Functional Analysis of Loss-development Patterns in P&C Insurance

Arthur Charpentier*

Université du Québec à Montréal. 405 Rue Sainte-Catherine Est, Montréal, QC H2L 2C4, Canada

Qiheng Guo†

Ball State University. 2000 W University Avenue. Muncie, IN 47306. USA

Michael Ludkovski

University of California, Santa Barbara. Santa Barbara, CA 93106. USA

September 2025

**Abstract**

We analyze loss development in NAIC Schedule P loss triangles using functional data analysis methods. Adopting the functional viewpoint, our dataset comprises 3300+ curves of incremental loss ratios (ILR) of workers' compensation lines over 24 accident years. Relying on functional data depth, we first study similarities and differences in development patterns based on company-specific covariates, as well as identify anomalous ILR curves. The exploratory findings motivate the probabilistic forecasting framework developed in the second half of the paper. We propose a functional model to complete partially developed ILR curves based on partial least squares regression of PCA scores. Coupling the above with functional bootstrapping allows us to quantify future ILR uncertainty jointly across all future lags. We demonstrate that our method has much better probabilistic scores relative to Chain Ladder and in particular can provide accurate functional predictive intervals.

*JEL Classification: G22, C14.*
*Keywords:* loss development; loss reserving; incremental loss ratios; unsupervised learning; functional data; functional depth; outlier detection; IBNR

*This work was supported for AC by the NSERC Discovery Grant; under Grant [RGPIN-2019-04190] and the SCOR Foundation for Science.

†Corresponding author. Phone: +1-(765)-285-8645. E-mail address: qguo@bsu.edu.

# 1 Introduction

## 1.1 Motivation and Literature

A property and casualty (P&C) insurer sells insurance policies in multiple businesses. The risks, or loss distributions, differ among business lines and spread across time as claims take multiple years to be fully paid off. When aggregating all claims within a company's given business line, paid losses gradually accumulate each year starting from the year claims are filed. Linking paid loss observations across years results in a *loss development* pattern, indexed by accident years and development lags. Studying loss development helps the insurer predict its future paid losses, or loss reserves, a crucial part in its balance sheet.

Different business lines have different patterns in loss development. Generally speaking, property lines, such as auto physical damage, have a loss development of 1-3 years, while casualty businesses, such as private passenger auto liability and workers' compensation, can have a loss development of 5-30 years. Within each casualty line, the amount of losses in each development lag is heterogeneous across the industry and business lines. While loss development patterns are similar for some personal liability businesses, in commercial liability insurance lines, such as workers' compensation and medical malpractice, loss development varies a lot between companies depending on types of claims and length of litigation process. It remains an open question to analyze loss development heterogeneity and its links to insurance company characteristics.

Occasionally, there are anomalies in loss development. For example, a few companies might have extremely large incremental paid losses in one of the later development years. Whether it is accounting error, data corruption or large claims that coincide to have payments in the same year, such anomaly would not represent the average loss development experience. Hence, it is crucial when making conclusions about industry-wide loss development patterns, to identify and make proper treatments to such anomalies/outliers. Loss development patterns and anomalies have a direct impact on a P&C insurer's financials in terms of reserves and capital. For regulators overseeing the entire industry and reinsurers interested in exposures to the P&C market, knowing the similarity and particularity of loss development can be helpful in setting strategic plans and making decisions.

The study of loss development has largely revolved around historical paid loss in the form of *loss triangles*. In fact, every P&C insurer is required to record business line level loss triangles for the past 10 accident years as part of Schedule P in their annual regulatory NAIC (National Associate of Insurance Commissioners) filings. Table 1 lists a sample loss triangle of cumulative paid losses with a full 10-year loss development for the accident year nine years prior to the current year, and partial loss development for the years thereafter. Development triangles also feature in non-insurance applications, for example for tracking disease cases that often have a delay between diagnosis and reporting, or hospital data that tracks infections and appearance of symptoms. Loss triangles are key for actuarial reserving (Mack, 1993; Wűthrich and Merz, 2008; Radtke et al., 2016), which involves predicting losses to be paid out in the future years, a.k.a. the lower right part of the loss triangle. Reserving analyses use parametric or non-parametric methods based on a triangle of 55 data points. However, the limitations of such cell-based methods motivate a functional perspective, as developed in this paper.

In the later development lags, the scarcity of data points leads to unreliable point estimates and their standard errors. With loss triangle data available at business line level for many years, insurers can borrow information from industry data to improve its individual forecasts while accounting for heterogeneity of triangles. In our analysis below we explore runoff loss triangles in workers' compensation business line for 100+ companies in the U.S., aiming to pool multiple triangles. There is an emerging strand of literature exploring dependence structure between business lines (Shi and Frees, 2011; Zhang and Dukic, 2013) and studying the triangles in multi-company context (Shi, 2017). We also mention the related works that deal with individual claims development, aiming to leverage policy-level covariates to predict loss development (Wüthrich, 2018; Taylor, 2019; Kuo, 2019; Gabrielli, 2021; Cai et al., 2024). Recently a variety of neural network methods (Wüthrich, 2018; Schneider and Schwab, 2025) have been proposed.

Beyond projecting the most likely trajectory, it is important to also quantify uncertainty regarding future loss development factors. A common strategy for that purpose is bootstrapping (Steinmetz and Jentsch, 2024). Alternatively, there are probabilistic predictive models, such as Gaussian-process based methods (Ludkovski and Zail, 2022; Lally and Hartman, 2018; Ang and Lee, 2022), and multi-output GPs (Griffith, 2025). From a parametric perspective, multivariate dependence can be captured by Sarmanov distributions (Abdallah and Wang, 2023), copula (Shi and Frees, 2011), categorical embedding (Shi and Shi, 2023), and time series techniques (Nieto-Barajas and Targino, 2021). Shi and Hartman (2016) considered a Bayesian hierarchical model where a single hyper-prior is used for loss developments. However, their method does not allow for grouping by covariates or any distance metrics we envision.

From the statistical direction, we leverage functional principal component analysis techniques (PCA) and functional depth methods. The version of functional PCA with penalized linear regression is a modification of one by Shang (2013); see also Shang (2017); Elías et al. (2022); Shang and Hyndman (2011). For depth methods we refer to Dai et al. (2020); Narisetty and Nair (2016); Sun et al. (2012).

## 1.2 Contributions and Roadmap

We present a study of the aggregate paid loss development in the property and casualty (P&C) insurance market using functional data analysis tools. In contrast to considering a given triangle as the base object to be modeled, we adopt a *functional data* viewpoint, thinking of loss development as curves, namely as a series of Incremental Loss Ratios (ILR) or Cumulative Loss Ratios (CLR), indexed by the lag year $x$. Moreover, instead of relying on a Markov property that underlies, say the Mack Chain Ladder method and assumes that the next ILR depends just on the previous year, we allow for a full path-dependency. This concept is critical for sorting among hundreds of triangles.

Our initial goals are (i) identifying commonalities in loss development patterns within a business line; (ii) detecting anomalies in loss development curves; (iii) investigating similarities and differences in development patterns pertaining to groups of company-specific factors (e.g. capital structure, region) and across different historical periods. In the second half of the paper we leverage the insights from (i)-(iii) to introduce a novel (iv) probabilistic forecasting of loss development given a partially developed curve.

Functional data depth provides a rigorous tool to identify central patterns and anomalies in the functional loss ratio data. Thus, from a practical perspective, it helps us learn the relationship between loss development curves through a single metric and to get a better picture of typical development shape, identify anomalous observations via statistical outlier techniques, and mitigate the impact of the respective outlier data. Similarly, we propose to exploit the functional viewpoint to probabilistically forecast future ILRs. Our approach below could be labeled semi-parametric: in addition to using regression to capture similarity of different curves and employing the resulting fit to project future loss ratios, we also directly "extend" the existing partially developed history in a functional manner rather than cell-by-cell. Moreover, we use functional bootstrapping to quantify future uncertainty jointly across all the future lags.

Our findings show that the loss development pattern can be dissected down to (a) near-term (development lag 0), (b) medium-term (lags 1-2) and (c) long-term (lags 3-9) components. Most outliers occur in the long-term segment, while the most variation in patterns happens in the medium-term segment. There are also some extreme short-term patterns, but given the variability across firms, most of these should not be classified as true outliers.

We then examine the aggregate patterns based on time-invariant company characteristics, including business focus (commercial vs. personal business), geographical focus (national vs. regional), capital structure (mutual vs. stock), and across historical periods from accident year 1987 to 2010. We find distinctive loss development patterns between commercial and personal business. The companies focusing on the Midwest region have a different development pattern than the rest, and the late 80s' pattern differs from the 90s' and from the 2000's. This study of association between company factors and loss development, to our knowledge, does not exist in either empirical insurance or the actuarial science literature.

For probabilistic completion of partially developed losses, we find that ILR analysis is more flexible, in particular to remove outliers prior to fusing many curves into a single dataset. We also demonstrate that our approach provides better predictive intervals than CL (which tends to underestimate future uncertainty), while providing comparable MSE on ultimate CLR. Last but not least, we show the utility of building functional predictive envelopes beyond the standard marginal credible intervals.

Empirically, our work lays the foundation for insurance companies to improve on loss reserving by pooling company specific information and industry-wide data. The tools are also useful for regulators to detect unusual loss development patterns and study patterns by region and company types. Our findings shed light on fusing information from industry-wide Schedule P data and on bootstrapping empirical triangles to make non-parametric probabilistic forecasts of future loss development.

The rest of the paper is organized as follows. Section 2 presents our data, and we perform an exploratory functional analysis of our dataset that spans 137 insurance companies in Section 3. Section 4 describes our main proposal of functional forecasting of incomplete development curves. Moving from some illustrative results in Section 4, Section 5 quantifies the performance of our method and benchmarks it against the classical Mack Chain-ladder (CL) approach. Section 6 concludes.

# 2 Data

Our empirical analysis relies on industry-wide loss development data for the U.S. workers' compensation line of business. The data[1] originate from Schedule P of the National Association of Insurance Commissioners (NAIC) statutory filings, which require every property and casualty (P&C) insurer to report annual business line loss triangles for the past ten accident years. These triangles form the foundation of regulatory reserving analyses and are widely used in academic and practitioner research (see, e.g., Mack, 1993; Wűthrich and Merz, 2008; Radtke et al., 2016).

## 2.1 Structure of the Data

For each company $i$ and accident year $t$, the cumulative paid losses are recorded across development lags $x = 0, \dots, 9$, denoted $C_{i,t}(x)$. Incremental losses are obtained by

$$Y_{i,t}(x) = \begin{cases} C_{i,t}(x) - C_{i,t}(x-1), & x > 0, \\ C_{i,t}(0), & x = 0, \end{cases}$$

and normalized by the corresponding net premium earned $P_{i,t}$ to form incremental and cumulative loss ratios:

$$y_{i,t}(x) = \frac{Y_{i,t}(x)}{P_{i,t}}, \qquad c_{i,t}(x) = \frac{C_{i,t}(x)}{P_{i,t}}.$$

These ratios provide scale-free measures of loss development, enabling meaningful cross-company comparisons (Shi and Frees (2011); Zhang and Dukic (2013); Shi (2017)). We analyze triangles of *cumulative paid losses* as displayed in Table 1, emphasizing our cross-sectional focus with 100+ companies and 3000+ company-accident year development curves.

We next describe how we selected the final dataset from these raw filings.

## 2.2 Sample Selection

We focus on workers' compensation triangles from 1987 to 2010. After preprocessing, the dataset comprises 3,362 company-year observations across more than 100 insurers, each with at least \$1 million in net premium earned per accident year. This threshold excludes marginal portfolios with negligible exposure, thereby ensuring the stability of loss ratio curves. To reduce distortions from extraordinary corporate events, we additionally exclude companies whose earned premium varies by more than a factor of 10 across accident years, since such

[1]Provided to us by a leading insurance analytics firm which undertook proprietary data wrangling and cleaning

| Accident Year | Earned Premium | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $t_i$ | $P_{i,t_i}$ | $C_{i,t_i}(0)$ | $C_{i,t_i}(1)$ | ... | ... | ... | ... | ... | ... | ... | $C_{i,t_i}(9)$ |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| $T-9$ | $P_{i,T-9}$ | $C_{i,T-9}(0)$ | $C_{i,T-9}(1)$ | ... | ... | ... | ... | ... | ... | ... | $C_{i,T-9}(9)$ |
| $T-8$ | $P_{i,T-8}$ | $C_{i,T-8}(0)$ | ... | ... | ... | ... | ... | ... | ... | $C_{i,T-8}(8)$ | |
| $T-7$ | $P_{i,T-7}$ | $C_{i,T-7}(0)$ | ... | ... | ... | ... | ... | ... | ... | | |
| $T-6$ | $P_{i,T-6}$ | $C_{i,T-6}(0)$ | ... | ... | ... | ... | ... | ... | | | |
| $T-5$ | $P_{i,T-5}$ | $C_{i,T-5}(0)$ | ... | ... | ... | ... | ... | | | | |
| $T-4$ | $P_{i,T-4}$ | $C_{i,T-4}(0)$ | ... | ... | ... | ... | | | | | |
| $T-3$ | $P_{i,T-3}$ | $C_{i,T-3}(0)$ | ... | ... | ... | | | | | | |
| $T-2$ | $P_{i,T-2}$ | $C_{i,T-2}(0)$ | ... | $C_{i,T-2}(2)$ | | | | | | | |
| $T-1$ | $P_{i,T-1}$ | $C_{i,T-1}(0)$ | $C_{i,T-1}(1)$ | | | | | | | | |
| $T$ | $P_{i,T}$ | $C_{i,T}(0)$ | | | | | | | | | |

Table 1: Sample cumulative paid losses triangle in calendar year $T$ for one line in one company.

swings typically reflect acquisitions, divestitures, or book transfers that break comparability of historical development.

