Journal of Mathematics and Modeling in Finance (JMMF)
Vol. 3, No. 2, Summer & Autumn 2023
Research paper

Cross-sectional estimation of loss reserve for cargo insurance market: the case of cargo insurance in Iran

Parissa Ghonji[1], **Ghadir Mahdavi**[2], **Mitra Ghanbarzadeh**[3]

[1] Department of Actuarial Science and Insurance Planning, ECO College of Insurance, Allameh Tabataba'i University, Tehran, Iran

parissaghf@gmail.com

[2] Department of Actuarial Science and Insurance Planning, ECO College of Insurance, Allameh Tabataba'i University, Tehran, Iran

mahdavi@atu.ac.ir

[3] Insurance Research Center, Tehran, Iran
ghanbarzadeh@irc.ac.ir

Abstract:
Insurance companies regularly estimate loss reserves due to delays in settling claims. These delays depend on the time taken from claim filing to settlement. The study aims to estimate reported loss reserves through cross-sectional regression using cargo insurance market data. The model considers written premiums, paid claims, reinsurance issued premiums, inflation rates, and return on investment. The analysis demonstrates a non-significant negative association between inflation rates and loss reserves, as well as a negative correlation between paid claims and loss. While revealing a statistically significant positive relationship between written premiums and loss reserves.

Keywords: Loss Reserve, Cargo Insurance, General Insurance, Premium, Regression Analysis.
Classifications: G22, C02, E43.

1 Introduction

Cargo insurance is a vital financial safeguard in the world of international trade and shipping. It is a specialized form of insurance that provides coverage for goods and commodities being transported across oceans and other waterways. This type of insurance protects businesses and individuals from potential financial losses due to various risks such as accidents, theft, damage, or loss of cargo during transit. Cargo insurance policies can be tailored to meet the specific needs of shippers, whether they are importers, exporters, or logistics companies. In an increasingly interconnected global economy, cargo insurance plays a crucial role in ensuring the smooth

[1] Corresponding author
Received: 24/11/2023 Accepted: 22/02/2024
https://doi.org/10.22054/JMMF.2024.76913.1110

flow of goods, mitigating risks, and providing peace of mind to those involved in international trade. It serves as a safety net that helps keep the wheels of commerce turning, even in the face of unforeseen challenges on the high seas.
Cross-sectional estimation of loss reserves is a critical aspect of risk management in the cargo insurance market. It involves predicting the funds needed to cover potential cargo damage or loss during transit, considering factors like cargo type, transportation mode, and historical loss data. This estimation is vital for insurers as it helps them effectively manage risks, ensuring they have adequate funds to settle claims, maintain financial stability, and uphold their market reputation. Factors like written premiums, paid claims, reinsurance, inflation rates, and investment rates are also important variables to consider. Studies in other insurance sectors, such as fire insurance, emphasize the significance of accurate loss reserve estimation for long-term solvency and profitability, advocating the use of advanced statistical models and data analytics tools.
Loss reserves in insurance are estimates set aside to cover future claims, impacting an insurer's profitability and solvency. They are influenced by unpredictable loss events, and regulators require them to be reported at nominal value rather than discounted present value. Loss reserves affect an insurer's tax liabilities and can be used for income smoothing. In the context of cargo insurance, they are determined based on the likelihood of cargo loss or damage during transit, with insurance companies using statistical models and historical data. In Iran, cargo insurance is typically provided by licensed insurance companies but is not mandatory.
The aim of this article is to investigate and examine the crucial notion of loss reserves in the insurance sector, with a specific focus on its importance in the realm of cargo insurance. The article endeavors to offer a thorough comprehension of the nature of loss reserves, their calculation methods, and their impact on the financial stability and profitability of insurance firms. Additionally, it seeks to clarify the regulatory and tax considerations related to these reserves. In the context of cargo insurance in Iran, the article aims to clarify the role of loss reserves and the dynamics of cargo insurance within the Iranian insurance industry. The study also introduces and examines hypotheses regarding the connections between written premiums, paid claims, inflation rates, and loss reserves. Ultimately, this article provides a valuable resource for insurance professionals, researchers, and policymakers seeking a deeper understanding of the intricate domain of loss reserves and their multifaceted implications.
The research is structured into five main sections, which include the introduction, literature review, theoretical framework, data analysis, and conclusion.
The proposed hypotheses are as follows:

(i) There exists a positive correlation between written premiums/paid claims and loss reserves.

