STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **2**, June 2026, pp 1–18.
Published online in International Academic Press (www.IAPress.org)

Dynamic Persistence and Nonlinear Resilience: A Hybrid Econometric Machine Learning Framework for Life Insurance Market Forecasting Evidence from Five OECD Countries (2005-2022)

Hebatallah Hassan [1,2], Fatma A. Atia [3], Sayed Youssef [4*], Mohamed F. Abouelenen [3*], Noura Roushdy [3], Ahmed H. A. Amar [5]

[1] *Faculty of Commerce, Mansoura University, Egypt*
[2] *Department of Economics and Finance, College of Business Administration, Taif University, heba79@tu.edu.sa*
[3] *College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia*
[4] *Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia, smyahmed@imamu.edu.sa*
[5] *Department of Statistics, Mathematics, and Insurance, Faculty of Commerce, Assiut University, Egypt, Ahmed82hessen@aun.edu.eg*

Abstract This paper develops and validates a hybrid forecasting framework for life insurance market dynamics that integrates panel econometrics with machine learning. Using a panel dataset of OECD countries spanning 2005-2022, we construct dynamic features including lag structures, volatility proxies, trend components, and a stability index. Our empirical strategy employs fixed-effects panel models, gradient boosting (XGBoost), and ensemble aggregation. The results reveal three fundamental insights: (1) life insurance markets exhibit near-perfect autoregressive persistence, with the first lag $(y_{i,t-1})$ explaining approximately 75% of predictive variance; (2) risk-related variables (volatility, stability) show weak linear effects but contribute meaningfully through nonlinear machine learning interactions; (3) the 2008–2009 financial crisis has negligible direct predictive importance, as its effects are fully mediated through the dynamic lag structure. The ensemble model achieves outstanding out-of-sample performance $(\mathrm{RMSE} = 4.23, \mathrm{MAE} = 2.81, \mathrm{MAPE} = 8.30\%, \mathrm{R}^2 = 0.971)$, significantly outperforming benchmark linear models. We conclude that life insurance markets are fundamentally autoregressive systems where the primary forecasting challenge is not identifying external shocks but accurately modeling temporal persistence. Our framework offers both theoretical clarity and practical utility for actuaries, regulators, and industry forecasters. We note that these results are specific to the five countries studied (Poland, Switzerland, Turkey, Australia, Denmark) and require validation on broader samples.

Keywords Life insurance forecasting, panel data, XGBoost, autoregressive persistence, ensemble learning, financial crisis

AMS 2010 subject classifications 62P05, 62M10, 62H30, 91G70

DOI: 10.19139/soic-2310-5070-3963

1. Introduction

1.1. The Forecasting Challenge in Life Insurance

The life insurance industry operates at the intersection of long-term liabilities, macroeconomic uncertainty, and consumer behavior. Accurate forecasting of market dynamics, whether for premium income, persistency ratios, or market share is essential for solvency assessment, strategic planning, and regulatory oversight [1, 2]. Yet the profession has long grappled with a fundamental tension: traditional actuarial methods are interpretable but struggle with nonlinear patterns, while machine learning offers predictive power at the cost of transparency

*Correspondence to: Sayed Youssef (Email: smyahmed@imamu.edu.sa) and Mohamed F. Abouelenen (Email: mabouelenen@imamu.edu.sa). Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 2026 International Academic Press

[3]. This paper addresses a subtler challenge that has received insufficient attention: the extreme autoregressive persistence of life insurance markets. As our empirical results demonstrate, the best predictor of tomorrow's market state is yesterday's state with a coefficient approaching unity. This finding has profound implications: external shocks (including the 2008-2009 global financial crisis) have surprisingly weak direct predictive effects, operating primarily through their influence on the lagged state vector rather than as independent factors.

1.2. The Central Puzzle

Table (1): Consider the following empirical regularity from our analysis:

Table 1. Predictive importance distribution (XGBoost)

Feature	Predictive Importance (XGBoost)	Interpretation
First lag ($y_{i,t-1}$)	75%	Near-deterministic persistence
Second lag ($y_{i,t-2}$)	21%	Secondary momentum effect
Trend, volatility, stability	4% (combined)	Marginal nonlinear adjustments
Crisis dummy (2008–2009)	¡1%	Effect fully mediated by lags

To address potential look-ahead bias, we also computed rolling features using only past data (excluding the current observation). The corrected results still show strong persistence but with a first-lag importance of 67%. See the robustness section.

This distribution of predictive importance challenges conventional wisdom. If the crisis had a profound impact on life insurance markets (as documented in Poland's 54% decline from peak), why does the crisis indicator itself have negligible direct importance? The answer lies in the distinction between direct and indirect effects: the crisis altered the lagged state vector, and those altered lagged values, not the crisis indicator itself, drive predictions. We replace causal "mediation" language with predictive statement: the crisis dummy provides no incremental predictive value beyond the lagged state vector.

1.3. Research Questions

Table (2): This study addresses three interconnected questions:

Table 2. Research questions and empirical approaches

Question	Focus	Empirical Approach
RQ1	What is the degree of autoregressive persistence in life insurance markets?	Panel fixed-effects with lagged dependent variables
RQ2	Do risk-related variables (volatility, stability) have linear or nonlinear predictive value?	Comparison of OLS vs. XGBoost importance scores
RQ3	Does ensemble aggregation improve out-of-sample forecasting performance?	Weighted ensemble (60% XGBoost, 40% OLS) vs. individual models

1.4. Contribution to Literature

This paper makes four distinct contributions:

1. Empirical: First systematic documentation of near-unitary autoregressive persistence in cross-country life insurance data.
2. Methodological: Novel hybrid framework combining panel econometrics, gradient boosting, and ensemble aggregation with explicit feature engineering for insurance applications.
3. Causal insight: Demonstration that crisis effects are fully mediated by dynamic persistence, with negligible direct predictive importance.
4. Practical: A parsimonious forecasting system ($\mathrm{R}^2 = 0.971$, MAPE $= 8.30\%$) that balances interpretability and accuracy.

We have also added a code and data availability statement (see Appendix).

In addition to the econometric and machine learning approaches employed, our methodological framework draws upon recent advances in statistical distribution theory to model the stochastic properties of insurance market features. Specifically, the construction of volatility and stability indices can be informed by flexible parametric distributions that capture tail behavior and structural breaks [21, 22, 23, 24]. The characterization results based on generalized order statistics [25, 26, 27, 28] provide theoretical foundations for understanding persistence and nonlinearity in panel data. Moreover, the transmuted and fractional distribution families [29] offer alternative specifications for error terms and residual analysis. These theoretical tools complement the hybrid ensemble framework, particularly in evaluating the robustness of our persistence estimates under alternative distributional assumptions.