Each insurer is identified by an NAIC code (e.g., "C121"), and we enrich the triangle data with time-invariant company characteristics: (i) *Geographic focus* (National, Midwest, Northeast, South, West); (ii) *Ownership structure* (Stock, Mutual, Other); (iii) *Business focus* (Personal, Commercial, Workers' Compensation). (those covariate are *time-invariant*, as opposed to earned premiums, that is inherently *time-varying*). This augmentation allows us to investigate heterogeneity in development patterns across company types, consistent with earlier multi-company reserving studies (e.g., Shi, 2017; Gabrielli, 2021).

## 2.3 Descriptive Features

Figure 1 summarizes the distribution of observations by region, ownership, and business focus. Across accident years, the number of reporting insurers varies between 70 and 224, with 115 companies contributing at least 15 years of observations. This breadth of coverage provides a rich empirical basis for functional data analysis of loss development. In contrast to traditional single-triangle reserving studies, our cross-sectional perspective allows pooling of information across insurers while accounting for systematic differences in underwriting portfolios and corporate structure.

In sum, our dataset combines the regulatory completeness of NAIC Schedule P triangles with additional company-level covariates, thereby enabling us to study not only typical loss development patterns but also the heterogeneity and anomalies that arise across insurers and time. This design sets the stage for the functional data methods developed in the following sections.

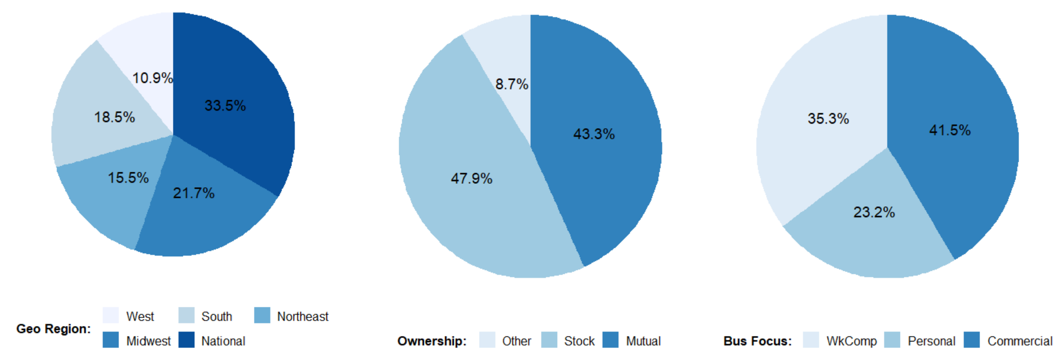



Figure 1: Distribution of sample development triangles by Geographical Region, Ownership Structure and Business Focus.

# 3 Exploratory Analysis of Runoff Triangles

Before developing our forecasting framework, we first investigate the empirical properties of incremental loss ratio curves in our dataset. Exploratory analysis is essential for understanding both the typical shape of loss development and the anomalies that may distort statistical modeling. This section provides descriptive statistics, discusses outlier detection, and examines heterogeneity across company characteristics.

## 3.1 Descriptive Statistics

Table 2 reports summary statistics of ILRs across development lags $x = 0, \ldots, 9$ for workers' compensation paid loss triangles from 1987–2010. The non-robust mean and standard deviation at each lag are systematically larger than the robust measures (median and median absolute deviation, MAD, see Rousseeuw and Croux (1993)). This discrepancy highlights the presence of outlying curves, consistent with earlier findings in multi-company analyses (Shi, 2017; Gabrielli, 2021).

The ILR distribution exhibits a characteristic shape: losses peak at lag 1, then decline sharply at lags 2–3, followed by a long right tail extending up to lag 9. Beyond lag 7, mean ILRs fall below $1\%$, reflecting the long-tailed settlement patterns in workers' compensation claims (Taylor, 2019). Negative ILRs, corresponding to downward adjustments in cumulative paid losses, occur in later lags and introduce skewness and heavy tails.

## 3.2 Outlier Detection

Outliers in development curves may arise from data errors, reinsurance events, or extraordinary claims. Their presence can substantially bias traditional reserving techniques (Wüthrich, 2018). To address this, we adopt two complementary approaches: robust principal component analysis (RPCA, Croux et al., 2007) and functional data depth methods (Sun et al., 2012; Narisetty and Nair, 2016; Dai et al., 2020).

| Lag $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 0.1669 | 0.1900 | 0.1003 | 0.0581 | 0.0344 | 0.0219 | 0.0145 | 0.0109 | 0.0078 | 0.0062 |
| Std. Dev. | 0.0598 | 0.0636 | 0.0429 | 0.0342 | 0.0256 | 0.0212 | 0.0184 | 0.0187 | 0.0166 | 0.0146 |
| Median | 0.1628 | 0.1843 | 0.0948 | 0.0547 | 0.0323 | 0.0196 | 0.0128 | 0.0086 | 0.0062 | 0.0047 |
| MAD | 0.0476 | 0.0556 | 0.0340 | 0.0249 | 0.0184 | 0.0137 | 0.0108 | 0.0084 | 0.0068 | 0.0054 |
| Min | 0.0008 | -0.1038 | -0.1677 | -0.3966 | -0.1881 | -0.2891 | -0.2785 | -0.2808 | -0.3155 | -0.2311 |
| Max | 0.6936 | 0.6647 | 0.4961 | 0.3981 | 0.3643 | 0.3492 | 0.2958 | 0.3224 | 0.2697 | 0.3293 |
| 95% quantile | 0.0817 | 0.1025 | 0.0419 | 0.0150 | 0.0039 | 0.0009 | 0.0000 | 0.0000 | -0.0002 | -0.0001 |
| 99% quantile | 0.2639 | 0.2988 | 0.1773 | 0.1121 | 0.0726 | 0.0510 | 0.0378 | 0.0297 | 0.0234 | 0.0192 |
| Skewness | 1.4431 | 0.7565 | 0.9805 | 0.7105 | 1.0798 | 0.6508 | -0.0472 | 3.5355 | -2.9519 | 2.6788 |
| Kurtosis | 7.84 | 3.19 | 5.12 | 21.72 | 19.34 | 51.10 | 61.89 | 110.70 | 137.08 | 139.89 |
| 5% winsorized mean | 0.1651 | 0.1891 | 0.0995 | 0.0572 | 0.0338 | 0.0214 | 0.0144 | 0.0104 | 0.0077 | 0.0060 |
| 5% winsorized sd | 0.0479 | 0.0532 | 0.0354 | 0.0253 | 0.0182 | 0.0135 | 0.0105 | 0.0084 | 0.0067 | 0.0055 |
| # Positive | 3362 | 3358 | 3355 | 3328 | 3290 | 3240 | 3173 | 3114 | 3051 | 3015 |
| # Zero | 0 | 0 | 0 | 2 | 10 | 24 | 47 | 82 | 133 | 171 |
| # Negative | 0 | 4 | 7 | 32 | 62 | 98 | 142 | 166 | 178 | 176 |

Table 2: Summary statistics of ILRs for workers' compensation paid loss triangles from 1987 to 2010

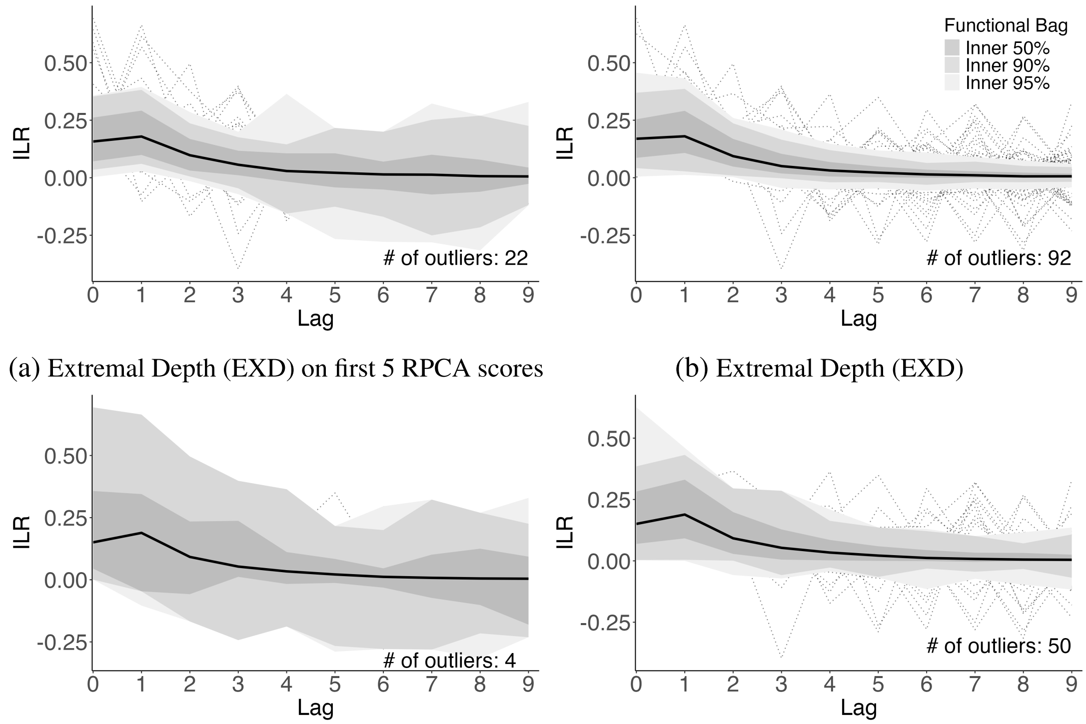
RPCA identifies curves with extreme component scores relative to the cross-sectional distribution. While effective at detecting early-lag anomalies, RPCA tends to produce overly wide tolerance bands at longer lags, limiting its discriminating power. In contrast, functional depth provides an ordering of curves from central to extreme, yielding interpretable notions of functional medians and quantile envelopes. We consider three widely used depth notions: Band Depth (BD), Modified Band Depth (MBD), and Extremal Depth (EXD). Consistent with the recommendations in Narisetty and Nair (2016), we find EXD to be most effective, producing tight central envelopes and flagging curves with unusual long-tail behavior (Figure 2 ).

The number of curves classified as outliers varies significantly by method: MBD flags only a handful of series, BD captures approximately 50 curves, while EXD identifies over 90. This sensitivity underscores the importance of choosing an outlier detection method consistent with actuarial intuition: wide variability in short lags but tighter envelopes in later lags where reserves are most sensitive. Additional technical details (with proper definitions of BD, MBD, and EXD) are provided in Appendix A.1.

## 3.3 Impact of Covariates

Having identified central patterns and outliers, we next investigate systematic heterogeneity driven by insurer characteristics. Beyond industry-wide averages, loss development is influenced by systematic company- and market-level characteristics. Understanding these covariates is crucial both for regulators, who monitor systemic risks across heterogeneous insurers, and for actuaries, who must ensure that reserving models remain calibrated to the specific portfolio under study. In this subsection, we stratify ILR curves by several company features and discuss the resulting heterogeneity.

**Business focus.** We first distinguish insurers according to their primary line of business: personal, commercial, and workers' compensation specialists. Figure 3(a) shows that median




(a) Extremal Depth (EXD) on first 5 RPCA scores

(b) Extremal Depth (EXD)

(c) Modified Bandwidth Depth (MBD)

(d) Bandwidth Depth (BD)

Figure 2: Identification of outlying loss development curves. We compare three different functional depth metrics to identify extreme ILR curves.

ILR curves are broadly similar across groups, with overlapping EXD-based 50% envelopes. This suggests that, within the workers' compensation line, loss development is relatively stable across broader business strategies. This finding echoes the results of Shi (2017), who documented strong within-line consistency of development patterns even for diversified insurers. Nevertheless, slight differences at lag 0 and lag 1 may reflect different claims handling speeds across personal versus commercial insurers.

**Ownership structure.** Ownership type (stock, mutual, or other) can affect corporate governance, risk appetite, and reserving practices (Cummins and Weiss, 2000). In our dataset, Figure 3(b) shows modest differences: mutual companies tend to report slightly higher ILRs in early lags, potentially reflecting more conservative settlement practices, whereas stock companies display marginally faster runoff. Although the median differences are small, this dimension remains important for reserve adequacy studies, since ownership type may interact with capital management and regulatory scrutiny.

**Geographic focus.** Geographic diversification is another relevant factor. Regional insurers (Midwest, Northeast, South, West) may be more exposed to jurisdictional differences in workers' compensation law, medical inflation, or litigation practices (Weiss, 2001). As shown in Figure 3(c), median curves exhibit mild regional variation, with Midwest-focused insurers displaying slightly higher lag-3 and lag-4 ILRs. Such effects, although subtle in aggregate, can accumulate to material differences in ultimate loss ratios, and highlight the importance of incorporating jurisdictional heterogeneity into multi-company reserving analyses.

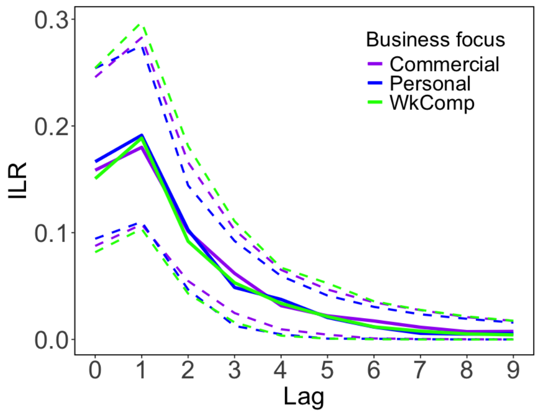
**Earned premium (company size).** Company size, proxied by annual net premium earned (NPE), reveals a more striking pattern (Figure 3(d)). We partition firms into three groups: small (<\$10m annual NPE), medium (\$10m–\$100m), and large (>\$100m). Small insurers display the greatest volatility, with wider envelopes across nearly all lags, consistent with classical credibility theory (Bühlmann, 1967). Medium-sized insurers exhibit the sharpest peak at lag 1, followed by steep declines, while large insurers show smoother curves and persistently higher ILRs beyond lag 4. This suggests that larger insurers, possibly due to more complex claims portfolios and protracted litigation, settle a higher proportion of losses at longer horizons. The empirical evidence thus aligns with practitioner intuition: smaller companies tend to close claims more quickly but with greater variability, whereas larger companies carry longer tails but exhibit more stable patterns.

