Università
Ca'Foscari
Venezia

Master's Degree Programme

in Data Analytics for Business and Society

Final Thesis

Modelling Regime Shifts in Insurance Dynamics

A Latent Markov Approach with Macroeconomic Covariates

Supervisor

Prof. Gaia Bertarelli

Graduand

Alessia Pizzolla

Matriculation Number 891453

Academic Year

2024 / 2025

ABSTRACT

The European Life insurance market has recently faced challenges mainly derived from macroeconomic shocks, notably the 2022-2023 inflationary spike and rapid interest rate hikes. While traditional literature often examines these shocks using static panel regressions, this thesis introduces a dynamic Latent Markov Model (LMM) approach to capture unobserved heterogeneity and regime shifts within the sector. Utilizing a comprehensive quarterly panel dataset of 29 European countries from 2017 to 2025, the empirical analysis identifies a fundamental dual-state market structure ($\boldsymbol{k = 2}$): a "Resilient" core and a "Volatile/Liquidity Stress" regime. The baseline model reveals a significant structural break in 2023-2024, characterized by a mass migration of historically stable markets (e.g., Italy, Germany) into the volatile state due to severe liquidity pressures. By incorporating macroeconomic and socio-demographic covariates directly into the measurement model, this study isolates the exogenous drivers of this shift, demonstrating that the observed volatility was not a symptom of chronic operational inefficiency, but a mechanical reaction to sudden surrender risks driven by rising 10-Year Government Bond yields. Ultimately, the estimated transition probabilities highlight the sector's robust internal resilience, showing a higher probability of recovery (19.22%) than of falling into distress (9.07%). The findings provide critical policy implications for regulatory bodies and insurance management, emphasizing the need for dynamic liquidity risk monitoring.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to Professor Gaia Bertarelli, my supervisor, for her precious guidance, providing me with resources and support. I really appreciate the understanding she expressed to me in the last months, also considering the challenging times I had to face and my path as a working student.

I want to thank my family, especially my parents, for always being by my side, even if $\boldsymbol{n}$ kilometers (where $\mathbf{600} \leq \boldsymbol{n} < \mathbf{1000}$) physically separate us. They're always there to help me in a practical way (e.g. the many relocations I have had to face in the latest years) as well as in exchanging ideas and giving emotional support.

Lastly, I want to acknowledge my few lifelong friends that are still by my side and proving that distance is not an obstacle for us. I'm also grateful to the friends and former colleagues I've met in this beautiful region, that welcomed me almost five years ago; thank you for encouraging me and for all the memories we shared.

I don't know where this journey is leading me after this, but I'll always carry all the experiences and people with me.

Contents

ABSTRACT 3

ACKNOWLEDGEMENTS 4

TABLES 7

FIGURES 7

1 THE ROLE OF INSURANCE IN EUROPE AND REGULATORY FRAMEWORK 8

1.1 THE ROLE OF INSURANCE IN THE ECONOMY 8
1.2 THE MECHANISMS UNDERNEATH INSURANCE CONTRACTS 9
1.3 THE BENEFITS OF INSURANCE 11
1.3.1 BUSINESS MODELS AND A DEEP-DIVE INTO LIFE INSURANCE BUSINESS 12
1.4 THE SOLVENCY II REGIME 17
1.4.1 PILLAR 1: QUANTITATIVE REQUIREMENTS AND CAPITAL ADEQUACY 17
1.4.2 PILLAR 2: GOVERNANCE AND RISK MANAGEMENT 18
1.4.3 PILLAR 3: DISCLOSURE AND MARKET DISCIPLINE 18
1.4.4 STRATEGIC IMPLICATIONS FOR LIFE INSURANCE BUSINESS 19
1.5 AIMS AND OBJECTIVES 20
1.6 THE LITERATURE GAP AND METHODOLOGICAL INNOVATION 22

2 DATA DESCRIPTION AND EXPLORATORY DATA ANALYSIS (EDA) 23

2.1 DATA SOURCE 23
2.2 RESPONSE VARIABLES DEFINITION 25
2.3 COVARIATES DEFINITION 27
2.4 EXPLORATORY DATA ANALYSIS 30
2.4.1 DATASET STRUCTURE AND PRE-PROCESSING 30
2.4.2 DESCRIPTIVE STATISTICS 33
2.4.3 CORRELATION ANALYSIS 37
2.4.4 TEMPORAL EVOLUTION AND SEASONALITY 39
2.4.5 SUMMARY OF EDA FINDINGS 44

3 METHODOLOGY: THE LATENT MARKOV FRAMEWORK 46

3.1 INTRODUCTION TO MARKOV CHAINS AND HIDDEN MARKOV MODELS 46
3.2 LATENT MARKOV MODEL (LMM) FOR PANEL DATA 48
3.3 FOCUS ON THE ROLE OF COVARIATES IN LMM 50
3.4 ESTIMATION AND MODEL SELECTION 52

4 RESULTS 54

4.1 MODEL WITHOUT COVARIATES **54**
4.1.1 MODEL SELECTION 54
4.1.2 ECONOMIC INTERPRETATION 55
4.2 MODEL WITH COVARIATES **63**
4.2.1 MODEL SELECTION 63
4.2.2 ECONOMIC INTERPRETATION 65
4.2.3 REGRESSION COEFFICIENTS 68
4.3 MODEL COMPARISON **69**

CONCLUSION **70**

REFERENCES **73**

APPENDIX **78**

R CODE **78**

TABLES

Table 2.4.1-1: Dataset Structure. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 31
Table 2.4.2-1: Descriptive statistics. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 33
Table 2.4.2-2: Descriptive statistics for Italy. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 35
Table 2.4.2-3: Descriptive statistics for Slovakia. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 36
Table 4.1.1-1: Information criteria comparison. Author's calculation on EIOPA (2025) data. 54
Table 4.1.2-1: Initial probabilities. Author's calculation on EIOPA (2025) data. 55
Table 4.1.2-2: Transition probability matrix for model with k = 2. Author's calculation on EIOPA (2025) data. 57
Table 4.1.2-3: Frequency of Latent State Occupancy by country. Author's calculation on EIOPA (2025) data. 61
Table 4.2.1-1: Information criteria comparison (model with covariates). Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 63
Table 4.2.2-1: Latent States intercepts (model with covariates). Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 65
Table 4.2.2-2: Initial probabilities (model with covariates). Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 66
Table 4.2.2-3: Transition Probability Matrix (model with covariates). Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 67
Table 4.2.3-1: Regression Coefficients. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 68

FIGURES

Figure 1.5.1-1: representation of the Loss Function (source: ScienceDirect, https://www.sciencedirect.com/topics/mathematics/loss-distribution) 13
Figure 2.4.3-1: Correlation matrix. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 37
Figure 2.4.4-1: Temporal Evolution of Insurance Response Variables. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 40
Figure 2.4.4-2: Macroeconomic Shocks vs Insurance Responses. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 41
Figure 2.4.4-3: Time Series Decomposition of Premium Growth. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 43
Figure 2.4.4-4: Time Series Decomposition of Expense Ratio. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data. 44
Figure 4.1.2-1: Standardized Mean Values for responses of model with k = 2. Author's calculation on EIOPA (2025) data. 56
Figure 4.1.2-2: Number of countries in Latent States.. Author's calculation on EIOPA (2025) data. 58
Figure 4.1.2-3: Individual country trajectories across latent states (2017-2025). Author's calculation on EIOPA (2025) data. 59

1 THE ROLE OF INSURANCE IN EUROPE AND REGULATORY FRAMEWORK

1.1 THE ROLE OF INSURANCE IN THE ECONOMY

The insurance sector has recently been at the center of significant financial and public discussion. Using the Italian market as an example, the industry has been brought into the spotlight: from significant corporate finance plans, such as the strategic operations involving Banca Generali (Banca Generali 2025), to the growing necessities of managing natural disaster risks following extreme climate events (Osservatorio Nazionale Città Clima 2025). Furthermore, the regulatory landscape has evolved significantly, as evidenced by the implementation of the "Gelli-Bianco Law", which reformed professional liability in the healthcare sector (Ministero delle Imprese e del Made in Italy 2025), and the technological disruption introduced by Internet of Things and "black boxes" in the automotive industry (Jeniot 2025, AtenaBrokers 2025).

Despite these events, the fundamental nature and utility of insurance are not always fully understood by the public. Insurance is, in fact, fundamental for risk transfer: it allows an individual or an entity (the insured) to transfer the financial burden of a potential loss to a specialized company (the insurer). By paying a determined, relatively small amount (premium), the policyholder protects themselves against the possibility of a much larger, unpredictable financial loss (Insurance Europe, 2012).

1.2 THE MECHANISMS UNDERNEATH INSURANCE CONTRACTS

Insurance is based on the core principle of risk pooling, that consists in spreading the financial losses of few people among a large group of policyholders. The technical process through which the level of risk of a potential policyholder is estimated is called underwriting, consisting in predicting the probability that a risky event will occur in the future. This measure makes it possible to study tailored premiums for each policyholder, which are based both on individual characteristics, but also on the comparison between the individual and homogeneous groups that present similar characteristics to the potential policyholder (Insurance Europe 2012).

The characteristics analyzed, which are also called "rating factors", depend specifically on the type of contract. There are two types of direct1 insurance, as defined at European level: Life insurance and non-Life insurance. Life insurance contracts involve regular payments by the policyholder, that in return is guaranteed with an agreed sum or an annuity (periodic payment), if an event such as survival or death happens. In the case of non-Life insurance, instead, the policyholder makes regular payment in exchange for an agreed amount if the event, such as accidents, sickness, fire, among others occurs (Eurostat 2010).

Based on this high-level separation, in the Life sector, pricing is driven by personal and biometric variables such as age, occupation, and health behavior. Conversely, non-Life relies on asset-specific characteristics: for instance, Property coverage takes into consideration elements related to building, while motor insurance considers vehicle-specific information about claims, costs of restoration, etc. (Insurance Europe 2012). This separation will be discussed in greater detail.

The risk pooling mechanism can be subject to information asymmetries, leading to adverse selection and moral hazard. Adverse selection occurs before the contract is signed, when high-risk individuals are more likely to seek insurance coverage than low-risk ones, thereby destabilizing the pooling equilibrium. On the other hand, moral hazard arises after the contract is signed; it describes the behavioral change where an insured party, after getting the coverage, may engage in riskier behavior or reduce preventive efforts, thus increasing the insurer's exposure. Furthermore, insurance contracts operate under the Indemnity Principle, a legal constraint ensuring that the payment serves only to restore the financial

[1] Direct Insurance is the main form of insurance, that consists in a constitutional unit (individual, or a collective entity) insuring themselves against specific risk. The other form of insurance is reinsurance, which makes it possible for the insurer to reinsure with other insurers to share a particularly damaging risk (Eurostat 2010). Reinsurance won't be seen in details in this thesis, as it is not a core topic of it.

position of the insured prior to the loss, avoiding any speculative gain (Insurance Europe, 2012).

Not all risks can be covered by Insurance, as there are some insurability criteria to follow: the risk must be definable and financially measurable, that is, can be clearly identified and can be assigned a monetary value. The risk must also be random and independent, meaning that it happens by chance and it's outside policyholder control; moreover the policyholder is required to have a direct relationship (measurable interest) with the insured item (Insurance Europe 2012).

1.3 THE BENEFITS OF INSURANCE

The benefits of Insurance include helping economic stability and individual well-being, by providing coverage for individuals to buy home and start businesses, protecting them against unpredictable events. Moreover, Insurance provide safety by offering a risk-based price, that is, lower prices to people that choose a safer behavior (Insurance Europe 2015).

Beyond its primary function of risk transfer, the insurance sector acts as a basis of macroeconomic stability, by being a major institutional investor. As of the second quarter of 2024, European insurers managed aggregate assets exceeding €8,000 billion (Assets under Management - AuM). The asset allocation strategy demonstrates a diversified profile, weighted towards supporting the real economy and public debt: approximately 35% of the portfolio is allocated to investment funds, 22% to sovereign debt securities (government bonds), and 21% to corporate bonds. Through this massive capital deployment, insurers provide essential long-term funding to the market, financing critical infrastructure and healthcare projects. In terms of underwriting performance, the aggregate Gross Written Premiums (GWP) within the European Economic Area (EEA) reached €1,321 billion in 2024, registering a significant year-on-year growth of approximately 11%. The Life Insurance segment continues to dominate the portfolio mix, accounting for the majority of volumes with €758 billion collected (+14%), whereas the non-Life sector contributed €555 billion (+7%). Some countries distinguish among others in terms of premium volume, being France (€332 billion), Germany (€264 billion), and Italy (€153 billion) (Insurance Europe 2024; EIOPA 2025).

This relevance can also make the insurance sector an engine enhancing economic growth and the competitiveness of Europe in the world, by reducing gaps in investments necessary for the ecological and digital transition, as explained in the Mario Draghi's report "The future of European competitiveness"; in order to do so, a review of laws related to insurance, together with the design of innovative products are necessary, personalized based on the new necessities of coverage against natural disasters (Generali 2025).

1.3.1 Business models and a deep-dive into Life Insurance business

The insurance industry is characterized by a duality in its business models, which is regulated at the European level by the Solvency II Directive (Directive 2009/138/EC). This regulatory framework separates **Life Insurance** and **Non-Life Insurance** (often referred to as General Insurance or Property and Casualty). This distinction reflects differences in the nature of the risks covered, the duration of the contracts, the underlying actuarial techniques, and the financial management of the collected premiums (Insurance Europe 2012). The core of this work is about Life Insurance, however it is essential to provide a comprehensive overview of the specific Lines of Business (LoBs) that constitute the Non-Life segment.

The primary distinction lies in the event triggering the coverage. **Life insurance** contracts are generally long-term agreements where the insured event is linked to human Life duration, specifically death, survival, or distinct biometric risks. These contracts often combine a protection element with a savings or investment component, often focused on a long-term perspective (Eurostat 2010). Consequently, the risk for the insurer is primarily biometric (longevity or mortality risk) and financial (market risk related to the guaranteed returns).

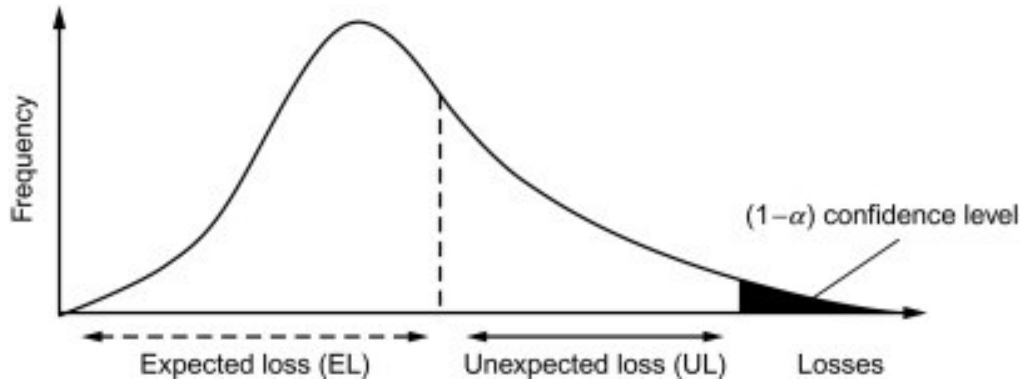
On the other hand, **Non-Life insurance** operates on the Indemnity Principle. As introduced in Section 1.2, the objective is to restore the policyholder's financial position with respect to the loss suffered, without generating any profit. Non-Life contracts are typically short-term (usually renewable annually), and the insurer's risk is focused on the frequency and severity of claims (underwriting risk). Unlike Life insurance, where a claim payment is often a fixed sum (e.g., value of the policy which has been taken out), Non-Life payments are uncertain in amount, depending on the actual economic damage sustained (Insurance Europe 2012).

Within the Non-Life sector, the market is segmented into several key branches, each requiring distinct underwriting approaches and rating factors. The most relevant categories in the European market are Motor insurance, Property insurance, Liability insurance, and Accident and Health.

Motor insurance represents the most significant segment of the Non-Life market in Europe in terms of volume and penetration. It is divided into two components: Motor Third-Party Liability (MTPL; commonly known as "RC Auto" in Italian) and Motor Own Damage (MOD), including, for example Kasko in the Italian context.

Property insurance protects physical assets—residential, commercial, and industrial—against damage or loss. The core guarantee usually covers Fire and Natural Forces. In this domain, the insurer's exposure is characterized by "low frequency, high severity" events. While minor claims (e.g., water damage) are frequent, the solvency (the concept of solvency will be largely discussed in chapter 1.5) of the insurer is tested by catastrophic events, which are difficult to predict and lead to higher costs.

Figure 1.3.1-1: representation of the Loss Function (source: ScienceDirect, https://www.sciencedirect.com/topics/mathematics/loss-distribution)



In recent years, Property Insurance Line of Business has become central to the debate on climate change. As highlighted by the Osservatorio Nazionale Città Clima (2025), the frequency of extreme weather events (floods, hailstorms) has increased the technical loss ratios for property insurers. This has led to a hardening of the market, with insurers revising their catastrophe models to account for non-linear climate trends. The "protection gap", meaning the difference between total economic losses and insured losses, remains a critical issue in Europe, prompting discussions on public-private partnerships to manage natural disaster risks (Generali 2025).

Liability insurance (excluding motor) protects the insured against legal liabilities arising from negligence that causes harm to third parties. The complexity in this segment lies in the legal framework. A prime example is the healthcare sector. The implementation of the "Gelli-Bianco Law" in Italy (Ministero delle Imprese e del Made in Italy 2025), already disclosed in the first chapter, enforces mandatory insurance and the definition of clear boundaries for gross negligence committed by doctors and healthcare professionals.

The Accident and Health line occupies a unique position, often lying between Life and Non-Life classification, providing indemnity in the event of death or permanent disability caused by an external event, as well as coverage of medical expenses and hospitalization.

Having defined Non-Life structure and the main risks covered, ranging from the liability of a driver to the destruction of a building, it is now crucial to return to the other pillar of the industry: Life Insurance. While Non-Life focuses on asset protection and liability, Life insurance is about two main blocks: longevity risk (risks of living more than predicted and not having sufficient resources) and premature death (risk related to expenses, debts left and insufficient income for relatives).

The classification of Life insurance activities is defined in Annex II of Directive 2009/138/EC:

Class I – Life Insurance: Standard policies covering biometric risks. Includes Term Life insurance (assures a payment to the beneficiary in the case of the death of the insured[2] during the time of the contract), Whole Life insurance (duration is not limited and there's an investment component; the payment is always done when insured's death happens), and Annuities.

Class II – Marriage and Birth Assurance: Policies that provide a lump sum upon marriage or the birth of a child. Class II is today considered obsolete; it's nonetheless disciplined by law.

Class III – Life Insurance Linked to Investment Funds: Policies where the investment risk is borne by the policyholder. Includes: Unit-linked and Index-linked policies. The payout depends entirely on the value of the reference units or share indices.

Class IV – Permanent Health Insurance (PHI): Long-term health coverage. Includes: Income protection against disability or critical illness (long-term sickness) where the insurer cannot cancel the policy as long as premiums are paid.

Class V – Tontines: Operations whereby associations of members are set up with a view to jointly capitalising their contributions and subsequently distributing the assets thus accumulated among the survivors or among the beneficiaries of the deceased.

Class VI – Capital Redemption Operations: Financial products issued by insurers with no biometric risk (no link to human Life duration).

[2] Such products typically allow for a distinction between the contracting parties. Specifically, the Beneficiary is frequently a different individual from the Insured, who, in turn, may differ from the Policyholder (the party responsible for premium payments).

Class VII – Management of Group Pension Funds: Management of investments and reserves for entities providing retirement, death, or disability benefits to employees (e.g., severance pay funds or company pension schemes).

Class VIII – Social insurance operations (Public/Private mix): These classes cover risks related to sickness, accidents, and disability, managed by private insurers but within the strict legal framework of national social security systems (primarily relevant for France).

Class IX – Social insurance operations (Life/Retirement): Similar to Class VIII but specifically related to Life duration and retirement benefits prescribed by social security legislation, when practiced by insurance undertakings [3] at their own risk (European Parliament and Council 2009).

These are the classes reported at a European level, that are slightly different from what the Italian law assesses, reducing, for example, the number of classes from nine to six, and inserting Group Pension Funds in Class VI instead of VII (COVIP n.d.).

Based on data from the main supervisory bodies and industry associations (ANIA, EIOPA, IVASS), the European Life insurance market—and the Italian one in particular—is characterized by a dual nature: a predominant financial component (savings and investment) and a residual biometric component (protection).

Insurance-Based Investment Products represent the majority of Gross Written Premiums (GWP) in Continental Europe. These products can be found mostly in Class I and Class III. Class I (traditional Life policies) assures capital-guaranteed products linked to funds which are not directly exposed to markets, with an investment risk on the behalf of the insurer, offering then stable returns to the beneficiary; on the contrary, Class III includes financial products where the investment risk is fully borne by the policyholder (ANIA 2024). In the case of Italy and France (Southern Europe in general), Class I products have been the most common ones in 2022, while UK and Northern Europe had a prevalence of Class III products in their portfolios (EIOPA 2023).

Protection Products covering Biometric Risk, such as Term Life and Disability account for a minor share of total premiums in Europe (IVASS 2023)

The post-pandemic economic environment has drastically reshaped the sector dynamics: the sharp rise in interest rates in years 2022-2023 caused a wave of early surrenders (lapses). Policyholders liquidated Class I policies (which were yielding low historical

[3] Undertaking is the technical term used to define an insurance company.

returns) to invest in higher-yielding Government Bonds (e.g., Italian BTPs). This created a liquidity strain for insurers, exemplified by the Eurovita case in Italy (EIOPA 2023).

In response to the liquidity crisis, insurers in 2024 launched new Traditional products with boosted yields to retain customers. At the same time, Hybrid products, combining Class I guarantees with Class III potential returns, remain a key strategy to balance capital absorption, which is has to be high for Class I policies, under Solvency II Directive, that will be discussed in the next chapter.

1.4 THE SOLVENCY II REGIME

The insurance industry in the European Union operates under a regulatory framework known as **Solvency II**, introduced by Directive 2009/138/EC and fully effective since January 1, 2016. Unlike its predecessor (Solvency I), which relied on static, volume-based capital requirements, Solvency II aims at shifting towards a risk-based approach. This framework is designed not merely to ensure that insurers hold sufficient capital to cover generic liabilities, but to align the capital requirements with the specific risk profile of each undertaking, hence suggesting a more dynamic and heterogeneous approach (European Parliament and Council 2009). The regime's primary objectives are to enhance consumer protection, ensuring that policyholders receive their payments even in adverse economic scenarios, and to promote financial stability across the single market (EIOPA n.d.).

From a structural perspective, Solvency II is built upon three "Pillars", in some ways similar to the Basel II accord in banking. Solvency is an "architecture" that encompasses quantitative metrics, qualitative governance and market discipline.

The Pillars are shortly described to provide a brief overview of how this important framework is, however this is not the main focus of the current work.

1.4.1 Pillar 1: Quantitative Requirements and Capital Adequacy

Pillar 1 establishes the fundamental quantitative rules for financial stability, requiring insurers to adopt a "Market Consistent Balance Sheet". In this framework, assets are valued at current market prices, while liabilities, which are technically known as Technical Provisions, represent the best estimate of the future payments the insurer will have to make to policyholders (European Parliament and Council 2009). Calculating these values is a complex task that requires actuaries to project future cash flows by modeling key factors such as mortality trends and policyholder behavior (e.g., the likelihood of early contract termination).

To ensure resilience against adverse events, the regulation defines two distinct capital thresholds:

Solvency Capital Requirement (SCR) is the target capital level, regulated to ensure the insurer can withstand a severe financial shock over the next 12 months. It covers all major risks, including Market Risk (fluctuations in interest rates or equity prices) and Underwriting Risk (uncertainty in insurance liabilities).

Minimum Capital Requirement (MCR) is a lower, absolute minimum threshold. Falling below this level triggers immediate regulatory intervention to protect policyholders,

potentially leading to the withdrawal of the company's license, that won't be able to operate (EIOPA n.d.).

Insurers can calculate their capital needs using a Standard Formula provided by the regulator or, for a more tailored approach, an Internal Model. This flexibility highlights the strategic importance of advanced statistical analysis: the ability to accurately map and quantify risks is crucial for optimizing capital allocation.

