Journal of Advanced Sciences and Mathematics Education
Volume 6, Issue 4, 423-445.
e_ISSN: 2798-1606
DOI: 10.58524/jasme.v6i4.1417

IBNR estimation method on reinsurance portfolio: evaluation of chain ladder, bornhuetter-ferguson, and cape cod

Ramdhan Fazrianto Suwarman*
Universitas Negeri Malang, INDONESIA

Aurellia Nurrahmalika
Universitas Negeri Malang, INDONESIA

Andi Daniah Pahrany
Universitas Negeri Malang, INDONESIA

Lita Wulandari Aeli
Universitas Negeri Malang, INDONESIA

Article Info

Article history:

Received: July 30, 2026
Revised: September 30, 2026
Accepted: November 11, 2026

Keywords:

Actuarial Reserving;
Comparative Analysis;
Incurred But Not Reported (IBNR);
Reinsurance Portfolios;
Relative Error.

Abstract

Background: Accurate estimation of Incurred But Not Reported (IBNR) reserves remains challenging in reinsurance because of heterogeneous claim development patterns and portfolio volatility. Inaccurate reserve estimates may compromise financial stability and undermine effective risk management. Differences in actuarial assumptions further complicate the selection of appropriate estimation methods.
Aims: This study compares the performance of Chain Ladder (CL), Bornhuetter–Ferguson (BF), and Cape Cod (CC) methods in estimating IBNR reserves across different reinsurance portfolios.
Method: A quantitative comparative analysis was conducted using Munich Re's cumulative paid-loss triangles and earned premium data for 2010–2021. Nine portfolios were evaluated using deterministic reserving methods with simple-average development factors, without tail factors or additional data adjustments. Estimation accuracy was assessed using Relative Error (RE) against reported IBNR reserves.
Results: Cape Cod achieved the lowest RE in four portfolios, Chain Ladder in four, and Bornhuetter–Ferguson in one. Across all portfolios, Cape Cod recorded the lowest mean RE (58.33%), followed by Bornhuetter–Ferguson (74.78%) and Chain Ladder (85.89%). Excluding three long-tail portfolios reversed this ranking, favoring Bornhuetter–Ferguson (25.67%) and Chain Ladder (26.67%) over Cape Cod (32.67%).
Conclusion: Reserving accuracy varies across portfolio characteristics, indicating that no single method consistently performs best. Portfolio-specific evaluation and additional validation are essential for reliable actuarial reserve estimation.

To cite this article: Suwarman, R. F., Nurrahmalika, A., Pahrany, A. D., & Aeli, L. W. (2026). IBNR estimation method on reinsurance portfolio: evaluation of chain ladder, bornhuetter-ferguson, and cape cod. *Journal of Advanced Sciences and Mathematics Education, 6*(4), 423-445.

INTRODUCTION

Accurate estimation of Incurred But Not Reported (IBNR) claim reserves remains a persistent challenge in actuarial risk management, particularly when reinsurance portfolios exhibit heterogeneous claim development patterns and substantial variability in financial exposures. The reliability of reserve estimates is essential for maintaining financial stability because underestimated liabilities may weaken an insurer's capacity to fulfill future obligations, whereas excessive reserves may distort assessments of financial performance (Taha Farahat et al., 2021). Reinsurance operations introduce additional complexity because claim liabilities are distributed across business lines with different underwriting arrangements, settlement periods, and risk characteristics. These differences complicate the application of standardized reserving procedures, especially when

***Corresponding author:**
Ramdhan Fazrianto Suwarman, Universitas Negeri Malang, Indonesia
ramdhan.fazrianto.fmipa@um.ac.id

historical claim observations provide incomplete information about ultimate liabilities (Avanzi et al., 2025; Chen, Cao, et al., 2026; Crevecoeur et al., 2023; Vanegas et al., 2025). Financial risk management studies emphasize that effective risk assessment requires quantitative approaches capable of accommodating uncertainty and differences in underlying financial conditions (Maulana et al., 2024). However, the accuracy of actuarial reserve estimates depends not only on the availability of historical information but also on the compatibility between estimation assumptions and observed claim development patterns. Portfolios characterized by irregular reporting delays, concentrated losses, or substantial changes in exposure may produce unstable development factors that influence projected reserve values (Ajang et al., 2026; Semieniuk et al., 2021). Such conditions are particularly relevant in reinsurance, where the distribution of large claims and differences in contractual arrangements may generate considerable variation in cumulative loss development (Heinrich et al., 2022). Consequently, applying identical estimation procedures across heterogeneous portfolios may produce inconsistent levels of accuracy despite the use of established actuarial techniques. These considerations establish the need for systematic quantitative comparisons that evaluate reserving performance across portfolios with different claim development characteristics rather than relying on generalized assumptions about methodological superiority.

The estimation of outstanding claim liabilities becomes technically demanding because reported and paid claims represent only the observable portion of financial obligations at a particular valuation date. Claims that have occurred but remain unreported contribute to IBNR reserves, while reported claims awaiting settlement create additional outstanding liabilities that must be distinguished in actuarial assessments. Although deterministic reserving methods provide structured mathematical procedures for projecting ultimate claims, their underlying assumptions may respond differently to variations in claim maturity and historical development patterns. The Chain Ladder method estimates future claim development by extrapolating historical cumulative loss patterns, making its performance dependent on the representativeness and stability of observed development factors (Rismawati, 2024). In contrast, the Bornhuetter–Ferguson method combines observed claim development with prior expectations of ultimate losses, potentially reducing dependence on immature historical observations (Amini & Hikmah, 2022). The Cape Cod method incorporates earned premium information and observed loss experience to estimate expected loss ratios, thereby introducing sensitivity to the relationship between premium exposure and realized claims (Pahlevi et al., 2024). These methodological differences suggest that portfolios with stable development patterns may favor historical extrapolation, whereas portfolios with incomplete or irregular claim development may respond differently to premium-based estimation approaches. Nevertheless, incorporating premium information does not necessarily improve estimation accuracy when historical premiums are poorly aligned with realized losses or when portfolio composition changes substantially over time. Previous investigations have demonstrated that reserve estimates can vary considerably according to the selected actuarial approach and the characteristics of the underlying insurance data (Wilandari et al., 2021; Hikmah & Hikmah, 2022). Therefore, evaluating these methods under comparable portfolio-level conditions is necessary to distinguish observed differences in estimation performance from differences arising from data sources, calculation procedures, and valuation periods.

Empirical comparisons of actuarial reserving methods have produced heterogeneous findings, suggesting that estimation accuracy cannot be attributed exclusively to the mathematical structure of a particular technique. Pahlevi et al. (2024) examined Chain Ladder and Cape Cod in reinsurance reserve estimation and reported comparatively favorable performance for Chain Ladder within the investigated dataset. Conversely, Firdausy and Oktaviarina (2025) identified advantages associated with Bornhuetter–Ferguson when compared with an alternative Chain Ladder approach under their study conditions. Similar comparisons conducted by Riyadi and Suradi (2024) further

illustrate how differences in insurance products and development characteristics may influence the relative performance of deterministic reserving techniques. These contrasting findings may arise from variations in claim maturity, development-factor stability, premium structures, and the treatment of incomplete claim information. Studies emphasizing relatively consistent historical claim development may obtain favorable results from Chain Ladder because its projections are derived directly from observed cumulative loss patterns. Conversely, approaches incorporating expected loss ratios may provide different estimates when historical observations are immature, although their reliability remains dependent on the appropriateness of premium-based assumptions. Differences in the composition of insurance portfolios may also influence estimation errors because low-frequency, high-severity claims can produce development patterns that differ substantially from those observed in more stable business lines (Kundu et al., 2023). Consequently, the findings of individual comparative studies should be interpreted within their respective data structures rather than generalized across all reinsurance portfolios. A broader portfolio-level evaluation is therefore warranted to clarify how established deterministic methods perform when applied consistently to heterogeneous reinsurance data.

Despite extensive applications of deterministic actuarial reserving techniques, the literature considered in this study provides limited directly comparable evidence regarding the performance of Chain Ladder, Bornhuetter–Ferguson, and Cape Cod across multiple reinsurance business lines evaluated under consistent analytical conditions. Previous investigations have frequently concentrated on pairwise methodological comparisons or individual insurance products, restricting their ability to characterize differences in estimation performance across heterogeneous portfolios (Pahlevi et al., 2024; Firdausy & Oktaviarina, 2025). Although these studies provide valuable evidence concerning the relative accuracy of particular methods, differences in data maturity, premium structures, and implementation assumptions complicate comparisons across research settings. Furthermore, the relationship between portfolio characteristics and method-specific estimation errors has often been interpreted within individual case studies rather than examined through a common portfolio-level comparative framework (Anderson et al., 2026; Chen, Wang, et al., 2026). Existing evidence also provides limited insight into how aggregate method rankings may change when portfolios with structurally distinct claim development characteristics are evaluated separately (Al-Azazi, 2026; Canton, 2025; Cricelli et al., 2021; Guidetti & Insana, 2026; Rahman, 2025; Sharkey et al., 2023). This issue is particularly important because Relative Error can become disproportionately large when reported reserve benchmarks are small, potentially influencing comparisons across portfolios with different liability magnitudes. Moreover, identifying the lowest-error method within a portfolio does not necessarily demonstrate that its estimates are sufficiently accurate for operational reserving decisions. The present study addresses these limitations by applying three established deterministic methods to nine Munich Re portfolio-level units using consistent data sources, development horizons, and evaluation procedures. Its contribution lies in providing a transparent comparison of portfolio-specific estimation errors and examining the sensitivity of aggregate performance rankings to the inclusion of structurally distinct portfolios, rather than introducing a new actuarial estimation technique. Accordingly, the study seeks to strengthen empirical understanding of the methodological trade-offs associated with deterministic IBNR reserving while acknowledging the limitations of single-source data and error-based performance evaluation.

