Patterns and Anomalies of Loss Development in P&C Insurance Market

Steve Guo
Department of Mathematical Sciences
Ball State University

June 17, 2021

Based on joint work with Arthur Charpentier (UQAM) and Michael Ludkovski (UCSB)
Data Provided by Ledger Investing

Introduction: Functional Data with Reporting Delay

P&C loss triangle:

- Workers' compensation: 10+ years loss development $\Rightarrow$ data with reporting delay.
- Example loss triangle.

Related: Medical cases reporting

- Delay between diagnosis, onset of symptoms, and reporting cases.
- Example AIDS reporting data.

Introduction: P&C Loss Development

Regulators and reinsurers on industry-wide loss (ratio) development.

- Study the pattern of loss development on a business line.
- Look for "anomalous" loss development for particular insurers.
- Increase / reduce exposure in particular business lines / types of company.

Loss reserving beyond a single triangle:

- Historical pattern of loss development may inform future.
- Loss development from other "similar" insurers' may provide additional information.

Our Research

Using statistical learning approaches, we can explore:

- What is the general pattern of loss development? How is it linked to company features?
- We use unsupervised learning to discover the patterns and anomalies of industry-wide loss development in terms of functional curves.

We also visit the problem of loss reserving:

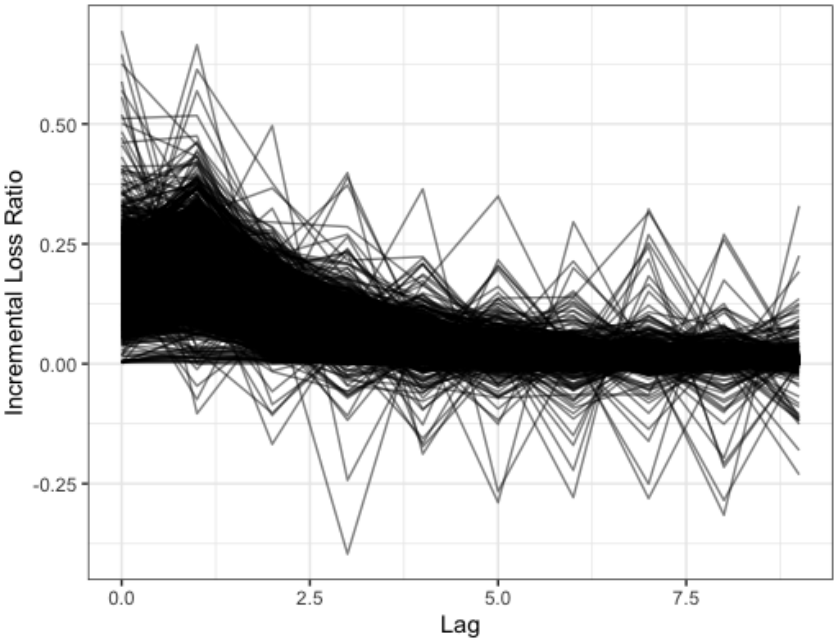
- How to obtain forecast for the incomplete loss development curves?
- We use a non-parametric, data-driven procedure to forecast future losses by matching a partial loss development experience against similar completed patterns in the database.

Triangle Data

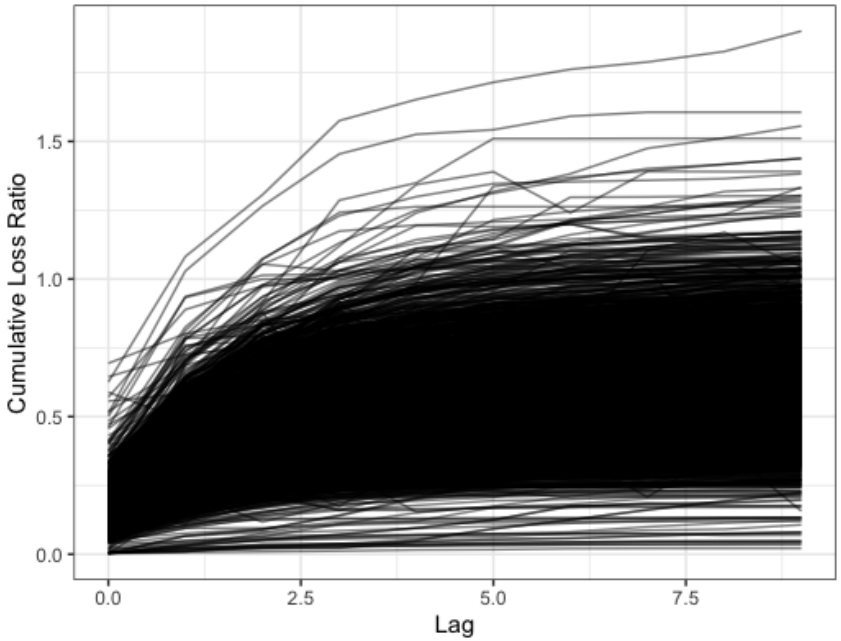
Paid loss triangle and normalizing:

- Incremental Loss Ratio (ILR): incremental paid loss in a development lag / net premium earned.
- Cumulative Loss Ratio (CLR): cumulative sum of ILRs.
- Unsupervised learning: A sample of 100+ firm-year complete ILR function curves from 1987 to 2010 in Workers' Compensation.

Data Summary Plots



(a) Incremental Loss Ratios



(b) Cumulative Loss Ratios

Methodology

Principal Component Analysis (PCA)

- Explore data structure in high dimensional and functional data.
- Good for interpretation and visualization.
- Apply a robust PCA algorithm (Croux, Filzmoser and Oliveira, 2007).
- Produce bivariate and functional bagplots using first two PCA scores (Hyndman and Shang, 2010).
- Produce visualizations that track companies' historical loss development patterns.