(ii) There is an inverse relationship between the inflation rate and loss reserves.

2 Literature review

Within the realm of insurance, the estimation of loss reserves holds paramount importance for ensuring financial stability and regulatory compliance. This literature review amalgamates insights from four distinctive studies, each offering valuable perspectives on the cross-sectional estimation of loss reserves, particularly in the context of cargo insurance.

In the fall of 2019, Xin Che's study, featured in The Journal of Insurance Issues, investigated the interplay between market competition and loss-reserving practices within the U.S. property-liability insurance sector [8]. The findings of the study suggest that heightened competition compels insurers to adopt less conservative loss reserve estimation strategies. This alignment with theoretical predictions accentuates the influence of managerial career concerns in competitive markets, potentially leading to earnings manipulation for short-term firm performance. Additionally, the study highlights nuances in the competition-conservatism relation, emphasizing that insurers with a tendency to over-reserve adjust their practices more significantly due to competitive pressures [8]. This contribution underscores the pivotal role of market dynamics in shaping loss-reserving practices.

Explored in The Journal of Finance and Data Science by Song and Heo in November 2022, this study introduces an innovative approach to loss reserve estimation using unsupervised-supervised machine learning techniques [9]. The incorporation of hierarchical clustering and artificial neural networks (ANN) aims to address statistical shortcomings in traditional linear estimation methods. Through a combined approach, especially leveraging Boosting and ANN, the study establishes a more robust and efficient framework for improving insurers' reserve error prediction. Notably, the research underscores the consistency of influential variables over time, signifying the predictability of future loss reserve errors [9]. This advancement contributes significantly to the existing literature by presenting a more advanced and reliable prediction method.

The work by Badounas and Pitselis in February 2020, featured in Risks, introduces a loss reserving model applicable to a general insurance portfolio [10]. This innovative model integrates quantile regression for longitudinal data and considers correlated run-off triangles. The method combines between- and within-subportfolio estimating functions, enhancing parameter estimation efficiency. Notably, the approach has proven resilient to error correlation structures. Furthermore, the study extends its application beyond traditional realms, demonstrating utility in estimating the reserve risk margin and value at risk (VaR) in both actuarial and finance contexts. Consequently, the research significantly contributes by offering an approach that not only improves parameter estimation efficiency but also provides valuable insights into risk margin and VaR estimation.

Greg Taylor's survey, featured in Risks in May 2019, offers a comprehensive overview of recent developments in granular and machine learning models for loss reserv-

ing [11]. The study meticulously traces the historical evolution of these models, drawing comparisons with older, more primitive methods. Taylor's evaluation extends to assessing the relative merits of granul ar and machine learning models, taking into account factors that influence the choice between them. The paper concludes by delving into potential future developments in these models, providing a holistic understanding of the historical evolution and the current landscape of loss reserving models.
In conclusion, these four studies collectively enrich our understanding of the cross-sectional estimation of loss reserves, offering insights into market dynamics, innovative machine learning approaches, quantile regression models, and the evolving landscape of loss reserving methodologies. This literature review provides a nuanced and comprehensive perspective, contributing valuable insights for practitioners, researchers, and policymakers in the insurance industry.

3 Theoretical Foundation

3.1 Cargo Insurance And Loss Reserve

Cargo insurance is a type of insurance that protects goods or merchandise while it is being transported from one place to another. It provides coverage for any loss or damage to the cargo during shipment, whether it is transported by land, sea, or air. Cargo insurance can be purchased by the buyer, the seller, or the transporter, and it typically covers various risks such as damage from accidents, theft, natural disasters, or other unforeseen circumstances during transit. This type of insurance is crucial for businesses involved in the import and export of goods, as it helps mitigate financial losses that may occur due to damaged or lost cargo.
Cargo insurance clauses A, B, and C are standardized terms commonly used in cargo insurance policies to define the extent of coverage and the perils for which cargo is protected during transit. Clause A offers the broadest and most comprehensive coverage, providing protection against nearly all risks of physical loss or damage to the cargo, except for specific excluded perils listed in the policy. Clause B, on the other hand, offers intermediate coverage, protecting cargo against a defined list of perils, typically including common risks like fire, collision, and sinking, but excluding other perils like theft or leakage. Lastly, Clause C is the most limited form of coverage, offering protection against a narrow set of perils such as sinking, collision, and stranding. cargo owners often choose between these clauses based on the value of their cargo, the perceived risks of their specific shipping route, and their budget constraints, tailoring the insurance coverage to their unique needs and circumstances.
In the context of cargo insurance, a loss reserve refers to the funds set aside by an insurance company to cover potential claims or losses that may arise from insured cargo. It represents the estimated amount of money that the insurer anticipates

