mathematics

MDPI

Article

Identifiable Regime Detection in Pension-Fund Networks via Sticky Hidden Markov Models

Megang Nkamga Junile Staures * and Audrius Kabašinskas

Faculty of Mathematics and Natural Sciences, Kaunas University of Technology, 44249 Kaunas, Lithuania; audrius.kabasinskas@ktu.lt

* Correspondence: junmeg@ktu.lt

Abstract

We document a network-level vulnerability in pension-fund systems: lifecycle allocation regulation, designed to protect individual participants, compresses cross-fund return dynamics to the point where provider choice offers little diversification. Using daily net asset value data from second-pillar pension funds in Lithuania over 2019–2025, we found that a single common factor explains more than 72% of total return variance even in the calmest observed periods. We develop an unsupervised regime-detection framework that combines a PCA-based absorption ratio, DTW-based hierarchical clustering, and a Gaussian hidden Markov model with a data-driven crisis threshold. The HMM specification is supported by a dual empirical calibration of the stickiness prior, and its emission estimates agree closely with a fully Bayesian sticky-HMM specification. The framework identifies three latent regimes in which elevated systemic co-movement is the structural norm rather than an exceptional state and shows that funds separate first by lifecycle segment (conservative versus growth cohorts) and only secondarily by provider, with no single label axis reproducing the structure on its own. The absorption ratio has no significant relationship with global equity benchmarks in either calm or High-Concentration regimes, indicating that the detected regimes are not explained by the external benchmarks considered, including global equity indices and a euro-area rate/bond proxy. Cluster-level mean-absolute-active-return amplification of $1.06\times$ to $1.33\times$ during High-Concentration episodes confirms that even conservative funds serving retirement-age participants are not insulated.

Keywords: systemic risk; pension funds; absorption ratio; Hidden Markov Model; regime detection; lifecycle regulation

MSC: 62M05; 91G45; 62P05; 62M10; 91B84

Academic Editor: Manuel Alberto M. Ferreira

Received: 30 May 2026
Revised: 1 July 2026
Accepted: 6 July 2026
Published: 8 July 2026

1. Introduction

Systemic risk, the potential for shocks to propagate across financial systems, has been prominent since the 2008 crisis, revealing how the failure of individual institutions can destabilise economies [1]. While traditionally associated with banks because of their central role in credit intermediation, attention is now shifting to large non-bank institutional investors, especially pension funds, which hold over $63 trillion in assets globally [2]. The risk arises not simply from scale; research demonstrates that overlapping portfolios among institutions allow asset sales or losses to depress prices and transmit shocks, amplifying stress throughout the system [3–5]. Notably, pension funds form networks characterised by such overlaps, making indirect contagion possible even without direct financial links.

Unlike banks, pension funds typically have longer investment horizons, limited leverage, and focus on hedging [6]. Yet, these safeguards do not dissolve the vulnerability that Cont and Schaanning [5] identify, because that vulnerability lives in the overlap structure rather than in leverage or direct exposure. Fire-sale dynamics and portfolio overlap can still create indirect contagion channels even when no direct financial links exist between funds [3,4].

The question is therefore not whether such financial networks can transmit systemic stress, but how one can identify when they have entered a High-Concentration state. The Financial Stability Board identifies counterparty exposures, asset-liquidation dynamics, and the provision of critical functions as key channels through which distress can propagate across financial institutions [7]. Recent episodes such as the 2022 UK liability-driven investment (LDI) pension-fund crisis demonstrate how liquidity stress, collateral calls, and forced asset sales can amplify systemic stress across interconnected markets [8].

Indeed, existing literature increasingly frames the risks confronting pension schemes as fundamentally unhedgeable, system-level risks that cannot be diversified away through portfolio construction alone [9]. The enterprise risk management tradition reinforces this view [10,11]: Ref. [11] shows formally that a firm sponsoring a defined-benefit pension plan incurs compounding vulnerabilities unless pension risk is integrated with all other business risks in a unified optimisation. Their model demonstrates that managing pension risk in a silo, rather than holistically, reduces expected firm value by over 12%. This is not merely an actuarial accounting problem; it reveals that portfolio homogeneity and shared liability structures across pension participants create interdependencies that travel well beyond the balance sheet of any single fund.

Before turning to how systemic risk has been modelled, it is worth naming the obstacle that (we will argue) every existing approach must confront, and that ultimately defines the gap this paper fills. Systemic episodes in pension funds are rare and slow-moving. Ref. [12] addresses this through modular modelling but notes that traditional balance-sheet-based indicators often reflect past conditions and cannot reliably predict the onset of stress. Ref. [13] surveying the broader machine learning for financial risk management literature arrives at a complementary diagnosis from the opposite direction: across market, credit, operational, and insurance risk [14], the dominant paradigm is supervised classification trained on labelled crises or externally specified targets, and the field's principal open challenges, data inaccessibility, label scarcity, imbalanced datasets, and biased portfolio-level estimates, all trace back to the same root that [12] identifies. Where [12] argues from systems-design first principles that the labels cannot be there, ref. [13] documents, study by study, that the field has been quietly assuming they are. Four main modelling traditions respond to this constraint, but each leaves gaps, particularly for pension-fund networks.

Financial network frameworks often model systemic risk using explicit matrices of bilateral obligations [15,16], which are well-suited to banking but less applicable to pension funds, whose connections are indirect. A complementary strand models systemic risk through network density and contagion probability directly, showing that the relationship between interconnectedness and expected loss is non-monotonic; neither fully connected nor ring networks are optimal because denser connections simultaneously enable risk sharing and multiply contagion pathways [17]. Formalising when and how such contagion becomes systemic, however, requires a rigorous measurement framework. The mathematical foundations for measuring such systemic risk rigorously were laid by [18], who developed a general framework for systemic-risk measures via multi-dimensional acceptance sets and aggregation functions, unifying both the *first aggregate, then inject capital*, and the *first inject capital, then aggregate* paradigms into a single axiomatic structure. This framework makes explicit that the choice of aggregation rule and acceptance set jointly

https://doi.org/10.3390/math14142463

determines where the contagion boundary lies, a theoretical insight that later computational work has sought to operationalise. This foundation has recently been extended in two influential directions. Ref. [19] generalises systemic-risk measures to graph-structured data, modelling the system as a random asset vector together with a random interbank liability matrix, and proving the existence of optimal bailout-capital allocations that secure the network before contagion occurs. In a closely related vein, ref. [20] operationalises the *first allocate, then aggregate* paradigm from [18], using a deep learning scheme to compute the minimal cash needed to secure a system when no closed-form solution exists. These models are mainly normative, asking how to allocate capital given a detected crisis, rather than diagnosing when a High-Concentration regime emerges. The supervised machine learning tradition emerged partly to fill this diagnostic gap but, in doing so, introduced a different and equally binding constraint: the need for labelled crises. Recent advances integrating network analysis with extreme dependency modelling and machine learning [21–27] improve early warning capabilities and interpretability but remain reliant on labelled historical crises.

Deep neural networks and hybrid architectures achieve high predictive accuracy but deepen rather than dissolve this label dependency [23,28,29]. The relationship between risk factors and crisis outcomes is nonlinear [28], and more sophisticated architectures confirm rather than escape this; a hybrid XGBoost-LSTM model can issue warnings 2.3 quarters in advance with 83.7% accuracy [29], yet every such figure is conditional on prior crisis labelling. This label dependency is not merely a technical inconvenience but a structural constraint on the entire supervised paradigm: Ref. [30] demonstrates that even well-specified ensemble classifiers such as random forests, which outperform panel logit models on global banking crisis data with AUROC exceeding 81%, require crisis labels drawn from externally defined databases such as [31], and their generalisability across heterogeneous economies remains contingent on the completeness and accuracy of those labels. Interpretability tools do not eliminate label dependence: Ref. [32] use Random Forest variable-importance measures to identify real exchange-rate appreciation, overvaluation, and exchange-rate volatility as major crisis predictors, but these insights remain conditional on a predefined crisis-labelling scheme, illustrating that explanation methods operate within, rather than circumvent, the supervised paradigm. Such a crisis label is unavailable in the pension fund setting, where NAV-based returns are the primary observable. These requirements are feasible in banking and broad equity markets, but much less so for pension fund networks, where systemic episodes are rare, slow-moving, and not cleanly labelled, and where the central question is whether elevated-risk regimes can be recognised without prior labels.

Unsupervised data-driven methods present a promising alternative. Ref. [33] introduces the absorption ratio, the fraction of total asset-return variance explained by a finite number of eigenvectors from a rolling PCA, as an unsupervised, threshold-free indicator of market fragility. The logic is direct: when a single dominant factor absorbs the bulk of return variation across assets, those assets are tightly coupled, and a shock propagates more quickly and broadly than when risks are dispersed. Applied to U.S. equity industries over 1998–2010, the absorption ratio rose to its highest recorded level during the global financial crisis of 2008, and spikes in the standardised shift in the absorption ratio preceded all of the 1% worst monthly drawdowns. Crucially, ref. [33] derives its signal entirely from observed returns: no crisis labels, no externally chosen thresholds, and no imposed regime classification. The same PCA-based logic has since been applied to broader contagion analysis [1] and forms the methodological core of our own absorption ratio component. Studies also apply clustering and anomaly detection to textual or market data [34,35], but typically do not operate directly on institutional financial positions. Their network is a network of what is *said about* firms, not of what firms *hold*, which is exactly the overlap-based contagion substrate that [5] shows to be the operative one for institutions with common asset holdings. The broader systemic-risk machine learning literature confirms how isolated this position is. The authors of Ref. [36], in their survey of machine learning for systemic risk, classify the field into four branches and identify data-driven systemic-risk analysis as a major future direction precisely because real interbank-network data are difficult to obtain and most existing work relies either on proxy networks or on externally specified targets. Endogenous regime detection operating directly on actual financial positions does not appear in their taxonomy as an established branch; it appears as a gap.

The argument across the preceding traditions can now be stated crisply, and it defines the specific opening this paper fills. Ref. [16] shows *formally* that network topology governs how shocks clear but requires observed bilateral liabilities. Ref. [37] shows *empirically* that PageRank and interbank-exposure ratios are the dominant drivers of systemic impact and vulnerability, respectively, but requires observed bilateral exposures and synthetic DebtRank labels. Ref. [28] shows that nonlinear classifiers substantially outperform linear ones for crisis detection but amplify rather than resolve the label dependency. Each step increases explanatory power while deepening the data requirements. The pension case sits at the intersection of all three constraints: no bilateral liabilities, no DebtRank-computable exposures, and no labelled crises. The natural response is to step back to an earlier, harder question: can the *state* of the network be identified directly from returns, without any of these inputs?

Within unsupervised methods applied directly to financial returns, the toolkit is well established: PCA-based connectedness measures [1,33]; DTW-based time-series clustering, robust to temporal misalignment and well suited to financial series [38]; and Markov regime-switching, whose foundations [39] motivate the HMM approach we adopt. The CRISP-DM framework [40] provides a reproducible scaffold for assembling these components. Notably, the most relevant pension-specific precedents identify market regimes via HMM and stress-testing frameworks, anchored to external benchmarks [41,42]. Together, these establish both the network embeddedness of pension funds and the usefulness of regime-switching methods in this exact setting, yet remain conditional on external states rather than allowing risk regimes to emerge endogenously from the pension network itself. In [12]'s terms, a benchmark-anchored regime is a measurement of a previous, externally defined state rather than a forward-looking read of the network's own condition.

The framework is designed for funded pension systems in which several pension funds or investment options publish regular NAV or return series. It is therefore directly applicable to second-pillar defined-contribution systems, lifecycle pension funds, target-date funds, and other multi-fund funded arrangements. It is not intended as a standalone tool for pure PAYGO systems, single defined-benefit plans without comparable fund-level return data, or pension systems where the relevant risk is primarily actuarial rather than market-valued portfolio co-movement.

We address this gap by developing an unsupervised regime-detection framework that treats systemic risk as an emergent property of the pension-fund network, rather than as a response to externally defined market conditions. The proposed approach integrates three components within a unified empirical pipeline: a PCA-based absorption ratio following the logic of [33] to measure system-wide co-movement; DTW-based hierarchical clustering to capture cross-sectional similarity structure; and a Gaussian HMM to identify latent risk regimes over time. Importantly, and in the spirit of the threshold-free, data-derived discipline shared by [33–35], the crisis threshold is derived endogenously from the HMM emission distributions, ensuring internal consistency between regime classification and crisis identification.

This paper is a measurement contribution rather than a structural model of pension-system equilibrium. Overlapping-generations (OLG) models [43,44], DSGE models [45], and lifecycle portfolio models [46] are designed to analyse household saving, portfolio choice, and policy counterfactuals under explicit behavioural assumptions. Our framework asks a narrower question: whether latent co-movement regimes and cohort-dominant clustering can be detected from observed fund returns without labelled crisis episodes or holdings data. It therefore identifies a measurable network pattern, but it does not predict how providers would re-optimise under alternative regulation.

Three methodological contributions distinguish this framework. First, the number of latent states is selected by combining BIC with a posterior identifiability analysis in a finite sticky HMM [47], addressing BIC's tendency to over-select when mixture components overlap. Second, the stickiness prior is calibrated empirically from the data through dual anchors corresponding to the baseline and the BIC-optimal specification, reducing the arbitrariness of prior selection. Third, emission estimates from the frequentist Baum–Welch algorithm are compared against a fully Bayesian sticky HMM estimated via the No-U-Turn Sampler, with the two agreeing closely, indicating the three-regime structure is driven by the data rather than by either estimation route. The framework is applied to daily NAV data from $N = 40$ second-pillar pension funds in Lithuania managed by five providers over January 2019 to September 2025, augmented with fund-specific benchmarks, global market indices and euro-area interest-rate factors.

The empirical analysis yields four main findings. First, systemic co-movement is persistently high: the absorption ratio computed using the method of [33] never falls below 0.728, indicating that a single common factor explains more than 72% of total return variance even in the calmest periods. Clustering results show that funds separate first by lifecycle segment and only secondarily by provider, suggesting that lifecycle allocation regulation, rather than managerial differences, is the dominant driver of cross-fund dependence, a structural reading consistent with the finding that pension risk behaviour is regulation-driven [11,48]. Second, three latent concentration regimes are identified, with moderate- and High-Concentration states accounting for nearly 87% of trading days, indicating that elevated systemic interaction is the network's structural norm. Third, systemic stress is not a passive reflection of global markets: the relationship between the absorption ratio and global benchmarks is statistically insignificant across regimes, while mean absolute active return amplification, the precise quantity practitioners monitor [48], is concentrated in the High-Concentration state. Fourth, the empirically derived crisis threshold $\tau = 0.845$ exceeds the values commonly used in network-based contagion studies, where thresholds are typically chosen heuristically [49]. This is consistent with [5]'s demonstration that contagion ignites only beyond a system-specific tipping point, one that depends on leverage, concentration, and the commonality of holdings rather than on any universal shock level. More broadly, the finding that an internally derived threshold sits above the heuristic values routinely used in banking-network studies echoes the methodological conclusion that [16] reaches from the opposite direction: the clearing structure of a specific network, its liabilities, its cash-flow distribution, and the topology of its obligations determine where the contagion boundary lies, and that boundary cannot be read off from first principles or borrowed from other sectors.

The remainder of the paper is structured as follows. Section 2 develops the theoretical framework and describes the empirical methodology, including the data, the absorption ratio, DTW clustering, and the Hidden Markov Model. Section 3 presents the main results, including the regime detection, the cluster structure, and the cross-step analysis connecting them. Section 4 discusses the findings and their policy implications. Section 5 concludes and outlines directions for future research.

2. Theoretical Framework and Methodology

This section develops the theoretical framework and explains how it is implemented empirically. The analysis follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) [40], which structures the workflow into data understanding, preparation, modelling, and evaluation. The three-step empirical pipeline, PCA absorption ratio, DTW-based hierarchical clustering, and HMM regime detection are summarised in Figure 1. Each step is presented together with the theoretical results that justify it.