1.4.2 Pillar 2: Governance and Risk Management

Pillar 2 focuses more on the qualitative aspects of management. It requires insurers to implement a robust System of Governance, establishing key independent functions: Risk Management, Compliance, Internal Audit, and the Actuarial Function (European Parliament and Council 2009).

A central element of this pillar is the Own Risk and Solvency Assessment (ORSA). The ORSA is a continuous process where the insurer assesses its overall solvency needs based on its specific risk profile, looking beyond the standard regulatory formulas and considering a medium-term time horizon of 3-5 years (EIOPA, 2015). The regulations explicitly task the Actuarial Function with ensuring the sufficiency and quality of the data used in decision-making processes. Since real-world data is complex and often contain imperfections, data validation and robust models are necessarily required.

1.4.3 Pillar 3: Disclosure and Market Discipline

Pillar 3 imposes strict transparency requirements to enhance market discipline. Insurers are obliged to publish annual reports detailed in the Solvency and Financial Condition Report (SFCR), which is accessible to the public, alongside confidential reports for the regulator.

These documents must detail the company's capital management, risk profile, and valuation methodologies. The "Solvency Ratio" (the ratio of available capital to the required capital) is considered a key competitive metric used by investors and analysts to judge an insurer's financial health. Therefore, the ability to manage capital consumption through precise strategic planning is vital. In this context, advanced analytical tools that can capture market dynamics and trends become valuable assets for supporting dividend policies and investment strategies.

Reports provided by undertakings are essential for this thesis, as primary source of the data analyzed in the subsequent sections.

1.4.4 Strategic Implications for Life Insurance Business

The introduction of Solvency II has fundamentally reshaped the Life insurance market described in the previous sections. The capital charges for Market Risk are particularly high for traditional products with guaranteed capital, as they expose insurers to significant interest rate risks.

This regulatory pressure has driven the industry's product innovation towards Unit-Linked and Hybrid products (previously described in chapter 1.4.2), where the investment risk is carried by the policyholder (ANIA 2024). While this shift reduces certain capital requirements, it introduces new challenges related to customer behavior and lapse risk. In a market where profitability is increasingly linked to fees and long-term customer relationships, understanding the dynamics of portfolio growth and stability becomes essential. Solvency II is then a comprehensive management framework that integrates risk into every business decision. For the modern insurer, this necessitates moving beyond simple deterministic calculations towards more sophisticated modeling approaches that can better interpret the complex economic environment.

1.5 AIMS AND OBJECTIVES

The relationship between insurance market development and macroeconomic stability has been greatly debated in the financial literature. Many studies suggest a causal link between insurance penetration and economic growth, highlighting how insurance activities act as a buffer against financial shocks and promote long-term stability (Outreville 2013a, Pradhan et al. 2016). In this context, insurance penetration is defined as the ratio of Gross Written Premiums (GWP), representing the total volume of premiums issued by insurers, to the country's GDP (OECD, 2025).

Furthermore, recent empirical evidence indicates that the insurance sector is deeply linked to structural socio-economic indicators such as inflation, interest rates, and GDP per capita (Verma and Bala 2013).

However, while traditional econometric approaches often treat these variables as direct predictors of growth, there is a need to understand how they characterize distinct "unobserved" phases of market evolution, defined as latent states, which can be assumed, for instance, by premiums growth (latent states will be further analyzed in the next chapter).

Markets may switch across different latent regimes over time, driven by underlying economic conditions (Maruotti 2011). Therefore, including socio-economic covariates directly into the measurement of these states allows for a more detailed clustering of countries, distinguishing, for instance, between mature markets and developing ones, based on their structural financial health.

Based on this premise, the primary objective of this thesis is to investigate the temporal dynamics of the insurance sector by applying a Latent Markov Model (LMM) framework to quarterly panel data. This research mainly aims to understand how to model and detect unobserved heterogeneity (latent states) in insurance data over time. Secondly, it is essential to understand what the role of socioeconomic and financial factors in the defined latent states is.

The empirical analysis will be conducted in two sequential stages, on a high-level distinction:

- Firstly, a homogeneous LMM will be constructed without covariates. The goal here is to identify the optimal number of latent states and to analyze the pure transition probabilities, isolating the internal dynamics of the system without the influence of external factors.
- Then, the model will be extended by introducing specific socioeconomic and structural financial indicators as covariates. A distinctive feature of this analysis is

that these covariates will be included in the measurement model, rather than in the latent transition model. By using this approach, covariates have an impact on the clustering of statistical units within each latent state, rather than modelling the transition from one state to another.

Lastly, this thesis seeks to provide an analysis of the trajectories of latent states over time to better understand their evolution.

1.6 THE LITERATURE GAP AND METHODOLOGICAL INNOVATION

While the relationship between macroeconomic variables—such as inflation, interest rates, and GDP—and insurance market performance has been widely explored in existing financial literature (Outreville 2013a, Verma and Bala 2013, Pradhan, Arvin and Norman 2016), traditional econometric approaches predominantly rely on static panel regressions. These standard methods generally treat macroeconomic indicators as direct predictors, implicitly assuming a homogeneous impact over time and assuming that the underlying market structure remains constant. Consequently, a significant gap exists in the ability to capture the unobserved heterogeneity and dynamic regime shifts that characterize the insurance sector during periods of severe economic stress (Maruotti 2011).

This thesis aims to bridge this gap by introducing a Latent Markov Model (LMM) framework to the analysis of the European Life insurance market (Bartolucci, Farcomeni and Pennoni 2012). While much of the literature has focused on measuring the direct impact of inflation or interest rates through standard regressions, the use of an LMM to isolate latent regimes and structural market phases represents an innovative methodological approach (Pandolfi, Bartolucci and Pennoni 2023). By modeling the transition probabilities between two latent states (Bertarelli 2015), this research moves beyond static impact assessments, providing a necessary and dynamic perspective on how systemic macroeconomic shocks structurally alter the equilibrium of the insurance sector (Colombi, Giordano and Kateri 2024).

2 DATA DESCRIPTION AND EXPLORATORY DATA ANALYSIS (EDA)

2.1 DATA SOURCE

The empirical analysis of this study relies on the comprehensive quarterly statistics provided by the **European Insurance and Occupational Pensions Authority (EIOPA)**. This dataset represents the primary source of harmonized prudential data for the European insurance sector, covering the entire European Economic Area (EEA) (EIOPA 2025).

The data is extracted from the Solvency II reporting framework, already discussed in chapter 1.5, which mandates a uniform reporting standard for all insurance undertakings operating within the EU (European Parliament and Council 2009). Unlike aggregated national statistics, which may suffer from methodological differences due to local accounting standards, Solvency II data ensures strict cross-country comparability and statistical robustness. The specific dataset used is the "European Insurance Overview", available via EIOPA's official statistical repository (EIOPA 2025). The official data source hence is EIOPA Insurance Statistics and the specific dataset corresponds to: *Quarterly Solo Prudential Reporting* (2017 Q1 – 2025 Q2).

This dataset is not estimated by external analysts but is reported directly by the insurance undertakings (insurers) to their respective National Competent Authorities (NCAs), such as IVASS in Italy or ACPR in France. These authorities then validate and transmit the data to EIOPA.

The values are derived from the Quantitative Reporting Templates, specifically from template **S.05.01** (*Premiums, claims and expenses by line of business*). The reporting is based on a "best estimate" valuation consistent with market-consistent balance sheet principles under Solvency II (EIOPA 2025).

For this study, the analysis focuses strictly on the Life Insurance line of business. The raw data points—extracted to compute the response variables—correspond to the following official Solvency II labels:

1. **Gross Written Premiums (GWP):** Officially labeled as *Premiums written - Gross - Direct Business*. This represents the total volume of premiums issued before any deduction for reinsurance. It serves as the primary indicator of market activity and turnover.

2. **Claims Incurred:** Officially labeled as *Claims incurred - Gross - Direct Business*. In the Life sector, this figure aggregates payments for death benefits, maturities (end of contract), and surrenders (early withdrawals).
3. **Expenses Incurred:** Officially labeled as *Expenses incurred*. This includes administrative costs, acquisition costs (commissions), and overheads associated with insurance operations.

While the raw dataset provides absolute monetary values (in Euros), relying on these would introduce a significant size bias. A cluster analysis based on absolute GWP would simply group countries by their economic dimension (e.g., separating "Big 5" economies from smaller states) rather than by their performance or risk profile (Gatzert and Wesker, 2012).

To address this, the raw values were transformed into relative measures (rates and ratios). This approach allows for the assessment of solvency, technical solidity, and efficiency, which are key metrics under the Solvency II framework, independent of market size. By standardizing the variables, we can effectively compare the performance of a mature, large market like France with a smaller, high-growth market like Estonia.

2.2 RESPONSE VARIABLES DEFINITION

To capture the multidimensional nature of the Life insurance market, four specific response variables were computed based on the raw EIOPA data. These variables serve as the foundation for the subsequent clustering analysis.

- **Premium Growth Rate ($premium_growth$)** This variable is calculated as the percentage change in Gross Written Premiums compared to the same quarter of the previous year (Year-on-Year, YoY):

$$\text{Premium Growth}_t = \frac{GWP_t - GWP_{t-4}}{GWP_{t-4}}$$

Using YoY transformation is essential to resolve the seasonality issue, which is particularly pronounced in the Life insurance sector. Fiscal incentives and commercial deadlines often lead to peaks in premium collection in the fourth quarter, followed by a drop in Q1. A simple quarter-on-quarter comparison would capture this seasonal noise rather than the underlying market trend (EIOPA 2025).

- **Loss Ratio ($loss_ratio$)** The Loss Ratio is defined as the ratio of claims incurred to premiums earned, expressed as a percentage:

$$\text{Loss Ratio} = \frac{\text{Claims Incurred}}{\text{Premiums Earned}}$$

In the Life sector, unlike Non-Life insurance where claims are driven by accidents or damages, claims mainly consist of maturities and surrenders (lapses). Therefore, this metric is a critical indicator of liquidity risk. A Loss Ratio consistently exceeding 100% does not merely indicate technical losses; it may signal a "run-on-the-bank" scenario. In such cases, policyholders, driven by panic or the search for higher yields elsewhere, withdraw funds before contract maturity, representing a potential challenge for the insurer's solvency (Feodoria and Förstemann 2015).

- **Expense Ratio (exp_ratio)** Calculated as the ratio of operating expenses to premiums:

$$\text{Expense Ratio} = \frac{\text{Operating Expenses}}{\text{Premiums Earned}}$$

This variable serves as a proxy for the operational efficiency of the distribution channels. It indicates the weight of technical costs, such as commissions paid to agents and brokers, on overall profitability. A lower ratio typically suggests a more

digitized and efficient distribution model, whereas high ratios may indicate legacy cost structures or high acquisition costs (Eling and Luhnen 2010).

- **Technical Margin ($\boldsymbol{margin}$)** The Technical Margin isolates the capacity to generate profit strictly from underwriting activities, excluding financial income from investments. It is computed as:

$$\text{Margin} = \frac{\text{Premiums} - (\text{Claims} + \text{Expenses})}{\text{Premiums}}$$

This metric provides a "pure" measure of technical profitability. In a low interest rate environment, where investment returns are compressed, the ability to generate a positive technical margin becomes a vital determinant of an insurer's long term sustainability.

2.3 COVARIATES DEFINITION

To explain the latent market states identified by the response variables (e.g., expansion, stagnation, crisis), a robust set of covariates was selected. These are divided into Macro-Financial Drivers and Socio-Demographic Drivers.

Covariate data was primary retrieved from Eurostat, the statistical office of the European Union, ensuring consistency with the EIOPA geographical scope. Specific integrations were required to ensure complete time-series coverage for the 10-Year Government Bond Yields (Y10). Standard Eurostat repositories provided incomplete historical data for Norway, Iceland, and Estonia for the 2017-2025 period. Consequently, these gaps were filled by merging external datasets retrieved from the European Central Bank (ECB) Data Portal and respective National Central Bank historical archives, ensuring a continuous macro-financial timeline for all analyzed countries.

The denomination between brackets corresponds to the one used in the code and figures.

A. Macro-Financial Covariates

- **HICP Monthly Inflation (*hicp*)** *Source: Eurostat (prc_hicp_midx)* Converted into quarterly data; the Harmonized Index of Consumer Prices measures the rate of inflation. Inflation has a double nature for Life insurance: it erodes real disposable income, reducing the demand for new premiums (GWP), and simultaneously increases the future cost of claims and operational expenses. Furthermore, high inflation often triggers a lapse risk, when policyholders liquidate policies before reaching maturity to cover rising living costs (Eurostat 2025a).
- **Nominal GDP (*gdp*)** *Source: Eurostat (namq_10_gdp)* This variable represents the business cycle. A positive GDP trend is expected to correlate with an "expansion state", boosting premium income. Economic growth is expected to provoke an increase in household wealth, thereby expanding the capacity to invest in long-term savings products (Eurostat 2025b).
- **Unemployment Rate (*une_rate*)** *Source: Eurostat (une_rt_q)* The unemployment rate acts as a proxy for household economic stability. Higher unemployment lowers the demand for savings and insurance products, potentially triggering a "crisis state". Conversely, low unemployment suggests a robust labor market where individuals have the surplus income necessary for life protection products (Eurostat 2025c).

- **10-Year Interest Rates ($y10$)** *Source: ECB / Eurostat (irt_lt_mcby)* Long-term interest rates determine the competitiveness of traditional Life insurance products (like With-Profits policies, i.e. Class I products that provide returns and whose risk is borne to the insurer) relative to other risk-free investments such as government bonds. Rapidly rising rates can lead to an increase in the Loss Ratio (lapses) and a decrease in Premium Growth (disintermediation risk, i.e. withdrawal from policy contract to invest the liquidity into other financial instruments), reflecting the market's sensitivity to European monetary policy (Eurostat 2025d).

B. Socio-Demographic Covariates

To deepen the analysis, specific socio-demographic indicators were included to capture structural demand drivers.

- **Old Dependency Ratio (old_dep)** *Source: Eurostat (demo_pjanind)* This ratio measures the weight of the population aged 65 and over relative to the working-age population (15-64). A higher ratio implies a greater structural need for longevity protection and annuities. In particular, key markets like Italy and France exhibit significant aging trends, with the elderly population representing 24% and 21% of the total, respectively (Eurostat, 2025). This demographic shift is a primary driver for the "pension gap" narrative in insurance marketing (Eurostat 2025f).
- **Gender Employment Gap ($gender_gap$)** *Source: Eurostat (lfsi_emp_q)* Defined as the difference between the employment rates of men and women aged 20-64. A lower gap indicates higher female participation in the labor force, which correlates with dual-income households and a higher aggregate capacity to purchase insurance products. furthermore, women are statistically more risk-averse than men, potentially showing a higher propensity for pure protection products (term life) (Eurostat 2025g).
- **Life Expectancy at Birth ($life_exp$)** *Source: Eurostat (demo_mlexpec)* This metric is fundamental for the pricing and profitability of annuity products. An increase in life expectancy raises the "longevity risk" for insurers (paying annuities for longer than expected) but also increases the consumer's perceived need for products that guarantee income in old age. Variations in life expectancy across Europe (e.g., Eastern vs. Western Europe) help explain structural differences in product mix (Eurostat 2025h).

- **Immigration Rate ($immigration_rate$)** *Source: Eurostat (migr_pop1ctz)* Calculated as the percentage of foreign citizens relative to the total population. This variable serves as a proxy for financial inclusion challenges within the potential customer base. While immigration positively contributes to the labor force supply in aging economies, foreign nationals often encounter structural barriers to accessing complex financial products (e.g., language differences, lack of credit history, or shorter planning horizons). Consequently, a higher share of foreign residents may structurally limit the immediate penetration of long-term Life insurance products, identifying a market segment that is often "underserved" by traditional distribution channels (Eurostat 2025i).
- **NEET Rate ($neet$)** *Source: Eurostat (edat_lfse_20)* The percentage of young people "Not in Education, Employment, or Training." A higher rate indicates a "lost generation" unable to generate income or accumulate savings. This limits the future customer base for life protection and signals structural economic weakness (Eurostat 2025j).
- **Tertiary Education (t_edu)** *Source: Eurostat (edat_lfse_03)* Used as a proxy for financial literacy. Higher education levels are associated with a better understanding of complex insurance products and a higher propensity to invest in policies rather than simple bank deposits (Eurostat 2025k).

C. Structural and Welfare Covariates

- **Digital Banking Penetration (dig_bank)** *Source: Eurostat (isoc_bde15cu)* A high adoption rate of internet banking facilitates the distribution of online policies, helping to reduce the Expense Ratio. It also enhances the customer's ability to manage savings products autonomously (Eurostat 2025l).
- **Out-of-Pocket Expenditure (oop_exp)** *Source: Eurostat (hlth_sha11_hf)* A proxy for the efficiency of the Welfare State. High private health expenditure suggests a gap in public coverage. In countries where citizens must pay significantly out-of-pocket for care, there is a latent, often unfulfilled, necessity for private life and health insurance solutions to act as a safety net (Eurostat 2025m).
- **At-Risk-of-Poverty Rate ($arop$)** *Source: Eurostat (ilc_li02)* A measure of income inequality. High inequality may restrict the insurance market to a smaller, elite segment of the population, limiting overall penetration rates. When a significant portion of the population is at risk of poverty, insurance

becomes a luxury good rather than a basic financial planning tool (Eurostat 2025n).

2.4 EXPLORATORY DATA ANALYSIS

The Exploratory Data Analysis (EDA) serves as a fundamental preliminary step in this research, providing a comprehensive understanding of the dataset's structure, the distribution of key variables, and the presence of any anomalies or patterns that could influence the subsequent LMM estimation. This section details the data processing workflow, the handling of missing values, and the descriptive statistics of the variables involved.

2.4.1 Dataset Structure and Pre-Processing

Before proceeding with the descriptive statistics, it is essential to outline the architecture of the final dataset obtained after the preprocessing and integration phases described above. The analysis relies on a balanced longitudinal panel dataset covering 29 European countries over a time horizon of 31 quarters (from 2017 Q4 to 2025 Q2), resulting in a total of 899 observations.

The dataset is organized in a "long" format, where each row corresponds to a unique Country-Quarter pair ($Country_i, Time_t$). This structure is the standard requirement for panel data analysis in R and specifically for the LMest package used in the LMM estimation.

The dataframe comprises 22 variables, which can be categorized into three logical blocks:

- Ids: country_region (chr) and date (Date/Year-Quarter).
- Insurance Response Variables: Including raw volumes (premiums, claims, expenses) and the computed performance indicators (premium_growth, loss_ratio, exp_ratio, margin).
- Covariates: A set of 13 continuous variables covering macroeconomic drivers and socio-demographic structural indicators.

Table 2.4.1-1 provides a technical overview of the dataset structure, displaying the variable types and a preview of the stored values.

Table 2.4.1-1: Dataset Structure. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.

Dataset Structure

Observations: 899 | Variables: 22

Variable	Class	First Values
country_region	character	AUSTRIA, AUSTRIA, AUSTRIA, AUSTRIA, ...
date	yearqtr	2017 Q4, 2018 Q1, 2018 Q2, 2018 Q3, ...
premiums	numeric	7921.073, 2119.947, 4026.267, 5826.043, ...
claims	numeric	7750.325, 2051.259, 4000.757, 5612.119, ...
expenses	numeric	914.788, 247.432, 507.592, 725.894, ...
premium_growth	numeric	NA, NA, NA, NA, ...
loss_ratio	numeric	97.844, 96.76, 99.366, 96.328, ...
exp_ratio	numeric	11.549, 11.672, 12.607, 12.459, ...
margin	numeric	-9.393, -8.432, -11.973, -8.788, ...
gdp	numeric	2.5, 2, 2.5, 2.3, ...
une_rate	numeric	5.8, 5.4, 5, 5.3, ...
hicp	numeric	2.367, 1.933, 2.133, 2.233, ...
neet	numeric	8.6, 8.9, 9, 8.6, ...
gender_gap	numeric	8.2, 8.9, 9.2, 9, ...
oop_exp	numeric	20.11, 19.29, 19.29, 19.29, ...
old_dep	numeric	49.1, 49.5, 49.5, 49.5, ...
t_edu	numeric	32.4, 32.7, 32.7, 32.7, ...
arop	numeric	14.4, 14.3, 14.3, 14.3, ...
dig_bank	numeric	57.49, 58.41, 58.41, 58.41, ...
life_exp	numeric	81.7, 81.8, 81.8, 81.8, ...
immigration_rate	numeric	0.152, 0.157, 0.157, 0.157, ...
y10	numeric	0.54, 0.77, 0.74, 0.62, ...

As shown in the above table, all financial and macroeconomic variables are stored as numeric, precisely double-precision vectors (dbl), ensuring computational precision. The country_region variable is encoded as a factor to facilitate the grouping of units during the iterative likelihood maximization process.

The final dataset is constructed by merging insurance-specific micro-data—namely Premiums, Claims, and Expenses—with a set of macroeconomic and socio-demographic covariates. The primary data source for the insurance variables is the quarterly prudential reporting provided by the European Insurance and Occupational Pensions Authority (EIOPA). The dataset follows a "long" format structure, where each row represents a unique observation defined by the pair $(Country, Quarter)$.

To ensure the robustness of the subsequent econometric analysis, a rigorous data cleaning and integration protocol was implemented.

During the initial screening of the covariates, a structural assessment of data availability was conducted. The variable representing *Gross Disposable Household Income* (originally retrieved from Eurostat) exhibited a significant rate of missingness across multiple jurisdictions, particularly for smaller economies. Since imputing this variable would have introduced considerable bias without adding substantial explanatory power compared to GDP, it was excluded from the final model configuration.

Similarly, the country Liechtenstein was removed from the sample. Despite having available insurance data, the lack of consistent quarterly macroeconomic indicators for this jurisdiction would have compromised the convergence of the model.

While Eurostat served as the primary source for standard macroeconomic indicators, specific integrations were required to ensure complete time-series coverage for the 10-Year Government Bond Yields (y10), a critical proxy for the risk-free rate and investment returns in the Life sector.

Standard repositories provided incomplete historical data for Norway, Iceland, and Estonia for the 2017-2025 period. To address this gap, data were manually integrated using historical datasets from the European Central Bank (ECB) Statistical Data Warehouse and respective National Central Banks. For Norway and Iceland, monthly yield data were aggregated into quarterly averages to match the temporal resolution of the panel. For Estonia, specific gaps in the 2017-2019 period were reconstructed based on ECB convergence reports, ensuring a continuous financial timeline for all 30 analyzed countries.

Furthermore, the socio-demographic covariate Gender Employment Gap (gender_gap) was explicitly derived to capture labor market disparities. It was computed as the arithmetic difference between the male and female employment rates (percentage of total population aged 20-64) provided by Eurostat:

$$\text{Gender Gap}_{it} = \text{EmpRate}_{Male,it} - \text{EmpRate}_{Female,it}$$

This metric serves as a proxy for the structural differences in household income composition and female financial independence across Europe.

A specific preprocessing step was applied to align annual covariates (e.g., *Old Dependency Ratio*, *Tertiary Education*, *Digital Banking*) with the quarterly frequency of the response variables. Annual figures were duplicated across the four quarters of the respective year to maintain dimensional consistency. Finally, to handle minor missing values in the very last

quarters of the series (2025 Q1-Q2) for specific slow-reporting covariates, a "Last Observation Carried Forward" (LOCF) approach was adopted, under the assumption of short-term stability for structural indicators (OECD, 2025).

2.4.2 Descriptive Statistics

To characterize the market dynamics, we computed descriptive statistics: Minimum, Maximum, 1st and 3rd Quartiles, Mean, Median, Standard Deviation (SD), and Coefficient of Variation (CV), for all continuous variables. The Coefficient of Variation ($CV = \frac{SD}{Mean}$) allows for a normalized comparison of volatility across different variables and countries.

The table below presents the aggregate descriptive statistics for the entire dataset, encompassing all included European countries.

Table 2.4.2-1: Descriptive statistics. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.