will be needed to settle future claims for cargo damage or loss during transit. The purpose of establishing a loss reserve is to ensure that the insurance company has adequate resources to fulfill its obligations to policyholders in the event of a claim. Setting aside a loss reserve helps insurers manage their financial risks and maintain the necessary liquidity to handle claims promptly and efficiently. The amount allocated to the loss reserve is determined based on various factors, including historical claims data, risk assessment, and industry trends. Adjustments to the loss reserve may be made over time as new information becomes available and as claims are settled.

3.2 Panel Data

Panel data structures are essentially collections of time series data for multiple individuals over various time periods. These datasets can be either balanced or unbalanced, with balanced datasets recording all individuals at every time period and unbalanced datasets not observing all individuals consistently. There is a presumption of correlation or clustering over time for each individual, while the observations for different individuals are assumed to be independent.
The Chow test, also known as the F-limer test, is utilized for panel data to ascertain if regression models estimated on different sub-samples of a panel dataset are statistically distinct. It was proposed by economist Gregory Chow in 1960. The model for this test is based on a panel dataset with n cross-sectional units observed over T time periods.

$$Y_{it} = \beta_0 + \beta_1 X_{it1} + \beta_2 X_{it2} + \cdots + \beta_k X_{itk} + u_{it} \tag{1}$$

where Y_{it} is the dependent variable for cross-sectional unit i at time X_{it1}, X_{it2}, $\cdots, X_{itk}$ are the k independent variables for cross-sectional unit i at time $t, \beta_0, \beta_1, \beta_2$, $\cdots, \beta_k$ are the unknown parameters to be estimated, and u_{it} is the error term.
the F-statistic for Chow test in panel data can be defined as:

$$F = \frac{(RSS_R - RSS_{UR})/q}{(RSS_{UR})/(n-k-2q)} \tag{2}$$

In the comparison between a restricted (pooled regression) and an unrestricted (separate regressions for each group) model, the terms are defined as follows:

(i) RSSR stands for the residual sum of squares for the restricted model.

(ii) RSSUR stands for the residual sum of squares for the unrestricted model.

(iii) q denotes the number of restrictions imposed on the coefficients of the interaction term.

(iv) n indicates the total number of observations in the dataset.

(v) k represents the total number of coefficients in the unrestricted model, excluding the constant term.

The null hypothesis assumes that the coefficients and error variances are the same across different subsets of the data, while the alternative hypothesis suggests that they are different. If the calculated F-statistic is greater than the critical value at a particular significance level, it implies that the null hypothesis can be rejected, indicating the presence of a structural break in the data.
The fixed effects regression model can be expressed as:

$$Y_{it} = \alpha_i + \beta_k X_{k,it} + \varepsilon_{it} \tag{3}$$

The Rational approach focuses on studying how variables change over time, accounting for potential biases or impacts caused by internal factors within individuals. It emphasizes controlling for the correlation between the error term and independent variables. The Fixed Effect Model, within this framework, eliminates the influence of unchanging individual characteristics, allowing for a clear assessment of the independent variables' impact on the dependent one.
The model's components include T for entities and t for time, with α_i representing the unknown intercept for each entity, Y_{it} as the dependent variable, β_k as the coefficients for the independent variables, and ε_{it} as the error term. Each group or entity is assigned a distinct intercept, enabling the model to consider specific unobserved factors that may affect the dependent variable. This approach is valuable for analyzing the effects of particular policy changes or events on the observed phenomenon, such as loss reserves in insurance portfolios.