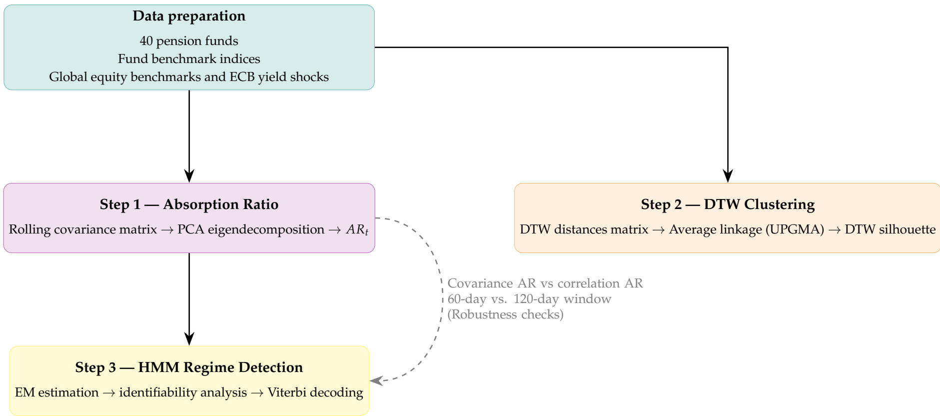

Figure 1. Empirical workflow. The analysis combines pension-fund returns with fund-specific benchmarks, global equity benchmarks, and ECB yield shocks. The pension-fund return matrix $R^{(f)} \in \mathbb{R}^{T\times N}$ is analysed along two complementary branches. The first branch constructs a rolling absorption-ratio series from the covariance matrix of fund returns and uses this series as the input to a Gaussian Hidden Markov Model for regime estimation, identifiability assessment, and concentration-regime classification. Global equity returns and ECB yield shocks are employed only for the economic interpretation and validation of the inferred regimes and are not included in the HMM estimation itself. The second branch characterises cross-sectional similarity among pension funds using Dynamic Time Warping (DTW) dissimilarities and average-linkage (UPGMA) hierarchical clustering, with the number of clusters selected using the DTW silhouette coefficient. The dashed path denotes robustness analyses based on alternative absorption-ratio specifications, including covariance- versus correlation-based estimation and alternative rolling-window lengths (60 versus 120 days).

2.1. Data

The dataset comprises daily NAV observations from pension funds operating in Lithuania. The sample covers $N = 40$ funds managed by five pension-fund providers, anonymised as Providers 1 through 5 to comply with data confidentiality requirements. Each provider manages eight age-cohort funds, yielding a structured 5×8 cross-sectional panel. The observation period runs from January 2019 to September 2025, yielding approximately 1700 daily trading observations per fund and a total of 67,794 fund-day observations. The panel is strongly balanced, with each of the 40 funds contributing between 1692 and 1696 observations over the sample period. After aligning all pension-fund, benchmark, and global factor series on their common sample period, the final estimation dataset contains approximately $T = 1634$ daily observations.

The age-cohort design reflects differences in lifecycle allocation across funds. Seven cohorts are defined by participant birth decade: cohorts AG1 through AG7 cover birth years from 1954–1960 to 1996–2002, capturing participants at progressively earlier career stages. The eighth cohort, AG8, corresponds to the Target Income Pension Fund (TIPF), a specialist vehicle for participants who have reached retirement age and are actively drawing income. This distinction is empirically consequential: preliminary descriptive analysis of the sample reveals that TIPF funds exhibit substantially lower daily return volatility than age-cohort funds, consistent with their capital preservation mandate (see Table 1). This distinction anticipates the finding in Section 3.2 that extreme-cohort funds (AG1 and AG8) separate sharply from middle cohorts across all five providers, forming dedicated clusters with mean absolute active return three to six times lower than those of middle-cohort clusters The presence of TIPF funds is therefore analytically significant, as they provide a conservative baseline against which the behaviour of age-cohort funds during stress periods can be assessed.

Table 1. Descriptive statistics of pension-fund returns and global factors. Pension-fund series are reported by age cohort (pooled across providers) and for the full sample. The global factor set consists of three equity-market benchmark returns and two interest-rate shock factors. Skewness and excess kurtosis are reported. All series exhibit negative skewness and substantial excess kurtosis, consistent with fat-tailed return distributions.

Series	Mean	Std	Min	Max	Skewness	Kurtosis
Pension funds by age cohort						
AG1 (54/60)	0.0001	0.0022	−0.0301	0.0159	−1.703	17.977
AG2 (61/67)	0.0002	0.0043	−0.0551	0.0371	−1.526	17.447
AG3 (68/74)	0.0004	0.0069	−0.0867	0.0661	−1.208	13.941
AG4 (75/81)	0.0004	0.0080	−0.0868	0.0662	−1.050	10.825
AG5 (82/88)	0.0004	0.0081	−0.0864	0.0659	−1.042	10.720
AG6 (89/95)	0.0004	0.0081	−0.0866	0.0663	−1.039	10.730
AG7 (96/02)	0.0004	0.0083	−0.0902	0.0695	−1.004	10.668
AG8 (TIPF)	0.0001	0.0019	−0.0235	0.0132	−1.495	14.549
Full sample	0.0003	0.0065	−0.0902	0.0695	−1.217	16.882
Global equity factors						
MSCI World	0.0004	0.0128	−0.1239	0.0871	−1.183	15.784
MSCI Europe	0.0005	0.0121	−0.1208	0.0871	−0.952	17.299
S&P 500	0.0006	0.0127	−0.1277	0.0909	−0.679	15.519
Interest-rate factors (basis points)						
ΔY_{10Y}	0.145	4.803	−26.464	23.615	−0.288	3.201
ΔY_{2Y}	0.152	4.559	−43.467	26.754	−0.733	10.795

For each fund and trading day, the dataset contains the fund unit price, an associated comparative benchmark index value, a unique fund identifier, and the corresponding age-group classification. The dataset is further augmented with $M = 5$ global market factors. Three factors are represented by broad equity-market benchmarks (MSCI World, MSCI Europe, and the S&P 500), while two factors capture changes in long-term and short-term euro-area interest rates using daily changes in the 10-year and 2-year AAA euro-area government bond yields. The sample period includes the COVID-19 market shock and the subsequent monetary-policy tightening cycle, providing natural benchmarks for validating the regime-detection framework against major episodes of financial stress.

2.1.1. Notation and Setup

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Consider N pension funds observed over T trading days, and let $X_{i,t}$ denote the NAV of fund i at time t, for $i = 1, \ldots, N$ and $t = 0, \ldots, T$. Each fund i is associated with a comparative benchmark index $Z^{(f)}_{i,t}$, reflecting its underlying investment strategy.

Daily pension-fund and benchmark returns are defined by

$$\begin{aligned} r^{(f)}_{i,t} &= \log(X_{i,t}) - \log(X_{i,t-1}), \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \\ r^{(b,f)}_{i,t} &= \log(Z^{(f)}_{i,t}) - \log(Z^{(f)}_{i,t-1}), \quad i = 1, \ldots, N. \end{aligned} \tag{1}$$

In addition, let $g_{m,t}$, $m = 1, \ldots, M$ denote a set of global market factors. These factors include both equity-market benchmark returns and interest-rate factors. Specifically,

$$g_{m,t} = \begin{cases} \log(G_{m,t}) - \log(G_{m,t-1}), & \text{for equity-market benchmarks,} \\ 100(G_{m,t} - G_{m,t-1}), & \text{for interest-rate factors.} \end{cases}$$

Let $R^{(f)} \in \mathbb{R}^{T \times N}$ denote the pension-fund return matrix and $R^{(b,f)} \in \mathbb{R}^{T \times N}$ the fund-specific benchmark return matrix. Further, let

$$G^{(b)} = \left[R_{\text{MSCI World}},\ R_{\text{MSCI Europe}},\ R_{\text{S\&P500}},\ \Delta Y_{10Y},\ \Delta Y_{2Y}\right] \in \mathbb{R}^{T \times M}$$

denote the global factor matrix, where $M = 5$. The first three columns correspond to daily log returns of major global equity benchmarks, while the final two columns correspond to daily changes in long-term and short-term euro-area interest rates measured in basis points.

After aligning all return and factor series on their common observation period, let $\hat{T}$ denote the number of remaining observations. When required, we define the augmented return matrix

$$A = \left[R^{(f)} \quad R^{(b,f)} \quad G^{(b)}\right] \in \mathbb{R}^{\hat{T} \times (2N+M)}.$$

In the empirical application, $N = 40$ and $M = 5$, implying

$$A \in \mathbb{R}^{\hat{T} \times 85}.$$

For a rolling window of length ω, define the windowed return matrix A_t for $t = \omega, \ldots, \hat{T}$, containing observations from periods $t - \omega + 1$ through t.

2.1.2. Descriptive Statistics

Table 1 reports descriptive statistics for the daily pension-fund return series, aggregated by age cohort, together with the global factor set used in the empirical analysis. The global factor set consists of three equity-market benchmark returns (MSCI World, MSCI Europe, and the S&P 500) and two interest-rate factors represented by daily changes in the 10-year and 2-year AAA euro-area government bond yields. Across nearly all series, negative skewness and substantial excess kurtosis indicate return distributions that deviate markedly from normality and exhibit pronounced tail risk. These distributional characteristics are consistent with the presence of extreme market events during the sample period, including the COVID-19 shock and the subsequent monetary-policy tightening cycle.

A clear lifecycle gradient is visible across the pension-fund cohorts. The oldest active cohort (AG1, born 1954–1960) and the Target Income Pension Fund (AG8) exhibit the lowest daily volatility, consistent with their conservative asset allocations and capital-preservation objectives. Volatility increases progressively with cohort youth, reaching 0.0083 for the youngest cohort (AG7, born 1996–2002), reflecting the higher equity exposures associated with longer investment horizons. Mean returns also increase slightly across cohorts, although the differences are economically modest relative to volatility.

The pension-fund return distributions exhibit substantial asymmetry and tail risk. Skewness is most negative for AG1, AG2, and AG8, indicating a greater sensitivity to adverse market movements relative to positive return episodes. Excess kurtosis ranges from approximately 10.7 to 18.0 across cohorts, suggesting that extreme return observations occur considerably more frequently than would be expected under a Gaussian distribution. The highest kurtosis values are observed among the most conservative cohorts, indicating that infrequent but pronounced portfolio revaluations may have a disproportionately large effect on funds with otherwise low day-to-day volatility.

Importantly, the high unconditional kurtosis reported in Table 1 does not invalidate the Gaussian emission assumption of the Hidden Markov Model introduced in Section 2.4. A regime-switching process may generate strongly non-Gaussian unconditional distributions even when the conditional distribution within each latent regime is approximately Gaussian. The observed fat tails, therefore, motivate, rather than contradict, the proposed regime-detection framework.

The three equity-market factors display similar statistical characteristics. Daily volatility ranges between 1.2% and 1.3%, while all three series exhibit negative skewness and substantial excess kurtosis. These results indicate that the extreme return behaviour observed in pension-fund portfolios is mirrored by broad global equity markets over the sample period.

The interest-rate factors are centred near zero, with mean daily changes of 0.145 and 0.152 basis points for the 10-year and 2-year maturities, respectively. Both series exhibit negative skewness, indicating a tendency toward larger downward than upward yield movements. The 2-year factor exhibits substantially higher excess kurtosis (10.795 versus 3.201), indicating a greater propensity for extreme short-term rate movements. This finding is consistent with the greater sensitivity of shorter maturities to changes in monetary-policy expectations and interest-rate repricing.

2.2. Step 1: The PCA Absorption Ratio

2.2.1. Definition and Computation

The rolling window is set to $\omega = 60$ trading days, approximately one calendar quarter. This balances two competing requirements: shorter windows react faster to stress but yield noisier covariance estimates for $N = 40$ funds, while longer windows stabilise the covariance matrix at the cost of blurring transitions and mechanically inflating regime persistence. We treat 60 days as the baseline and report sensitivity to 40-, 90-, and 120-day windows in Section 3.3.6.

Definition 1 (Rolling Covariance Matrix)**.** *The rolling sample covariance matrix at time t is*

$$\Sigma_t = \frac{1}{\omega - 1} R_t^\top M R_t,$$

where $M = I_\omega - \frac{1}{\omega}\mathbf{1}_\omega \mathbf{1}_\omega^\top$ is the centering matrix. Since Σ_t is real, symmetric, and positive semi-definite, it admits an eigendecomposition. For the baseline fund-only matrix $\Sigma_t^{(f)} \in \mathbb{R}^{N \times N}$, let $\lambda_1(t) \geq \cdots \geq \lambda_N(t)$ denote its eigenvalues; the first principal component corresponds to $\lambda_1(t)$. For the augmented robustness matrix $\Sigma_t^{(aug)} \in \mathbb{R}^{(2N+M)\times(2N+M)}$, the eigenvalues are indexed $\lambda_1^{aug}(t), \ldots, \lambda_{2N+M}^{aug}(t)$.

PCA applied to the return covariance matrix is a standard tool for identifying the dominant common factors driving asset co-movement [50–52].

Definition 2 (PCA Absorption Ratio)**.** *The absorption ratio at time t is the fraction of total fund-return variance explained by the first principal component:*

$$AR_t = \frac{\lambda_1(t)}{\sum_{j=1}^{N} \lambda_j(t)} = \frac{\lambda_1(t)}{\mathrm{tr}(\Sigma_t)}.$$

A high AR_t means a single common factor drives most of the variation across the N funds, the signature of systemic co-movement.

2.2.2. Empirical Implementation

At each t from ω to $\hat{T}$, the windowed matrix $R_t^{(f)}$ is extracted, its unbiased covariance Σ_t formed, and $AR_t = \lambda_1(t)/\mathrm{tr}(\Sigma_t)$ computed by eigendecomposition. The resulting series $\{AR_t\}_{t=\omega}^{\hat{T}}$ measures the share of return variance absorbed by the leading factor at quarterly resolution; a rising AR_t signals greater commonality, consistent with the monotone relationship of Proposition 3.

2.2.3. Key Properties

Proposition 1 (Boundedness)**.** *For all t, $AR_t \in \left[\frac{1}{N}, 1\right]$.*

Proof. By eigenvalue ordering, $\lambda_1(t) \leq \mathrm{tr}(\Sigma_t)$, giving $AR_t \leq 1$. Since $\lambda_1(t) \geq \lambda_j(t)$ for all j, summing gives $N\lambda_1(t) \geq \mathrm{tr}(\Sigma_t)$, so $AR_t \geq \frac{1}{N}$. $\square$

Proposition 2 (Interpretation of Extremes)**.** *$AR_t = \frac{1}{N}$ iff all eigenvalues are equal. $AR_t = 1$ iff* $\mathrm{rank}(\Sigma_t) = 1$.

Proof. $AR_t = \frac{1}{N}$ implies $\lambda_1 = \mathrm{tr}(\Sigma_t)/N$; equal eigenvalues are the unique PSD solution. $AR_t = 1$ implies $\lambda_j = 0$ for $j \geq 2$, so $\mathrm{rank}(\Sigma_t) = 1$. $\square$

Proposition 3 (Monotonicity in Cross-Fund Correlation)**.** *Let $\bar{\rho}_t$ be the average pairwise return correlation. Under the equicorrelation model $\Sigma_t = \sigma^2[(1-\bar{\rho}_t)I_N + \bar{\rho}_t \mathbf{1}_N \mathbf{1}_N^\top]$,*

$$AR_t = \frac{1 + (N-1)\bar{\rho}_t}{N},$$

strictly increasing in $\bar{\rho}_t$ for $\bar{\rho}_t \in \left(-\frac{1}{N-1}, 1\right]$.

Proof. Under equicorrelation, Σ_t has eigenvalue $\lambda_1 = \sigma^2(1 + (N-1)\bar{\rho}_t)$ and $N-1$ repeated eigenvalues $\sigma^2(1-\bar{\rho}_t)$, so $\mathrm{tr}(\Sigma_t) = N\sigma^2$ and $AR_t = (1 + (N-1)\bar{\rho}_t)/N$. The derivative $(N-1)/N > 0$ gives strict monotonicity. $\square$

Proposition 3 links the absorption ratio directly to cross-fund dependence: greater co-movement raises the variance share absorbed by the leading component. The mapping is exact only under equal variances. Because fund volatilities differ across lifecycle cohorts in the empirical panel, AR_t should be read as an approximate, rather than exact, transform of $\bar{\rho}_t$; the correlation-based robustness check below restores the exact mapping. This single caveat applies wherever AR_t is interpreted in correlation terms.