**Calendar time.** Finally, we investigate accident-year effects, which capture market-wide cycles in workers' compensation. Figure 3(e) reveals recurrent peaks in median ILRs during the late 1980s, mid-1990s, and late 2000s. These coincide with well-documented underwriting cycles (Cummins et al., 1992; Harrington and Niehaus, 2000) and suggest that external factors—such as interest rate environments, medical cost inflation, and regulatory changes—play a significant role in shaping development patterns. Notably, while levels fluctuate across periods, we do not observe a monotone trend in ILRs, underscoring the cyclical rather than secular nature of these effects.

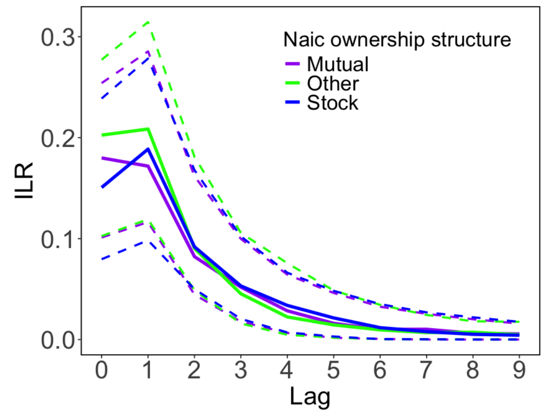
Taken together, these results highlight three key insights. First, ownership and business focus appear to exert only limited influence, consistent with the dominance of line-specific settlement dynamics. Second, company size is a primary driver of heterogeneity, with small insurers displaying the most volatile development. Third, accident-year effects reveal cyclical variation linked to market conditions. These findings reinforce the need for models that can flexibly account for both cross-sectional and temporal heterogeneity, motivating our functional approach in Section 4.

# 4 Forecasting Incremental Loss Ratios (ILR) Curves

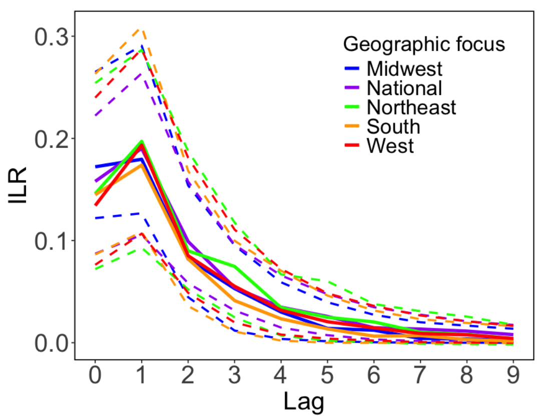
We now develop a probabilistic forecasting framework for ILR curves based on partially observed loss development triangles. Our aim is to forecast the continuation of a curve $y_{n^*}(x)$, observed only up to lag $s-1$, by exploiting information contained in the historical database. The methodology builds on functional principal component analysis (FPCA), regression of functional scores on covariates, and penalized curve completion. We also construct bootstrap-based predictive intervals to quantify uncertainty.



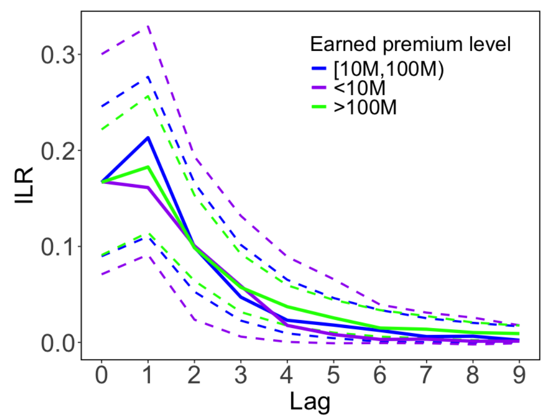

(a) Business focus




(b) Ownership structure

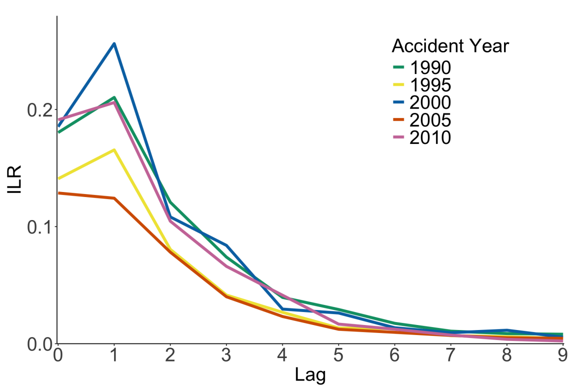



(c) Geographic region

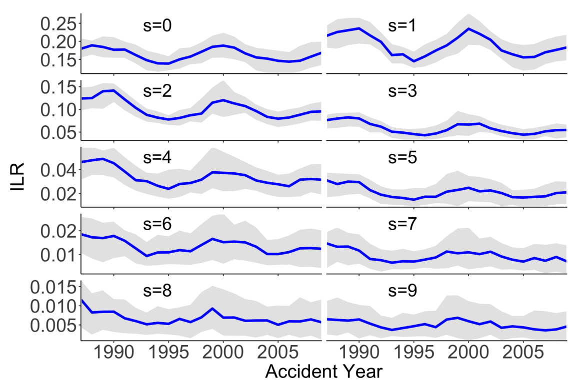



(d) Earned premium level

Figure 3: Median ILR curves and EXD-based 50% envelopes stratified by company characteristics: (a) business focus, (b) ownership, (c) geographic region, (d) earned premium category.




(a) ILR median curve for select AYs




(b) Median and interquartile range of ILR by lag $s$. $y$-axes vary by row.

Figure 4: ILR in relation to accident year

## 4.1 Functional Representation and Matrix Formulation

Let $\mathcal{X} = \{0, \dots, 9\}$ denote the set of development lags. For each company–accident year pair $n \in \mathcal{N} = \{i = 1, 2, \dots, I\} \times \{t = t_i, t_i + 1, \dots, T - 9\}$ , we observe a complete ILR curve $\{y_n(x) : x \in \mathcal{X}\}$. We view $y_n$ as a function in the Hilbert space $L^2(\mathcal{X})$ with inner product $\langle f, g \rangle = \sum_{x \in \mathcal{X}} f(x)g(x)$. Let $\mu(x) = \mathbb{E}[y_n(x)]$ denote the mean function and $\Gamma(x, x') = \mathrm{Cov}(y_n(x), y_n(x'))$ the covariance kernel.

By the Karhunen–Loève decomposition (see Ramsay and Silverman (2005)),

$$y_n(x) = \mu(x) + \sum_{k=1}^{\infty} \phi_k(x)\beta_{k,n}, \qquad \beta_{k,n} = \langle y_n - \mu, \phi_k \rangle, \tag{1}$$

where $\{\phi_k\}$ are orthonormal eigenfunctions of $\Gamma$ with decreasing eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. In practice, we truncate to the first $K$ components, consistently with Equation (4),

$$y_n(x) = \mu(x) + \sum_{k=1}^{K} \phi_k(x)\beta_{k,n} + \varepsilon_n(x) \approx \mu(x) + \sum_{k=1}^{K} \phi_k(x)\beta_{k,n}, \tag{2}$$

where $K$ is chosen via cross-validation or variance explained criterion (Shang and Hyndman, 2011; Elías et al., 2022). The residual term represents high-order variation not captured by the factor structure.

Let $Y \in \mathbb{R}^{N \times |\mathcal{X}|}$ be the data matrix with entries $Y_{n,x} = y_n(x)$. Let $\Phi \in \mathbb{R}^{|\mathcal{X}| \times K}$ denote the matrix of estimated eigenfunctions $\hat{\phi}_k$ evaluated on $\mathcal{X}$. Then (2) may be written as

$$Y \approx \mathbf{1}\mu^\top + B\Phi^\top, \tag{3}$$

where $B \in \mathbb{R}^{N \times K}$ contains the scores $\hat{\beta}_{k,n}$. Thus FPCA is equivalent to low-rank factorization of $Y$.

## 4.2 Regression Priors and Partial Curve Completion

For a mathematical formulation, the idea is that each complete ILR curve is represented through functional PCA as

$$y_n(x) = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\phi}_k(x)\, \hat{\beta}_{k,n} + \varepsilon_n(x), \qquad x \in \{0, \dots, 9\}, \tag{4}$$

where $\hat{\mu}$ is the empirical mean function, $\hat{\phi}_k$ are the estimated functional principal components, and $\hat{\beta}_{k,n}$ are the scores for curve $n$.

Each score vector $\beta_n = (\beta_{1,n}, \dots, \beta_{K,n})^\top$ encodes the shape of curve $n$. To incorporate company features $X_n$ (earned premium, ownership, region, etc.), we model

$$\beta_{k,n} = f_k(X_n) + \xi_{k,n}, \qquad k = 1, \dots, K, \tag{5}$$

where $f_k$ is approximated by a linear form $f_k(X) = X^\top \theta_k$ estimated via LASSO (Tibshirani, 1996) to encourage sparsity. Let $\hat{\beta}^{\mathrm{RM}}_{k,n^*} = X_{n^*}^\top \hat{\theta}_k$ denote the regression-based prior score for a new curve $n^*$.

Suppose curve $n^*$ is observed up to lag $s-1$, i.e., we observe $y_{n^*}(x)$ for $x \in \mathcal{X}_e = \{0, \ldots, s-1\}$. We complete the curve by solving a penalized least squares problem:

$$\hat{\beta}^{\mathrm{PLS}}_{n^*} = \mathrm{argmin}_{\beta \in \mathbb{R}^{K(s)}} \left\{ \sum_{x \in \mathcal{X}_e} \left( y_{n^*}(x) - \hat{\mu}(x) - \sum_{k=1}^{K(s)} \hat{\phi}_k(x) \beta_k \right)^2 + \lambda(s) \sum_{k=1}^{K(s)} \left( \beta_k - \hat{\beta}^{\mathrm{RM}}_{k,n^*} \right)^2 \right\}. \quad (6)$$

The first term enforces fit to the partial curve, while the second shrinks towards the regression prior.

**Closed form.** Let $y_e$ be the vector of observed lags, $\Phi_e \in \mathbb{R}^{s \times K}$ the restriction of $\Phi$, and $\mu_e$ the restricted mean. Then (6) has solution

$$\hat{\beta}^{\mathrm{PLS}}_{n^*} = \left( \Phi_e^\top \Phi_e + \lambda(s) I_{K(s)} \right)^{-1} \left( \Phi_e^\top (y_e - \mu_e) + \lambda(s) \hat{\beta}^{\mathrm{RM}}_{n^*} \right). \quad (7)$$

This highlights the shrinkage effect of $\lambda(s)$: for small $s$, regression priors dominate, while as $s$ increases, the fit to partial observations becomes more important.

## 4.3 Forecasting and Prediction Intervals

The predicted ILR curve for future lags $\mathcal{X}_\ell = \{s, \ldots, 9\}$ is

$$\hat{y}_{n^*}(x) = \hat{\mu}(x) + \sum_{k=1}^{K(s)} \hat{\phi}_k(x) \, \hat{\beta}^{\mathrm{PLS}}_{k,n^*}, \qquad x \in \mathcal{X}_\ell. \quad (8)$$

The corresponding cumulative loss ratio (CLR) forecast is

$$\hat{c}_{n^*}(x) = c_{n^*}(s-1) + \sum_{j=s}^{x} \hat{y}_{n^*}(j), \qquad x \in \mathcal{X}_\ell. \quad (9)$$

To quantify forecast uncertainty, we construct bootstrap prediction intervals. Let $\epsilon_n(x) = y_n(x) - \hat{\mu}(x) - \sum_{k=1}^{K(s)} \hat{\phi}_k(x) \hat{\beta}_{k,n}$ denote residuals. We assume $\epsilon_n$ are independent across $n$ with mean zero.

**Algorithm.** For $b = 1, \ldots, B$:

1. Resample company features and curves $(X_n, y_n)$.
2. Refit regressions (5), obtain $\hat{\beta}^{\mathrm{RM},(b)}$.
3. Solve (7) to obtain $\hat{\beta}^{\mathrm{PLS},(b)}_{n^*}$.

4. Resample residual function $\epsilon^{(b)}$ and set

$$\hat{y}_{n^*}^{(b)}(x) = \hat{\mu}(x) + \sum_{k=1}^{K(s)} \hat{\phi}_k(x)\, \hat{\beta}_{k,n^*}^{\mathrm{PLS},(b)} + \epsilon^{(b)}(x). \tag{10}$$

Pointwise $(1-\alpha)$ prediction intervals are obtained from the empirical quantiles of $\{\hat{y}_{n^*}^{(b)}(x)\}$. Simultaneous bands are formed by functional depth (Narisetty and Nair (2016); Dai et al. (2020)), defining the central $100(1-\alpha)\%$ region. More precisely, from our bootstrapped ILR curves, we construct $100(1-\alpha)\%$ central region using several functional depth methods. Let $\mathcal{X}_e = \{0, \ldots, s-1\}$ be observed lags and $\mathcal{X}_\ell = \{s, \ldots, 9\}$ the missing lags. The construction follows equations (2) and (5) in Narisetty and Nair (2016):

$$C_{1-\alpha} = \{y \in S : y_n^L(x) \leq y(x) \leq y_n^U(x),\ \forall x \in \mathcal{X}_\ell\}, \tag{11}$$

where $S = \{\hat{y}_{n^*}^{PLS,1}(x), \ldots, \hat{y}_{n^*}^{PLS,B}(x)\}, x \in \mathcal{X}_\ell$ is the collection of all bootstrapped predicted ILR curves and

$$\hat{y}_{n^*}^{D,L}(x) := \inf\left\{y_{n^*}(x) : y \in S,\ D(y_{n^*}, \mathbb{P}_{n^*}) > \alpha\right\},$$
$$\hat{y}_{n^*}^{D,U}(x) := \sup\left\{y_{n^*}(x) : y \in S,\ D(y_{n^*}, \mathbb{P}_{n^*}) > \alpha\right\},$$

with $D$ representing a functional depth measure, and $\mathbb{P}_{n^*}$ the empirical distribution of bootstrapped predicted ILR curves. The lower and upper bounds of CLR can be obtained similarly from $S = \{\hat{c}_{n^*}^1(x), \ldots, \hat{c}_{n^*}^B(x)\}, x \in \mathcal{X}_\ell$. For $s = 9$, since the forecast is one additional point, the depth-based methods essentially yield the point-wise prediction interval. We consider extremal depth (EXD) (Narisetty and Nair, 2016) for $D(y, \cdot)$.

