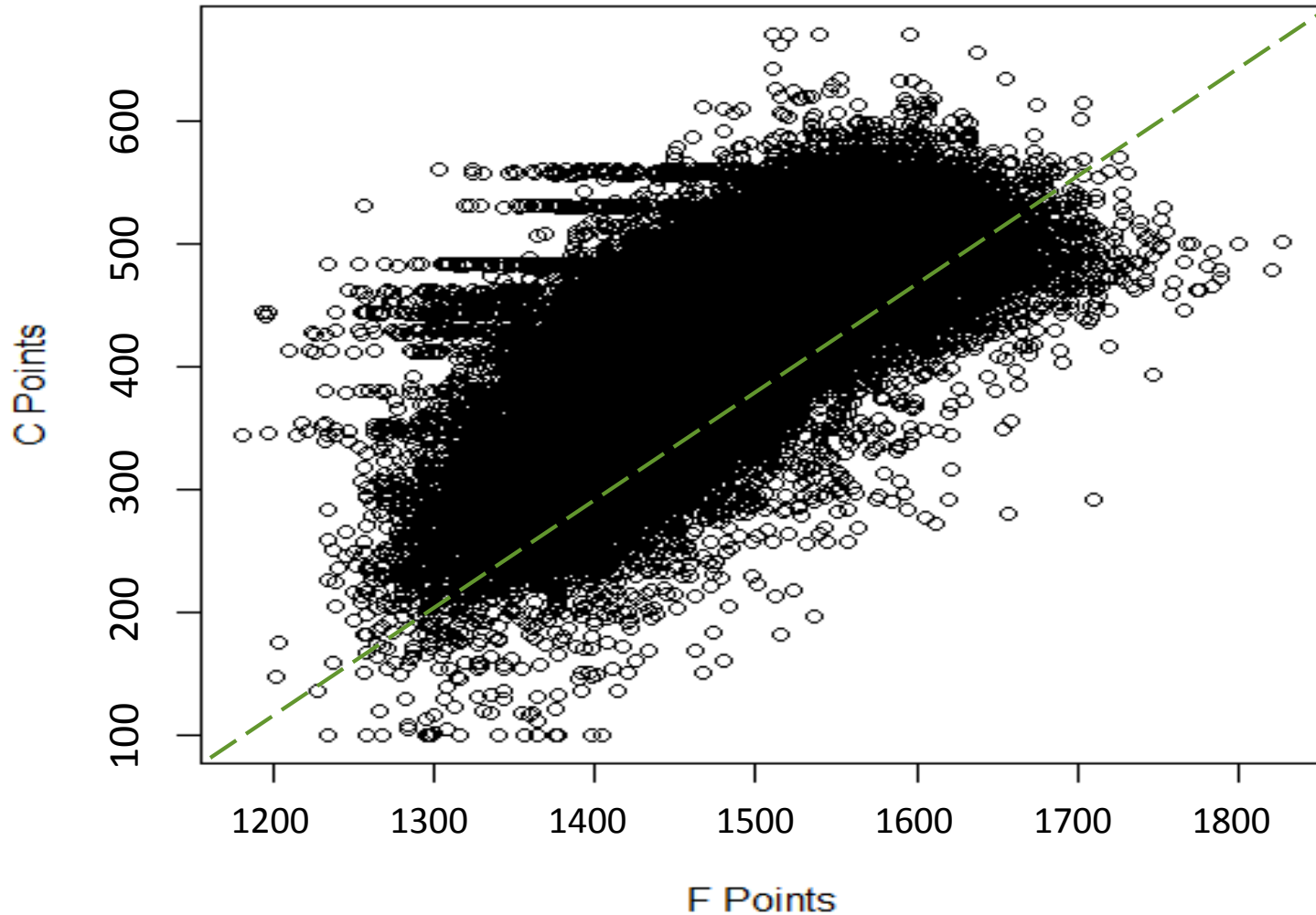
## F POINTS (risk of financial stress)

### LR Relativities by F Points



# EXPLORATORY ANALYSIS

## C POINTS VS. F POINTS – SCATTER PLOT



# EXPLORATORY ANALYSIS (SUMMARY)

---

- 35.3% of policies were missing credit score information.
- Noticed unexpected values for some variables:
  - Rerated premiums (below \$500, the minimum premium amount)
  - Incurred losses (below \$0)
- Many policies had data given as N/A: fleet size, liability group, policy type and incurred loss.
- 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
- Future Considerations
- Acknowledgements

# DATA PREPARATION (INITIAL STEPS)

1. We began with 165,142 policies
2. We removed the following
  - 5,728 policies with no premium
  - 58,309 policies with no credit information (“no-hits”)
  - 5,599 manually written policies
  - 330 policies with market segment of “Other” (207) and policies with no liability coverage (123)
  - 3 duplicated records

	Original data	Remove No-premiums	Remove No-hits	Cleaned data
Incurred Loss Ratio	36.02%	34.09%	36.59%	36.72%
Number of Policies	165,142	159,414	101,105	95,173



# DATA PREPARATION

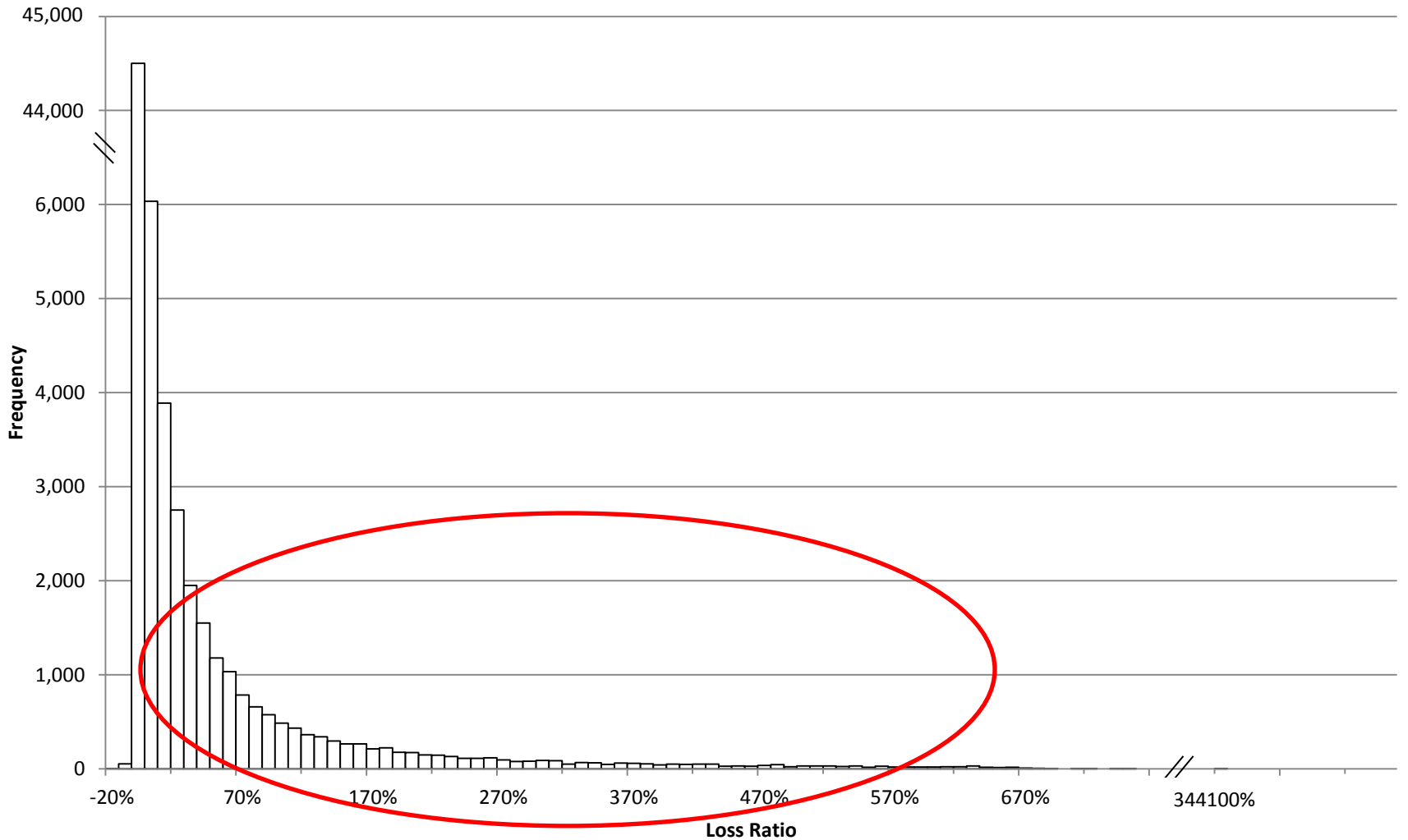
Once we reduced the data to 95,173 policies, we made additional adjustments:

1. A minimum \$500 premium was imposed, impacting 4,514 policies.
2. Policies with negative losses were set to zero, impacting 51 policies.
3. The top 1% of incurred loss ratios were capped to the 99<sup>th</sup> percentile (622%), impacting 952 policies.

	LR before any adjustments	\$500 minimum premium	No negative losses	Capping top 1% of LR's
Incurred Loss Ratio	36.72%	36.65%	36.65%	27.57%

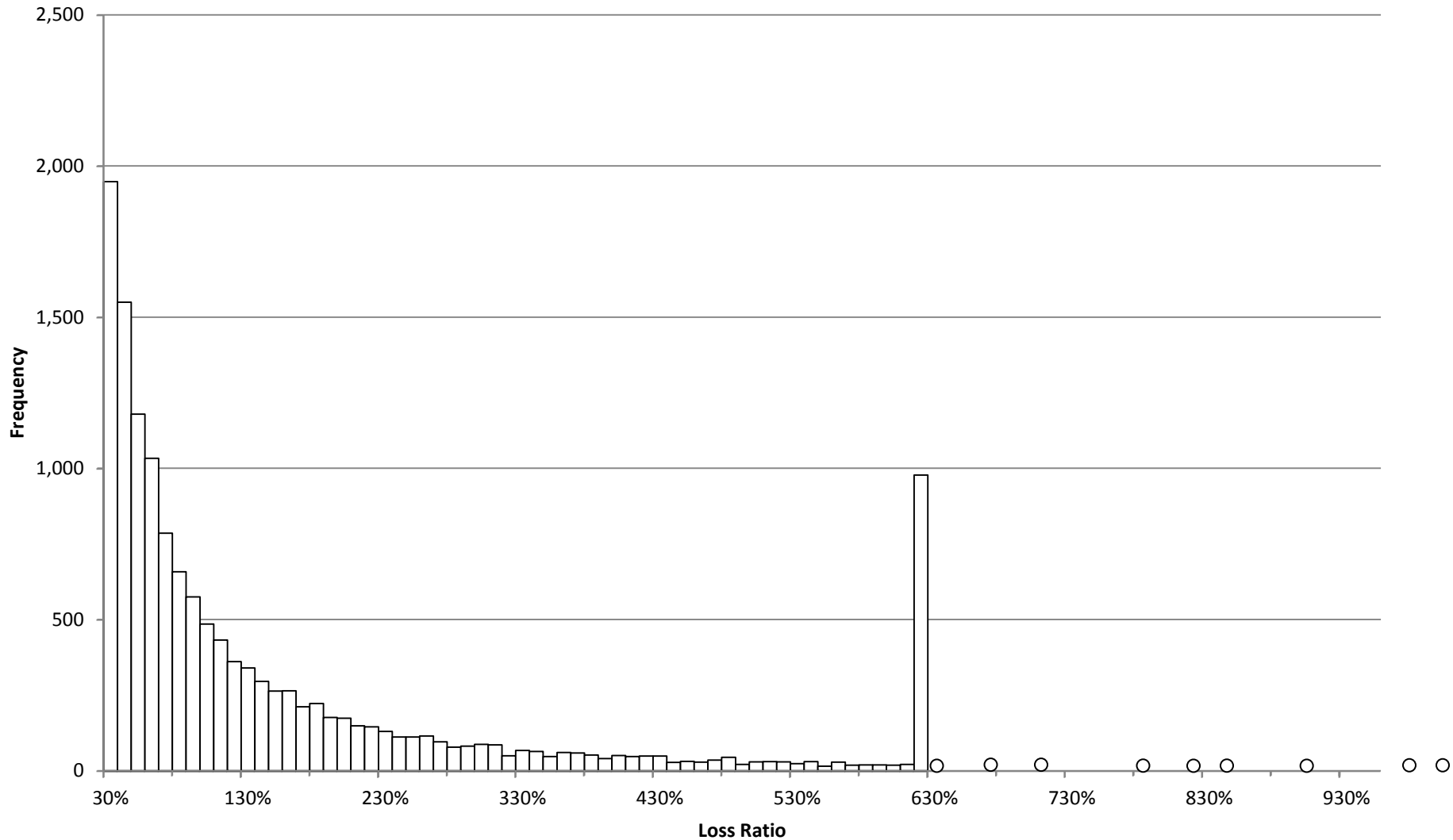
# DATA PREPARATION

## LOSS RATIO DISTRIBUTION - UNADJUSTED



# DATA PREPARATION

## LOSS RATIO DISTRIBUTION (ZOOMED)





# DATA PREPARATION

---

- After cleaning and adjusting the data, we had to determine the appropriate variables to use in the model.
- Based on exploratory analysis, we found that C points and F points were correlated which had to be corrected.
- Factor Analysis:
  - enabled analysis of multi-collinearity among variables.
  - an uncorrelated factor was created as a weighted sum of the standardized variables (C Points and F Points).
  - this factor was named Financial Stability and used as an input variable.

# DATA PREPARATION (SUMMARY)

---

1. Deleted policies with missing, invalid or inconsistent data (e.g., policies with no credit information, missing rerated premium and manually written policies).
  
2. Data Adjustments:
  - Negative incurred losses were set to zero
  - Incurred loss ratios were capped
  - Rerated premiums were set to at least \$500
  
3. Used factor analysis to create a new credit variable (Financial Stability) and assigned this variable to 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
- Future Considerations
- Acknowledgements

# MODEL DESIGN

---

Response variable: Incurred loss ratio

Model distribution: Tweedie (member of exponential family)

Model type: Generalized linear model (GLM)

Input variables:

- Policy type
- Business type
- Market segment
- Fleet size
- Financial stability
- Policy effective age (current year minus effective year)



# MODEL DESIGN

$$\log(E[\text{loss ratio}]) = \beta_0 + \beta_1 \times \text{Policy Type} + \beta_2 \times \text{Fleet Size} + \beta_3 \times \text{Financial Stability} + \dots + \varepsilon$$

## 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  $\log(\mu)$  where  $\mu$  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
- Future Considerations
- Acknowledgements

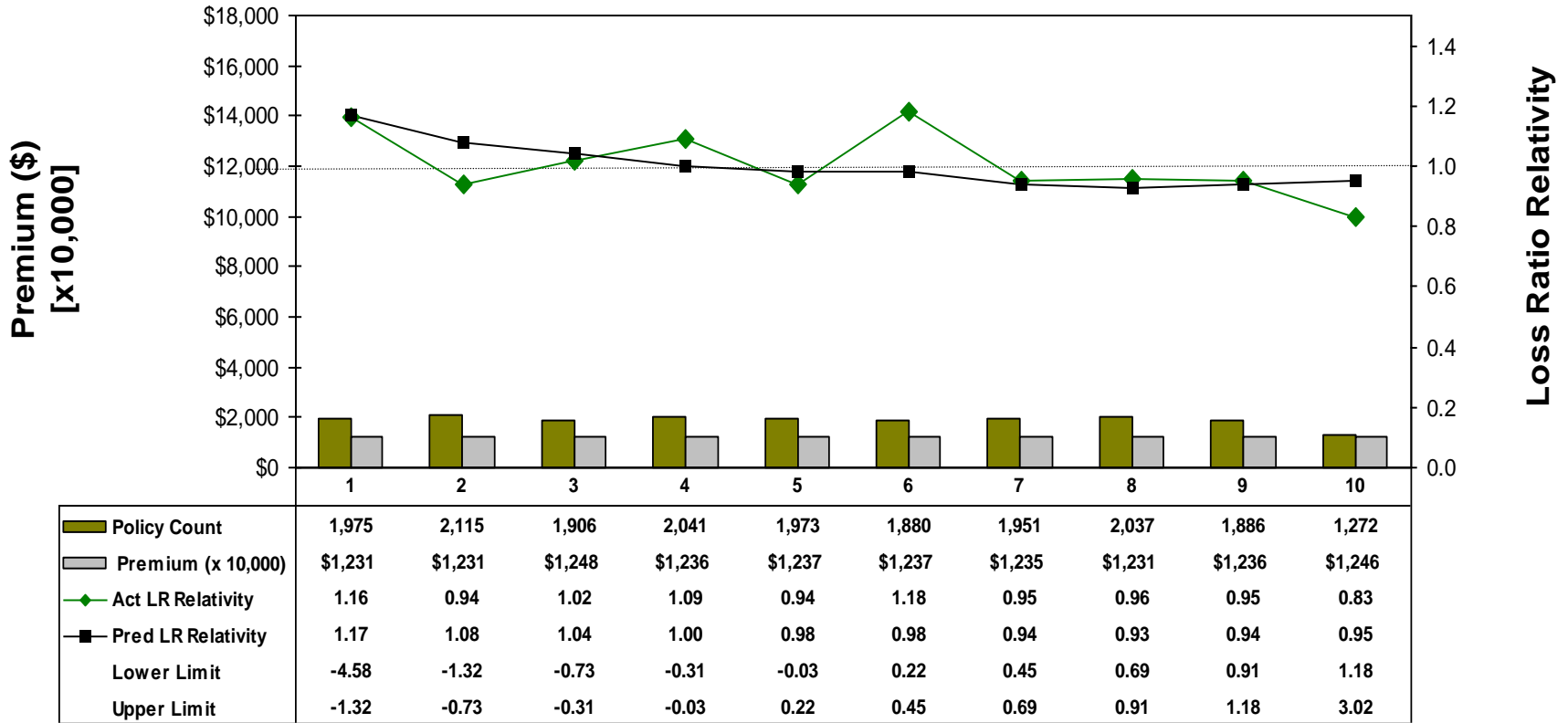
# CREDIT MODEL LIFT BY FINANCIAL STABILITY (BUILD SAMPLE)