Theorem 1 (Crisis Identification under Equicorrelation)**.** *Let $C_t = \mathbb{I}\{AR_t > \tau\}$, $\tau \in \left(\frac{1}{N}, 1\right)$, and $\bar{\rho}_{\mathrm{crit}} = (N\tau - 1)/(N-1)$. Under equicorrelation, $C_t = 1 \iff \bar{\rho}_t > \bar{\rho}_{\mathrm{crit}}$.*

Theorem 1 maps the threshold τ to a critical average correlation. Subject to the approximation noted above, it supplies the economic reading of τ; the threshold itself is fixed not a priori but from the HMM emission distributions in Step 3 (Section 2.4).

2.2.4. Relationship to Standard Risk Decomposition

In factor-model terms, network risk decomposes as

$$\underbrace{\mathrm{tr}(\Sigma_t)}_{\text{Total}} = \underbrace{\lambda_1(t)}_{\text{Systematic}} + \underbrace{\textstyle\sum_{j=2}^{N} \lambda_j(t)}_{\text{Specific}}, \qquad AR_t = \frac{\text{Systematic}}{\text{Total}}.$$

This connects the spectral measure to standard portfolio risk metrics [52,53], but the economic reading inverts. In portfolio construction, a high systematic share signals that idiosyncratic risk has been diversified away [51,52]. In a pension network, it signals the opposite: when $AR_t = 0.87$, a single factor drives 87% of return variance and provider choice offers only 13% of fund-specific differentiation, a structural homogeneity that is a vulnerability, not an achievement.

2.2.5. Robustness: Estimator and Benchmark Inclusion

The baseline AR_t uses $R^{(f)}$ alone, keeping the measure fund-focused. The benchmark and global factors then serve two roles: as interpretive series compared against elevated-AR_t episodes, and in an augmented robustness specification that recomputes AR_t on $A = [R^{(f)}\ R^{(b,f)}\ G^{(b)}]$. A ridge term εI ($\varepsilon = 10^{-6}$) is added to Σ_t only in the augmented exercise, where $2N + M = 85 > \omega = 60$ makes Σ_t rank-deficient. Applying it to the full-rank baseline shifts AR_t by 0.01–0.05 across windows, confirming it must not be used there (see Section 3.1).

Volatility heterogeneity.

Because the covariance-based AR_t weights funds by variance, the leading eigenvalue reflects both common dependence and the higher volatility of younger cohorts. To isolate dependence, AR_t is recomputed from the rolling *correlation* matrix, which standardises each fund to unit variance and restores the exact AR–$\bar{\rho}$ mapping of Proposition 3. All downstream analyses are repeated on this scale-free series. As reported in Section 3.3.6, it lowers absorption levels and thresholds, as expected, but leaves the ordered low/moderate/high regime structure and the timing of elevated episodes unchanged, so the conclusions are not artefacts of volatility weighting.

Covariance estimator.

The unbiased sample covariance is the baseline because it preserves the full concentration of variance in $\lambda_1(t)$, the quantity AR_t exists to detect. Shrinkage (Ledoit–Wolf [54]), robust (MCD), and Random-Matrix-Theory filtering [55,56] each target a different goal, none aligned with detection. Recomputing AR_t under all three (Appendix B) leaves the conclusions intact: the MP-filtered series tracks the baseline almost exactly (correlation 0.98), confirming that $\lambda_1(t)$ sits well outside the noise bulk and reflects genuine co-movement; MCD agrees closely on average; and Ledoit–Wolf produces only a level shift consistent with its shrinkage design.

2.3. Step 2: DTW Hierarchical Clustering

2.3.1. Formal Definition

Let $\mathcal{X} = \{x = (x_1, \ldots, x_T) : x_t \in \mathbb{R}\}$ be the space of length-T series. For $x, y \in \mathcal{X}$, the local cost matrix has entries $C_{ij} = (x_i - y_j)^2$.

Definition 3 (Warping Path)**.** *A warping path* $W = \{(i_1, j_1), \ldots, (i_L, j_L)\}$ *satisfies: (i)* $(i_1, j_1) = (1, 1)$, $(i_L, j_L) = (T, T)$*; (ii)* $i_{k+1} \geq i_k$, $j_{k+1} \geq j_k$*; (iii)* $(i_{k+1} - i_k, j_{k+1} - j_k) \in \{(1, 0), (0, 1), (1, 1)\}$.

Definition 4 (DTW Distance [57])**.** $\mathrm{DTW}(x, y) = \min_W \sqrt{\sum_{(i,j)\in W} C_{ij}}$, *computed via* $D_{ij} = C_{ij} + \min\{D_{i-1,j}, D_{i,j-1}, D_{i-1,j-1}\}$, $D_{11} = C_{11}$, *with* $\mathrm{DTW}(x, y) = \sqrt{D_{TT}}$.

2.3.2. Semi-Metric Properties and Consequences

Theorem 2. *DTW satisfies non-negativity, identity, and symmetry, but not the triangle inequality, and so is not a metric.*

Proof. Non-negativity from $C_{ij} \geq 0$; identity from the zero-cost diagonal path at $x = y$; symmetry from $(x_i - y_j)^2 = (y_j - x_i)^2$. Triangle-inequality failure is standard [58]. □

The non-metric structure dictates two choices. First, the linkage: we use average linkage (UPGMA), which is defined directly on a pairwise dissimilarity matrix and so needs neither a Euclidean embedding nor the triangle inequality, unlike Ward's variance-minimising criterion [59], whose justification fails on DTW distances. Average linkage also yields a fully nested hierarchy for dendrogram inspection and resists the chaining that affects single linkage. Second, the validation: the number of clusters is chosen by the silhouette coefficient computed on the same DTW dissimilarity matrix, so construction and validation share one geometry. We avoid the Davies–Bouldin index because it is centroid-based and would evaluate a DTW partition in Euclidean space. Robustness against algorithm (DTW k-means) and distance (correlation, Euclidean) follows in Section 3.2.

2.3.3. Empirical Implementation

Each return series is standardised to zero mean and unit variance so clustering reflects dynamics rather than scale. The $N \times N$ DTW distance matrix is computed with `dtaidistance` (no warping-window constraint) and clustered by average linkage. The cluster count K_c^* maximises the silhouette over $K_c \in \{2, \ldots, 15\}$ on the precomputed matrix, reaching $K_c^* = 10$ (0.438); beyond this point the score declines, so finer partitions merely subdivide existing groups.

2.4. Step 3: Hidden Markov Model Regime Detection

2.4.1. The Absorption Ratio as a Stochastic Process

Before introducing a probabilistic model, we make explicit the sense in which the absorption ratio is a random object. On the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ introduced in Section 2.1.1, let $\{\mathcal{F}_t\}$ denote the filtration generated by the fund-return process, $\mathcal{F}_t = \sigma\big(r_s^{(f)} : s \leq t\big)$, where $r_s^{(f)} = (r_{1,s}^{(f)}, \ldots, r_{N,s}^{(f)})^\top$ is the s-th row of $R^{(f)}$.

For a rolling window of length ω, the windowed return matrix $R_t^{(f)}$ collects the returns from periods $t - \omega + 1$ through t, and the rolling covariance matrix of Step 1,

$$\Sigma_t = \frac{1}{\omega - 1} \, (R_t^{(f)})^\top M \, R_t^{(f)},$$

is a measurable function of $\{r_s^{(f)}\}_{s=t-\omega+1}^{t}$ and is therefore itself a random matrix adapted to $\mathcal{F}_t$. Its ordered eigenvalues $\lambda_1(t) \geq \cdots \geq \lambda_N(t)$ are continuous and hence measurable functions of Σ_t, so the absorption ratio

$$AR_t = \frac{\lambda_1(t)}{\mathrm{tr}(\Sigma_t)}$$

is a measurable transformation of random quantities. Consequently, $\{AR_t\}_{t=\omega}^{\hat{T}}$ is a real-valued stochastic process adapted to $\{\mathcal{F}_t\}$ rather than a deterministic series, which is what legitimises treating it as the observable output of a probabilistic generating mechanism and modelling it with the Hidden Markov specification of Section 2.4.2.

This stochastic character does not imply serial independence, and the two should not be conflated. The series inherits genuine temporal dependence from the persistence of the underlying covariance structure. It also inherits a *mechanical* component of dependence from the overlap of successive estimation windows: two windows separated by fewer than ω trading days share raw observations, while windows separated by ω days or more are disjoint. This overlap-induced dependence therefore has bounded support of at most $\omega - 1$ lags, a point we exploit in the robustness analysis of Section 2.4 to separate it from the longer-horizon economic persistence captured by the latent regimes.

2.4.2. Model Specification

Let $\{O_t\}_{t=1}^{T}$ denote the observed absorption ratio sequence, where $O_t = AR_t \in [\frac{1}{N}, 1]$. We model $\{O_t\}$ as generated by an unobserved discrete-time Markov chain $\{S_t\}_{t=1}^{T}$ taking values in the finite state space $\mathcal{S} = \{1, \ldots, K\}$.

Definition 5 (Hidden Markov Model)**.** *A K-state Hidden Markov Model (HMM) is characterised by the parameter set* $\lambda = (\pi, A, B)$*, where*

$$\begin{aligned} \pi_k &= \mathbb{P}(S_1 = k), && \textit{initial distribution}, \\ A_{ij} &= \mathbb{P}(S_t = j \mid S_{t-1} = i), && \textit{transition probabilities}, \\ B_k(o) &= \mathcal{N}(o; \mu_k, \sigma_k^2), && \textit{Gaussian emission density}. \end{aligned}$$

Conditional on the latent state $S_t = k$, observations therefore follow a Gaussian distribution, $AR_t \mid S_t = k \sim \mathcal{N}(\mu_k, \sigma_k^2)$.

2.4.3. Estimation, Decoding, and Long-Run Behaviour

Parameter estimation.

The absorption-ratio series obtained in Step 1 is modelled using Gaussian Hidden Markov Models estimated by maximum likelihood via the Baum–Welch expectation-maximisation (EM) algorithm using the forward–backward recursions. The resulting updates maximise the likelihood $\mathbb{P}(O \mid \lambda)$ and converge to a stationary point of the likelihood function under standard regularity conditions [60].

Conditional on state k, observations follow a Gaussian emission distribution with state-specific mean and variance. Gaussian HMMs with diagonal covariance matrices are estimated for

$$K \in \{2, \ldots, 8\},$$

using multiple random initialisations to reduce sensitivity to local optima. For each value of K, the model with the highest log-likelihood across initialisations is retained and subsequently evaluated using information-criterion and Bayesian identifiability diagnostics.

State decoding.

The most likely latent state sequence $S^* = (s_1^*, \ldots, s_T^*)$ is obtained using the Viterbi algorithm,

$$S^* = \arg\max_S \mathbb{P}(S \mid O, \lambda),$$

with computational complexity $\mathcal{O}(K^2 T)$. The states are subsequently ordered according to their estimated emission means $\hat{\mu}_k$. For the baseline three-state specification (K = 3), the ordered states are labelled "Low-Concentration", "Moderate-Concentration", and "High-Concentration" regimes. These labels describe the degree of common variation captured by the absorption ratio; economic interpretation is introduced subsequently through comparisons with benchmark market indicators.

Forecasting and long-run behaviour.

Conditional on the current state S_t, the distribution of S_{t+h} for any horizon $h \geq 1$ is obtained by iterating the transition kernel:

$$\mathbb{P}(S_{t+h} = k \mid S_t = i) = (A^h)_{ik},$$

where A^h denotes the hth matrix power of A. Letting $\mathbf{e}_i$ be the ith unit vector, the forward distribution over states is given by $\mathbf{e}_i^\top A^h$. If the chain is irreducible and aperiodic, A^h converges as $h \to \infty$ to a rank-one matrix whose rows equal the stationary distribution $\boldsymbol{\pi}^*$, defined as the unique probability vector satisfying

$$\boldsymbol{\pi}^{*\top} A = \boldsymbol{\pi}^{*\top}, \qquad \boldsymbol{\pi}^{*\top}\mathbf{1} = 1.$$

Equivalently, $\boldsymbol{\pi}^*$ is the left eigenvector of A associated with the eigenvalue 1, and represents the long-run unconditional probability of occupying each regime, independently of the starting state. Forward state probabilities are reported at horizons of one day, one week (five trading days), and one month (twenty trading days). In addition, the stationary distribution is reported to characterise the long-run unconditional occupancy of each regime.

2.4.4. Model Selection

BIC-based screening.

Candidate models are first evaluated using the Bayesian Information Criterion (BIC),

$$\mathrm{BIC}(K) = -2\log \mathbb{P}(O \mid \hat{\lambda}^{(K)}) + d(K)\log T,$$

where $d(K) = (K-1) + K(K-1) + 2K = K^2 + 2K - 1$ is the number of free parameters: K^2 transition probabilities (including the initial state distribution) and $2K$ Gaussian emission parameters. While BIC provides a standard likelihood-based selection criterion, it tends to favour more complex models that may include weakly identified states [61,62]. A complementary Bayesian identifiability analysis is therefore conducted to discriminate genuine regime structure from over-parameterisation.

Calibration of regime persistence.

The Bayesian specification employs a sticky transition prior to encourage persistence of latent regimes. For each state k, the transition probabilities follow a Dirichlet prior,

$$\boldsymbol{\pi}_k \sim \mathrm{Dirichlet}(\alpha\mathbf{1} + \kappa\mathbf{e}_k),$$

corresponding to the finite-state analogue of the sticky HDP-HMM prior of [47], where κ governs the strength of self-transition persistence. Under this prior, the expected self-transition probability is

$$\mathbb{E}[\pi_{kk}] = \frac{\alpha+\kappa}{K\alpha+\kappa}, \tag{2}$$

which, with $\alpha = 1$, inverts to $\kappa = (K\bar{p} - 1)/(1-\bar{p})$, where $\bar{p}$ denotes the target average self-transition probability.

Rather than fixing κ arbitrarily, it is calibrated from the transition matrices obtained from the EM estimates. Two anchors are employed: the operational baseline model ($K = 3$) and the BIC-optimal specification ($K = 7$). For each model, the average self-transition probability implied by the estimated transition matrix is substituted into (2) to obtain a calibrated value of κ. The resulting pair of persistence parameters is subsequently used in the Bayesian identifiability analysis.

Bayesian identifiability analysis.

A Bayesian sticky HMM is estimated for $K \in \{3,4,5,6,7\}$. Posterior inference is performed using the No-U-Turn Sampler (NUTS) with four parallel chains. The procedure is repeated under both calibrated values of κ to ensure that the identifiability conclusions do not depend on the EM specification used to anchor the prior. Three diagnostics are used: (1) the Gelman-Rubin statistic ($\hat{R}$), reported both before and after label-switching correction

https://doi.org/10.3390/math14142463

obtained by ordering posterior draws according to the emission means; (2) the minimum separation between posterior mean emission parameters; and (3) the number of divergent transitions. For each specification, the number of identified states is defined as the number of states whose label-switching-corrected emission-mean estimate satisfies $\hat{R} < 1.1$.

Sensitivity analysis.

To verify that the identifiability results are not artefacts of the calibrated persistence prior, the analysis is repeated across a grid of additional values of κ spanning weak, moderate, and strong assumptions regarding regime persistence. Regime configurations that remain identifiable across this range are regarded as robust, whereas specifications whose identification depends on a narrow prior region are interpreted more cautiously.

Selected specification.

The final specification is determined by jointly considering likelihood-based model fit, Bayesian identifiability diagnostics, posterior sensitivity to the persistence prior, and economic interpretability. Preference is given to the most parsimonious specification that is statistically identifiable, robust to prior assumptions, and capable of capturing the principal dynamics of the absorption-ratio series. Higher-dimensional specifications are retained for robustness analysis and comparison. All computations are implemented in Python using `hmmlearn`, `NumPyro`, and the No-U-Turn Sampler (NUTS).