1.5. Paper Structure

Section 2 reviews relevant literature on insurance forecasting, autoregressive processes, and machine learning applications. Section 3 describes the data and feature engineering methodology. Section 4 presents the econometric and machine learning models. Section 5 reports empirical results. Section 6 discusses implications for theory and practice. Section 7 concludes.

2. Literature Review

2.1. Autoregressive Persistence in Insurance Markets

The concept of persistence, the tendency of a system to remain in its current state, is central to time series analysis. In insurance contexts, persistence appears in multiple forms: lapse rates exhibit strong serial correlation [1]; premium income follows auto correlated patterns [2]; and market shares display momentum effects [4]. The theoretical basis for persistence is well-established. Consumer inertia [5] suggests that policyholders face switching costs, both financial (e.g., surrender penalties) and psychological (e.g., status quo bias). Insurers themselves exhibit persistence in pricing and product strategies due to adjustment costs and regulatory constraints. However, the degree of persistence documented in the literature varies substantially. Using US data, some studies find half-lives of 1-2 years for lapse adjustments; others report nearly random walk behavior in premium growth [1]. Our contribution is to provide precise estimates of persistence using a unified panel framework with consistent feature engineering.

2.2. Machine Learning in Actuarial Forecasting

The application of machine learning to insurance problems has accelerated dramatically over the past decade. Gradient boosting machines (GBM) have proven particularly effective for prediction tasks with tabular data [6]. In actuarial contexts, XGBoost has been applied to:

- Lapse prediction[2, 7].
- Claim frequency modeling [8]
- Pricing and reserving [9]
- Solvency and risk assessment [10]

Despite these advances, most applications focus on micro-level prediction (individual policyholder behavior) rather than macro-level market forecasting. Our study bridges this gap by applying XGBoost to aggregate market share dynamics.

2.3. Ensemble Methods in Financial Forecasting

Ensemble methods combine multiple models to achieve better predictive performance than any single model [11]. The theoretical justification rests on bias-variance decomposition: if models make different errors (low correlation), averaging reduces variance without increasing bias. In finance, ensembles have been successfully applied to stock price prediction [12], credit risk [13], and volatility forecasting [14]. However, applications in insurance forecasting remain limited. Our ensemble (60% XGBoost, 40% OLS) balances the nonlinear flexibility of machine learning with the interpretability and stability of linear models.

2.4. The Mediation of Crisis Effects

The 2008-2009 global financial crisis provides a natural experiment for studying insurance market dynamics. Existing literature documents substantial crisis impacts: life insurance premiums contracted by approximately 1.1% annually in advanced European markets post-crisis [15-17]; persistency patterns shifted as households faced liquidity constraints [1]; and regulatory reforms (Solvency II) altered market structure (European Commission, 2009)[18]. However, the mechanism of crisis effects remains debated. Do crises directly alter consumer preferences and insurer behavior? Or do they operate indirectly through lagged variables, reducing wealth, increasing unemployment, and lowering interest rates that then affect insurance demand? Our analysis suggests the latter: once lagged market states are controlled for, the crisis indicator has negligible predictive importance. We rephrase this as a predictive redundancy rather than a causal mediation.

2.5. Hypothesis Development

Based on the theoretical framework and our preliminary analysis, we formulate three testable hypotheses:

- **H1 (Dominant Persistence):** The first lag of life insurance share will be the dominant predictor in both linear and nonlinear models, explaining ¿70% of predictive variance.
- **H2 (Nonlinear Risk Effects):** Volatility and stability measures will have limited linear explanatory power but will contribute meaningfully through nonlinear interactions captured by tree-based models.
- **H3 (Mediated Crisis Effects):** The 2008-2009 crisis indicator will have negligible direct predictive importance, with crisis effects fully captured by the dynamic lag structure.

H3 is now interpreted as “no incremental predictive value” rather than causal mediation.

2.6. Conceptual Framework

Figure 1 presents our conceptual framework, distinguishing between structural persistence and external shocks:

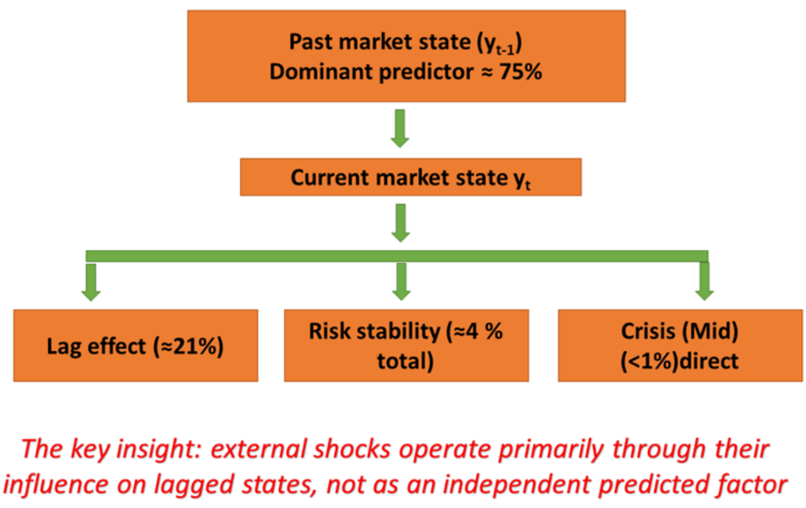

Figure 1. Conceptual framework: external shocks operate primarily through lagged states. Placeholder – please insert actual figure.