Descriptive Statistics (Response Variables vs Covariates)

Variable	Min	Q1	Median	Mean	Q3	Max	SD	CV
premiums	9.18	293.21	2082.62	14456.60	13009.08	210866.91	30330.16	209.80
claims	3.44	196.52	1744.40	12124.64	10963.76	174844.58	25560.78	210.82
expenses	1.31	63.29	309.44	1545.12	1177.26	25618.68	3696.51	239.24
premium_growth	-52.26	-3.92	3.56	5.25	10.76	186.36	18.85	359.00
loss_ratio	9.23	64.62	75.75	78.64	90.83	185.43	25.13	31.95
exp_ratio	3.04	10.50	14.40	15.78	21.80	64.82	7.75	49.11
margin	-98.60	-6.34	7.98	5.58	21.25	85.44	24.25	434.84
gdp	-21.50	0.40	2.20	2.36	4.30	22.70	4.73	200.72
une_rate	1.90	4.20	5.70	6.14	7.10	21.10	2.90	47.21
hicp	-2.27	1.33	2.50	3.78	4.72	25.87	4.02	106.51
y10	-0.61	0.44	1.94	2.07	3.23	9.23	1.86	89.96
old_dep	43.80	51.50	55.10	54.43	57.40	62.40	4.45	8.18
gender_gap	-0.50	5.70	7.80	9.01	11.35	24.20	5.10	56.64
neet	2.50	8.50	10.30	11.07	13.05	24.30	4.02	36.33
t_edu	17.60	29.80	38.30	36.93	44.00	56.60	8.95	24.24
dig_bank	5.49	53.36	66.92	65.67	83.34	97.76	21.42	32.62
arop	8.60	12.80	15.00	15.96	19.30	23.80	3.90	24.43
life_exp	71.20	78.20	81.60	80.41	82.60	84.10	2.92	3.63
immigration_rate	0.01	0.05	0.09	0.10	0.14	0.48	0.09	89.71
oop_exp	8.39	12.92	18.37	19.70	24.15	44.76	8.39	42.61

The descriptive analysis highlights the extreme heterogeneity of the European Life insurance market, however for all variables mean and median values are sufficiently close to assess that there are no specific countries acting as outliers in the analysis. The Premium Growth variable exhibits a high standard deviation, supported by a CV value of 359,

reflecting the volatile nature of premium collection in the Life sector, which is sensitive to commercial cycles and fiscal deadlines. The Loss Ratio, averaging approximately 79, shows a significant range of variation (from 9.23 to 185.43), however limited with respect to Premium Growth, suggesting that Loss Ratio in the Life sector is not so dispersive, as Lapse Risk mainly depends on demographic and behavioral dynamics.

Macroeconomic covariates reflect heterogeneous economic environments, as evidenced by the high Coefficient of Variation (CV) values. Specifically, GDP exhibits the greatest dispersion (CV = 201%), followed by HICP (107%) and the 10-year yield (90%). For instance, the 10-Year Interest Rate (y10) ranges from negative values (observed in core Eurozone countries like Germany during the quantitative easing era) to peaks exceeding 9.23% in Eastern European economies or during the 2023 inflationary spike. This variability is essential for the LMM to identify distinct market regimes driven by the monetary environment.

In contrast, socio-demographic and structural indicators present a more nuanced, although generally lower, degree of heterogeneity. Focusing on the socio-demographic covariates, the level of dispersion varies significantly among the variables. The immigration rate (immigration_rate) stands out with a notably high CV of 89.71%, indicating stark demographic differences across the sample. The gender gap (gender_gap) and the share of NEETs (neet) also show moderate-to-high variability, with CVs of 56.64% and 36.33%, respectively. Conversely, the old-age dependency ratio (old_dep) is characterized by remarkable stability, displaying the lowest heterogeneity in this category with a CV of just 8.18%, which is comprehensible considering that Europe sees a progressively growing share of individuals aged 65 and over (21.6% in 2024) (Eurostat 2024).

Regarding the structural covariates, the dataset reveals a moderate and relatively consistent level of variation. Both out-of-pocket health expenditure (oop_exp) and the adoption of digital banking (dig_bank) exhibit similar dispersion patterns, with CVs of 42.61% and 32.62%, respectively.

To better illustrate these disparities, we compare two distinct market archetypes within the sample: Italy, representing a mature, high-volume market, and Slovakia, representing a smaller, developing market.

Table 2.4.2-2: Descriptive statistics for Italy. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.

Descriptive Statistics: ITALY

Variable	Min	Q1	Median	Mean	Q3	Max	SD	CV
premiums	25781.81	40562.78	62529.98	64844.16	87405.56	112074.45	28707.73	44.27
claims	18926.72	35996.51	57852.65	55898.63	76349.09	114897.93	26294.53	47.04
expenses	1281.69	1928.46	3116.28	3232.10	4461.17	5874.33	1463.65	45.28
premium_growth	-10.74	-5.18	-0.44	2.30	9.88	20.96	10.03	435.29
loss_ratio	69.20	73.98	77.81	87.15	100.03	122.90	18.53	21.26
exp_ratio	4.09	4.74	4.95	4.97	5.11	5.68	0.34	6.80
margin	-28.57	-5.18	17.79	7.88	21.35	25.92	18.77	238.05
gdp	-18.10	0.45	0.70	1.11	2.00	19.30	5.83	525.40
une_rate	6.00	7.70	8.90	8.72	9.95	11.00	1.54	17.66
hicp	-0.40	0.87	1.17	2.67	2.95	12.50	3.34	124.93
y10	0.64	1.29	2.76	2.58	3.70	4.38	1.30	50.26
old_dep	56.10	56.40	57.30	57.02	57.50	57.60	0.56	0.98
gender_gap	18.70	19.30	19.50	19.51	19.70	20.00	0.32	1.64
neet	13.40	16.05	22.30	20.08	23.45	24.30	3.77	18.79
t_edu	18.70	19.70	20.00	20.53	21.60	22.30	1.11	5.41
dig_bank	30.83	36.31	45.15	44.50	51.55	55.01	8.03	18.05
arop	18.90	18.90	20.10	19.73	20.10	20.30	0.59	2.99
life_exp	82.30	82.75	83.40	83.25	83.60	84.10	0.60	0.72
immigration_rate	0.08	0.08	0.09	0.09	0.09	0.09	0.00	3.12
oop_exp	21.48	22.31	23.32	23.04	23.87	24.13	0.86	3.75

As shown in the table above, Italy represents the archetype of a mature and consolidated insurance sector, with absolute Premiums volumes reaching an average value of about 64,844. Its technical indicators show high volatility in years, possibly due to external shocks, reaching very high levels. The macroeconomic variables track the Eurozone average, acting as a stabilizing anchor. For Italy, the primary risk captured by the data is related to the interest rate cycle (Y10) and inflation, and its impact on the surrender of Life products, rather than structural inefficiency.

Table 2.4.2-3: Descriptive statistics for Slovakia. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.

Descriptive Statistics: SLOVAKIA

Variable	Min	Q1	Median	Mean	Q3	Max	SD	CV
premiums	180.63	308.82	503.67	547.11	754.47	1072.96	272.32	49.77
claims	99.84	198.42	367.67	382.26	554.62	763.33	196.21	51.33
expenses	39.46	71.73	116.52	130.63	177.44	249.34	65.60	50.22
premium_growth	-15.99	-6.86	-4.28	-3.49	1.48	6.09	6.29	180.09
loss_ratio	55.27	67.05	69.26	69.43	73.96	76.20	5.73	8.25
exp_ratio	19.18	21.83	23.59	24.00	25.43	30.54	3.11	12.96
margin	-5.62	2.64	6.08	6.57	8.91	22.89	6.99	106.43
gdp	-9.00	0.75	2.10	2.01	3.35	13.20	3.22	160.38
une_rate	5.20	5.70	6.00	6.15	6.60	7.60	0.64	10.40
hicp	1.03	2.42	3.03	5.04	5.93	15.10	4.25	84.30
y10	-0.42	0.05	0.84	1.56	3.40	3.80	1.59	101.52
old_dep	43.80	46.60	49.20	49.05	51.40	52.30	2.63	5.36
gender_gap	7.10	7.55	8.10	8.10	8.40	10.10	0.68	8.37
neet	10.50	11.20	13.30	12.84	13.95	15.40	1.57	12.24
t_edu	23.10	25.80	27.90	27.40	29.00	29.20	1.81	6.62
dig_bank	48.08	50.45	57.72	55.00	57.99	58.20	3.97	7.22
arop	11.40	12.05	12.30	12.99	14.30	14.50	1.19	9.13
life_exp	74.60	77.00	77.40	77.32	78.20	78.60	1.22	1.57
immigration_rate	0.01	0.01	0.01	0.01	0.01	0.02	0.00	11.61
oop_exp	18.71	18.91	19.26	19.42	20.15	20.15	0.55	2.83

In contrast, Slovakia presents significantly lower absolute volumes. As expected, Premium Growth is stable and remains on a relatively low level, ranging from ~-16% to a max of ~6%. Interestingly, the Expense Ratio in Slovakia (mean ~24%) appears higher than in Italy (~5%), likely due to the lack of economies of scale that larger markets benefit from. This distinction validates the necessity of including country-specific covariates in the LMM to control for market size and maturity.

From the Socio-Demographic point of view, Italy presents a chronical and structural rigidity. Comparing the average values, Italy shows worse values for almost all social indicators: it displays a notably higher old-age dependency ratio (old_dep: 57.02), a wider gender gap (19.51), and a higher share of NEETs (20.08) compared to Slovakia. Crucially, in Italy, these socio-demographic variables present near-zero volatility (with CVs ranging between 0.98% and 1.87%), indicating that these are deep-rooted, chronic national characteristics rather than temporary fluctuations.

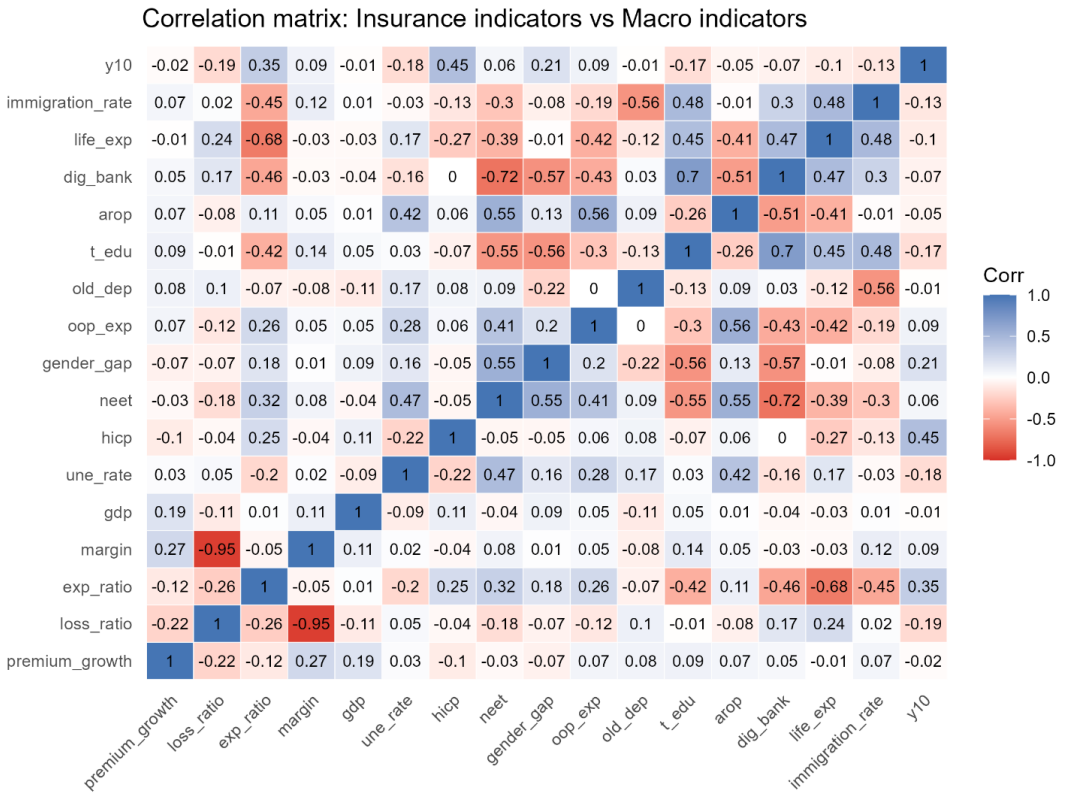
Conversely, Slovakia portrays a more dynamic social framework with a significantly lower gender gap (8.10) and fewer NEETs (12.84). Furthermore, from a structural standpoint,

Slovakia shows a higher propensity for digital adoption, with a digital banking penetration (dig_bank) averaging 55.00%, outperforming the Italian average of 44.50%.

2.4.3 Correlation Analysis

To investigate the relationships between the continuous variables and detect potential multicollinearity, a correlation matrix was computed. Figure below visualizes these relationships using a correlation plot.

Figure 2.4.3-1: Correlation matrix. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.



The correlation analysis reveals several predictable patterns. As expected, a strong negative correlation (-0.95) is observed between Loss Ratio and Margin, consistent with the mathematical expression behind their computation. Interestingly, Expense Ratio is highly negatively correlated with Life Expectancy (-0.68), Digital Banking (-0.46) and Immigration Rate (-0.45): this acts as a proxy for structural market maturity, suggesting that economies with higher life expectancies benefit from greater operational efficiency and digitalization, rather than implying a spurious direct causal link. Furthermore, macroeconomic variables such as GDP show a positive correlation with premium growth (0.19), supporting the hypothesis that economic growth drives insurance demand (Outreville 2013). Conversely, variables like Unemployment Rate tend to show a weak relationship with premium growth

(0.03), reflecting the complex and often counter-cyclical nature of Life insurance products, which are sometimes viewed as discretionary or precautionary savings.

The visual inspection of the correlation matrix highlights meaningful economic relationships. Specifically:

- A positive correlation is evident between gdp and premium_growth (0.19), confirming that economic expansion fuels the capacity for savings and insurance investment. Conversely, the une_rate (unemployment rate) exhibits a near-zero correlation with premium growth (0.03). This suggests that the theoretical expectation—where higher unemployment reduces household disposable income and contracts Life insurance demand—is mitigated by other structural factors in the current dataset.

- The structural nature of the industry is confirmed by the almost perfect inverse relationship between margin and loss_ratio (-0.95). However, the exp_ratio shows a much weaker correlation with the macroeconomic cycle (e.g., 0.01 with gdp), suggesting that operational efficiency is heavily dependent on company-specific management and structural market maturity rather than on external economic fluctuations.

- The Harmonised Index of Consumer Prices (hicp) shows a nuanced relationship. While moderate inflation typically accompanies growth, the matrix reveals a negative correlation (-0.10) between hicp and premium_growth. This indicates that inflationary spikes tend to erode the real purchasing power of households, effectively displaying a negative effect on the growth rate of Life insurance premiums.

- Finally, the correlation matrix reveals a striking negative relationship (-0.72) between Digital Banking and the share of NEET. This role of digitalization as a proxy for comprehensive modernization is further confirmed by its relationships with other key demographic and structural indicators. Specifically, Digital Banking shows a strong positive correlation of 0.70 with tertiary education, confirming that digital financial penetration thrives in highly educated societies. Moreover, it exhibits a significant negative correlation of -0.57 with the gender employment gap, indicating that economies with broad technological adoption also tend to foster greater gender parity in the labor market. Additionally, the negative association with out-of-pocket health expenditure (-0.43) suggests that digitally advanced nations generally benefit from more robust public welfare systems or more efficient private health

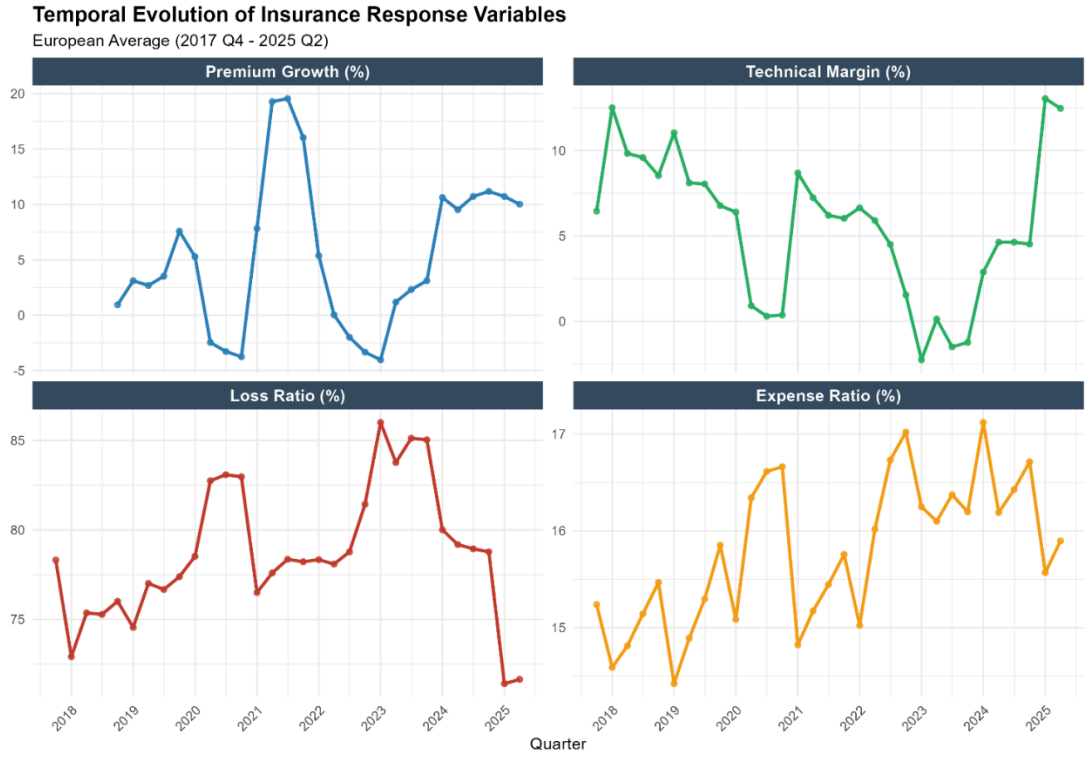
insurance networks, thereby reducing the direct financial burden on citizens. Collectively, these strong correlations serve as a bridge between the demographic and structural dimensions of the dataset. They suggest that a higher penetration of digital financial services is characteristic of more dynamic, modern economies where educational systems are robust, and labor markets are inclusive and capable of absorbing the younger population. Consequently, broad technological adoption (acts not merely as a structural indicator of the banking and insurance ecosystem, but as a powerful proxy for broader social mobility, gender equality, and systemic welfare efficiency.

In summary, the correlation matrix highlights that the Life insurance sector is governed by a dual dynamic. On one hand, internal technical performances (such as Margin and Loss Ratio) are dictated by strict, structural industry mechanisms rather than short-term macroeconomic fluctuations. On the other hand, the sector's operational efficiency and capacity for growth are deeply embedded in the broader socio-demographic fabric. Indicators such as the Expense Ratio and Digital Banking transcend their technical definitions, emerging as powerful proxies for a country's systemic maturity, social inclusivity, and economic modernization. Ultimately, these findings confirm that a resilient and efficient insurance market is not an isolated financial ecosystem, but rather a direct reflection of the structural development of the society it serves.

2.4.4 Temporal Evolution and Seasonality

Given the longitudinal nature of the panel dataset, which covers 31 consecutive quarters from 2017 Q4 to 2025 Q2, analyzing the temporal evolution of key variables is essential to identify underlying trends, cyclical behaviors, and structural breaks.

Figure 2.4.4-1: Temporal Evolution of Insurance Response Variables. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.



To provide a comprehensive yet detailed snapshot of the Life insurance sector's temporal dynamics, the cross-sectional European averages of the four response variables were plotted using a multi-panel visualization.

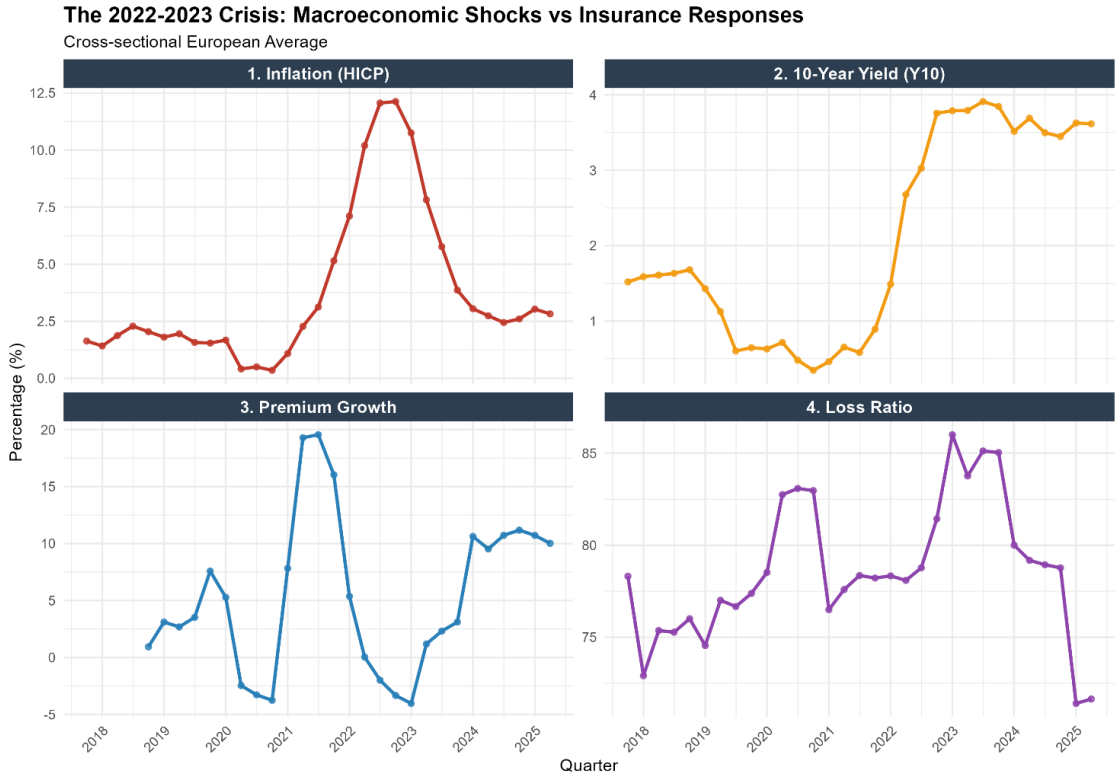
Following this peak, the 2022-2023 period reveals a drastic contraction in premium collection, which perfectly mirrors an unprecedented spike in the Loss Ratio. To validate the hypothesis that this specular shock is exogenously driven, we can visually correlate these technical responses with the prevailing macroeconomic environment (HICP and Y10). As inflation (HICP) surged, eroding households' real disposable income, the capacity for discretionary savings collapsed, driving Premium Growth into negative territory. Concurrently, the aggressive rate-hiking cycles initiated by central banks caused a rapid steepening of the 10-Year Government Bond yields (y10). This created a strong incentive for policyholders to surrender their existing Life contracts in favor of more profitable short-term sovereign debt, mechanically triggering a massive wave of claims (surrenders) that pushed the average Loss Ratio to historical highs.

Premium Growth increases again only in the last periods of observations: on the macroeconomic front, the gradual normalization of inflation (HICP) restored a portion of households' real purchasing power, re-enabling savings. Simultaneously, as the 10-Year Government Bond yields (y10) stabilized following the halt in the central banks' aggressive

rate-hiking cycles, insurance providers were able to adjust their product portfolios, offering more competitive guaranteed returns to match the new interest rate environment (Swiss Re Institute 2023; EIOPA 2024).

In contrast, the Expense Ratio trajectory shows a marked seasonality, probably due to the nature of contractual expenses (numerator effect). The panel also reveals that the level of expenses is increasing over time (drift), hence showing a slightly non-stationary time-series, suggesting that operational costs are influenced by external factors, mainly inflation.

Figure 2.4.4-2: Macroeconomic Shocks vs Insurance Responses. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.