2.4.5. Crisis Identification

In addition to regime classification, a crisis threshold is obtained analytically from the estimated emission densities. Let $f_M(\cdot)$ and $f_H(\cdot)$ denote the Gaussian emission densities associated with the Moderate-Concentration and High-Concentration regimes, respectively. The crisis threshold τ is defined as the absorption-ratio value at which the two regime densities intersect:

$$f_M(\tau) = f_H(\tau).$$

Under the Gaussian emission specification,

$$f_k(x) = \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right),$$

the equality condition becomes

$$\frac{1}{\sigma_M\sqrt{2\pi}} \exp\left(-\frac{(\tau-\mu_M)^2}{2\sigma_M^2}\right) = \frac{1}{\sigma_H\sqrt{2\pi}} \exp\left(-\frac{(\tau-\mu_H)^2}{2\sigma_H^2}\right).$$

Cancelling the common factor $\sqrt{2\pi}$ and taking logarithms yields

$$-\ln(\sigma_M) - \frac{(\tau-\mu_M)^2}{2\sigma_M^2} = -\ln(\sigma_H) - \frac{(\tau-\mu_H)^2}{2\sigma_H^2},$$

which can be rearranged as

$$\frac{(\tau-\mu_H)^2}{\sigma_H^2} - \frac{(\tau-\mu_M)^2}{\sigma_M^2} = 2\ln\left(\frac{\sigma_H}{\sigma_M}\right).$$

Expanding the quadratic terms gives

$$\left(\frac{1}{\sigma_H^2} - \frac{1}{\sigma_M^2}\right)\tau^2 + \left(-\frac{2\mu_H}{\sigma_H^2} + \frac{2\mu_M}{\sigma_M^2}\right)\tau + \left(\frac{\mu_H^2}{\sigma_H^2} - \frac{\mu_M^2}{\sigma_M^2} - 2\ln\left(\frac{\sigma_H}{\sigma_M}\right)\right) = 0.$$

Hence, τ is obtained as the solution of

$$A\tau^2 + B\tau + C = 0,$$

where

$$A = \frac{1}{\sigma_H^2} - \frac{1}{\sigma_M^2},$$

$$B = -\frac{2\mu_H}{\sigma_H^2} + \frac{2\mu_M}{\sigma_M^2},$$

$$C = \frac{\mu_H^2}{\sigma_H^2} - \frac{\mu_M^2}{\sigma_M^2} - 2\ln\left(\frac{\sigma_H}{\sigma_M}\right).$$

The threshold is therefore

$$\tau = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}.$$

Although two mathematical roots generally exist, only one lies between the estimated regime means μ_M and μ_H; this root is selected as the crisis threshold. When the regime variances are equal, $\sigma_M = \sigma_H$, the solution simplifies to the midpoint

$$\tau = \frac{\mu_M + \mu_H}{2}.$$

Crisis observations are then classified as

$$C_t = \mathbb{I}\{AR_t > \tau\},$$

providing a threshold-based measure of extreme network concentration that complements the Viterbi regime assignments. The threshold is additionally translated into an implied average-correlation benchmark through the absorption-ratio representation discussed in Step 1. To assess economic relevance, the inferred regimes and crisis periods are compared with global equity-market returns, ECB yield shocks, and benchmark return dynamics.

The three modules are used together because they answer different parts of the same measurement problem. PCA-AR answers when the pension network becomes more concentrated by reducing the covariance matrix to a scalar co-movement indicator. DTW clustering answers which funds move together by characterising the cross-sectional similarity structure. The HMM answers how the network-level concentration evolves over time by mapping AR_t into persistent latent regimes. Without PCA-AR there is no network-level stress input; without DTW clustering there is no cross-sectional explanation of which funds generate the structure; without the HMM the absorption ratio remains a continuous indicator without estimated regime persistence or transition probabilities. The next section applies this pipeline to the Lithuanian second-pillar pension-fund panel and reports what it reveals about the structure of systemic risk in the network.

3. Main Results

The empirical analysis proceeds in three stages following the pipeline described in Section 2 and illustrated in Figure 1. Section 3.1 presents the absorption ratio dynamics and robustness results. Section 3.2 reports the DTW clustering structure. Section 3.3 presents the HMM regime detection results, including the emission distributions, the data-driven crisis threshold, and its economic interpretation. Section 3.4 connects the clustering and regime detection findings through a cross-step analysis of the mean absolute active return amplification under systemic stress.

3.1. Absorption Ratio Dynamics

Table 2 reports the baseline series $\{AR_t\}$ from $R^{(f)}$ on a 60-day window: 1575 observations, mean 0.837, ranging from 0.728 to 0.939. The series sits far above the theoretical floor $1/N = 0.025$, which would require near-independence across the 40 funds. Even at its minimum, the leading factor explains more than 72% of return variance, so a dominant common factor is present throughout the sample, despite differences across providers and cohorts.

Table 2. Descriptive statistics of the baseline absorption ratio AR_t computed from $R^{(f)}$ with rolling window $\omega = 60$ trading days. Theoretical bounds follow from Proposition 3: $AR_t \in [1/N, 1] = [0.025, 1.000]$ for $N = 40$.

Statistic	Value
Observations	1575
Mean	0.8366
Standard deviation	0.0363
Minimum	0.7280
Maximum	0.9387

Figures 2 and 3 make the co-movement visible. Fund returns across all five providers move almost in step during the COVID-19 crash and the 2022 macroeconomic episode; the conservative AG1 and AG8/TIPF funds are less volatile but move in the same direction under stress. The absorption ratio peaks at 0.939 in early 2020 and stays elevated through the 2022 inflation and tightening period, confirming that stress coincides with tighter covariance concentration and weaker effective diversification.

Crucially, this co-movement is not simply imported. Although pension returns are contemporaneously correlated with global equities, AR_t measures the *internal* covariance structure, not the level of returns, and its statistical link to external factors is weak. Adding the euro-area 10-year and 2-year sovereign-yield shocks to the interpretive set does not change this: yield shocks move individual valuations, especially for conservative cohorts, but their association with AR_t stays small relative to the magnitude and persistence of the concentration regimes, as do those of MSCI World, MSCI Europe, and the S&P 500. The concentration regimes are therefore not well explained by the external factors considered, evidence consistent with, though not proof of, an internal driver. Lead–lag results are in Appendix A.

Robustness: Augmented Absorption Ratio

Recomputing AR_t on the augmented matrix $A \in \mathbb{R}^{T\times 85}$ (rank-deficient, so correlation-based with minimal ridge) gives a series ranging 0.428–0.876, mean 0.673. The lower level is mechanical: the denominator now spans 85 rather than 40 assets, so it is not evidence of weaker risk. The two series correlate at 0.539 overall (Figure 4); after standardisation, their crisis peaks (COVID-19, 2022, the 2023 banking stress) align almost exactly, while they diverge in calm periods, where benchmark and macro factors add variation. The crisis-versus-calm split makes this precise: correlation rises to 0.746 within crisis windows against 0.542 outside them, a gap a permutation test rejects as random ($p = 0.030$). Systemic episodes are thus detected consistently across both constructions, with benchmark information mattering mainly in calm conditions.

Finally, deliberately applying the augmented exercise's ridge term to the full-rank baseline distorts AR_t by 0.013–0.047 across windows (Table 3), which is why regularisation is confined to the rank-deficient augmented case.

https://doi.org/10.3390/math14142463

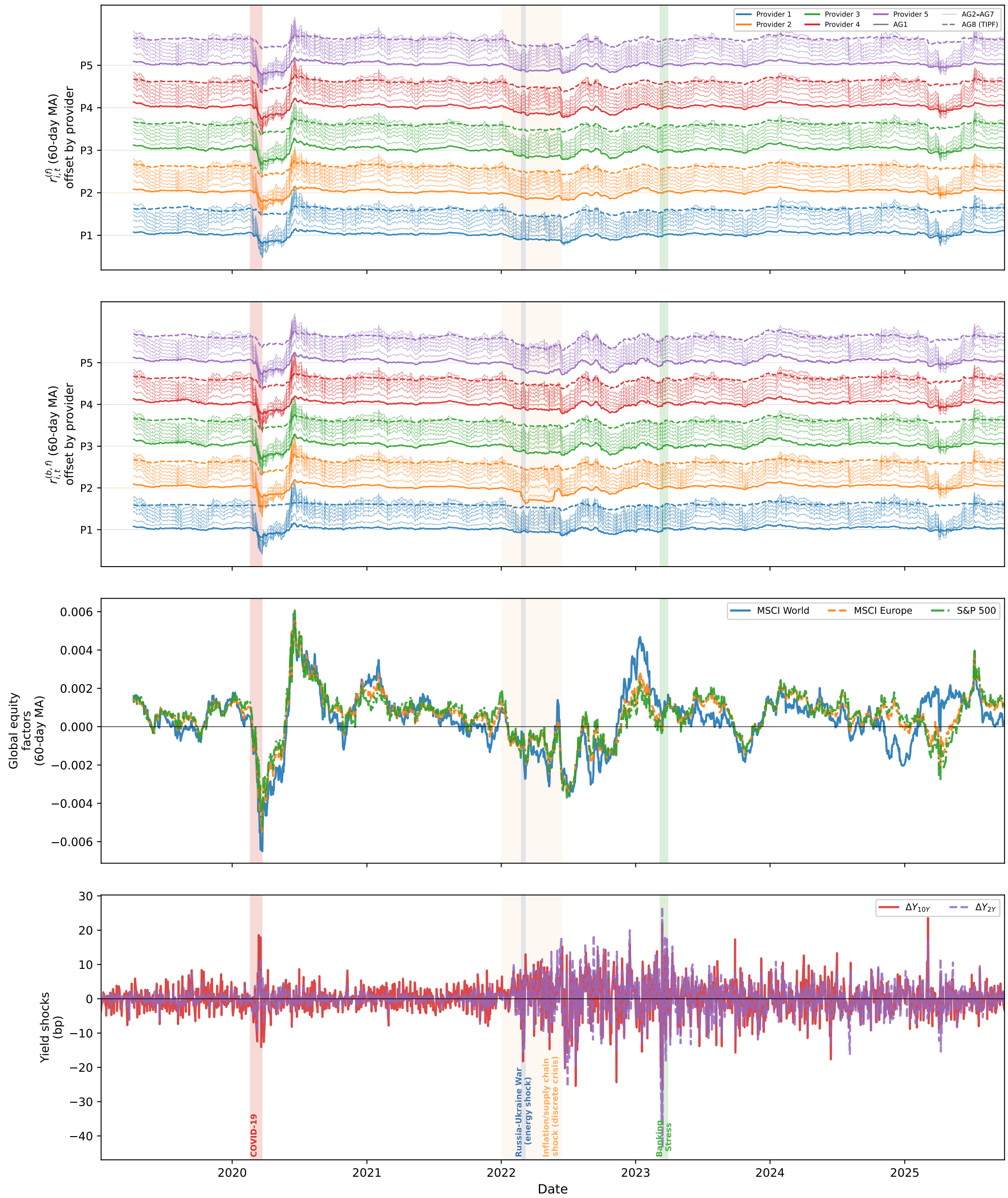

Figure 2. Sixty-day moving averages of pension-fund returns, fund-specific benchmark returns, global equity factors, and euro-area interest-rate shocks. Panel 1 displays all $N = 40$ pension-fund return series $R^{(f)}$, offset by the provider for clarity. Panel 2 shows the corresponding fund-specific benchmark returns $R^{(b,f)}$. Panel 3 presents the global equity factors used in the study (MSCI World, MSCI Europe, and S&P 500), while Panel 4 reports euro-area yield shocks measured as daily changes in 10-year and 2-year AAA sovereign yields. Shaded regions denote the COVID-19 market crash, the Russia–Ukraine war, the 2022 inflation and supply-chain shock, and the 2023 banking stress episode.

Table 3. Effect of ridge regularisation ($\varepsilon = 10^{-6}$) on the baseline absorption ratio across representative rolling windows.

Window	AR_t (Unregularised)	AR_t (Regularised)	Difference
Early window	0.8993	0.8573	0.0420
Mid window	0.8604	0.8479	0.0125
Late window	0.8285	0.7817	0.0468

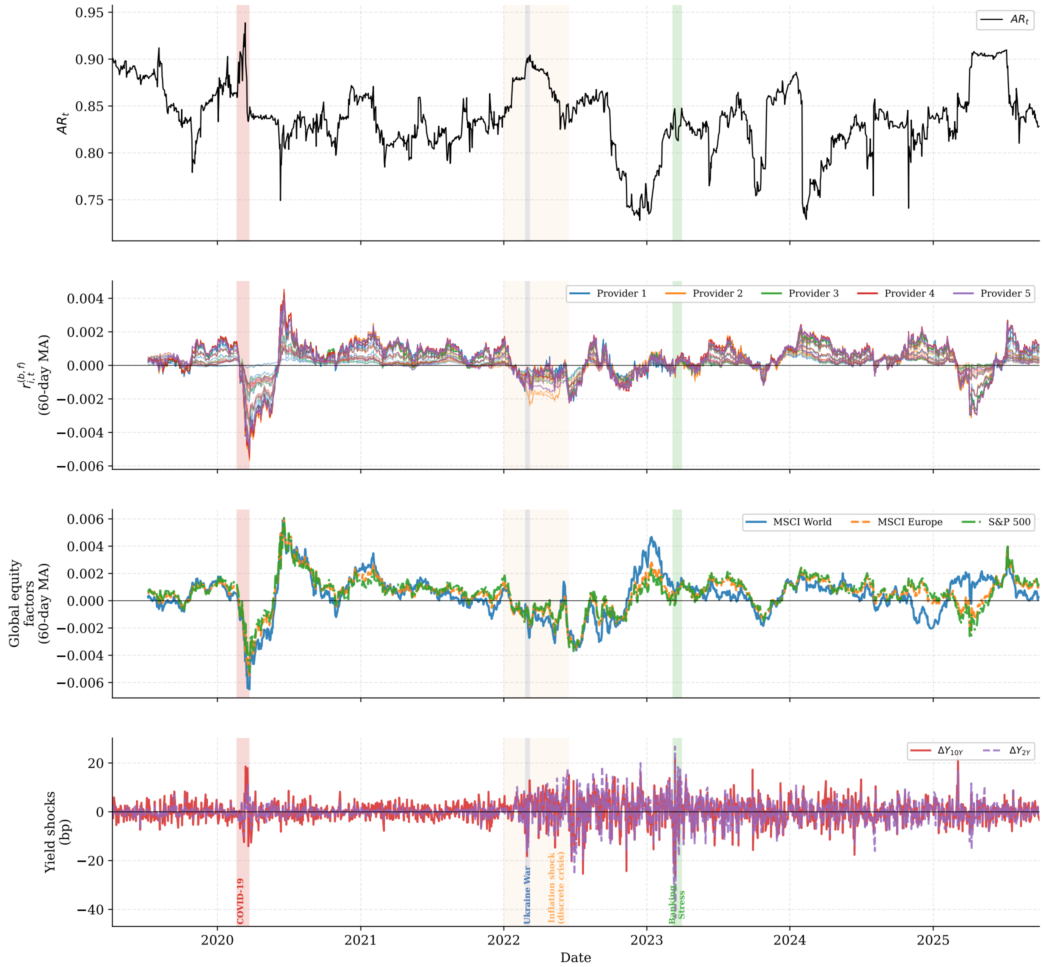

Figure 3. Panel 1: Baseline absorption ratio $\{AR_t\}$ computed from $R^{(f)}$ using a 60-day rolling covariance matrix. Panel 2: Sixty-day moving averages of all fund-specific benchmark returns $R^{(b,f)}$. Panel 3: Global equity factors (MSCI World, MSCI Europe, and S&P 500). Panel 4: Euro-area yield shocks based on daily changes in 10-year and 2-year AAA sovereign yields. Shaded regions denote major market stress episodes. The absorption ratio is constructed exclusively from pension-fund returns; benchmark and macro-financial series are included for interpretation only.