Principal Components

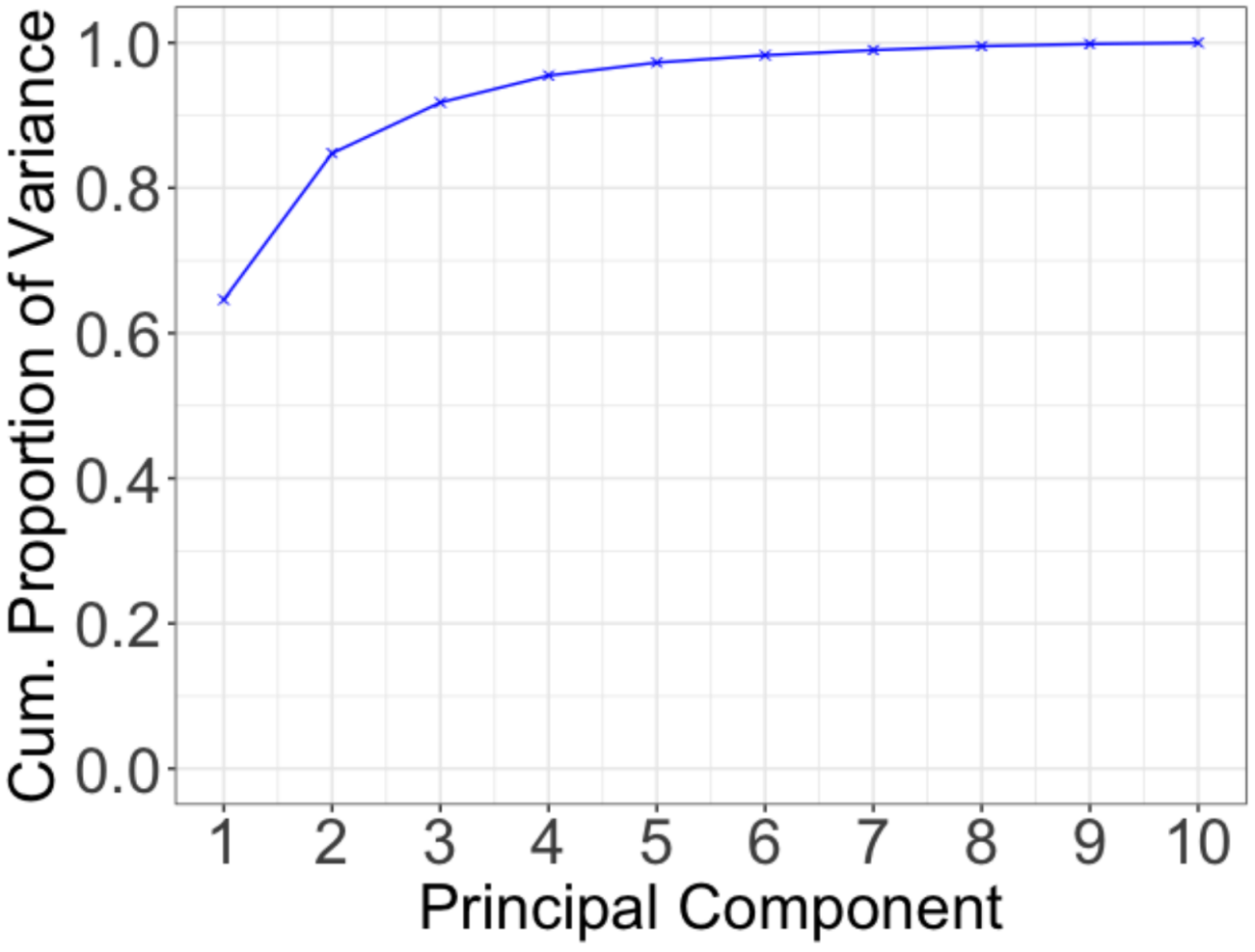


Figure: Cumulative Scree Plot

Biplot: Principal Components and Variable Mapping

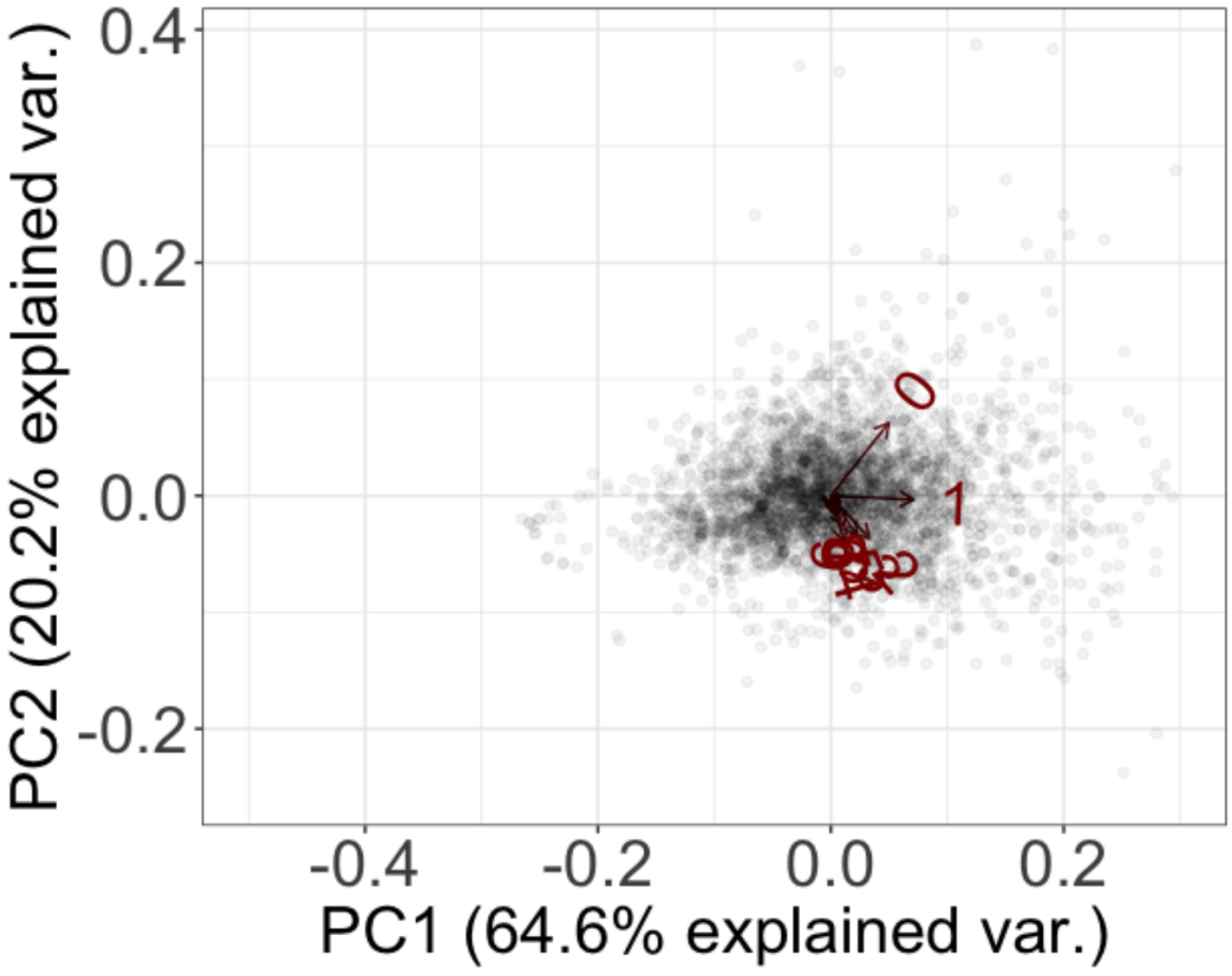


Figure: Biplot PC1-PC2

Biplot: Principal Components and Variable Mapping

Lag 0 positively influences PC2 and somewhat correlated with Lag 1.

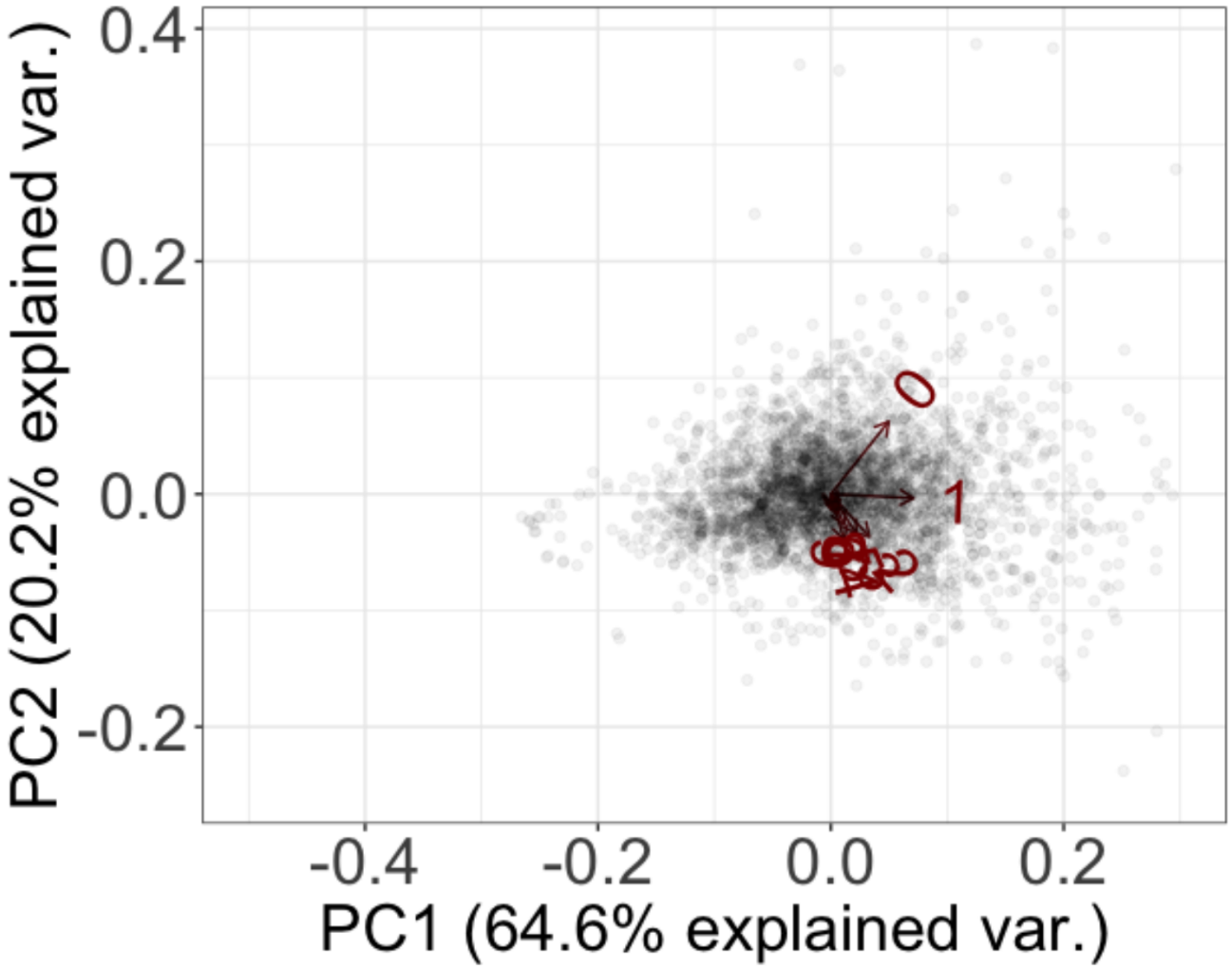


Figure: Biplot PC1-PC2

Biplot: Principal Components and Variable Mapping

Lag 1 positively influences PC1.