3.3 Cross-Sectional Estimation (Regression Analysis)

The Ordinary Least Squares (OLS) regression method is commonly used to determine parameters in multiple linear equations. It is applied for cross-sectional estimates, such as those related to loss reserves in fire insurance, by analyzing the correlation between these reserves and selected independent variables. However, before utilizing OLS, it's crucial to assess if the linear regression model's assumptions are met, which involves diagnostic tests for factors like normal distribution of residuals, multicollinearity, heteroscedasticity, and serial correlation.
When employing the cross-sectional regression method, the analysis focuses on one dimension of the data, without considering factors like time or spatial dimensions. This technique is applied to both regression and classification problems, like linear and logistic regression, and is also often seen in machine learning models.
The key assumptions in regression analysis include linearity, independence of observations, homoscedasticity, normality of errors, and the absence of multicollinearity. These assumptions form the basis for valid and reliable statistical inference and interpretation of the regression model's results. Violating these assumptions can im-

pact the credibility and interpretability of the findings, necessitating careful assessment and consideration of potential remedies or alternative modeling approaches.

Regression Equation

The regression equation model can be expressed as:

$$Y_i = \beta_0 + \sum_{i=1}^{5} \beta_i X_i + \varepsilon \quad where : \varepsilon \sim N(0, 1)$$

In the regression model, the dependent variable is denoted as Y, while the independent variables are represented as X_0, X_1 up to X_k. The coefficients $\beta_0, \beta_1, \beta_2, \cdots, \beta_k$ correspond to the regression coefficients for each independent variable, and ε stands for the error term, accounting for unexplained variability in the model.
The key assumptions in the regression model are:

(i) Linearity: The relationship between the dependent variable Y and the independent variables $X_1, X_2, \cdots, X_k$ is assumed to be linear.

(ii) Independence: The errors are independent of each other, represented as $\varepsilon_i \neq \varepsilon_j$ for $i \neq j$, where ε_i and ε_j represent the errors associated with observations i and j, respectively.

(iii) Homoscedasticity: The variance of the errors is consistent across all levels of the independent variables, expressed as $Var(\varepsilon_i) = \sigma_2$ for all i, where σ_2 denotes the constant variance of the error term.

(iv) Normality: The errors follow a normal distribution, indicated as $\varepsilon_i \sim N(0, \sigma_2)$ for all i, where ε_i represents the error associated with observation i, $N(0, \sigma_2)$ represents a normal distribution with a mean of 0 and a variance of σ_2.

To ensure the validity of regression results, it is essential to uphold these assumptions. Any violations should be addressed before interpreting the findings of the regression analysis.

Newey west estimator

The Newey-West estimator is a weighted least-squares method that corrects for heteroskedasticity and autocorrelation in a regression analysis.
Let Y be an $n \times 1$ vector of dependent variable observations, X be an $n \times k$matrix of independent variable observations, β be a $k \times 1$ vector of regression coefficients, and u be an $n \times 1$ vector of error terms.
The standard OLS estimator of β is given by:

$$\beta = (X'X)^{(-1)} X'Y \tag{4}$$

where $(X'X)^{(-1)}$ is the inverse of the matrix product $X'X$.
The Newey-West estimator modifies the OLS estimator by introducing a weight matrix, w, that captures the temporal dependence in the error terms. The weight matrix is defined as:

$$(h) = \left(1 - \frac{h}{T}\right)\frac{Z'Z}{T} \tag{5}$$

where h is the lag of the autocorrelation, T is the sample size, and Z is an n Œ 1 vector of residuals from the OLS regression. The weight matrix is a function of the lag h and is scaled by T to ensure that the weights sum to one.
The modified estimator, $\hat{\beta}_{NW}$, is given by:

$$\hat{\beta}_{NW} = (X'WX)^{-1}X'WY \tag{6}$$

The Newey-West estimator, despite the presence of heteroskedasticity and autocorrelation, remains reliable and asymptotically normally distributed under reasonable conditions. It is a popular choice in econometric and time series analyses, as well as other fields, where the assumptions of homoskedasticity and uncorrelated errors may not be met.

4 Methodology

The choice between pooled OLS and fixed effect models in panel data analysis can be determined through the F-Limer test, which assesses variation in entity-specific intercepts. Non-normality and heteroskedasticity issues can be resolved by transforming data using the logarithm to ensure normal distribution of residuals and employing the Newey-West estimator, which provides robust standard errors considering autocorrelation and heteroskedasticity for more accurate regression results.