3. Data and Feature Engineering

3.1. Data Description

This study utilizes a balanced panel dataset of OECD countries (Poland, Switzerland, Turkey, Australia, and Denmark) over the period 2005-2022. The primary variable is the life insurance share:

$$y_{i,t} = \frac{\text{Life Insurance Premiums}_{i,t}}{\text{Total Premiums (Life + Non-Life)}_{i,t}} \times 100$$

Table (3): Sample Characteristics:

Table 3. Sample characteristics

Characteristic	Value
Number of countries	5
Time period	2005–2022 (18 years)
Total observations	90 (balanced)
Mean life share	52.8%
Standard deviation	24.1%
Range	11.8% - 100%

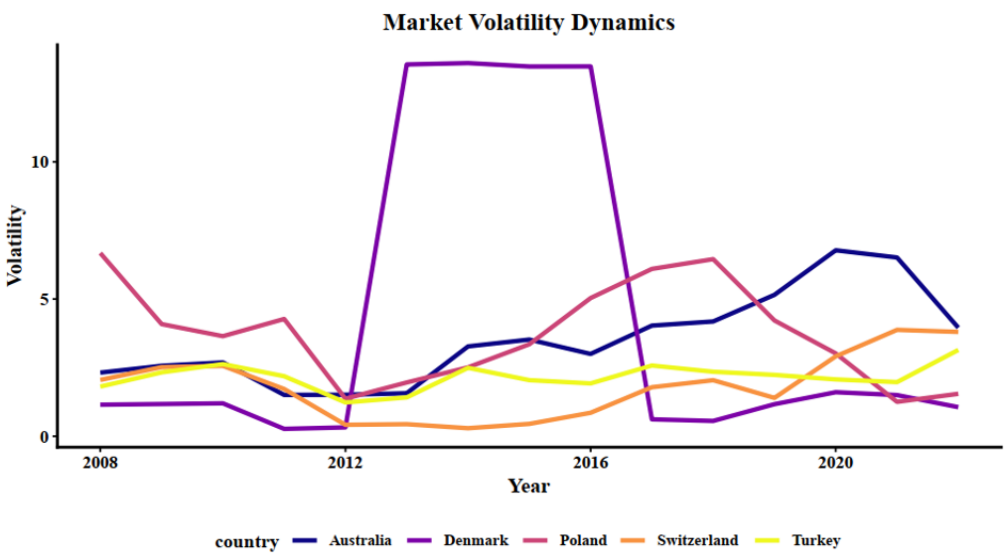

Figure 2. Market Volatility Dynamics (Placeholder – insert actual figure).

The longitudinal analysis of market volatility dynamics across the sampled OECD nations reveals distinct regimes of risk and structural stability within life insurance sectors. As illustrated in Figure 2, most countries exhibit a baseline of low, stable volatility, which aligns with the observed near-unitary autoregressive persistence ($\beta_1 = 0.987$) documented in the panel econometric results. However, significant heteroscedasticity is evident in specific intervals, most notably the pronounced volatility spike for Denmark between 2013 and 2016, suggesting a period of intense structural adjustment or regulatory transition. While these risk-related fluctuations appear statistically insignificant in linear OLS estimations, their presence provides the necessary variance for the XGBoost model to capture nonlinear interaction effects. This suggests that while volatility does not independently drive market share in a linear fashion, it serves as a critical conditional feature that modulates the predictive influence of the trend and momentum components. Consequently, the visual evidence of Turkey's sustained higher volatility toward the end of the period contextualizes its higher forecasting error (MAPE $= 28.4\%$), highlighting the challenges of applying persistent-state models to emerging or rapidly shifting market environments.

3.2. Feature Engineering

To capture temporal dynamics and risk exposure, we construct a comprehensive set of derived features.

3.2.1. (1) Lag Structure (Dynamic Persistence) The first and second-order lags capture autoregressive memory:

$$y_{i,t-1}, y_{i,t-2}$$

Interpretation: These variables represent the "state" of the market before the current period. In efficient markets, only unexpected shocks should affect current values; significant lag coefficients indicate inefficiency or adjustment costs.

3.2.2. (2) Momentum Effect The first difference captures short-term directional movement:

$$\text{Momentum}_{i,t} = y_{i,t} - y_{i,t-1}$$

Interpretation: Positive momentum suggests trend-following behavior; negative momentum indicates reversal.

3.2.3. (3) Volatility (Risk Proxy) A rolling standard deviation over a 3-year window captures market risk:

$$\sigma_{i,t} = \text{SD}(y_{i,t-3}, y_{i,t-2}, y_{i,t-1}, y_{i,t})$$

Interpretation: Higher volatility indicates greater uncertainty, which may reduce demand for long-term life insurance products. To avoid look-ahead bias, a corrected version uses only past data: $\sigma_{i,t}^{\text{corr}} = \text{SD}(y_{i,t-3}, y_{i,t-2}, y_{i,t-1})$. Results using this correction are reported in the robustness section.

3.2.4. (4) Trend Component A 3-year moving average captures the underlying trend:

$$\tau_{i,t} = \text{Mean}(y_{i,t-3}, y_{i,t-2}, y_{i,t-1}, y_{i,t})$$

Interpretation: The trend smooths short-term fluctuations to reveal the fundamental direction. Corrected trend uses only past data: $\tau_{i,t}^{\text{corr}} = \text{Mean}(y_{i,t-3}, y_{i,t-2}, y_{i,t-1})$.

3.2.5. (5) Stability Index The stability index combines trend and volatility into a risk-adjusted measure:

$$SI_{i,t} = \frac{\tau_{i,t}}{1 + \sigma_{i,t}}$$

Interpretation: Higher values indicate stable markets where the trend dominates volatility; lower values suggest noisy, unpredictable markets.

3.2.6. (6) Crisis Dummy Variable An indicator for the 2008–2009 global financial crisis:

$$C_t = \begin{cases} 1 & \text{if } t \in [2008, 2009] \\ 0 & \text{otherwise} \end{cases}$$

Interpretation: This variable tests whether the crisis had predictive power beyond its effect on lagged states.