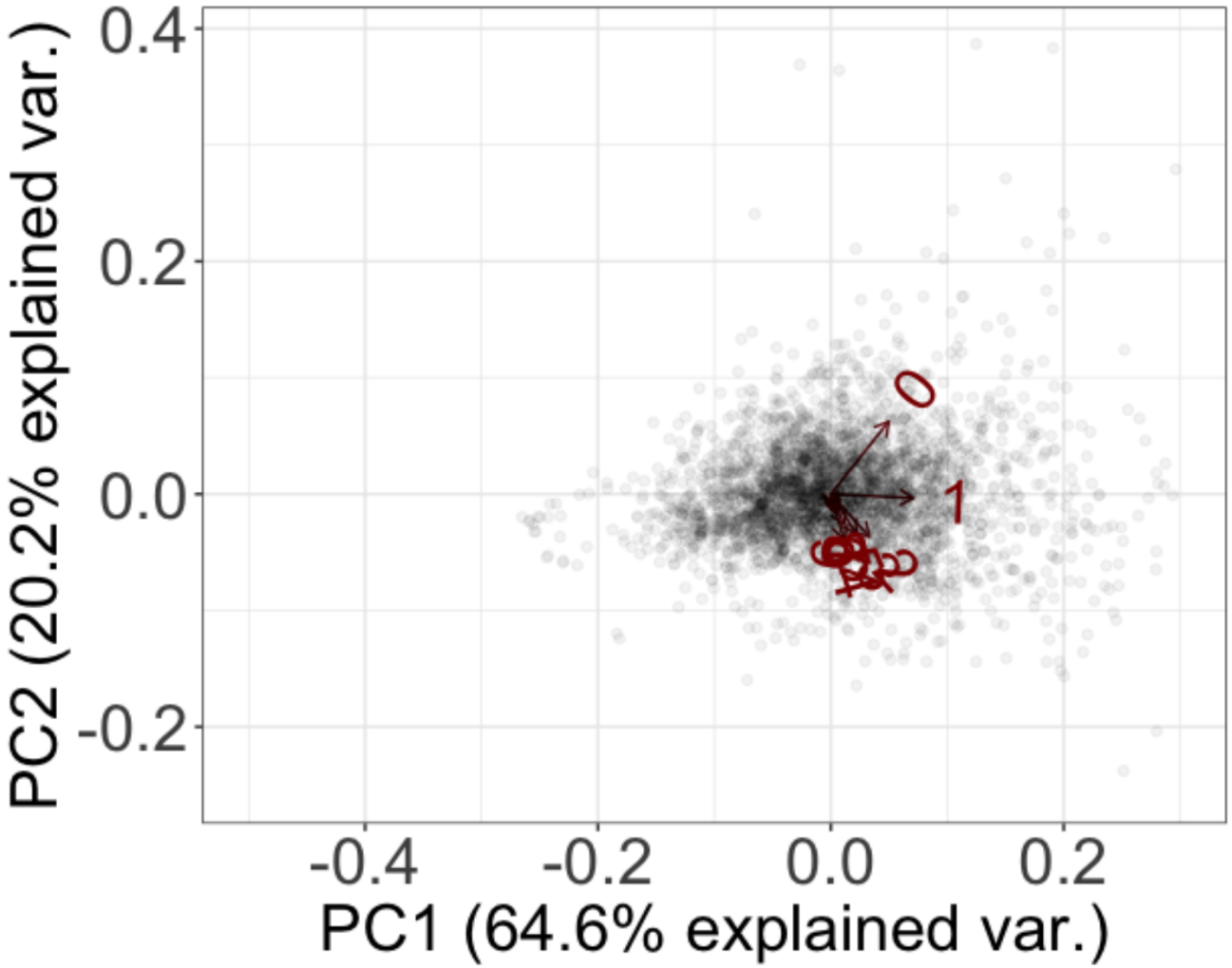


Figure: Biplot PC1-PC2

Biplot: Principal Components and Variable Mapping

Lag 2 to 9 are correlated and somewhat negatively correlated with Lag 0.