## Actual versus Predicted LR Relativities by Financial Stability



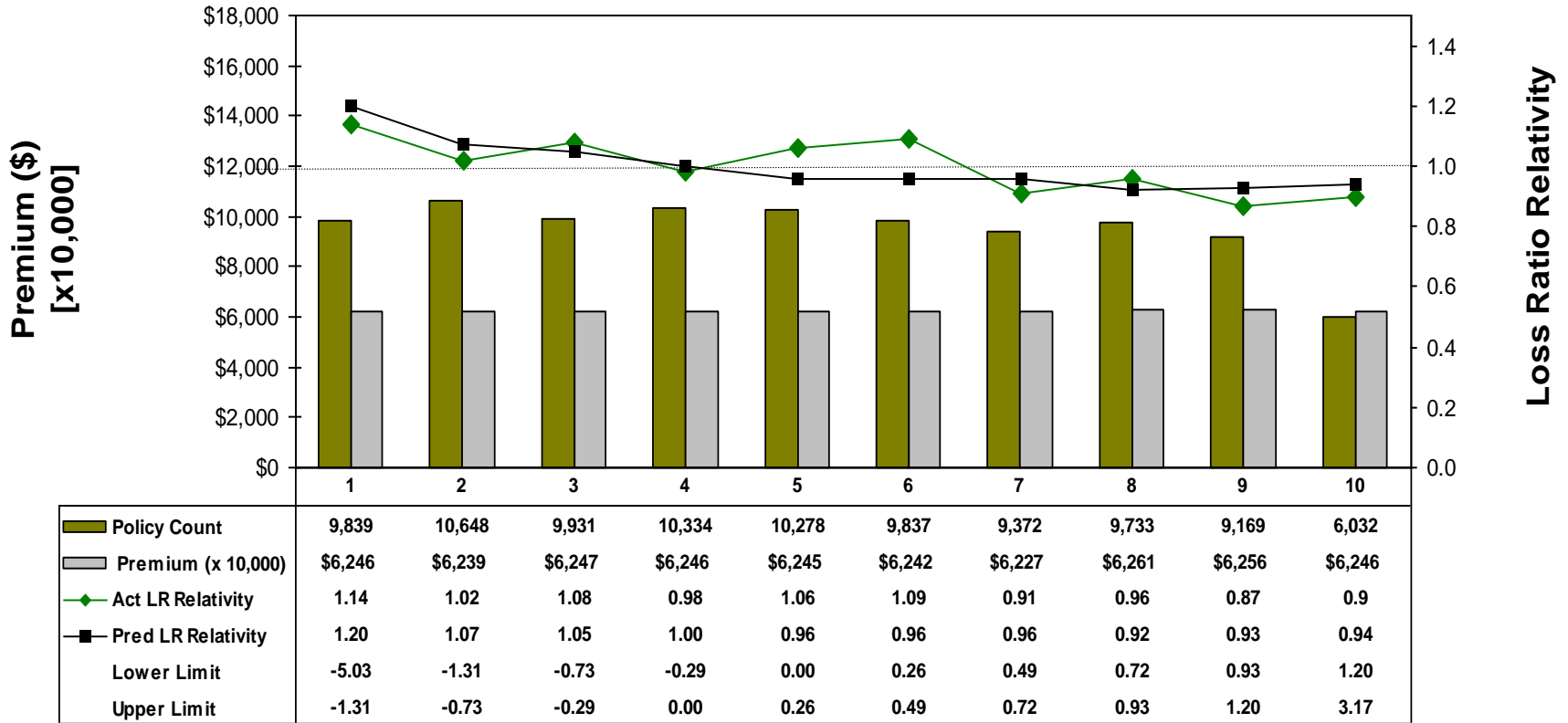
# CREDIT MODEL LIFT BY FINANCIAL STABILITY (VALIDATION SAMPLE)

Actual versus Predicted LR Relativities by Financial Stability



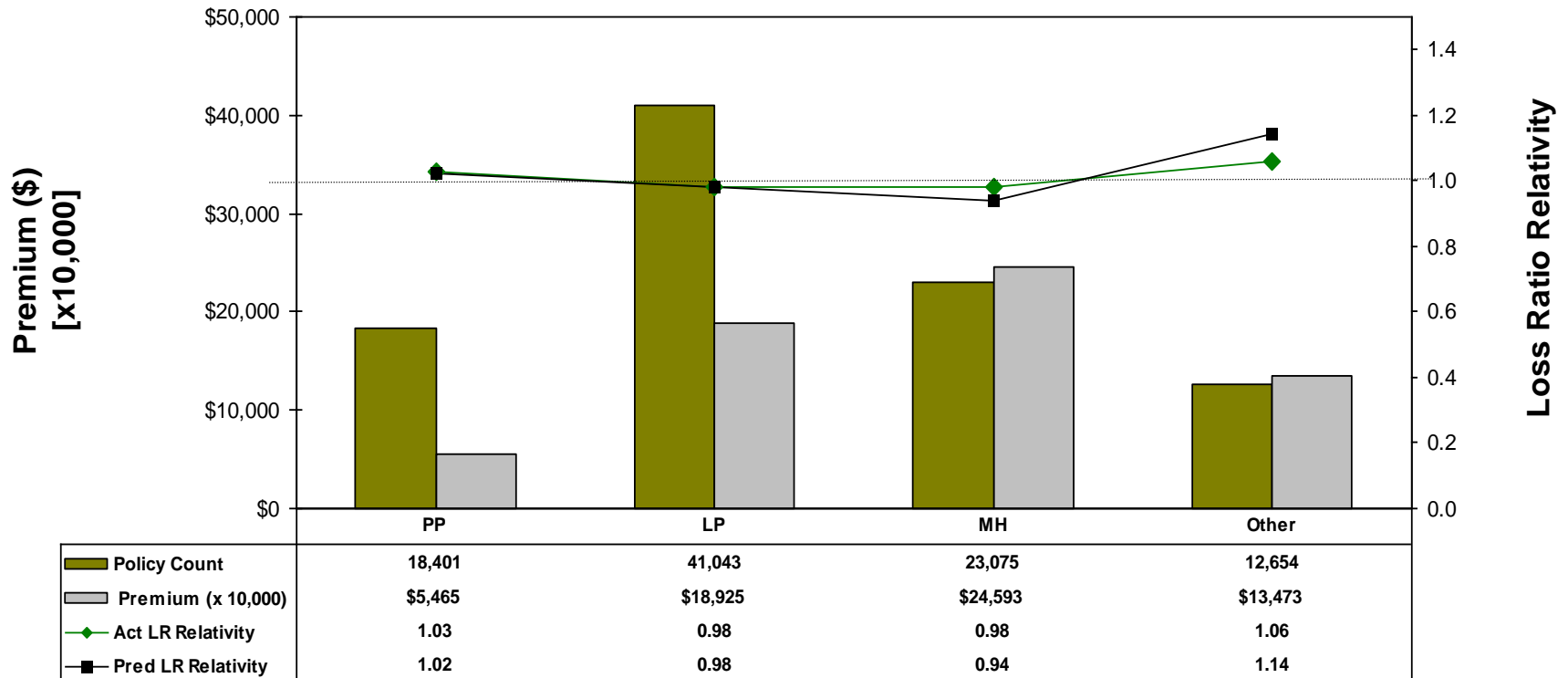
# CREDIT MODEL LIFT BY FINANCIAL STABILITY (FULL SAMPLE)

## Actual versus Predicted LR Relativities by Financial Stability



# CREDIT MODEL LIFT BY POLICY TYPE (FULL SAMPLE)

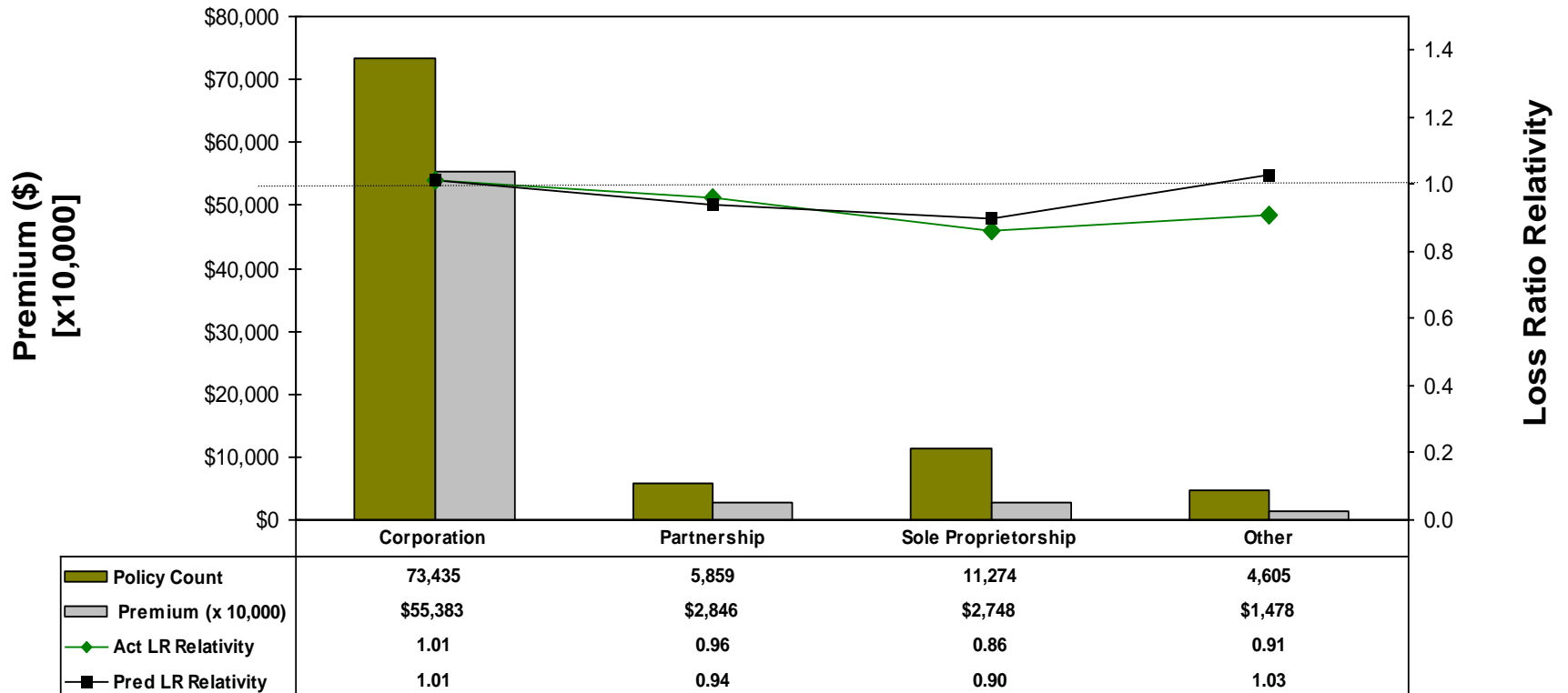
Actual versus Predicted LR Relativities by Policy Type





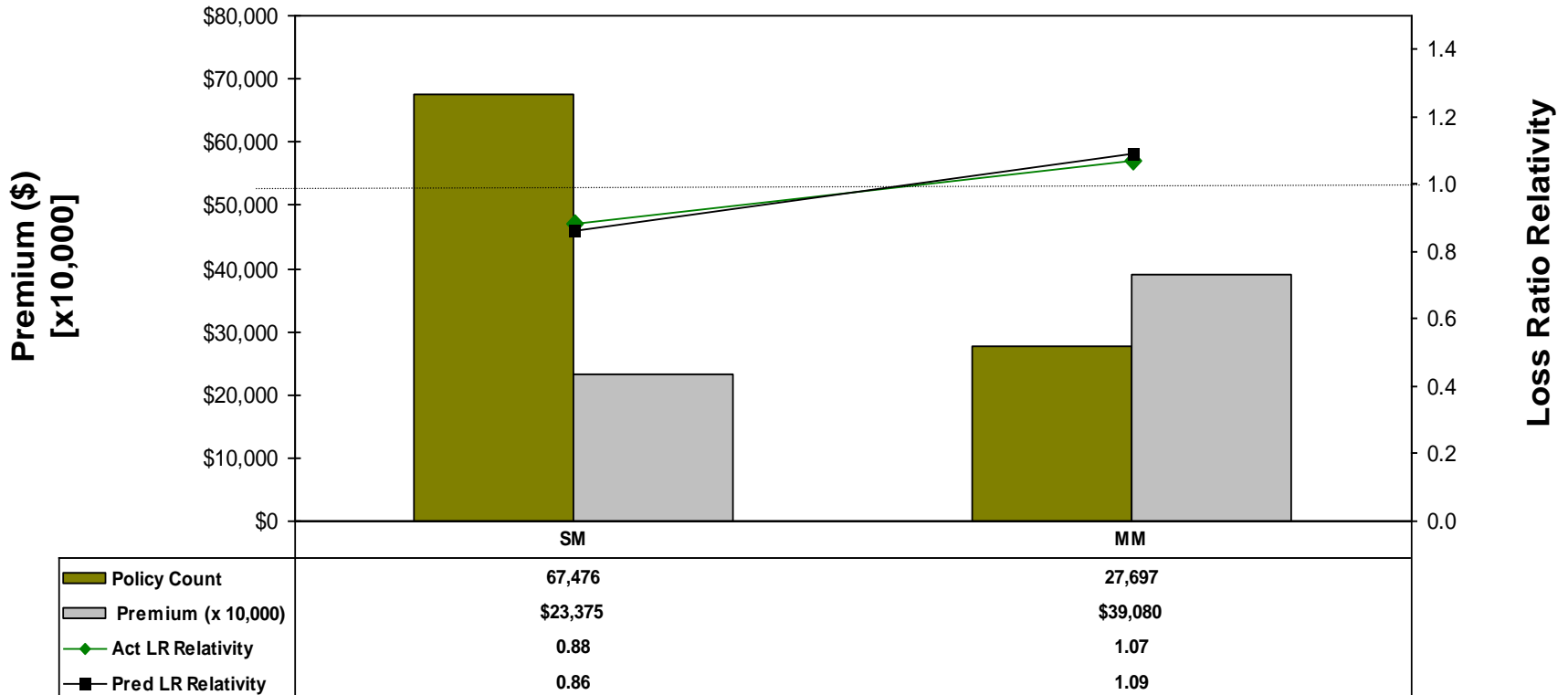
# CREDIT MODEL LIFT BY BUSINESS TYPE (FULL SAMPLE)