Our method can be viewed as a semi-parametric hierarchical model:

$$y_n(x) \mid \beta_n \sim \mathcal{N}\Big(\mu(x) + \sum_k \phi_k(x)\beta_{k,n}, \sigma^2(x)\Big), \tag{12}$$

$$\beta_{k,n} \mid X_n \sim f_k(X_n) + \xi_{k,n}. \tag{13}$$

Thus, uncertainty arises from both regression residuals and functional residuals. The bootstrap mimics this hierarchical structure by resampling both components.

## 4.4 Fixed-Origin Backtesting

To test out the forecast and bootstrap algorithm, we implement a fixed–origin backtesting procedure. Specifically, we fix the calendar year $T$ and treat accident year $T$ as "current," with ILR curves available only up to lag $s-1$. We then forecast the continuation of accident year $T$ using our proposed method and compare against realized values once the full triangle is known. This exercise mimics the practical reserving task of predicting partially developed cohorts. We implement this workflow on three representative companies P52, C260 and P1406. Table 3 lists their covariates.

| Company | Business Focus | Ownership Structure | Geographic Focus | Net Premium Earned in 2010 |
|---|---|---|---|---|
| P53 | WkComp | Stock | West | 321,786,000 |
| C260 | Personal | Mutual | South | 177,078,000 |
| P1406 | Personal | Mutual | National | 310,728,000 |

Table 3: Company covariates for P53, C260 and P1406, cf. Figure 9.

**Setup.** Let $\mathcal{D}_{\text{train}}(s) = \{y_n(x) : n \in \mathcal{N}, t \leq T - 9, x \in \mathcal{X}\}$ denote the training set comprising fully developed curves up to accident year $T - 9$. For the focal accident year $T$, we observe only $y_{i,T}(x)$ for $x \in \{0, \ldots, s - 1\}$. Denote the predicted continuation as $\hat{y}_{i,T}(x), x \in \{s, \ldots, 9\}$, with corresponding cumulative forecast $\hat{c}_{i,T}(9)$.

**Sequential updating.** We conduct the procedure for $s = 1, \ldots, 9$. At each $s$, the following steps are performed:

1. Estimate FPCA basis $\{\hat{\mu}, \hat{\phi}_k\}$ and regression models $f_k$ from $\mathcal{D}_{\text{train}}(s)$.
2. For accident year $T$, obtain regression prior scores $\hat{\beta}^{\text{RM}}_{k,i,T}$.
3. Apply penalized least squares (6) using the observed partial curve.
4. Compute forecasts $\hat{y}_{i,T}(x)$ and $\hat{c}_{i,T}(9)$.

As $s$ increases, the algorithm is updated with longer partial curves and larger training sets, so forecast precision should improve monotonically.

**Empirical findings.** Figure 5 illustrates typical results for three representative companies. As expected, forecasts are least accurate for $s = 1$–$2$, where limited early development provides weak signal and predictions are dominated by regression priors. Accuracy and calibration improve steadily as $s$ increases, with predictive intervals narrowing around the realized CLR once $s \geq 5$. This confirms that our penalized least squares adjustment balances prior information with observed development, leading to stable and progressively accurate forecasts.

For the full evolution of the partial curve and future forecasts across development years for the three companies both in terms of ILR and CLR, see Figure 9 in Appendix A.3.

## 4.5 Application to Loss Triangle Completion

We collect loss history from $I = 224$ companies from the earliest accident year 1987 labeled as $t = 1$ to the most recent accident year 2010 labeled as $T = 24$. We train the PLS model on $y_{i,t}(x), t \leq T - 9$ and features $X_{i,t}$ including accident year $t$, earned premium $P_{i,t}$ and time-invariant company and business line characteristics. We then forecast the lower right triangle $y_{i,t}(x), x > T - t$ using the model and upper left triangle $y_{i,t}(x), 0 \leq x \leq T - t$ and evaluate the performance against the held-out true value of the lower right triangle. Before carrying

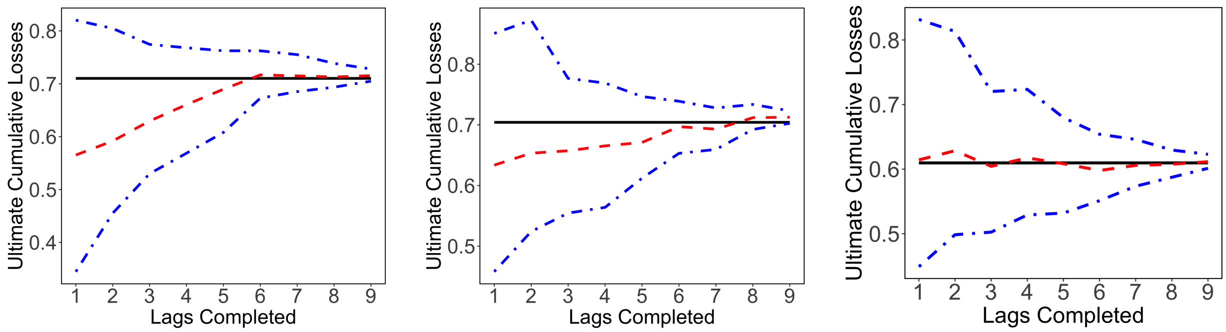



Figure 5: Evolution of ultimate CLR forecast for AY 2010 as a function of developed lag $s$ for three representative triangles listed in Table 3. Black line— ground truth ultimate CLR; red line —mean forecast ultimate CLR; blue lines—95% EXD prediction interval.

out the PCA and regression, we remove the outliers, namely curves with outlying bandwidth depth (BD) scores. A total of 50 outlying ILR curves are removed, see Figure 2.

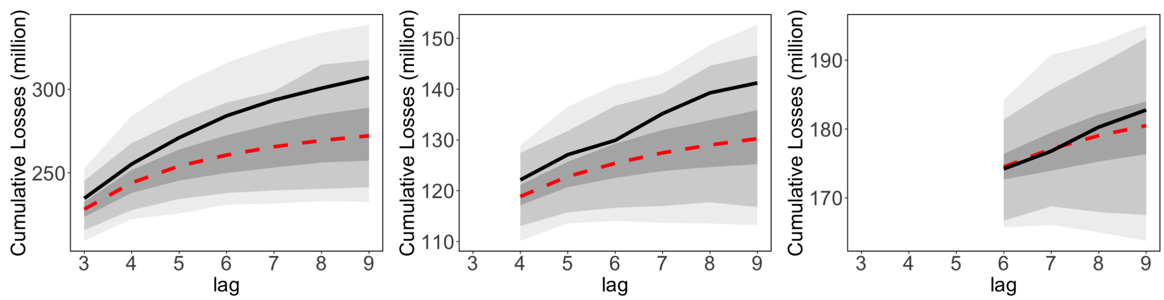
The procedure is similar to the previous section, but we start from accident year 2002, or $t = 16$, with $s = 9$, forecast the remainder of the ILR curves, augment the training set with the completed ILR curves, and update $s$ to $s - 1$. We then obtain the CLR and multiply by earned premium to recover prediction for cumulative losses for comparison with traditional reserving methods.

We report the optimal $K(s)$ and $\lambda(s)$ tuning results in Table 4. The collection of loading vectors used in the model can be found in Table 7 in Appendix A.2. The coefficients of LASSO regression can be found in Table 8 in Appendix A.2. We observe that the LASSO uses all the covariates for shorter lags $s < 7$ and drops most covariates for $s = 7, 8, 9$.

| Accident Year (Lag $s$) | Training Curves | Optimal $K(s)$ | Optimal $\lambda(s)$ | MAPE |
|---|---|---|---|---|
| 2002 (9) | 1550 | 8 | 0.0471 | 0.0065 |
| 2003 (8) | 1710 | 8 | 0.0853 | 0.0117 |
| 2004 (7) | 1882 | 6 | 0.0387 | 0.0192 |
| 2005 (6) | 2063 | 6 | 0.1111 | 0.0272 |
| 2006 (5) | 2251 | 5 | 0.0898 | 0.0396 |
| 2007 (4) | 2445 | 4 | 0.1000 | 0.0557 |
| 2008 (3) | 2653 | 3 | 0.1606 | 0.0820 |
| 2009 (2) | 2866 | 2 | 0.1762 | 0.1284 |
| 2010 (1) | 3088 | 1 | 0.2271 | 0.1825 |

Table 4: Optimal number of PCA factors $K(s)$ and penalty $\lambda(s)$ as a function of number of completed lags $s$.

In Figure 6 we show the resulting forecasts and bootstrapped intervals for several confidence levels. In the left and center panel, the predicted CLR is below the true CLR, but both are covered in the 90% EXD central region. The right panel where more lags are complete, shows more accurate prediction and the true CLR lies in the 50% EXD central region. In the next section, we will formally make assessment on the accuracy and precision of the forecast.




(a) P53: $s = 3$ (b) C260: $s = 4$ (c) P1406: $s = 6$

Figure 6: Illustrating PLS predictions for cumulative losses and respective EXD-based 50%, 90% and 95% functional regions (grey shades). Solid black lines: realized cumulative losses (million USD); dashed red lines: predicted cumulative losses.

# 5 Forecast Assessment

Having introduced the methodology, we now assess its empirical performance in forecasting loss development. Evaluation is performed along three dimensions: (i) point forecast accuracy, (ii) calibration and sharpness of probabilistic forecasts, and (iii) functional coverage of CLR trajectories. We benchmark our approach against the classical Mack chain-ladder (CL) model (Mack, 1993; see also Wűthrich and Merz, 2008), the most widely used reserving method in practice.

## 5.1 Experimental Design

We consider a subset of 137 insurers with fully observed triangles from accident years 2001–2010. This ensures availability of "future truth" for validating predictions. For each accident year $t \in \{T-8, \ldots, T\}$, and for each development horizon $s = 1, \ldots, 9$, we fit the models using only information up to lag $s-1$ and then forecast the continuation of ILR and CLR curves. Our evaluation metrics include:

1. **MAPE of ultimate cumulative losses**

$$\mathrm{MAPE}_t = \frac{1}{I} \sum_{i=1}^{I} \left| \frac{\hat{c}_{i,t}(9) - c_{i,t}(9)}{c_{i,t}(9)} \right|, \quad t \in [T-8, T]$$

2. **Empirical coverage of prediction intervals**, i.e., the fraction of triangles where the

realized value falls within the $(1-\alpha)$ forecast band.

$$\text{Coverage}_{\alpha,t} = \frac{1}{I}\sum_{i=1}^{I} \mathbf{1}\{L_{i,t} \le c_{i,t}(9) \le U_{i,t}\}, \quad t \in [T-8, T]$$

where $[L_{i,t}, U_{i,t}]$ is the $(1-\alpha)$ predictive interval for ultimate CLR.

3. **Functional coverage** of CLR curves, assessed via depth-based envelopes (Narisetty and Nair, 2016; Dai et al., 2020).

$$\text{FuncCoverage}_{\alpha,t}(s) = \frac{1}{I}\sum_{i=1}^{I} \mathbf{1}\{c_{i,t}(x) \in C_{1-\alpha}\}, \; x \in X_l, t \in [T-8, T]$$

$C_{1-\alpha}$ is the depth-based prediction interval as in Equation (11).

4. **Interval Score** (see Gneiting and Raftery (2007)), which jointly evaluates coverage and sharpness. A general formula of interval score of true value $Y$ and prediction upper bound $U$ and lower bound $L$ is

$$IS_\alpha(L, U; Y) = (U - L) + \tfrac{2}{\alpha}(L - Y)\mathbf{1}\{Y < L\} + \tfrac{2}{\alpha}(Y - U)\mathbf{1}\{Y > U\}.$$

We will evaluate the interval score for ultimate cumulative losses and the predicted cumulative loss curve.

This combination of metrics follows best practice in forecast evaluation (Dawid, 1984; Diebold and Mariano, 1995; Patton, 2011). These metrics jointly assess accuracy and calibration, and allow comparison against benchmarks such as Mack's chain-ladder (Mack, 1993), Gaussian process reserving models (Ludkovski and Zail, 2022; Ang and Lee, 2022), and machine learning approaches (Wüthrich, 2018; Schneider and Schwab, 2025).

## 5.2 Point Forecast Accuracy

Table 5 reports MAPE for ultimate cumulative losses. We find that CL tends to perform better for short development horizons ($s = 1$–$4$), where the information content of the partial curve is limited and CL's multiplicative structure is effective. However, our functional PLS method outperforms CL once more lags are observed ($s \ge 5$), reflecting its ability to exploit the functional similarity across triangles and covariate information.