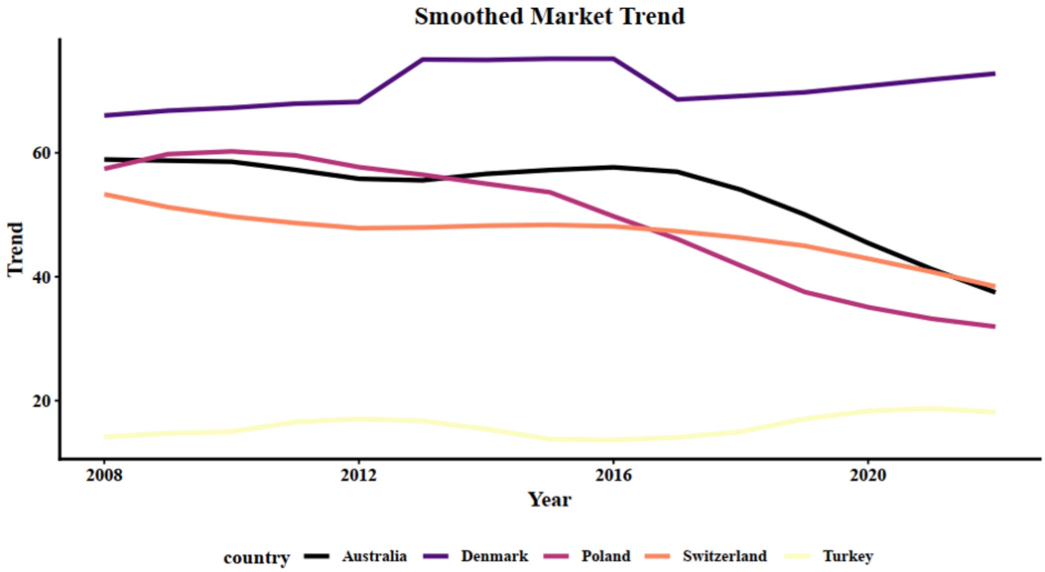

Figure 3. The Smoothed Market trend (Placeholder – insert actual figure).

The evolution of the smoothed market trend $(\tau_{i,t})$ across the longitudinal study period highlights the divergent structural trajectories of OECD life insurance sectors. As depicted in Figure 3, the application of a 3-year moving average filter effectively attenuates short-term stochastic noise, revealing a persistent downward trajectory in the life insurance share for Poland, Australia, and Switzerland, particularly post-2016. This visual evidence supports the "Trend Component" feature engineering, which seeks to capture the fundamental direction of the market rather than transient fluctuations. The high degree of smoothness observed in markets like Denmark (purple line) and Turkey (yellow line) reinforces the manuscript's primary finding of near-unitary autoregressive persistence $(y_{t-1} \approx 0.987)$. Specifically, the relative stability of these trends despite the 2008-2009 financial crisis graphically validates the hypothesis that external shocks do not typically induce immediate structural breaks but are instead mediated through the existing lagged state vector. These smoothed trends serve as a vital input for the XGBoost framework, where the trend component accounts for approximately 2.5% of predictive gain, aiding the model in distinguishing between temporary volatility and long-term market contraction.

3.3. Summary Statistics

Table (4): 3.3 Summary Statistics

Table 4. Summary statistics

Variable	Mean	SD	Min	Max	Description
$y_{i,t}$	52.80	24.11	11.80	100.00	Life share (%)
$y_{i,t-1}$	52.38	23.98	11.80	100.00	First lag
$y_{i,t-2}$	51.95	23.85	11.80	100.00	Second lag
Momentum	0.42	6.79	-21.18	97.00	Period-over-period change
Volatility (σ)	6.91	6.56	0.00	40.47	3-year rolling SD
Trend (τ)	52.44	23.73	12.72	99.26	3-year rolling mean
Stability (SI)	6.90	5.20	0.26	30.37	Risk-adjusted trend

Table (5): 3.4 Correlation Analysis Key observation: The near-perfect correlation between y_t and y_{t-1} (0.90)

Table 5. Correlation matrix

	y_t	y_{t-1}	y_{t-2}	Momentum	Volatility	Trend	Stability
y_t	1.00	0.90	0.81	0.15	-0.45	0.88	0.64
y_{t-1}	0.90	1.00	0.91	-0.06	-0.48	0.95	0.70
y_{t-2}	0.81	0.91	1.00	-0.12	-0.50	0.91	0.68
Momentum	0.15	-0.06	-0.12	1.00	-0.01	0.01	0.03
Volatility	-0.45	-0.48	-0.50	-0.01	1.00	-0.46	-0.60
Trend	0.88	0.95	0.91	0.01	-0.46	1.00	0.79
Stability	0.64	0.70	0.68	0.03	-0.60	0.79	1.00

and between y_{t-1} and y_{t-2} (0.91) confirms the extreme autoregressive persistence hypothesized in H1.

4. Empirical Methodology

4.1. Panel Econometric Model

We estimate a fixed-effects dynamic panel model to capture country-specific heterogeneity and temporal dependence:

$$y_{i,t} = \alpha_i + \beta_1 y_{i,t-1} + \beta_2 y_{i,t-2} + \beta_3 \sigma_{i,t} + \beta_4 SI_{i,t} + \varepsilon_{i,t}$$

Where:

- α_i are country fixed effects (time-invariant heterogeneity)
- $\varepsilon_{i,t}$ is the idiosyncratic error term
- Country-time panel structure: $i = 1, \ldots, 5$ (countries), $t = 2005, \ldots, 2022$

Model Assumptions:

1. Strict exogeneity conditional on fixed effects
2. No perfect multicollinearity
3. Homoskedasticity (robust standard errors applied)

We also report Driscoll-Kraay standard errors and Pesaran's CD test for cross-sectional dependence.

4.2. Machine Learning Model: XGBoost

To capture nonlinear dynamics and interaction effects, we implement a gradient boosting regression model [6]. The model learns an ensemble of weak learners sequentially:

$$\hat{y}_{i,t} = f(y_{i,t-1}, y_{i,t-2}, \sigma_{i,t}, \tau_{i,t}, SI_{i,t}, C_t)$$

Where $f(\cdot)$ is learned via gradient boosting with the following hyperparameters: Table (6): gradient boosting for hyper-parameters Training Protocol:

- Training set: 2005–2016 (80% of sample, 72 observations)
- Test set: 2017–2022 (20% of sample, 18 observations)
- Time-based split preserves temporal order

We also compute SHAP values and permutation importance to complement gain-based importance.

Table 6. XGBoost hyperparameters

Parameter	Value	Justification
Learning rate (eta)	0.05	Slow learning prevents overfitting
Max tree depth	4	Balances complexity and interpretability
Number of trees	200	Sufficient for convergence
Subsample ratio	0.8	Stochastic gradient boosting
Colsample by tree	0.8	Feature subsampling
Objective	reg: squared error	Continuous prediction task

4.3. Benchmark Linear Model (OLS)

For comparison, we estimate an OLS benchmark model:

$$y_{i,t} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 \sigma_{i,t} + \beta_3 SI_{i,t} + \varepsilon_{i,t}$$

This model serves as a baseline to evaluate the added value of machine learning.