Actual versus Predicted LR Relativities by Business Type



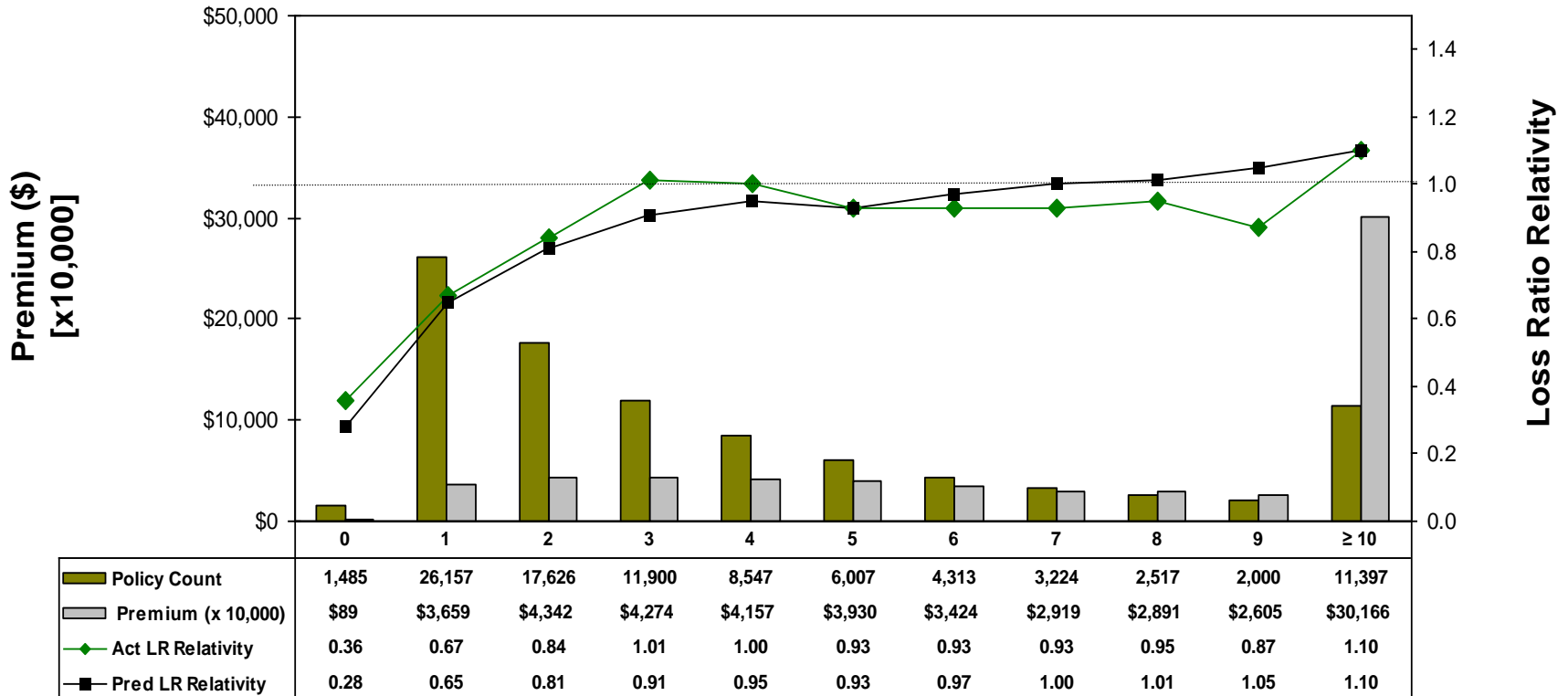
# CREDIT MODEL LIFT BY MARKET SEGMENT (FULL SAMPLE)

Actual versus Predicted LR Relativities by Market Segment



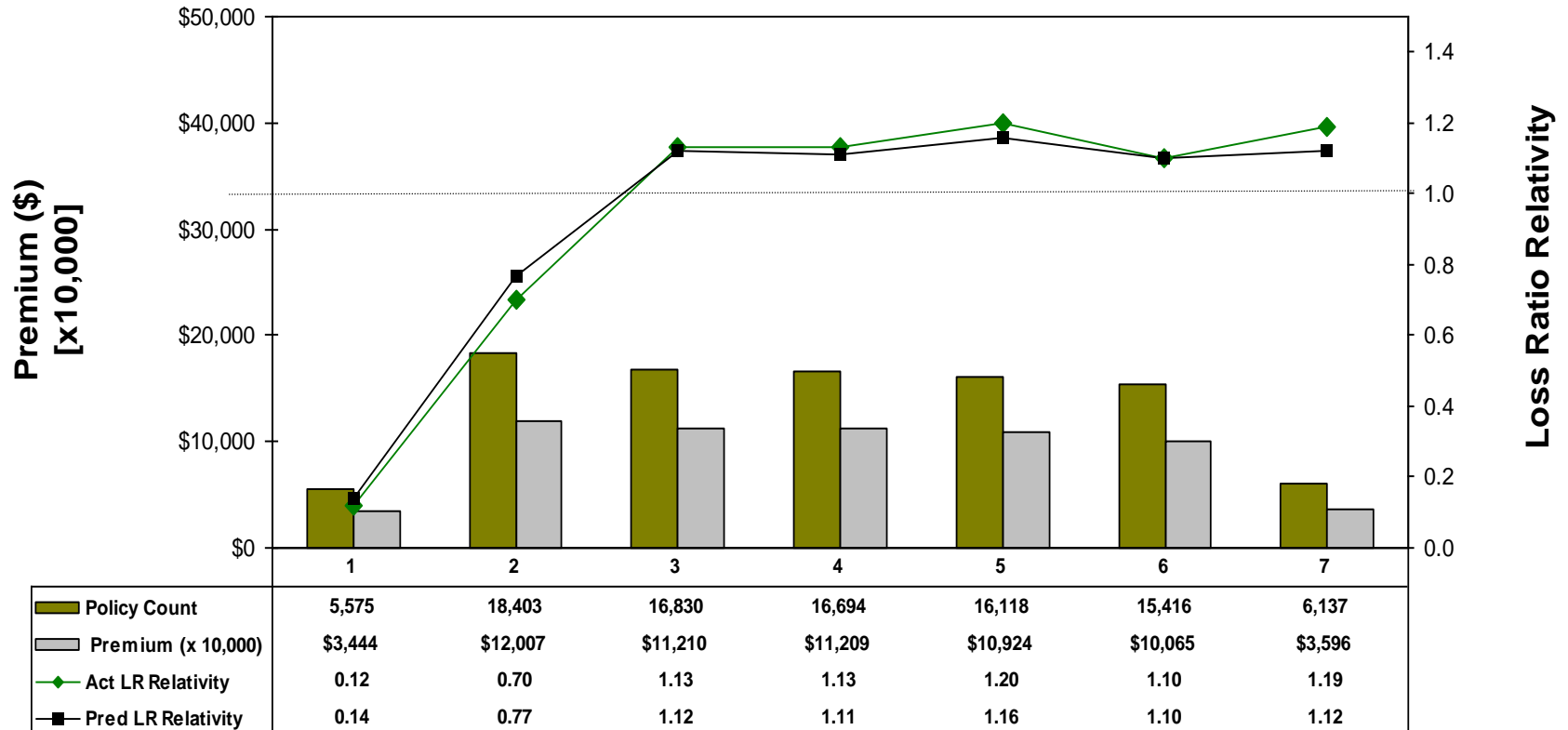
# CREDIT MODEL LIFT BY FLEET SIZE (FULL SAMPLE)

Actual versus Predicted LR Relativities by Fleet Size



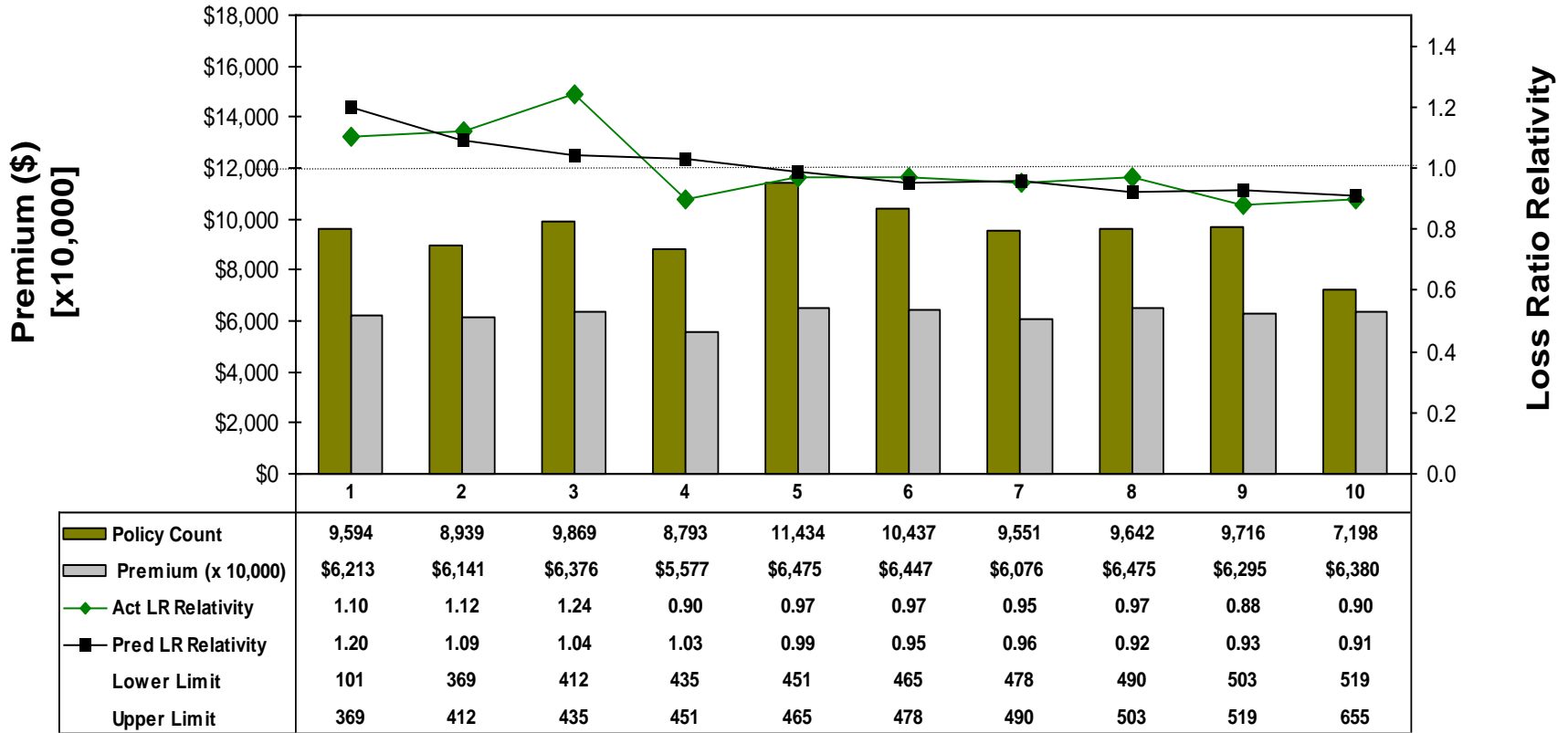
# CREDIT MODEL LIFT BY POLICY EFFECTIVE AGE (FULL SAMPLE)

Actual versus Predicted LR Relativities by Policy Effective Age



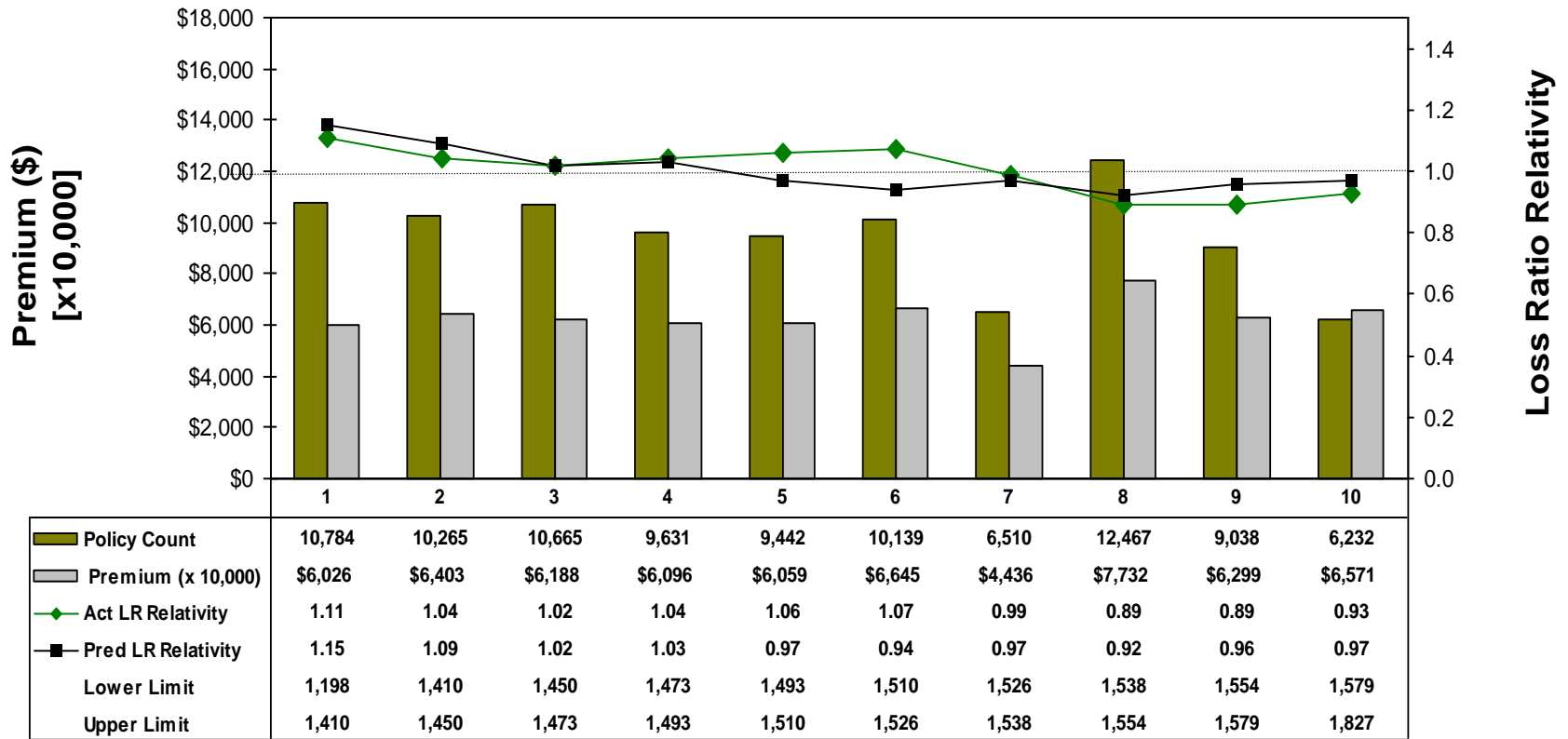
# CREDIT MODEL LIFT BY C POINTS (FULL SAMPLE)

## Actual versus Predicted LR Relativities by C Points



# CREDIT MODEL LIFT BY F POINTS (FULL SAMPLE)

## Actual versus Predicted LR Relativities by F Points



# PRESENTATION OUTLINE

---

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

# CONCLUSIONS

---

- The credit variable Financial Stability is a 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 and pricing 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
- **Future Considerations**
- Acknowledgements

# FUTURE CONSIDERATIONS

---

There are adjustments that could be made to improve the usability of the model:

## Creating Credit Factors

- Break down Financial Stability scores into predicted loss ratio groups.
- Assign each tier a credit factor for application in pricing.

## Policy Underwriting

- Determine rules for policy to be referred or declined.
- Implement corresponding rules.

## Options for handling No-Hits

1. Do not change the policies; assign them a credit factor of 1.
2. Take average of the actual loss ratios for no-hits and determine appropriate credit factor range.

# PRESENTATION OUTLINE

---

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

Jaris Wicklund

Andrew Evans

Marc Cournoyer

Professor Jon Abraham

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

THANK YOU!