The average interval scores of ultimate cumulative losses based on $s$ developed lags are calculated by weighing each curve by the size of the business line (earned premium),

$$\overline{UIS}_\alpha(s) := \frac{\sum_i IS_\alpha(C_{i,t}(9))P_{i,t}}{\sum_i P_{i,t}}. \tag{14}$$

The two right-most columns of Table 5 compare the PLS method with point-wise and EXD-based intervals against the Mack (1993) chain ladder method in terms of the ultimate interval

| Lag | MAPE | | Coverage % | | | Interval score $\overline{UIS}_{0.95}$ | | |
|---|---|---|---|---|---|---|---|---|
| $s$ | PLS | CL | PLS | EXD | CL | PLS | EXD | CL |
| 1 | 21.86% | 16.53% | 94.16% | 89.78% | 86.13% | 0.7132 | 1.0490 | 0.7098 |
| 2 | 13.11% | 10.31% | 98.54% | 97.08% | 79.56% | 0.4857 | 0.4286 | 1.0422 |
| 3 | 9.06% | 7.21% | 98.54% | 93.43% | 71.53% | 0.2969 | 0.5009 | 0.9904 |
| 4 | 5.31% | 5.18% | 98.54% | 96.35% | 70.80% | 0.2379 | 0.1854 | 0.4047 |
| 5 | 3.90% | 4.10% | 99.27% | 99.27% | 73.72% | 0.1772 | 0.1434 | 0.1808 |
| 6 | 2.38% | 2.69% | 100.00% | 98.54% | 79.56% | 0.1254 | 0.0894 | 0.1176 |
| 7 | 1.80% | 2.07% | 97.81% | 97.81% | 84.67% | 0.1263 | 0.1208 | 0.0900 |
| 8 | 1.14% | 1.41% | 97.81% | 97.81% | 83.94% | 0.0586 | 0.0561 | 0.0669 |
| 9 | 0.56% | 0.67% | 94.89% | 94.89% | 92.70% | 0.0299 | 0.0299 | 0.0354 |

Table 5: Statistical scoring metrics for forecasting ultimate cumulative losses. We report MAPE, coverage at 95% level, and average ultimate interval scores $\overline{UIS}_\alpha$ with $\alpha = 0.95$. PLS refers to point-wise interval, CL refers to chain ladder, and EXD to extremal depth.

score $\overline{UIS}_{0.95}$ on 137 test triangles. The point-wise or EXD method has a favorable interval score for all lags except for lags 1 and 7.

The results reveal a clear trade-off: CL yields slightly lower MAPE at early lags, but significantly underestimates predictive uncertainty, as seen in low coverage rates. In contrast, our functional method provides consistently higher coverage and lower interval scores, indicating better-calibrated probabilistic forecasts.

## 5.3 Calibration and Predictive Distribution

Figure 7 shows the empirical distribution of realized CLR quantiles relative to our forecast distributions (probability integral transforms). If forecasts are well calibrated, this empirical CDF should be close to uniform (Dawid, 1984; Wilks, 1990; Denuit et al., 2021). We observe that the functional method produces distributions of quantiles that are close to uniform, while CL exhibits systematic undercoverage, especially for intermediate lags. This diagnostic confirms that CL's variance assumptions underestimate true uncertainty, while our bootstrap-based functional forecasts are more reliable.

Figure 7 shows the combined lag 7-9 CLR predictive results for accident years 2005-2008 (with 6,5,4,3 lags complete respectively). The reverse S-shape in all three panels suggests the predictions tend to be at either lower or higher quantiles of the distribution. Further test of uniformity of the realized quantiles is done by the Kolmogorov Smirnov (KS) statistic. The KS score for the three ECDF is $0.0741, 0.0869$, and $0.0954$. Based on the critical value $1.35/\sqrt{137 \cdot 4} = 0.0577$, we reject the null hypothesis and the predictive distribution quantile is not uniform.

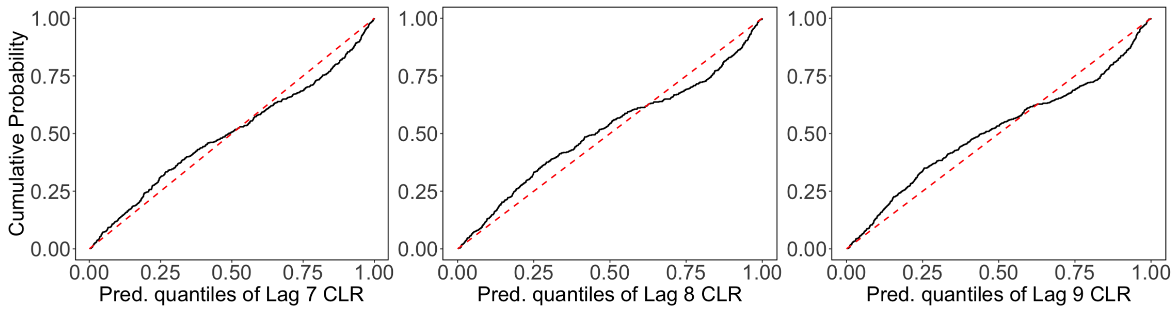



Figure 7: Empirical CDF comparing realized cumulative loss ratios to predicted quantiles across the 137 triangles for 3-6 lags complete.

## 5.4 Functional Coverage of CLR Curves

Beyond pointwise metrics, we evaluate the coverage of entire CLR trajectories. We compute functional prediction envelopes using extremal depth (EXD), which provide simultaneous bands over development lags (Narisetty and Nair, 2016). Table 6 reports functional coverage rates and cumulative interval scores (CIS), defined as

$$CIS_\alpha(s) = \frac{1}{I}\sum_{i=1}^{I}\frac{1}{10-s}\sum_{j=s}^{9} IS_\alpha\big(L_{i,t}^{(s)}(j), U_{i,t}^{(s)}(j); c_{i,t}(j)\big). \tag{15}$$

We find that CL consistently delivers coverage well below the nominal 95% level, especially for $s < 6$, whereas our method achieves coverage close to nominal, with tighter scores for $s \geq 4$.

| $s$ | Coverage (%) at $\alpha = 95\%$ PLS | EXD | CL | Interval Score PLS | EXD | $\overline{CIS}_{0.95}$ CL |
|---|---|---|---|---|---|---|
| 1 | 70.80% | 72.26% | 82.48% | 0.4813 | 0.6031 | 0.5125 |
| 2 | 96.35% | 95.62% | 72.26% | 0.3465 | 0.3241 | 0.4802 |
| 3 | 91.24% | 89.78% | 64.23% | 0.2005 | 0.2177 | 0.4728 |
| 4 | 94.89% | 94.16% | 66.42% | 0.1589 | 0.1331 | 0.2003 |
| 5 | 95.62% | 97.08% | 70.07% | 0.1188 | 0.1092 | 0.1121 |
| 6 | 96.35% | 95.62% | 75.91% | 0.0906 | 0.0769 | 0.0734 |
| 7 | 97.81% | 97.81% | 78.10% | 0.0870 | 0.0877 | 0.0573 |
| 8 | 95.62% | 97.08% | 82.48% | 0.0476 | 0.0474 | 0.0501 |
| 9 | 94.89% | 94.89% | 92.70% | 0.0299 | 0.0299 | 0.0354 |

Table 6: Statistical scores for forecasting cumulative loss curves from given lag $s$ to lag 9. We report coverage at 95% level, and average cumulative interval scores $\overline{CIS}_\alpha$ with $\alpha = 0.95$. PLS refers to point-wise interval, CL to chain ladder and EXD to extremal depth.

Taken together, these results highlight three conclusions. First, CL achieves good point accuracy in early lags, but its variance assumptions lead to systematically under-dispersed predictive distributions. Second, our functional method achieves competitive or superior accuracy at later lags, while providing better-calibrated uncertainty quantification. Third, functional coverage metrics show that entire CLR trajectories are captured with higher reliability by our approach. From a practical perspective, this implies that regulators and insurers can use our forecasts not only for reserve point estimates but also for realistic predictive envelopes, supporting solvency analysis and capital adequacy under uncertainty.

We next show a few illustrative predictions and functional envelopes of PLS versus CL. Figure 8a shows the case where the PLS method gives better prediction than chain ladder and sufficiently wide predictive intervals. Figure 8b show the case where PLS method gives worse prediction, but sufficiently covers the true curve, while CL does not cover the realized losses. Figure 8c show the case where PLS method gives better prediction and tighter but sufficient prediction interval. We also note that removing outlier curves prior to running the model helps with improving the MAPE and interval score, cf. Table 9 and 10 in Appendix A.2.

## 5.5 Extensions to Other Domains

While our empirical study focused on P&C loss development, the framework is general. For example, epidemiology modeling, or mortality analysis, also involves incomplete cohort structures.

### Epidemiology

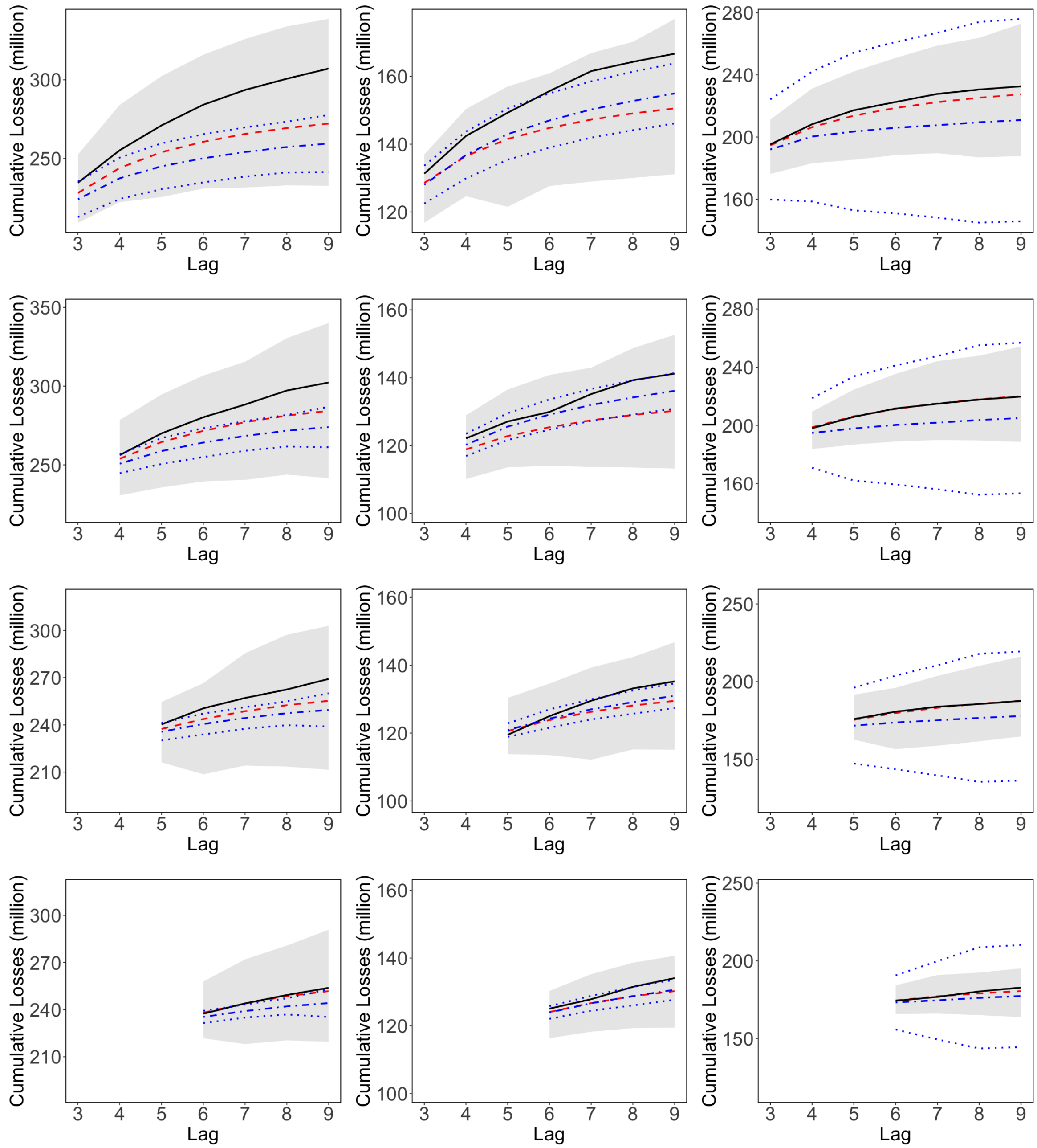
Beyond insurance applications, the proposed functional–probabilistic framework naturally extends to epidemiology, where delayed reporting structures produce data highly analogous to run-off triangles. In infectious disease surveillance, cases are often registered with a lag between infection, diagnosis, and reporting, generating incomplete "reporting triangles" (Lawless, 1994; Höhle and an der Heiden, 2014). This setting raises similar methodological challenges as loss development: incomplete curves, strong dependence across development lags, and the presence of reporting anomalies. A concrete illustration is provided by the analysis of acquired immune deficiency syndrome (AIDS) in England and Wales, where reported case counts also form incomplete development triangles with a fixed 12-quarter horizon, structurally identical to the loss triangles considered here (De Angelis and Gilks, 1994).

Let $t \in \mathcal{T}$ denote calendar time (e.g., diagnosis week) and $x \in \mathcal{X}$ the reporting delay (lag). Define $C_t(x)$ as the cumulative number of cases diagnosed in period $t$ that are reported within $x$ weeks. The incremental counts are

$$Y_t(x) := C_t(x) - C_t(x-1), \quad x \geq 1,$$

with $Y_t(0) = C_t(0)$ denoting the immediately reported cases. Analogous to incremental loss ratios, we may normalize by exposures or population size to obtain scale-free curves $y_t(x)$, suitable for comparison across regions or age groups.

The reporting triangle $\{Y_t(x) : t \leq T,\, 0 \leq x \leq T - t\}$ is then structurally identical to insurance loss triangles. Forecasting unreported cases corresponds to predicting the lower-right




(a) P53 (b) C260 (c) P1406

Figure 8: Comparing PLS prediction and EXD coverage against chain-ladder method. Red curves: PLS prediction; black curves: realized actual ILRs; blue lines: chain ladder predictions and 95% prediction intervals; grey intervals: 95% EXD coverage region.

corner of the reporting triangle. Traditional approaches include chain-ladder-type estimators (Höhle, 2017) or Bayesian hierarchical models (Stoner and Economou, 2020). Our functional data analysis (FDA) approach provides a nonparametric alternative: case-delay profiles are treated as functional curves, decomposed via FPCA, with partially observed trajectories completed through penalized least squares and bootstrap-based predictive envelopes.