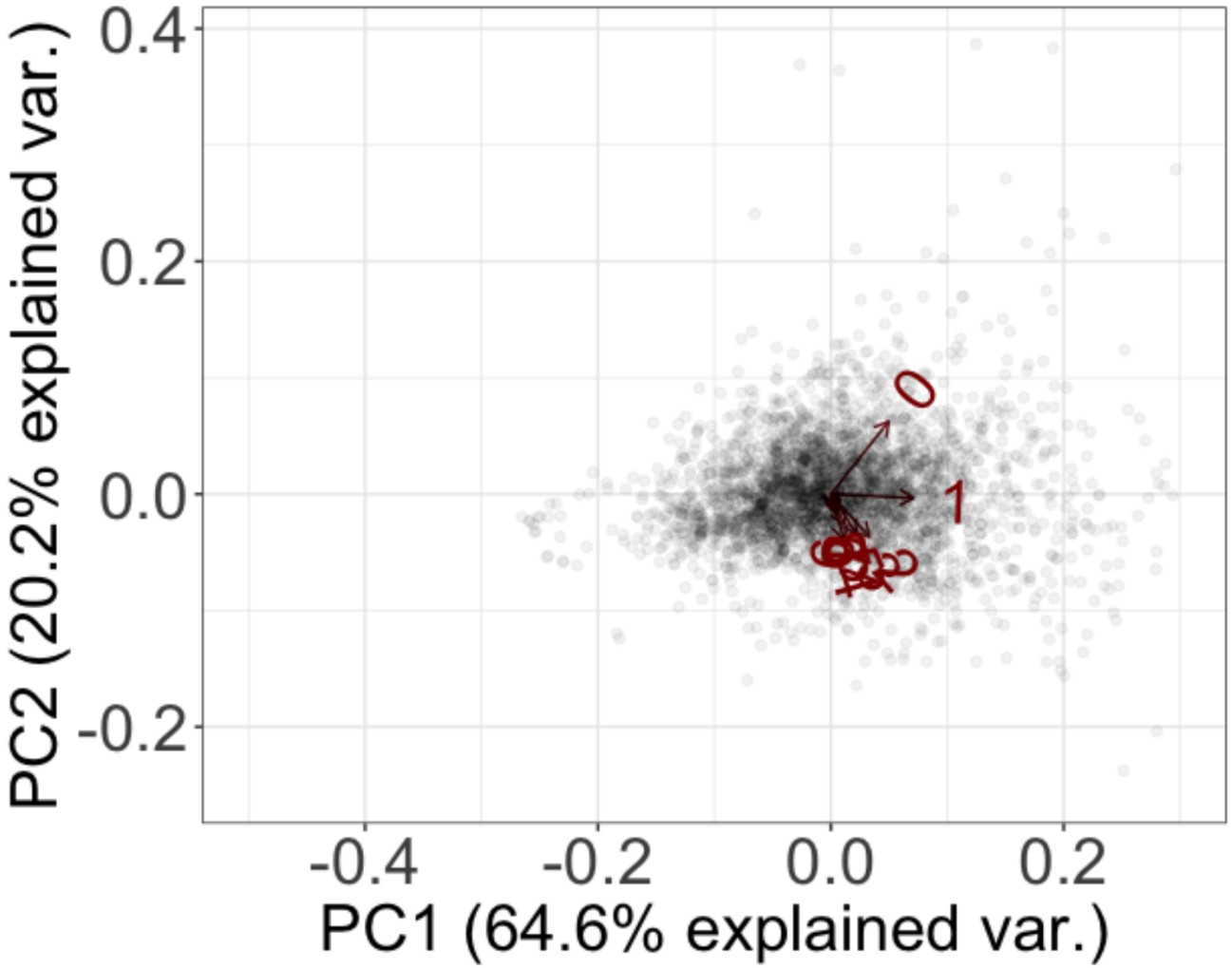


Figure: Biplot PC1-PC2

Bivariate and Functional Bagplots

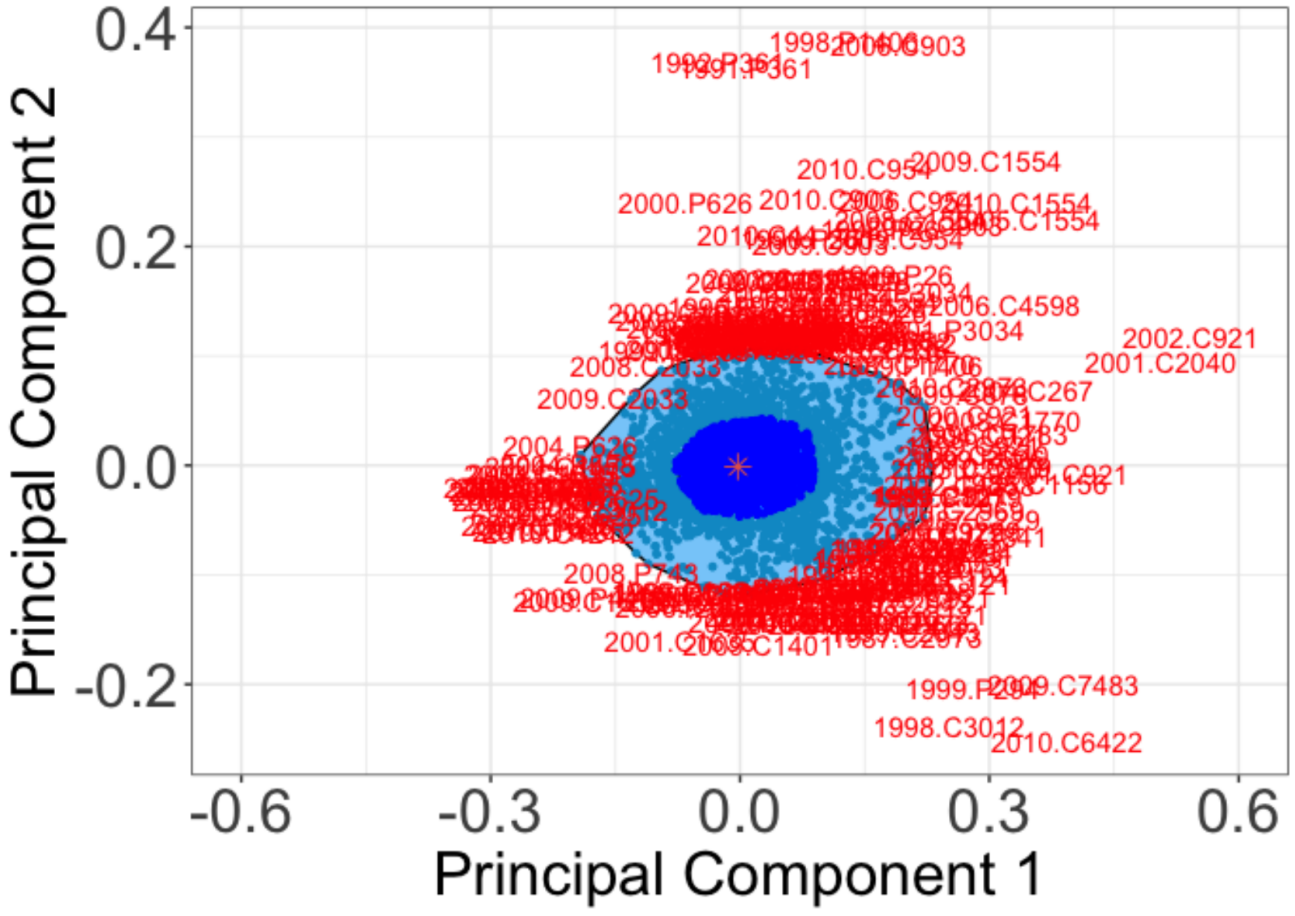
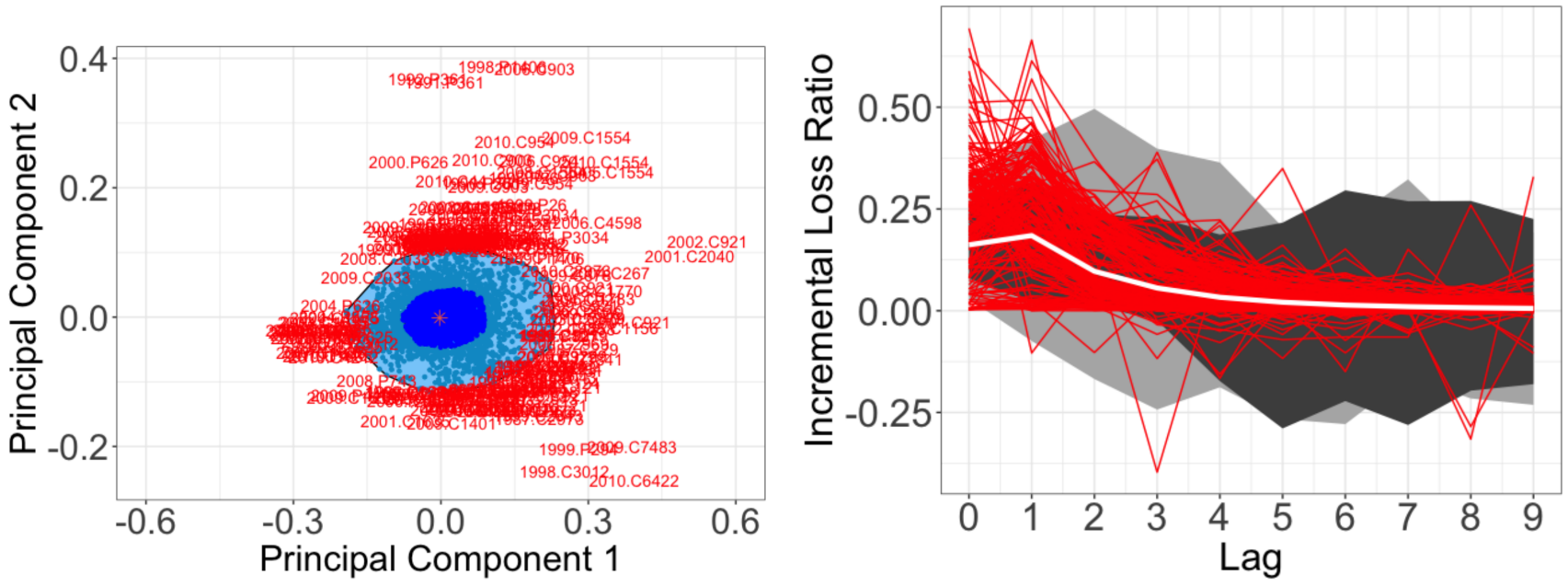


Figure: Bivariate bagplot of Incremental Loss Ratios

Bivariate and Functional Bagplots



(a) Bivariate bagplot

(b) Functional bagplot

Figure: Incremental Loss Ratios

Explore Patterns within Company Covariates

Company Covariates Data

- Business focus: commercial, personal, minimum.
- Capital structure: stock, mutual, other.
- Geographical focus: Northeast, South, Midwest, West, National.