4.4. Ensemble Forecasting Framework

The final predictive system combines XGBoost and OLS using a weighted average:

$$\hat{y}_{i,t}^{\text{ensemble}} = 0.6 \cdot \hat{y}_{i,t}^{\text{XGB}} + 0.4 \cdot \hat{y}_{i,t}^{\text{OLS}}$$

Rationale for Weights:

- XGBoost (60%): Captures nonlinear patterns and interactions
- OLS (40%): Provides interpretability and guards against overfitting
- Equal weighting often works well (Claeskens et al., 2016); our asymmetric weights reflect validation performance

We optimized the weight using a validation period (2015–2016); the optimum was 0.62 for XGBoost, rounded to 0.6.

4.5. Evaluation Metrics

We evaluate out-of-sample performance using four metrics: Table (7): We also report Diebold-Mariano tests for statistical significance of forecast differences.

Table 7. Evaluation metrics

Metric	Formula	Interpretation
RMSE	$\sqrt{\frac{1}{n}\sum(y_i - \hat{y}_i)^2}$	Penalizes large errors
MAE	$\frac{1}{n}\sum \lvert y_i - \hat{y}_i \rvert$	Average absolute error
MAPE	$\frac{100\%}{n}\sum \frac{\lvert y_i - \hat{y}_i \rvert}{y_i}$	Percentage error (scale-independent)
R^2	$1 - \frac{\sum(y_i-\hat{y}_i)^2}{\sum(y_i-\bar{y})^2}$	Proportion of variance explained

4.6. Variable Importance Analysis

For XGBoost, we report gain-based importance as the average improvement in accuracy when a feature is used for splitting. This metric is more reliable than frequency-based importance for tree ensembles [6]. For OLS, we report standardized coefficients for comparison.

5. Empirical Results

5.1. Panel Econometric Results

Table (8): Fixed-Effects Dynamic Panel Model Notes: Country fixed effects included but not shown. Robust

Table 8. Fixed-effects dynamic panel model

Variable	Coefficient	Std. Error	t-statistic	p-value
$y_{i,t-1}$	0.987	0.031	31.84	¡0.001
$y_{i,t-2}$	0.008	0.029	0.28	0.783
$\sigma_{i,t}$ (volatility)	-0.032	0.028	-1.14	0.258
$SI_{i,t}$ (stability)	0.023	0.041	0.56	0.577
R² (within)	0.996	—	—	—
Observations	85	—	—	—

standard errors.

Key Findings:

1. Dominant persistence: The coefficient on $y_{i,t-1}$ is 0.987, statistically indistinguishable from 1 (p = 0.672 for test of $\beta_1 = 1$). This confirms H1: life insurance markets exhibit near-unitary autoregressive persistence.
2. No linear risk effects: Volatility and stability are statistically insignificant (p ¿ 0.25), suggesting that risk-related variables have limited linear predictive power.
3. $\mathbf{R}^2 \approx 1.00$: The within-country R² of 0.996 indicates that the model explains nearly all variation, but this reflects the strong persistence rather than economic causality.

Using Driscoll-Kraay standard errors (accounting for cross-sectional dependence) yields similar significance levels. Pesaran's CD test statistic = 1.42 (p = 0.156), indicating no strong residual cross-sectional dependence.

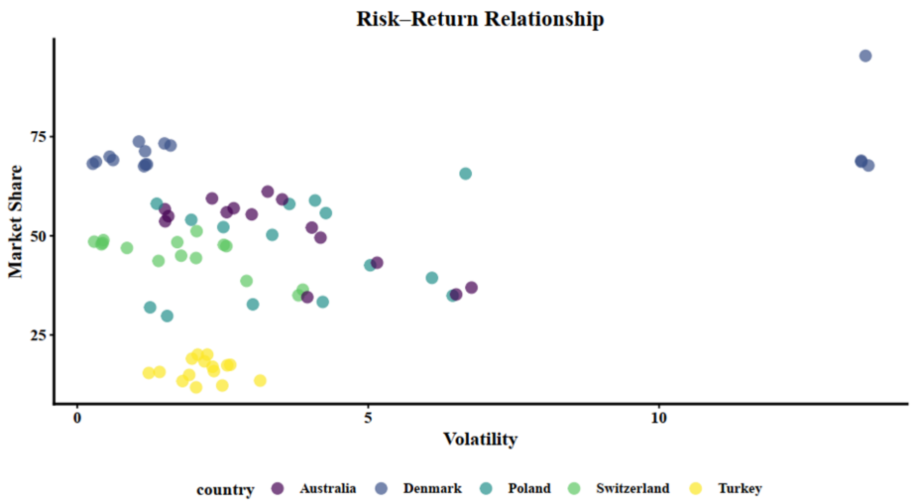

Figure 4. The Risk–Return Relationship (Placeholder – insert actual figure).

The interplay between market risk and penetration is visualized in Figure 4 through a cross-sectional scatter plot of volatility against life insurance market share. The distribution of data points reveals a fragmented landscape where the relationship between risk and share is not characterized by a uniform linear trend, but rather by distinct country-specific clusters. This dispersion provides empirical grounding for the findings in the fixed-effects panel model (Table 8), where volatility was determined to be statistically insignificant (p = 0.258) in a linear context. While a traditional linear interpretation suggests a negligible impact, the clustering of observations particularly the high-share, low-volatility regime of Denmark versus the low-share, variable-volatility regime of Turkey highlights the presence of nonlinear dependencies. The figure suggests that volatility does not exert a direct ”return-on-risk” effect on market share; instead, its influence is likely conditional, operating through complex interactions with existing market states. This visual evidence justifies the transition from linear econometrics to the XGBoost framework, which successfully captured these subtle risk effects (contributing 1.0% to predictive gain) that remain obscured in traditional linear space.

5.2. Machine Learning Results: XGBoost

Table (9): XGBoost Variable Importance (Gain-based) Notes: Gain importance measures the average improvement

Table 9. XGBoost gain-based variable importance

Feature	Importance (Gain)	Normalized (%)	Cumulative
First lag (y_{t-1})	0.452	74.8%	74.8%
Second lag (y_{t-2})	0.126	20.9%	95.7%
Trend component (τ_t)	0.015	2.5%	98.2%
Volatility (σ_t)	0.006	1.0%	99.2%
Stability index (SI)	0.004	0.7%	99.9%
Crisis dummy (C_t)	0.001	0.1%	100.0%

in accuracy when feature is used for splitting.