As explicitly illustrated by the multi-panel chart, the temporal alignment between exogenous macroeconomic shocks and internal insurance dynamics is striking. The 2022-2023 period reveals a drastic contraction in Premium Growth, which perfectly mirrors an unprecedented spike in the Loss Ratio. This specular shock was fundamentally driven by the prevailing macroeconomic environment. As inflation (HICP) surged to historical peaks, it severely eroded households' real disposable income, collapsing the capacity for discretionary savings. Concurrently, the aggressive rate-hiking cycles initiated by central banks caused a rapid steepening of the 10-Year Government Bond yields (y10). This forward-looking expectation of higher yields created a strong "substitution effect": policyholders were heavily incentivized to surrender their existing Life contracts (which carried older, lower guaranteed rates) in favor of more lucrative short-term sovereign debt.

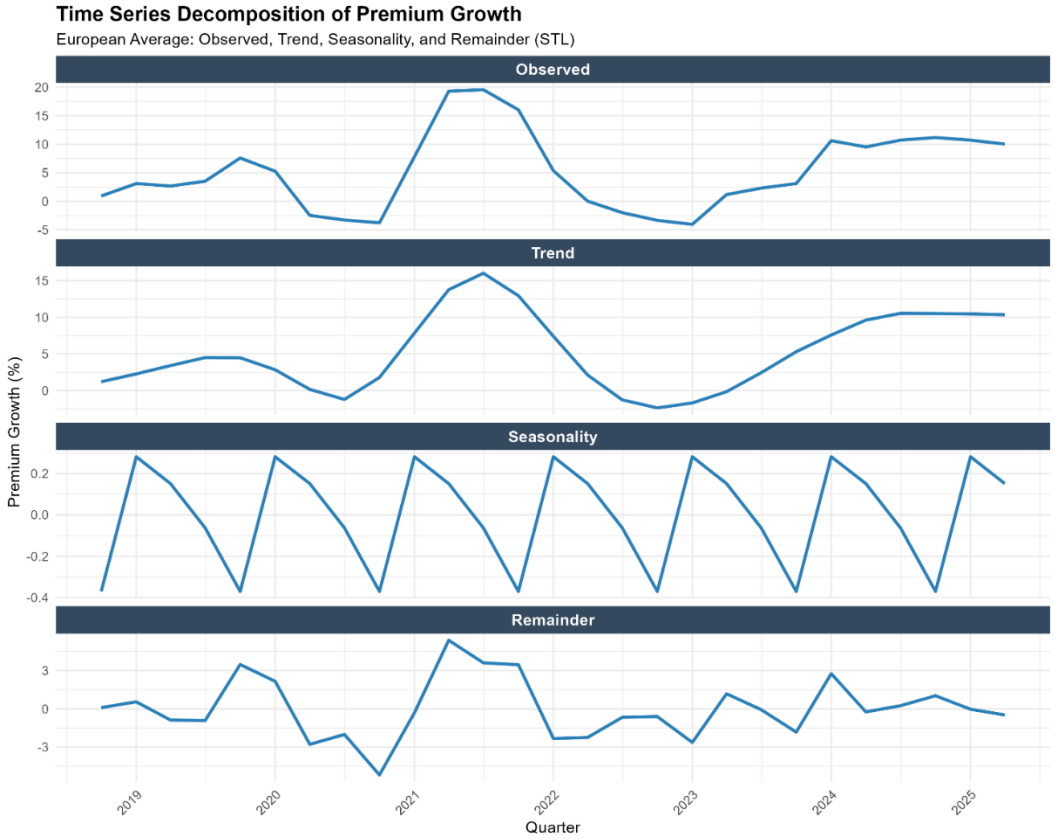
In the Life sector, this massive wave of early surrenders mechanically translates into an explosion of paid claims, pushing the average Loss Ratio to historical highs.

However, the final period of observation (2024-2025) demonstrates the sector's structural resilience following the normalization of these shocks. As inflation stabilized, a portion of households' purchasing power was restored, re-enabling retail savings flows. Simultaneously, while the Y10 yields plateaued at a relatively high level, the cessation of continuous rate hikes allowed the market to find a new equilibrium. Insurance providers adapted to this "higher-for-longer" regime by adjusting their product portfolios, offering new Life policies with competitive guaranteed returns that successfully rivaled alternative financial instruments (Swiss Re Institute 2023, EIOPA 2024).

Most notably, the stabilization of the yield curve triggered a drastic collapse in the Loss Ratio. This sharp decline indicates the exhaustion of the lapse risk wave: the cohort of policyholders highly sensitive to interest rate arbitrage had already surrendered their policies during the 2023 panic. With the portfolio stabilized and new, higher-yielding products retaining capital, the surrender rate physiologically plummeted, instantly restoring the sector's technical margins.

To further isolate the impact of macroeconomic shocks, a Seasonal-Trend decomposition (STL) was applied to the Premium Growth and Expense Ratio series.

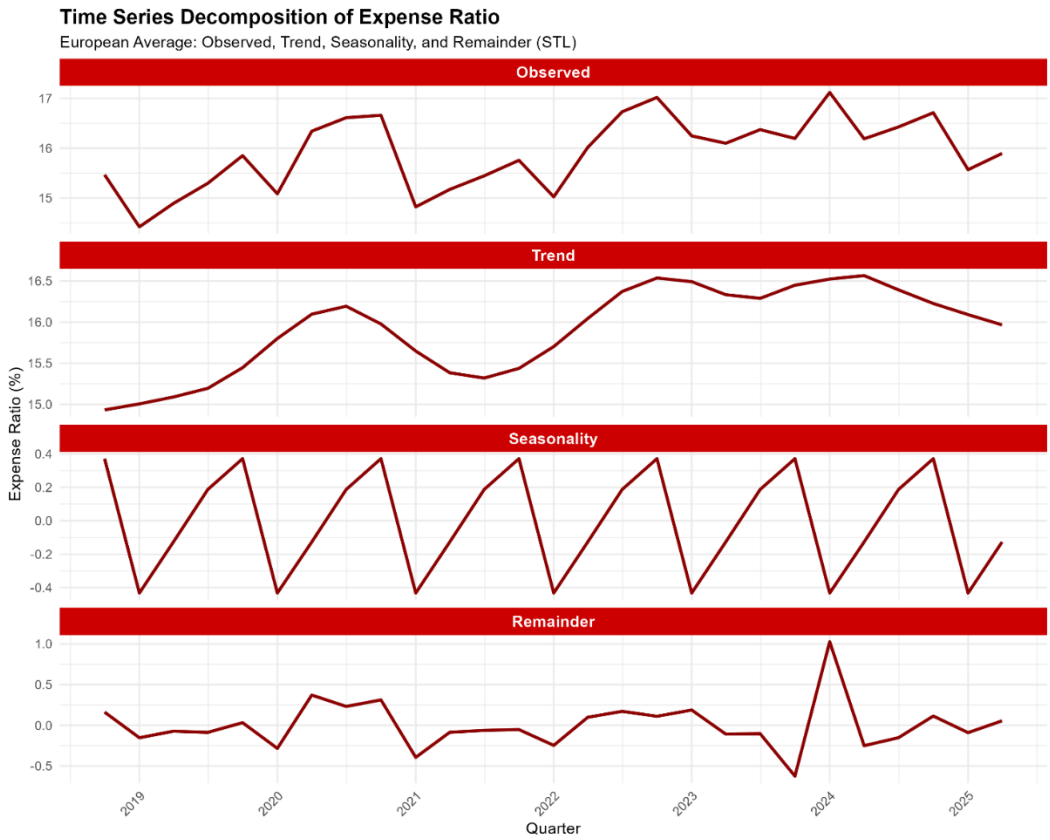
Figure 2.4.4-3: Time Series Decomposition of Premium Growth. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.



By looking at the scales on the y-axis, it is clear that the **Trend** component drives most of the variable's fluctuations. Ranging from negative values to a peak above 15%, the trend captures the main market phases: the 2021 post-pandemic savings boom, the sharp drop during the 2022-2023 macroeconomic crisis, and the final stabilization. On the other hand, the **Seasonality** component confirms a recurring calendar effect, with regular drops in Q1 and Q3 and peaks in Q4. However, its overall impact (varying by less than 1 percentage point) is very small compared to the macroeconomic trend. This shows that Life insurance

premium collection is primarily influenced by the broader economic cycle rather than just seasonal variations.

Figure 2.4.4-4: Time Series Decomposition of Expense Ratio. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.



A similar STL decomposition was performed on the Expense Ratio. Unlike Premium Growth, the variations in the Expense Ratio are strongly driven by both its Trend and Seasonal components. The Seasonality panel shows a very clear and predictable 'sawtooth' pattern, which explains much of the quarter-by-quarter volatility. This regular fluctuation is largely due to the timing of administrative reporting, annual bonuses, and accounting settlements. At the same time, the Trend component reveals a persistent upward movement, shifting the baseline of the Expense Ratio from approximately 15.0% to 16.5% over the observed period. This steady increase confirms that the prolonged inflationary environment has permanently raised the fixed operating costs for insurance companies, reducing their overall operational efficiency regardless of short-term seasonal cycles.

2.4.5 Summary of EDA Findings

The exploratory analysis confirms that the dataset, following the applied preprocessing and imputation protocols, is robust and highly suitable for advanced econometric modeling. The final panel successfully integrates 899 observations across 29 European countries.

Methodological choices, such as the duplication of annual covariates and the application of a "Last Observation Carried Forward" (LOCF) approach, effectively aligned structural socio-demographic factors—like education and digital banking penetration—with the quarterly frequency of the response variables without compromising data integrity.

The descriptive and correlation analyses yielded three primary conclusions that will guide the LMM estimation:

- The European Life insurance landscape is highly fragmented. The comparative analysis of market archetypes—such as the mature, high-volume Italian market versus the smaller, developing Slovakian market —demonstrates distinct operational scales, varying levels of structural rigidity, and different sensitivities to demographic factors.
- The correlation matrix highlighted that variables defining the sector's operational efficiency and capacity for growth (such as the Expense Ratio and Premium Growth) act as powerful proxies for a country's systemic maturity, showing deep connections with macroeconomic cycles and socio-demographic indicators like Digital Banking penetration.
- The temporal analysis explicitly confirmed that the European insurance sector does not operate in a static environment. The 2022-2023 macroeconomic crisis triggered massive systemic shocks, highlighted by the collapse in premium collection together with the spike in surrenders, while the inflationary pressure introduced a permanent upward drift in operational costs.

These three empirical findings confirm that a simple pooled regression would likely mask underlying structural divergences and fail to capture the temporal shifts in the economy. Consequently, this Exploratory Data Analysis validates the necessity of employing a LMM. By incorporating country-specific covariates and allowing for unobserved heterogeneity, the LMM is an optimal method to identify the distinct regimes that govern the complex dynamics of the European Life insurance market over time.

3 METHODOLOGY: THE LATENT MARKOV FRAMEWORK

The empirical analysis presented in this study was conducted using the R statistical computing environment (R Core Team 2024). Specifically, the estimation of the Latent Markov Models (LMM) was performed using the LMest package (Bartolucci, Pandolfi, and Pennoni 2017). This package provides a comprehensive framework for the analysis of longitudinal data, implementing the Expectation-Maximization (EM) algorithm for maximum likelihood estimation. The LMest library was selected for its flexibility in handling continuous response variables and its capability to estimate both the measurement model (relating the latent states to the observed outputs) and the structural model (governing the temporal evolution of the latent process), as detailed in the following sections.

3.1 INTRODUCTION TO MARKOV CHAINS AND HIDDEN MARKOV MODELS

The statistical analysis of longitudinal must be implemented through a model able to recognize the evolution of stochastic processes over time. In many empirical applications, particularly within econometrics and social sciences, the primary objective is to study the transitions of a specific phenomenon across discrete time points. A foundational approach to modeling such dynamics is the Markov chain, a stochastic process where the probability of a future state only depends on the current state, satisfying the Markov property (Maruotti 2011). However, standard Markov chain models rely on a critical assumption: that the state of the system is directly observable. This assumption is often weak in complex research contexts where the variables of interest are not observable but rather invisible and can only be inferred from a set of manifest indicators. For instance, there can be situations in which the true health status of a person is not known, and can only be assumed by looking at the individual interview or a questionnaire.

Consequently, relying exclusively on observable Markov chains can lead to significant inference errors. As noted by Dotto (n.d.), concepts like material deprivation are latent in nature; researchers observe binary or ordinal indicators (e.g., the inability to afford heating, arrears on utility bills), but these can be considered as proxies for the underlying condition of poverty. Modeling the transitions of these observed variables directly ignores measurement error and fails to distinguish between real structural changes in the subject's status and mere fluctuations in the reporting of specific items. To address this limitation, the Hidden Markov Model (HMM) framework introduces an architecture that separates the unobserved dynamics from the observed data.

In this context, the HMM assesses that the observed sequence of data depends on an unobserved (latent) stochastic process. As described by Pandolfi et al. (2023), the model consists of two distinct components: a latent model, which describes the evolution of the unobserved characteristic over time, and a measurement model, which links the latent state to the manifest indicators. This separation is essential. For instance, when analyzing material deprivation, the HMM allows the researcher to define "deprivation" as a latent state that generates the probabilities of observing specific hardships. This implies that a subject can remain in a state of "severe deprivation" even if, due to measurement noise or temporary relief, one specific indicator changes from negative to positive. By shifting the focus from the observable chain to the latent one, the HMM provides a rigorous mechanism to correct for selection bias and measurement error, hence offering a more accurate representation of the subject's true status (Pandolfi et al. 2023; Dotto n.d.).

Furthermore, the necessity of this approach becomes even more clear when dealing with continuous longitudinal data subject to complex missingness patterns. Traditional methods often require list-wise deletion (deleting the entire observation if a value of a variable is missing) or assume data are Missing Completely at Random (MCAR), i.e. data are missing for a human mistake; however, these assumptions are rarely met in panel surveys. The HMM framework can still accommodate "Missing-At-Random" (MAR) mechanisms where missingness depends on the latent state itself (Pandolfi et al. 2023). This capability ensures that the estimation of the underlying phenomenon remains unbiased even when subjects drop out or fail to respond to specific items, a common occurrence in longitudinal studies of vulnerable populations.

3.2 LATENT MARKOV MODEL (LMM) FOR PANEL DATA

Building upon the general HMM structure, the Latent Markov Model (LMM) is specifically adapted for the analysis of longitudinal panel data, where a sample of N subjects is observed over T discrete time occasions. This extension is particularly powerful for capturing unobserved heterogeneity and temporal dynamics simultaneously. Let $\boldsymbol{Y_{it}}$ denote the vector of response variables for **subject** $\boldsymbol{i}$ at **time** $\boldsymbol{t}$. The core assumption of the LMM is that the distribution of these observed variables is conditional on a time-varying latent variable U_{it} which follows a first-order Markov chain with a finite state space $\{1, \dots, k\}$ (Pandolfi et al. 2023; Maruotti 2011).

A distinguishing feature of the LMM applied to panel data is its ability to model the temporal evolution of the latent trait explicitly. Unlike static latent class models, which cluster subjects based on cross-sectional similarities, the LMM focuses on the transitions between states. This is critical for analyzing phenomena such as material deprivation, where the research interest lies not only in counting the poor but in understanding the dynamics of poverty persistence and social mobility. For example, the model can identify distinct latent profiles ranging from "well-being" to "severe deprivation" and, crucially, estimate the probability of a subject moving from a state of well-being at time $t-1$ to a state of poverty at time t (Dotto n.d.).

The dynamic component of the LMM is ruled by two fundamental sets of probabilities that define the latent structure:

Initial Probabilities $(\boldsymbol{\pi_u})$: Defined as $P(U_{i1} = u)$, these probabilities represent the likelihood of a subject belonging to latent state u at the very start of the observation period $(t = 1)$. This vector captures the baseline heterogeneity of the population before any temporal evolution occurs.

Transition Probabilities ($\boldsymbol{\pi_{uv}}$): Defined as $\pi_{uv} = P\left(U_{it} = v \middle| U_{i,t-1} = u\right)$, these probabilities form the transition matrix that condenses the temporal dynamics of the process. This matrix is the core engine of the LMM, describing the probability of switching from state u at the previous occasion to state v at the current occasion (Pandolfi et al. 2023).

The interpretation of the transition matrix provides deep insights into the structural nature of the phenomenon under study. The diagonal elements (π_{uu}) represent the probability of persistence— an immobility within the same latent state. In the context of economic analysis, high probabilities on the diagonal for a "deprived" state indicate the presence of a "poverty trap," suggesting that once a subject enters a state of deprivation, it is structurally difficult to exit (Dotto n.d.). On the other hand, the off-diagonal elements $(\pi_{uv}, where\ u \neq v)$

measure mobility between states. Significant probabilities in these cells would indicate fluid transitions, suggesting that the deprivation experienced by the population might be temporary rather than chronic.

Moreover, the LMM framework for panel data is highly flexible regarding the distribution of the response variables. While early applications focused on categorical data, recent advancements allow for continuous responses (e.g., income, BMI) using multivariate Gaussian distributions within the measurement model (Pandolfi et al. 2023). This flexibility extends to the handling of "borrowing strength" across time and units, meaning that a specific data point can be interpreted in the context of other data related to the same unit (across time) and corresponding data from other units at the same time. As highlighted by Bertarelli (2015) in the context of aggregate data, the dynamic nature of the latent process allows the model to use information from adjacent time points to stabilize estimates. This is particularly beneficial in panels with irregular sampling or sparse data, as the Markov structure implies that the state at time t is informative about the state at time $t + 1$, effectively smoothing the inferred trajectories of the subjects.

In summary, the LMM transforms the analysis of panel data from a static comparison of snapshots into an identification of trajectory pattern. Unlike observable Markov chains, this approach uses the transition matrix to separate true state dependence from unobserved heterogeneity, leading to a more accurate understanding of social mobility.

3.3 FOCUS ON THE ROLE OF COVARIATES IN LMM

While the basic Latent Markov Model provides a powerful framework for analyzing dynamic processes, it assumes a certain degree of homogeneity among subjects, postulating that the transition probabilities are constant across the population. However, in most socio-economic and epidemiological applications, this assumption is often unrealistic. Individuals differ in their demographic profiles, educational backgrounds, and geographic locations, all of which may significantly influence their latent transitions. Consequently, incorporating individual-specific covariates X_{it} into the model structure is essential to account for observed heterogeneity and to move from a purely descriptive analysis to a causal inference framework (Maruotti 2011). Formally, it is crucial to distinguish between covariates affecting the **latent model**—which governs the distribution and evolution of the unobserved construct—and those affecting the **measurement model**—which links the latent states to the observed responses. As outlined by Bartolucci et al. (2013), this separation allows researchers to unravel whether individual characteristics influence the true underlying status of a subject or merely the way that status is manifested in the observed data.

Within the latent model, the inclusion of covariates typically occurs at two distinct levels. First, covariates influence the initial probabilities, π_u, allowing the researcher to model the baseline heterogeneity at the beginning of the observation period ($t = 1$). Through a multinomial logit parameterization, one can estimate how factors such as gender or region affect the likelihood of starting in a specific latent state, such as "severe deprivation" versus "well-being" (Pandolfi et al. 2023). Secondly, and perhaps most critically for longitudinal analysis, covariates within the latent model affect the transition probabilities. This parametrization allows the model to identify the risk factors associated with state-switching. As described by Pandolfi et al. (2023), the probability of transitioning from state u to state v is modeled as a function of time-varying or time-constant covariates:

$$\log \frac{P(U_{it} = v | U_{i,t-1} = u, \boldsymbol{X_{it}})}{P(U_{it} = u | U_{i,t-1} = u, \boldsymbol{X_{it}})} = \boldsymbol{x_{it}^{\top} \gamma_{uv}}$$

In this context, the coefficients γ_{uv} quantify the effect of a predictor on the log-odds of moving to state v relative to remaining in state u. For instance, this specification allows researchers to test whether an increase in income (covariate) significantly reduces the probability of transitioning into a "poverty" state, or whether higher education acts as a protective factor against "downward mobility" (Colombi et al. 2024).

Finally, concerning the measurement model, recent methodological developments emphasize the importance of accounting for Differential Item Functioning (DIF). Standard HMMs assume measurement invariance, implying that the probability of observing a specific response given the latent state is the same for all individuals. However, as argued by Dotto (n.d.), this assumption may be violated if the relationship between the latent trait and the observed response varies across sub-populations due to cultural or structural differences. By including covariates Z_{it} directly in the measurement model, the LMM corrects for this reporting bias. This ensures that the estimated latent states reflect the true underlying construct rather than artifacts of the respondent's demographic profile, distinguishing, for example, between true economic deprivation and culturally driven consumption preferences.

3.4 ESTIMATION AND MODEL SELECTION

The estimation of Latent Markov Models presents specific computational challenges due to the presence of unobserved variables (U_{it}). Since the latent states are not directly observed, the complete-data likelihood function cannot be maximized analytically. Instead, the standard approach relies on the **Expectation-Maximization (EM) algorithm** (Dempster et al., 1977), an iterative procedure that alternates between two steps. In the *Expectation* (E) step, the algorithm computes the expected value of the log-likelihood given the current parameter estimates and the observed data, effectively reconstructing the posterior probabilities of the latent states. Subsequently, in the *Maximization* (M) step, these expected values are used to update the model parameters (Pandolfi et al., 2023; Bertarelli, 2015). A key advantage of this framework is its ability to handle missing data under the Missing-At-Random (MAR) assumption. As noted by Pandolfi et al. (2023), the likelihood function can be defined by integrating over the missing components, allowing the model to utilize all available information without discarding subjects with incomplete records.

Once the estimation framework is established, a fundamental issue is model selection, which primarily concerns the determination of the optimal number of latent states k. This choice is critical: specifying too few states may obscure relevant sub-populations (underfitting), while specifying too many can lead to the identification of spurious clusters and difficult interpretation (overfitting) (Pennoni et al., 2024).

The primary metric resulting from this maximization process is the **Log-Likelihood** ($\boldsymbol{l(\hat{\theta})}$). Formally, this value represents the natural logarithm of the likelihood function evaluated at the estimated parameters $\hat{\theta}$. It serves as a measure of "goodness-of-fit," quantifying how well the model describes the observed data; a higher (less negative) Log-Likelihood indicates that the model is more likely to have generated the observed dataset. However, in the context of model selection, the Log-Likelihood cannot be used in isolation. Since it is a non-decreasing function of model complexity, adding more latent states (k) or covariates will strictly increase the value of $l(\hat{\theta})$, regardless of whether the additional complexity provides a meaningful improvement (Maruotti 2011).

Consequently, relying solely on the maximization of the Log-Likelihood would inevitably lead to overfitting, selecting the most complex model available. To address this trade-off between goodness-of-fit (maximized l) and model parsimony, Information Criteria are universally employed.

While the **Akaike Information Criterion** (**AIC**) is a common option, the **Bayesian Information Criterion (BIC)** is generally preferred in the context of mixture models due to its stricter penalty for model complexity. The BIC is defined as:

$$BIC = -2l(\hat{\theta}) + p\ln(n)$$

where $l(\hat{\theta})$ is the maximized log-likelihood obtained from the EM algorithm, p is the number of free parameters, and n represents the sample size (Pandolfi et al., 2023; Maruotti, 2011). The term $-2l(\hat{\theta})$ rewards the model for fitting the data well, while the term $p\ln(n)$ imposes a penalty that grows with the number of parameters. The optimal model is typically the one that minimizes the BIC value, identifying the solution that best explains the data with the fewest number of latent states.

4 RESULTS

4.1 MODEL WITHOUT COVARIATES

The first stage of the empirical analysis involves the estimation of a homogeneous Latent Markov Model (LMM) without covariates. As outlined in the methodology, this baseline model serves a dual purpose: first, to identify the optimal number of latent states ($\boldsymbol{k}$) that best describe the heterogeneity of the European Life insurance market; and second, to analyze the intrinsic temporal dynamics of these states before introducing external socioeconomic drivers.

The model assumes that the vector of response variables $\boldsymbol{Y_{it}}$ (Premium Growth, Loss Ratio, Expense Ratio, and Technical Margin) is conditional on a latent process following a first-order Markov chain. By excluding covariates at this stage, the resulting clusters are determined solely by the performance and technical stability of the insurance undertakings.

4.1.1 Model Selection

The selection of the optimal number of latent states, denoted $\boldsymbol{k}$, is a critical step in the LMM framework. The goal is to balance goodness-of-fit and structural validity, without building a too complex model. To determine the optimal $\boldsymbol{k}$ the model was estimated for a range of states $k = 2, \dots, 5$. The selection process relied on the analysis of Information Criteria (BIC, AIC, LogLik) and, crucially, on the size of the resulting latent classes.