Explore Patterns within Company Covariates

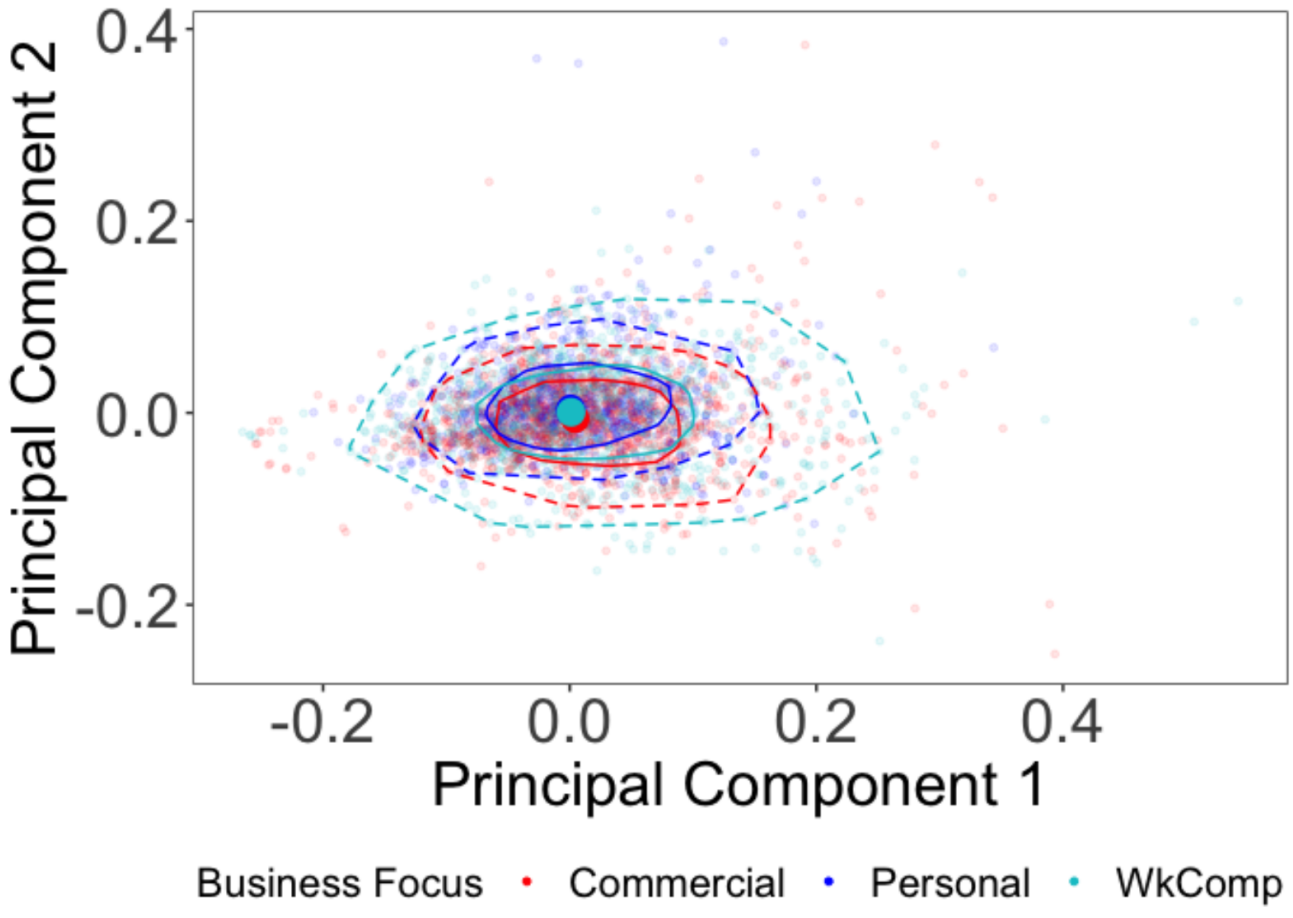


Figure: Biplot by business focus

Explore Patterns within Company Covariates

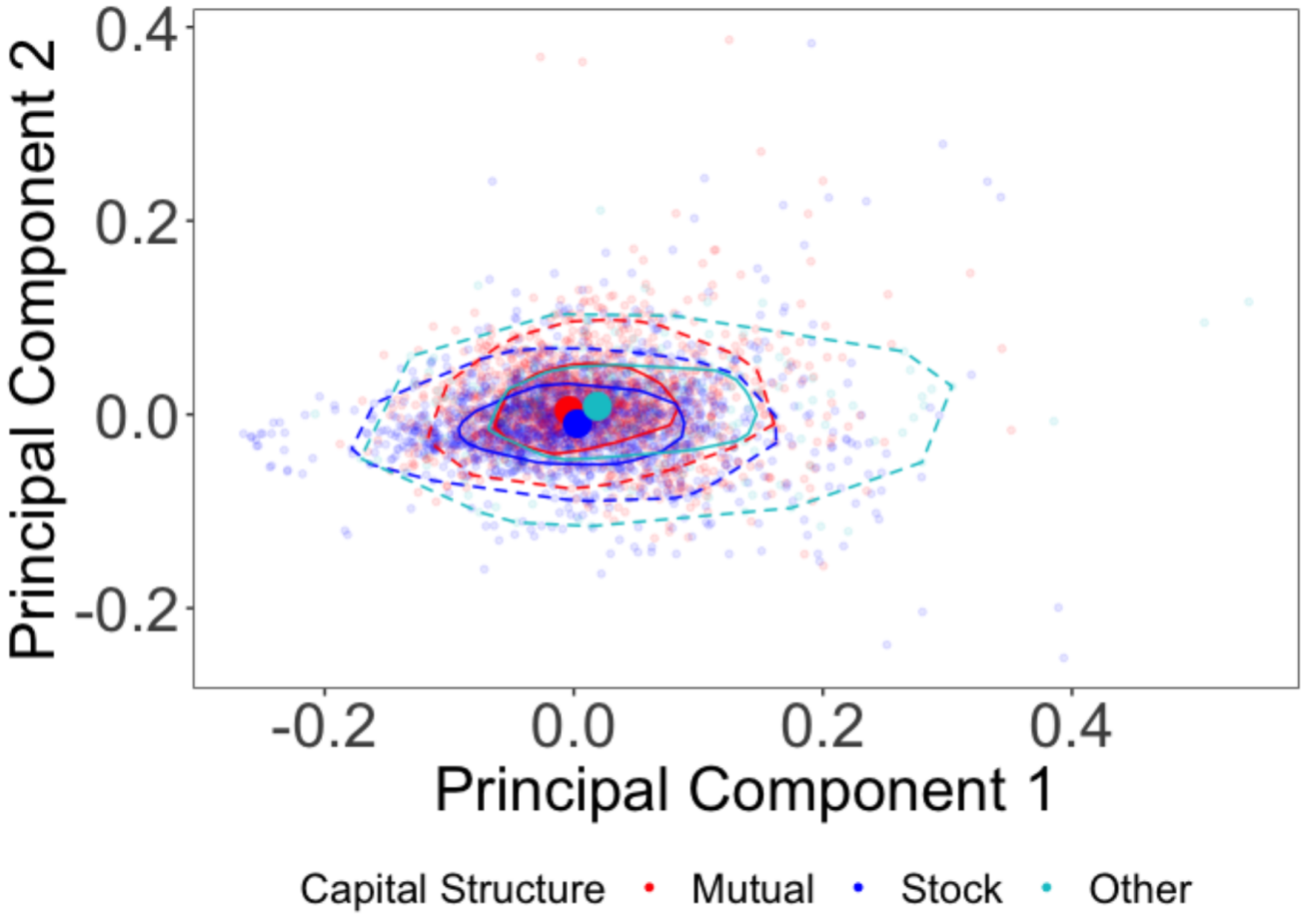


Figure: Bagplot by capital structure

Explore Patterns within Company Covariates