Epidemiological reporting is further complicated by systemic anomalies, such as reporting delays during holidays or sudden diagnostic surges during epidemics. These correspond to outliers in our FDA framework. Depth-based detection methods (Sun et al., 2012; Narisetty and Nair, 2016) can robustify the analysis by down-weighting atypical years (e.g., pandemic waves) when estimating mean reporting patterns. This mirrors our treatment of anomalous insurers in Section 3.2.

Forecasting incomplete cohorts of infected individuals can be viewed through a cohort–development lens. For a birth-cohort analogue, consider infection cohorts $c = t - x$; right-truncated infection cohorts (those diagnosed recently) resemble open triangles in reserving. Illness–death models formulated in Lexis space provide another powerful representation of such cohort dynamics (Brinks, 2014), and age–period–cohort regression on the Lexis diagram is a natural extension (Carstensen, 2007). Estimation of delay distributions via robust functional regression allows prediction of unreported cases and quantification of associated uncertainty. Probabilistic envelopes provide epidemiologists with calibrated real-time case estimates, an important input for policy responses during outbreaks.

In summary, epidemiological reporting triangles share the same mathematical structure as insurance loss triangles. By adapting our functional–probabilistic methodology, we would obtain flexible, robust forecasts of unreported cases, enhancing both public health surveillance and real-time epidemic response (Bastos et al., 2019).

### Mortality

Another natural domain of application is cohort-driven mortality modeling. Age-specific death rates arranged on a Lexis diagram give rise to incomplete birth cohorts that mirror right-truncated loss triangles. The functional–probabilistic methods developed here can therefore be used to forecast partially observed age–cohort mortality profiles, in close analogy to how we complete insurance development curves. This parallel has been widely recognized in the actuarial literature (Gan and Lin, 2015; Hunt and Villegas, 2015; Tsai and Kim, 2022), reinforcing the unity of reserving problems across nonlife insurance and demography.

# 6 Conclusion

This paper introduced a functional perspective on the analysis and forecasting of loss development in property and casualty insurance. By reinterpreting run-off triangles as collections of incremental and cumulative loss ratio curves, we moved beyond the traditional cell-based reserving paradigm and established a framework that leverages modern functional data analysis (FDA). This reconceptualization provides three key advances: (i) robust detection of outliers and anomalies, (ii) systematic incorporation of insurer- and market-level covariates, and (iii) probabilistic forecasting of incomplete curves with calibrated uncertainty quantification.

Empirically, we documented heterogeneous loss development patterns across more than one hundred U.S. workers' compensation insurers over 24 accident years. Functional data depth allowed us to identify central development shapes and to flag atypical curves arising from data errors, unusual claims, or structural portfolio shifts. We showed that outliers are concentrated in longer development horizons and that company size is a major driver of heterogeneity, with smaller insurers exhibiting greater volatility and larger insurers displaying longer-tailed runoff. We also highlighted cyclical calendar-year effects aligned with well-known underwriting cycles, underlining the importance of accounting for market-wide shocks when benchmarking development patterns.

On the forecasting side, our methodology combines functional principal component analysis with regression-based priors and penalized curve completion. This approach allows partial curves to be extrapolated in a principled manner, balancing historical cross-sectional similarity with company-specific covariates. Forecast evaluation against the industry standard Chain-Ladder model (Mack, 1993; Wűthrich and Merz, 2008) demonstrated that while CL performs competitively at very short horizons, our functional method dominates once a modest number of development lags are observed. In particular, it delivers sharper yet well-calibrated predictive intervals, as confirmed by interval scores (Gneiting and Raftery, 2007) and functional coverage diagnostics. This improvement reflects the dual strength of FDA: its ability to exploit the global shape of development curves and its flexibility to integrate heterogeneous information sources.

Beyond actuarial reserving, the functional–probabilistic framework generalizes to other domains characterized by incomplete cohort data. We illustrated parallels with mortality forecasting where Lexis diagrams yield truncated age–cohort structures analogous to open triangles and epidemiological surveillance data organized into reporting triangles, with development corresponding to delayed case registration. rThese connections emphasize that the actuarial problem of projecting evolution of incurred-but-not-reported claims is part of a broader class of statistical forecasting problems on partially observed functional data.

Looking forward, several extensions are promising. First, the regression step can be enriched with additional financial or operational covariates, including external macroeconomic indicators, to capture drivers of long-tailed liabilities. Second, alternative resampling and conformal prediction methods (Graziadei et al., 2023; Hong, 2025) can further strengthen distributional calibration. Third, linking the FDA approach with machine learning architectures (Kuo, 2019; Schneider and Schwab, 2025) opens a path toward hybrid models that balance interpretability with predictive power. Finally, broader comparative studies across multiple lines of business and international datasets would help validate the robustness and generality of our framework.

In conclusion, reframing loss development as a functional forecasting problem yields both theoretical and practical gains. Our empirical study demonstrates that functional methods not only improve predictive performance relative to the Chain-Ladder benchmark but also provide a unified language for detecting anomalies, quantifying uncertainty, and extending actuarial techniques to adjacent fields. We hope that this work stimulates further cross-pollination between actuarial science, epidemiology, and demography, and contributes to the development of forecasting methodologies that are both statistically rigorous and operationally relevant.

# References


Abdallah, Anas and Lan Wang (2023), "Rank-based multivariate Sarmanov for modeling dependence between loss reserves." *Risks*, 11, 187.

Ang, Zi Qing and See Keong Lee (2022), "Hierarchical Bayesian Gaussian process regression model for loss reserving using combinations of squared exponential kernels." *Insurance: Mathematics and Economics*, 105, 54–63.

Bastos, Leonardo S, Theodoros Economou, Marcelo FC Gomes, Daniel AM Villela, Flávio C Coelho, Oswaldo G Cruz, Owen Stoner, Tom Bailey, and Claudia T Codeço (2019), "A modelling approach for correcting reporting delays in disease surveillance data." *Statistics in Medicine*, 38, 4363–4377.

Brinks, Ralph (2014), "Illness–death model: an efficient tool for non-parametric analysis of chronic disease epidemiology." *BMC Public Health*, 14, 644.

Bühlmann, Hans (1967), "Experience rating and credibility." *ASTIN Bulletin*, 4, 199–207.

Cai, Pengfei, Anas Abdallah, and Pratheepa Jeganathan (2024), "Recurrent neural networks for multivariate loss reserving and risk capital analysis." *arXiv preprint arXiv:2402.10421*.

Carstensen, Bendix (2007), "Age–period–cohort models for the Lexis diagram." *Statistics in Medicine*, 26, 3018–3045.

Croux, Christophe, Peter Filzmoser, and Maria Rosario Oliveira (2007), "Algorithms for projection–pursuit robust principal component analysis." *Chemometrics and Intelligent Laboratory Systems*, 87, 218–225.

Cummins, J. David, Neil A. Doherty, and James Lo (1992), "The economics of insurance regulation." *Journal of Risk and Insurance*, 59, 623–654.

Cummins, J. David and Mary A. Weiss (2000), "The global market for reinsurance: Consolidation, capacity, and efficiency." *Brookings-Wharton Papers on Financial Services*, 2000, 159–222.

Dai, Wenlin, Tomáš Mrkvička, Ying Sun, and Marc G Genton (2020), "Functional outlier detection and taxonomy by sequential transformations." *Computational Statistics & Data Analysis*, 149, 106960.

Dawid, A. Philip (1984), "Present position and potential developments: Some personal views: Statistical theory: The prequential approach." *Journal of the Royal Statistical Society: Series A*, 147, 278–292.

De Angelis, D. and W. R. Gilks (1994), "Estimating acquired immune deficiency syndrome incidence accounting for reporting delay." *Journal of the Royal Statistical Society Series A: Statistics in Society*, 157, 31–40.

Denuit, Michel, Arthur Charpentier, and Julien Trufin (2021), “Autocalibration and Tweedie-dominance for insurance pricing with machine learning.” *Insurance: Mathematics and Economics*, 101, 485–497.

Diebold, Francis X and Roberto S Mariano (1995), “Comparing predictive accuracy.” *Journal of Business & Economic Statistics*, 13, 253–263.

Elías, Antonio, Raúl Jiménez, and Han Lin Shang (2022), “On projection methods for functional time series forecasting.” *Journal of Multivariate Analysis*, 189, 104890.

Gabrielli, Andrea (2021), “An individual claims reserving model for reported claims.” *European Actuarial Journal*, 11, 541–577.

Gan, Guojun and X. Sheldon Lin (2015), “Valuation of large variable-annuity portfolios under nested simulation: a functional data approach.” *Insurance: Mathematics and Economics*, 62, 138–150.

Gneiting, Tilmann and Adrian E. Raftery (2007), “Strictly proper scoring rules, prediction, and estimation.” *Journal of the American Statistical Association*, 102, 359–378.

Graziadei, Helton, Eduardo FL de Melo, Rodrigo S Targino, et al. (2023), “Conformal prediction for frequency-severity modeling.” *arXiv preprint arXiv:2307.13124*.

Griffith, Devan (2025), “Multioutput Gaussian processes for loss ratio development.” *Variance*, 18.

Harrington, Scott E. and Greg Niehaus (2000), “Volatility and underwriting cycles.” *Handbook of Insurance*, 657–686.

Höhle, Michael (2017), “An R package for the analysis of infectious disease surveillance data.” *Journal of Statistical Software*, 77, 1–54.

Höhle, Michael and Matthias an der Heiden (2014), “Bayesian nowcasting during the STEC o104:H4 outbreak in Germany, 2011.” *Biometrics*, 70, 993–1002.

Hong, Liang (2025), “Conformal prediction of future insurance claims in the regression problem.” *arXiv preprint arXiv:2503.03659*.

Hunt, Andrew and Andres M. Villegas (2015), “Robustness and convergence in the Lee–Carter model with cohort effects.” *Insurance: Mathematics and Economics*, 64, 186–202.

Kuo, Kevin (2019), “Deeptriangle: A deep learning approach to loss reserving.” *Risks*, 7, 97.

Lally, Nathan and Brian Hartman (2018), “Estimating loss reserves using hierarchical Bayesian Gaussian process regression with input warping.” *Insurance: Mathematics and Economics*, 82, 124–140.

Lawless, Jerald F (1994), “Adjustments for reporting delays and the prediction of occurred but not reported events.” *Canadian Journal of Statistics*, 22, 15–31.

Ludkovski, Mike and Howard Zail (2022), "Gaussian process models for incremental loss ratios." *Variance*, 15.

Mack, Thomas (1993), "Distribution-free calculation of the standard error of chain ladder reserve estimates." *ASTIN Bulletin*, 23, 213–225.

Narisetty, Naveen N. and Vijayan N. Nair (2016), "Extremal depth for functional data and applications." *Journal of the American Statistical Association*, 111, 1705–1714.

Nieto-Barajas, Luis E and Rodrigo S Targino (2021), "A Gamma moving average process for modelling dependence across development years in run-off triangles." *ASTIN Bulletin: The Journal of the IAA*, 51, 245–266.

Patton, Andrew J (2011), "Volatility forecast comparison using imperfect volatility proxies." *Journal of Econometrics*, 160, 246–256.

Radtke, Michael, Klaus D. Schmidt, and Anja Schnaus (2016), *Handbook on Loss Reserving*. Springer.

Ramsay, J. O. and B. W. Silverman (2005), *Functional Data Analysis*, 2nd edition. Springer.

Rousseeuw, Peter J and Christophe Croux (1993), "Alternatives to the median absolute deviation." *Journal of the American Statistical Association*, 88, 1273–1283.

Schneider, Judith C and Brandon Schwab (2025), "Advancing loss reserving: A hybrid neural network approach for individual claim development prediction." *Journal of Risk and Insurance*, 92, 389–423.

Shang, Han Lin (2013), "Functional time series approach for forecasting very short-term electricity demand." *Journal of Applied Statistics*, 40.

Shang, Han Lin (2017), "Forecasting intraday S&P 500 index returns: A functional time series approach." *Journal of forecasting*, 36, 741–755.

Shang, Han Lin and Rob J Hyndman (2011), "Nonparametric time series forecasting with dynamic updating." *Mathematics and Computers in Simulation*, 81, 1310–1324.

Shi, Peng (2017), "A multivariate analysis of intercompany loss triangles." *Journal of Risk and Insurance*, 84, 717–737.

Shi, Peng and Edward W Frees (2011), "Dependent loss reserving using copulas." *ASTIN Bulletin: The Journal of the IAA*, 41, 449–486.

Shi, Peng and Brian M Hartman (2016), "Credibility in loss reserving." *North American Actuarial Journal*, 20, 114–132.

Shi, Peng and Kun Shi (2023), "Non-life insurance risk classification using categorical embedding." *North American Actuarial Journal*, 27, 579–601.

Steinmetz, Julia and Carsten Jentsch (2024), "Bootstrap consistency for the Mack bootstrap." *Insurance: Mathematics and Economics*, 115, 83–121.

Stoner, Owen and Theodoros Economou (2020), "A hierarchical framework for correcting reporting delays in disease surveillance data." *Biostatistics*, 21, 555–572.

Sun, Ying, Marc G Genton, and Douglas W Nychka (2012), "Exact fast computation of band depth for large functional datasets: How quickly can one million curves be ranked?" *Stat*, 1, 68–74.

Taylor, Greg (2019), "Loss reserving models: Granular and machine learning forms." *Risks*, 7, 82.

Tibshirani, Robert (1996), "Regression shrinkage and selection via the lasso." *Journal of the Royal Statistical Society: Series B*, 58, 267–288.

Tsai, Chung-Cheng Lee and Sojung Kim (2022), "Model mortality rates using property-and-casualty insurance reserving methods." *Insurance: Mathematics and Economics*, 106, 326–340.

Weiss, Mary A. (2001), "Factors affecting workers' compensation medical costs." *Journal of Risk and Insurance*, 68, 365–384.

Wilks, Daniel S (1990), "On the combination of forecast probabilities for consecutive precipitation periods." *Weather and forecasting*, 5, 640–650.