This study aims to quantitatively evaluate and compare the performance of Chain Ladder, Bornhuetter–Ferguson, and Cape Cod in estimating Incurred But Not Reported claim reserves across heterogeneous reinsurance portfolios. The analysis utilizes cumulative paid-loss run-off triangles and earned premium information covering accident years 2010–2021, obtained from publicly disclosed Munich Re portfolio data. Specifically, the first objective is to calculate IBNR reserve

estimates for nine portfolio-level units using standardized deterministic estimation procedures. The second objective is to assess the accuracy of each method by calculating Relative Error against the corresponding reported IBNR reserve benchmarks. The third objective is to identify the method producing the minimum estimation error within each portfolio and describe variations in method rankings across different business lines. The fourth objective is to examine whether aggregate performance rankings remain consistent when three structurally distinct long-tail portfolios are excluded through sensitivity analysis. The study further considers how observed differences in claim development patterns, portfolio maturity, and premium–loss relationships may help explain variations in estimation performance, without treating these interpretations as independently established causal relationships. By maintaining consistent data sources and analytical procedures, the comparison seeks to reduce methodological inconsistencies that could obscure differences among the three reserving approaches. The expected contribution is a reproducible portfolio-level assessment that clarifies the relative strengths and limitations of established actuarial methods within the observed reinsurance dataset. Ultimately, the findings are intended to support more informed interpretation of deterministic reserve estimates and provide an empirical foundation for subsequent research incorporating additional accuracy measures, stochastic approaches, and independent validation datasets.

LITERATURE REVIEW

Actuarial reserving constitutes a fundamental component of insurance and reinsurance financial management because it provides quantitative estimates of liabilities arising from claims that have occurred but remain partially or entirely unsettled. Within this framework, Incurred But Not Reported (IBNR) reserves represent an important source of estimation uncertainty, particularly when reporting delays and incomplete claim development obscure the ultimate financial obligations of insurers and reinsurers (Taha Farahat et al., 2021). The reliability of reserve estimates depends on the availability, completeness, and representativeness of historical claims information, which may differ considerably across insurance products and reinsurance arrangements. Reinsurance portfolios introduce additional complexity because their underlying exposures may involve different claim frequencies, severity distributions, contractual structures, and settlement periods (Heinrich et al., 2022). These differences can generate heterogeneous development patterns that complicate the application of uniform actuarial assumptions across multiple business lines. Deterministic reserving methods address this challenge by projecting outstanding liabilities through mathematical relationships derived from observed claims, premium information, or combinations of historical and expected losses (Mahohoho et al., 2025; Schneider & Schwab, 2025). However, deterministic estimates generally provide point estimates rather than complete probability distributions of future reserve outcomes, limiting their ability to represent parameter uncertainty and process variability. Consequently, differences between estimated and reported reserve values may reflect not only methodological limitations but also variations in the underlying characteristics of individual portfolios. Previous research has emphasized that inaccurate reserving can affect financial reporting, risk assessment, and the interpretation of an insurer's financial position (Hikmah & Hikmah, 2022). Therefore, comparative evaluation of actuarial reserving methods is necessary to establish a clearer understanding of their relative performance under heterogeneous reinsurance conditions.

The Chain Ladder method is among the most widely applied deterministic approaches to estimating outstanding claim liabilities because it extrapolates historical development patterns to project ultimate losses. Its mathematical framework typically relies on cumulative claims arranged in run-off triangles, where observations are organized according to accident years and development

periods (Rismawati, 2024). Development factors are calculated from observed changes in cumulative claims and subsequently used to estimate the unobserved portions of each accident year's development trajectory (Abdelrahman et al., 2026; Usman & Chan, 2022; Xie & Gan, 2025). The method assumes that historical development patterns provide a sufficiently representative basis for projecting future claim emergence and settlement behavior. Under relatively stable claim development conditions, this assumption may support consistent estimation because historical observations contain information about the progression of claims toward ultimate settlement. However, irregular reporting patterns, exceptional claim events, and changes in settlement practices may distort development factors and reduce the reliability of projected liabilities. The sensitivity of Chain Ladder to historical observations becomes particularly important in reinsurance portfolios characterized by infrequent but financially significant claims. Moreover, immature accident years contain fewer observed development periods, making their ultimate loss projections more dependent on estimated development factors. Comparative evidence reported by Pahlevi et al. (2024) indicates that Chain Ladder can produce relatively favorable estimation accuracy under particular reinsurance data conditions, although such findings cannot establish universal methodological superiority. Accordingly, the performance of Chain Ladder should be evaluated in relation to portfolio-specific development stability, data maturity, and the consistency of historical claim patterns.

The Bornhuetter–Ferguson method combines observed claim development with prior expectations of ultimate losses to estimate outstanding claim liabilities (Amini & Hikmah, 2022). Unlike Chain Ladder, which derives ultimate loss projections primarily from historical development patterns, Bornhuetter–Ferguson incorporates an expected loss component that is not determined exclusively by the observed cumulative claims. This approach is particularly relevant when claim development is incomplete because the unreported or unpaid portion of expected losses can be estimated using an assumed ultimate loss ratio and corresponding exposure information. The expected loss ratio generally reflects prior actuarial assumptions concerning the relationship between anticipated claims and earned premiums, making its specification an important determinant of estimation performance (N. Nguyen, 2026; Yu et al., 2026). When historical claims are immature or subject to substantial fluctuations, the incorporation of prior expectations may reduce the sensitivity of reserve estimates to individual observations. Nevertheless, the method may produce inaccurate estimates when the assumed loss ratio does not adequately represent the underlying risk exposure or when portfolio characteristics change over time. Consequently, the reliability of Bornhuetter–Ferguson depends on both the estimated development pattern and the appropriateness of the prior expected loss assumptions. Empirical investigations by Firdausy and Oktaviarina (2025) and Riyadi and Suradi (2024) reported favorable performance for Bornhuetter–Ferguson relative to alternative reserving approaches within their respective analytical settings. These findings suggest that incorporating expected loss information can be advantageous under certain conditions, although the magnitude and consistency of such advantages remain dependent on the characteristics of the investigated portfolios. Therefore, evaluating Bornhuetter–Ferguson alongside other deterministic techniques requires explicit consideration of the assumptions underlying expected ultimate losses and their relationship with observed claim development.

The Cape Cod method represents a deterministic reserving approach that integrates historical claim development with earned premium information to estimate expected ultimate losses. Its underlying framework shares conceptual similarities with Bornhuetter–Ferguson but differs in the estimation of expected loss ratios, which are derived from observed claim experience and exposure-adjusted premium information. By incorporating premium-based exposure measures, Cape Cod provides an alternative mechanism for estimating outstanding liabilities when historical cumulative claims alone may not adequately represent ultimate loss expectations. The method

generally assumes that a meaningful relationship exists between premium exposure and expected claim costs across the observations used to estimate the loss ratio. This assumption becomes important when evaluating heterogeneous reinsurance portfolios because differences in underwriting arrangements, risk composition, and premium structures may influence the relationship between earned premiums and realized claims. Portfolios exhibiting relatively consistent premium–loss relationships may provide a more suitable basis for estimating expected loss ratios, whereas substantial exposure changes or exceptional claims may introduce estimation instability (Anders & Tang, 2026; Clemente et al., 2025; Gentile & Calvi, 2023; Vazov & Hristozov, 2026). Furthermore, the use of observed claim experience to estimate expected loss ratios creates a dependence on the representativeness of the available data, particularly when development periods are incomplete. Comparative research by Pahlevi et al. (2024) demonstrates that Cape Cod does not necessarily outperform Chain Ladder in every reinsurance context, highlighting the importance of evaluating methodological performance empirically. Nevertheless, the incorporation of premium information offers a distinct analytical perspective that may produce different reserve estimates from methods relying predominantly on historical claim development. Accordingly, Cape Cod should be assessed not only through its overall estimation accuracy but also through the extent to which premium-based assumptions remain appropriate across different reinsurance business lines.

Comparative evaluation of actuarial reserving methods requires consideration of both methodological assumptions and the characteristics of the portfolios to which those methods are applied. Previous studies have demonstrated that differences in claim development patterns, data maturity, and expected loss assumptions may contribute to variations in estimation accuracy across deterministic reserving approaches (Wilandari et al., 2021; Pahlevi et al., 2024). Chain Ladder relies heavily on the stability of historical development factors, whereas Bornhuetter–Ferguson incorporates prior expectations and Cape Cod estimates expected loss ratios using observed claims and premium exposure. These distinctions imply that methods may respond differently to irregular claim emergence, incomplete development periods, and inconsistencies between premiums and realized losses. Relative Error provides a standardized measure of proportional deviation between estimated and benchmark reserves, allowing method performance to be compared across portfolios with different liability magnitudes (Diffey et al., 2022; Hamel et al., 2026; M. D. Nguyen, 2026; Santos et al., 2026; Schneider & Schwab, 2025; Zhu et al., 2025). However, the interpretation of Relative Error requires caution because small benchmark values can magnify proportional deviations and disproportionately influence aggregate performance rankings. Moreover, a method producing the lowest error within one portfolio may not demonstrate comparable performance when applied to portfolios with different claim development characteristics. The heterogeneity of reinsurance business lines therefore provides a rationale for evaluating reserving methods individually rather than selecting a single approach solely on the basis of aggregate estimation accuracy. Within this framework, sensitivity analysis can help determine whether comparative rankings remain consistent when portfolios with structurally distinct development characteristics are excluded from the evaluation. Consequently, a portfolio-level comparison of Chain Ladder, Bornhuetter–Ferguson, and Cape Cod offers an empirical basis for identifying context-dependent estimation performance while recognizing the limitations of deterministic methods and the need for additional validation.