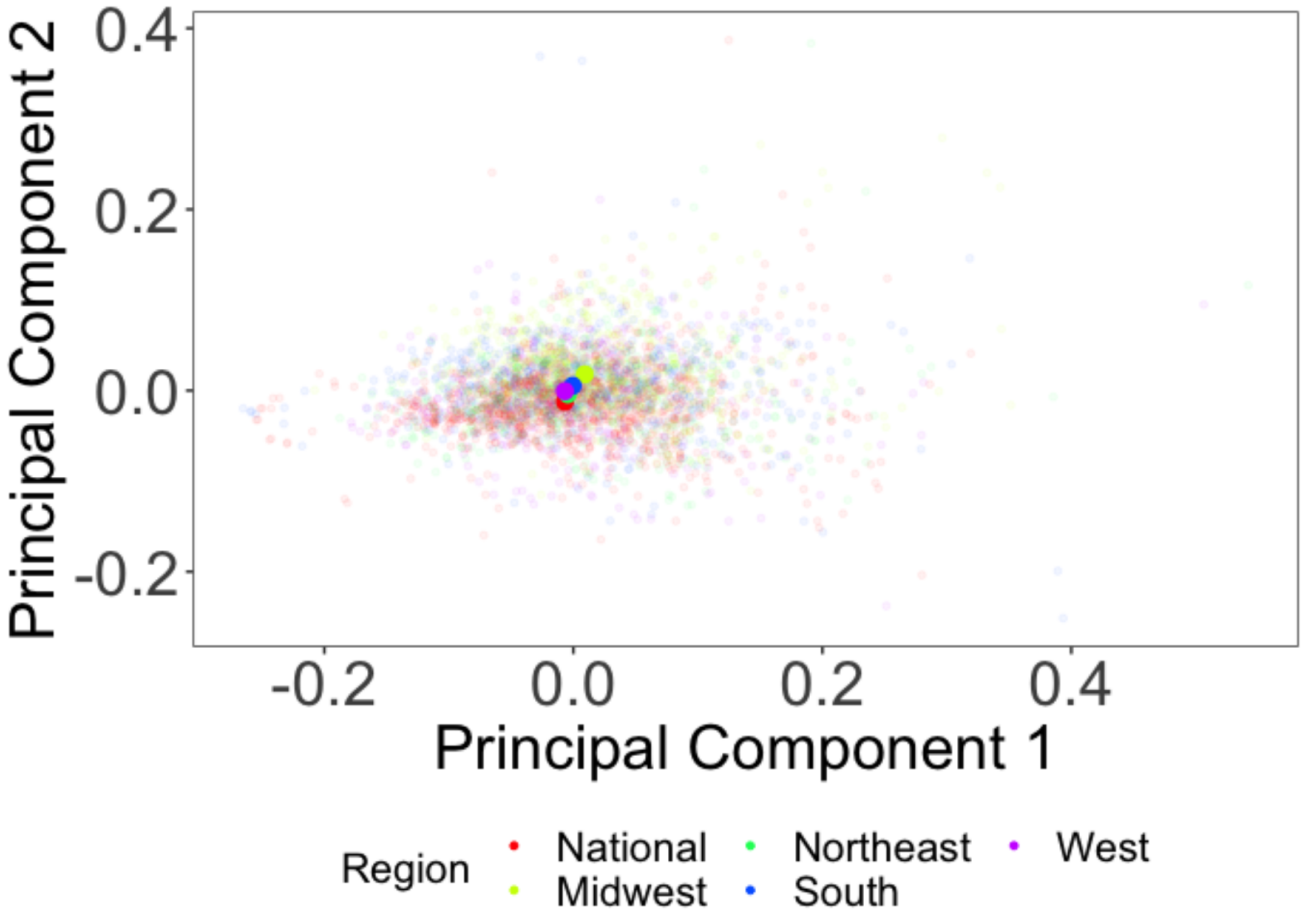


Figure: Bagplot by region

Explore Patterns within Company Covariates

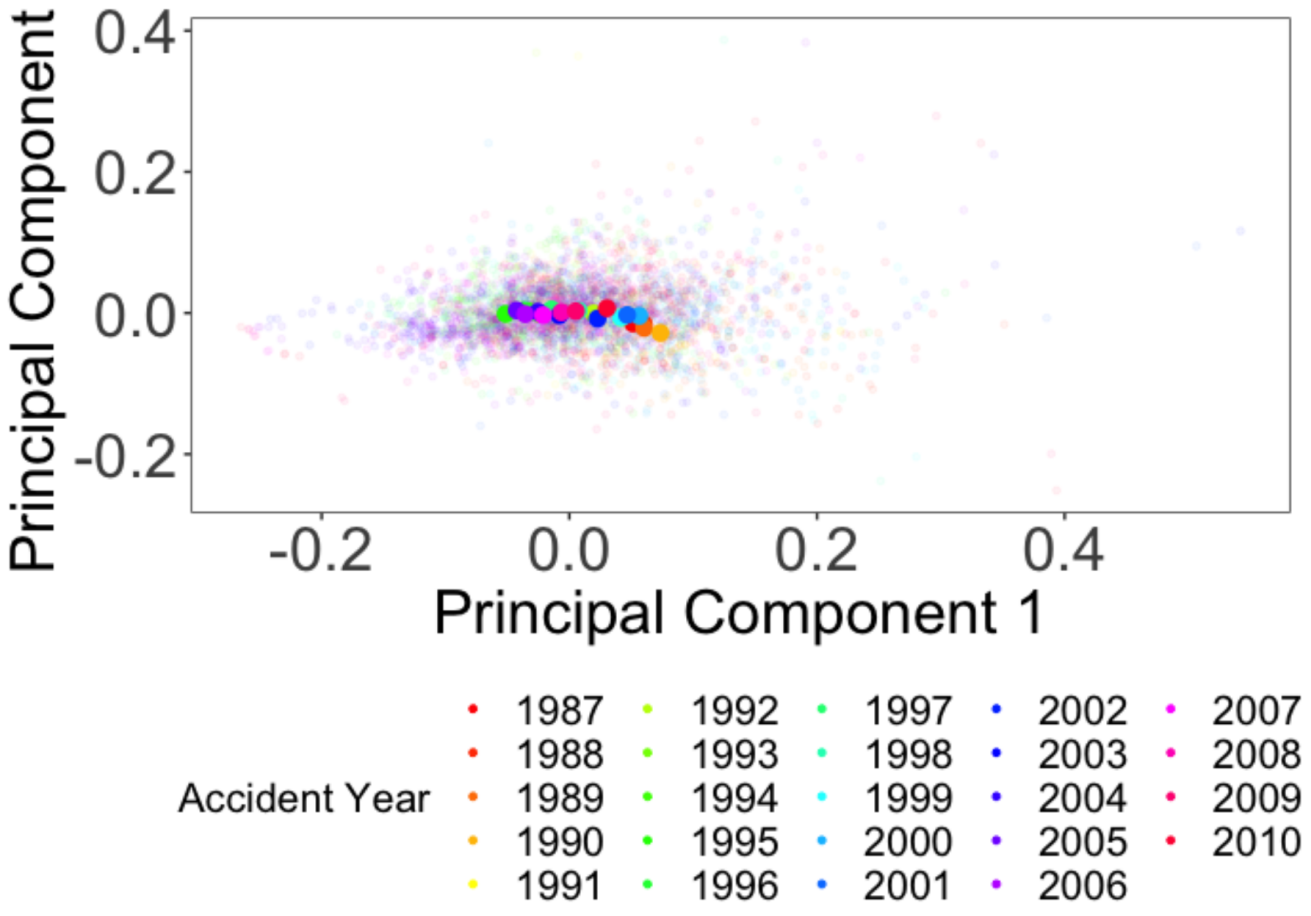


Figure: Bagplot by accident year

Representative Companies

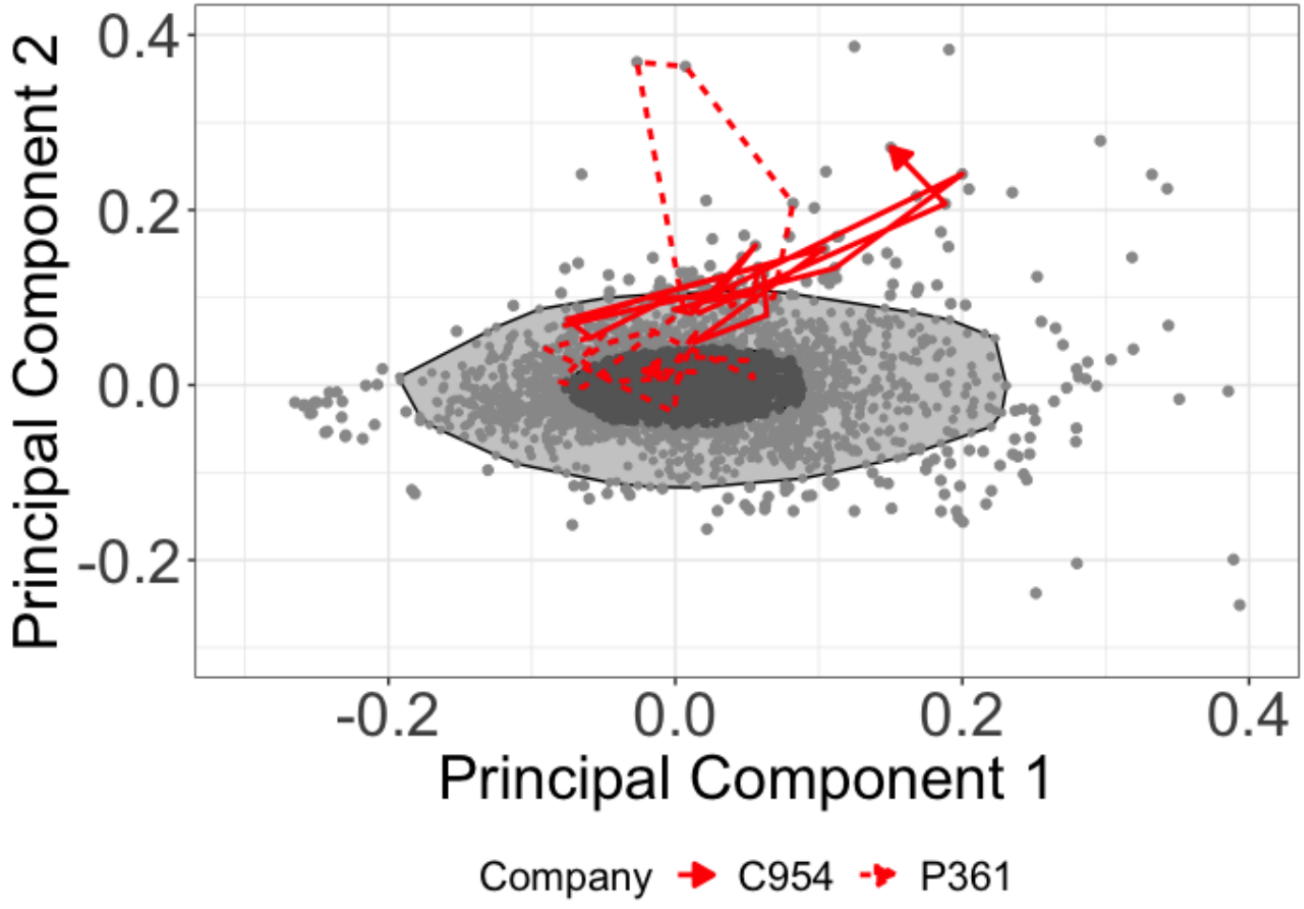
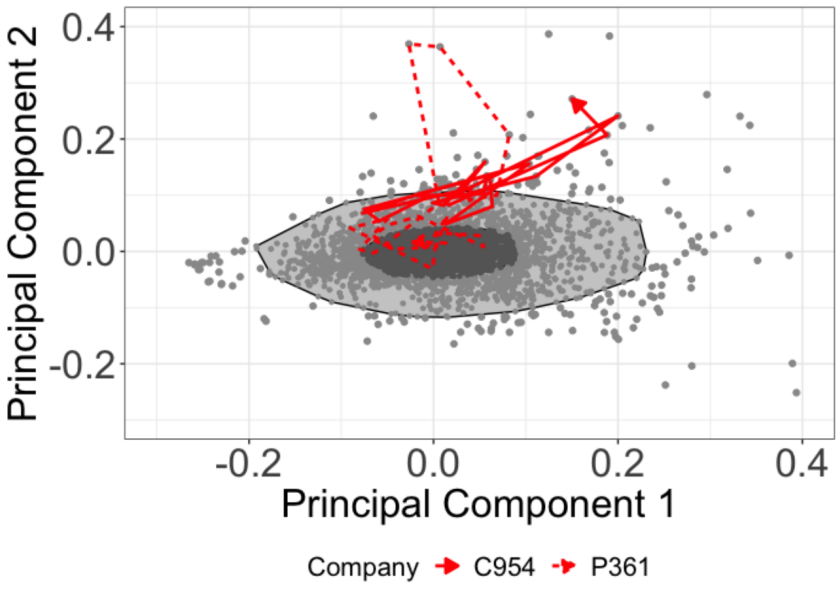