---

# QUESTIONS?

# APPENDIX

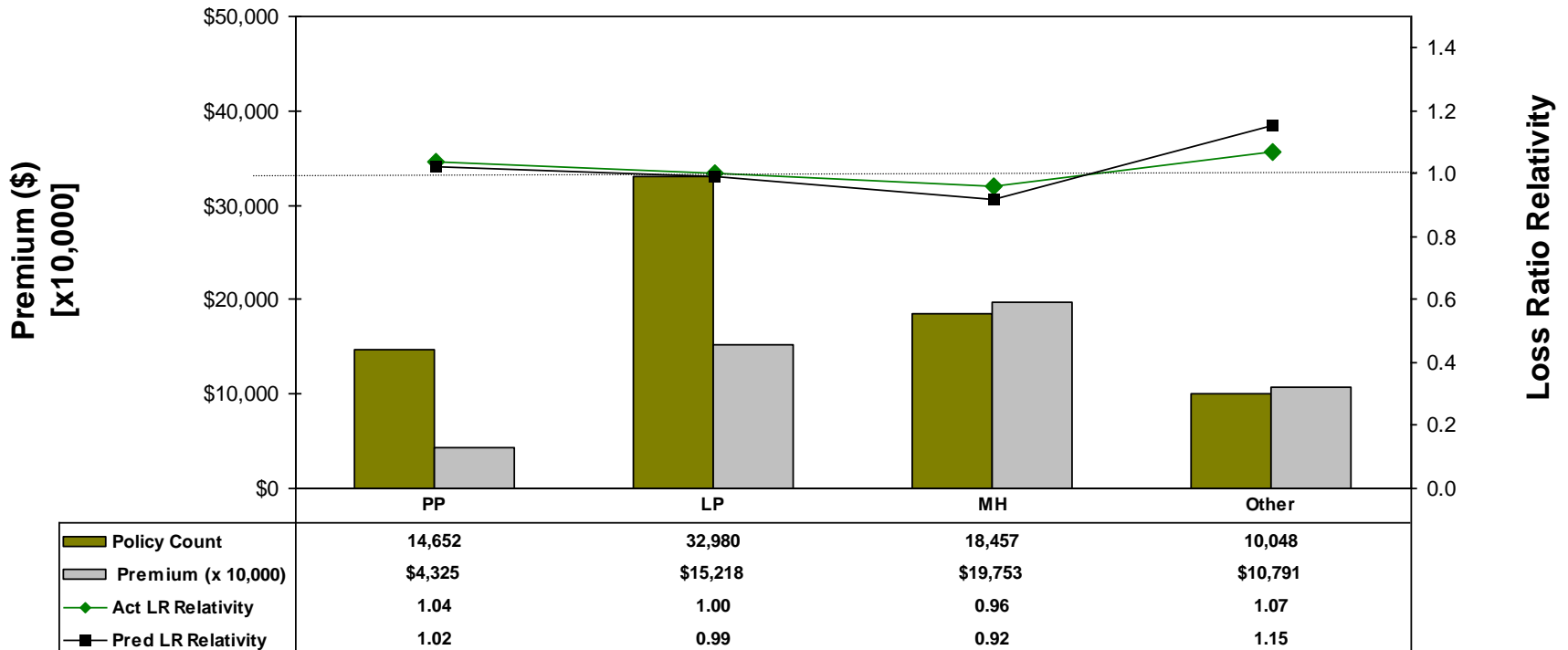
---

This Section is broken up into the following sections:

- Lift Charts (Build Sample, Validation Sample and Full Sample)
- Final Model Coefficient results

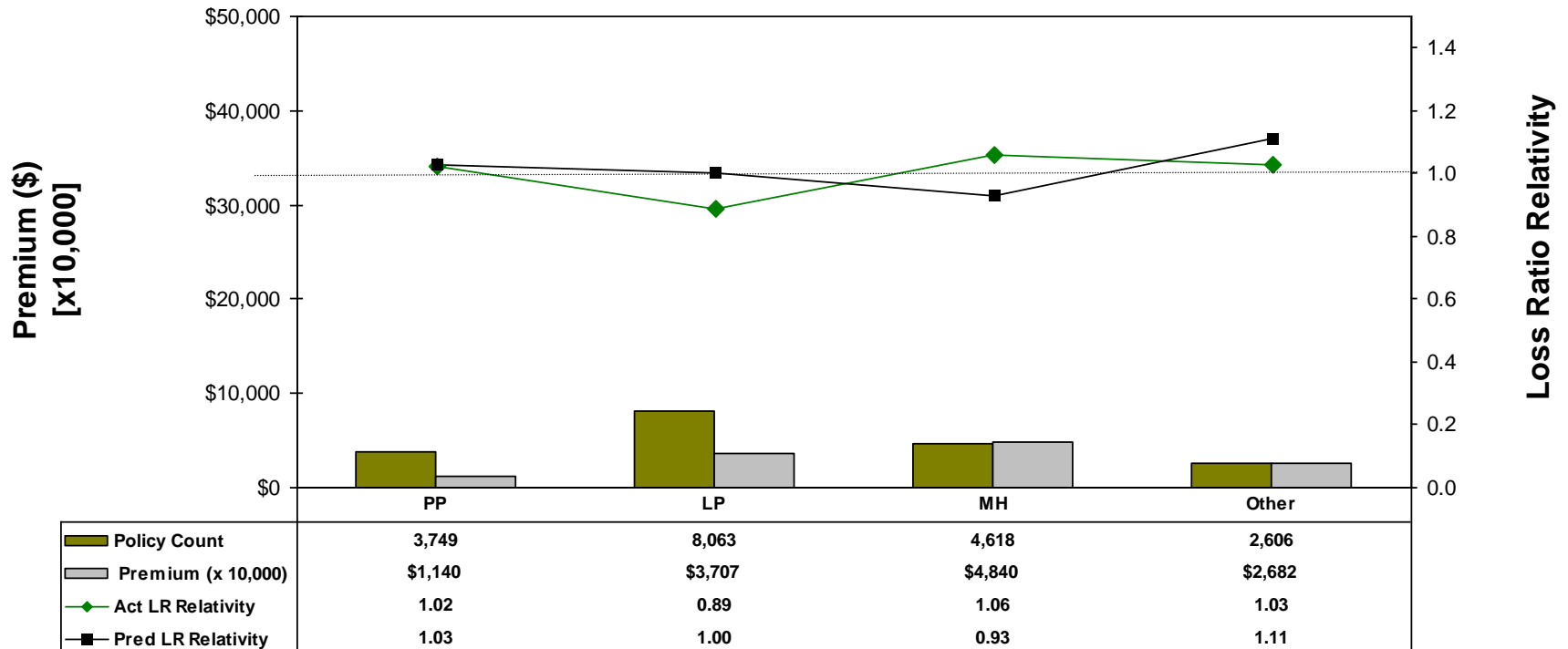
# CREDIT MODEL LIFT BY POLICY TYPE (BUILD SAMPLE)

Actual versus Predicted LR Relativities by Policy Type



# CREDIT MODEL LIFT BY POLICY TYPE (VALIDATION SAMPLE)

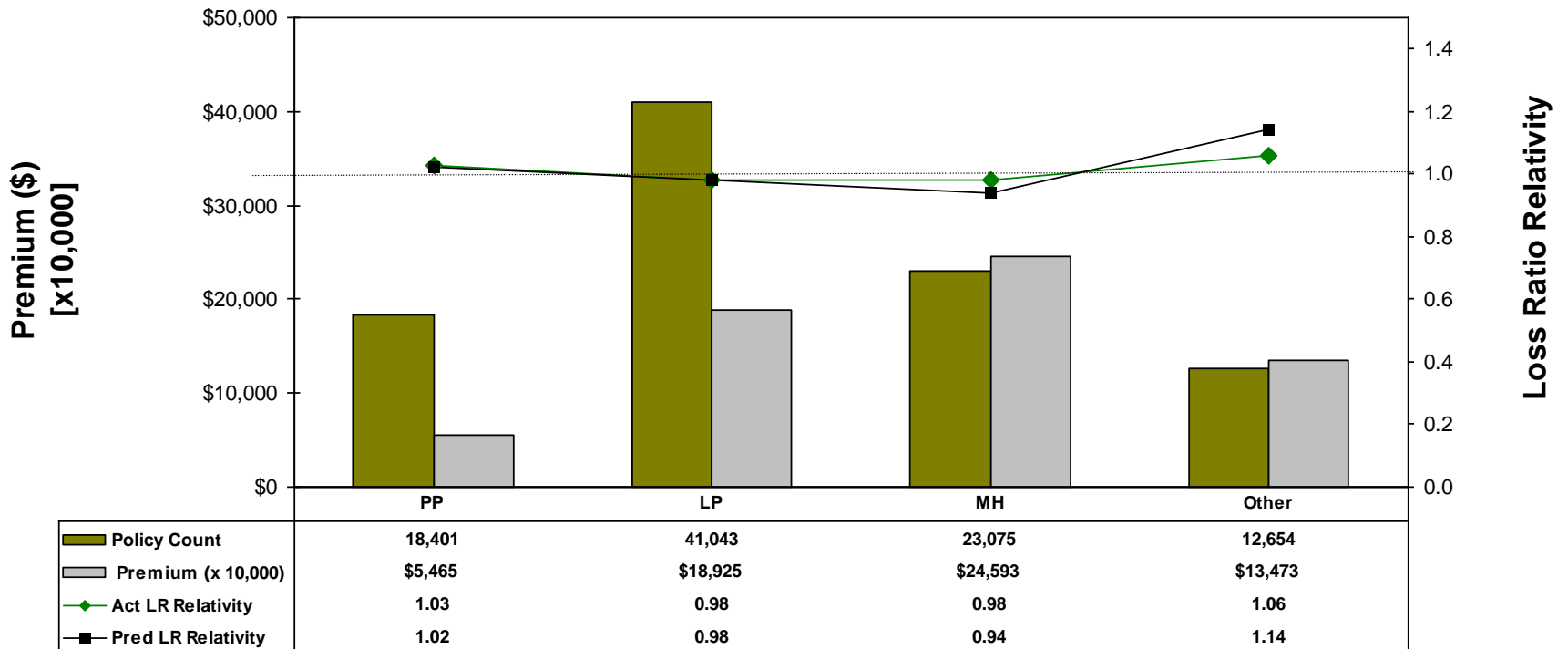
Actual versus Predicted LR Relativities by Policy Type





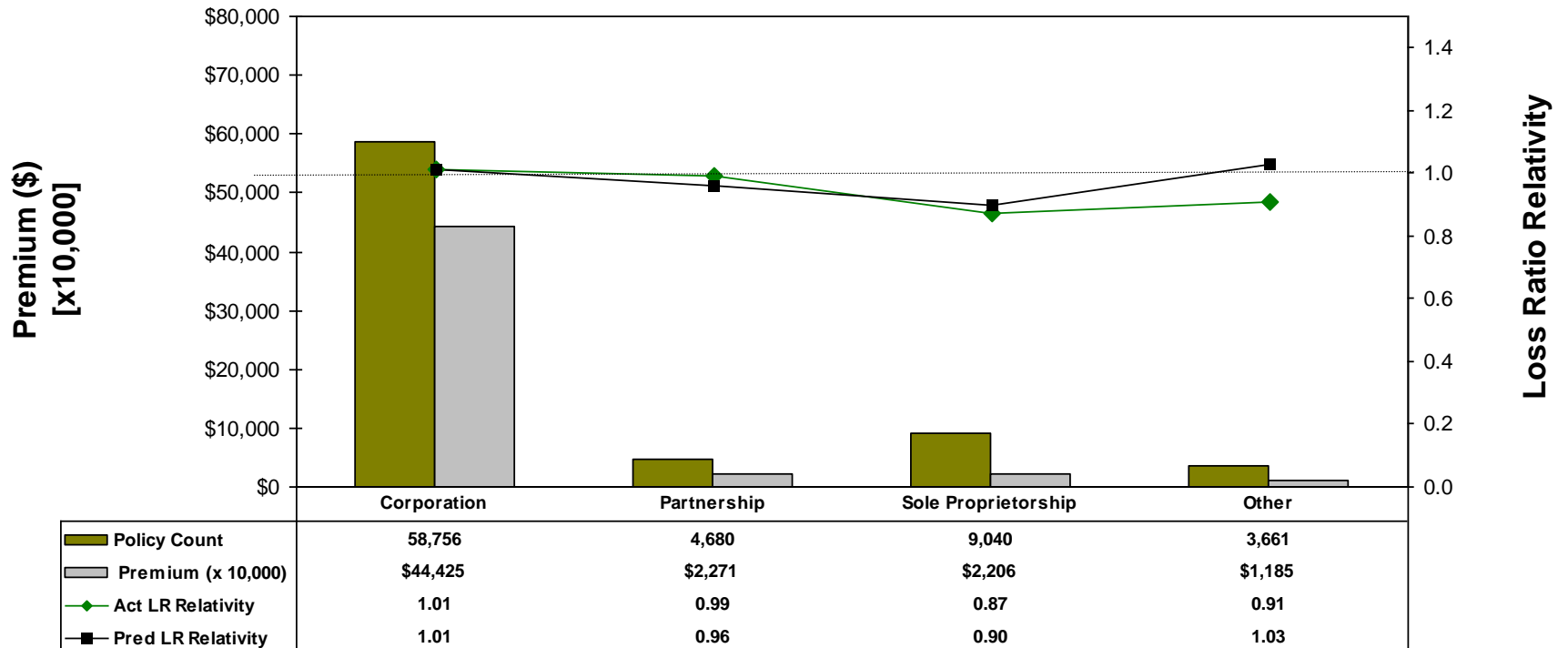
# CREDIT MODEL LIFT BY POLICY TYPE (FULL SAMPLE)

Actual versus Predicted LR Relativities by Policy Type



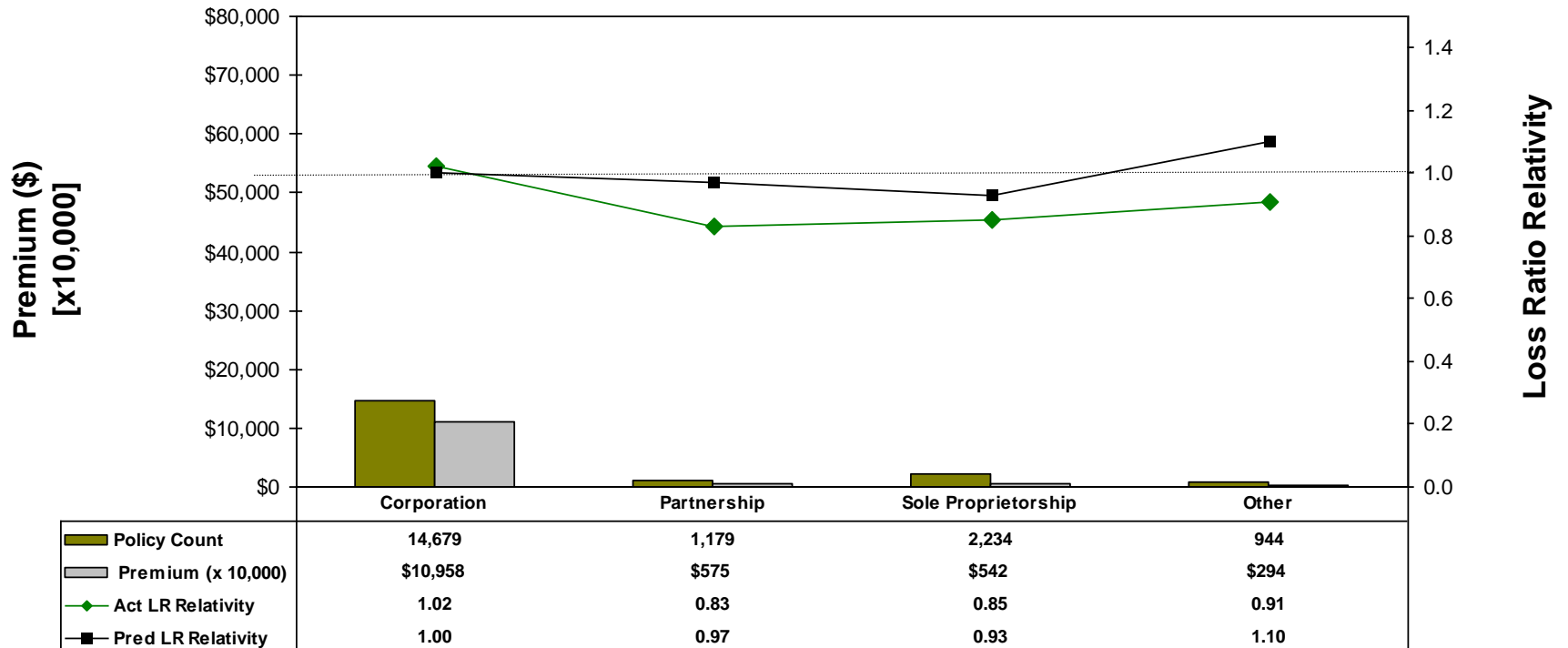
# CREDIT MODEL LIFT BY BUSINESS TYPE (BUILD SAMPLE)