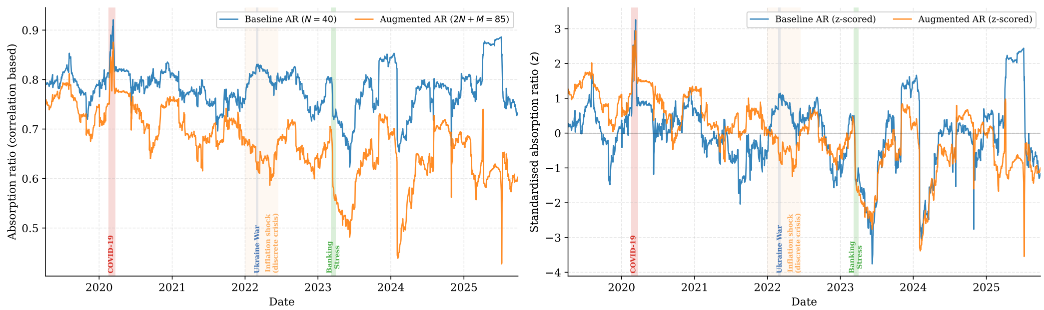

Figure 4. Comparison of baseline and augmented absorption ratios. The left panel shows the original series levels. The right panel reports standardised (z-score) versions of both series. The correlation between the standardised series is $\rho = 0.539$.

3.2. DTW Hierarchical Clustering

The analysis selects the cluster count, characterises the structure, confirms it is recovered from returns rather than labels, and tests robustness to algorithm and distance.

3.2.1. Number of Clusters

The silhouette coefficient on the DTW dissimilarity matrix (Table 4) is maximised at $K_c^* = 10$ and declines monotonically thereafter, so finer partitions over-segment. Figure 5 shows the corresponding cut; Table 5 summarises the result.

Table 4. Silhouette coefficient for $K_c \in \{2, \ldots, 15\}$, computed on the DTW dissimilarity matrix under average linkage. Higher is better; the maximum is in bold.

K_c	**Silhouette**
2	0.348
3	0.371
4	0.358
5	0.343
6	0.396
7	0.401
8	0.403
9	0.397
10	**0.438**
11	0.426
12	0.412
13	0.400
14	0.388
15	0.377

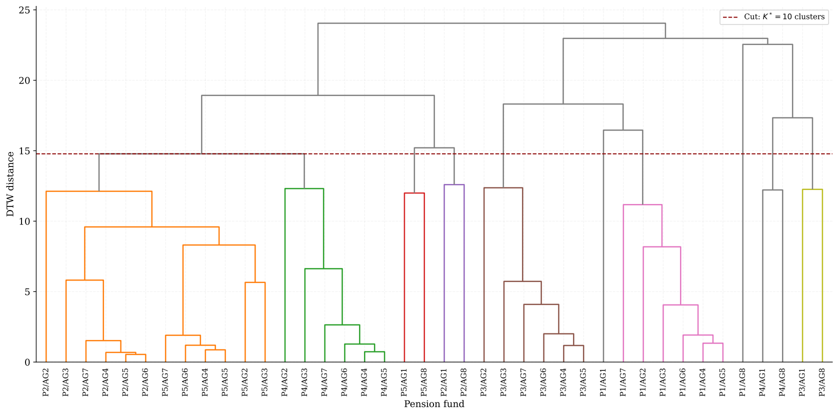

Figure 5. Hierarchical clustering dendrogram from pairwise DTW distances under average linkage. The dashed line marks the cut yielding $K_c^* = 10$ clusters, with branch colors indicating distinct cluster membership below the cut; singleton clusters and links above the cut are shown in gray/black. Fund labels follow the convention Pi/AGj for Provider i, age cohort j.

Table 5. Cluster composition and mean absolute active return at $K_c^* = 10$ (average linkage on DTW distances). Mean absolute active return (MAAR) is the mean absolute daily deviation of fund returns from the fund-specific benchmark $R^{(b,f)}$. Clusters are ordered by lifecycle segment.

Cluster	**Composition**	**Size**	**MAAR**
Growth segment (middle cohorts AG2–AG7)			
C1	Providers 2 & 5, AG2–AG7	12	0.0036
C6	Provider 1, AG2–AG7	6	0.0066
C5	Provider 3, AG2–AG7	6	0.0046
C2	Provider 4, AG2–AG7	6	0.0041
Conservative segment (extreme cohorts AG1, AG8)			
C4	Provider 2, AG1 & AG8	2	0.0008
C9	Provider 3, AG1 & AG8	2	0.0011
C8	Provider 4, AG1 & AG8	2	0.0012
C3	Provider 5, AG1 & AG8	2	0.0013
C7	Provider 1, AG1	1	0.0017
C10	Provider 1, AG8	1	0.0011

3.2.2. Cluster Structure

Two patterns emerge. The dominant one is the lifecycle split: conservative extreme-cohort funds (AG1; AG8/TIPF) separate sharply from growth middle cohorts (AG2–AG7) across all five providers. Conservative clusters carry MAAR of 0.0008–0.0017 against 0.0036–0.0066 for growth clusters; the tightest conservative cluster (0.0008) tracks roughly eight times more closely than the loosest growth cluster (0.0066), matching the low-turnover, capital-preservation mandates of funds serving participants at or near retirement.

Within each segment, funds group by provider, revealing which providers run the most similar strategies. Providers 2 and 5 are closest: their growth funds merge into a single 12-fund cluster (C1). Provider 1 is the most distinct: its growth funds form their own cluster (C6) with the highest MAAR in the sample (0.0066), and its two conservative funds do not group with any others, forming the singleton clusters C7 (AG1) and C10 (AG8). Because the analysis uses returns rather than holdings, these patterns indicate similarity in fund *behaviour*, not directly in asset allocation; the institutional reading is deferred to Section 4.

3.2.3. Recovery of Structure Beyond Fund Labels

The clustering uses only standardised returns; labels enter only afterward, for comparison. Table 6 compares the return-based partition with label-only references via the adjusted Rand index. No single label axis reproduces it: provider alone is moderate, lifecycle segment weak, and age cohort below chance. Only the joint provider $\times$ segment design is recovered, exactly under DTW k-means (ARI $= 1.00$) and substantially under hierarchical clustering (ARI $= 0.78$), the latter differing only by merging Providers 2 and 5's growth funds and splitting Provider 1's conservatives into singletons. The structure is genuine recovery, and the discrepancies flag funds whose behaviour departs from their nominal classification.

Table 6. Non-circularity check: adjusted Rand index (ARI) between the return-based clusters and partitions constructed purely from the nominal fund labels. The labels are not used in clustering. Only the joint provider $\times$ segment design is recovered; no single label axis predicts the partition, and age cohort alone scores below chance.

Label-Only Reference Partition	Hierarchical	DTW k-Means
Provider only (5 groups)	0.55	0.69
Lifecycle segment (2 groups)	0.20	0.13
Age cohort only (8 groups)	−0.07	−0.11
Provider $\times$ segment (10 groups)	0.78	**1.00**

3.2.4. Robustness: Algorithm and Distance

DTW k-means agrees with the hierarchical solution at ARI $= 0.78$ (Figure 6, Tables 7 and 8): both place every provider's conservative funds in one group and growth funds in another, with no fund crossing the conservative–growth boundary. The two differences are the Provider 2/5 growth merge and the treatment of Provider 1's conservatives.

Distance choice matters only as expected (Table 9). DTW and correlation-based clustering, both scale-free, are nearly identical (ARI $= 0.97$) and yield the same provider-by-lifecycle structure. Raw Euclidean k-means diverges (ARI $= 0.49$–0.52) because it is driven by volatility level, pooling all low-volatility conservative funds and fragmenting the growth funds. The close DTW-correlation agreement confirms the clusters reflect shared return *dynamics*, not volatility differences.

https://doi.org/10.3390/math14142463

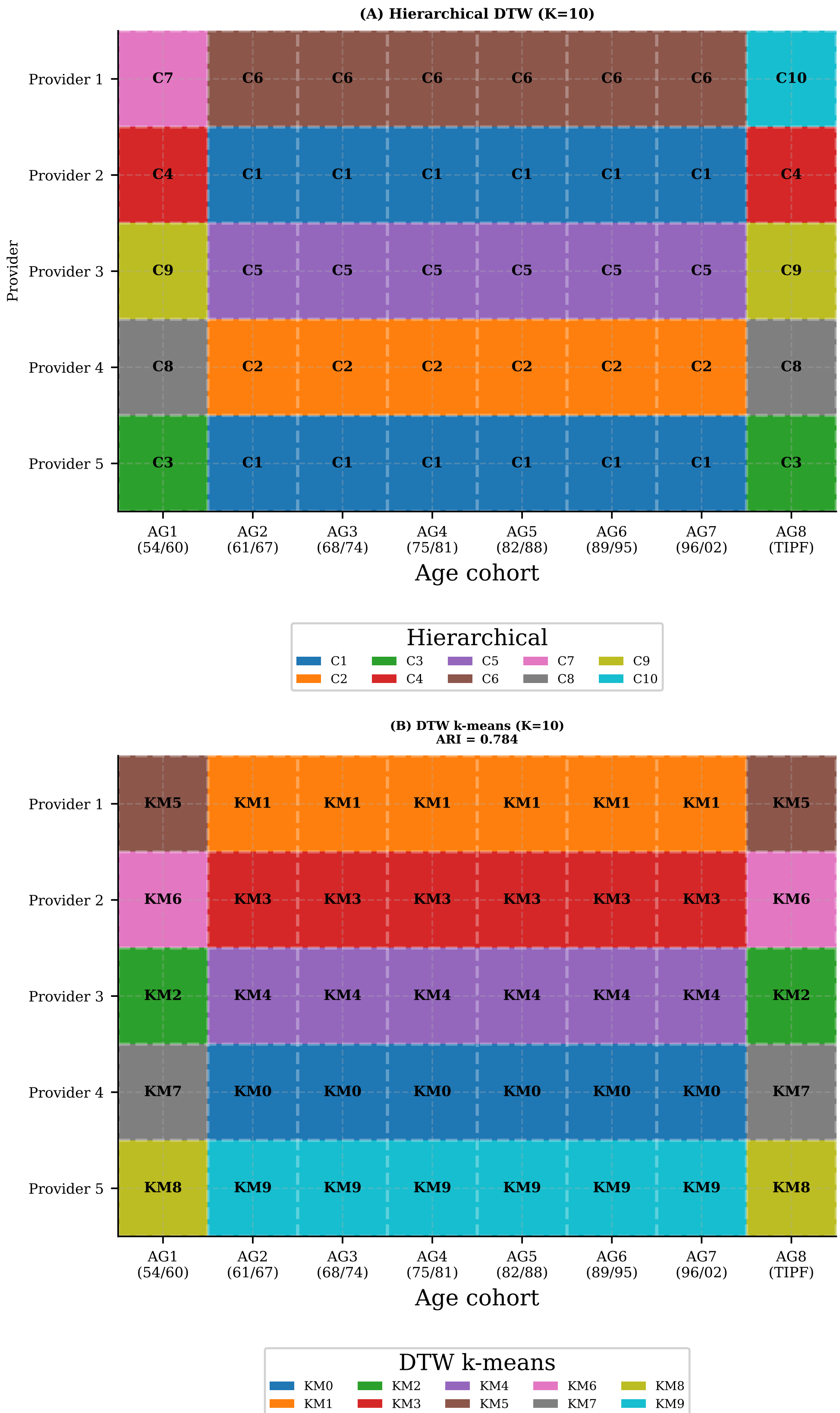

Figure 6. Comparison of hierarchical DTW clustering and DTW k-means clustering at $K_c = 10$. Cell labels denote cluster assignments. The adjusted Rand index between the two partitions is 0.784, indicating substantial agreement. Differences are limited to the separation of Providers 2 and 5 growth funds under DTW k-means and the treatment of Provider 1's conservative funds.

Table 7. Cluster composition under DTW-barycenter k-means at $K_c = 10$. Numbering is method-specific; the mapping to the average-linkage clusters of Table 5 is given in Table 8.

Cluster	Composition	Size	MAAR
KM0	Provider 4, AG2–AG7	6	0.0041
KM1	Provider 1, AG2–AG7	6	0.0066
KM2	Provider 3, AG1 & AG8	2	0.0011
KM3	Provider 2, AG2–AG7	6	0.0037
KM4	Provider 3, AG2–AG7	6	0.0046
KM5	Provider 1, AG1 & AG8	2	0.0014
KM6	Provider 2, AG1 & AG8	2	0.0008
KM7	Provider 4, AG1 & AG8	2	0.0012
KM8	Provider 5, AG1 & AG8	2	0.0013
KM9	Provider 5, AG2–AG7	6	0.0035

Table 8. Cross-tabulation of cluster assignments at $K_c = 10$. Rows are the average-linkage clusters of Table 5; columns are the DTW k-means clusters of Table 7. Empty cells denote zero funds. The only departures from a one-to-one mapping are C1, which splits across KM3 and KM9, and C7 and C10, which both fall in KM5.

Average Linkage	DTW *k*-Means									
	KM0	**KM1**	**KM2**	**KM3**	**KM4**	**KM5**	**KM6**	**KM7**	**KM8**	**KM9**
C1				6						6
C2	6									
C3									2	
C4							2			
C5					6					
C6		6								
C7						1				
C8								2		
C9			2							
C10						1				

Table 9. Distance-measure robustness at $K_c = 10$. Adjusted Rand index (ARI) between partitions. DTW and correlation are scale-free (standardised returns); Euclidean is computed on raw returns and retains volatility level.

Partition A	Partition B	ARI
DTW (average linkage)	DTW k-means	0.78
DTW k-means	Correlation k-means	0.97
DTW k-means	Euclidean k-means (raw)	0.49
Correlation k-means	Euclidean k-means (raw)	0.52

3.3. HMM Regime Detection

The third step of the pipeline (Section 2.4) maps the absorption-ratio dynamics of Section 3.1 onto a small number of latent risk regimes. The only observed input is the baseline series $\{AR_t\}$ computed from the covariance matrix of pension-fund returns $R^{(f)}$ (Table 2); its 1575 daily values are treated as draws from a Gaussian HMM whose hidden states index distinct levels of systemic concentration. Following the labelling convention of Section 2.4, the ordered states are termed the Low-, Moderate-, and High-Concentration regimes.

3.3.1. Model Selection

Selecting the number of latent regimes K in a Gaussian HMM requires balancing in-sample fit against the risk of over-parameterisation. We combine the standard BIC with a Bayesian posterior identifiability analysis under two data-driven κ calibrations, leading to a defensible choice of $K = 3$ as the baseline.

Information-criterion comparison.

Table 10 reports the maximised log-likelihood, AIC, and BIC for HMM models with $K \in \{2, \ldots, 8\}$ states. Each model is fitted from 200 random initialisations. The maximum-likelihood solution is retained for every K. Two of the larger specifications (K = 6 and K = 8) retain marginally lower log-likelihoods than their immediate predecessors, indicating residual sensitivity to local optima in the higher-dimensional models; because model selection here relies on BIC together with the posterior identifiability analysis rather than on the raw likelihood ordering, this does not affect the retained K = 3 specification. The penalty is computed using the corrected parameter count $d(K) = (K-1) + K(K-1) + 2K = K^2 + 2K - 1$, comprising $K-1$ free initial-state probabilities, $K(K-1)$ free transition probabilities, and $2K$ Gaussian emission parameters. The Bayesian (Schwarz) information criterion and the SIC coincide under this penalty, so a single BIC column is reported.

Table 10. Information criteria for HMM models with $K \in \{2, \ldots, 8\}$ states fitted to the absorption-ratio series $\{AR_t\}$. The number of free parameters is $d(K) = K^2 + 2K - 1$.

K	Log-Likelihood	$d(K)$	AIC	BIC
2	3573.89	7	−7133.78	−7096.25
3 †	4194.61	14	−8361.22	−8286.15
4	4559.99	23	−9073.98	−8950.66
5	4841.03	34	−9614.07	−9431.76
6	4838.96	47	−9583.93	−9331.91
7 *	4944.66	62	−9765.32	−9432.88
8	4944.57	79	−9731.14	−9307.55

* BIC- and AIC-optimal. † baseline.