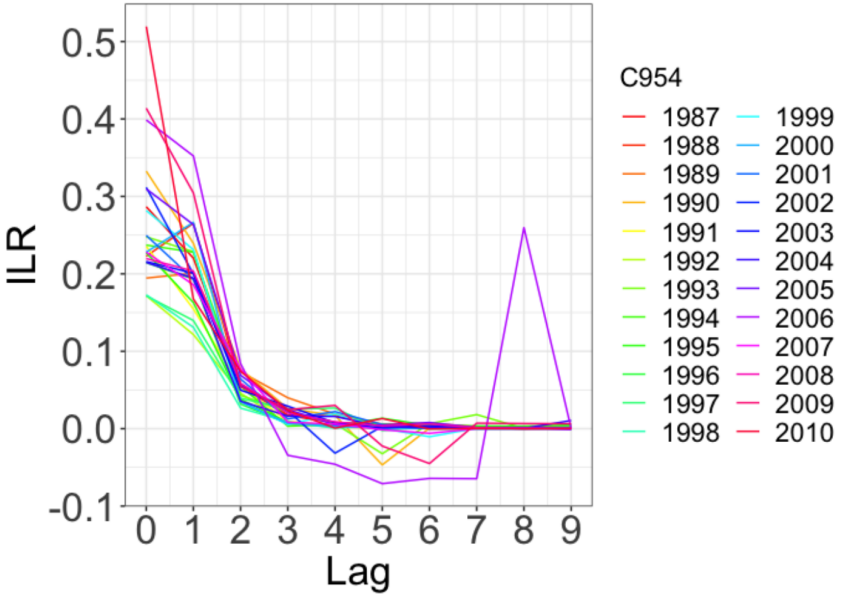


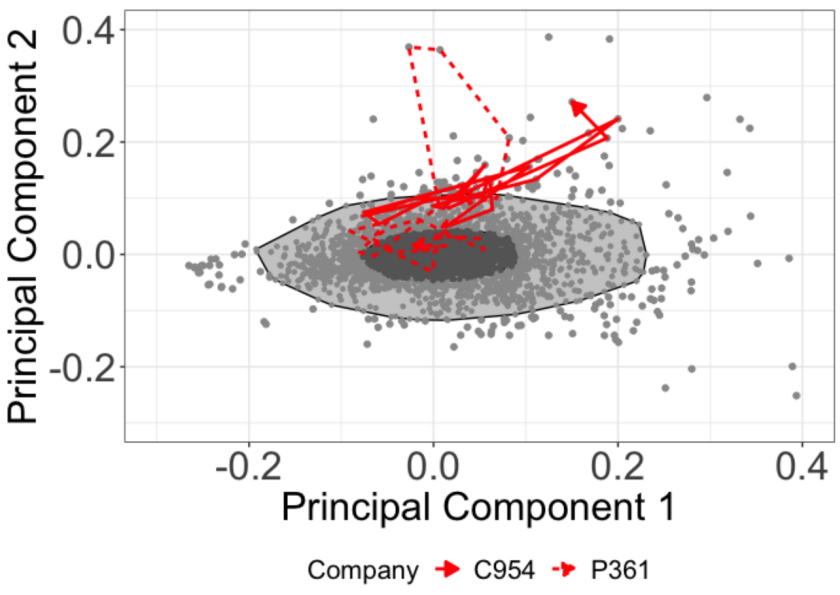
Figure: Tracking two companies from 1987 to 2010

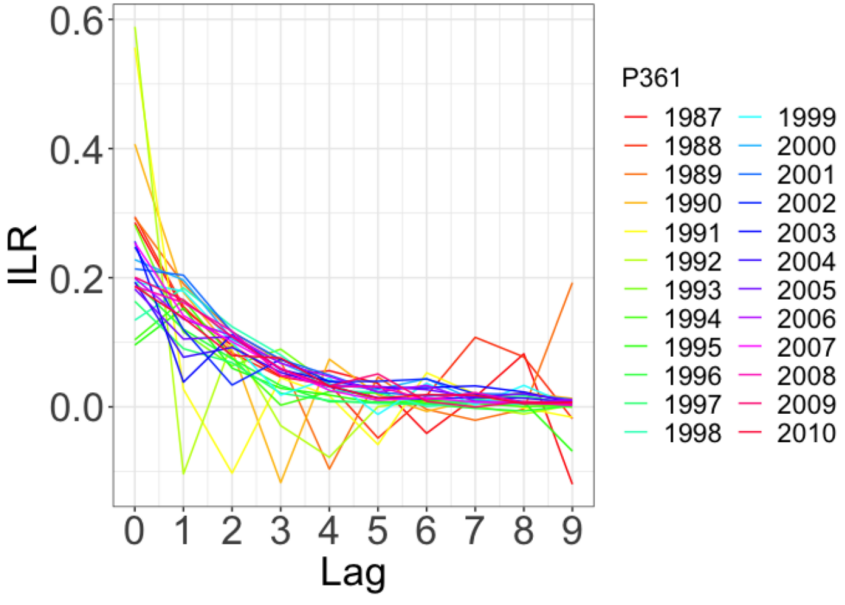
Representative Companies





Representative Companies





Loss Reserving Model

Goal: forecast partially complete CLR curves.