Actual versus Predicted LR Relativities by Business Type



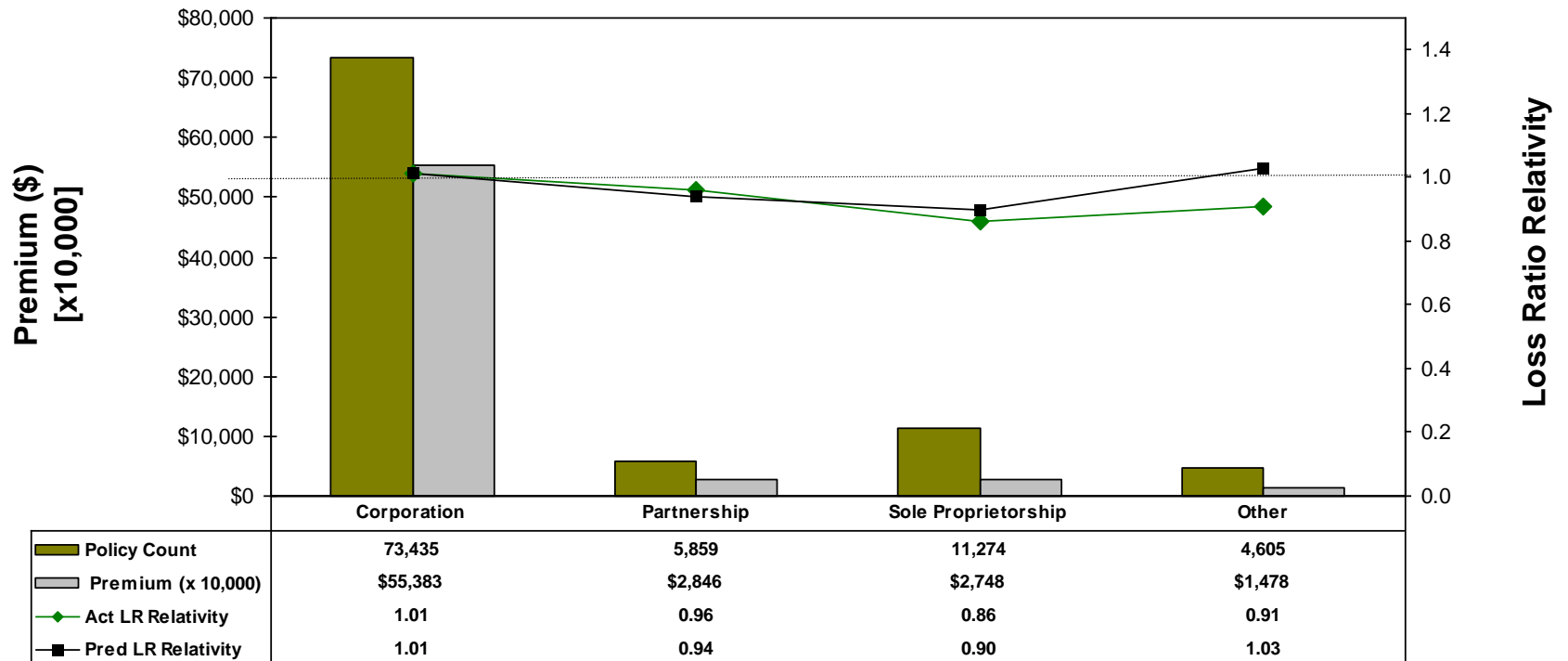
# CREDIT MODEL LIFT BY BUSINESS TYPE (VALIDATION SAMPLE)

Actual versus Predicted LR Relativities by Business Type



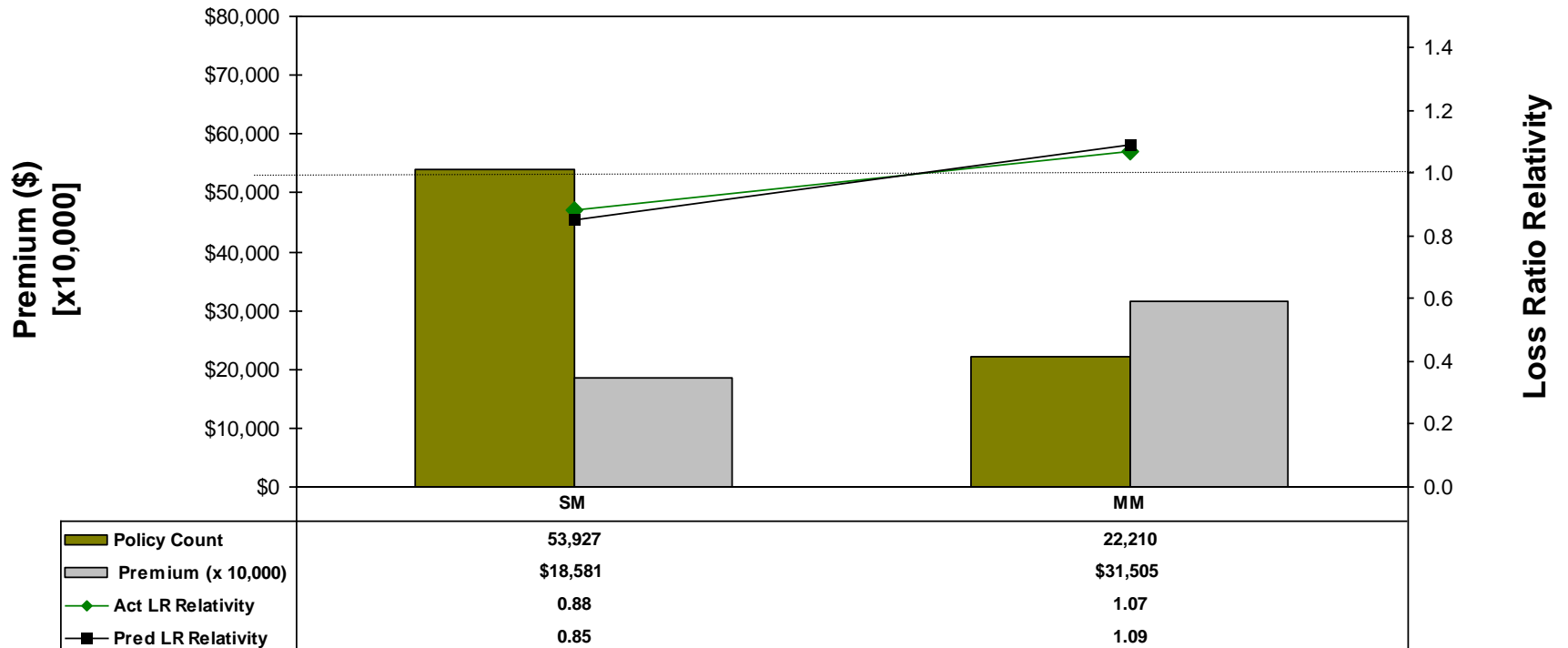
# CREDIT MODEL LIFT BY BUSINESS TYPE (FULL SAMPLE)

Actual versus Predicted LR Relativities by Business Type



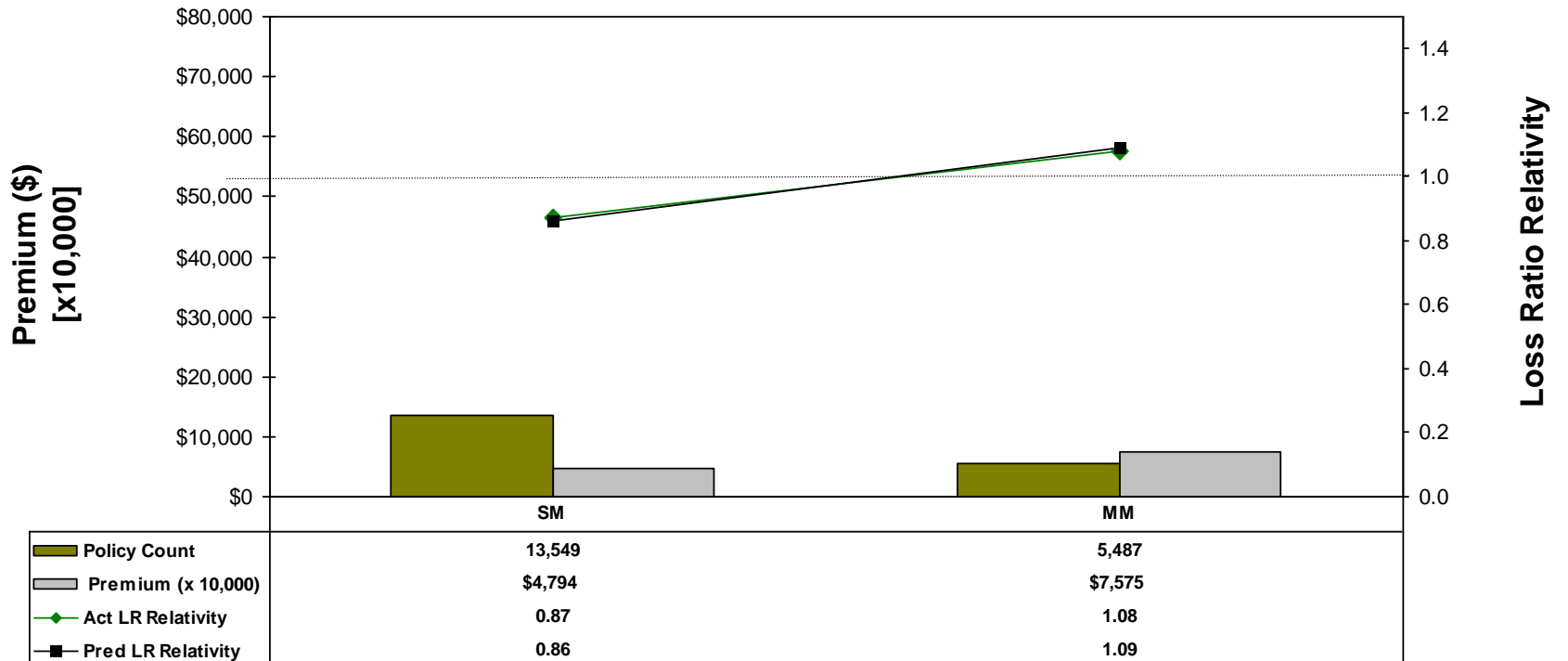
# CREDIT MODEL LIFT BY MARKET SEGMENT (BUILD SAMPLE)

Actual versus Predicted LR Relativities by Market Segment



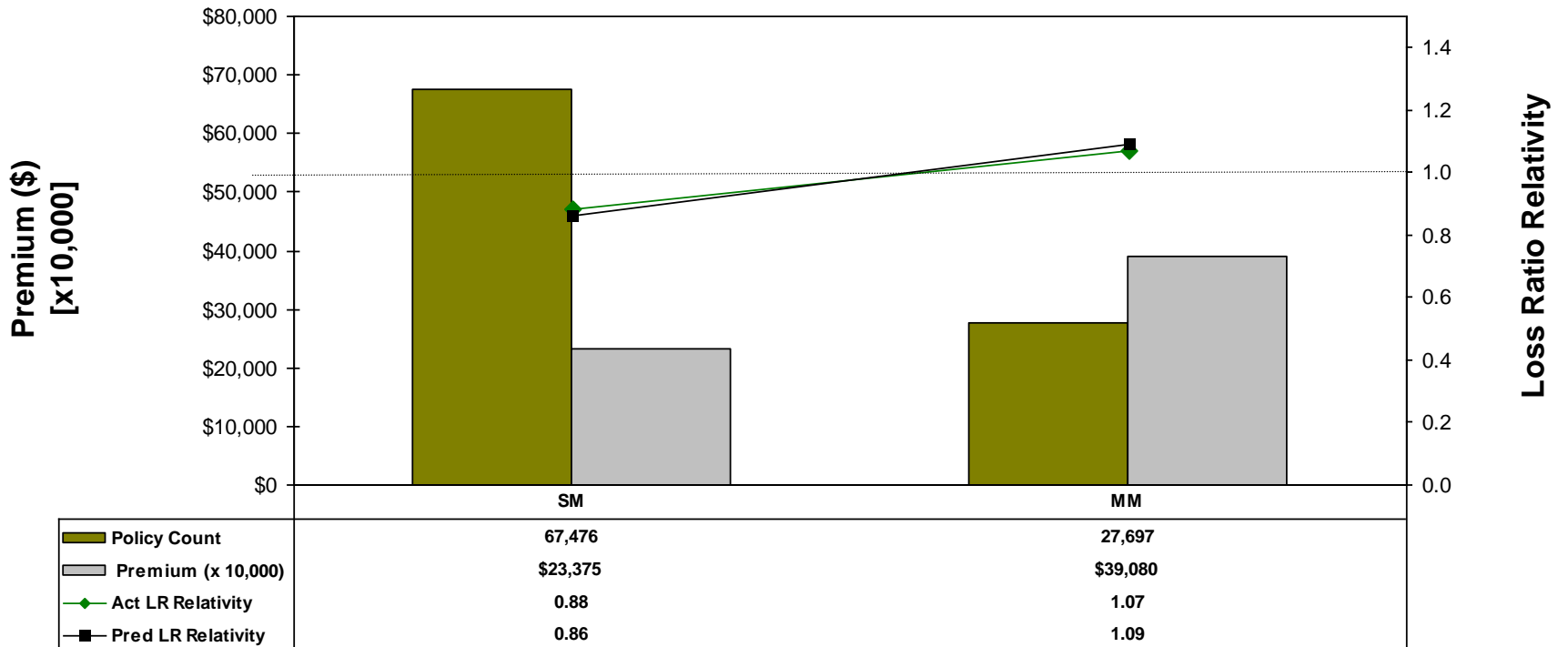
# CREDIT MODEL LIFT BY MARKET SEGMENT (VALIDATION SAMPLE)

Actual versus Predicted LR Relativities by Market Segment



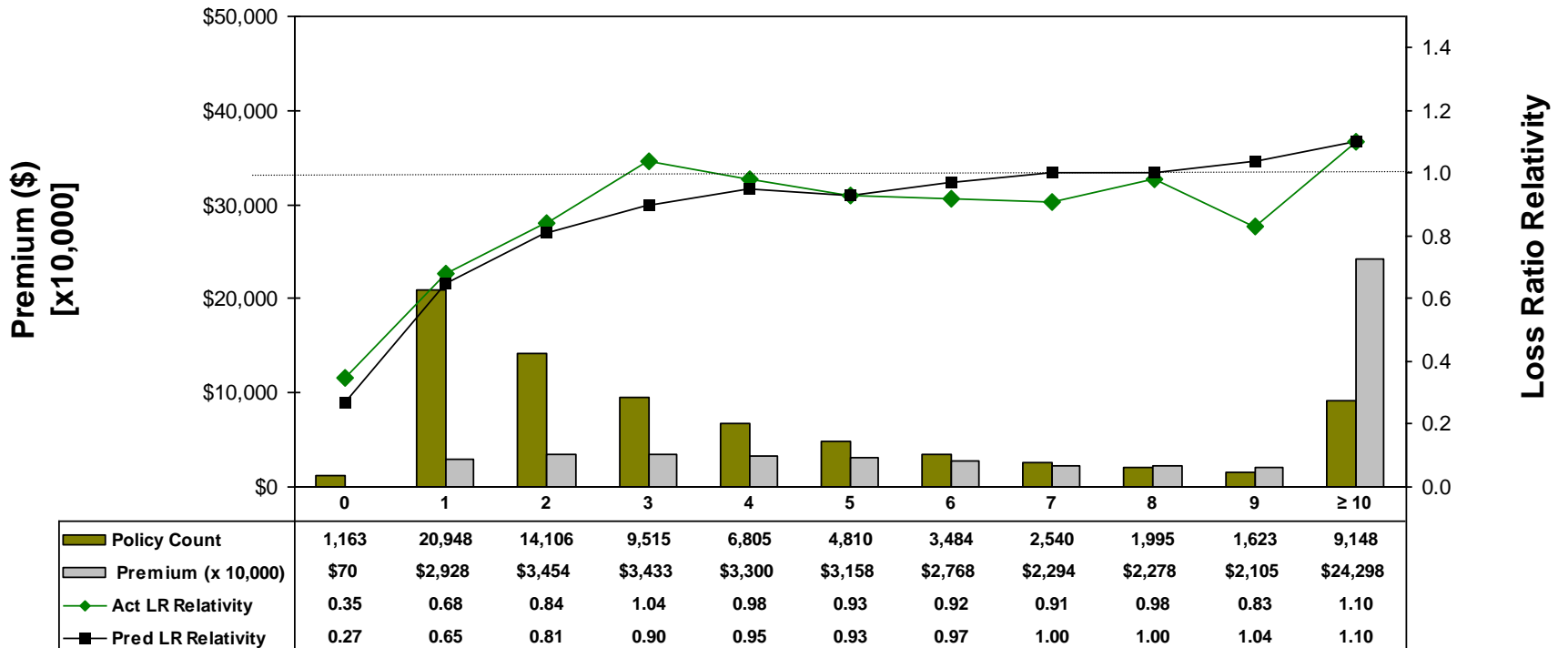
# CREDIT MODEL LIFT BY MARKET SEGMENT (FULL SAMPLE)