4.1 Codebook and Data Requirements

In our comprehensive dataset, we elaborate on each variable's definition and units. We provide insights into our methodological decisions and the experimental setup used.
The variables in our dataset include:

(i) WP (Written Premium): This denotes the insurance premium specified in the policy, derived from the sum of direct and indirect insurance premiums. Direct premiums are associated with policies offered by insurance companies, while indirect premiums refer to amounts issued by representatives of insurance companies and brokers.

(ii) RIP (Reinsurance Issued Premium): This represents the portion of risk accepted by a reinsurer from the original insurer in exchange for a specified premium.

(iii) Cost of Loss or Claim: This encompasses the expenses paid for reported damages in direct insurances and relies on the information provided by the ceding insurer in acceptance reinsurance.

(iv) Loss Reserve: This signifies the amount reserved by insurers annually for incurred, unreported, and unpaid losses, reflecting past experience and potential future losses.

(v) ROI (Return on Investment): As an alternative to the absence of investment rates in Iranian financial statements, we utilize ROI derived from the Profit and Loss Statement. It quantifies the efficiency of an investment or compares returns from various investments, calculated as the ratio of investment returns to costs.

$$ROI = \frac{Revenue - Expenses}{Expenses} \times 100$$

In this study, a straightforward sampling approach is employed. The primary objective is to employ a multiple cross-sectional regression model to assess the potential estimation of loss reserves based on specific chosen explanatory variables. The examined variables include written premiums (WP), paid claims (PCLAIM), reinsurance issued premiums (RIP), inflation rate (INFRATE), and the interest rate (ROI).
The dataset used for regression analysis consists of annual data spanning nine years from nine distinct insurance companies operating in Iran.

4.2 Main Hypothesis of the Thesis

(i) Hypothesis 1: There is positive relationship between written premium/paid claims and loss reserve.

(ii) Hypothesis 2: There is negative relationship between inflation rate and loss reserve.

5 Data Analysis

The raw data is the unprocessed source material, which, in this context, is obtained from Codal.ir, housing the audited financial reports of Iranian insurance firms. However, this raw data necessitates preprocessing or cleaning using the R programming language before analysis.
Processed data, cleaned and primed for analysis, undergoes various transformations, merges, and standardization procedures from the raw dataset. It is imperative to document all processing steps for the sake of transparency and reproducibility.
Tidy data conforms to a specific structured format, where each variable occupies

its column, and each observation resides in its row. This organization facilitates convenient data manipulation and analysis.
Panel data, observing entity behavior over time, can be longitudinal (tracking the same entities over time) or cross-sectional (observing different entities simultaneously). Panel data analysis aims to manage unobserved entity-specific variables or those changing over time but not across entities.
Our study involves a balanced panel dataset, encompassing data from nine companies across nine years. The balanced panel data condition is satisfied when n equals N multiplied by T; in contrast, unbalanced panel data satisfies the condition n less than N multiplied by T. $(n < N \times T)$
Our research objective revolves around exploring the relationship between Loss Reserve and other factors, namely Written Premium, Paid Claims, and Inflation rate.

5.1 Descriptive Statistics

The data set contains 81 valid values for the variables WP, RIP, Claims, InfRate, ROI, and Reserve. Missing values are not mentioned. The subsequent columns display the minimum, maximum, and mean of each variable, while the last column represents the standard deviation (σ), indicating the spread of the data in relation to the mean. A low standard deviation implies that the data are closely grouped around the mean, while a high standard deviation suggests a more dispersed distribution.

Table 1: Descriptive Statistics (all units are million rials)

	N	Minimum	Maximum	Mean	Std. Deviation
WP	81	11194	3372952	309091.19	478225.669
RIP	81	10	195239	12983.04	23837.674
Claims	81	129	780148	68147.15	133773.828
InfRate	81	7	36	22.68	11.699
ROI	81	0.01	28	1.90	3.746
Reserve	81	221	654100	72938.95	116837.182
Valid N (listwise)	81				