Table 4.1.1-1: Information criteria comparison. Author's calculation on EIOPA (2025) data.

Information criteria comparison (k)

Latent States	Log-Likelihood	BIC	AIC	Min. Class Size (%)
k = 2	-2688.65	5448.02	5419.31	43.05
k = 3	-2674.51	5450.04	5409.02	0.33
k = 4	-2634.37	5406.81	5350.75	2.22
k = 5	-2533.43	5248.70	5174.87	0.33

The selection of the optimal number of latent states involves a careful trade-off between mathematical fit and the structural validity of the resulting clusters. Table 4.1.1-1 summarizes the goodness-of-fit statistics for k values ranging from 2 to 5. From a strictly computational standpoint, the model with $k = 5$ states appears superior, as it maximizes the Log-Likelihood (-2533.4) and achieves the lowest Bayesian Information Criterion (BIC = 5248.7).

However, as highlighted by Pennoni et al. (2024), relying solely on information criteria in complex panel datasets can lead to overfitting, where the model captures noise rather than systemic patterns. The rejection of these more complex models, despite their favorable information criteria, is justified by their lack of economic interpretability and structural fragility. As evidenced by the "Minimum Class Size (%)" column, models with 3, 4, or 5 states produce solutions where at least one latent state is populated by a negligible fraction of the sample—falling to a critical 0.33% for $k = 3$ and $k = 5$. In a panel of 899 observations, such a marginal representation indicates that these "ghost states" do not correspond to distinct economic regimes but serve merely to isolate a few extreme outliers.

In contrast, the model with $k = 2$ states shows remarkable structural integrity. While its BIC (5448.0) is higher than that of the 5-state model, it offers a highly balanced and symmetrical clustering structure, with a division of observations between State 1 (56.95%) and State 2 (43.05%). This distribution suggests a robust economic segmentation that allows for a coherent narrative regarding the European Life insurance market. Therefore, prioritizing parsimony and inferential stability over mere likelihood maximization, the 2-state model is selected as the optimal baseline for the analysis.

4.1.2 Economic interpretation

Having established $k = 2$ as the optimal number of latent states, we now proceed to characterize the economic nature of these regimes. The interpretation relies on the analysis of the conditional means ($\hat{\mu}$) of the response variables and the initial probabilities ($\hat{\pi}$) estimated at the beginning of the observation period. The estimation process begins by defining the market structure at the starting point of the time series (2017 Q4). In this section the vector of initial probabilities is shown, representing the likelihood of a country belonging to each state at $t = 1$.

Table 4.1.2-1: Initial probabilities. Author's calculation on EIOPA (2025) data.

Initial Probabilities

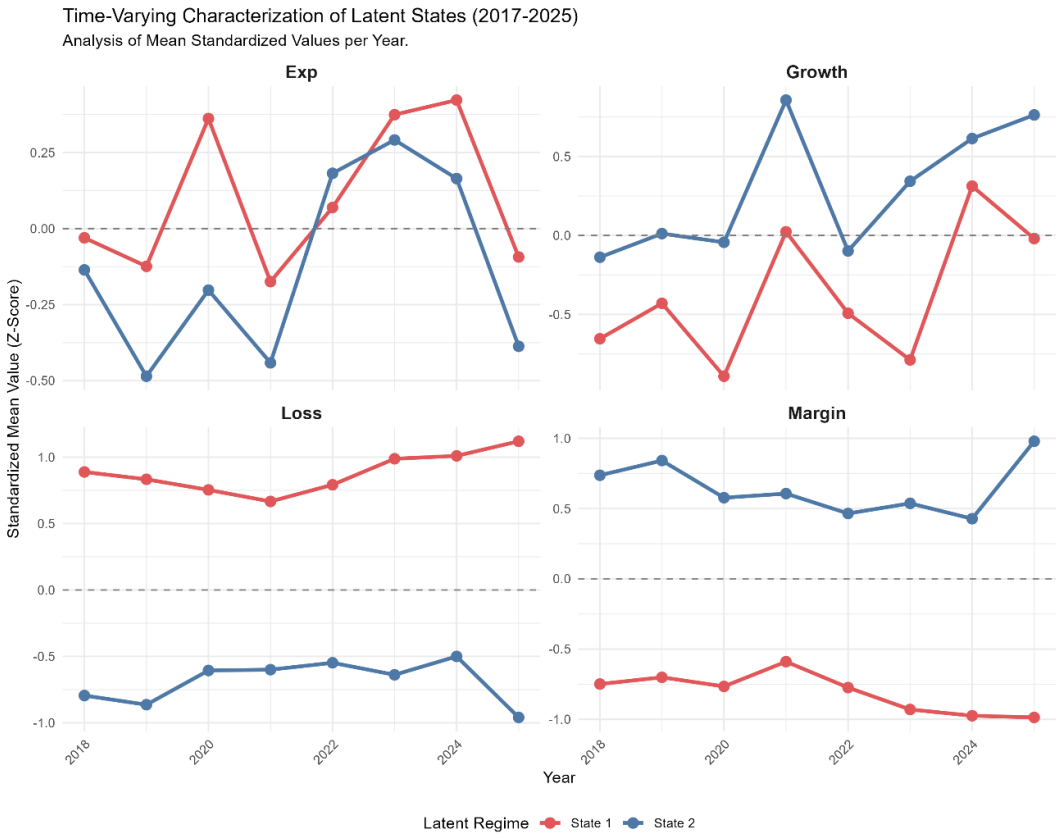
Regime	Probability
State 1	0.3873
State 2	0.6127

As shown in Table 4.1.2-1, the European market in late 2017 did not display a monolithic structure but was already segmented. A significant portion of the probability mass is

assigned to State 2, suggesting that this regime was already the prevalent condition for the core European markets before the pandemic and inflationary shocks.

To understand the economic meaning of the two identified regimes, we analyze the estimated conditional means ($\hat{\mu}$) of the standardized response variables for each state. This is useful to isolate external phenomena that may happen in specific time intervals.

Figure 4.1.2-1: Standardized Mean Values for responses of model with k = 2. Author's calculation on EIOPA (2025) data.



Based on the estimated parameters and the posterior probabilities, the two identified states exhibit distinct economic and technical profiles, reflecting a dichotomy between "Stability" and "Stress/Volatility" rather than a simple geographic divide.

- State 1: The "Volatile" Cluster: this state is characterized by extreme values in the response variables. Historically (2018-2021), it gathered emerging CEE markets (e.g., Bulgaria, Hungary) exhibiting high Premium Growth rates due to a catching-up effect. However, in the post-2022 period, this state evolved to capture markets under liquidity pressure. The defining feature becomes a deteriorated Loss Ratio, serving as a proxy for surrender risk and run-on-the-bank scenarios. Countries in this state experience heightened outflows (claims/surrenders) relative to premiums collected, often accompanied by volatile premium adjustments.

- State 2: The "Resilient" Cluster: this state represents the structural equilibrium of the European insurance market. It includes traditional markets (e.g., France, Spain, Belgium) characterized by moderate Premium Growth (typical of saturated markets), controlled Expense Ratios, and, crucially, a stable Loss Ratio. The persistence of countries in this state indicates a capacity to maintain technical profitability and contain lapse risk even during macroeconomic turbulence, likely due to more resilient product structures (e.g., profit participation mechanisms) or lower exposure to interest rate-driven shocks.

The estimated transition probability matrix ($\widehat{\Pi}$) reveals the temporal dynamics of these regimes.

Table 4.1.2-2: Transition probability matrix for model with k = 2. Author's calculation on EIOPA (2025) data.

Transition Probability Matrix

Regime	State 1	State 2
State 1	0.8504	0.1496
State 2	0.1104	0.8896

The estimated transition probability matrix provides crucial insights into the stability and dynamic nature of the identified regimes. A key finding is the high degree of persistence observed in both states, evidenced by dominant diagonal elements.

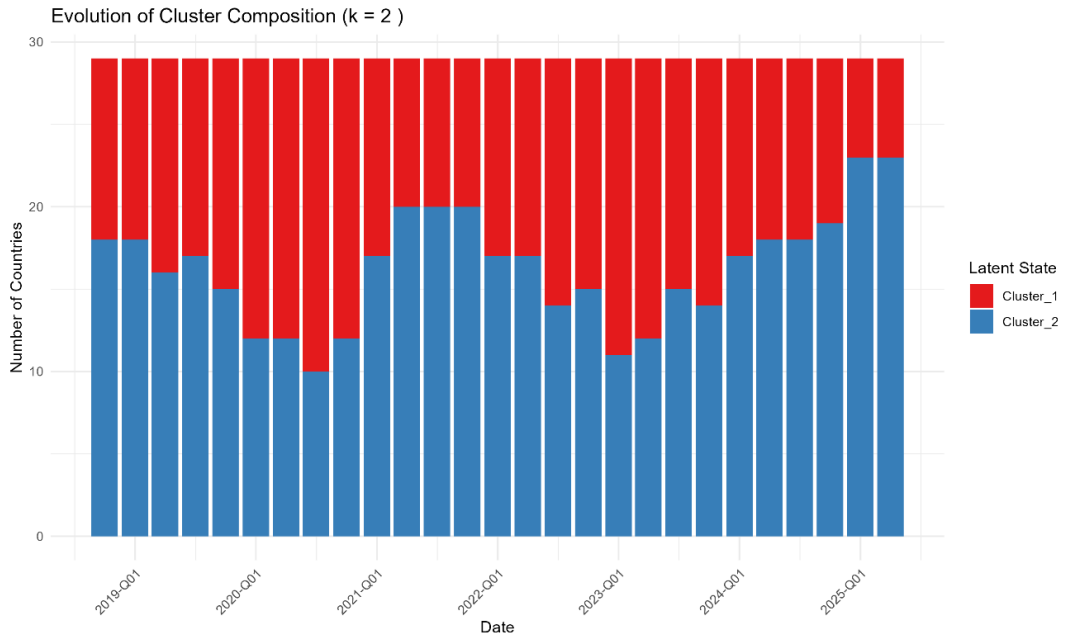
The probability of remaining in State 2 given that a market was previously in State 2 is historically high. This confirms that for the majority of the observation period, the European insurance market operates in a structural equilibrium characterized by low volatility and steady margins.

Crucially, the model estimates a significant persistence for State 1 as well. This implies that the "Volatile" state is not merely a transient shock (e.g., a one-quarter outlier) but a persistent economic regime. Once a market migrates into this state—driven by a surge in the Loss Ratio or Premium volatility—it tends to remain "trapped" there for consecutive quarters.

This is economically consistent with the nature of the surrender risk crisis observed in 2023-2024. Liquidity strains caused by interest rate hikes are not resolved instantaneously, instead, they require structural portfolio adjustments and time. Consequently, the

transition from State 2 to State 1 represents a structural break, while the reverse transition (recovery) is a slower process, explaining the accumulation of countries in the "Volatile" State towards the end of the sample.

Figure 4.1.2-2: Number of countries in Latent States. Author's calculation on EIOPA (2025) data.



By inspecting these trajectories over the observation period, we can distinguish between structural cluster membership and temporary regime shifts driven by macroeconomic shocks.

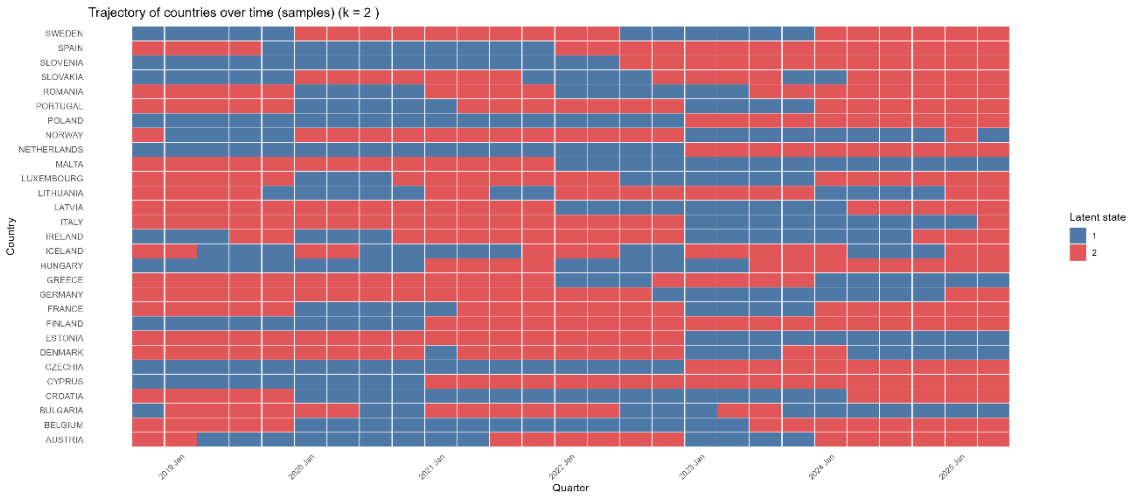
In stable market conditions, such as the pre-pandemic period of 2018-2021, the European landscape exhibits a clear structural dichotomy. The majority of the sample, comprising the "Resilient" State (2), includes mature markets like France, Spain, and Belgium. These markets are characterized by saturation, moderate growth, and, crucially, a controlled Loss Ratio, reflecting their deep integration into the real economy and stable liability structures (Swiss Re 2024).

Conversely, a subset of countries, including Ireland, the Baltic states, and CEE nations, tends to reside structurally in the "Volatile" State (1). This reflects distinctive drivers: for instance, Ireland acts as a financial hub with cross-border premium volumes often decoupled from local GDP, while CEE markets are undergoing a structural deepening phase, naturally exhibiting higher volatility in growth rates and claims compared to the Western European average (Allianz 2025).

The trajectory analysis reveals a critical "mass migration" in the 2023-2024 period, where previously stable countries shifted from the Resilient Core to the Volatile State. This shift

was not uniform: while France and Spain remained largely anchored in State 2, major markets like Italy and Germany migrated to State 1. This divergence suggests that the recent regime shift was driven by a specific vulnerability to the interest rate shock: markets with higher surrender sensitivity were pushed into the high-volatility cluster, aligning their statistical profile with that of emerging or financial-hub markets, although for different reasons.

Figure 4.1.2-3: Individual country trajectories across latent states (2017-2025). Author's calculation on EIOPA (2025) data.



Starting in Q1 2021, the market witnessed a consolidation of the "Resilient" State (2). Rather than shifting into volatility, major markets like France, Italy, and Germany firmly anchored themselves in the stable cluster. This alignment reflects the post-COVID economic recovery: the easing of lockdowns and the normalization of financial markets boosted the performance of Unit-Linked products (EIOPA 2021).

The most dramatic changes occur in the final quarters of the sample (2023-2025), where State 1 (the Volatile cluster) becomes the dominant regime. This phenomenon requires an adjusted interpretation. Unlike previous expansions driven by growth, this migration is driven by liquidity stress.

Recent industry dynamics suggest that the transition of mature markets (specifically Italy and Germany) into State 1 was triggered by a surge in surrender rates (lapses). As the European Central Bank raised interest rates to combat inflation, traditional savings products and government bonds became more attractive than life insurance policies, leading to a wave of withdrawals.

According to recent sector analyses (e.g., Fitch Ratings 2024), this "mass lapse" event caused a deterioration in the Loss Ratio for many insurers, forcing them to sell assets at a loss to

cover outflows. Consequently, the LMM identifies these markets as "Volatile" (State 1) not because they are growing aggressively, but because their liquidity indicators (Loss Ratio) spiked above the stability threshold, mimicking the volatility typically seen in emerging markets.

The trajectory analysis confirms that the latent states captured by the model represent dynamic economic phases rather than static geographic labels. State 2 acts as the "Safe Harbor", characterizing markets with stable liability structures, low surrender rates, and solid technical margins. State 1 emerges as a "Turbulence" regime, capturing both structurally volatile economies (high growth) and mature economies undergoing liquidity shocks (high lapses).

The "mass migration" observed in 2024 represents a structural break: the high-interest-rate environment effectively pushed the European "Core" out of its equilibrium, exposing it to liquidity risks previously associated only with less structured markets. In the next section, covariates will be included in the model to explicitly isolate these external drivers (e.g., interest rates and inflation).

While the aggregate analysis highlights a general trend towards the expansionary regime, a granular examination reveals significant heterogeneity across national markets. The table below ranks the already mentioned European countries based on their persistence in the two States over the period.

Table 4.1.2-3: Frequency of Latent State Occupancy by country. Author's calculation on EIOPA (2025) data.

Frequency of Latent State Occupancy by Country (%)

Country	State 1 (%)	State 2 (%)
CYPRUS	33.3	66.7
DENMARK	33.3	66.7
FINLAND	33.3	66.7
FRANCE	33.3	66.7
GERMANY	33.3	66.7
GREECE	33.3	66.7
ITALY	33.3	66.7
LATVIA	33.3	66.7
LUXEMBOURG	33.3	66.7
PORTUGAL	33.3	66.7
SPAIN	33.3	66.7
ESTONIA	37.0	63.0
ROMANIA	37.0	63.0

Country	State 1 (%)	State 2 (%)
LITHUANIA	40.7	59.3
SLOVAKIA	40.7	59.3
SWEDEN	40.7	59.3
AUSTRIA	48.1	51.9
BULGARIA	48.1	51.9
ICELAND	48.1	51.9
IRELAND	48.1	51.9
NORWAY	48.1	51.9
BELGIUM	51.9	48.1
MALTA	51.9	48.1
HUNGARY	55.6	44.4
SLOVENIA	55.6	44.4
CROATIA	63.0	37.0
CZECHIA	63.0	37.0
NETHERLANDS	63.0	37.0
POLAND	63.0	37.0

Table 4.1.2-3 visually identifies different behaviors. The upper section of the table includes markets that spent most of the timeframe (approximately 67%) in the Resilient regime (State 2). As expected, this group comprises the Eurozone's leaders: Italy, Germany, and France, alongside stable Nordic economies like Denmark and Finland.

The high persistence of these countries in State 2 confirms that the "Resilient" regime is the standard equilibrium for the integrated European market. It is important to note that while the trajectory analysis (Figure 4.1.2-3) showed Italy and Germany migrating to the Stress state in the very last quarters (2023-2025), their aggregate history places them firmly in the Resilient Market. This indicates that their recent volatility is a specific reaction to the interest rate (or other external) shock rather than a structural characteristic of their markets.

The lower section lists countries with a higher prevalence of the Volatile/Stress regime (State 1). Contrary to the core markets, this group is dominated by:

- Countries like Poland, Czechia, and Croatia (showing >60% persistence in State 1), belonging to the Central and Eastern Europe (CEE). For these economies, the "Stress" state likely reflects high-growth volatility typical of catching-up markets rather than liquidity crises.

- The Netherlands also appears in this volatile cluster (63% in State 1). Unlike the CEE countries, this placement is likely driven by the Dutch market's unique sensitivity to interest rate movements and technical provisions, which generates volatility in the Loss Ratio similar to that of emerging markets, although for different structural reasons.

In conclusion, the table demonstrates that while the temporal trend pushes towards convergence during crises, structural differences remain. The European market is dichotomous: a synchronized "Resilient Core" (Western Europe and Nordics) that generally maintains stability but is vulnerable to severe liquidity shocks, and a "Structurally Volatile" periphery (Eastern Europe) driven by dynamic expansion.

4.2 MODEL WITH COVARIATES

Following the comprehensive baseline analysis presented in the previous section, the second stage of the empirical strategy introduces a set of exogenous covariates into the LMM framework. The primary objective of this methodological advancement is to understand precisely how the underlying macroeconomic environment, coupled with structural socio-demographic factors, drives the technical performance and the overall stability of the European Life insurance sector. In this advanced specification, the selected covariates are included directly in the measurement model. This means that the external variables are assumed to directly influence the conditional distribution of the response variables within each of the identified latent states, rather than affecting the transition probabilities between the states themselves.

4.2.1 Model Selection

Similarly to the baseline approach adopted for the homogeneous model, the selection process for determining the optimal number of latent states (k) was meticulously repeated. The model was estimated for a range of possible states, specifically for $k = 2, \dots, 5$, while simultaneously including the full, comprehensive set of available covariates.

Table 4.2.1-1: Information criteria comparison (model with covariates). Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.

Information criteria comparison (k)

Latent States	Log-Likelihood	AIC	BIC
k=2	-4940.9	10131.7	10302.6
k=3	-4870.0	10112.1	10366.4
k=4	-4879.5	10257.0	10597.4
k=5	-4734.9	10097.9	10527.2

The empirical results derived from this stage confirmed and reinforced the structural tendencies already observed during the preliminary phase of the analysis. Here, BIC strictly preferred the two-state specification, returning the lowest value (BIC = 10,302.6) among all tested configurations. While AIC and Log-Lik mathematically suggested that a five-state model ($k = 5$) might provide a better fit, it is widely acknowledged in econometric literature that the these metrics tend to heavily favor over-parameterized models, especially when applied to complex longitudinal panel datasets. Relying solely on these metrics in this context would inevitably lead to severe overfitting, capturing statistical noise rather than genuine, systemic market patterns.

The decision to reject the $k = 3$ specification is further confirmed by the analysis of its estimated transition probability matrix. A fundamental requirement for a valid Latent Markov Model is that the identified states represent persistent and theoretically sound economic regimes. However, the $k = 3$ transition matrix revealed a critically unstable structure: State 1 exhibited a persistence probability of 0.00%, meaning that any country identified in this state would systematically migrate to a different regime in the subsequent quarter. Furthermore, the probabilities of entering State 1 from the other regimes are insignificant, standing at 0.14% from State 2 and 10.04% from State 3. Such a high exit rate coupled with a near-zero entry rate indicates that State 1 lacks structural 'mass' and does not correspond to a meaningful market condition. Instead, it functions as a temporary labeling state, likely capturing a few isolated outliers or statistical noise rather than a distinct economic phase. This evidence, combined with the previously mentioned convergence issues and singularity errors, confirms that increasing the number of states beyond two leads to severe overfitting without providing any additional inferential value. Consequently, the 2-state model is maintained as the only parsimonious and structurally valid representation of the European insurance dynamics

In order to structurally refine the $k = 2$ model and ensure the highest degree of statistical robustness, a rigorous backward step-wise selection procedure was implemented. This iterative analytical process systematically eliminated statistically non-significant variables from the initial pool of covariates. The procedure successfully returned an optimal, highly parsimonious formulation that retained seven fundamental drivers: Gross Domestic Product (GDP), Unemployment Rate, Old Dependency Ratio, Tertiary Education, Life Expectancy, Immigration Rate, and the 10-Year Government Bond Yield.

It is worth noting that the structural indicator for Digital Banking was excluded during the backward stepwise selection process. While this variable proved to be a significant proxy for modernization during the exploratory phase, its removal was dictated by the minimization of the BIC indicator, meaning that its inclusion provided marginal explanatory power relative to the increased model complexity. From an analytical standpoint, this exclusion is further justified by the high correlation (0.70) observed between digital penetration and Tertiary Education. Since the latter was retained in the final model, it likely captures the underlying effects of socio-economic development and human capital more effectively, rendering the digital banking variable redundant within this specific LMM configuration.

To definitively and scientifically rule out the presence of a third distinct latent state, this optimized formula was subsequently tested for a $k = 3$ configuration. However, the

estimation process consistently encountered a "computationally singular system" error. This specific statistical friction is fundamentally caused by severe multicollinearity within the dataset. Specifically, the near-perfect inverse correlation between the Technical Margin and the Loss Ratio (previously calculated at -0.95) creates a highly collinear environment that the estimation algorithm struggles to process mathematically when forced to divide the sample into too many small latent classes.

Even after artificially attempting to drop the Loss Ratio from the equation to forcefully achieve algorithmic convergence, the resulting $k = 3$ model proved to be structurally invalid and economically meaningless. It produced highly unbalanced initial probabilities—with one specific "ghost" state populated by less than 0.01% of the total sample—and generated completely erratic, uninterpretable transition dynamics. Consequently, the $k = 2$ model is definitively confirmed as the most statistically robust, symmetrically balanced, and economically meaningful specification for representing this complex dataset.