Actual versus Predicted LR Relativities by Market Segment



# CREDIT MODEL LIFT BY FLEET SIZE (BUILD SAMPLE)

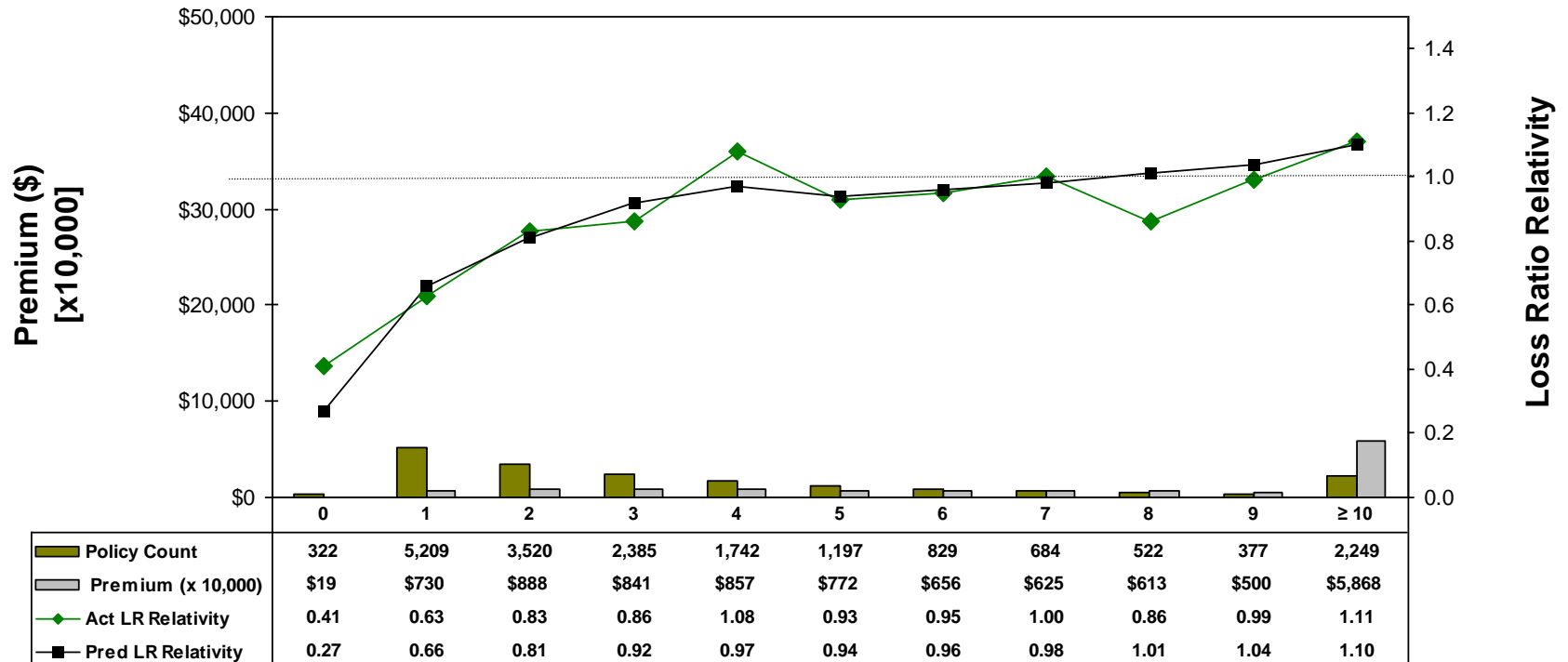
Actual versus Predicted LR Relativities by Fleet Size





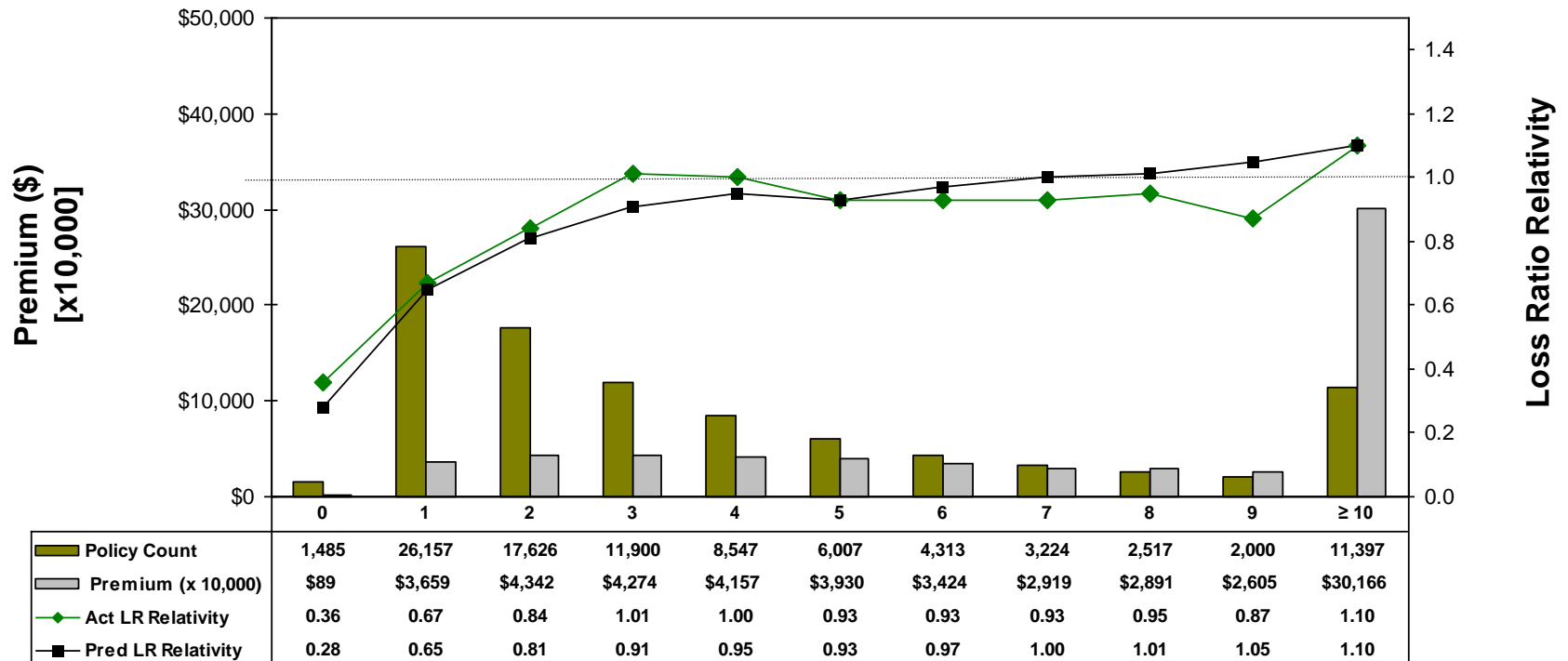
# CREDIT MODEL LIFT BY FLEET SIZE (VALIDATION SAMPLE)

Actual versus Predicted LR Relativities by Fleet Size



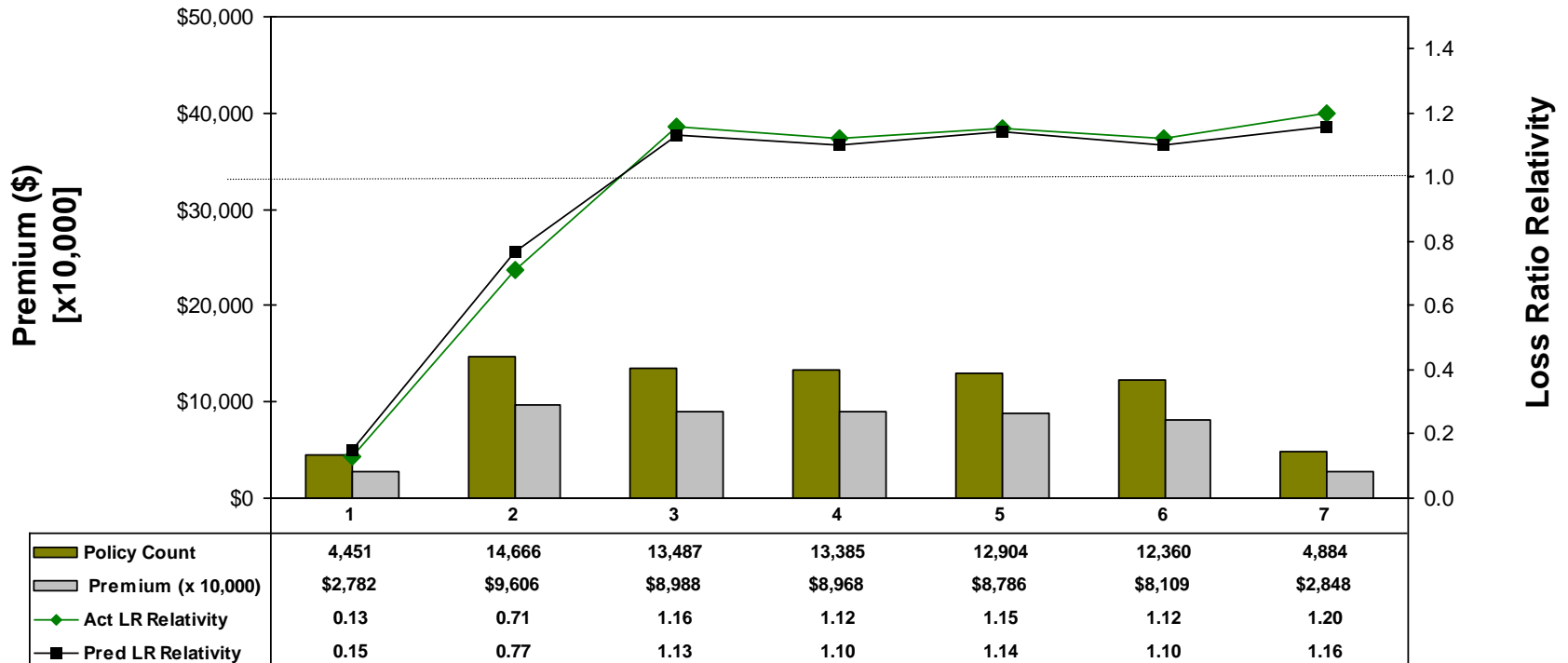
# CREDIT MODEL LIFT BY FLEET SIZE (FULL SAMPLE)

## Actual versus Predicted LR Relativities by Fleet Size



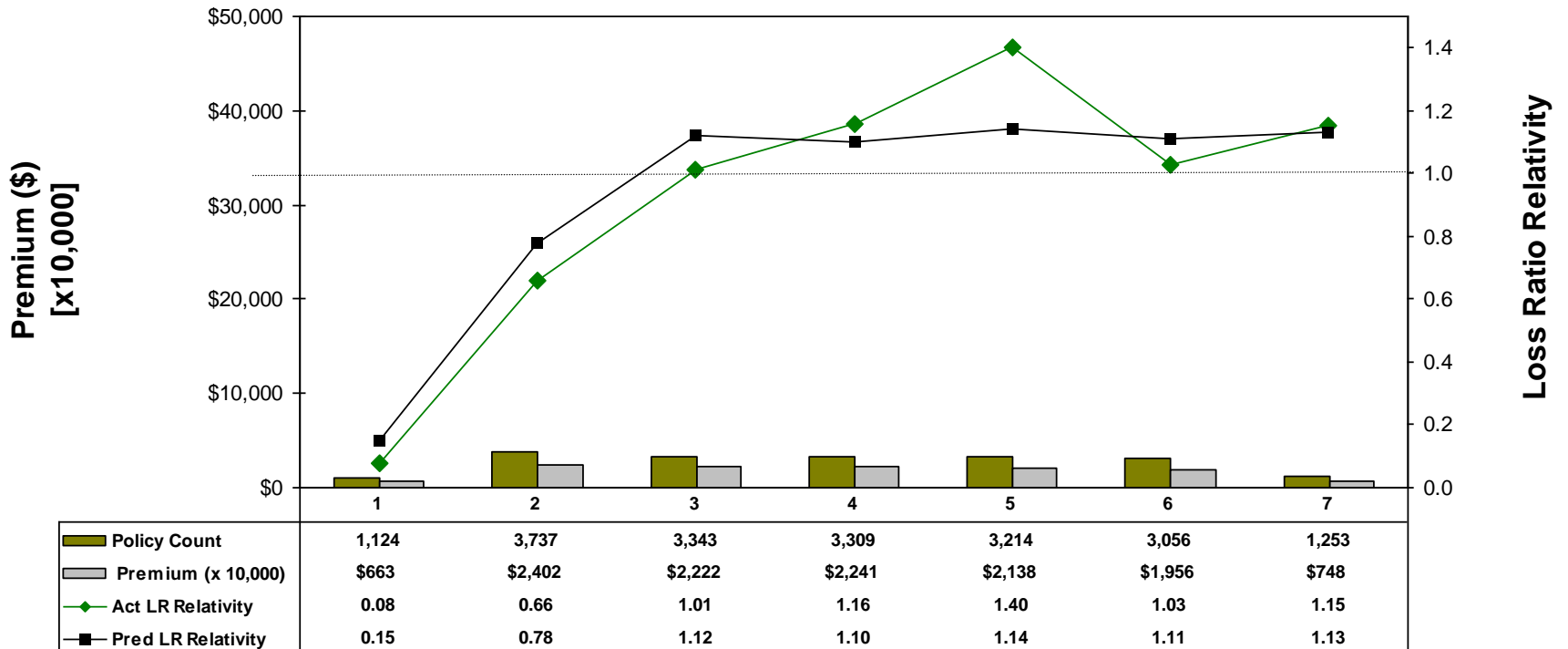
# CREDIT MODEL LIFT BY POLICY EFFECTIVE AGE (BUILD SAMPLE)

Actual versus Predicted LR Relativities by Policy Effective Age



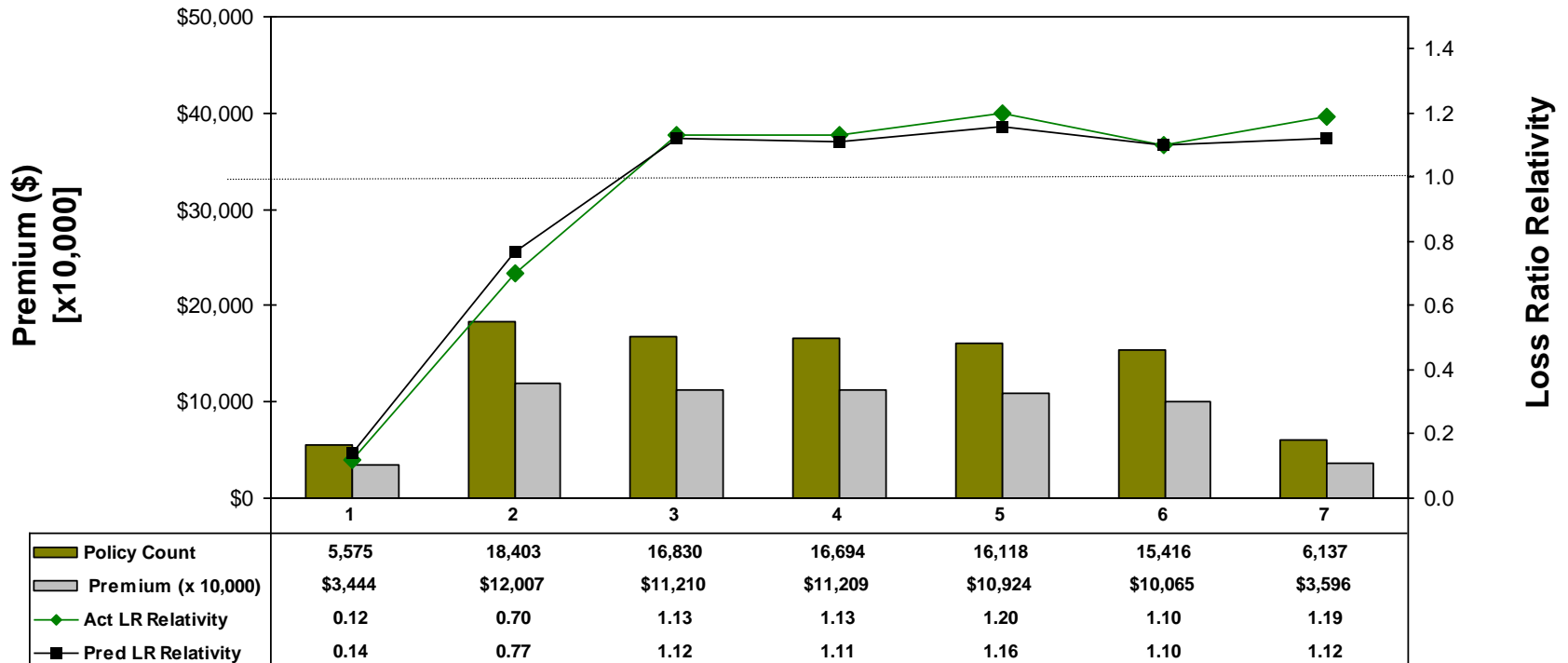
# CREDIT MODEL LIFT BY POLICY EFFECTIVE AGE (VALIDATION SAMPLE)