Both BIC and AIC are minimised at $K = 7$, but the criterion does not sharply resolve the dimension: the BIC of $K = 7$ is lower than that of $K = 5$ by only about one unit, a difference well within the range attributable to estimation noise. This non-separation is symptomatic of a known limitation: BIC penalises the number of free parameters but does not assess whether the additional components correspond to interpretable or well-separated states [61,62], and it can favour extra components when emission distributions overlap. We therefore complement the information criteria with a posterior identifiability analysis that asks not which K maximises penalised fit, but which states are statistically distinguishable across independent Markov chains.

Dual empirical calibration of κ.

The identifiability analysis relies on a Bayesian finite sticky HMM with the Dirichlet prior specified in Section 2.4. Rather than fixing κ arbitrarily, it is calibrated empirically by inverting the prior expected self-transition probability (2) against the EM-fitted transition matrix. We interpret this as an empirical-Bayes procedure: anchoring the persistence prior to EM estimates reduces arbitrariness in prior selection, but it does not render the Bayesian fit an independent validation of the EM model, and we no longer describe it as such. Because the calibration depends on K, we use two anchors that span a wide persistence range: the operational baseline ($K = 3$), whose three states have mean self-transition $\bar{p}_3 = 0.971$ (diagonals $0.948, 0.980, 0.984$) and yield $\kappa_3 \approx 66$; and the information-criterion-optimal specification ($K = 7$), whose mean self-transition is lower at $\bar{p}_7 = 0.809$ (one transient state with near-zero persistence pulls the mean down) and yields $\kappa_7 \approx 24$. The two anchors therefore bracket a strong-persistence prior (κ_3) and a substantially weaker one (κ_7), ensuring that conclusions do not hinge on the EM anchor.

Identifiability under both calibrations.

We fit the finite sticky HMM via NUTS at each calibrated κ for $K \in \{3, 4, 5, 6, 7\}$, using two parallel chains with 1000 warmup and 1000 post-warmup iterations per chain.

The Gelman–Rubin $\hat{R}$ statistic for emission means is reported both raw and after label-switching correction (sorting emission means within each posterior draw); a state is identified when its corrected $\hat{R} < 1.1$. Table 11 reports the results.

Table 11. Posterior identifiability of the finite sticky HMM under the two data-calibrated stickiness priors $\kappa_3 = 66$ and $\kappa_7 = 24$. Raw and label-switching-corrected Gelman–Rubin statistics ($\hat{R}$) are reported alongside the number of identified states (# Id, corrected $\hat{R} < 1.1$) and the minimum separation between sorted emission means. Zero divergent transitions across all (K, κ) confirm that identifiability failures reflect over-parameterisation rather than sampler pathology.

K	Raw Min/Max $\hat{R}$	Corr Min/Max $\hat{R}$	# Id	Min μ Gap	Div.
Panel A: $\kappa_3 = 66$ (baseline anchor)					
3	23.82/37.34	0.999/0.999	3	0.0449	0
4	1.15/30.59	1.222/16.059	0	0.0245	0
5	1.13/45.12	1.195/26.399	0	0.0226	0
6	1.12/29.89	1.016/27.525	1	0.0156	0
7	8.88/66.27	0.999/25.207	2	0.0156	0
Panel B: $\kappa_7 = 24$ (IC-optimal anchor)					
3	1.00/1.00	0.999/1.001	3	0.0449	0
4	1.15/31.60	1.219/16.425	0	0.0245	0
5	1.00/42.97	0.999/1.000	5	0.0188	0
6	1.32/57.44	11.859/43.475	0	0.0216	0
7	2.90/40.93	1.014/37.506	1	0.0140	0

Both calibrations identify $K = 3$ cleanly (corrected $\hat{R} = 0.999$, minimum emission-mean gap 0.044 absorption-ratio units), and both reject $K = 4$ entirely (corrected $\hat{R}$ above 16). The two priors diverge on $K = 5$: it is fully identified under the weaker $\kappa_7 = 24$ prior (5 of 5 states, $\hat{R} \approx 1.000$) but collapses under the stronger $\kappa_3 = 66$ prior (0 of 5, $\hat{R}$ up to 26.4). $K = 6$ and $K = 7$ are at most partially identified under either prior. This prior-sensitivity of $K = 5$, together with its small emission-mean separations (≤ 0.019), distinguishes it sharply from the robustly identified three-state structure; it is examined as a refinement in Section 3.3.6.

Sensitivity to the stickiness prior.

To verify that these conclusions extend beyond the two anchors, the analysis is repeated across $\kappa \in \{10, 15, 24, 66, 80\}$. Table 12 reports the number of identified states at each (K, κ).

The grid makes the contrast explicit. **$K = 3$ is the only specification identified at every κ value tested**, from the weakest ($\kappa = 10$) to the strongest ($\kappa = 80$). $K = 4$ fails at every κ except a single degenerate point at $\kappa = 15$, where the chains collapse rather than separate genuine regimes. $K = 5$ is identified at four of the five values but fails precisely at $\kappa = 66$, the prior calibrated to the empirical persistence of the three-state model, so its identification is contingent on the prior rather than robust to it. $K \geq 6$ are erratic. The uniform identifiability of $K = 3$ across the entire prior range is therefore a property of the data, not of any single prior choice.

Selected specification.

We retain $K = 3$ as the baseline. It is the smallest specification and the only one cleanly identified under both data-calibrated priors and across the full empirically supported κ range; it follows the Markov regime-switching framework of [39], with the low-, moderate-, and High-Concentration labels assigned here by ordering the three estimated emission means rather than imported from any prior taxonomy; and the EM and Bayesian estimates agree to within 0.001 across all emission parameters (Section 3.3.2). The prior-sensitive $K = 5$ specification is examined as a finer refinement in Section 3.3.6.

Table 12. Sensitivity of posterior identifiability to the stickiness prior κ. Cells report the number of states with corrected $\hat{R} < 1.1$. The data-calibrated anchors $\kappa_3 = 66$ and $\kappa_7 = 24$ (highlighted) are used in the main identifiability analysis.

K	$\kappa = 10$	$\kappa = 15$	$\kappa_7 = 24$	$\kappa = 66^{\kappa_3}$	$\kappa = 80$
3	**3**	**3**	**3**	**3**	**3**
4	0	4 ‡	0	0	0
5	5	5	5	0	5
6	1	6	0	1	6
7	1	1	1	2	3

‡ Degenerate identification: chains collapse to a near-uniform mean. configuration (single isolated κ value).

3.3.2. Regime Characterisation

Table 13 reports the emission parameters, state occupancy, and persistence statistics for the three regimes under the $K = 3$ baseline, together with mean daily log returns of global equity benchmarks over the regime-assigned days.

Table 13. HMM regime characterisation ($K = 3$). Emission parameters $(\hat{\mu}_k, \hat{\sigma}_k)$, regime persistence (transition-matrix diagonal), and expected duration $1/(1 - p_{kk})$ in trading days. Global-benchmark statistics are mean daily log returns over regime-assigned days. Persistence and expected duration describe the rolling absorption-ratio signal and are reported descriptively; they are not durations of independent daily economic states (Section 2.4.1).

Regime	Days	Share	$\hat{\mu}_k$	$\hat{\sigma}_k$	Persistence	Duration	S&P 500/MSCI World
Low-Concentration	209	13.3%	0.7731	0.0237	0.9479	19.2	$+0.0004/+0.0009$
Moderate-Concentration	853	54.2%	0.8294	0.0127	0.9804	51.1	$+0.0011/+0.0008$
High-Concentration	513	32.6%	0.8743	0.0194	0.9843	63.7	$-0.0003/-0.0006$

The three regimes are cleanly ordered by their absorption-ratio level. The Low-Concentration regime ($\hat{\mu} = 0.773$) is the least persistent (0.948, expected duration ≈ 19 days) and the rarest, occupying only 13.3% of trading days. The Moderate-Concentration regime is the modal state (54.2% of days), the narrowest in emission spread ($\hat{\sigma} = 0.013$), and corresponds to the network's business-as-usual level of co-movement. The High-Concentration regime ($\hat{\mu} = 0.874$) is the most persistent (0.984, expected duration ≈ 64 days) and occupies just under a third of the sample. Its negative average equity returns (-0.0003 S&P 500, -0.0006 MSCI World) confirm that High-Concentration days, on average, coincide with weak rather than strong global markets, but, as Section 3.4 shows, this association is far too weak to make the regime a restatement of equity conditions.

To verify that the regime estimates are not artefacts of the EM optimisation, the emission means are re-estimated under the Bayesian sticky HMM at the data-calibrated prior $\kappa_3 = 66$. Table 14 compares the two estimators after label-switching correction.

Table 14. EM vs Bayesian estimates of emission means at $K = 3$. The Bayesian posterior is computed via NUTS at $\kappa_3 = 66$ after label-switching correction; 90% credible intervals are reported.

Regime	$\hat{\mu}_k^{\text{EM}}$	$\hat{\mu}_k^{\text{Bayes}}$	$\lvert\text{Diff}\rvert$	Bayes 5%	Bayes 95%
Low-Concentration	0.7731	0.7734	0.0003	0.7704	0.7766
Moderate-Concentration	0.8294	0.8294	0.0000	0.8285	0.8302
High-Concentration	0.8743	0.8742	0.0001	0.8727	0.8758

The EM and Bayesian point estimates agree to within 0.001 across all three regimes, with narrow 90% credible intervals (widths 0.002–0.006). This convergence between a frequentist and a fully Bayesian estimator confirms that the data sharply identify the

https://doi.org/10.3390/math14142463

three-regime structure rather than an artefact of either estimation route. Because the calibration is empirical-Bayes (Section 3.3.1), this agreement should be read as confirming that the data dominate the calibrated prior at $K = 3$, rather than as a fully independent cross-validation.

The Gaussian emission assumption is adopted as a parsimonious device for regime classification rather than as a claim that the unconditional absorption-ratio distribution is exactly Gaussian. The excess kurtosis documented in Table 1 is consistent with regime switching itself: a mixture of approximately Gaussian within-regime densities generates heavy unconditional tails, and the visible mass in the 0.80–0.85 band of Figure 7 is the Moderate-High overlap rather than evidence against the within-regime specification. Replacing the Gaussian emissions with Student-t or other heavy-tailed densities is a natural extension for future work.

Figure 7 plots the three estimated Gaussian emission distributions relative to the crisis threshold τ derived in Section 3.3.3.

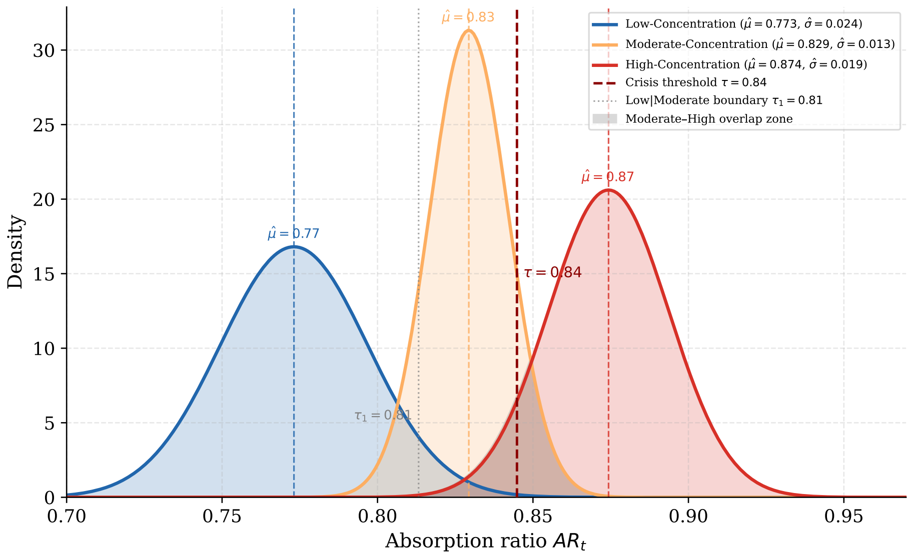

Figure 7. Gaussian emission densities estimated by the BW algorithm for the three-state HMM ($K = 3$). Each density $\mathcal{N}(\hat{\mu}_k, \hat{\sigma}_k^2)$ corresponds to one latent regime; dashed vertical lines mark the emission means. The Moderate-Concentration regime is the narrowest ($\hat{\sigma} = 0.013$), reflecting the tightly clustered business-as-usual level of the absorption ratio. The dark dashed line marks the crisis threshold $\tau = 0.845$, the absorption-ratio level at which the Moderate- and High-Concentration densities have equal density; the dotted line marks the Low | Moderate boundary $\tau_1 = 0.813$. The shaded region is the Moderate–High overlap zone, where regime assignment is governed by the transition matrix rather than emission density alone.

A regime-wise reading of the decoded sequence along the sample timeline reinforces this characterisation. The pre-COVID window of 2019 is already classified predominantly as Moderate- and High-Concentration rather than Low: the network enters the sample in an elevated state, consistent with the persistently high absorption ratio documented in Section 3.1. The COVID-19 crash of February–March 2020 is the single most severe episode, classified as High-Concentration and containing the sample maximum $AR_t = 0.939$. The network then relaxes intermittently through 2020–2021 before re-entering High-Concentration during the 2022–2023 inflation and monetary-tightening cycle. The rare Low-Concentration spells cluster in the comparatively calm windows of 2023–2024, when the ratio dips toward 0.73, and a renewed High-Concentration episode appears in 2025. The decomposition therefore shows that elevated concentration is not confined to the COVID shock but recurs across distinct macro-financial episodes, while genuinely low concentration is the exception.

Figure 8 overlays the decoded regime classification onto the absorption ratio and the benchmark and macro-financial series across the full sample. The decoupling between the recovery of global benchmarks and the persistence of High-Concentration shading, particularly visible in late 2022, indicates that pension-network stress is not a mirror of equity-market dynamics but a structural synchronisation that persists beyond the initial shock. Crucially, the HMM is fitted to the absorption ratio alone; the macro-financial panels enter for interpretation only, and the visual correspondence with known crises therefore validates the regime assignment against external events without those events ever entering the estimation.

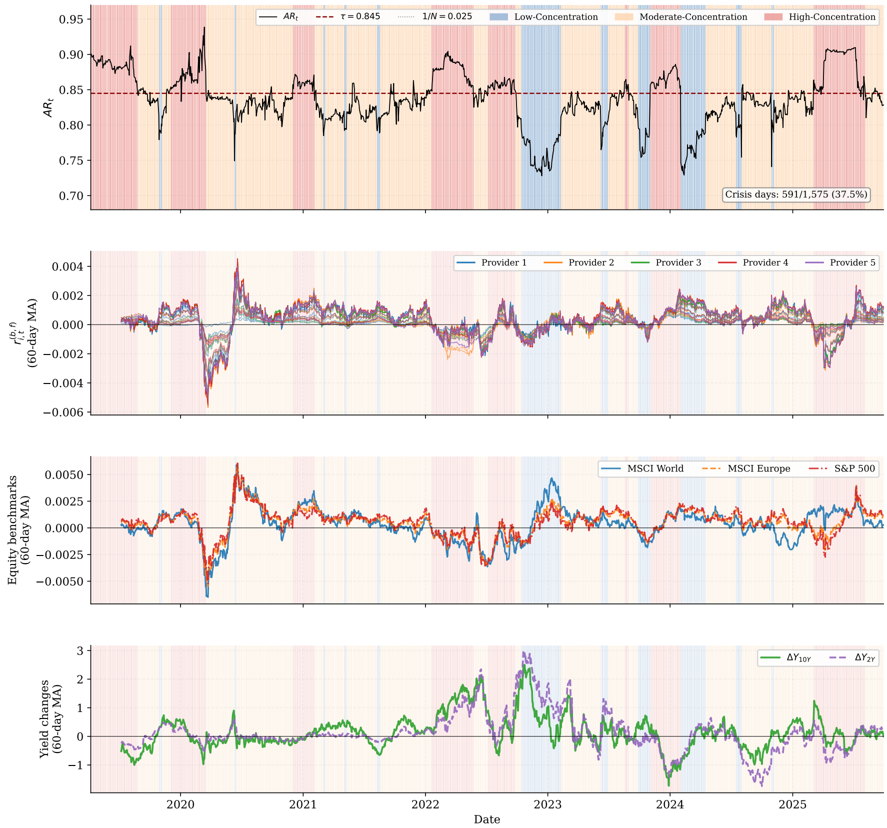

Figure 8. Decoded regime classification over time. **Top panel**: Absorption ratio $\{AR_t\}$ with HMM regime classification (low: blue, moderate: amber, high: red), empirical crisis threshold $\tau = 0.845$ (dashed red line), and theoretical lower bound $1/N = 0.025$ (dotted line). The annotation reports the threshold-rule crisis count $C_t = \mathbb{I}\{AR_t > \tau\}$ (591 of 1575 trading days, 37.5%), which exceeds the Viterbi-decoded High-Concentration count (513 days, 32.6%) for the reasons in Section 3.3.3. **Middle panel**: Sixty-day moving average of all $N = 40$ fund-specific benchmark returns $r_{i,t}^{(b,f)}$, by provider. **Bottom panel**: Global equity benchmarks and euro-area yield changes. Regime shading is consistent across panels.