METHOD

Research Design and Analytical Framework

This study employed a quantitative comparative research design using secondary actuarial data to evaluate the performance of three deterministic methods for estimating Incurred But Not Reported (IBNR) claim reserves across heterogeneous reinsurance portfolios. A comparative design

was considered appropriate because the selected methods generate numerical reserve estimates from a common set of historical claims and premium information, allowing their estimation deviations to be evaluated under consistent data conditions. The analytical framework focused on differences in estimation performance rather than causal relationships between portfolio characteristics and reserve accuracy. Three established deterministic approaches were examined: Chain Ladder (CL), Bornhuetter–Ferguson (BF), and Cape Cod (CC). These methods were selected because they represent distinct approaches to incorporating historical claim development and expected ultimate losses into reserve estimation. Chain Ladder relies primarily on observed claim development patterns, Bornhuetter–Ferguson combines historical development information with prior expectations, and Cape Cod estimates expected losses using observed claims and premium-based exposure information (Amini & Hikmah, 2022; Rismawati, 2024). Their selection enabled a focused comparison of alternative deterministic assumptions without introducing additional model structures that would require different data inputs or estimation frameworks. More complex stochastic and simulation-based approaches were excluded because the objective was to compare established deterministic point-estimation procedures rather than quantify the complete probability distribution of reserve uncertainty. Accordingly, the analytical framework emphasized consistency in input data, valuation periods, development horizons, and benchmark comparisons to support interpretable differences in estimation outcomes.

The comparative framework was implemented at the individual portfolio level to preserve differences in claim development characteristics across reinsurance business lines. Each method was applied separately to the same cumulative paid-loss triangle and corresponding earned premium information within a given portfolio. Development factors were calculated using the available historical observations, and claim projections were restricted to the development periods disclosed in the source dataset. The resulting IBNR estimates were subsequently compared with the corresponding reported reserve benchmarks using a standardized evaluation procedure. This approach allowed differences in estimation errors to be examined without introducing inconsistencies arising from alternative data sources or valuation dates. The analytical framework also incorporated a sensitivity assessment to determine whether aggregate method rankings changed when three structurally distinct long-tail portfolios were excluded. This supplementary comparison was intended to assess the dependence of aggregate rankings on portfolio composition rather than establish statistical robustness or predictive superiority. Consequently, the study provides a descriptive comparative evaluation of deterministic reserving performance under the observed data conditions.

Data Sources and Portfolio Selection

The study utilized secondary data obtained from Munich Re's official publication *Paid and Reported Loss Triangles 2021*, published in 2022 (Munich Re, 2022). The dataset contained cumulative paid-loss run-off triangles, earned premium information, and reported IBNR reserve figures associated with reinsurance business lines covering accident years 2010–2021. Financial values were expressed in millions of euros, following the reporting units used in the original publication. The same official publication was used to obtain both the model inputs and the reported reserve benchmarks, thereby maintaining consistency in the source and reporting period of the compared values. The use of publicly disclosed actuarial information also supports transparency because the underlying observations can be identified from the referenced financial reporting materials. However, the analysis was restricted to the information available in the published triangles and did not incorporate confidential claims records, individual claim histories, or subsequent valuation updates. Therefore, the findings represent a retrospective evaluation of the disclosed portfolio data rather than an assessment of the insurer's current reserving position.

The analysis covered seven reinsurance business lines: Group, Property, Credit, Marine, Liability, Motor, and Personal Accident and Workers' Compensation. Liability and Motor were each separated into Proportional and Non-Proportional & Facultative categories, resulting in nine portfolio-level units. All nine units disclosed in the relevant Munich Re run-off triangle publication were included, and no portfolio was excluded from the primary comparative analysis. This inclusion approach was adopted to preserve the complete portfolio structure available within the selected publication and avoid discretionary selection based on estimation outcomes. Each portfolio was analyzed independently because aggregation across business lines could obscure differences in claim development patterns and premium exposure structures. The resulting dataset provided variation in portfolio characteristics relevant to evaluating the performance of deterministic reserving methods. Nevertheless, the portfolios originated from a single reinsurance company and therefore cannot be assumed to represent the full diversity of the international reinsurance market. Differences in underwriting policies, geographic exposure, claims management, and reporting practices may limit the generalizability of the findings to other insurers and reinsurers. Accordingly, the study was designed as a portfolio-level comparison within a defined institutional dataset rather than a representative market-wide evaluation.

Data Preparation and Computational Procedures

Data preparation involved organizing the cumulative paid-loss observations into run-off triangles according to accident year and development period for each portfolio. Earned premium information was aligned with the corresponding accident years to support the premium-based calculations required by Bornhuetter–Ferguson and Cape Cod. The original financial reporting units were retained throughout the calculations to avoid inconsistencies arising from currency conversion or alternative scaling procedures. Reserve calculations were conducted using Microsoft Excel, with the three actuarial methods applied to the same portfolio-specific input data. Development factors were derived from the observed cumulative paid-loss information using the calculation procedures specified for the respective methods. The calculations were based on the available development periods, without extending projections beyond the disclosed triangle horizon. No independent computational replication using alternative statistical software was performed, which limits the extent to which implementation errors can be independently excluded. Consequently, the reproducibility of the numerical results depends on the correct specification of spreadsheet formulas, cell references, and actuarial calculation procedures. The study therefore interprets the resulting estimates as deterministic outputs generated under the documented implementation assumptions rather than independently validated actuarial benchmarks.

No tail factor was incorporated into the reserve estimation procedures because the available dataset did not provide additional claim development observations beyond the disclosed triangle horizons. The observed development periods extended approximately 10–12 years, depending on the portfolio, and the calculations were restricted to those available periods. This restriction established a common treatment of unobserved development across the three methods within each portfolio. However, the absence of tail factors may lead to incomplete estimates of ultimate liabilities when claims continue developing beyond the available observation period. This limitation is particularly relevant for long-tail business lines, where reporting and settlement processes may extend substantially beyond the disclosed development horizon. Consequently, the resulting reserve estimates should not be interpreted as incorporating all possible future claim development. The omission of tail factors was therefore treated as a limitation of the available data rather than evidence that additional development was immaterial.

The study also retained the original published observations without applying additional outlier treatment, inflation adjustments, discounting, or other data transformations. This decision

was intended to preserve the disclosed data structure and maintain a consistent basis for comparing the three methods. Applying different adjustments across methods could introduce additional assumptions that complicate the interpretation of estimation differences. Nevertheless, using unadjusted data does not eliminate the possibility of bias arising from exceptional claims, changes in settlement practices, or economic conditions. Large observations may influence historical development factors and premium-based loss ratio estimates differently across methods. Therefore, the absence of adjustments should be understood as a standardized implementation condition rather than a demonstration that the underlying data were unaffected by irregularities. These considerations were incorporated into the interpretation of portfolio-specific estimation errors and the limitations of the comparative findings.

IBNR Estimation Procedures

The Chain Ladder method was implemented using cumulative paid-loss run-off triangles to estimate ultimate claims from observed historical development patterns. Development factors were calculated for successive development periods using the available cumulative loss observations. These factors were applied to project the unobserved portions of the triangles, and estimated outstanding liabilities were obtained from the differences between projected ultimate claims and observed cumulative paid claims. The procedure assumed that historical development patterns provided a suitable basis for extrapolating future claim development within the available observation horizon. However, this assumption may be sensitive to irregular cumulative losses, changes in settlement behavior, and the limited maturity of recent accident years (Rismawati, 2024). The method was implemented without tail-factor adjustments, consistent with the treatment applied to the other reserving approaches.

The Bornhuetter–Ferguson method combined observed claim development information with expected ultimate losses derived from premium-based assumptions. The observed development pattern was used to estimate the proportion of ultimate losses that had already emerged and the corresponding proportion remaining outstanding. Expected ultimate losses were incorporated into the reserve calculation through the loss ratio framework described in the original estimation procedure (Amini & Hikmah, 2022). Unlike Chain Ladder, this approach did not rely exclusively on observed cumulative paid claims to determine the outstanding component of ultimate losses. Its estimates were therefore influenced by both the development pattern and the expected loss assumptions incorporated into the calculation. The reliability of the resulting reserves depends on the appropriateness of these assumptions for each portfolio, particularly when observed claim development is incomplete. Because the available methodological description does not establish an independently sourced prior expected loss ratio, the results should be interpreted within the specific premium-based parameterization used in the calculations. This limitation is relevant when comparing Bornhuetter–Ferguson with approaches that derive expected loss ratios directly from observed portfolio experience.

The Cape Cod method integrated cumulative paid-loss information and earned premiums to estimate ultimate losses through an experience-based loss ratio approach. Development factors were calculated from the observed cumulative loss triangles, while earned premiums were incorporated into the estimation of expected ultimate losses. The method used the relationship between observed claims and premium exposure to derive the loss ratio parameters required for reserve estimation. Estimated outstanding liabilities were subsequently obtained by applying the corresponding development proportions to the premium-based ultimate loss estimates. This approach assumes that the observed premium–loss relationship provides a sufficiently representative basis for estimating ultimate claims within each portfolio. However, changes in underwriting composition, premium adequacy, or exceptional claims may influence the stability of the estimated loss ratios. Accordingly,

the Cape Cod estimates were interpreted in relation to the observed portfolio characteristics and the limitations of the available historical information. The method was implemented using the same source data and available development horizon as Chain Ladder and Bornhuetter–Ferguson.

Performance Evaluation and Sensitivity Analysis

The performance of the three reserving methods was evaluated by comparing their estimated IBNR reserves with the corresponding reported reserve benchmarks for each portfolio. Relative Error (RE) was employed as the primary accuracy measure because it expresses the absolute difference between estimated and benchmark values as a percentage of the benchmark reserve. The metric was calculated using the following equation:

$$Re_{p,m} = \frac{|Rp,m - Rp|}{|Rp|} \; x \; 100\%$$

where $Re_{p,m}$ represents the Relative Error for method m in portfolio p, $Re_{p,m}$ denotes the estimated IBNR reserve, and R_p represents the corresponding reported reserve benchmark. The calculation assumes a nonzero benchmark value, and lower RE values indicate smaller proportional deviations between estimated and reported reserves. For each portfolio, the method producing the minimum RE was identified as the best-performing method under the specified evaluation criterion. This classification was limited to estimation accuracy against the reported benchmark and was not interpreted as evidence of universal actuarial superiority. Although RE facilitates comparisons across portfolios with different reserve magnitudes, it may disproportionately amplify deviations when benchmark values are relatively small. Therefore, the results were interpreted with attention to the potential influence of portfolio size and benchmark magnitude.