Actual versus Predicted LR Relativities by Policy Effective Age



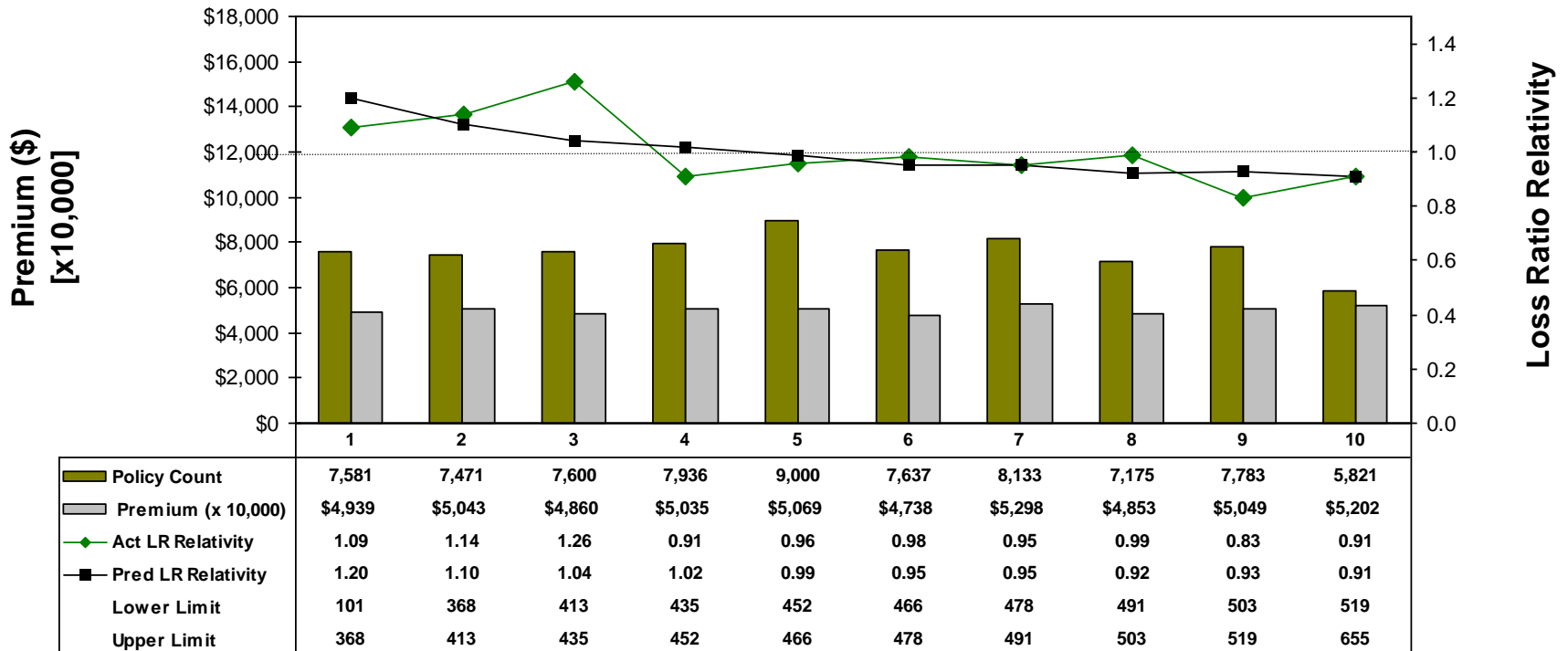
# CREDIT MODEL LIFT BY POLICY EFFECTIVE AGE(FULL SAMPLE)

Actual versus Predicted LR Relativities by Policy Effective Age



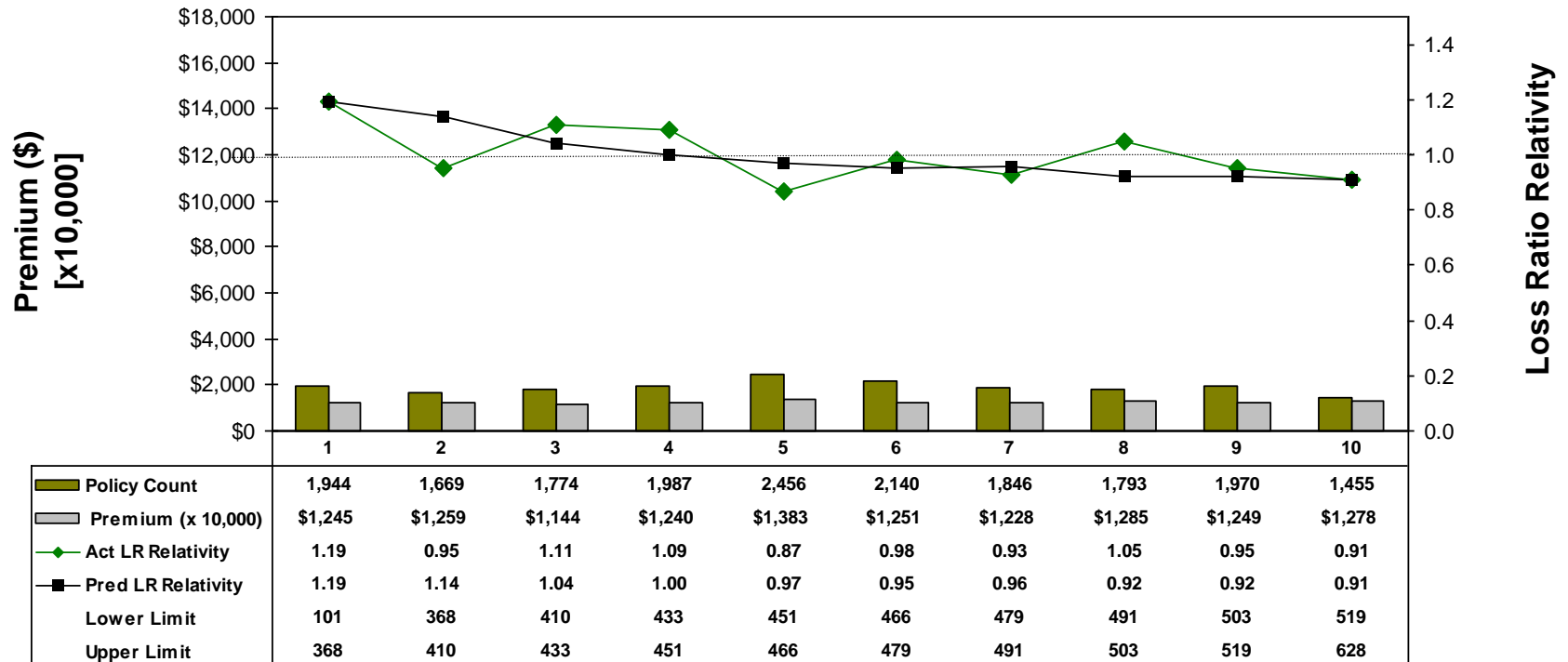
# CREDIT MODEL LIFT BY C POINTS (BUILD SAMPLE)

## Actual versus Predicted LR Relativities by C Points



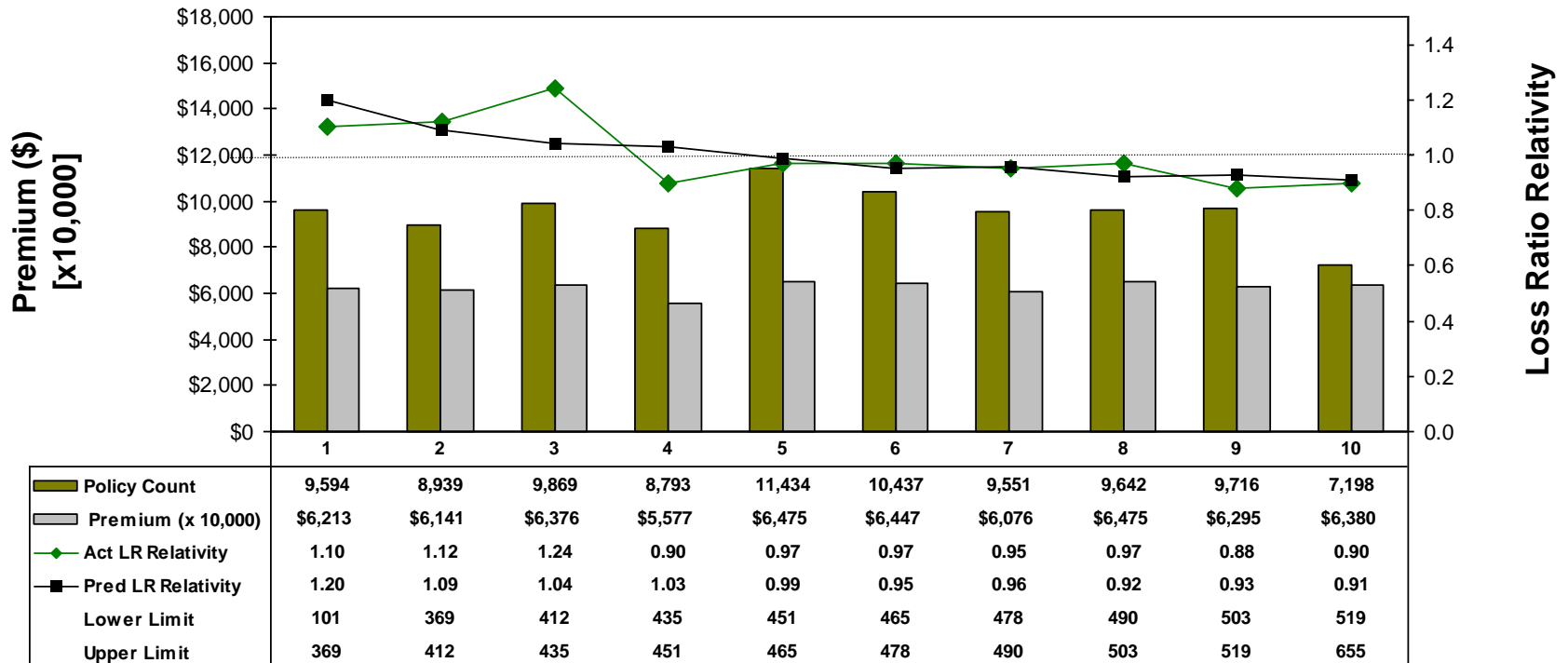
# CREDIT MODEL LIFT BY C POINTS (VALIDATION SAMPLE)

Actual versus Predicted LR Relativities by C Points



# CREDIT MODEL LIFT BY C POINTS (FULL SAMPLE)

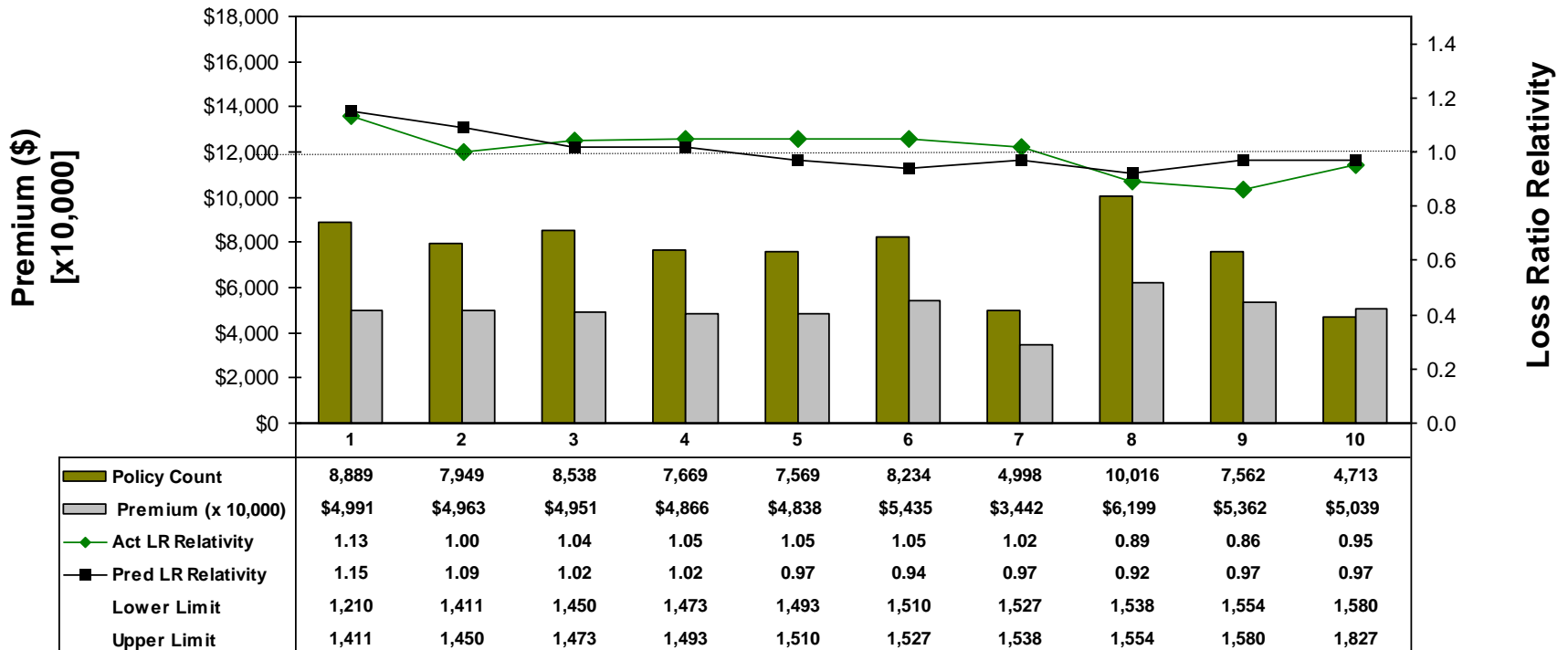
## Actual versus Predicted LR Relativities by C Points





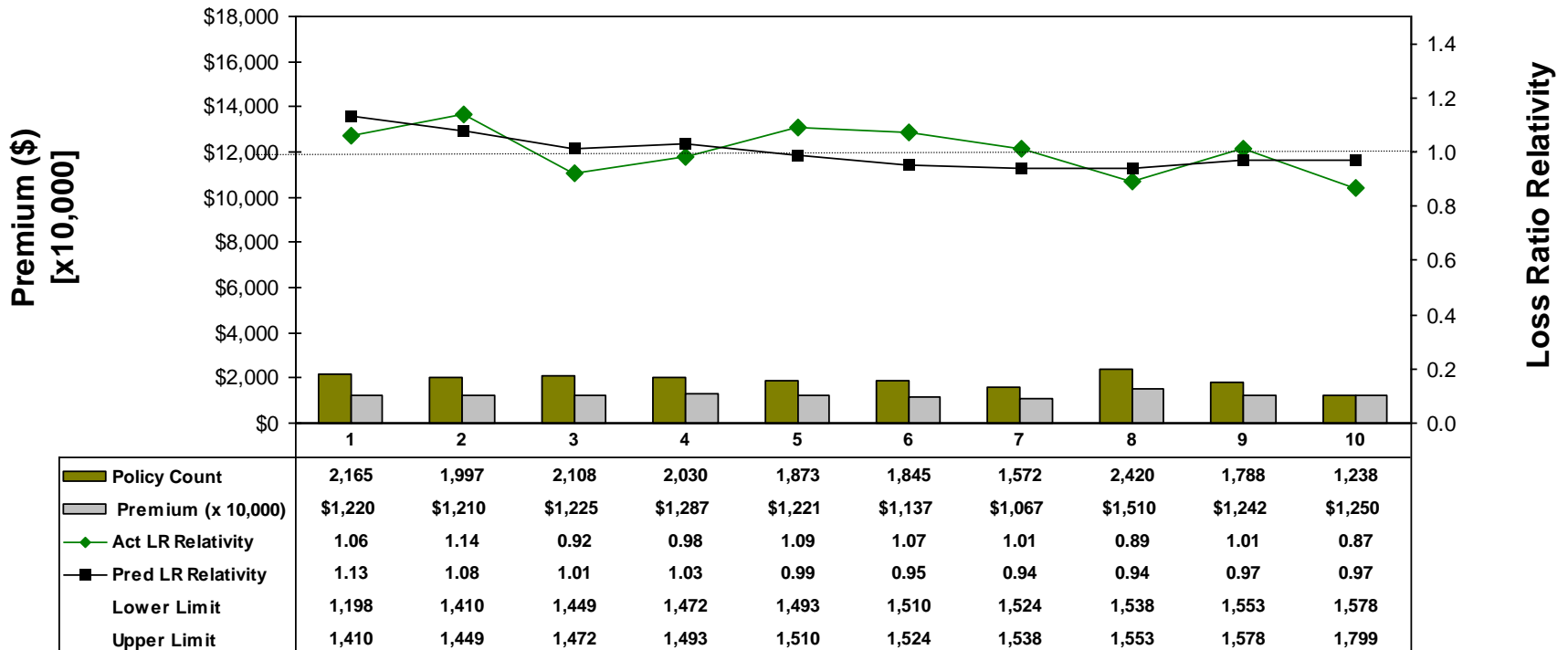
# CREDIT MODEL LIFT BY F POINTS (BUILD SAMPLE)