4.2.2 Economic Interpretation

In the LMM framework with covariates included in the measurement model (modBasic = 1), the estimated intercepts (α) serve as the fundamental tool for identifying the baseline identity of the latent states. These parameters represent the expected value of the standardized response variables—Premium Growth, Loss Ratio, Expense Ratio, and Technical Margin—when all exogenous macroeconomic and socio-demographic covariates are held at their mean levels. By filtering out the immediate volatility caused by external shocks, the intercept matrix isolates the intrinsic technical performance characteristic of each regime.

Table 4.2.2-1: Latent States intercepts (model with covariates). Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.

Latent States Intercepts

Regime	Premium Growth	Loss Ratio	Expense Ratio	Margin
State 1	8.0806	-0.1177	-0.4848	0.1920
State 2	-0.7519	0.2497	1.0290	-0.4075

As shown in Table 4.2.2-1, the estimated intercepts reveal a stark structural dichotomy between two distinct market conditions: State 1 is characterized by an exceptionally strong baseline for Premium Growth (α = 8.0806), indicating a robust capacity for expansion independent of the economic cycle. Furthermore, the negative intercepts for both the Loss Ratio (-0.1177) and the Expense Ratio (-0.4848) suggest that markets in this regime

naturally operate with technical claims and administrative costs significantly below the European average. This efficiency translates into a structurally positive Technical Margin (0.1920), identifying State 1 as the primary equilibrium for consolidated and profitable markets.

Conversely, the second state represents a regime of technical friction and contraction. It exhibits a negative baseline for Premium Growth (α = -0.7519), accompanied by a significant deterioration in operational efficiency, as evidenced by a massive spike in the baseline Expense Ratio (1.0290). The Loss Ratio also stands above the mean (0.2497), reflecting higher sensitivity to claims or surrender waves. Consequently, the baseline Technical Margin falls into negative territory (-0.4075), marking State 2 as a "stress" regime where insurance undertakings struggle to maintain profitability. This clear mathematical separation between a high-growth/low-cost state and a contractionary/high-cost state confirms that the LMM has successfully captured two meaningful economic "regimes" rather than mere statistical noise.

Table 4.2.2-2: Initial probabilities (model with covariates). Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.

Initial Probabilities

Regime	Probability
State 1	0.8972
State 2	0.1028

The estimated initial probabilities indicate a highly concentrated starting point for the time series. At the very beginning of the observation period (2017 Q4), the Latent Markov Model estimates an overwhelming 89.72% probability that the European insurance markets were firmly positioned within the "Healthy" regime (State 1).

Table 4.2.2-3: Transition Probability Matrix (model with covariates). Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.

Transition Probability Matrix

Regime	State 1	State 2
State 1	0.9093	0.0907
State 2	0.1922	0.8078

The dominant values on the main diagonal indicate a high degree of structural inertia in both regimes.

There is a 90.93% probability that a country in a "Healthy" state will remain there in the subsequent quarter. This confirms that the expansionary regime is the natural equilibrium for most European markets, showing remarkable stability even in the face of moderate economic fluctuations.

The "Distressed" state also shows significant persistence (80.78%). This suggests that technical crises—such as those driven by liquidity strains or expense spikes—are not fleeting shocks but persistent regimes. Once a market enters a period of low performance, it tends to remain "trapped" there for several quarters, reflecting the time required for structural portfolio adjustments to take effect.

The most remarkable feature of the matrix is the asymmetry between the off-diagonal elements, which provides a measure of the sector's resilience:

The probability of a healthy market slipping into a distressed regime is relatively low, at 9.07%. This indicates that the transition into crisis is typically triggered only by severe exogenous shocks, such as the unprecedented interest rate hikes seen in 2023.

Conversely, the probability of returning from a distressed state to a healthy one is 19.22%.

The probability of recovery (19.22%) is more than double the risk of falling into a crisis (9.07%). This finding is statistically significant as it demonstrates a fundamental structural resilience in the European Life insurance market: while “crises” are persistent in the overall period of observation, the systemic tendency toward recovery is stronger than the vulnerability to contraction.

4.2.3 Regression Coefficients

Table 4.2.3-1: Regression Coefficients. Author's calculation on ECB (2025), EIOPA (2025), Eurostat (2025) data.

Regression Coefficients (Beta)

Covariate	Premium Growth	Loss Ratio	Expense Ratio	Margin
GDP	4.012***	0.682	1.229	1.256
Unemployment Rate	4.35***	3.898***	-5.388***	-0.11***
Old Dependency Ratio	-0.009	-0.167***	0.217***	-0.107**
Tertiary Education	-0.142***	0.099*	-0.13***	-0.013
Life Expectancy	0.024	0.04	-0.2***	-0.011
Immigration Rate	0.165***	0.121***	-0.011	0.157***
Y10	-0.222***	0.135***	0.15***	-0.12**

The regression matrix provides a clear view of how external factors impact the insurance sector's performance. The results highlight several key relationships.

Data show that Premium Growth is strongly and positively linked to GDP (4.012^{***}) and the Unemployment Rate (4.35^{***}). This confirms that a growing economy supports insurance demand. On the other hand, the 10-Year Bond Yield (Y10) has a significant negative impact (-0.222^{***}). This illustrates the substitution effect with other financial instruments: when government bonds offer higher returns, life insurance products become less attractive to savers, leading to lower premium volumes.

The Loss Ratio increases significantly with higher Unemployment (3.898^{***}) and rising Y10 yields (0.135^{***}). This supports the idea of surrender risk: in difficult economic times or when interest rates rise, households are more likely to withdraw their policies to get cash or seek better investments elsewhere. Conversely, a higher Old Dependency Ratio helps lower the loss ratio (-0.167^{***}), likely because older policyholders are less likely to cancel their contracts before they mature.

The Expense Ratio is lower in countries with higher Life Expectancy (-0.2^{***}). This suggests that more developed societies often have better digital infrastructures that reduce administrative costs. However, rising interest rates increase these expenses (0.15^{***}), probably because higher rates are often tied to inflation, which raises the cost of doing business.

Finally, the Technical Margin is squeezed by Unemployment (-0.11^{***}), an aging population (Old Dependency: -0.107^{**}), and rising interest rates (y10: -0.12^{**}). In this environment, the Immigration Rate (0.157^{***}) is the main factor supporting profitability. This suggests

that a growing labor force helps keep the insurance market profitable by bringing in new customers and balancing the effects of an aging population.

4.3 Model Comparison

To conclude the empirical analysis, it is essential to compare the findings from the baseline Latent Markov Model (without covariates) and the advanced specification (with covariates). This comparison highlights the dual nature of the European Life insurance market and provides a comprehensive understanding of its structural and dynamic vulnerabilities.

In the baseline model, the latent states absorbed the total variance of the dataset, thereby capturing a long-term structural and geographic divide (e.g., mature Western markets versus dynamic, highly volatile Eastern European economies). Conversely, by introducing exogenous covariates in the measurement model, the macroeconomic noise was effectively filtered out. Consequently, the latent states in the advanced specification represent pure technical efficiency, clearly distinguishing between a "Healthy/High-Performance" regime and a "Distressed" one, irrespective of national borders or market maturity.

The two specifications answer distinct but highly complementary research questions. The baseline model maps the temporal trajectories, revealing *who* was vulnerable and when the systemic fracture occurred—most notably, the mass migration towards the volatile state in 2023. The advanced model, through the estimated regression coefficients (β), explains why this fracture happened. It mathematically quantifies the exogenous causes: the unprecedented spike in 10-Year Government Bond Yields (Y10) and systemic inflation, which triggered severe surrender risk, depressed premium collection, and artificially inflated the technical Loss Ratios across the continent.

Ultimately, this comparative analysis demonstrates the fundamental resilience of the European Life insurance sector. The advanced model proves that the widespread crisis captured during the 2022-2023 period was not the result of chronic operational inefficiency or a failing internal business model. Instead, it was a mechanical, specular reaction to a severe and global exogenous shock—namely, the central banks' aggressive monetary tightening cycle. As the macroeconomic environment begins to normalize, the transition probability matrix reveals a strong inherent tendency (19.22%) for the market to recover its equilibrium and return to the high-performance state, definitively confirming the sector's structural solidity.

CONCLUSION

This thesis aims to investigate the complex temporal dynamics and structural heterogeneity of the European Life insurance market. By applying a LMM to a quarterly panel dataset (2017-2025), the empirical analysis went beyond traditional static evaluations, effectively capturing the unobserved regimes that drive the sector's performance.

The first phase of the analysis, based on a homogeneous LMM without covariates, revealed that the European landscape is fundamentally built around two distinct latent states ($k = 2$). This baseline model captured the market's "symptoms", identifying a "Mature and Resilient" core (characterized by stable margins and low loss ratio) and a "High Volatility and Liquidity Stress" regime. Crucially, the trajectory analysis uncovered a structural break during the 2023-2024 period. Historically stable markets, such as Italy and Germany, experienced a mass migration towards the stress state, suddenly aligning their statistical profile with that of peripheral economies due to severe liquidity pressures.

The subsequent inclusion of exogenous covariates in the measurement model provided the analytical key to understanding the causes of this migration, creating a direct link with the findings from the Exploratory Data Analysis (EDA). As highlighted in the preliminary visual analysis, the market had suffered a dual shock: inflation (HICP) eroded purchasing power, causing a collapse in Premium Growth, while the increase in 10-year bond yields (Y10) led to early surrenders, and a consequent spike in the Loss Ratio.

The advanced LMM mathematically validated these insights. By including covariates, the model filtered out macroeconomic noise from the regimes, isolating the sector's "pure technical efficiency". The regression coefficients confirmed the EDA dynamics: the rise in the Y10 rate has a significant negative impact on Premium Growth and mechanically drives up the Loss Ratio through the substitution effect. This connection clearly explains why the movements observed in the first model occurred. The volatility recorded in 2023 was not a symptom of chronic inefficiency or failing internal business models; instead, it was a mechanical reaction to a global exogenous shock driven by the central banks' monetary tightening.

However, despite the severity of the crisis described by both models, the advanced LMM specification demonstrated the fundamental robustness of the European Life insurance sector. The estimated transition probabilities reveal a strong intrinsic resilience: the probability of a distressed market recovering its healthy equilibrium (19.22%) is more than double the risk of a stable market falling into a crisis (9.07%).

Ultimately, this analysis highlights the delicate balance between structural stability and vulnerability within the market. Although the sector rests on a solid and highly persistent technical equilibrium, it remains acutely vulnerable to rapid changes in monetary policy. As the macroeconomic environment normalizes, the sector's proven ability to absorb extreme liquidity risks and naturally return to profitability confirms its vital role in the real economy. It proves that the true resilience of the market lies in its ability to navigate—and quickly recover from—inevitable regime shifts.

The empirical findings of this study carry significant implications for both industry practitioners and regulatory authorities. From a managerial perspective, the strong sensitivity of the Loss Ratio to the 10-Year Government Bond Yield highlights the acute vulnerability of traditional Life portfolios to lapse risk. To mitigate the "run-on-the-bank" phenomenon observed during the 2023 rate shock, insurance companies must proactively structure products that are less sensitive to rapid interest rate hikes (Swiss Re Institute 2023). This objective can be achieved by designing hybrid policies with dynamic surrender penalties or more flexible profit-participation mechanisms to enhance resilience (Swiss Re Institute 2024). From a regulatory standpoint, authorities such as EIOPA should shift towards a more dynamic monitoring of liquidity risk (EIOPA 2023b). Although the current Solvency II regulatory framework effectively ensures long-term capital adequacy (European Parliament and Council 2009), the rapid migration of historically stable markets into the stress state demonstrates that capital reserves alone cannot prevent sudden liquidity crises (Fitch Ratings 2024). Supervisory frameworks must therefore integrate forward-looking and regime-switching stress tests to preemptively identify markets at risk of mass surrenders (EIOPA 2025a).

Despite the robustness of the methodology employed, this study presents certain inherent limitations that serve as a baseline for future exploration. Primarily, the empirical analysis relies on aggregated national data extracted from EIOPA's prudential reporting (EIOPA 2025b). While this approach ensures macroeconomic comparability and captures systemic trends, it inevitably obscures firm-level heterogeneity (Bertarelli 2015); future research would greatly benefit from applying the Latent Markov framework to micro-data of specific insurance undertakings to analyze the resilience of individual business models (Pandolfi, Bartolucci and Pennoni 2023).

Furthermore, from a statistical and modeling perspective, estimating models with a higher number of latent states ($k \geq 3$) presented structural identification challenges. In complex longitudinal frameworks incorporating multiple exogenous covariates, expanding the state space often leads to severe over-parameterization and data sparsity within specific

transition matrices (Pennoni, Bartolucci and Pandolfi 2024). Consequently, rather than identifying true economic phases, forcing additional states causes the estimation algorithm to capture idiosyncratic noise or temporary country-specific shocks. The robust convergence of the binary specification ($k = 2$) confirms that a dual-state architecture is the most statistically sound representation of the market's underlying systemic dichotomy (Colombi, Giordano and Kateri 2024). Future methodological extensions could explore penalized estimation techniques or incorporate larger global datasets, potentially providing the sample depth required to safely isolate more nuanced micro-regimes without compromising structural stability

REFERENCES

Allianz (2025). Allianz Global Insurance Report 2025: Rising demand for protection. Munich: Allianz SE Economic Research. Available at: https://www.allianz.com/en/economic-research/publications.html [Accessed 5 January 2026].

ANIA (2024). Italian Insurance Market Trends 2023-2024. Rome: Associazione Nazionale fra le Imprese Assicuratrici. Available at: https://www.ania.it/pubblicazioni [Accessed 29 December 2025].

AtenaBrokers (2025). Decreto attuativo Legge Gelli-Bianco: i requisiti minimi delle coperture assicurative. Available at: https://www.atenabrokers.it [Accessed 29 December 2025].

Banca Generali (2025). Investor Relations: Public Exchange Offer (OPS) Mediobanca. Available at: https://www.bancagenerali.com/investors [Accessed 29 December 2025].

Bartolucci, F., Farcomeni, A. and Pennoni, F. (2012). Latent Markov Models for Longitudinal Data. Boca Raton: CRC Press.

Bartolucci, F., Pandolfi, S. and Pennoni, F. (2017). 'LMest: An R Package for Latent Markov Models for Longitudinal Categorical Data', Journal of Statistical Software, 81(4), pp. 1–25. Available at: https://doi.org/10.18637/jss.v081.i04.

Bertarelli, G. (2015). Latent Markov models for aggregate data: application to disease mapping and small area estimation. PhD Thesis. University of Milano-Bicocca.

Colombi, R., Giordano, S. and Kateri, M. (2024). 'Hidden Markov models for longitudinal rating data with dynamic response styles', Statistical Methods and Applications, 33, pp. 1–36.

COVIP (n.d.). Glossario: Rami assicurativi Vita. Commissione di Vigilanza sui Fondi Pensione. Available at: https://www.covip.it/per-il-cittadino/educazione-previdenziale/glossario/rami-assicurativi-vita [Accessed 29 December 2025].

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). 'Maximum Likelihood from Incomplete Data via the EM Algorithm', Journal of the Royal Statistical Society. Series B (Methodological), 39(1), pp. 1–38.

Dotto, F. (n.d.). Modelli Latent Markov per dati longitudinali: utilizzo in ambito econometrico e possibili estensioni. [Thesis/Book Chapter].

Draghi, M. (2024). The future of European competitiveness. Brussels: European Commission. Available at: https://commission.europa.eu [Accessed 29 December 2025].

ECB (2025). Statistical Data Warehouse: Long-term interest rate statistics for EU Member States. Frankfurt: European Central Bank. Available at: https://sdw.ecb.europa.eu [Accessed 3 January 2026].

EIOPA (2015). Guidelines on Own Risk and Solvency Assessment. Frankfurt: EIOPA.

EIOPA (2021). European Insurance Overview 2021. Luxembourg: Publications Office of the European Union.

EIOPA (2023a). European Insurance Overview 2023. Luxembourg: Publications Office of the European Union. Available at: https://www.eiopa.europa.eu/ [Accessed 29 December 2025].

EIOPA (2023b). Financial Stability Report - December 2023. Luxembourg: Publications Office of the European Union. Available at: https://www.eiopa.europa.eu/ [Accessed 29 December 2025].

EIOPA (2025a). Annual Report 2024. Luxembourg: Publications Office of the European Union. Available at: https://www.eiopa.europa.eu/about/governance-structure/annual-activity-reports_en [Accessed 29 December 2025].

EIOPA (2025b). European Insurance Overview: Quarterly Solo Prudential Reporting (2017-2025). European Insurance and Occupational Pensions Authority. Available at: https://www.eiopa.europa.eu/tools-and-data/insurance-statistics_en [Accessed 30 January 2026].

EIOPA (n.d. a). Solvency II: Regulatory Framework. Available at: https://www.eiopa.europa.eu/browse/regulation-and-policy/solvency-ii_en [Accessed 30 December 2025].

EIOPA (n.d. b). Solvency II Overview - Architecture. Available at: https://www.eiopa.europa.eu/ [Accessed 29 January 2026].

Eling, M. and Luhnen, M. (2010). 'Efficiency in the international insurance industry: A cross-country comparison', Journal of Banking and Finance, 34(7), pp. 1497-1509.

European Parliament and Council (2009). 'Directive 2009/138/EC of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)', Official Journal of the European Union, L 335, 17 December, pp. 1–155. Available at:

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02009L0138-20230101 [Accessed 29 December 2025].

Eurostat (2013). European system of accounts: ESA 2010. Luxembourg: Publications Office of the European Union. Available at: https://ec.europa.eu/eurostat/esa2010/chapter/view/16/ [Accessed 29 December 2025].

Eurostat (2024). Population structure and ageing. European Commission. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Population_structure_and_ageing [Accessed 29 December 2025].

Eurostat (2025a). At-risk-of-poverty rate by poverty threshold [ilc_li02]. Luxembourg: Eurostat.

Eurostat (2025b). Employment and activity by sex and age - quarterly data [lfsi_emp_q]. Luxembourg: Eurostat.

Eurostat (2025c). GDP and main components (output, expenditure and income) [namq_10_gdp]. Luxembourg: Eurostat.

Eurostat (2025d). Health care expenditure by financing scheme [hlth_sha11_hf]. Luxembourg: Eurostat.

Eurostat (2025e). HICP - monthly data (index) [prc_hicp_midx]. Luxembourg: Eurostat. Available at: https://ec.europa.eu/eurostat/databrowser/view/prc_hicp_midx/default/table?lang=en [Accessed 30 January 2026].

Eurostat (2025f). Individuals using the internet for internet banking [isoc_bde15cu]. Luxembourg: Eurostat.

Eurostat (2025g). Life expectancy by age and sex [demo_mlexpec]. Luxembourg: Eurostat.

Eurostat (2025h). Long-term interest rates - Maastricht criterion [irt_lt_mcby]. Luxembourg: Eurostat.

Eurostat (2025i). Macroeconomic and socio-demographic databases. European Commission.

Eurostat (2025j). Non-financial transactions [nasa_10_nf_tr]. Luxembourg: Eurostat.

Eurostat (2025k). Population by educational attainment level [edat_lfse_03]. Luxembourg: Eurostat.

Eurostat (2025l). Population on 1 January by age, sex and broad group of citizenship [migr_pop1ctz]. Luxembourg: Eurostat.

Eurostat (2025m). Population structure and ageing [demo_pjanind]. Luxembourg: Eurostat.

Eurostat (2025n). Unemployment by sex and age – quarterly data [une_rt_q]. Luxembourg: Eurostat.

Eurostat (2025o). Young people neither in employment nor in education and training (NEET) [edat_lfse_20]. Luxembourg: Eurostat.

Feodoria, M. and Förstemann, T. (2015). 'Lethal lapses – how a positive interest rate shock might stress German life insurers', Discussion Paper No. 12/2015. Deutsche Bundesbank.

Fitch Ratings (2024). Western European Life Insurance Outlook 2024: Rising Rates and Lapse Risk Challenge Liquidity. Fitch Ratings Research.

Gatzert, N. and Wesker, H. (2012). 'The impact of natural hedging on a life insurer's risk situation', Journal of Risk and Insurance, 79(4), pp. 1069-1101.

Generali (2025). Climate Change and Insurance: A Strategic Outlook. Trieste: Assicurazioni Generali S.p.A.

Insurance Europe (2012). How insurance works. Brussels: Insurance Europe aisbl. Available at: https://www.insuranceeurope.eu/publications [Accessed 27 December 2025].

Insurance Europe (2015). The benefits of insurance. Brussels: Insurance Europe aisbl. Available at: https://www.insuranceeurope.eu/publications [Accessed 27 December 2025].

IVASS (2023). Relazione annuale sull'attività svolta. Rome: Istituto per la Vigilanza sulle Assicurazioni.

Jeniot (2025). Black Box in auto: cos'è e a cosa serve. Generali Jeniot Blog. Available at: https://www.jeniot.it/blog/black-box-in-auto-cosa-e-cosa-serve [Accessed 29 December 2025].

Mapfre Economics (2025). Economic and Industry Outlook 2025: Fourth-Quarter Update. Madrid: Mapfre Economic Research.

Maruotti, A. (2011). 'Mixed Hidden Markov Models for Longitudinal Data: An Overview', International Statistical Review, 79(3), pp. 427-454.

Ministero delle Imprese e del Made in Italy (2025). Polizze per rischi catastrofali: normativa per le imprese. Available at: https://www.mimit.gov.it [Accessed 29 December 2025].

OECD (2025a). Pensions at a Glance 2025: OECD and G20 Indicators. Paris: OECD Publishing. Available at: https://doi.org/10.1787/pension_glance-2025-en [Accessed 29 December 2025].

OECD (2025b). Short-term economic statistics and structural indicators. Paris: OECD.

Osservatorio Nazionale Città Clima (2025). Il clima è cambiato: Bilancio 2025. Legambiente. Available at: https://cittaclima.it [Accessed 29 December 2025].

Outreville, J.F. (2013a). 'The Relationship Between Insurance and Economic Growth: A Survey', Ocean and Coastal Management, 17(1), pp. 71–99.

Outreville, J.F. (2013b). 'The relationship between insurance and economic development: 85 empirical papers for a review of the literature', Risk Management and Insurance Review, 16(1), pp. 71-122.

Pandolfi, S., Bartolucci, F. and Pennoni, F. (2023). 'A hidden Markov model for continuous longitudinal data with missing responses and dropout', Biometrical Journal, 65(5), 2200016.

Pennoni, F., Bartolucci, F. and Pandolfi, S. (2024). 'Variable Selection for Hidden Markov Models with Continuous Variables and Missing Data', Journal of Classification, 41, pp. 568-589.

Pradhan, R.P., Arvin, M.B. and Norman, N.R. (2016). 'Insurance penetration and economic growth nexus: Cross-country evidence from ASEAN', Research in International Business and Finance, 36, pp. 447–458.

R Core Team (2024). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Swiss Re Institute (2023). 'Raising the bar: the global macro and insurance outlook 2024/25', Sigma, No 6/2023. Zurich: Swiss Re Institute.

Swiss Re Institute (2024). 'World insurance: strengthening global resilience with a new lease of life', Sigma, No 3/2024. Zurich: Swiss Re Institute.

Verma, A. and Bala, R. (2013). 'Relationship between Life Insurance and Economic Growth: Evidence from India', Global Journal of Management and Business Studies, 3(4), pp. 413–420.