3.3.3. Crisis Threshold

The empirical crisis threshold τ is defined as the absorption-ratio level at which the Moderate- and High-Concentration emission densities are equal, $f_M(\tau) = f_H(\tau)$, obtained as the root of the Gaussian-crossing quadratic of Section 2.4 that lies between the two regime means:

$$\tau = 0.845, \qquad \tau_1 = 0.813,$$

for the Moderate | High and Low | Moderate boundaries respectively. The threshold is an equal-*density* decision boundary and does not incorporate regime-prevalence (prior) weights. Because the two regimes have unequal variances ($\hat{\sigma}_{\text{Mod}} = 0.013$, $\hat{\sigma}_{\text{High}} = 0.019$), the exact crossing is obtained analytically from the quadratic rather than by averaging the means; it differs from the equal-variance midpoint approximation $(\hat{\mu}_{\text{Mod}} + \hat{\mu}_{\text{High}})/2 = 0.852$, and we report the exact crossing $\tau = 0.845$ throughout. By Theorem 1 and Proposition 3, $\tau = 0.845$ corresponds to a critical average inter-fund correlation

$$\bar{\rho}_{\text{crit}} = \frac{N\tau - 1}{N - 1} = \frac{40 \times 0.845 - 1}{39} \approx 0.841,$$

with $\tau_1 = 0.813$ mapping to $\bar{\rho}_{\text{crit},1} \approx 0.808$.

The two boundaries partition the absorption-ratio axis into three concentration zones used for threshold-based classification:

$$\underbrace{AR_t \leq \tau_1,}_{\text{Low}} \qquad \underbrace{\tau_1 < AR_t \leq \tau,}_{\text{Moderate}} \qquad \underbrace{AR_t > \tau,}_{\text{High}}$$

Values $\tau_1 = 0.813$ define the Low-Concentration zone; the interval $\tau_1 < AR_t \leq \tau$, i.e., $0.813 < AR_t \leq 0.845$, defines the Moderate-Concentration zone; and values $AR_t > \tau = 0.845$ define the High-Concentration zone. These thresholds give the absorption-ratio levels at which a supervisor would reclassify the network between zones.

The threshold rule and the Viterbi decoder give two related but distinct classifications. The daily crisis indicator $C_t = \mathbb{I}\{AR_t > \tau\}$ flags 591 days (37.5% of the sample), whereas the Viterbi-decoded High-Concentration regime spans 513 days (32.6%). The 78-day difference reflects the Moderate–High overlap zone of Figure 7: because $\tau = 0.845$ lies below the High-Concentration emission mean, the threshold rule flags overlap-zone days that the Viterbi decoder, weighting the persistence encoded in the transition matrix, still assigns to the Moderate-Concentration state. The downstream analysis in Section 3.4 uses the Viterbi-decoded regime, while τ provides the economic interpretation of where the High-Concentration boundary lies in observation space.

The empirical value $\tau = 0.845$ substantially exceeds the heuristic thresholds typically assumed in network-based contagion studies [49], and the gap is informative rather than incidental. It is exactly what [5] predicts: contagion ignites only beyond a system-specific tipping point determined by the commonality of holdings, not at any universal shock level. The same logic appears from the opposite direction in [16], whose clearing analysis shows that the contagion boundary is a property of a particular network's obligation structure and cannot be borrowed across sectors. A threshold of 0.845 would be implausibly high in a banking network with dispersed exposures; in a lifecycle-regulated pension system whose calm-state correlation approaches 0.77, it is precisely where the additional concentration becomes systemically meaningful.

3.3.4. Regime Persistence and Transition Structure

Beyond regime characterisation, the full transition matrix reveals an asymmetric escalation pathway. Table 15 reports the estimated matrix.

Table 15. Estimated HMM transition probability matrix ($K = 3$). Diagonal entries (persistence) also appear in Table 13; off-diagonal entries give one-day transition probabilities between regimes.

From\To	Low-Conc.	Moderate-Conc.	High-Conc.
Low-Concentration	0.9479	0.0475	0.0046
Moderate-Concentration	0.0122	0.9804	0.0074
High-Concentration	0.0018	0.0139	0.9843

Direct transitions from the Low- to the High-Concentration regime are very rare (probability 0.0046): the system almost always passes through the Moderate-Concentration regime when escalating, and likewise de-escalates through it ($P(\text{High} \to \text{Low}) = 0.0018$). This asymmetric pathway gives the Moderate zone a precursor interpretation: a sustained rise of AR_t above τ_1 typically precedes a full High-Concentration episode. We frame this as a structural-precursor property of the *retrospective* regime decomposition rather than as a validated real-time alarm; the requirements for an operational early-warning implementation are discussed in Section 3.3.5. The same strong serial dependence justifies the HAC (Newey–West) adjustment applied to the regime-conditional regressions of Section 3.4.

3.3.5. Current State, Forward Probabilities, and Real-Time Use

As of the final observation in the sample (29 September 2025), the network is in the **Moderate-Concentration** regime, with $AR_t = 0.829$, closely aligned with the Moderate emission mean $\hat{\mu} = 0.829$. The system has remained in this regime for 38 consecutive trading days, against an expected total duration of $1/(1 - 0.980) \approx 51$ trading days. Forward state probabilities $\mathbb{P}(S_{t+h} = k \mid S_t = \text{Moderate})$ are obtained by iterating the transition matrix of Table 15. Table 16 reports them alongside the long-run stationary distribution.

Table 16. Forward regime probabilities conditional on the current Moderate-Concentration state (29 September 2025) and the long-run stationary distribution.

Horizon	**Low-Conc.**	**Moderate-Conc.**	**High-Conc.**
$t + 1$ (next day)	0.0122	0.9804	0.0074
$t + 5$ (1 week)	0.0529	0.9122	0.0349
$t + 20$ (1 month)	0.1308	0.7537	0.1155
Stationary ($h \to \infty$)	0.1409	0.5565	0.3027

The short-run dynamics are dominated by persistence: the network has a 98% probability of remaining in the Moderate-Concentration regime the next day and still 75% over the following month. At the one-month horizon, the system has only begun to relax, with the Low-Concentration probability (13.1%) marginally above the High-Concentration probability (11.6%). Over the long run, however, the stationary distribution places more than twice as much mass on the High- as on the Low-Concentration regime (30.3% versus 14.1%): because the High-Concentration regime is the most persistent of the three (self-transition 0.984) while the Low-Concentration regime decays fastest (self-transition 0.948, exiting predominantly back to the Moderate state, Table 15), the network gravitates toward high rather than low concentration over long horizons. The stationary distribution—30.3% High, 55.6% Moderate, 14.1% Low—implies that the network spends approximately 86% of its time in the Moderate or High regimes, reinforcing the characterisation of elevated systemic co-movement as the structural norm rather than an exceptional state.

A caveat on real-time use is warranted. The Viterbi state sequence is a *smoothed* ex-post classification: the label at day t conditions on the entire sample, including days after t, and therefore differs from a filtered classification that uses only information available up to t, especially near transitions. The decoded regimes of Sections 3.3 and 3.4 should accordingly be read as retrospective decompositions of the concentration process. The forward probabilities of Table 16 are the appropriate real-time object, since they condition only on the currently inferred state, and they make quantitative, falsifiable predictions for the months following September 2025. We therefore present the framework primarily as a monitoring and measurement tool; a fully validated early-warning system would additionally require the filtered (online) regime probabilities and out-of-sample evaluation beyond the single labelled COVID-19 episode.

3.3.6. Robustness: Higher-State Specifications and Alternative Constructions

The model-selection analysis retains $K = 3$ as the smallest cleanly identified specification. This subsection asks what richer models, alternative absorption-ratio constructions, and alternative rolling windows add, and whether the three-regime reading survives them.

Higher-state specifications.

The BIC runner-up $K = 5$ is identified under the weaker calibration $\kappa_7 = 24$ (and at $\kappa = 10, 15, 80$) but fails precisely at $\kappa_3 = 66$, the prior calibrated to the persistence of the three-state model (Table 12). Where it is identified, the five ordered means do not introduce qualitatively new states: they partition the broad Moderate-Concentration region into finer gradations of stress intensity, while the extreme states still correspond to the baseline Low- and High-Concentration regimes. Repeating the analyses of Section 3.4 under $K = 5$, after aggregating the intermediate states into a single Moderate regime, preserves the substantive conclusions. The five-state model therefore refines rather than overturns the three-regime interpretation.

The $K = 4$ specification, by contrast, fails identification across the full empirically supported κ range. We read this as a statistical identification phenomenon rather than a structural feature of the network: splitting the broad central Moderate component into two heavily overlapping states (emission-mean gaps of approximately 0.025) induces label-switching that mean-sorting cannot repair. We therefore draw no economic conclusion from it, noting only that the data support a robust three-state partition and a finer, prior-sensitive five-state refinement, with no stable four-state configuration between them.

Correlation-based absorption ratio.

Re-estimating the HMM on the correlation-based absorption ratio, the scale-free construction for which the AR–$\bar{\rho}$ mapping of Proposition 3 holds exactly, preserves the ordered, persistent Low-, Moderate-, and High-Concentration regimes. As with the covariance series, BIC favours a richer specification ($K = 6$, against $K = 7$ for the covariance AR), but the extra states again only subdivide the Moderate region; the re-fitted three-state baseline retains the Low–Moderate–High ordering. Standardisation removes the volatility concentration embedded in the covariance estimator, so the regime means shift downward mechanically and the crisis threshold falls from 0.845 to 0.802, while the qualitative ordering is unchanged.

Rolling-window length.

Because AR_t is computed from overlapping windows, the inferred regimes describe persistent patterns in concentration rather than independent daily states. Repeating the analysis with 40-, 60-, 90-, and 120-day windows (Table 17) shows that the absorption-ratio distribution is stable in level: the mean ranges only from 0.8354 to 0.8368. As expected, longer windows smooth the series, so the standard deviation falls monotonically from 0.0444 (40-day) to 0.0284 (120-day), producing longer but less volatile concentration episodes.

Information criteria favour somewhat richer specifications at the alternative windows—BIC selects $K = 6$ (40-day), $K = 7$ (60-day), and $K = 5$ (90- and 120-day)—but the additional states only subdivide the Moderate-Concentration region while preserving the same ordered Low–Moderate–High structure. The economic interpretation is therefore unchanged even when the statistically preferred K exceeds three.

The crisis threshold is especially stable, varying only between $\tau = 0.8449$ and $\tau = 0.8587$—a range of 1.4 percentage points—across windows that differ substantially in smoothing: 0.8550 (40-day), 0.8449 (60-day), 0.8587 (90-day), and 0.8475 (120-day). The High-Concentration share varies more (31.4%, 37.5%, 19.0%, 29.2% respectively), as expected: longer windows smooth short-lived fluctuations, so fewer observations cross the boundary but identified episodes last longer.

Exact daily classifications agree less than the thresholds: relative to the 60-day baseline, agreement is 70.2% (40-day), 52.3% (90-day), and 59.1% (120-day), with adjusted Rand indices of 0.315, 0.186, and 0.162. This is unsurprising, since each specification smooths the same underlying process differently. The framework's robustness is therefore better judged by the stability of the concentration ordering and the crisis threshold than by exact day-to-day agreement: across estimators, window lengths, and higher-state specifications, low concentration is consistently separated from elevated concentration by a broad intermediate regime, and the systemic-concentration threshold remains close to 0.845.

Table 17. Rolling-window robustness of the absorption-ratio HMM. Longer windows mechanically smooth the series, but the mean level, regime ordering, and crisis threshold remain stable.

Window	Mean AR	Std. AR	BIC-Optimal K	Threshold τ	Crisis-Day Share
40-day	0.8354	0.0444	6	0.8550	31.4%
60-day	0.8366	0.0363	7	0.8449	37.5%
90-day	0.8367	0.0315	5	0.8587	19.0%
120-day	0.8368	0.0284	5	0.8475	29.2%

Overlap-induced versus economic persistence.

Because AR_t is built from overlapping 60-day windows, part of its serial dependence is mechanical: two windows fewer than ω days apart share raw observations, which alone would produce an autocorrelation decaying linearly to zero by lag ω.

Figure 9 shows that the empirical autocorrelation of $\{AR_t\}$ lies above this triangular overlap-only benchmark within the window, the gap reflecting genuine persistence beyond the mechanical component; beyond lag ω, where pure overlap predicts exactly zero, the empirical mean is only 0.008, so almost no dependence survives once the windows are disjoint. To confirm that the regimes are not an overlap artefact, the HMM is re-estimated on the 60 disjoint subsamples obtained by taking every ω-th observation. The ordered three-regime structure (Low $<$ Moderate $<$ High) and the separation between regimes are recovered in every subsample. The regime means shift modestly upward in the thinned samples—most for the rare Low-Concentration regime (0.773 $\to$ 0.818), and by about 0.015 for the Moderate mean and the crisis threshold ($\tau = 0.859$, s.d. 0.020, against 0.845 for the overlapping series)—while the High-Concentration mean is essentially unchanged. The upward shift arises because disjoint sampling unevenly retains the short-lived Low-Concentration episodes; it leaves the regime ordering, separation, and economic interpretation intact.

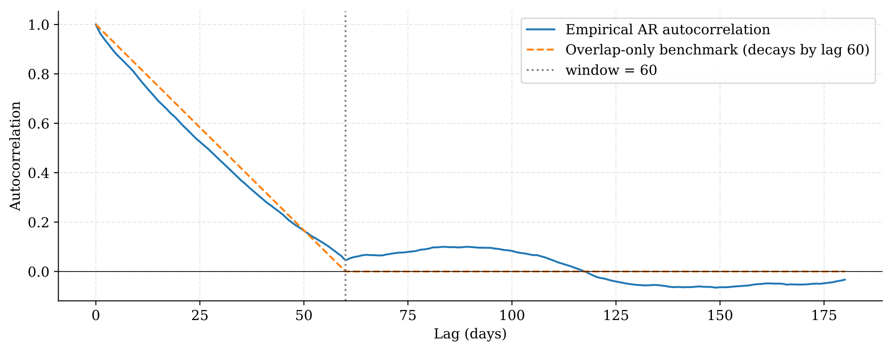

Figure 9. Empirical autocorrelation of the absorption-ratio series $\{AR_t\}$ (solid) against the overlap-only benchmark (dashed), which decays linearly to zero by the window length $\omega = 60$. The empirical curve lies above the benchmark within the window, genuine persistence beyond the mechanical overlap and its mean beyond lag ω is only 0.008, confirming that windows separated by at least ω trading days are close to independent.

3.4. Correlation Analysis and Regime Conditioning

Having established the three-regime structure (Section 3.3), we now ask what the regimes mean economically. A natural objection is that the High-Concentration regime merely re-expresses the level of global markets, that High-Concentration days are simply days on which world equities fell. This section rejects that reading. It first shows that the absorption ratio has no robust relationship with any external benchmark, then demonstrates through a regime-conditional regression that the only quantity which responds to systemic concentration is the funds' own mean absolute active return, and finally localises that response at the cluster level.