To complement the primary portfolio-level comparison, the study examined aggregate method rankings across the nine portfolios and conducted a sensitivity analysis excluding three structurally distinct long-tail portfolios. Mean RE values were calculated separately for the complete portfolio set and the reduced portfolio set. The purpose of this analysis was to determine whether the comparative ranking of Chain Ladder, Bornhuetter–Ferguson, and Cape Cod remained consistent under different portfolio compositions. Changes in aggregate rankings were interpreted as evidence that the overall comparison was sensitive to the portfolios included in the analysis. However, the sensitivity assessment did not constitute independent out-of-sample validation because it relied on subsets of the same historical dataset. Additional measures, such as Absolute Error, reserve-weighted aggregation, and retrospective backtesting, were not incorporated into the reported analysis. Consequently, the findings provide an exploratory assessment of relative estimation performance rather than a comprehensive validation of reserving accuracy.

Research Procedure

The research procedure consisted of three sequential stages, as illustrated in Figure 1. The first stage involved extracting cumulative paid-loss triangles, earned premium information, and reported IBNR reserve benchmarks from the official Munich Re publication. The extracted observations were organized into nine portfolio-level datasets while retaining their original accident-year structure and reporting units. The second stage involved calculating reserve estimates using Chain Ladder, Bornhuetter–Ferguson, and Cape Cod under consistent portfolio-specific input conditions. The third stage involved evaluating the resulting estimates through Relative Error, identifying the minimum-error method for each portfolio, and comparing aggregate rankings. A supplementary sensitivity analysis was subsequently conducted to examine the influence of structurally distinct portfolios on the overall method rankings. This procedure established a consistent basis for comparing the three deterministic approaches while preserving the limitations associated with the available data and estimation assumptions.

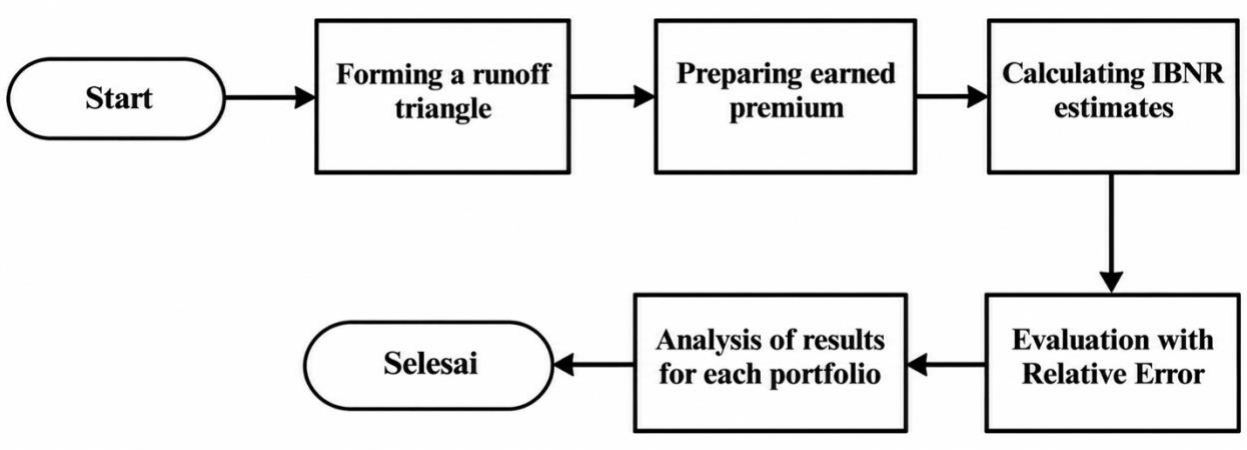


Figure 1. Research Flowchart

The research flowchart summarizes the progression from data extraction and preparation to actuarial reserve estimation, performance evaluation, and sensitivity analysis. The initial stage establishes the portfolio-specific input datasets used consistently across the three methods. The estimation stage produces separate IBNR reserve values for Chain Ladder, Bornhuetter–Ferguson, and Cape Cod. The evaluation stage compares these estimates with the reported reserve benchmarks to determine portfolio-level estimation errors. Finally, the sensitivity analysis examines whether aggregate performance rankings remain consistent after excluding the three structurally distinct long-tail portfolios.

RESULTS AND DISCUSSION

Results

Portfolio-Level Comparison of IBNR Reserve Estimates

The comparative analysis generated Incurred But Not Reported (IBNR) reserve estimates for nine Munich Re reinsurance portfolio units using the Chain Ladder (CL), Bornhuetter–Ferguson (BF), and Cape Cod (CC) methods. Each method was applied to cumulative paid-loss run-off triangles and earned premium information covering accident years 2010–2021. The estimated reserves were compared with the corresponding reported IBNR values using Relative Error (RE) as the primary evaluation metric. The analysis produced 27 portfolio–method estimates, allowing the accuracy of the three approaches to be examined under consistent portfolio-specific data conditions. Across the nine portfolios, the minimum observed RE was 0.59%, obtained by Bornhuetter–Ferguson for the Group portfolio, whereas the maximum was 220.36%, obtained by Chain Ladder for Motor Non-Proportional & Facultative. Cape Cod achieved the lowest RE in four portfolios, Chain Ladder in four portfolios, and Bornhuetter–Ferguson in one portfolio. The results also showed that all three methods produced estimates exceeding the reported benchmark in seven portfolios, while the Credit portfolio exhibited underestimation by Chain Ladder and Bornhuetter–Ferguson. The Group portfolio presented estimates on both sides of the benchmark, with Chain Ladder and Bornhuetter–Ferguson producing lower estimates and Cape Cod producing a higher estimate. These findings establish the observed distribution of estimation errors across the analyzed portfolios without attributing the differences to specific actuarial mechanisms. The portfolio-level reserve estimates and corresponding Relative Error values are presented in Table 1.

Table 1. Comparison of Actual IBNR Reserves, Estimated Reserves, and Relative Errors Across Reinsurance Portfolios

Portfolio	Actual IBNR	IBNR CL	RE CL (%)	IBNR BF	RE BF (%)	IBNR CC	RE CC (%)	Best Method
Group	21,284	20,940	1.62	21,159	0.59	23,731	11.50	BF
Property	8,057	10,719	33.04	10,742	33.33	10,953	35.94	CL
Credit	688	524	23.84	577	16.13	748	8.72	CC
Marine	571	623	9.11	627	9.81	649	13.66	CL
Liability Proportional	2,729	8,217	201.10	7,794	185.60	6,959	155.00	CC
Liability NP&FAC	968	2,828	192.15	2,329	140.60	1,590	64.26	CC
Motor Proportional	3,236	5,353	65.42	5,391	66.59	5,623	73.76	CL
Motor NP&FAC	329	1,054	220.36	961	192.10	690	109.73	CC
Personal Accident & Workers' Compensation	207	262	26.57	263	27.05	315	52.17	CL

Table 1 demonstrates substantial variation in estimation errors across the nine portfolios, with observed RE values ranging from 0.59% to 220.36%. The Group portfolio recorded the smallest estimation deviations, with RE values of 1.62%, 0.59%, and 11.50% for CL, BF, and CC, respectively. Conversely, Liability Proportional and Motor NP&FAC recorded RE values exceeding 100% under all three methods. Cape Cod produced the lowest RE in four portfolios, although its minimum errors in Liability Proportional, Liability NP&FAC, and Motor NP&FAC remained relatively large at 155.00%, 64.26%, and 109.73%, respectively. Chain Ladder produced the lowest RE in Property, Marine, Motor Proportional, and Personal Accident & Workers' Compensation, while Bornhuetter–Ferguson achieved the minimum RE exclusively in the Group portfolio.

Group Portfolio

The Group portfolio recorded an actual IBNR reserve of €21,284 million, representing the largest reported reserve benchmark among the nine analyzed portfolios. Chain Ladder estimated €20,940 million, corresponding to an RE of 1.62%. Bornhuetter–Ferguson generated an estimate of €21,159 million, producing the lowest RE of 0.59%. Cape Cod estimated €23,731 million, resulting in an RE of 11.50%. The absolute differences between estimated and reported reserves were €344 million, €125 million, and €2,447 million for CL, BF, and CC, respectively. Bornhuetter–Ferguson therefore achieved the closest estimate to the reported benchmark within this portfolio.

Property Portfolio

The Property portfolio had a reported IBNR reserve of €8,057 million. Chain Ladder estimated €10,719 million, representing an RE of 33.04%. Bornhuetter–Ferguson produced an estimate of €10,742 million, corresponding to an RE of 33.33%. Cape Cod generated the highest estimate of €10,953 million, resulting in an RE of 35.94%. All three methods exceeded the reported benchmark, with absolute deviations of €2,662 million, €2,685 million, and €2,896 million, respectively. Chain Ladder achieved the minimum RE, although its advantage over Bornhuetter–Ferguson was limited to 0.29 percentage points.

Credit Portfolio

The Credit portfolio recorded an actual IBNR reserve of €688 million. Chain Ladder generated an estimate of €524 million, corresponding to an RE of 23.84%. Bornhuetter–Ferguson estimated €577

million, producing an RE of 16.13%. Cape Cod estimated €748 million, resulting in the lowest RE of 8.72%. Chain Ladder and Bornhuetter–Ferguson underestimated the reported reserve by €164 million and €111 million, respectively, whereas Cape Cod overestimated it by €60 million. Cape Cod therefore produced the smallest absolute and proportional deviations within the Credit portfolio.