Wüthrich, Mario V (2018), "Machine learning in individual claims reserving." *Scandinavian Actuarial Journal*, 2018, 465–480.

Wűthrich, Mario V. and Michael Merz (2008), *Stochastic Claims Reserving Methods in Insurance*, volume 435. John Wiley and Sons.

Zhang, Yanwei and Vanja Dukic (2013), "Predicting multivariate insurance loss payments under the Bayesian copula framework." *Journal of Risk and Insurance*, 80, 891–919.

# A Appendix

## A.1 Mathematical Study of Functional Outliers

For comparison, we carry out outlier detection using the alternative concept of *functional data depth*. This approach defines a metric that directly compares a given curve $y_{i,t}(x)$ to a collection of curves $\mathcal{F} = \{I^m(x) : m = 1, \dots, M\}$ in the sense of "depth". The deepest curve is interpreted as the functional median, and the shallowest curves are outliers, being "outside" or on the edge of the collection. In particular, we consider outliers and center curves defined in terms of Band Depth (BD), Modified Band Depth (MBD) and Extremal Depth (EXD, Narisetty and Nair (2016)).

Functional depth is a statistical framework for ranking functional observations from most outlying to most typical, providing a center-outward ordering on multivariate data. While there is a canonical way to sort and rank real-valued scalars, there is a wide range of possible functional depths. The marginal rank $R(x)$ of $y_{i,t}$ at lag $x$ with respect to the collection $\mathcal{F}$ is

$$R(x) := \sum_m 1_{\{I^m(x) < y_{i,t}(x)\}} \in \{1, \dots, M\}. \tag{16}$$

The (rank-2) Band Depth (BD) quantifies the frequency that $y_{i,t}$ falls into the interval defined by another pair $y_{j,s}, y_{j',t'}$ in $\mathcal{F}$:

$$BD(y_{i,t}) := \frac{(\min_x R(x) - 1) \cdot (M - \max_x R(x)) + M - 1}{M(M+1)}. \tag{17}$$

Instead of looking only at the smallest and highest rank in Equation (17), Modified Band Width averages the product $(R(x) - 1)(M - R(x))$ of the ranks of $y_{i,t}$ across all the 10 lags:

$$MBD(y_{i,t}) := \frac{1}{M(M+1)} \sum_{x=1}^{10} \frac{(R(x) - 1)\,(M - R(x))}{M - 10 - 1}. \tag{18}$$

The MBD is symmetric around the median and underlies the commonly used functional boxplot for identifying magnitude outliers.

Extremal Depth (EXD), originally due to Narisetty and Nair (2016), focuses on the extremal behavior of a curve, recording how *often* it attains very high/low values relative to $\mathcal{F}$. To this end, EXD employs the point-wise depth

$$d(x) := 1 - \frac{|2R(x) - M - 1|}{M}. \tag{19}$$

Rather than integrating $d.$, EXD constructs the depth Cumulative Density Function (d-CDF)

$$\Phi(r; y_{i,t}) := \frac{1}{10} \sum_{x=1}^{10} 1_{\{d(x) \le r\}}, \qquad r \in (0,1) \tag{20}$$

and then ranks an ILR curve $y_{i,t}$ according to the left-tail stochastic ordering of $\Phi$'s:

$$EXD(y_{i,t}) := \frac{1}{M} \sum_{m=1}^{M} 1_{\{\Phi \succcurlyeq \Phi_m\}}, \tag{21}$$

where the stochastic order $\Phi \succcurlyeq \Phi_m$ intuitively means that the frequency (across lags $x$) of $y_{i,t}(\cdot)$ achieving an extreme rank is higher than that for $I^m(\cdot)$.

For all of the above methods, the median curve in $\mathcal{F}$ is defined as the one with the highest depth, and the respective ranking also provides a natural notion of a functional quantile interval.

All our analysis is performed within the `R` statistical environment. The RPCA analysis is done using `R` package `pcaPP` and functional depth using `fdaoutlier`. Visualization is done using `ggplot2` package. A GitHub repository containing the codes to replicate the result is available upon request.

## A.2 Additional Tables

| $s$ | $k$ | PC0 | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 | PC8 | PC9 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | −0.5341 | −0.7187 | −0.3734 | −0.2053 | −0.0951 | −0.0677 | −0.0402 | −0.0259 | −0.0180 | −0.0189 |
| 2 | 1 | −0.5215 | −0.7230 | −0.3798 | −0.2093 | −0.0972 | −0.0691 | −0.0410 | −0.0264 | −0.0184 | −0.0192 |
| 3 | 1 | −0.5124 | −0.7225 | −0.3861 | −0.2177 | −0.1021 | −0.0724 | −0.0429 | −0.0278 | −0.0194 | −0.0199 |
| 3 | 2 | −0.7897 | 0.1979 | 0.3955 | 0.3352 | 0.2029 | 0.1328 | 0.0690 | 0.0509 | 0.0405 | 0.0248 |
| 3 | 3 | 0.3366 | −0.6424 | 0.3761 | 0.4290 | 0.2881 | 0.1895 | 0.1329 | 0.0856 | 0.0605 | 0.0305 |
| 4 | 1 | −0.5034 | −0.7258 | −0.3911 | −0.2178 | −0.1031 | −0.0732 | −0.0431 | −0.0279 | −0.0195 | −0.0201 |
| 4 | 2 | −0.7968 | 0.1941 | 0.3911 | 0.3282 | 0.2005 | 0.1312 | 0.0678 | 0.0501 | 0.0400 | 0.0244 |
| 4 | 3 | 0.3332 | −0.6402 | 0.3847 | 0.4259 | 0.2892 | 0.1905 | 0.1333 | 0.0857 | 0.0606 | 0.0307 |
| 4 | 4 | 0.0177 | 0.1577 | −0.7325 | 0.4317 | 0.3929 | 0.2041 | 0.1686 | 0.1216 | 0.1015 | 0.0478 |
| 5 | 1 | −0.4813 | −0.7332 | −0.3988 | −0.2258 | −0.1065 | −0.0763 | −0.0446 | −0.0291 | −0.0205 | −0.0207 |
| 5 | 2 | −0.8026 | 0.1612 | 0.3940 | 0.3220 | 0.2077 | 0.1339 | 0.0699 | 0.0514 | 0.0409 | 0.0251 |
| 5 | 3 | 0.3511 | −0.6454 | 0.4017 | 0.4021 | 0.2761 | 0.1834 | 0.1281 | 0.0818 | 0.0573 | 0.0294 |
| 5 | 4 | 0.0238 | 0.1396 | −0.7047 | 0.3696 | 0.4733 | 0.2297 | 0.1896 | 0.1346 | 0.1116 | 0.0579 |
| 5 | 5 | −0.0027 | −0.0098 | −0.1622 | 0.7164 | −0.6447 | −0.1293 | −0.1208 | −0.0702 | −0.0553 | −0.0743 |
| 6 | 1 | −0.4582 | −0.7381 | −0.4096 | −0.2333 | −0.1120 | −0.0801 | −0.0471 | −0.0309 | −0.0218 | −0.0217 |
| 6 | 2 | −0.8254 | 0.1595 | 0.3683 | 0.3051 | 0.1982 | 0.1282 | 0.0663 | 0.0491 | 0.0394 | 0.0241 |
| 6 | 3 | 0.3283 | −0.6404 | 0.4139 | 0.4096 | 0.2828 | 0.1859 | 0.1308 | 0.0837 | 0.0587 | 0.0303 |
| 6 | 4 | 0.0276 | 0.1384 | −0.7087 | 0.3823 | 0.4591 | 0.2278 | 0.1885 | 0.1341 | 0.1112 | 0.0570 |
| 6 | 5 | 0.0024 | 0.0129 | 0.1472 | −0.7101 | 0.6531 | 0.1320 | 0.1250 | 0.0730 | 0.0576 | 0.0757 |
| 6 | 6 | 0.0076 | −0.0098 | −0.0055 | 0.1803 | 0.4644 | −0.7457 | −0.3049 | −0.2174 | −0.1752 | −0.1569 |
| 7 | 1 | −0.4407 | −0.7430 | −0.4144 | −0.2386 | −0.1135 | −0.0863 | −0.0500 | −0.0331 | −0.0236 | −0.0230 |
| 7 | 2 | −0.8390 | 0.1537 | 0.3538 | 0.2930 | 0.1936 | 0.1245 | 0.0647 | 0.0478 | 0.0385 | 0.0235 |
| 7 | 3 | 0.3174 | −0.6371 | 0.4243 | 0.4074 | 0.2842 | 0.1918 | 0.1338 | 0.0857 | 0.0603 | 0.0318 |
| 7 | 4 | 0.0288 | 0.1335 | −0.7088 | 0.3897 | 0.4518 | 0.2242 | 0.1971 | 0.1347 | 0.1117 | 0.0570 |
| 7 | 5 | 0.0029 | 0.0179 | 0.1377 | −0.7129 | 0.6456 | 0.1455 | 0.1354 | 0.0782 | 0.0620 | 0.0793 |
| 7 | 6 | 0.0118 | −0.0094 | 0.0001 | 0.1712 | 0.4823 | −0.7339 | −0.3128 | −0.2161 | −0.1745 | −0.1557 |
| 8 | 1 | −0.4172 | −0.7439 | −0.4256 | −0.2503 | −0.1207 | −0.0934 | −0.0544 | −0.0351 | −0.0257 | −0.0244 |
| 8 | 2 | −0.8526 | 0.1331 | 0.3383 | 0.2898 | 0.1853 | 0.1225 | 0.0608 | 0.0456 | 0.0371 | 0.0223 |
| 8 | 3 | 0.3129 | −0.6411 | 0.4210 | 0.4134 | 0.2858 | 0.1857 | 0.1265 | 0.0810 | 0.0575 | 0.0294 |
| 8 | 4 | 0.0284 | 0.1306 | −0.7194 | 0.4397 | 0.3912 | 0.2247 | 0.1904 | 0.1290 | 0.1095 | 0.0533 |
| 8 | 5 | 0.0011 | 0.0250 | 0.0873 | −0.6822 | 0.6621 | 0.1954 | 0.1612 | 0.0970 | 0.0785 | 0.0905 |
| 8 | 6 | 0.0085 | 0.0024 | −0.0095 | 0.1469 | 0.5074 | −0.7499 | −0.2709 | −0.1935 | −0.1606 | −0.1479 |
| 8 | 7 | 0.0151 | −0.0014 | −0.0398 | 0.0225 | 0.1420 | 0.5134 | −0.7627 | −0.2758 | −0.2156 | −0.0981 |
| 8 | 8 | 0.0033 | −0.0034 | 0.0130 | −0.0074 | −0.0195 | −0.1314 | −0.4810 | 0.8093 | 0.2953 | 0.0931 |
| 9 | 1 | −0.3790 | −0.7542 | −0.4389 | −0.2515 | −0.1282 | −0.0959 | −0.0579 | −0.0366 | −0.0267 | −0.0257 |
| 9 | 2 | −0.8715 | 0.1102 | 0.3064 | 0.2873 | 0.1737 | 0.1210 | 0.0599 | 0.0416 | 0.0382 | 0.0210 |
| 9 | 3 | 0.3061 | −0.6380 | 0.4560 | 0.4049 | 0.2772 | 0.1688 | 0.1169 | 0.0737 | 0.0490 | 0.0257 |
| 9 | 4 | 0.0535 | 0.1067 | −0.7059 | 0.4696 | 0.3902 | 0.2208 | 0.1867 | 0.1310 | 0.1061 | 0.0514 |
| 9 | 5 | −0.0082 | 0.0217 | 0.0728 | −0.6584 | 0.7049 | 0.1469 | 0.1536 | 0.0779 | 0.0766 | 0.0816 |
| 9 | 6 | 0.0027 | 0.0085 | −0.0167 | 0.1855 | 0.4653 | −0.7146 | −0.3575 | −0.2096 | −0.1991 | −0.1634 |
| 9 | 7 | 0.0160 | −0.0007 | −0.0358 | 0.0015 | 0.1024 | 0.5825 | −0.7338 | −0.2648 | −0.1827 | −0.0824 |
| 9 | 8 | 0.0016 | −0.0039 | 0.0176 | −0.0144 | −0.0194 | −0.1233 | −0.4816 | 0.7963 | 0.3293 | 0.0966 |