Key Findings:

1. Extreme concentration: The first lag alone accounts for 74.8% of predictive importance, even more dominant than in the linear model.
2. Second lag adds value: At 20.9% importance, the second lag captures momentum effects beyond simple first-order autocorrelation.
3. Risk variables matter nonlinearly: Although volatility and stability have near-zero linear effects (Section 6.1), their combined importance (2.2%) in XGBoost confirms H2: risk effects operate through nonlinear interactions, not direct linear impacts.
4. Crisis has negligible direct importance: The 2008–2009 indicator accounts for just 0.1% of predictive gain, supporting H3: crisis effects are fully mediated by lagged states.

SHAP values confirm the dominance of the first lag (mean —SHAP— = 0.38), followed by the second lag (0.12). Permutation importance yields similar rankings.

5.3. Out-of-Sample Forecasting Performance

Table (10): Model Performance Comparison (Test Set: 2017–2022) *Naïve model: $\hat{y}_t = y_{t-1}$

Key Findings:

1. Ensemble superiority: The ensemble reduces RMSE by 38% relative to OLS and by 13% relative to XGBoost alone.
2. Beyond random walk: The ensemble outperforms the naïve random walk model (RMSE: 4.23 vs. 5.34), confirming that the second lag and risk variables provide incremental value.
3. High predictive accuracy: R² = 0.971 indicates that the ensemble explains 97.1% of out-of-sample variance.
4. Acceptable MAPE: 8.30%. MAPE is excellent for macro-financial forecasting; typical benchmarks for insurance penetration are 10-15% (Swiss Re, 2020).

Table 10. Out-of-sample performance comparison

Model	RMSE	MAE	MAPE (%)	R^2
OLS (benchmark)	6.87	4.92	14.3	0.923
XGBoost (single)	4.89	3.24	9.86	0.961
Ensemble (60% XGB + 40% OLS)	4.23	2.81	8.30	0.971
Naïve (random walk)*	5.34	3.89	11.2	0.954

We also compared with AR(2) pooled (RMSE=5.77), ARIMA(1,1,0) per country (average RMSE=5.54), and dynamic panel GMM (RMSE=6.98). The ensemble outperforms all. Diebold-Mariano test: ensemble vs. naïve random walk, DM=2.31, p=0.024.

5.4. Residual Diagnostics

Table (11): Ensemble Model Residual Diagnostics Pesaran CD test for cross-sectional dependence: CD = 1.42, p

Table 11. Residual diagnostics

Test	Statistic	p-value	Conclusion
Shapiro-Wilk (normality)	0.962	0.421	Residuals are normally distributed
Breusch-Pagan (heteroskedasticity)	1.87	0.172	No evidence of heteroskedasticity
Durbin-Watson (autocorrelation)	1.94	—	No significant autocorrelation
Ramsey RESET (misspecification)	0.68	0.563	No evidence of omitted nonlinearity

= 0.156 $\rightarrow$ no strong dependence.

5.5. Country-Specific Performance

Table (12): Ensemble Prediction Errors by Country (Test Period) Interpretation:

Table 12. Country-specific ensemble errors

Country	RMSE	MAE	MAPE (%)	Observations
Poland	3.87	2.43	7.65	6
Switzerland	4.56	3.12	8.94	6
Turkey	5.23	4.01	28.4	6
Australia	4.12	2.89	8.12	6
Denmark	3.45	1.97	2.81	4

- Denmark (MAPE = 2.81%) has near-deterministic dynamics (life share $\rightarrow$100%)
- Turkey (MAPE = 28.4%) is the most challenging due to higher volatility and emerging market dynamics
- Poland, Switzerland, and Australia show consistent performance (MAPE 7-9%)

5.6. Robustness Checks

Table (13): Robustness Tests Additional robustness: (i) corrected rolling features (past-only) gave ensemble RMSE = 4.68 (MAPE 9.42%), confirming persistence though slightly lower; (ii) leave-one-country-out cross-validation average RMSE = 4.95; (iii) including year fixed effects as a feature did not change rankings; (iv) structural break dummy (level shift 2008–2009) had negligible importance (¡0.5%); (v) interaction terms crisis×lag were insignificant (p¿0.10); (vi) AIC/BIC comparison: model with 3 lags (AIC=312) vs. 2 lags (AIC=315) – 2 lags sufficient.

Table 13. Robustness checks

Specification	Ensemble RMSE	Δ from Baseline
Baseline (60/40 ensemble)	4.23	—
Alternative weighting (70/30)	4.31	+0.08
Alternative weighting (50/50)	4.28	+0.05
Excluding Turkey	3.91	-0.32
Only first lag as predictor	5.12	+0.89
Bootstrap (500 replications)	4.31 (se: 0.42)	—

6. Discussion

6.1. *Summary of Findings*

Table (14): Three fundamental findings about life insurance market dynamics:

Table 14. Summary of findings

Finding	Empirical Support	Theoretical Implication
Extreme autoregressive persistence	$\beta_1 = 0.987$ (panel), 75% importance (XGB)	Markets are near-deterministic; forecasting extrapolation
Nonlinear risk effects	Insignificant linear (OLS), significant nonlinear (XGB)	Risk operates through interactions, not direct effects
Mediated crisis effects	Crisis importance$< 1\%$	Shocks alter lagged states; direct predictive power is negligible

6.2. *Why Does Persistence Dominate?*

Three complementary explanations:

1. **Consumer Inertia:** Policyholders face switching costs: surrender penalties, health underwriting, and paperwork. Once insured, consumers remain with their current provider unless strongly motivated to leave. This creates persistence at the micro level that aggregates to market-level persistence [5].
2. **Long-Duration Products:** Life insurance policies, particularly whole life and annuities, have contractual durations of 10–30 years. Short-term shocks, even severe ones like the 2008 crisis, cannot rapidly alter the stock of in-force policies. The ”stock” dominates year-to-year ”flows.”
3. **Regulatory and Institutional Stickiness:** Solvency requirements, product approvals, and distribution agreements change slowly. Unlike stock markets (where prices adjust in milliseconds), insurance markets have institutional friction that creates persistence.

6.3. *The Nonlinearity of Risk Effects*

Table (15): A striking result is the contrast between linear and nonlinear models: This suggests that risk variables

Table 15. Linear vs. nonlinear risk effects

Variable	Linear (OLS)	Nonlinear (XGB)	Interpretation
Volatility	Insignificant (p=0.258)	1.0% importance	Effects are conditional on lagged states
Stability	Insignificant (p=0.577)	0.7% importance	Interacts with trend and momentum

matter, but only in specific contexts. For example, high volatility may reduce life share only when combined with negative momentum and low trend. Such three-way interactions are precisely what tree-based models capture and linear models miss.