- Incorporating information from the past complete curves.
- Incorporating information from similar curves of other companies.

Functional data forecast methods:

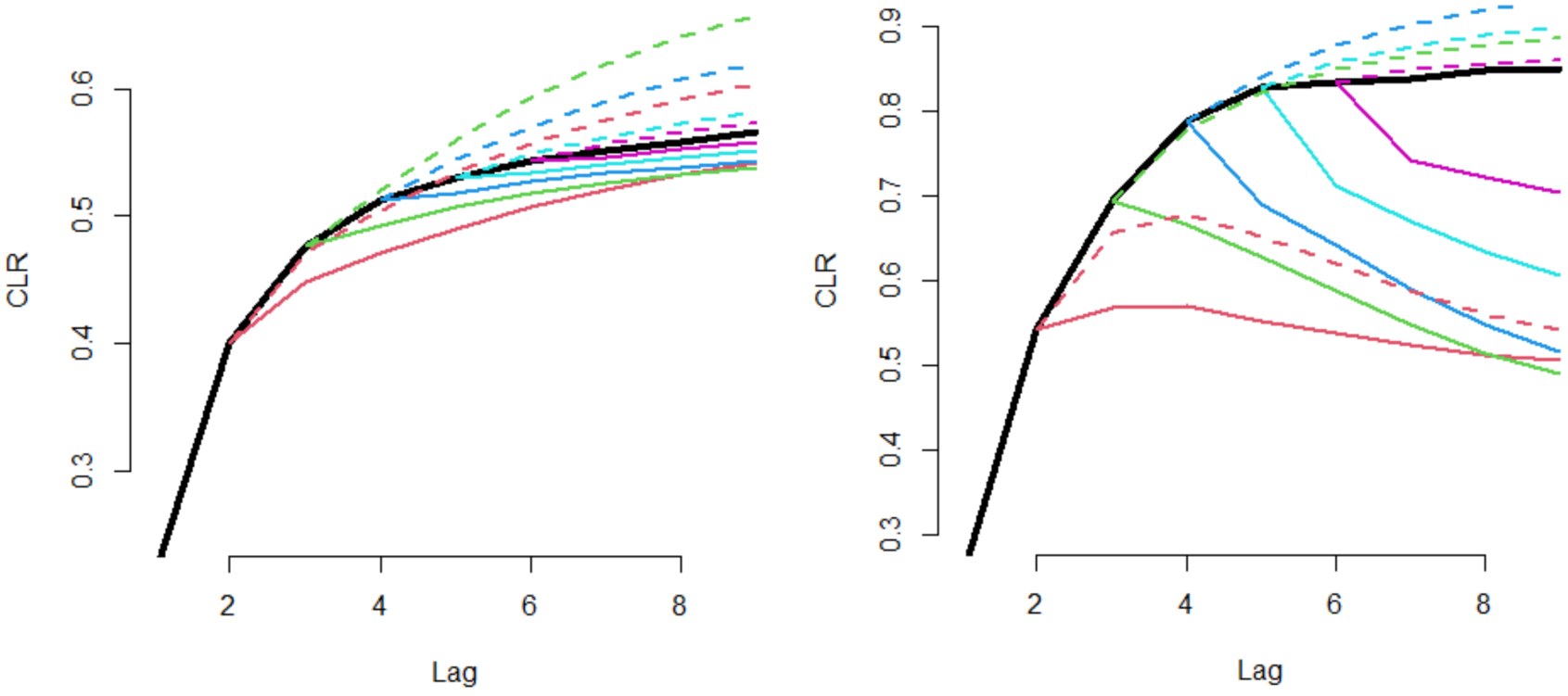
- Diebold and Li (2006): parametric modeling of treasury yield curves; forecast level, slope and curvature parameters; recover yield curve forecast.
- Shang (2013): nonparametric modeling of electricity demand curves using FPCA; forecast FPCA scores; allowing forecast of partially complete curves.

Loss Reserving Model

Modeling Idea:

- Obtain PCA decomposition of complete CLR curves.
  - Mean curves, loadings and scores.
- Forecast PCA scores.
  - Within one company: time-series forecast.
  - Among all companies: generalized linear model.
- Forecast the incomplete portion of the CLR curve using Penalized Least Squares.
  - Least square on partially complete curve with PCA score forecasts as regularization constraint.
  - Hyperparameters: number of PCA loadings; shrinkage factor.
- Calculate ultimate cumulative loss and loss reserve estimates.
- Construct prediction interval using nonparametric bootstrap.

Point Forecast Samples



(a) Incremental Loss Ratios

(b) Cumulative Loss Ratios

Figure: Number of PCA factors: 3; solid curves: shrinkage parameter = 0; dashed curves: shrinkage parameter = 0.08

Conclusion

- Application of statistical learning method to traditional insurance data.
- Find different development pattern in company covariates and across historical periods.
- Propose loss reserving model based on nonparametric functional forecast.

Appendix: Loss Triangle

	Development Lag									
Accident Year	0	1	2	3	4	5	6	7	8	9
1999	2739	3942	4515	4988	5429	5418	5421	5430	5463	5463
2000	1964	3019	3407	3816	3969	4061	4062	4134	4212	
2001	1813	2880	3368	3888	3900	3934	3943	3954		
2002	1309	2121	2648	2811	2811	2830	2893			
2003	961	1558	1765	1816	1846	1876				
2004	916	1223	1254	1426	1663					
2005	790	1221	1290	1290						
2006	849	1354	1476							
2007	874	1171								
2008	792									

Table: Sample cumulative paid losses triangle in calendar year 2008

Back to slides

Appendix: Medical Data with Delayed Reporting

Back to slides

Diagnosis period		Reporting delay interval (quarters):									Total reports to end
Year	Quarter	0†	1	2	3	4	5	6	···	≥14	of 1992
1988	1	31	80	16	9	3	2	8	···	6	174
	2	26	99	27	9	8	11	3	···	3	211
	3	31	95	35	13	18	4	6	···	3	224
	4	36	77	20	26	11	3	8	···	2	205
1989	1	32	92	32	10	12	19	12	···	2	224
	2	15	92	14	27	22	21	12	···	1	219
	3	34	104	29	31	18	8	6	···		253
	4	38	101	34	18	9	15	6	···		233
1990	1	31	124	47	24	11	15	8	···		281
	2	32	132	36	10	9	7	6	···		245
	3	49	107	51	17	15	8	9	···		260
	4	44	153	41	16	11	6	5	···		285
1991	1	41	137	29	33	7	11	6	···		271
	2	56	124	39	14	12	7	10			263
	3	53	175	35	17	13	11				306
	4	63	135	24	23	12					258
1992	1	71	161	48	25						310
	2	95	178	39							318
	3	76	181								273
	4	67									133

Figure: Number of AIDS reported in England and Wales at the end of 1992[1]

Appendix: Data Preprocessing

Data and prepocessing

- Company level triangles in workers' compensation line from 1996 to 2008.
- At least \$1 million net premium earned (NPE) in every accident year.
- The highest NPE is at most 10 times more than the lowest NPE.
- Take the fully developed accident year (top row).

Appendix: Data Description and Preprocessing

Company Covariates Data (SNL)

- Business focus: commercial vs personal.
- Capital structure: stock vs mutual.
- Geographical focus: Northeast, South, Midwest, West, National.
- Remove NAs.

Appendix: Additional Figures