Marine Portfolio

The Marine portfolio had a reported IBNR reserve of €571 million. Chain Ladder produced an estimate of €623 million, corresponding to an RE of 9.11%. Bornhuetter–Ferguson generated an estimate of €627 million, resulting in an RE of 9.81%. Cape Cod estimated €649 million, producing an RE of 13.66%. The corresponding absolute deviations were €52 million, €56 million, and €78 million. Chain Ladder achieved the lowest RE, followed closely by Bornhuetter–Ferguson, with a difference of 0.70 percentage points between the two methods.

Liability Proportional Portfolio

The Liability Proportional portfolio recorded an actual IBNR reserve of €2,729 million. Chain Ladder generated an estimate of €8,217 million, corresponding to an RE of 201.10%. Bornhuetter–Ferguson produced an estimate of €7,794 million, resulting in an RE of 185.60%. Cape Cod estimated €6,959 million, producing the lowest RE of 155.00%. All three methods substantially exceeded the reported reserve benchmark, with absolute deviations of €5,488 million, €5,065 million, and €4,230 million, respectively. Although Cape Cod achieved the minimum RE, its estimate remained approximately 2.55 times the reported IBNR value. This portfolio exhibited the highest minimum RE among the nine analyzed portfolio units.

Liability Non-Proportional & Facultative Portfolio

The Liability Non-Proportional & Facultative (NP&FAC) portfolio recorded a reported IBNR reserve of €968 million. Chain Ladder estimated €2,828 million, corresponding to an RE of 192.15%. Bornhuetter–Ferguson produced an estimate of €2,329 million, resulting in an RE of 140.60%. Cape Cod generated an estimate of €1,590 million, corresponding to the lowest RE of 64.26%. The absolute deviations from the benchmark were €1,860 million, €1,361 million, and €622 million for CL, BF, and CC, respectively. Cape Cod reduced the proportional estimation error by 127.89 percentage points compared with Chain Ladder. Nevertheless, all three methods produced estimates exceeding the reported reserve.

Motor Proportional Portfolio

The Motor Proportional portfolio had a reported IBNR reserve of €3,236 million. Chain Ladder estimated €5,353 million, corresponding to an RE of 65.42%. Bornhuetter–Ferguson produced an estimate of €5,391 million, resulting in an RE of 66.59%. Cape Cod generated an estimate of €5,623 million, corresponding to an RE of 73.76%. The absolute deviations were €2,117 million, €2,155 million, and €2,387 million, respectively. Chain Ladder achieved the lowest RE, with an advantage of 1.17 percentage points over Bornhuetter–Ferguson. Despite this ranking, all three methods produced estimates substantially higher than the reported benchmark.

Motor Non-Proportional & Facultative Portfolio

The Motor Non-Proportional & Facultative (NP&FAC) portfolio recorded an actual IBNR reserve of €329 million. Chain Ladder generated an estimate of €1,054 million, corresponding to an RE of 220.36%, the highest individual error observed in the study. Bornhuetter–Ferguson estimated €961 million, producing an RE of 192.10%. Cape Cod generated an estimate of €690 million, resulting in the lowest RE of 109.73%. The absolute deviations were €725 million, €632 million, and €361 million for CL, BF, and CC, respectively. Cape Cod reduced the proportional estimation error by 110.63 percentage points compared with Chain Ladder. However, even the lowest-error estimate exceeded twice the reported reserve benchmark.

Personal Accident and Workers' Compensation Portfolio

The Personal Accident and Workers' Compensation portfolio recorded an actual IBNR reserve of €207 million. Chain Ladder generated an estimate of €262 million, corresponding to an RE of 26.57%. Bornhuetter–Ferguson estimated €263 million, resulting in an RE of 27.05%. Cape Cod produced an estimate of €315 million, corresponding to an RE of 52.17%. The absolute deviations were €55 million, €56 million, and €108 million, respectively. Chain Ladder achieved the minimum RE, although its advantage over Bornhuetter–Ferguson was only 0.48 percentage points. Both methods produced substantially smaller deviations than Cape Cod within this portfolio.

Aggregate Comparison of Estimation Performance

The portfolio-level results indicate that the three methods achieved different rankings across the nine analyzed reinsurance portfolios. Chain Ladder produced the lowest RE in four portfolios, Cape Cod in four portfolios, and Bornhuetter–Ferguson in one portfolio. The number of portfolio-level wins therefore did not identify a single method that consistently outperformed the alternatives. To summarize aggregate performance, the arithmetic mean of the nine portfolio-specific RE values was calculated for each method. This aggregation assigned equal weight to each portfolio regardless of its reported reserve magnitude and was used as a descriptive measure rather than a portfolio-size-adjusted estimate of overall reserving accuracy. The resulting mean RE values were 85.89% for Chain Ladder, 74.78% for Bornhuetter–Ferguson, and 58.33% for Cape Cod. Under this equal-weighted aggregation, Cape Cod produced the lowest average proportional deviation. However, the interpretation of these averages requires consideration of the substantial variation in individual portfolio errors, particularly those exceeding 100%. The distribution of Relative Error values across the nine portfolios is illustrated in Figure 2.

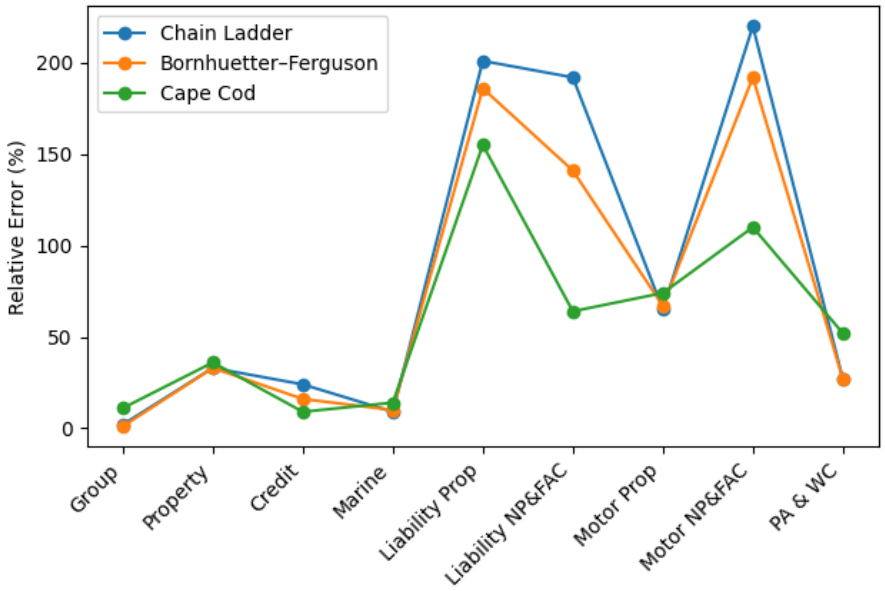


Figure 2. Relative Error of Chain Ladder, Bornhuetter–Ferguson, and Cape Cod Across Nine Reinsurance Portfolios

Figure 2 illustrates the differences in estimation accuracy among the three actuarial methods across the analyzed portfolios. The Group and Marine portfolios recorded comparatively low RE values, whereas Liability Proportional, Liability NP&FAC, and Motor NP&FAC exhibited substantially higher errors. Cape Cod produced the lowest RE in the three latter portfolios, although its deviations remained considerable. Chain Ladder achieved the lowest RE in four portfolios, while Bornhuetter–Ferguson recorded the minimum error in the Group portfolio. The figure therefore highlights the heterogeneity of estimation performance and the absence of a uniformly lowest-error method across all nine portfolios.

Sensitivity Analysis of Aggregate Method Rankings

A supplementary sensitivity analysis was conducted to examine the influence of portfolio composition on aggregate method rankings. The analysis excluded three portfolios identified as structurally distinct long-tail business lines: Liability Proportional, Liability Non-Proportional & Facultative, and Motor Non-Proportional & Facultative. These exclusions were used to evaluate the sensitivity of the descriptive mean RE rather than to redefine the primary study population. The reduced dataset therefore comprised six portfolios: Group, Property, Credit, Marine, Motor Proportional, and Personal Accident & Workers' Compensation. Mean RE values were recalculated using the same equal-weighted aggregation procedure applied to the complete dataset. The results showed that the ranking of the three methods changed when the three portfolios were excluded.

Table 2. Sensitivity Analysis of Mean Relative Error Across Portfolio Compositions

Method	Mean RE: All Nine Portfolios (%)	Mean RE: Six Portfolios (%)	Ranking: Nine Portfolios	Ranking: Six Portfolios
Chain Ladder	85.89	26.67	3	2
Bornhuetter–Ferguson	74.78	25.67	2	1
Cape Cod	58.33	32.67	1	3

Table 2 shows that Cape Cod achieved the lowest mean RE of 58.33% when all nine portfolios were included. After excluding the three structurally distinct portfolios, Bornhuetter–Ferguson recorded the lowest mean RE of 25.67%, followed by Chain Ladder at 26.67% and Cape Cod at 32.67%. The exclusion resulted in reductions of 59.22 percentage points for Chain Ladder, 49.11 percentage points for Bornhuetter–Ferguson, and 25.66 percentage points for Cape Cod. The ranking consequently changed from Cape Cod, Bornhuetter–Ferguson, and Chain Ladder in the complete dataset to Bornhuetter–Ferguson, Chain Ladder, and Cape Cod in the reduced dataset. These findings demonstrate that the descriptive aggregate rankings were sensitive to the inclusion of the three excluded portfolios.