5.2 Inferential Statistics in Eviews

In the previous section, panel data was introduced, and two regression models were discussed: ordinary least squares (OLS) and fixed effects (FE). To determine the preferred model, an F-Limer test was conducted, assessing whether the coefficients on individual-specific variables are collectively equal to zero. The obtained p-value

from the F-Limer test is 0.6443, which fails to reject the null hypothesis, indicating that the appropriate model to use is pooled OLS. The tables below provide detailed information on the findings. The symbol C denotes the average intercept value across all companies, precisely calculated as 0.4242.
To ascertain the appropriate model for our dataset, we utilized the F-Limer test, which aids in selecting between a pooled ordinary least squares (OLS) model and a fixed-effects model. Our findings indicate that the p-value doesn't provide sufficient evidence to reject the null hypothesis, implying that the preferable model is the pooled OLS. (prob = 0.6443)
Considering our dataset, we observe 81 entries comprising independent variables such as Written Premium (WP), Reinsurance Issued Premium (RIP), Paid Claims (Claims), Inflation Rate (InfRate), Investment Rate (ROI), and the dependent variable, Loss Reserve. To facilitate Cross-Sectional Regression, the temporal dimension and company codes were eliminated, as discussed in the earlier section.

5.3 Diagnosis Test

Normality

Our primary regression model can be expressed as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

Where: Y : loss reserve
X_1: Written Premiums
X_2: Paid Claims
X_3: Inflation Rate
This section focuses on checking the normality of residuals in a regression model, emphasizing their random distribution around zero with a horizontal pattern. Diagnostic plots, including residual plots, normal probability (Q-Q) plots, scale-location plots, and Cook's distance plots, are used to evaluate the model's quality. The interpretation of these plots helps identify potential issues such as nonlinearity, heteroscedasticity, or influential observations, aiming to enhance the model's accuracy and reliability. However, in the provided diagnostic plots, while the top-left graph suggests near-zero residuals, indicating possible normality, the Q-Q plot displays some extreme values, making it challenging to definitively confirm normality. The histogram indicates that most residuals are concentrated near zero, with few observations in the tails representing extreme values. This suggests that the residuals of our regression model follow a normal distribution. Using the Kolmogorov-Smirnov test on a dataset of 81 observations, it was found that the p-value is less than 0.05. Consequently, the null hypothesis was rejected, leading to the conclusion that the residuals do not conform to a normal distribution.

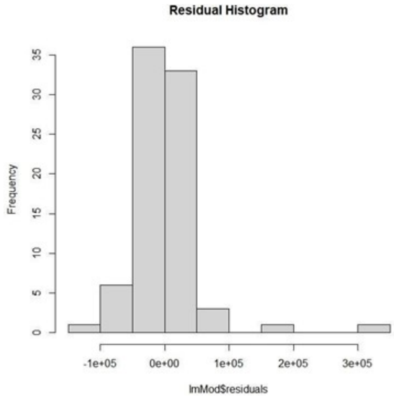

Figure 1: Histogram of Residuals in R

Table 2: Results of Normality Tests

Shapiro-Wilk	0.680	0.000
Kolmogorov-Smirnov	0.215	9e-04
Cramer-von Mises	6.827	0.000
Anderson-Darling	5.738	0.000

Linearity and Homoscedasticity

The provided graphs demonstrate that not all predictors exhibit a linear relationship with the response variable. To verify error variance stability, the Breusch-Pagan test is utilized. Rejection of the null hypothesis implies the presence of heteroscedasticity, suggesting that the covariates are correlated. In this specific case, the calculated p-value of 0.01202 indicates the absence of homoscedasticity, affirming the presence of correlated covariates.

$$BP = 14.639, \quad df = 5, \quad p - value = 0.1202$$

As the p-value shows, there is no homoscedasticity.

Autocorrelation

The correlation analysis initially reveals significant correlations between Loss Reserves (LR) and Written Premiums (WP), Paid Claims (Claims), and Reinsurance Issued Premiums (RIP). Furthermore, indications of serial correlation are present. Notably, there are noteworthy correlations between independent variables, such as WP and PCLAIM, WP and REINIPR, as well as RIP and Claims. These findings suggest the possibility of multicollinearity, emphasizing the need for thorough investigation into the correlations among the independent variables.