Table 7: PCA Loadings by Component (PC0–PC9)

| s | k | Business Focus | | NAIC Ownership | | Geographic Focus | | | | Time-variant Features | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | personal | wkcomp | stock | other | midwest | northeast | south | west | time | log prem | log prem × time |
| 1 | 1 | 0.0166 | −0.0026 | 0.0021 | −0.0304 | −0.0228 | −0.0062 | −0.0044 | 0.0112 | −0.0106 | −0.0085 | 0.0007 |
| 2 | 1 | 0.0159 | −0.0037 | 0.0014 | −0.0338 | −0.0206 | −0.0057 | −0.0005 | 0.0148 | −0.0110 | −0.0089 | 0.0008 |
| 3 | 1 | 0.0154 | −0.0042 | 0.0005 | −0.0361 | −0.0185 | −0.0047 | 0.0012 | 0.0171 | −0.0126 | −0.0098 | 0.0009 |
| 3 | 2 | −0.0028 | 0.0019 | 0.0050 | −0.0044 | −0.0197 | −0.0002 | −0.0081 | | | 0.0025 | |
| 3 | 3 | 0.0017 | −0.0041 | 0.0017 | −0.0028 | −0.0119 | 0.0134 | −0.0111 | −0.0011 | −0.0003 | 0.0009 | |
| 4 | 1 | 0.0150 | −0.0047 | −0.0005 | −0.0377 | −0.0144 | −0.0049 | 0.0025 | 0.0194 | −0.0134 | −0.0100 | 0.0010 |
| 4 | 2 | −0.0029 | 0.0033 | 0.0049 | −0.0066 | −0.0199 | −0.0006 | −0.0095 | 0.0001 | | 0.0026 | |
| 4 | 3 | 0.0013 | −0.0040 | 0.0023 | −0.0023 | −0.0121 | 0.0143 | −0.0107 | −0.0020 | −0.0004 | 0.0010 | |
| 4 | 4 | | −0.0014 | −0.0009 | 0.0003 | −0.0026 | 0.0020 | −0.0039 | | | 0.0001 | |
| 5 | 1 | 0.0138 | −0.0060 | −0.0028 | −0.0390 | −0.0105 | −0.0018 | 0.0056 | 0.0216 | −0.0130 | −0.0091 | 0.0009 |
| 5 | 2 | −0.0037 | 0.0032 | 0.0055 | −0.0088 | −0.0198 | 0.0012 | −0.0095 | 0.0027 | | 0.0028 | |
| 5 | 3 | 0.0016 | −0.0032 | 0.0027 | −0.0020 | −0.0123 | 0.0137 | −0.0105 | −0.0030 | −0.0004 | 0.0009 | |
| 5 | 4 | −0.0008 | −0.0024 | −0.0017 | 0.0013 | −0.0045 | 0.0030 | −0.0049 | −0.0005 | −0.0014 | −0.0004 | 0.0001 |
| 5 | 5 | | 0.0004 | | | | | | | | −0.0004 | |
| 6 | 1 | 0.0134 | −0.0084 | −0.0054 | −0.0416 | −0.0074 | 0.0023 | 0.0080 | 0.0245 | −0.0114 | −0.0077 | 0.0008 |
| 6 | 2 | −0.0042 | 0.0030 | 0.0051 | −0.0128 | −0.0190 | 0.0007 | −0.0092 | 0.0061 | | 0.0027 | 0.0000 |
| 6 | 3 | 0.0010 | −0.0029 | 0.0031 | −0.0034 | −0.0138 | 0.0139 | −0.0114 | −0.0030 | −0.0004 | 0.0007 | 0.0000 |
| 6 | 4 | −0.0012 | −0.0028 | −0.0020 | 0.0013 | −0.0052 | 0.0035 | −0.0045 | −0.0008 | −0.0020 | −0.0009 | 0.0001 |
| 6 | 5 | | −0.0005 | | | | | | | | 0.0005 | |
| 6 | 6 | 0.0006 | 0.0006 | 0.0021 | 0.0011 | 0.0026 | 0.0025 | 0.0029 | −0.0005 | | −0.0009 | |
| 7 | 1 | 0.0132 | −0.0077 | −0.0078 | −0.0428 | −0.0028 | 0.0056 | 0.0086 | 0.0226 | | −0.0008 | 0.0001 |
| 7 | 2 | −0.0043 | 0.0032 | 0.0040 | −0.0169 | −0.0184 | 0.0006 | −0.0099 | 0.0099 | 0.0036 | 0.0048 | −0.0002 |
| 7 | 3 | 0.0011 | −0.0017 | 0.0036 | −0.0040 | −0.0148 | 0.0131 | −0.0124 | −0.0033 | −0.0005 | 0.0009 | |
| 7 | 4 | | −0.0021 | −0.0006 | | −0.0044 | 0.0035 | −0.0027 | | | 0.0001 | |
| 7 | 5 | | −0.0004 | | | | | | | | 0.0005 | |
| 7 | 6 | | | | | | | 0.0004 | | | −0.0008 | |
| 8 | 1 | 0.0142 | −0.0047 | −0.0094 | −0.0411 | | 0.0066 | 0.0089 | 0.0189 | | −0.0009 | |
| 8 | 2 | −0.0037 | 0.0035 | 0.0028 | −0.0203 | −0.0193 | 0.0014 | −0.0100 | 0.0115 | 0.0053 | 0.0055 | −0.0004 |
| 8 | 3 | | −0.0012 | 0.0031 | −0.0038 | −0.0140 | 0.0123 | −0.0113 | −0.0011 | −0.0005 | 0.0005 | |
| 8 | 4 | | −0.0014 | | | −0.0038 | 0.0033 | −0.0020 | | | | |
| 8 | 5 | | −0.0012 | | | | | | | | 0.0006 | |
| 8 | 6 | | | | | 0.0003 | | 0.0012 | | | −0.0006 | |
| 8 | 7 | | | | | | 0.0014 | | | | −0.0008 | |
| 8 | 8 | | | 0.0007 | | −0.0003 | −0.0021 | 0.0002 | | | 0.0002 | |
| 9 | 1 | 0.0147 | | −0.0105 | −0.0335 | | 0.0077 | 0.0053 | 0.0129 | 0.0006 | −0.0015 | |
| 9 | 2 | −0.0023 | 0.0050 | 0.0013 | −0.0225 | −0.0191 | 0.0028 | −0.0108 | 0.0108 | 0.0088 | 0.0070 | −0.0006 |
| 9 | 3 | | −0.0008 | 0.0035 | −0.0053 | −0.0146 | 0.0121 | −0.0118 | −0.0011 | −0.0005 | 0.0005 | |
| 9 | 4 | | −0.0008 | | | −0.0044 | 0.0039 | −0.0018 | | | | |
| 9 | 5 | | | | | | | | | | 0.0004 | |
| 9 | 6 | | 0.0002 | 0.0010 | 0.0003 | 0.0024 | 0.0019 | 0.0028 | | | −0.0008 | |
| 9 | 7 | | | | | | 0.0011 | | | | −0.0005 | |
| 9 | 8 | | | 0.0009 | | | −0.0019 | | | | 0.0002 | |

Table 8: LASSO Coefficients by Variable. Each row represents the coefficient estimates for a given development lag $s$ and factor index $k$.

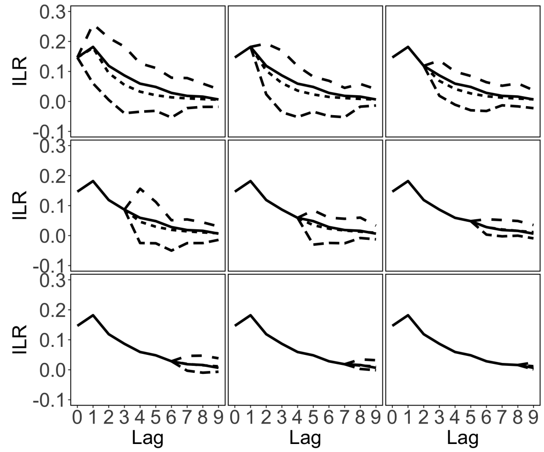
| Lag | MAPE | | Coverage % | | | Interval score | | |
|---|---|---|---|---|---|---|---|---|
| $s$ | PLS | CL | PLS | EXD | CL | PLS | EXD | CL |
| 1 | 21.84% | 16.53% | 95.62% | 90.51% | 86.13% | 0.6684 | 0.9469 | 0.7098 |
| 2 | 13.18% | 10.31% | 97.81% | 97.08% | 79.56% | 0.5130 | 0.4494 | 1.0422 |
| 3 | 9.69% | 7.21% | 99.27% | 97.08% | 71.53% | 0.3675 | 0.3097 | 0.9904 |
| 4 | 6.12% | 5.18% | 100.00% | 99.27% | 70.80% | 0.2590 | 0.2142 | 0.4047 |
| 5 | 4.08% | 4.10% | 100.00% | 97.08% | 73.72% | 0.2037 | 0.1783 | 0.1808 |
| 6 | 2.52% | 2.69% | 100.00% | 98.54% | 79.56% | 0.1529 | 0.1253 | 0.1176 |
| 7 | 2.09% | 2.07% | 100.00% | 100.00% | 84.67% | 0.1090 | 0.1096 | 0.0900 |
| 8 | 1.33% | 1.41% | 100.00% | 99.27% | 83.94% | 0.0766 | 0.0709 | 0.0669 |
| 9 | 0.70% | 0.67% | 95.62% | 95.62% | 92.70% | 0.0334 | 0.0334 | 0.0354 |

Table 9: Statistical scores for forecasting ultimate cumulative losses without outliers removed. We report MAPE, coverage at 95% level, and interval scores $IS_\alpha$ with $\alpha = 0.95$. PLS refers to point-wise interval, CL to chain ladder and EXD to extremal depth.

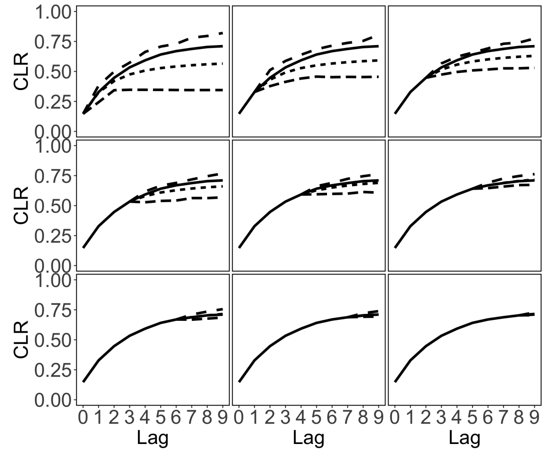
| $s$ | Coverage % | | | Interval Score | | |
|---|---|---|---|---|---|---|
| | PLS | EXD | CL | PLS | EXD | CL |
| 1 | 69.34% | 75.18% | 82.48% | 0.4728 | 0.5759 | 0.5125 |
| 2 | 94.16% | 95.62% | 72.26% | 0.3651 | 0.3429 | 0.4802 |
| 3 | 94.16% | 94.16% | 64.23% | 0.2538 | 0.2346 | 0.4728 |
| 4 | 97.08% | 97.81% | 66.42% | 0.1724 | 0.1575 | 0.2003 |
| 5 | 96.35% | 97.08% | 70.07% | 0.1356 | 0.1349 | 0.1121 |
| 6 | 96.35% | 97.08% | 75.91% | 0.1082 | 0.1011 | 0.0734 |
| 7 | 100.00% | 100.00% | 78.10% | 0.0785 | 0.0854 | 0.0573 |
| 8 | 98.54% | 99.27% | 82.48% | 0.0604 | 0.0617 | 0.0501 |
| 9 | 95.62% | 95.62% | 92.70% | 0.0334 | 0.0334 | 0.0354 |

Table 10: Statistical scores for forecasting cumulative loss curves from given lag $s$ to lag 9 without outliers removed. We report coverage at 95% level, and interval scores $IS_\alpha$ with $\alpha = 0.95$. PLS refers to point-wise interval, CL to chain ladder and EXD to extremal depth.

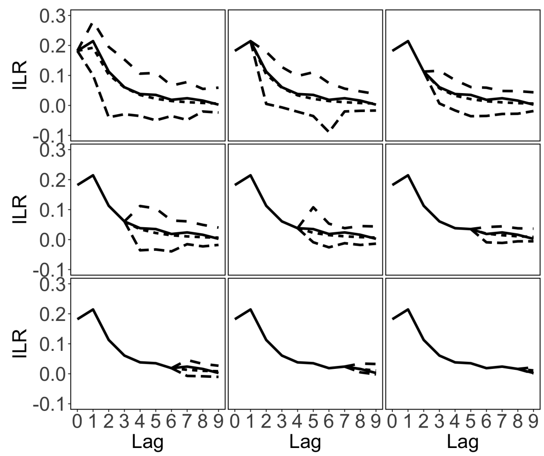
## A.3 Additional Plots



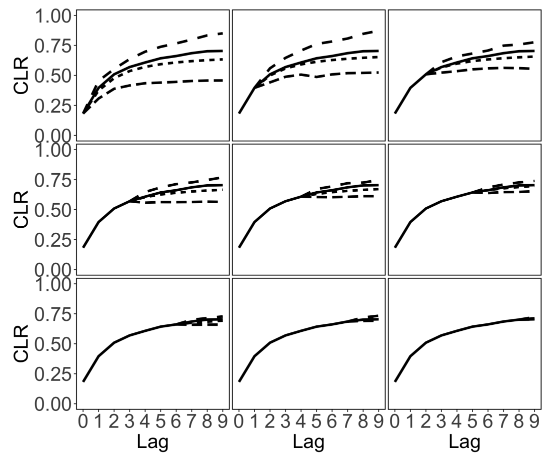

(a) Company P53 ILR



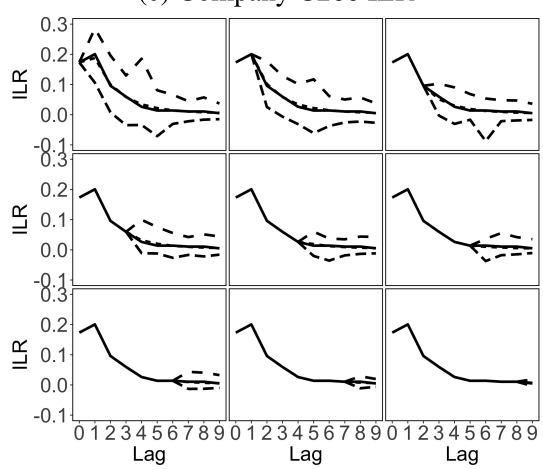

(b) Company P53 CLR



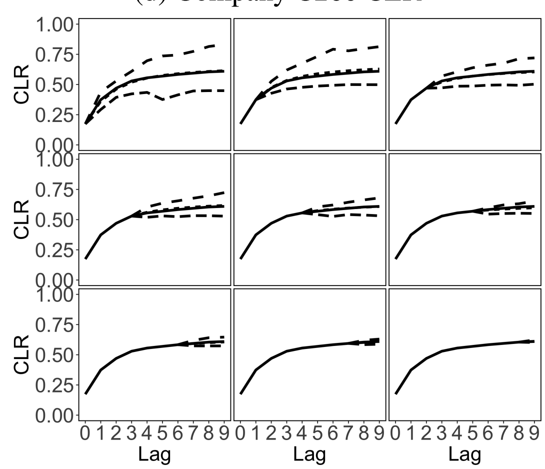

(c) Company C260 ILR




(d) Company C260 CLR




(e) Company P1406 ILR




(f) Company P1406 CLR

Figure 9: Forecast of ILR and CLR for AY 2010: solid line—actual ILR/CLR; short-dashed line—forecast ILR/CLR; long-dashed line—95% bootstrap prediction interval