Summary of Principal Findings

The results demonstrate that estimation accuracy varied substantially across the nine Munich Re portfolio units examined in this study. Cape Cod and Chain Ladder each achieved the lowest RE in four portfolios, while Bornhuetter–Ferguson performed best in the Group portfolio. Although Cape Cod recorded the lowest mean RE across the complete dataset, its aggregate advantage was not maintained in the reduced six-portfolio sensitivity analysis. The three portfolios excluded from that analysis exhibited substantial estimation deviations under all three methods, with minimum RE values ranging from 64.26% to 155.00%. These results indicate that the relative performance rankings depended on the portfolio units included in the comparison. Accordingly, the findings establish portfolio-specific differences in deterministic IBNR estimation accuracy within the analyzed dataset but do not independently identify the causal mechanisms underlying those differences or demonstrate generalizable superiority across the broader reinsurance market.

Discussion

The findings of this study demonstrate substantial heterogeneity in the estimation accuracy of Chain Ladder (CL), Bornhuetter–Ferguson (BF), and Cape Cod (CC) across the nine Munich Re reinsurance portfolios, indicating that the relative performance of deterministic reserving methods varies according to the portfolio being evaluated. Chain Ladder and Cape Cod each achieved the lowest Relative Error (RE) in four portfolios, whereas Bornhuetter–Ferguson produced the most accurate estimate in the Group portfolio, with an RE of 0.59%. This distribution indicates that no single method consistently generated the closest estimates to reported IBNR reserves across all

business lines. The findings are consistent with who compared Chain Ladder and Cape Cod using Munich Re data for 2013–2022 and reported lower relative error for Chain Ladder (12.8%) than Cape Cod (26.9%) in their analytical setting. However, the present study extends the comparative perspective by examining three deterministic approaches across nine portfolio-level units rather than relying exclusively on a two-method comparison. The observed variation also who examined the relationships among Chain Ladder, Bornhuetter–Ferguson, and Cape Cod within a broader actuarial modeling framework and highlighted the importance of their underlying development assumptions. Nevertheless, the methodological differences among these approaches should not be interpreted as independently established explanations for the observed estimation errors because the present analysis did not formally test the effects of individual portfolio characteristics. Instead, the results suggest that the relative accuracy of deterministic reserving techniques should be assessed within the specific data conditions under which their estimates are generated. This distinction is important because a method producing the smallest proportional deviation in one portfolio may generate substantially larger errors when applied to another portfolio with different claims experience. Accordingly, the principal contribution of this study is the identification of portfolio-specific differences in estimation performance within a common reinsurance dataset rather than the establishment of a universally superior actuarial reserving method.

The comparative performance of Chain Ladder provides further insight into the potential relationship between historical claim development information and reserve estimation accuracy. Chain Ladder generated the lowest RE in the Property, Marine, Motor Proportional, and Personal Accident and Workers' Compensation portfolios, with respective errors of 33.04%, 9.11%, 65.42%, and 26.57%. These results indicate that a development-based approach can remain competitive across several business lines, although its estimation accuracy varies substantially even among portfolios in which it achieves the lowest error. The relatively favorable result for Marine is compatible with the theoretical premise that historical development patterns can provide useful projection information when the observed claims experience is sufficiently representative (Rismawati, 2024). However, the substantially higher errors recorded for Liability Proportional, Liability NP&FAC, and Motor NP&FAC suggest that the reliability of historical extrapolation may be limited under certain portfolio conditions. Research on the impact of outliers in loss reserving has demonstrated that Chain Ladder estimates can be sensitive to individual observations within development triangles, providing a theoretical basis for examining whether unusual claims influence reserve projections (Avanzi et al., 2025; Chang et al., 2024; Diffey et al., 2022; Vanegas et al., 2025). Similarly, that a Bornhuetter–Double Chain Ladder approach produced more stable and accurate estimates than classical Chain Ladder in their investigated dataset, although their alternative model differs from the conventional Bornhuetter–Ferguson method evaluated here. The present findings should therefore be interpreted as compatible with potential development-factor sensitivity rather than as direct evidence that specific outliers caused the observed errors. Furthermore, the absence of tail-factor adjustments may influence estimates in portfolios where claim development continues beyond the available observation horizon, although the direction and magnitude of this influence were not independently quantified. These considerations suggest that the practical application of Chain Ladder should be accompanied by an assessment of development-factor stability, claim maturity, and the adequacy of the available observation period before its estimates are adopted for reserving decisions.

Bornhuetter–Ferguson demonstrated a different performance pattern, achieving the lowest RE in the Group portfolio while producing estimates relatively close to Chain Ladder in several other business lines. Its Group estimate of €21,159 million differed from the reported reserve of €21,284 million by only 0.59%, compared with 1.62% for Chain Ladder and 11.50% for Cape Cod. The relatively small deviation is compatible with the possibility that combining observed claim

development with expected ultimate losses can provide useful estimates in an aggregated portfolio, although the contribution of individual model components was not separately evaluated. Bornhuetter–Ferguson also produced RE values of 33.33%, 9.81%, and 27.05% in Property, Marine, and Personal Accident and Workers' Compensation, respectively, indicating performance relatively close to Chain Ladder in these portfolios. Demonstrated that Bornhuetter–Ferguson and classical Chain Ladder can produce substantially different reserve estimates under a specified expected loss ratio, emphasizing the importance of assumptions used to construct expected ultimate losses. This distinction is relevant because Bornhuetter–Ferguson incorporates an expected loss component whose reliability depends on the appropriateness of the underlying assumptions for the portfolio being evaluated. Previous research by Firdausy and Oktaviarina (2025) and Riyadi and Suradi (2024), as cited in the reviewed literature, similarly reported favorable results for Bornhuetter–Ferguson relative to alternative reserving approaches within their respective study contexts. Nevertheless, the present analysis does not establish whether its comparatively favorable performance resulted from the expected loss ratio, the development pattern, or interactions between these components. The absence of an independently documented source for the prior expected loss ratio also limits the extent to which the implementation can be interpreted as a conventional externally parameterized Bornhuetter–Ferguson model. Consequently, the findings support the importance of transparent expected-loss assumptions and portfolio-specific validation rather than the unconditional preference for Bornhuetter–Ferguson in aggregated or relatively mature portfolios.

Cape Cod achieved the lowest RE in Credit, Liability Proportional, Liability NP&FAC, and Motor NP&FAC, indicating comparatively favorable performance in several portfolios where the other methods generated substantial deviations. Its relative errors were 8.72%, 155.00%, 64.26%, and 109.73%, respectively, demonstrating that the method's favorable ranking did not necessarily correspond to an acceptably small estimation error. This distinction is particularly important in Liability Proportional and Motor NP&FAC, where even the lowest-error estimates substantially exceeded the reported reserve benchmarks. The comparatively smaller deviations generated by Cape Cod may be consistent with the influence of its experience-based loss ratio framework, which combines observed claims and premium exposure when estimating ultimate losses. However, the present analysis did not independently quantify premium–claim consistency, development-factor volatility, or the contribution of individual loss ratio parameters to estimation error. Accordingly, the observed performance cannot establish that Cape Cod systematically mitigates extreme claims or provides greater resistance to outliers than the alternative methods. This interpretation is consistent with the broader actuarial literature, which recognizes the importance of exposure-based assumptions while emphasizing that reserve estimates remain sensitive to model specification and the underlying data structure (Ajang et al., 2026; Mahohoho et al., 2025). For example, investigated bootstrap procedures for Cape Cod and compared their implications for reserve risk estimation, illustrating the importance of evaluating uncertainty beyond deterministic point estimates. The present results therefore suggest that incorporating premium-based information may produce comparatively closer estimates in certain reinsurance portfolios, but they do not demonstrate that this information consistently improves reserve accuracy. From a practical perspective, Cape Cod estimates should be assessed against the representativeness of historical premium exposure, the stability of observed loss ratios, and the magnitude of the remaining deviation from reported liabilities before the method is considered suitable for a particular portfolio.

The aggregate comparison and sensitivity analysis provide additional evidence that the apparent superiority of a reserving method depends on the composition of the evaluated portfolios and the criteria used to measure estimation accuracy. Across all nine portfolios, Cape Cod recorded the lowest arithmetic mean RE of 58.33%, followed by Bornhuetter–Ferguson at 74.78% and Chain

Ladder at 85.89%, suggesting an aggregate advantage for Cape Cod under equal portfolio weighting. However, excluding Liability Proportional, Liability NP&FAC, and Motor NP&FAC changed the ranking, with Bornhuetter–Ferguson recording the lowest mean RE of 25.67%, followed by Chain Ladder at 26.67% and Cape Cod at 32.67%. This reversal demonstrates that the aggregate comparison is sensitive to the inclusion of portfolios with exceptionally large proportional deviations and should not be interpreted as evidence of stable methodological superiority. Moreover, the use of arithmetic mean RE assigns equal importance to portfolios with substantially different reserve magnitudes, potentially limiting its suitability for evaluating financial materiality across an entire reinsurance portfolio. The sensitivity analysis provides useful descriptive evidence regarding ranking stability, although the exclusion of selected portfolios does not constitute independent predictive validation and cannot establish that the remaining portfolios are statistically homogeneous. These findings reinforce the importance of considering portfolio-level estimation errors alongside aggregate performance measures when evaluating deterministic reserving techniques (Bhattacharjee et al., 2026; Cheng et al., 2026; Etcil et al., 2026; Li et al., 2026; Tseng et al., 2026). The study is also limited by its reliance on a single reinsurer's published data, the absence of tail-factor adjustments, the use of unadjusted historical observations, and the lack of independent computational validation. Future investigations should incorporate additional valuation periods, external reinsurance datasets, absolute and reserve-weighted error measures, and retrospective backtesting to determine whether the observed rankings remain consistent under alternative evaluation conditions. Stochastic reserving approaches and uncertainty quantification could further complement deterministic comparisons by distinguishing estimation differences from the variability inherent in future claim development, as emphasized in recent actuarial research. Ultimately, the findings support a portfolio-specific and evidence-based approach to IBNR reserve estimation while restricting conclusions about comparative accuracy to the Munich Re portfolios, historical observations, and deterministic implementation assumptions examined in this study.