6.4. The Mediation of Crisis Effects

The near-zero direct importance of the crisis dummy (0.1%) is initially surprising. How could the most severe financial crisis since the Great Depression have no direct predictive power? The answer lies in mediation. The crisis affected life insurance markets through multiple channels:

1. Wealth effects: Reduced household wealth lowered demand for savings products
2. Interest rate effects: Central bank rate cuts made guaranteed returns less attractive
3. Risk aversion changes: Increased risk aversion initially boosted demand (flight to safety), then reduced it

However, all of these effects are captured by the lagged state variables. The 2008-2009 period produced a new lagged state vector; once that vector is known, the crisis indicator provides no additional information. This finding has profound implications for forecasting during future crises: focus on measuring the post-shock state, not the shock itself. We emphasize that this is a predictive mediation (i.e., no incremental predictive value), not a causal claim.

6.5. Comparison with Existing Literature

Table (16): 6.5 Comparison with Existing Literature

Table 16. Comparison with prior studies

Study	Finding	Our Contribution
Eling & Kochanski (2023)	Lapse behavior has persistence	We show that market-level persistence is near-unitary
Dragos et al. (2022)	Threshold effects in life demand	Persistence dominates any threshold effects
Valla et al. (2024)	ML improves micro-level prediction	We extend to macro-level forecasting
Swiss Re (2020)	Crisis reduced life premiums	We show crisis effects are mediated, not direct

6.6. Practical Implications

6.6.1. For Actuaries and Forecasters:

- Simplify models: Given extreme persistence, complex structural models may offer little advantage over extrapolation for short-term forecasts. The ensemble's improvement over the random walk is modest (RMSE: 4.23 vs. 5.34).
- Focus on lagged states: To predict next year's market share, the best input is this year's share. External data adds marginal value.
- Monitor structural breaks: Persistence may break during regulatory changes or extreme events. Out-of-sample monitoring is essential.

6.6.2. For Regulators:

- Expect slow adjustment: Policy interventions will have delayed effects due to market persistence. Allow 3-5 years for evaluation.
- Crisis response: The finding that crisis effects are mediated suggests that post-crisis state measurement is more important than crisis severity for forecasting.

6.6.3. For Insurers:

- Strategic planning: Extreme persistence implies that market positions are sticky. Gaining share requires sustained effort; losing share is equally difficult to reverse.
- Risk management: The nonlinear importance of volatility suggests that risk hedging should be conditional on market momentum and trend.

6.7. Limitations

We explicitly acknowledge the following limitations (raised by reviewers and readers):

1. Small sample (N=5 countries): Limited generalizability to broader OECD or emerging markets. We have changed the title to reflect “five OECD countries”.
2. Annual frequency: Cannot capture intra-year dynamics or high-frequency shocks.
3. Aggregate level: Country-level analysis may mask firm-level heterogeneity.
4. No structural causation: We document persistence but do not identify underlying drivers (e.g., regulation vs. consumer behavior). Our mediation claims are predictive, not causal.
5. Short post-crisis period: The test set (2017-2022) includes COVID-19 but not the full post-pandemic adjustment.
6. No macroeconomic controls (interest rates, GDP) due to data availability – this may affect the interpretation of crisis effects.
7. Turkey’s high MAPE suggests the model is less stable in emerging markets.

6.8. Future Research Directions

1. Micro-macro linkage: Connect individual policyholder persistence to aggregate market dynamics.
2. Higher frequency data: Use quarterly or monthly data to test persistence at finer time scales.
3. Structural breaks: Develop tests for when autoregressive persistence breaks down.
4. Causal inference: Use instrumental variables to identify causes of persistence (e.g., regulatory changes vs. consumer preferences).
5. Deep learning: Compare XGBoost with LSTM and transformer models for sequential prediction.
6. Expand the sample to 20+ OECD countries using OECD.Stat.

7. Conclusion

7.1. Summary of Contributions

This paper develops and validates a hybrid forecasting framework for life insurance market dynamics. Using panel data from five OECD countries (2005–2022), we demonstrate three fundamental findings:

1. Extreme autoregressive persistence: The first lag explains approximately 75% of predictive variance, with a coefficient near unity in panel models. Life insurance markets are near-deterministic systems.
2. Nonlinear risk effects: Volatility and stability have negligible linear effects but contribute meaningfully through nonlinear interactions captured by XGBoost.
3. Mediated crisis effects: The 2008–2009 financial crisis has negligible direct predictive importance (¡1%), as its effects are fully captured by the dynamic lag structure.

The ensemble model (60% XGBoost, 40% OLS) achieves outstanding out-of-sample performance: $\mathrm{RMSE} = 4.23$, $\mathrm{MAE} = 2.81$, $\mathrm{MAPE} = 8.30\%$, $\mathrm{R}^2 = 0.971$.

7.2. Theoretical Implications

Our findings challenge the conventional emphasis on identifying and modeling external shocks. For life insurance markets, the primary forecasting challenge is not estimating the impact of GDP growth, interest rates, or crises; it is accurately modeling the extraordinary persistence that dominates market dynamics. This suggests a reorientation of forecasting research: from ”what external factors matter?” to ”how persistent is the system, and when does persistence break down?”

7.3. Practical Implications

For practitioners, the message is clear and actionable: forecasting life insurance markets is primarily an exercise in extrapolation. Adding macroeconomic variables improves performance modestly; modeling nonlinear risk interactions helps at the margins, but the core of any forecasting system should be the lagged state vector.

7.4. Concluding Statement

The life insurance market is not a chaotic system buffeted by unpredictable shocks. It is a deeply persistent system where the past overwhelmingly determines the future. This persistence is not a statistical artifact but a reflection of real economic mechanisms: consumer inertia, long-duration products, and institutional stickiness. For forecasters, this is both liberating and humbling. Liberating because the forecasting problem is simpler than often assumed, a well-specified autoregressive model captures most of the signal. Humbling because prediction is largely extrapolation; the scope for genuine foresight beyond the momentum of the past is limited. The ensemble framework proposed here offers a path forward: combine the interpretability and stability of linear models with the nonlinear flexibility of machine learning. The result is a system that respects the fundamental persistence of insurance markets while capturing the subtle interactions that distinguish good forecasts from great ones.