APPENDIX

R CODE

```
# ----------------
# 1. PREPROCESSING

req_pkgs <- c("dplyr", "tidyr", "ggplot2", "pheatmap", "tidyverse", "readxl",
              "readr", "ggrepel", "corrplot", "viridis", "LMest", "zoo", "lubridate", "janitor",
              "ggcorrplot", "gt", "knitr", "reshape2" )
to_install <- setdiff(req_pkgs, rownames(installed.packages()))
if (length(to_install) > 0) install.packages(to_install, dependencies = TRUE)

suppressPackageStartupMessages({
  library(dplyr)
  library(tidyr)
  library(ggplot2)
  library(pheatmap)
  library(readxl)
  library(tidyverse)
  library(knitr)
  library(readr)
  library(ggrepel)
  library(LMest)
  library(zoo)
  library(lubridate)
  library(janitor)
  library(corrplot)
  library(gt) # exporting tables
  library(grid)
  library(gridExtra)
  library(gtable)
  library(reshape2)
  library(viridis)
})

Sys.setlocale("LC_TIME", "C")

# set wd
setwd("~/Dati tesi magistrale AP 891453/TESI_Insurance")

# 1. RESPONSE VARIABLES CALCULATION (PCEQ)
PCEQ <- read_excel("SQ_Premiums_Claims_Expenses.xlsx", sheet = 4) %>%
  filter(`Item code` %in% c("R1410", "R1900", "R1610")) %>%
  select(1, 2, 5, 6) %>%
  rename(country_region = 1, quarter = 2, item_code = 3, value = 4) %>%
  mutate(
    value_type = case_when(
      item_code == "R1410" ~ "Premiums",
      item_code == "R1610" ~ "Claims",
      TRUE                 ~ "Expenses"
    ),
    quarter = as.yearqtr(quarter, format = "%Y Q%q")
  ) %>%
  relocate(value_type, .before = value) %>%
  na.omit()

write.csv(PCEQ, "PCEQ.csv", row.names = FALSE)

PCEQ_y <- PCEQ %>%
  mutate(date = yq(quarter)) %>%
  select(country_region, date, value_type, value) %>%
  pivot_wider(names_from = value_type, values_from = value) %>%
  clean_names() %>%
  arrange(country_region, date) %>%
  group_by(country_region) %>%
  mutate(
    value_prev_year = lag(premiums, n = 4),
    premium_growth  = round(((premiums - value_prev_year) / value_prev_year) * 100, 2),
    loss_ratio      = (claims / premiums) * 100,
    exp_ratio       = (expenses / premiums) * 100,
    margin          = ((premiums - claims - expenses) / premiums) * 100
  ) %>%
  select(-value_prev_year) %>%
  ungroup() %>%
  mutate(date = as.yearqtr(date, format = "%Y-Q%q"))
```

```r
# 2. MASTER CLEANING FUNCTION
clean_data <- function(dataset, val_name, time_freq = "quarterly") {
  dataset[-1] <- lapply(dataset[-1], as.numeric)

  cleaned <- dataset %>%
    rename(country_region = 1) %>%
    pivot_longer(cols = -country_region, names_to = "date", values_to = val_name) %>%
    mutate(country_region = toupper(country_region))

  # Assegnazione automatica del formato data in base alla frequenza
  if (time_freq == "quarterly") {
    cleaned$date <- as.yearqtr(cleaned$date, format = "%Y-Q%q")

  } else if (time_freq == "monthly") {
    # Converte nel formato stringa Anno-Mese (es. "2017-01")
    cleaned$date <- format(ym(cleaned$date), "%Y-%m")

  } else if (time_freq == "annual") {
    cleaned$date <- as.numeric(cleaned$date)
  }

  return(cleaned)
}

# 3. IMPORTING COVARIATES
# Quarterly Macro
NOM_GDP_Q <- clean_data(read_excel("NOM_GDP_Q.xlsx", sheet = 2, skip = 9), "gdp")
UNE_Q    <- clean_data(read_excel("UNE_Q.xlsx", sheet = 3, skip = 10), "une_rate")
Y10_Q    <- clean_data(read_excel("Y10_Q.xlsx", sheet = 3, skip = 7), "y10")
NEET_Q   <- clean_data(read_excel("NEET_Q.xlsx", sheet = 3, skip = 10), "neet")
NASQ_Q   <- clean_data(read_excel("NASQ_Q.xlsx", sheet = 3, skip = 11), "nonfin_tr")

# HICP (Monthly to Quarterly transformation)
HICP_M <- clean_data(read_excel("HICP_M.xlsx", sheet = 3, skip = 8), "hicp_m", time_freq = "monthly")
HICP_Q <- HICP_M %>%
  mutate(
    full_date = ym(date),
    quarter_date = floor_date(full_date, unit = "quarter")
  ) %>%
  group_by(country_region, quarter_date) %>%
  summarise(hicp = mean(hicp_m, na.rm = TRUE), .groups = "drop") %>%
  mutate(date = as.yearqtr(quarter_date)) %>%
  select(country_region, date, hicp)

# Annual Macro
raw_OOP <- read_excel("OOP_Y.xlsx", sheet = 3, skip = 8)
OOP_Y   <- clean_data(raw_OOP[, c(1, seq(2, ncol(raw_OOP), by = 2))], "oop_exp", time_freq = "annual")
ODR_Y       <- clean_data(read_excel("ODR_Y.xlsx", sheet = 3, skip = 7), "old_dep", time_freq = "annual")
EDU_Y       <- clean_data(read_excel("EDU_Y.xlsx", sheet = 3, skip = 10), "t_edu", time_freq = "annual")
AROP_Y      <- clean_data(read_excel("AROP_Y.xlsx", sheet = 3, skip = 10), "arop", time_freq = "annual")
DIG_BANKING_Y <- clean_data(read_excel("DIG_BANKING_Y.xlsx", sheet = 3, skip = 9), "dig_bank", time_freq = "annual")
LIFE_EXP_Y   <- clean_data(read_excel("LIFE_EXP_Y.xlsx", sheet = 3, skip = 9), "life_exp", time_freq = "annual")

# Computed Macro: Employment Gender Gap (Quarterly) and
# EMPLOYMENT (quarter), seas adj, % of total pop -> to create gender gap
EMP_MALE = read_excel("TOTAL_EMP_Q.xlsx", sheet = 3, skip = 11)
EMP_FEMALE = read_excel("TOTAL_EMP_Q.xlsx", sheet = 5, skip = 11)

EMP_MALE_clean <- EMP_MALE %>%
  select(-1) %>%
  mutate(across(everything(), as.numeric))

EMP_FEMALE_clean <- EMP_FEMALE %>%
  select(-1) %>%
  mutate(across(everything(), as.numeric))

emp_cols = intersect(names(EMP_MALE_clean), names(EMP_FEMALE_clean)) # logical op

EMP_G_GAP_Q = EMP_MALE_clean[emp_cols] - EMP_FEMALE_clean[emp_cols]

# now we add Id
EMP_G_GAP_Q = cbind(EMP_MALE[,1], EMP_G_GAP_Q)
EMP_G_GAP_Q   <- clean_data(EMP_G_GAP_Q, "gender_gap")

# Computed Macro: Immigration Rate (Annual)
# FOREIGNERS as total numbers -> we need to compute an annual immigration rate as ->
# denominator: ALL citizens; numerator : only foreign citizenship

IMM_NUM = read_excel("NUM_IMM_Y.xlsx", sheet = 3, skip = 10)
IMM_DEN = read_excel("DEN_IMM_Y.xlsx", sheet = 3, skip = 10)
```

```
# computing the ratio
IMM_NUM_clean <- IMM_NUM %>%
  select(-1) %>%
  mutate(across(everything(), as.numeric))

IMM_DEN_clean <- IMM_DEN %>%
  select(-1) %>%
  mutate(across(everything(), as.numeric))

imm_cols = intersect(names(IMM_NUM_clean), names(IMM_DEN_clean))

IMM_RATE_Y = IMM_NUM_clean[imm_cols] / IMM_DEN_clean[imm_cols]

# control
as.numeric(IMM_NUM[1,2]) / as.numeric(IMM_DEN[1,2]) == IMM_RATE_Y[1,1]

# now we add Id
IMM_RATE_Y = cbind(IMM_NUM[,1], IMM_RATE_Y)
IMM_RATE_Y    <- clean_data(IMM_RATE_Y, "immigration_rate", time_freq = "annual")


# 4. FINAL MERGE
# quarter
covars_q_list <- list(NOM_GDP_Q, UNE_Q, Y10_Q, HICP_Q, NEET_Q, NASQ_Q, EMP_G_GAP_Q)
covars_q      <- covars_q_list %>% reduce(left_join, by = c("country_region", "date"))

# annual
covars_y_list <- list(OOP_Y, ODR_Y, EDU_Y, AROP_Y, DIG_BANKING_Y, LIFE_EXP_Y, IMM_RATE_Y)
covars_y      <- covars_y_list %>% reduce(left_join, by = c("country_region", "date"))

full_PCEQ <- PCEQ_y %>%
  left_join(covars_q, by = c("country_region", "date")) %>%
  mutate(year = as.integer(date)) %>%
  left_join(covars_y, by = c("country_region", "year" = "date")) %>%
  select(-year)

# Further cleaning after diagnostic analysis
vars_to_check <- colnames(full_PCEQ[10:length(full_PCEQ)])
countries_to_drop <- c("LIECHTENSTEIN")
cov_to_drop = c("nonfin_tr")
new_vars = vars_to_check[!vars_to_check %in% cov_to_drop]

PC_fixed <- full_PCEQ %>%
  filter(!country_region %in% countries_to_drop) %>%
  select(-cov_to_drop) %>%
  arrange(country_region, date) %>%
  # applying LOCF
  group_by(country_region) %>%
  fill(all_of(new_vars), .direction = "down") %>%
  ungroup()

# HANDLING MISSING DATA: 10-YEAR BOND YIELDS (NON-EU/SPECIFIC COUNTRIES)
# Focus: Imputing Y10 data for Norway, Iceland, and Estonia to prevent data loss.
# Note: Liechtenstein was dropped entirely due to excessive missing variables.
# A. NORWAY (Converting Monthly Data to Quarterly Averages)
y10_M_Nor <- read.csv("Norway 10-Year Bond Yield Historical Data.csv")

nor_clean <- y10_M_Nor %>%
  select(date = Date, y10 = Price) %>%
  mutate(
    date_dt = mdy(date),
    q_date = as.yearqtr(date_dt)
  ) %>%
  group_by(q_date) %>%
  summarise(y10 = mean(y10, na.rm = TRUE), .groups = "drop") %>%
  mutate(
    date = as.yearqtr(q_date),
    country_region = "NORWAY"
  ) %>%
  select(country_region, date, y10)

# B. ICELAND (Converting Monthly Data to Quarterly Averages)
y10_M_Ice <- read.csv("Iceland 10-Year Bond Yield Historical Data.csv")

ice_clean <- y10_M_Ice %>%
  select(date = Date, y10 = Price) %>%
  mutate(
    date_dt = mdy(date),
    q_date = as.yearqtr(date_dt)
  ) %>%
  group_by(q_date) %>%
```

```
  summarise(y10 = mean(y10, na.rm = TRUE), .groups = "drop") %>%
  mutate(
    date = as.yearqtr(q_date),
    country_region = "ICELAND"
  ) %>%
  select(country_region, date, y10)

# C. ESTONIA (Merging ECB Downloaded Data with Manual Convergence Imputations)
est_raw <- read.csv("ECB Data Portal_20260211120641.csv", sep = ",")
colnames(est_raw) <- c("int_date", "date", "y10")

# 1. Clean downloaded ECB data (2020 Q2 - 2026)
est_file_part <- est_raw %>%
  mutate(
    date_clean = sub("(\\d{4})Q(\\d)", "\\1 Q\\2", date),
    date = as.yearqtr(date_clean, format = "%Y Q%q"),
    country_region = "ESTONIA"
  ) %>%
  select(country_region, date, y10)

# 2. Manual imputation for missing historical data (based on Convergence Reports)
manual_quarters <- c(
  "2017 Q4",
  paste("2018", c("Q1", "Q2", "Q3", "Q4")),
  paste("2019", c("Q1", "Q2", "Q3", "Q4")),
  "2020 Q1"
)

# Assigned values: 2017->0.7, 2018->0.6, 2019->0.3
manual_values <- c(0.7, rep(0.6, 4), rep(0.3, 4), 0.14)

est_manual_part <- data.frame(
  country_region = "ESTONIA",
  date = as.yearqtr(manual_quarters, format = "%Y Q%q"),
  y10 = manual_values
)

# Bind Estonia data together
est_clean <- bind_rows(est_manual_part, est_file_part)

# D. MERGING NEW Y10 DATA INTO MAIN DATASET
Y10_Q_new <- Y10_Q %>%
  filter(!country_region %in% c("ESTONIA", "NORWAY", "ICELAND"))

# Row-bind the cleaned data and reorder
Y10_Q_final <- bind_rows(Y10_Q_new, nor_clean, ice_clean, est_clean) %>%
  arrange(country_region, date)

# MASTER DATASET UPDATE AND STRUCTURE VISUALIZATION

# Update Master Dataset and rename specific column
pc_master <- PC_fixed %>%
  select(-y10) %>%
  left_join(Y10_Q_final, by = c("country_region", "date")) %>%
  rename(oop_exp = 15)

dataset_to_plot <- pc_master

# 1. df structure
df_structure <- data.frame(
  Variable = names(dataset_to_plot),

  # extract the class of the variable
  Class = sapply(dataset_to_plot, function(x) paste(class(x), collapse = ", ")),

  # extraction of first values
  `First Values` = sapply(dataset_to_plot, function(x) {
    vals <- head(x, 4)
    if(is.numeric(vals)) vals <- round(vals, 3)
    paste0(paste(as.character(vals), collapse = ", "), ", ...")
  }),
  check.names = FALSE,
  row.names = NULL
)

# 2. graphic theme
theme_str <- ttheme_default(
  core = list(
    bg_params = list(fill = c("white", "#f9f9f9"), col = "grey90"),
    fg_params = list(fontface = "plain", fontsize = 9, hjust = 0, x = 0.05)
  ),
```

```
  colhead = list(
    bg_params = list(fill = "#2C3E50", col = "grey90"),
    fg_params = list(col = "white", fontface = "bold", fontsize = 10, hjust = 0, x = 0.05)
  )
)

t_str <- tableGrob(df_structure, rows = NULL, theme = theme_str)

title_text <- "Dataset Structure"
subtitle_text <- paste0("Observations: ", nrow(dataset_to_plot), " | Variables: ", ncol(dataset_to_plot))

title_grob <- textGrob(title_text, gp = gpar(fontsize = 12, fontface = "bold"))
subtitle_grob <- textGrob(subtitle_text, gp = gpar(fontsize = 10, fontface = "italic", col = "grey30"))

header_grob <- arrangeGrob(title_grob, subtitle_grob, ncol = 1, heights = unit(c(0.6, 0.4), "npc"))

final_str_plot <- arrangeGrob(
  header_grob,
  t_str,
  nrow = 2,
  heights = unit.c(unit(1.5, "cm"), unit(1, "npc") - unit(1.5, "cm"))
)

# dimensions and saving
h <- convertHeight(sum(t_str$heights), "in", valueOnly = TRUE) + 1.5
w <- convertWidth(sum(t_str$widths), "in", valueOnly = TRUE) + 1

ggsave("Table_Dataset_Structure.png", final_str_plot, width = w, height = h, dpi = 300)

# ----------------
# 2. EDA

# dataset for EDA
PC_clean <- pc_master %>%
  mutate(
    date_dt = yq(date),
    country_region = as.factor(country_region)
  ) %>%
  relocate(date_dt, .before = "premiums") %>%
  arrange(country_region, date_dt)

# ================================================================================
# Descriptive stats
# function to compute descriptive statistics
desc_stats <- function(x) {
  # if all row is na doesn't work
  if(all(is.na(x))) return(NULL)

  tibble(
    Min = min(x, na.rm = TRUE),
    Q1 = quantile(x, 0.25, na.rm = TRUE),
    Median = median(x, na.rm = TRUE), # La "mediava" (mediana)
    Mean = mean(x, na.rm = TRUE),
    Q3 = quantile(x, 0.75, na.rm = TRUE),
    Max = max(x, na.rm = TRUE),
    SD = sd(x, na.rm = TRUE),
    CV = (sd(x, na.rm = TRUE) / abs(mean(x, na.rm = TRUE))) * 100
  )
}

# building the overview table

# 1. personalized order
response_vars <- c("premiums", "claims", "expenses", "premium_growth",
                   "loss_ratio", "exp_ratio", "margin")

covariate_vars <- c("gdp", "une_rate", "hicp", "y10",
                    "old_dep", "gender_gap", "neet", "t_edu",
                    "dig_bank", "arop", "life_exp", "immigration_rate", "oop_exp")

ordered_vars <- c(response_vars, covariate_vars)

existing_vars <- intersect(ordered_vars, names(PC_clean))

# statistics computation
global_stats <- PC_clean %>%
  ungroup() %>%
  select(all_of(existing_vars)) %>%
  pivot_longer(cols = everything(), names_to = "Variable", values_to = "value") %>%
  mutate(Variable = factor(Variable, levels = existing_vars)) %>%
  group_by(Variable) %>%
```

```
  summarise(stats = list(desc_stats(value))) %>%
  unnest(stats) %>%
  mutate(across(where(is.numeric), \(x) round(x, 2))) %>%
  arrange(Variable)

# ================================================================================
# Descriptive Statistics
save_table_as_image <- function(df, filename, title = NULL) {

  my_theme <- ttheme_default(
    core = list(bg_params = list(fill = "white", col = NA),
                fg_params = list(fontface = "plain", fontsize = 9)),
    colhead = list(bg_params = list(fill = "#f0f0f0", col = NA),
                   fg_params = list(fontface = "bold", fontsize = 10)),
    rowhead = list(fg_params = list(fontface = "bold", fontsize = 10))
  )

  t1 <- tableGrob(df, rows = NULL, theme = my_theme)

  if (!is.null(title)) {
    title_grob <- textGrob(title, gp = gpar(fontsize = 12, fontface = "bold"))
    final_plot <- arrangeGrob(title_grob, t1, heights = unit.c(unit(1, "cm"), unit(1, "npc") - unit(1, "cm")))
  } else {
    final_plot <- t1
  }

  h <- convertHeight(sum(t1$heights), "in", valueOnly = TRUE) + 0.5
  w <- convertWidth(sum(t1$widths), "in", valueOnly = TRUE) + 0.5

  ggsave(filename, final_plot, width = w, height = h, dpi = 300)
}

save_table_as_image(
  global_stats,
  "Table_Descriptive_Statistics.png",
  "Descriptive Statistics (Response Variables vs Covariates)"
)

# ================================================================================
# Descriptive Statistics Italy vs Country with different premium structure
outlier_market <- PC_clean %>%
  filter(country_region != "ITALY") %>%
  group_by(country_region) %>%
  summarise(avg_pg = mean(premium_growth, na.rm = TRUE)) %>%
  arrange(avg_pg) %>%
  slice(1) %>%
  pull(country_region)

# generating country statistics
generate_country_stats <- function(dataset, country_name, file_suffix) {

  stats_df <- dataset %>%
    filter(country_region == country_name) %>%
    ungroup() %>%
    select(all_of(existing_vars)) %>%
    pivot_longer(cols = everything(), names_to = "Variable", values_to = "value") %>%

    mutate(Variable = factor(Variable, levels = existing_vars)) %>%

    group_by(Variable) %>%
    summarise(stats = list(desc_stats(value))) %>%
    unnest(stats) %>%
    mutate(across(where(is.numeric), \(x) round(x, 2))) %>%
    arrange(Variable)

  table_title <- paste0("Descriptive Statistics: ", country_name)

  filename <- paste0("Table_Stats_", file_suffix, ".png")
  save_table_as_image(stats_df, filename, table_title)
}

# Table for Italy
generate_country_stats(PC_clean, "ITALY", "Italy")

# Table for outlier (Slovakia)
generate_country_stats(PC_clean, outlier_market, outlier_market)

# ================================================================================
# Correlation Matrix
vars_matrix <- PC_clean %>%
  ungroup() %>%
```

```
  select(
    -(1:6)
  ) %>%
  na.omit()

# 2. computing corr
corr_values <- cor(vars_matrix)

# 3. from wide to long format
corr_melted <- as.data.frame(as.table(corr_values))

# heatmap
corr_plot = ggplot(corr_melted, aes(x = Var1, y = Var2, fill = Freq)) +
  geom_tile(color = "white") +

scale_fill_gradient2(low = "#d73027", mid = "white", high = "#4575b4",
                     midpoint = 0, limit = c(-1,1), space = "Lab",
                     name="Corr") +

  geom_text(aes(label = round(Freq, 2)), color = "black", size = 3) +

  labs(title = "Correlation matrix: Insurance indicators vs Macro indicators",
       x = "", y = "") +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, vjust = 1, hjust = 1),
    panel.grid.major = element_blank()
  )
ggsave("corr_plot.png", plot = corr_plot, width = 8, height = 6)

# ========================
# temporal evolution for response variables
# mean computation and pivot
response_data <- PC_clean %>%
  group_by(date_dt) %>%
  summarise(
    Premium_Growth = mean(premium_growth, na.rm = TRUE),
    Loss_Ratio = mean(loss_ratio, na.rm = TRUE),
    Expense_Ratio = mean(exp_ratio, na.rm = TRUE),
    Margin = mean(margin, na.rm = TRUE)
  ) %>%
  pivot_longer(
    cols = c(Premium_Growth, Loss_Ratio, Expense_Ratio, Margin),
    names_to = "Indicator",
    values_to = "Value"
  ) %>%
  mutate(Indicator = factor(Indicator,
                            levels = c("Premium_Growth", "Margin", "Loss_Ratio", "Expense_Ratio"),
                            labels = c("Premium Growth (%)", "Technical Margin (%)", "Loss Ratio (%)", "Expense Ratio (%)")))

# plot with facet_wrap
plot_responses_faceted <- ggplot(response_data, aes(x = date_dt, y = Value, color = Indicator)) +
  geom_line(linewidth = 1) +
  geom_point(size = 1.5, alpha = 0.8) +
  scale_color_manual(
    values = c(
      "Premium Growth (%)" = "#2980b9",
      "Technical Margin (%)" = "#27ae60",
      "Loss Ratio (%)" = "#c0392b",
      "Expense Ratio (%)" = "#f39c12"
    )
  ) +
  scale_x_date(date_breaks = "1 year", date_labels = "%Y") +
  facet_wrap(~ Indicator, scales = "free_y", ncol = 2) +
  labs(
    title = "Temporal Evolution of Insurance Response Variables",
    subtitle = "European Average (2017 Q4 - 2025 Q2)",
    x = "Quarter",
    y = NULL
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14),
    legend.position = "none",
    strip.text = element_text(face = "bold", size = 11, color = "white"),
    strip.background = element_rect(fill = "#34495e", color = NA),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )

ggsave("Figure_Faceted_Response_Variables.png", plot = plot_responses_faceted, width = 10, height = 7, dpi = 300)
```

```r
# ================================================================================
# The 2022-2023 Crisis (Macro Shocks vs Insurance Responses)
# computing mean values
crisis_data <- PC_clean %>%
  group_by(date_dt) %>%
  summarise(
    Premium_Growth = mean(premium_growth, na.rm = TRUE),
    Loss_Ratio = mean(loss_ratio, na.rm = TRUE),
    HICP = mean(hicp, na.rm = TRUE),
    Y10 = mean(y10, na.rm = TRUE)
  ) %>%
  pivot_longer(
    cols = -date_dt,
    names_to = "Indicator",
    values_to = "Value"
  ) %>%
  mutate(Indicator = factor(Indicator,
                            levels = c("HICP", "Y10", "Premium_Growth", "Loss_Ratio"),
                            labels = c("1. Inflation (HICP)", "2. 10-Year Yield (Y10)",
                                       "3. Premium Growth", "4. Loss Ratio")))

# plot with facet_wrap
plot_crisis <- ggplot(crisis_data, aes(x = date_dt, y = Value, color = Indicator)) +
  geom_line(linewidth = 1) +
  geom_point(size = 1.5, alpha = 0.8) +
  scale_color_manual(
    values = c(
      "1. Inflation (HICP)" = "#c0392b",
      "2. 10-Year Yield (Y10)" = "#f39c12",
      "3. Premium Growth" = "#2980b9",
      "4. Loss Ratio" = "#8e44ad"
    )
  ) +
  scale_x_date(date_breaks = "1 year", date_labels = "%Y") +
  facet_wrap(~ Indicator, scales = "free_y", ncol = 2) +
  labs(
    title = "The 2022-2023 Crisis: Macroeconomic Shocks vs Insurance Responses",
    subtitle = "Cross-sectional European Average",
    x = "Quarter",
    y = "Percentage (%)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14),
    legend.position = "none",
    strip.text = element_text(face = "bold", size = 11, color = "white"),
    strip.background = element_rect(fill = "#2c3e50", color = NA),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )

ggsave("Figure_Crisis_Macro_vs_Micro.png", plot = plot_crisis, width = 10, height = 7, dpi = 300)

# ================================================================================
# Time Series Decomposition of Premium Growth (STL)
ts_data_raw <- PC_clean %>%
  group_by(date_dt) %>%
  summarise(avg_premium_growth = mean(premium_growth, na.rm = TRUE),
            avg_exp_ratio = mean(exp_ratio, na.rm = TRUE)) %>%
  drop_na(avg_premium_growth) %>%
  drop_na(avg_exp_ratio) %>%
  arrange(date_dt)

# 2. creation of Time Series obj (ts)
# starting from last quarter of 2018
ts_premium <- ts(ts_data_raw$avg_premium_growth, start = c(2018, 4), frequency = 4)

# 3.STL decomposition
decomp <- stl(ts_premium, s.window = "periodic")