Implications

The findings of this study provide important theoretical, methodological, and practical implications for actuarial reserving practices, particularly in selecting appropriate deterministic approaches for estimating Incurred But Not Reported (IBNR) liabilities across heterogeneous reinsurance portfolios. From a theoretical perspective, the variation in estimation accuracy among Chain Ladder, Bornhuetter–Ferguson, and Cape Cod reinforces the importance of considering portfolio-specific characteristics when interpreting the performance of established actuarial reserving methods. The absence of a consistently superior method across all nine Munich Re portfolios suggests that methodological suitability cannot be determined exclusively by the underlying mathematical structure of a reserving technique. Instead, actuarial assessments should consider the compatibility between model assumptions, observed claim development patterns, premium exposure information, and the maturity of historical claims data. From a methodological perspective, the differences between portfolio-level and aggregate performance rankings highlight the limitations of relying exclusively on arithmetic mean Relative Error as an indicator of overall estimation accuracy. The reversal in method rankings following the exclusion of three structurally distinct portfolios further demonstrates the value of sensitivity analysis for evaluating how portfolio composition influences comparative conclusions. These findings also emphasize the need for complementary evaluation measures, including absolute estimation errors, reserve-weighted accuracy indicators, and retrospective validation procedures, to strengthen the reliability of future comparative assessments. From a practical perspective, reinsurance actuaries may benefit from evaluating multiple reserving methods within individual business lines rather than applying a uniform estimation approach across portfolios with different development characteristics. Although Cape Cod produced the lowest average Relative Error across the complete dataset, its substantial

deviations in several portfolios indicate that comparative superiority does not necessarily imply sufficient accuracy for financial reporting or operational reserving decisions. Similarly, the relatively favorable performance of Chain Ladder and Bornhuetter–Ferguson in the reduced portfolio analysis suggests that historical development information and expected loss assumptions should be evaluated jointly with portfolio-specific evidence rather than treated as universally reliable estimation inputs. For reinsurance institutions and financial risk managers, these implications support the development of transparent reserving frameworks that document model assumptions, examine estimation discrepancies, and incorporate independent validation before reserve estimates are used to inform financial decisions. Nevertheless, because the analysis was restricted to publicly disclosed Munich Re data, deterministic point estimates, and a single reporting period, these implications should be interpreted as context-specific guidance requiring further examination across additional reinsurers, valuation periods, and actuarial modeling frameworks.

Limitations and Suggestions for Future Research

This study has several limitations that should be considered when interpreting the comparative performance of Chain Ladder, Bornhuetter–Ferguson, and Cape Cod in estimating Incurred But Not Reported (IBNR) reserves. First, the analysis relied exclusively on Munich Re data covering accident years 2010–2021, limiting the generalizability of the findings to other reinsurance institutions. Second, the deterministic approaches provided point estimates without explicitly quantifying reserve uncertainty. Third, the absence of tail-factor adjustments and additional data treatments may have influenced estimation accuracy, particularly in long-tail portfolios. Fourth, Relative Error was the sole accuracy metric, making aggregate comparisons sensitive to differences in portfolio size and reported reserve values. Fifth, the sensitivity analysis assessed changes in method rankings but did not provide independent predictive validation. Furthermore, the lack of independent computational verification and fully documented expected loss ratio assumptions limited methodological reproducibility. Future research should incorporate multiple reinsurers, additional valuation periods, and more diverse portfolio characteristics. Stochastic reserving techniques should also be considered to quantify uncertainty surrounding estimated liabilities. Additional evaluations using Absolute Error, reserve-weighted measures, and retrospective backtesting would strengthen comparative accuracy assessments. Researchers should further examine the effects of development-factor stability, tail adjustments, and premium–loss relationships on reserve estimation. Ultimately, these improvements could support more reliable, transparent, and portfolio-specific actuarial reserving practices.

CONCLUSION

This study contributes to actuarial reserving research by providing a systematic portfolio-level comparison of Chain Ladder, Bornhuetter–Ferguson, and Cape Cod using consistent data sources, estimation procedures, and evaluation criteria across nine Munich Re reinsurance portfolios covering accident years 2010–2021. The principal contribution lies in demonstrating that comparative method rankings depend on portfolio composition and evaluation conditions, highlighting the limitations of selecting reserving methods solely on the basis of aggregate estimation accuracy. Chain Ladder and Cape Cod each achieved the lowest Relative Error (RE) in four portfolios, while Bornhuetter–Ferguson produced the closest estimate in the Group portfolio. Although Cape Cod recorded the lowest mean RE across all nine portfolios (58.33%), the sensitivity analysis excluding three structurally distinct portfolios changed the ranking, with Bornhuetter–Ferguson (25.67%) and Chain Ladder (26.67%) outperforming Cape Cod (32.67%). These findings suggest methodological trade-offs that may be associated with differences in claim development patterns,

portfolio maturity, and premium–loss relationships, although these associations were not independently tested. From a practical perspective, Chain Ladder warrants consideration when historical development factors are sufficiently stable and representative, whereas Bornhuetter–Ferguson may be appropriate when credible prior expected loss ratios are available to complement incomplete claims experience. Cape Cod may provide a useful alternative when earned premiums adequately represent underlying risk exposure and observed claims support reliable estimation of experience-based loss ratios. However, the substantial estimation errors observed in several portfolios demonstrate that achieving the lowest relative error does not necessarily indicate satisfactory reserve accuracy. Consequently, actuarial method selection should incorporate portfolio-specific diagnostics, transparent parameter assumptions, and comparisons against relevant reserve benchmarks rather than relying exclusively on aggregate method rankings. The findings are limited to the analyzed Munich Re portfolios, deterministic estimation procedures, and available development horizons, without independent validation of outlier sensitivity or predictive performance. Future research should incorporate additional reinsurance datasets, tail-factor adjustments, stochastic uncertainty assessment, and complementary accuracy measures to evaluate the consistency and practical reliability of these comparative findings.

AKNOWLEDGMENT

The authors would like to thank Universitas Negeri Malang for fully funding this research under Grant/Contract No. 13.4.278/UN32/KP/2026. This support was vital to the successful implementation and conclusion of the study.

AUTHOR CONTRIBUTIONS STATEMENT

Ramdhan Fazrianto Suwarman contributed to conceptualization, methodology, formal analysis, investigation, and original manuscript drafting. Aurellia Nurrahmalika contributed to data collection, data curation, actuarial calculations, and visualization. Andi Daniah Pahrany contributed to methodology, validation, interpretation of results, and manuscript review and editing. Lita Wulandari Aeli contributed to supervision, validation, critical review, and manuscript editing. All authors reviewed and approved the final version of the manuscript and agreed to be accountable for their respective contributions.

REFERENCES

Abdelrahman, H., Badescu, A., & Lin, X. S. (2026). An effective Bayesian GLM approach for IBNR claim count estimation. *North American Actuarial Journal*, *30*(3), 575–598. https://doi.org/10.1080/10920277.2025.2587884

Ajang, N. M., Ngesa, O., & Aduda, J. (2026). A segmentation-based hybrid Chain Ladder-XGBoost framework for incurred but not reported claims estimation. *Nigerian Journal of Technological Development*, *23*(1), 113–128. https://doi.org/10.63746/njtd.v23i1.4437

Al-Azazi, S. A. (2026). Task-dependent reliability evidence in maritime networks: An auditable evidence framework for port disruption and resilience decisions. *Reliability Engineering & System Safety*, *278*, 113682. https://doi.org/10.1016/j.ress.2026.113682

Amini, R., & Hikmah, Y. (2022). Estimasi cadangan klaim IBNR menggunakan metode Chain-Ladder dan Bornhuetter-Ferguson pada produk indemnity di PT. XYZ. *MAp (Mathematics and Applications) Journal*, *4*(1), 50–59. https://doi.org/10.15548/map.v4i1.4203

Anders, N. A., & Tang, M. (2026). Valuation of cyber catastrophe bonds and their role in portfolio efficiency: An analysis of model selection and investment implications. *The Geneva Papers on Risk and Insurance - Issues and Practice*, *51*(4), 920–961. https://doi.org/10.1057/s41288–026–00393–4

Anderson, R. M., Kim, B., & Ryu, D. (2026). Long-history principal component analysis in a dynamic factor model with weak loadings. *Operations Research*, *74*(4), 1770–1789. https://doi.org/10.1287/opre.2024.1134

A.S, J. B., Muchlian, M., & Rosita, S. (2025). Estimasi cadangan klaim IBNR dan RBNS produk asuransi kendaraan bermotor menggunakan metode Double Chain Ladder. *Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika*, *8*(2), 522–530. https://doi.org/10.30605/proximal.v8i2.5385

Avanzi, B., Tan, X., Taylor, G., & Wong, B. (2025). On the evolution of data breach reporting patterns and frequency in the United States: A cross-state analysis. *North American Actuarial Journal*, *29*(4), 833–864. https://doi.org/10.1080/10920277.2025.2457491

Bhattacharjee, A., Bhattacharjee, R., Bose, S. K., Xiong, Z., & Sikdar, B. (2026). A portfolio-theoretic framework for hierarchical resource allocation in ultra-dense mmWave networks. *IEEE Transactions on Network Science and Engineering*, *13*, 10595–10615. https://doi.org/10.1109/TNSE.2026.3715652

Canton, E. (2025). A portfolio approach to research funding. *Research Policy*, *54*(1), 105129. https://doi.org/10.1016/j.respol.2024.105129

Chang, L., Gao, G., & Shi, Y. (2024). Claims reserving with a robust generalized additive model. *North American Actuarial Journal*, *28*(4), 840–860. https://doi.org/10.1080/10920277.2023.2259445

Cheng, W., Li, N., Wang, L., Jiang, M., & Song, Y. (2026). Regime-conditioned reserve calibration and hydrogen flexibility for virtual power plant scheduling. *Energies*, *19*(19), 4554. https://doi.org/10.3390/en19194554