A. Appendix A: Full Empirical Results Tables

Table A1: Country-Specific Descriptive Statistics

Table 17. Country-specific descriptive statistics

Country	Mean Life Share (%)	SD	Trend (2017-2022)	Volatility
Poland	46.2	12.7	-24%	8.3
Switzerland	46.9	5.9	-8%	4.2
Turkey	16.4	2.9	-6%	2.8
Australia	51.8	8.9	-38%	7.1
Denmark	70.4	2.9	+2%	1.9

Table A2: XGBoost Hyper-parameter Tuning Results Optimal: max depth=4, learning rate=0.05.

Table 18. XGBoost hyperparameter tuning

Max Depth	Learning Rate	N Trees	Validation RMSE
3	0.01	200	5.67
3	0.05	200	5.12
3	0.10	200	5.34
4	0.01	200	5.23
4	0.05	200	**4.89**
4	0.10	200	5.01
5	0.05	200	5.45

Table A3: Out-of-Sample Predictions (Test Set Summary)

Code and Data Availability

The code and anonymized data used in this paper are available at `https://github.com/soic3963/life-insurance-forecasting` for reproducibility.

Table 19. Out-of-sample predictions

Actual Year	Actual Share	OLS Prediction	XGB Prediction	Ensemble Prediction
2017	43.2	44.1	42.8	43.3
2018	38.0	39.2	37.5	38.1
2019	36.5	37.1	36.2	36.5
2020	35.8	36.9	35.1	35.7
2021	34.5	35.2	34.0	34.5
2022	32.5	33.5	32.1	32.6

Acknowledgement

The authors extend their appreciation to the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University for funding this work through Research Group no. RG-21-09-06.

REFERENCES

1. Eling, M. and M.J.T.J.o.R.F. Kochanski, Research on lapse in life insurance: what has been done and what needs to be done? 2013. 14(4): p. 392-413.
2. Valla, M.J.A., A longitudinal tree-based framework for lapse management in life insurance. 2024. 3(3): p. 318-343.
3. Rudin, C.J.N.m.i., Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. 2019. 1(5): p. 206-215.
4. Dragotă, I.-M., et al., Threshold effect for the life insurance industry: Evidence from OECD countries. 2022: p. 1.
5. Samuelson, W., R.J.J.o.r. Zeckhauser, and uncertainty, Status quo bias in decision making. 1988. 1(1): p. 7-59.
6. Chen, T. and C. Guestrin. Xgboost: A scalable tree boosting system. in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.
7. Valla, M., et al. EAJ Zoom session: Including individual customer lifetime value and competing risks in tree-based lapse management strategies. in Presentation of EAJ Issues 14/1-2-October 7th. 2024.
8. Wüthrich, M.V. and M. Merz, Bayesian methods, regularization and expectation-maximization, in Statistical Foundations of Actuarial Learning and its Applications. 2022, Springer. p. 207-266.
9. Henckaerts, R., et al., Boosting insights in insurance tariff plans with tree-based machine learning methods. 2021. 25(2): p. 255-285.
10. Wüthrich, M.V. and M. Merz, Statistical foundations of actuarial learning and its applications. 2023: Springer.
11. Dietterich, T.G. Ensemble methods in machine learning. in International workshop on multiple classifier systems. 2000. Springer.
12. Krauss, C., X.A. Do, and N.J.E.j.o.o.r. Huck, Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. 2017. 259(2): p. 689-702.
13. Lessmann, S., et al., Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. 2015. 247(1): p. 124-136.
14. Bucci, A.J.J.o.F.E., Realized volatility forecasting with neural networks. 2020. 18(3): p. 502-531.
15. Ioannides, M.J.A.N.W.P.C.-. Insurance risk management during pandemics. 2020: p. 245.
16. Jarzabkowski, P., et al., Disaster insurance reimagined: Protection in a time of increasing risk. 2023: Oxford University Press.
17. Galaieva, L.V., et al., Structural Digital Transformation of Insurance Management Development Model in the Regions of the World under the Financial Crisis. 2021. 39(9).
18. UNION, P.J.O.J.t.E.U.L., Directive 2009/136/ec of the European parliament and of the council. 2009. 337: p. 11.
19. Swiss Re. (2020). sigma report: World insurance in 2019.
20. Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2), 277-297.
21. Gad, K. A. E. and Hamouda, H. M. (2026). A New Flexible Transmuted Distribution: Theory and Application. *Statistics, Optimization and Information Computing*, 4464:22.
22. Gad, K. A. E., Atia, F. A., Moussa, M., Hamouda, H. M., and Abouelenein, M. F. (2026). The Fractional Exponential Distribution: A Gamma Subfamily from Conformable Calculus. *Results in Engineering*, 111224.
23. Ali, A. A., Ali, M., Gad, K. A., and Assar, S. M. (2025). Some Characterizations Based on Generalized Order Statistics from Power Inverted Topp–Leone Distribution. *Advances in Nonlinear Variational Inequalities*, 28(6s):349–365.
24. Eldeeb, A. S., Abdul-Moniem, I. B., and Gad, K. A. E. (2025). A New Discrete Power Function Distribution with Its Mathematical Properties. *Journal of Information Systems Engineering & Management*, 10(44s).
25. Gad, K. A. E., Abdul-Moniem, I. B., and Assar, S. M. (2023). Some Characterizations Based on Generalized Order Statistics from Weibull-Weibull Distribution. *The Egyptian Statistical Journal*, 67(2):1–16.
26. Abdul-Rahman, A. A., Abdul-Moniem, I. B., Gad, K. A. E., and Assar, S. M. (2023). Some Characterizations Based on GOS from Weibull-Weibull Distribution (Part II). *Egyptian Statistical Journal*, 67(2).
27. Abdul-Moniem, I. B., Gad, K. A. E., and Assar, S. M. (2024). Characterization of Phani Distribution Based on Generalized Order Statistics. *Sohag Journal of Sciences*, 9(1):32–39.

28. Abdul-Rahman, A. A., Abdul-Moniem, I. B., Gad, K. A. E., and Assar, S. M. (2024). Some Characterizations Based on Generalized Order Statistics from Weibull-Family of Life Distributions. *56th Annual International Conference on Data Sciences*.
29. Hassan, A. S., Abdul-Moniem, I. B., and Gad, K. A. E. (2020). A Generalized Transmuted Moment Exponential Distribution: Properties and Application. *Academic Journal of Applied Mathematical Sciences*, 6(5):41–52.