## Actual versus Predicted LR Relativities by F Points



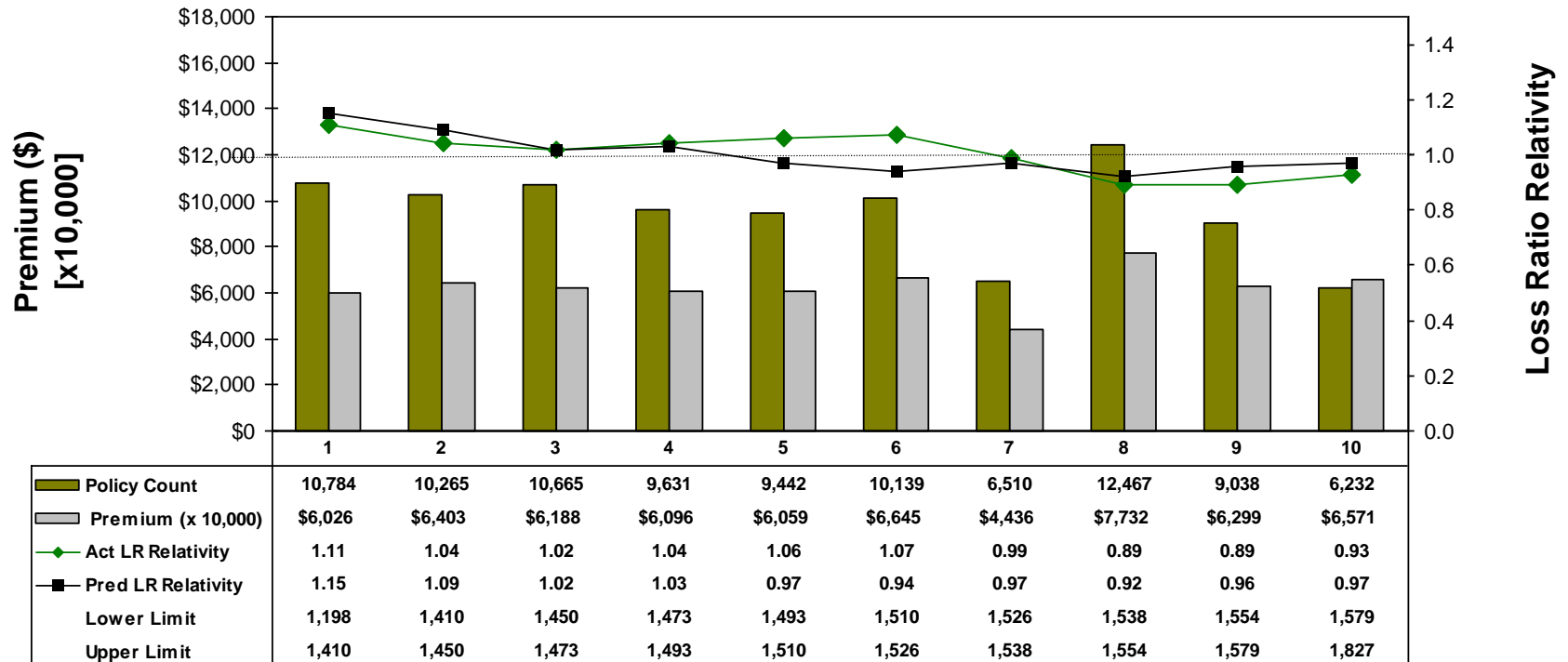
# CREDIT MODEL LIFT BY F POINTS (VALIDATION SAMPLE)

## Actual versus Predicted LR Relativities by F Points



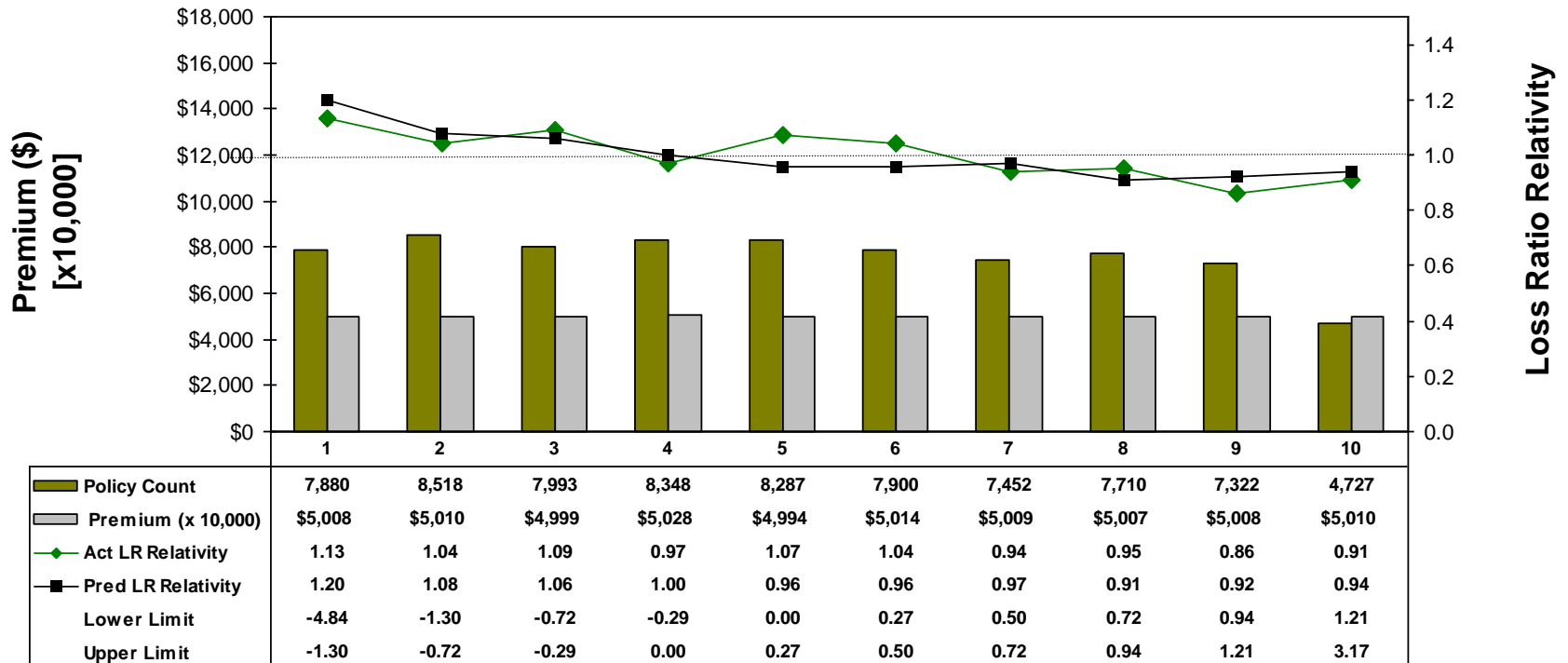
# CREDIT MODEL LIFT BY F POINTS (FULL SAMPLE)

## Actual versus Predicted LR Relativities by F Points



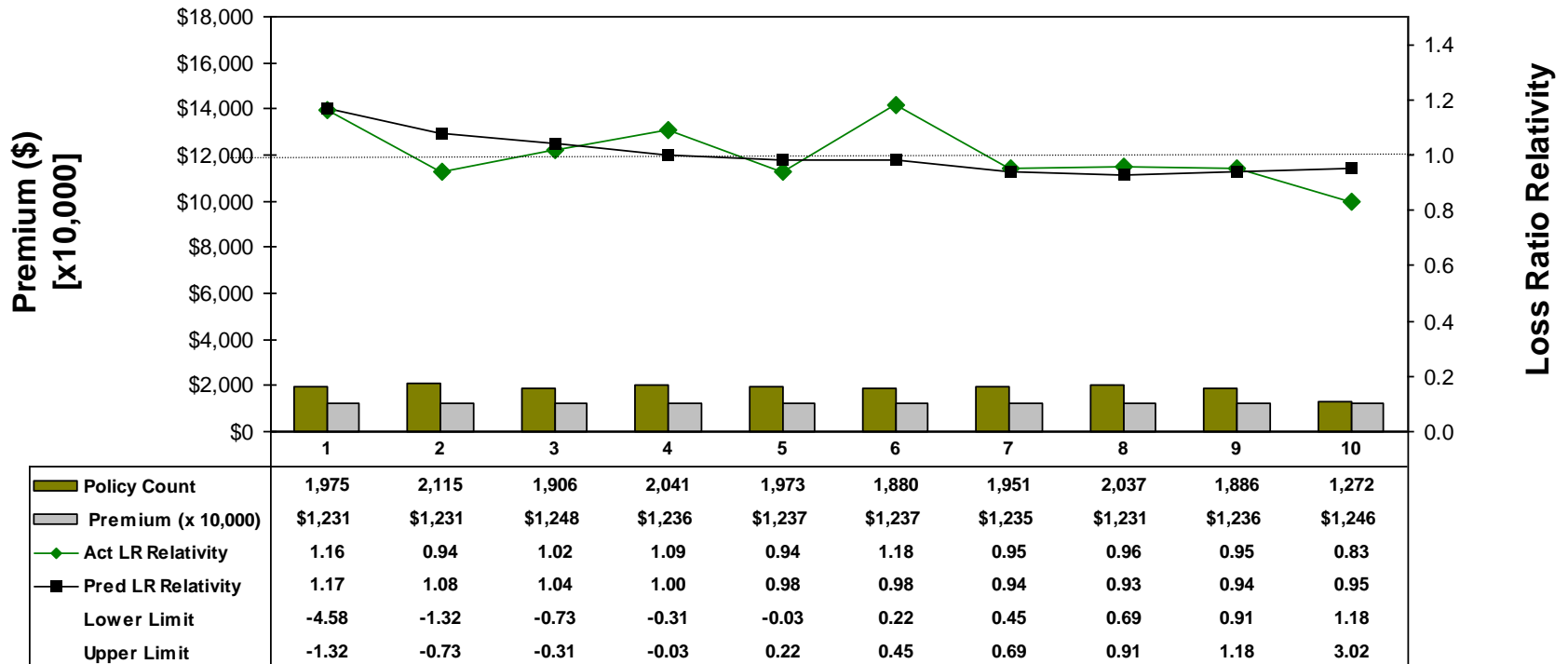
# CREDIT MODEL LIFT BY FINANCIAL STABILITY (BUILD SAMPLE)

## Actual versus Predicted LR Relativities by Financial Stability



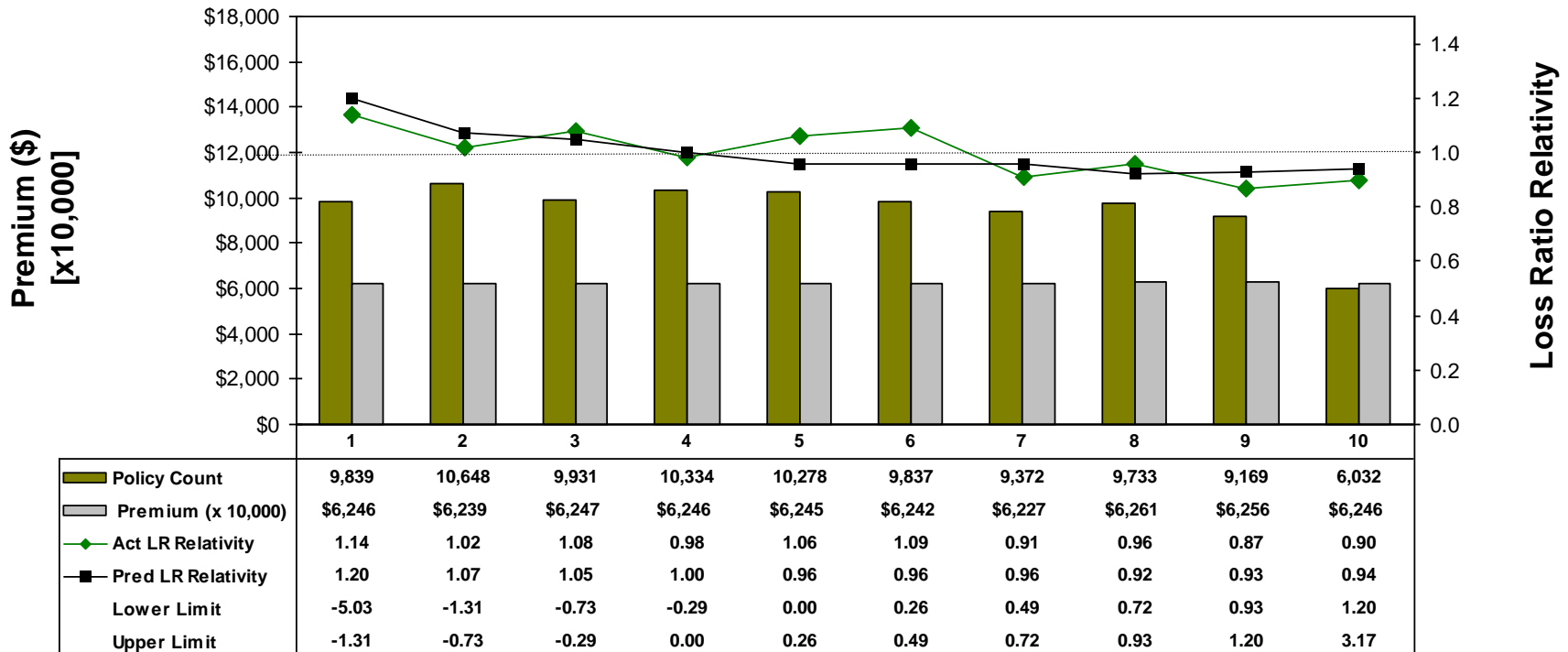
# CREDIT MODEL LIFT BY FINANCIAL STABILITY (VALIDATION SAMPLE)

Actual versus Predicted LR Relativities by Financial Stability



# CREDIT MODEL LIFT BY FINANCIAL STABILITY (FULL SAMPLE)

## Actual versus Predicted LR Relativities by Financial Stability



# Variable Coefficients

Variable	Coefficient	Chi-Squared	AIC
Intercept	-3.944		
Financial Stability	-0.106	1.746e-07	
Corp. Indicator	-0.015	0.0002085	
Policy Type 2 (LP)	-0.352	0.0002984	
Policy Type 3 (MM)	-1.358		
Policy Type 4 (MH)	-0.451		
Policy Type 5 (AO)	-0.573		
Market Segment (MM)	0.151	< 2.2e-16	

# Variable Coefficients (Continued)

Variable	Coefficient	Chi-Squared	AIC
Policy Effective Age (2)	1.648	< 2.2e-16	
Policy Effective Age (3)	2.020		
Policy Effective Age (4)	2.005		
Policy Effective Age (5)	2.042		
Policy Effective Age (6)	2.003		
Policy Effective Age (7)	2.049		
Fleet Size	-0.273	< 2.2 e-16	
Fleet Size <sup>2</sup>	0.007	4.291e-05	
Ln( Fleet Size + 1)	1.185	< 2.2e-16	
Fleet Size : Policy Type 2	0.035	1.216e-09	
Fleet Size : Policy Type 3	0.129		
Fleet Size : Policy Type 4	0.029		
Fleet Size : Policy Type 5	0.137		126180.3