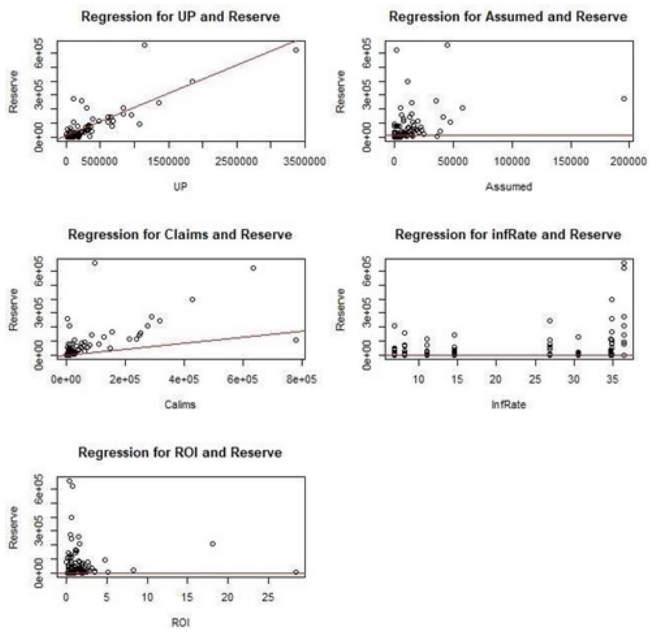

Figure 2: Diagnostics plots

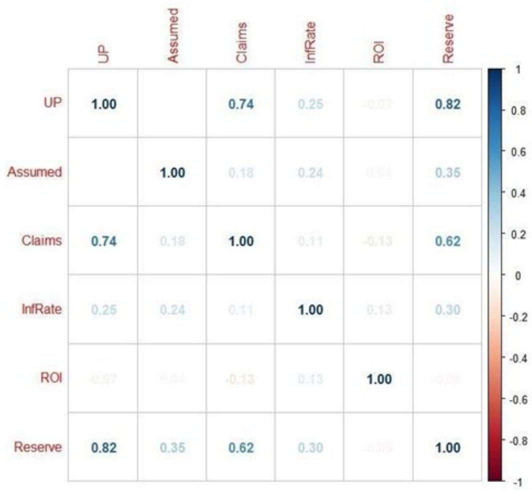

Figure 3: Correlation Matrix

5.4 Transforming

In regression analysis, it's assumed that data follows a normal distribution, but this isn't always the case. To handle this, data can be transformed. When using pooled

OLS and finding non-normal data distribution, one solution is to use the logarithm of the data. This can lead to a more normal distribution, helping meet the assumption for regression analysis. Using log(data) instead of the original data in pooled OLS often yields improved results and helps meet the assumptions.(according to figure 4)

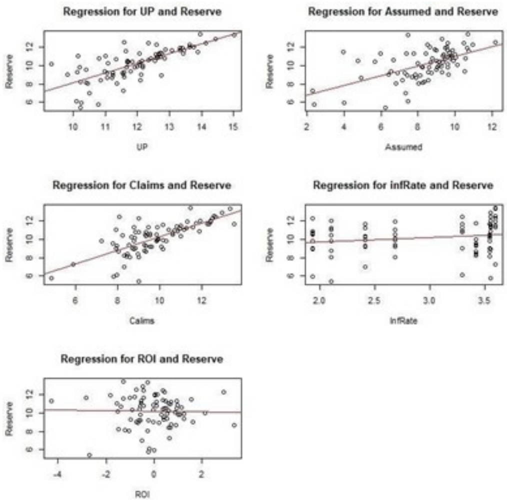

Figure 4: Diagnosis plots

Table 3: Results of Normality Tests after Transforming

Test	Statistic	p-value
Shapiro-Wilk	0.969	0.049
Kolmogorov-Smirnov	0.105	0.309
Cramer-von Mises	7.865	0.000
Anderson-Darling	1.057	0.008

Based on the Kolmogrov-Smirnov Test and Figure 7-4, the transformed data satisfies the assumptions of Normality and Linearity. However, the Breuc-Pagan Test indicates the presence of Heteroscedasticity (HET) in the data, meaning the error term's variance is not constant. Consequently, the Ordinary Least Squares (OLS) estimators may not be efficient. To address this, White's Heteroscedasticity-Corrected Standard Errors (SE) or Eicker-Huber-White SE are employed, resulting in larger SE and smaller p-values. In this case with 81 observations, the sample size is reasonably large, making the Newey-West SE a suitable choice, despite being more appropriate for larger samples. Due to the failure to meet the requirements for the initial regression model, the Robust Newey-West Estimator is chosen for further analysis.