# extraction of component for plot
decomp_df <- data.frame(
  date_dt = ts_data_raw$date_dt,
  Observed = as.numeric(ts_premium),
  Trend = as.numeric(decomp$time.series[, "trend"]),
  Seasonality = as.numeric(decomp$time.series[, "seasonal"]),
  Remainder = as.numeric(decomp$time.series[, "remainder"])
)

# pivot_longer
decomp_long <- decomp_df %>%
  pivot_longer(
```

```
    cols = c(Observed, Trend, Seasonality, Remainder),
    names_to = "Component",
    values_to = "Value"
  ) %>%
  mutate(Component = factor(Component, levels = c("Observed", "Trend", "Seasonality", "Remainder")))

# plot visualization
plot_decomp <- ggplot(decomp_long, aes(x = date_dt, y = Value)) +
  geom_line(color = "#2980b9", linewidth = 1) +
  scale_x_date(date_breaks = "1 year", date_labels = "%Y") +
  facet_wrap(~ Component, ncol = 1, scales = "free_y") +
  labs(
    title = "Time Series Decomposition of Premium Growth",
    subtitle = "European Average: Observed, Trend, Seasonality, and Remainder (STL)",
    x = "Quarter",
    y = "Premium Growth (%)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14),
    strip.text = element_text(face = "bold", size = 11, color = "white"),
    strip.background = element_rect(fill = "#34495e", color = NA),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )

ggsave("Figure_Decomposition_Premium_Growth.png", plot = plot_decomp, width = 10, height = 8, dpi = 300)

# ================================================================================
# Time Series Decomposition of Expense Ratio (STL)
# from first quarter of 2018
ts_exp <- ts(ts_data_raw$avg_exp_ratio, start = c(2018, 1), frequency = 4)

# STL
decomp_exp <- stl(ts_exp, s.window = "periodic")

# component extraction
decomp_df_exp <- data.frame(
  date_dt = ts_data_raw$date_dt,
  Observed = as.numeric(ts_exp),
  Trend = as.numeric(decomp_exp$time.series[, "trend"]),
  Seasonality = as.numeric(decomp_exp$time.series[, "seasonal"]),
  Remainder = as.numeric(decomp_exp$time.series[, "remainder"])
)

# pivot longer
decomp_exp_long <- decomp_df_exp %>%
  pivot_longer(
    cols = c(Observed, Trend, Seasonality, Remainder),
    names_to = "Component",
    values_to = "Value"
  ) %>%
  mutate(Component = factor(Component, levels = c("Observed", "Trend", "Seasonality", "Remainder")))

# plot visualization
plot_decomp_exp <- ggplot(decomp_exp_long, aes(x = date_dt, y = Value)) +
  geom_line(color = "darkred", linewidth = 1) +
  scale_x_date(date_breaks = "1 year", date_labels = "%Y") +
  facet_wrap(~ Component, ncol = 1, scales = "free_y") +
  labs(
    title = "Time Series Decomposition of Expense Ratio",
    subtitle = "European Average: Observed, Trend, Seasonality, and Remainder (STL)",
    x = "Quarter",
    y = "Expense Ratio (%)"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14),
    strip.text = element_text(face = "bold", size = 11, color = "white"),
    strip.background = element_rect(fill = "#CC0000", color = NA),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )

ggsave("Figure_Decomposition_exp_ratio.png", plot = plot_decomp_exp, width = 10, height = 8, dpi = 300)

# ================================================================================
# LATENT MARKOV MODEL (BASELINE MODEL)

PC_lmest <- pc_master %>%
  arrange(country_region, date) %>% # ORDERING
  group_by(country_region) %>%
  mutate(
```

```
    id_num = cur_group_id(),
    time_id = row_number()
  ) %>%
  ungroup() %>%
  select(date, id_num, time_id, country_region, premium_growth, loss_ratio, exp_ratio, margin) %>%
  na.omit()

# standardization

PC_scaled <- PC_lmest %>%
  group_by(country_region) %>%
  mutate(across(c(premium_growth, loss_ratio, exp_ratio, margin), scale)) %>%
  ungroup()

head(PC_scaled)

# checks

cat("N of countries:", n_distinct(PC_scaled$country_region), "\n")
cat("N of Quarters:", n_distinct(PC_scaled$time_id), "\n")
cat("Total observations:", nrow(PC_scaled), "\n")
cat("Time interval:", min(pc_master$date), "-", max(pc_master$date), "\n")

# ============================================================================
# model generation -> k from 2 to 5

# containers for results
models_list <- list() # to save models
comparison_table <- data.frame() # to save metrics

# k
k_values <- 2:5

# formulation for LMMs
my_formula <- premium_growth + loss_ratio + exp_ratio + margin ~ NULL


for (k in k_values) {
  cat("\nProcessing k =", k, "...\n")

  tryCatch({

    # fitting
    mod <- lmestCont(
      responsesFormula = my_formula,
      index = c("id_num", "time_id"),
      data = PC_scaled,
      k = k,
      modBasic = 1,
      output = FALSE,
    )

    # extracting the results
    decoding <- lmestDecoding(mod)

    # used results
    used_states <- length(unique(as.vector(decoding$Ul)))

    # saving metrics
    res_row <- data.frame(
      k = k,
      LogLik = round(mod$Lk, 2),
      BIC = round(mod$Bic, 2),
      AIC = round(mod$Aic, 2),
      N_Params = mod$np,
      Active_States = used_states
    )

    # adding to the table
    comparison_table <- rbind(comparison_table, res_row)

    # saving the model as mod1, mod2 etc
    models_list[[paste0("k", k)]] <- mod

    cat("-> completed k =", k, "| BIC:", round(mod$Bic, 2), "\n")

  }, error = function(e) {
    cat("-> error for k =", k, ":", e$message, "\n")
  })
}
```

```
# ================================================================================
# Table comparison

final_comparison <- data.frame()

for (k_name in names(models_list)) {
  mod <- models_list[[k_name]]
  if (is.null(mod) || is.null(mod$Lk)) next

  decoding <- try(lmestDecoding(mod), silent = TRUE)

  if (!inherits(decoding, "try-error") andand !is.null(decoding$UI)) {
    counts <- table(decoding$UI)
    min_pct <- round(min(counts) / sum(counts) * 100, 2)
  } else {
    min_pct <- NA
  }

  row <- data.frame(
    Model = paste0("k = ", mod$k),
    LogLik = round(mod$Lk, 2),
    BIC = round(mod$Bic, 2),
    AIC = round(mod$Aic, 2),
    `Min. Class Size (%)` = min_pct
  )
  final_comparison <- rbind(final_comparison, row)
}

# format
final_comparison_word <- final_comparison %>%
  rename(
    `Latent States` = Model,
    `Log-Likelihood` = LogLik,
    `Min. Class Size (%)` = Min..Class.Size....
  )


# PNG export
save_table_as_image(
  final_comparison_word,
  "Table_Model_Comparison.png",
  "Information criteria comparison (k)"
)

# ================================================================================
# K = 2 probabilities

# selecting model withk = 2
mod <- models_list[["k2"]]

if (!is.null(mod)) {

  # piv: initial probabilities
  df_piv <- data.frame(
    Regime = c("State 1", "State 2"),
    Probability = as.numeric(round(mod$piv, 4)),
    check.names = FALSE
  )
  save_table_as_image(df_piv, "Table_Initial_Probabilities.png", "Initial Probabilities")

  # Pi: transition matrix
  Pi_slice <- mod$Pi[,,2]
  Pi_mat <- round(Pi_slice, 4)

  # df
  df_pi <- data.frame(
    `Regime` = c("State 1", "State 2"),
    `State 1` = Pi_mat[, 1], # first column of the matrix
    `State 2` = Pi_mat[, 2], # second column of the matrix
    check.names = FALSE
  )
  save_table_as_image(df_pi, "Table_Transition_Matrix.png", "Transition Probability Matrix")

  # Mu: conditional means
  Mu_mat <- round(mod$Mu, 4)
  var_names <- rownames(Mu_mat)

  df_mu <- data.frame(
    Variable = var_names,
    `State 1` = Mu_mat[, 1], # La colonna 1 ha i valori "brutti" (loss ratio alto)
    `State 2` = Mu_mat[, 2],  # La colonna 2 ha i valori "buoni" (loss ratio basso)
```

```
    check.names = FALSE
  )
  save_table_as_image(df_mu, "Table__Conditional_Means.png", "Conditional Means of Response Variables")

} else {
  stop("error in the model")
}

# ================================================================================
# K = 2 clustering

k <- mod$k
# function for computation
calculate_state_counts <- function(mod, data_source) {
  decoding <- try(lmestDecoding(mod), silent = TRUE)
  if (inherits(decoding, "try-error") || is.null(decoding$UI)) return(NULL)

  temp_data <- data_source
  temp_data$Current_State <- decoding$UI[cbind(temp_data$id_num, temp_data$time_id)]

  count_table <- temp_data %>%
    group_by(date, Current_State) %>%
    summarise(n_countries = n(), .groups = "drop") %>%
    pivot_wider(id_cols = date, names_from = Current_State, values_from = n_countries, names_prefix = "Cluster_") %>%
    mutate(across(starts_with("Cluster_"), ~replace_na(., 0)))

  return(count_table)
}

# plot

plot_state_evolution <- function(counts_tbl, k) {
  plot_data <- counts_tbl %>%
    mutate(date_plot = as.Date(as.yearqtr(date))) %>%
    pivot_longer(cols = starts_with("Cluster_"), names_to = "Latent_State", values_to = "N_Countries")

  p <- ggplot(plot_data, aes(x = date_plot, y = N_Countries, fill = Latent_State)) +
    geom_col(position = "stack", width = 80) +
    scale_fill_brewer(palette = "Set1") +
    labs(title = paste("Evolution of Cluster Composition (k =", k, ")"),
         x = "Date", y = "Number of Countries", fill = "Latent State") +
    theme_minimal() +
    theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
    scale_x_date(date_labels = "%Y-Q%m", date_breaks = "1 year")

  return(p)
}
# countings

counts_tbl <- calculate_state_counts(mod, PC_scaled)

if (!is.null(counts_tbl)) {
    p <- plot_state_evolution(counts_tbl, k)
  print(p)

  plot_filename <- paste0("Evolution_Plot_k", k, ".png")
  ggsave(plot_filename, p, width = 8, height = 5, dpi = 300)

  table_filename <- paste0("Table_Image_k", k, ".png")
  my_title <- paste0("Cluster Counts Table (k=", k, ")")

  save_table_as_image(counts_tbl, table_filename, my_title)

} else {
}

# ================================================================================
# K = 2 trajectories

# pivot table with decoding
create_time_table <- function(mod, data_source) {

  decoding <- try(lmestDecoding(mod), silent = TRUE)
  if (inherits(decoding, "try-error") || is.null(decoding$UI)) {
    return(NULL)
  }

  data_source$Current_State <- decoding$UI[cbind(data_source$id_num, data_source$time_id)]

  time_table <- data_source %>%
    arrange(date, country_region) %>%
```

```
    group_by(date, Current_State) %>%
    summarise(
      Countries_List = paste(country_region, collapse = ", "),
      n_countries = n(),
      .groups = "drop"
    ) %>%
    pivot_wider(
      id_cols = date,
      names_from = Current_State,
      values_from = Countries_List,
      names_prefix = "STATE_"
    ) %>%
    mutate(across(starts_with("STATE_"), ~replace_na(., "-")))

  return(time_table)
}

# visualization
plot_heatmap <- function(mod, data) {

  data$date <- as.Date(data$date)
  decoding <- lmestDecoding(mod)
  data$State <- as.factor(decoding$Ul[cbind(data$id_num, data$time_id)])

  ggplot(data, aes(x = date, y = country_region, fill = State)) +
    geom_tile(color = "white") +
    scale_fill_manual(values = c("1" = "#4E79A7", "2" = "#E15759")) +
    scale_x_date(date_breaks = "1 year", date_labels = "%Y %b") +
    theme_minimal() +
    labs(title = paste("Trajectory of countries over time (k =", mod$k, ")"),
         x = "Quarter", y = "Country", fill = "Latent state") +
    theme(
      axis.text.x = element_text(angle = 45, hjust = 1, size = 8),
      panel.grid = element_blank()
    )
}

# table generation
tbl <- create_time_table(mod, PC_scaled)

if (!is.null(tbl)) {
  cat("First three quarters:\n")
  print(head(tbl, 3))

  csv_filename <- paste0("Chronological_Clusters_k", k, ".csv")
  write.csv(tbl, csv_filename, row.names = FALSE)
  p_heat <- plot_heatmap(mod, PC_scaled)

  # Saving PNG
  plot_filename <- paste0("trajectory_", k, ".png")
  ggsave(plot_filename, p_heat, width = 16, height = 7, dpi = 300)

} else {
  cat("Failed decoding\n")
}

# ================================================================================
# % of latent state occupancy

if (!is.null(mod)) {

  # decoding
  decoding <- lmestDecoding(mod)

  country_states <- PC_scaled %>%
    select(country_region, id_num, time_id) %>%
    arrange(id_num, time_id)

  country_states$Latent_State <- decoding$Ul[cbind(country_states$id_num, country_states$time_id)]

  # % computation
  state_persistence <- country_states %>%
    group_by(country_region, Latent_State) %>%
    summarise(Count = n(), .groups = "drop_last") %>%
    mutate(Percent = round((Count / sum(Count)) * 100, 1)) %>%
    ungroup() %>%
    select(country_region, Latent_State, Percent) %>%
    pivot_wider(
      names_from = Latent_State,
      values_from = Percent,
      names_prefix = "State ",
```

```
    values_fill = list(Percent = 0)
  ) %>%
  rename(
    Country = country_region,
    `State 1 (%)` = `State 1`,
    `State 2 (%)` = `State 2`
  ) %>%
  arrange(desc(`State 2 (%)`), Country)

# definition of the two groups
split_idx <- which(state_persistence$Country == "ROMANIA")
if(length(split_idx) == 0) split_idx <- ceiling(nrow(state_persistence) / 2)

df_left <- state_persistence[1:split_idx, ]
df_right <- state_persistence[(split_idx + 1):nrow(state_persistence), ]

# tables creation
my_theme_clean <- ttheme_default(
  core = list(
    bg_params = list(fill = c("white", "#f9f9f9"), col = "grey90"),
    fg_params = list(fontface = "plain", fontsize = 10)
  ),
  colhead = list(
    bg_params = list(fill = "#e6e6e6", col = "grey90"),
    fg_params = list(fontface = "bold", fontsize = 11)
  )
)

t_left <- tableGrob(df_left, rows = NULL, theme = my_theme_clean)
t_right <- tableGrob(df_right, rows = NULL, theme = my_theme_clean)

# merge of the two tables
tables_combined <- arrangeGrob(t_left, t_right, nrow = 1, padding = unit(1, "cm"))

# aesthetics
title_text <- "Frequency of Latent State Occupancy by Country (%)"
title_grob <- textGrob(title_text, gp = gpar(fontsize = 12, fontface = "bold"))

final_plot <- arrangeGrob(
  title_grob,
  tables_combined,
  nrow = 2,
  heights = unit.c(unit(1, "cm"), unit(1, "npc") - unit(1, "cm"))
)

h <- max(convertHeight(sum(t_left$heights), "in", valueOnly = TRUE),
         convertHeight(sum(t_right$heights), "in", valueOnly = TRUE)) + 1
w <- convertWidth(sum(t_left$widths), "in", valueOnly = TRUE) +
  convertWidth(sum(t_right$widths), "in", valueOnly = TRUE) + 2

ggsave("Table_State_Persistence_Wide.png", final_plot, width = w, height = h, dpi = 300)


} else {
  warning("error")
}

# ================================================================================
# LATENT MARKOV MODEL (COVARIATES MODEL)

PC_lmest_cov <- pc_master %>%
  arrange(country_region, date) %>%
  group_by(country_region) %>%
  mutate(
    id_num = cur_group_id(),
    time_id = row_number()
  ) %>%
  ungroup() %>%
  na.omit()

# standardization

PC_scaled_cov <- PC_lmest_cov %>%
  group_by(country_region) %>%
  mutate(across(6:21, scale)) %>%
  ungroup()

head(PC_scaled_cov)

# checks
```

```
cat("N of countries:", n_distinct(PC_scaled_cov$country_region), "\n")
cat("N of Quarters:", n_distinct(PC_scaled_cov$time_id), "\n")
cat("Total observations:", nrow(PC_scaled_cov), "\n")
cat("Time interval:", min(pc_master$date), "-", max(pc_master$date), "\n")

# model generation -> k from 2 to 5

# containers for results
models_list_cov <- list() # to save models
comparison_table_cov <- data.frame() # to save metrics

# formulation for LMMs with covariates (Complete)

my_formula_cov <- as.formula(premium_growth + loss_ratio + exp_ratio + margin ~ gdp + une_rate +
                        hicp + neet + gender_gap + oop_exp +
                        old_dep + t_edu + arop + dig_bank + life_exp + immigration_rate +
                        y10)

models_results <- list()
bic_values <- c()

for (k in k_values) {
  cat(paste0("\n--- k = ", k, " ---\n"))

  tryCatch({
    mod <- lmestCont(
      responsesFormula = my_formula_cov,
      data = PC_scaled_cov,
      index = c("id_num", "time_id"),
      k = k,
      modBasic = 1, # 1 = covariates in measurement model
      start = 0,
      out_se = TRUE
    )

    models_results[[paste0("k", k)]] <- mod
    bic_values <- c(bic_values, mod$Bic)

  }, error = function(e) {
  })
}

# results
summary_table <- data.frame(States = k_values, BIC = bic_values)

if(length(bic_values) > 0){
  best_k <- summary_table$States[which.min(bic_values)]
  cat(paste0("\nBest model for k = ", best_k, "\n"))

  best_model <- models_results[[paste0("k", best_k)]]
}

# output: comparison BIC, AIC, LogLik -> to choose the best k

final_comparison_cov <- data.frame()

for (k_name in names(models_results)) {
  mod <- models_results[[k_name]]

  if (is.null(mod) || is.null(mod$Lk)) next

  row <- data.frame(
    model = paste0("k=", mod$k),
    LogLik = round(mod$Lk, 1),
    AIC = round(mod$Aic, 1),
    BIC = round(mod$Bic, 1)
  )

  final_comparison_cov <- rbind(final_comparison_cov, row)
}

final_comparison_cov_word <- final_comparison_cov %>%
  rename(
    `Latent States` = model,
    `Log-Likelihood` = LogLik,
  )

# Saving PNG
save_table_as_image(
  final_comparison_cov_word,
  "Table_Model_Comparison_COVARIATES.png",
```

```
  "Information criteria comparison (k)"
)

# ================================================================================
# Backward stepwise selection (optimizing model k = 2)

all_covariates <- c("gdp", "une_rate", "hicp", "neet", "gender_gap",
                    "oop_exp", "old_dep", "t_edu", "arop", "dig_bank", "life_exp",
                    "immigration_rate", "y10")

current_covariates <- all_covariates
best_bic_so_far <- models_results[["k2"]]$Bic # k = 2
improved <- TRUE

# optimization loop
while(improved) {

  improved <- FALSE
  best_candidate_bic <- best_bic_so_far
  worst_var <- NULL

  # removing variables
  for(var in current_covariates) {

    test_covs <- setdiff(current_covariates, var)

    if(length(test_covs) == 0) next

    formula_str <- paste("premium_growth + loss_ratio + exp_ratio + margin ~",
                         paste(test_covs, collapse = " + "))
    test_formula <- as.formula(formula_str)

    tryCatch({
      mod_test <- lmestCont(
        responsesFormula = test_formula,
        data = PC_scaled_cov,
        index = c("id_num", "time_id"),
        k = 2,
        modBasic = 1,
        start = 0,
        out_se = FALSE
      )

      # if BIC gets better we keep it
      if(mod_test$Bic < best_candidate_bic) {
        best_candidate_bic <- mod_test$Bic
        worst_var <- var
      }

    }, error = function(e) NULL)
  }

  if(!is.null(worst_var)) {
    cat(paste0("Removed: '", worst_var, "' | new BIC: ", round(best_candidate_bic, 2), "\n"))
    current_covariates <- setdiff(current_covariates, worst_var)
    best_bic_so_far <- best_candidate_bic
    improved <- TRUE # go on
  } else {
    cat("Stop.\n")
  }
}

# best model
final_formula_str <- paste("premium_growth + loss_ratio + exp_ratio + margin ~",
                           paste(current_covariates, collapse = " + "))
final_formula <- as.formula(final_formula_str)

# Estimating again model k = 2
best_model_k2 <- lmestCont(
  responsesFormula = final_formula,
  data = PC_scaled_cov,
  index = c("id_num", "time_id"),
  k = 2,
  modBasic = 1,
  start = 0,
  out_se = TRUE
)

# trying to estimate model k = 3 with the new formula (doesn't work)

best_model_k3 <- lmestCont(
```

```
 responsesFormula = final_formula,
 data = PC_scaled_cov,
 index = c("id_num", "time_id"),
 k = 3,
 modBasic = 1,
 start = 0,
 out_se = TRUE
)

# ================================================================================
# Tables generation

if (!is.null(best_model_k2)) {

 # piv: Initial probabilities
 df_piv <- data.frame(
   Regime = c("State 1", "State 2"),
   Probability = as.numeric(round(best_model_k2$piv, 4)),
   check.names = FALSE
 )
 save_table_as_image(df_piv, "Table_Initial_Probabilities_COV.png", "Initial Probabilities")

 # Pi: transition probabilities
 Pi_slice <- best_model_k2$Pi[,,2]
 Pi_mat <- round(Pi_slice, 4)

 # df
 df_pi <- data.frame(
   `Regime` = c("State 1", "State 2"),
   `State 1` = Pi_mat[, 1], # first column
   `State 2` = Pi_mat[, 2], # second column
   check.names = FALSE
 )
 save_table_as_image(df_pi, "Table_Transition_Matrix_COV.png", "Transition Probability Matrix")
}

# intercepts for economic interpretation
resp_names <- c("Premium Growth", "Loss Ratio", "Expense Ratio", "Margin")

# as df
intercept_table_wider <- as.data.frame(round(best_model_k2$Al, 4))

colnames(intercept_table_wider) <- resp_names

intercept_table_wider <- cbind(Regime = c("State 1", "State 2"), intercept_table_wider)

save_table_as_image(intercept_table_wider, "intercept_table_cov.png", "Latent States Intercepts")

# beta coefficients regression
est_beta <- best_model_k2$Be
se_beta  <- best_model_k2$seBe
z_stats  <- est_beta / se_beta
p_values <- 2 * (1 - pnorm(abs(z_stats)))

cov_names <- c("GDP", "Unemployment Rate", "Old Dependency Ratio",
           "Tertiary Education", "Life Expectancy", "Immigration Rate", "Y10")
resp_names <- c("Premium Growth", "Loss Ratio", "Expense Ratio", "Margin")

# as df
results_df <- expand.grid(Variable = resp_names, Covariate = cov_names)
results_df$Estimate <- as.vector(est_beta)
results_df$PValue <- as.vector(p_values)

final_table_coeff <- results_df %>%
 mutate(
   Signif = case_when(
     PValue < 0.01 ~ "***",
     PValue < 0.05 ~ "**",
     PValue < 0.1  ~ "*",
     TRUE          ~ ""
   ),
   Formatted = paste0(round(Estimate, 3), Signif)
 ) %>%
 select(Covariate, Variable, Formatted) %>%
 pivot_wider(names_from = Variable, values_from = Formatted)

# saving PNG
save_table_as_image(final_table_coeff, "table_coeff_cov.png", "Regression Coefficients (Beta)")
```