Chen, Z., Cao, Y., & Quan, Z. (2026). Assessing insurer's litigation risk: Claim dispute prediction with actionable interpretations using machine learning techniques. North American Actuarial Journal, 1–27. https://doi.org/10.1080/10920277.2026.2722246

Chen, Z., Wang, M., Zeng, Z., & Ping, W. (2026). Uncertainty-aware financial forecasting: Leveraging conformal prediction for risk-adjusted models. *IEEE Access*, *14*, 90053–90077. https://doi.org/10.1109/ACCESS.2026.3702666

Clemente, G. P., Della Corte, F., Savelli, N., & Zappa, D. (2025). Market-consistent valuation and capital assessment for demographic risk in life insurance: A cohort approach. *North American Actuarial Journal*, *29*(2), 494–518. https://doi.org/10.1080/10920277.2024.2412675

Crevecoeur, J., Antonio, K., Desmedt, S., & Masquelein, A. (2023). Bridging the gap between pricing and reserving with an occurrence and development model for non-life insurance claims. *ASTIN Bulletin: The Journal of the IAA*, *53*(2), 185–212. https://doi.org/10.1017/asb.2023.14

Cricelli, L., Grimaldi, M., Rogo, F., & Strazzullo, S. (2021). Patent ranking indicators: A framework for the evaluation of a patent portfolio. *International Journal of Intellectual Property Management*, *11*(2), 185–218. https://doi.org/10.1504/IJIPM.2021.114619

Diffey, W., Hobern, L., Lauder, A., Cleugh, M., Wu, M., Mukherjee, S., Dan, A., Dharamshi, P., Atwal, R., Vijay, A., Dolan, F., Grewal, A., & Harrison, E. (2022). Your reserves may be best estimate, but are they valid? British Actuarial Journal, 27, e2. https://doi.org/10.1017/S1357321721000179

Etcil, M., Kolukısa, B., & Güngör, B. B. (2026). Uncertainty-aware stock selection and portfolio optimization with deep learning-based return prediction. Annals of Operations Research. https://doi.org/10.1007/s10479–026–07448–8

Firdausy, A. E., & Oktaviarina, A. (2025). Perbandingan keakuratan metode Mack Chain Ladder dan Bornhuetter-Ferguson dalam estimasi cadangan klaim. *Jurnal Ilmiah Matematika dan Pendidikan Matematika (JMP)*, *17*(1), 53–72. https://doi.org/10.20884/1.jmp.2025.17.1.16200

Gentile, R., & Calvi, G. M. (2023). Direct loss-based seismic design of reinforced concrete frame and wall structures. *Earthquake Engineering & Structural Dynamics*, *52*(14), 4395–4415. https://doi.org/10.1002/eqe.3955

Guidetti, V., & Insana, A. (2026). Trust anomalies' judgment: Social ranking theories for portfolio construction. *European Journal of Operational Research*, *329*(3), 1063–1072. https://doi.org/10.1016/j.ejor.2025.07.067

Hamel, E., Abdallah, A., & Léveillé, G. (2026). A unified micro-model for loss reserves, IBNR, and unearned premium risk with dependence, inflation, and discounting. North American Actuarial Journal, 1–21. https://doi.org/10.1080/10920277.2026.2743706

Heinrich, T., Sabuco, J., & Farmer, J. D. (2022). A simulation of the insurance industry: The problem of risk model homogeneity. *Journal of Economic Interaction and Coordination*, *17*(2), 535–576. https://doi.org/10.1007/s11403–021–00319–4

Hikmah, Y., & Hikmah, I. R. (2022). Perhitungan cadangan klaim dengan metode Chain Ladder menggunakan Excel dan RStudio. *MAp (Mathematics and Applications) Journal*, *4*(2), 122–131. https://doi.org/10.15548/map.v4i2.4837

Kundu, P., Kundu, A., & Hawlader, B. (2023). Stochastic comparisons of largest claim amounts from heterogeneous portfolios. *Statistica Neerlandica*, *77*(4), 497–515. https://doi.org/10.1111/stan.12296

Li, B., Athar, M., Khan, M. A., Muqtadir, A., Guo, Q., & Pan, H. (2026). Quantifying and monetizing demand-side potential at the grid edge: Methods for aggregation, bidding and retail optimization. *Energies*, *19*(4), 893. https://doi.org/10.3390/en19040893

Mahohoho, B., Chimedza, C., Matarise, F., & Munyira, S. (2025). The advanced actuarial data science based AI-driven solutions for automated loss reserving under IFRS 17 in non-life insurance. *Science, Engineering and Technology*, *5*(2), 75–115. https://doi.org/10.54327/set2025/v5.i2.210

Maulana, A., Dwita, M., Fitriyani, M., Sunaryo, D., & Adiyanto, Y. (2024). Risk management as a determinant of Indonesian banking financial performance: A systematic literature approach. *Indo-Fintech Intellectuals: Journal of Economics and Business*, *4*(5), 2523–2537. https://doi.org/10.54373/ifijeb.v4i5.2120

Munich Re. (2022). Paid and reported loss triangles 2021 [Data set]. Munich Re.

Nguyen, M. D. (2026). RMSE-triggered rebalancing for deep learning-guided portfolio optimization. *The Journal of Finance and Data Science*, *12*, 100200. https://doi.org/10.1016/j.jfds.2026.100200

Nguyen, N. (2026). Machine learning and frequency-severity decomposition for insurance pricing. *Mathematics*, *14*(10), 1640. https://doi.org/10.3390/math14101640

Pahlevi, Z. A., Nur, F., & Sudding, F. (2024). Comparison of Chain Ladder method and Cape Cod method in reinsurance incurred but not reported (IBNR) reserve estimation. *Journal of Actuarial, Finance and Risk Management*, *3*(2).

Rahman, M. H. (2025). Machine learning for nonperforming corporate loan recovery: A systematic review and meta-analysis of predictive performance. *American Journal of Scholarly Research and Innovation*, *4*(1), 1174–1202. https://doi.org/10.63125/a7v6f533

Rismawati, R. (2024). Estimasi besar cadangan klaim IBNR pada asuransi properti menggunakan metode Chain Ladder dan Bootstrapping [Thesis, UIN Sunan Gunung Djati].

Riyadi, A. A., & Suradi, L. (2024). Estimasi cadangan klaim IBNR pada asuransi kredit menggunakan metode Munich Chain Ladder dan Bornhuetter-Ferguson pada PT XYZ. *MATHunesa: Jurnal Ilmiah Matematika*, *12*(2).

Santos, S., Silva, N., Sebastião, H., & Godinho, P. (2026). Enhancing stock return prediction and portfolio performance with adaptive LASSO. Investment Analysts Journal, 1–25. https://doi.org/10.1080/10293523.2026.2636363

Schneider, J. C., & Schwab, B. (2025). Advancing loss reserving: A hybrid neural network approach for individual claim development prediction. *Journal of Risk and Insurance*, *92*(2), 389–423. https://doi.org/10.1111/jori.12501

Semieniuk, G., Campiglio, E., Mercure, J.-F., Volz, U., & Edwards, N. R. (2021). Low-carbon transition risks for finance. *WIREs Climate Change*, *12*(1), e678. https://doi.org/10.1002/wcc.678

Sharkey, A., Kovács, B., & Hsu, G. (2023). Expert critics, rankings, and review aggregators: The changing nature of intermediation and the rise of markets with multiple intermediaries. *Academy of Management Annals*, *17*(1), 1–36. https://doi.org/10.5465/annals.2021.0025

Taha Farahat, A., Cosgrave DOCOsoft, B., Rashwan, W., Taha, A., Cosgrave, B., McKeever, S., & Authors Ayman Taha Farahat, A. (2021). Insurance reserve prediction: Opportunities and challenges. https://doi.org/10.21427/trtz-8t14

Tseng, C.-J., Su, Y.-K., Shih, Y.-T., & Chang, M.-H. (2026). A leakage-controlled framework for explainable and calibrated telematics-based claim prediction and pricing in usage-based insurance. *IEEE Access*, *14*, 166849–166866. https://doi.org/10.1109/ACCESS.2026.3741369

Usman, F., & Chan, J. S. K. (2022). New loss reserve models with persistence effects to forecast trapezoidal losses in run-off triangles. *ASTIN Bulletin: The Journal of the IAA*, *52*(3), 877–920. https://doi.org/10.1017/asb.2022.17

Vanegas, S. C., Badescu, A. L., & Lin, X. S. (2025). Claim reserving via inverse probability weighting: A micro-level Chain-Ladder method. *European Actuarial Journal*, *15*(1), 89–135. https://doi.org/10.1007/s13385–024–00395–3

Vazov, R., & Hristozov, Z. (2026). Reinsurance as a mechanism for optimizing capital requirements under the Solvency II regime: An empirical analysis of a ten-year insurance portfolio. *Journal of Risk and Financial Management*, *19*(9), 687. https://doi.org/10.3390/jrfm19090687

Wilandari, Y., Gunardi, G., & Effendie, A. R. (2021). Estimasi cadangan klaim menggunakan metode deterministik dan stokastik. *Jurnal Statistika Universitas Muhammadiyah Semarang*, *9*(1), 56–63. https://doi.org/10.26714/jsunimus.9.1.2021.56–63

Xie, S., & Gan, C. (2025). Modeling age-to-age development factors in auto insurance through principal component analysis and temporal clustering. *Risks*, *13*(6), 100. https://doi.org/10.3390/risks13060100

Yu, F., Zhang, M., Li, Z., & Yang, M. (2026). Ship insurance in the era of AI: An intelligent risk profiling system under the POM principles. *Journal of Risk and Insurance*, *93*(1), 92–117. https://doi.org/10.1111/jori.70033

Zhu, Y., Taasim, S. I., & Daud, A. (2025). Volatility modeling and tail risk estimation of financial assets: Evidence from gold, oil, Bitcoin, and stocks for selected markets. *Risks*, *13*(7), 138. https://doi.org/10.3390/risks13070138