Table 4: Results of Normality Tests after Transforming

	Estimate	Std. Error	t value	Pr(> $\|t\|$)	Sig.
(Intercept)	-4.355	2.546	-1.710	0.091	
log(X) WP	0.887	0.251	3.528	0.000	***
log(X) Claims	0.059	0.117	0.511	0.610	
log(X) RIP	0.404	0.092	4.386	3.702	***
log(X) ROI	-0.017	0.118	-0.149	0.881	
log(X) InfRate	-0.034	0.170	-0.201	0.840	

The Regression Equation can be written as follows:

$$\begin{aligned}\log(Lossreserve) &= -4.355 + 0.887\log(WP) + 0.059\log(Claims)\\ &\quad + 0.404\log(RIP) - 0.017\log(ROI) - 0.034\log(InfRate)\end{aligned}$$

6 Conclusion

The text outlines a study on the cargo Insurance Market. It introduces the topic and reviews previous research in section one and two, followed by an analysis of data and presentation of results in section five. The results indicate a significant positive relationship between written premium and loss reserve, but no significant connections were found between paid claims or inflation rate and loss reserve. The study's limitations include its focus on Iran's cargo insurance, reliance on secondary data, a small sample size, and the omission of external factors. Suggestions for future research include broadening the study's scope, using primary data sources, increasing the sample size, and considering external factors' impacts.

Acknowledgement

This research has been designed and conducted in collaboration with the Insurance Research Center affiliated with the Central Insurance of the Islamic Republic of Iran.

Bibliography

[1] G. C. Chow, *Economic growth and air quality influences on energy sources depletion, forest sources and health in MENA*, Econometrica 28 (3), https://doi.org/10.2307/1910133 (1960), 591-605.

[2] Y. Ibrahim, T. A. E. A. Taha and M. S. Minai , *Cross Sectional Estimation of Loss Reserve for Egyptian Fire Insurance Market*, European Journal of Social Sciences, 25 (4), (2011), 456-464.

[3] T. Mack, *Measuring the Variability of Chain Ladder Reserve Estimates*, Casualty Actuarial Society, 1, (1993), 101-182.

[4] P. Narayan, T. Warthen , *A comparative study of the performance of loss reserving methods through simulation*, (Doctoral dissertation), Journal of Actuarial Practices, 8, (2000), 63-88.

[5] A. A. Soleha, Y. Hikmah , *Claim Reserve Estimation Using Double Chain Ladder Method*, International Journal of Industrial Engineering & Production Research , https://doi.org/10.22068/ijiepr.3333 33(33), (2022), 4-18.

[6] G. Taylor, G. McGuire , *Loss reserving with GLMs: a case study*, Presented at the Accident Compensation Seminar, Institute of Actuaries of Australia, November 28-December 1, (2004).

[7] A. Wolny-Dominiak , *The hierarchical generalized linear model and the bootstrap estimator of the error of prediction of loss reserves in a non-life insurance company*, (2016).

[8] X. Che, *Insurance Market Competition and Reserve Management*, Journal of Insurance Issues, 42 (2), (2019), 1-38.

[9] I. J. Song, W. Heo, *Improving insurers loss reserve error prediction: Adopting combined unsupervised-supervised machine learning techniques in risk management*, The Journal of Finance and Data Science, 8, https://doi.org/10.1016/j.jfds.2022.09.003, (2022), 233-254.

[10] I. Badounas, G. Pitselis, *Loss Reserving Estimation With Correlated Run-Off Triangles in a Quantile Longitudinal Model*, Risks, 8 (1), 14, https://doi.org/10.3390/risks8010014, (2020).

[11] G. Taylor, *Loss Reserving Models: Granular and Machine Learning Forms. School of Risk and Actuarial Studies, University of New South Wales*, Kensington, NSW 2052, Australia, (2019).

How to Cite: Parissa Ghonji[1], Ghadir Mahdavi[2], Mitra Ghanbarzadeh[3], *Cross-sectional estimation of loss reserve for cargo insurance market: the case of cargo insurance in Iran*, Journal of Mathematics and Modeling in Finance (JMMF), Vol. 3, No. 2, Pages:161–176, (2023).