Throughout, the interpretive object for the fund side is the mean absolute active return (MAAR), the average absolute deviation of fund returns from their fund-specific benchmarks, $|r_{i,t}^{(f)} - r_{i,t}^{(b,f)}|$. This is the mean absolute (rather than standard-deviation) form of active return; we use the explicit term "mean absolute active return" rather than "tracking error" to avoid confusion with the conventional definition of tracking error as the standard deviation of active returns [48]. Note that for brevity, some figure labels in this paper use "average tracking error"; this refers to MAAR as defined above, not the standard-deviation-based conventional definition. The external series are the three global equity factors and the two euro-area yield shocks of Section 2.2.

Table 18 reports three correlation measures between AR_t and each interpretive series over the full sample: Pearson (parametric), Spearman, and Kendall's τ (rank-based, less sensitive to outliers and ties). Each is reported with a p-value testing $H_0 : \rho = 0$. The interpretive set now includes the euro-area 10-year and 2-year sovereign-yield shocks, addressing the concern that a common imported macro factor might operate through rates rather than equities, particularly for the bond-sensitive conservative cohorts during the 2022–2023 tightening cycle.

The three tests yield a consistent picture. None of the six series exhibits a statistically significant rank-based association with AR_t under either Spearman or Kendall, with p-values uniformly above 0.25. The two Pearson coefficients that reach conventional significance, the mean absolute active return ($r = 0.086$, $p < 0.001$) and MSCI World ($r = -0.060$, $p = 0.018$), are corroborated by neither rank-based test. Because rank correlations are far less sensitive than Pearson to a handful of extreme observations, this divergence indicates that the parametric significance reflects a subset of high-leverage points rather than a continuous monotone relationship spanning the full distribution. Crucially, both euro-area yield shocks are statistically indistinguishable from zero under all three measures (Pearson $p = 0.92$ and 0.20), so the weak link between AR_t and external factors is not an artefact of the original equity-only benchmark set: interest-rate shocks affect the level of individual fund valuations but do not explain the concentration regimes. The conclusion is unchanged when lead–lag structure is allowed for, including the non-synchronous-pricing correction between US-close benchmarks and European-close NAVs (Appendix A).

Figure 10 provides the visual diagnostic. The mean-absolute-active-return panel reveals that the Pearson association is driven primarily by a subset of High-Concentration observations exhibiting both elevated AR_t and elevated active return, while the Low- and Moderate-Concentration observations form a diffuse cloud with no discernible trend. This concentration of the association in one regime explains the divergence between Pearson, which weights extreme values heavily, and the rank-based tests, which treat all observations equally. The equity and yield panels confirm the absence of any continuous relationship: the full-sample OLS lines are essentially flat, and observations are well mixed across regimes at every level of AR_t. Tellingly, the most extreme negative benchmark returns (-0.10 to -0.12) correspond to Low-Concentration rather than High-Concentration

https://doi.org/10.3390/math14142463

observations, an apparent paradox that reflects the 60-day rolling-window construction of AR_t, which captures sustained co-movement over the preceding window rather than instantaneous market shocks.

Table 18. Correlations between the baseline absorption ratio AR_t and the interpretive series, with p-values testing $H_0 : \rho = 0$. Pearson is parametric; Spearman and Kendall are non-parametric robustness checks. $T = 1575$ observations.

Series	Pearson		Spearman		Kendall	
	r	p	ρ	p	τ	p
Mean absolute active return	0.0862 ***	0.0006	0.0289	0.2512	0.0193	0.2504
MSCI World	−0.0597 *	0.0179	−0.0203	0.4217	−0.0138	0.4134
MSCI Europe	−0.0437	0.0830	−0.0128	0.6115	−0.0078	0.6428
S&P 500	−0.0356	0.1575	−0.0106	0.6747	−0.0062	0.7118
ΔY_{10Y}	−0.0026	0.9194	−0.0176	0.4846	−0.0120	0.4744
ΔY_{2Y}	−0.0326	0.1957	−0.0280	0.2672	−0.0186	0.2686

*** $p < 0.001$, * $p < 0.05$.

Regime-conditional regression.

The pattern in Figure 10 suggests that any genuine relationship between AR_t and the interpretive series operates within the High-Concentration regime rather than uniformly across the sample. To test this directly, each interpretive series is regressed on AR_t, a High-Concentration dummy $D_{\text{High},t}$, and their interaction:

$$y_t = b_0 + b_1 AR_t + b_2 D_{\text{High},t} + b_3 (AR_t \cdot D_{\text{High},t}) + \varepsilon_t.$$

This fits two lines simultaneously: a calm-regime line with slope b_1, and a High-Concentration line with slope $b_1 + b_3$. The interaction coefficient b_3 therefore measures the difference in slopes; a significant b_3 indicates a regime-dependent rather than continuous relationship. Standard errors are HAC (Newey–West) with 7 lags, to account for the strong serial dependence in AR_t induced by the 60-day rolling window and consistent with the high regime persistence in Table 15.

Table 19 and Figure 11 reinforce three points. First, the non-high-slope b_1 is small and statistically indistinguishable from zero for all six dependent variables, confirming the absence of a continuous relationship outside High-Concentration periods; this is the near-horizontal blue line in every panel. Second, the interaction b_3 for the mean absolute active return is positive (+0.036), flipping the slope from flat in calm periods ($b_1 = -0.001$) to clearly positive in the High-Concentration state ($b_1 + b_3 = 0.035$). Although b_3 falls short of conventional significance under the conservative HAC adjustment ($p = 0.171$), its sign and magnitude are economically coherent and match the cluster-level amplification documented below. The limited statistical power of the daily-frequency test is expected: the effect is concentrated in roughly one-third of observations, and idiosyncratic active-return variation across the 40 funds is substantial. Third, no global benchmark exhibits a regime-dependent relationship with AR_t; all six interaction coefficients are small, of inconsistent sign, and far from significance, with R^2 below 0.005 for every equity and yield series. The near-parallel calm and high lines in the benchmark panels make this visually unambiguous and rule out the alternative interpretation that systemic co-movement in the pension network is a passive reflection of global market conditions.

Together, the rank-based correlations and the regime-conditional regression resolve the apparent tension in the Pearson results: the parametric significance for the mean absolute active return reflects regime-specific amplification rather than a continuous linear effect, but the daily-frequency tests have limited power to confirm it directly. The cluster-level analysis below provides a more powerful test of the same hypothesis by aggregating observations within structurally homogeneous fund groups under each regime.

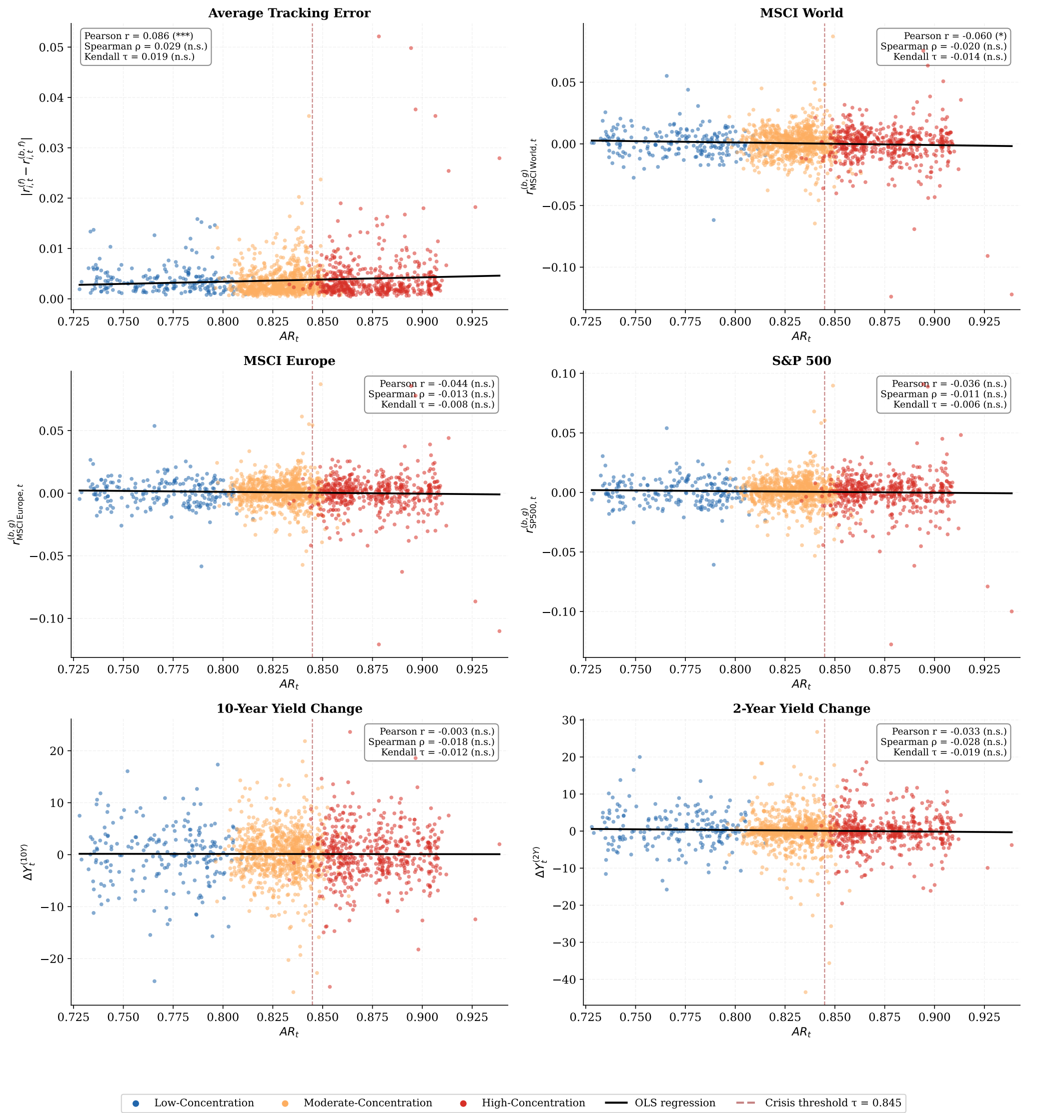

Figure 10. Scatter plots of the absorption ratio AR_t against each interpretive series, with observations coloured by HMM regime (Low: blue, Moderate: amber, High: red). Solid black lines are full-sample OLS fits. Pearson (r), Spearman (ρ), and Kendall (τ) coefficients are reported with significance levels (*** $p < 0.001$, * $p < 0.05$, n.s. not significant). The vertical dashed line marks the crisis threshold $\tau = 0.845$.

Table 19. Regime-conditional OLS regression of each interpretive series on AR_t, a High-Concentration dummy, and their interaction. b_1 is the slope in non-high regimes; b_3 is the additional slope during High-Concentration periods. Standard errors are HAC (Newey–West, 7 lags). $T = 1575$; High-Concentration observations: 513 (32.6%). Yield coefficients are in basis points.

Dependent Variable	b_1	**SE**	b_3	**SE**	$p(b_3)$	R^2
Mean absolute active return	−0.0014	0.0052	0.0364	0.0266	0.1714	0.0176
MSCI World	−0.0082	0.0133	−0.0471	0.0633	0.4571	0.0049
MSCI Europe	−0.0001	0.0115	−0.0314	0.0548	0.5662	0.0031
S&P 500	0.0020	0.0115	−0.0210	0.0494	0.6710	0.0023
ΔY_{10Y}	0.3909	7.2192	−0.3353	14.1861	0.9811	0.0000
ΔY_{2Y}	−11.6330	6.5339	1.1851	13.2372	0.9287	0.0037

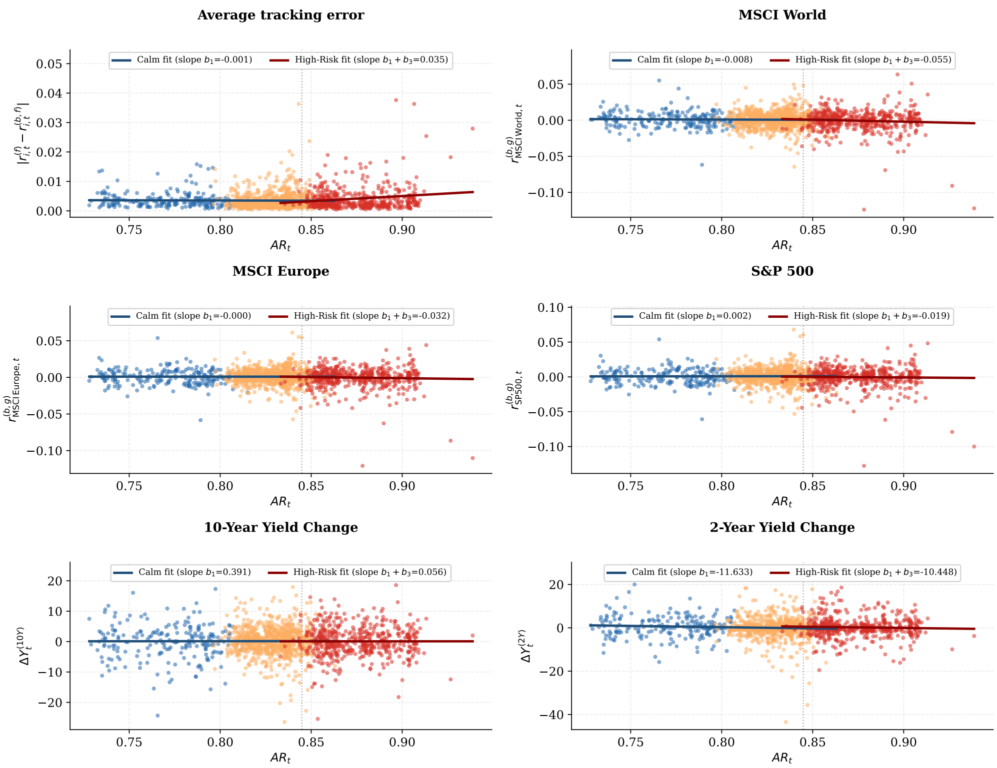

Figure 11. Regime-conditional scatter of AR_t against each interpretive series. The blue line is fitted only to non-high observations (slope b_1); the dark red line only to High-Concentration observations (slope $b_1 + b_3$), each over the AR_t range observed within its regime. The vertical dotted line marks the crisis threshold $\tau = 0.845$. Coefficients and HAC p-values are in Table 19.

Cluster-Level Mean Absolute Active Return Amplification

Figure 12 quantifies the amplification at the cluster level, joining the DTW clustering of Section 3.2 to the HMM regime classification of Section 3.3. For each of the ten clusters, the mean absolute active return is computed separately over High-Concentration days and over all other days, and the crisis amplification ratio is their quotient.

All ten clusters exhibit amplification ratios above one, confirming that the mean absolute active return rises universally during High-Concentration periods regardless of cohort or provider. The magnitude and composition, however, differ systematically between the two cluster types. The four middle-cohort (growth) clusters: C1 (Providers 2 and 5), C2 (Provider 4), C5 (Provider 3), and C6 (Provider 1) carry much higher absolute active returns in both regimes, reflecting their greater equity exposure, and amplify in a tight band of 1.15× to 1.22×. Cluster C6 records the highest absolute level in the sample, rising from 6.4×10^{-3} in normal periods to 7.5×10^{-3} under stress (1.18×), consistent with Provider 1's growth funds carrying the highest mean absolute active return identified in the clustering analysis of Section 3.2.

The six extreme-cohort (conservative) clusters carry far lower absolute active returns but span a wider range of amplification, from 1.06× (C3, Provider 5) to 1.33× (C4, Provider 2)—the largest ratio in the sample, albeit off a very low base (rising from 0.7×10^{-3} to 0.9×10^{-3}). The dispersion within the conservative segment is itself informative: even the most insulated cluster (C3, 1.06×) is not immune to systemic stress, merely less sensitive to it. This carries direct welfare implications. The AG8/TIPF funds embedded in clusters C4, C8, C9, and C10 serve participants who are actively drawing down retirement income, and the standalone TIPF cluster C10 still amplifies by 1.12×. For these participants, any amplification of active-return deviation translates immediately into income uncertainty at a life stage when recovery is not possible.

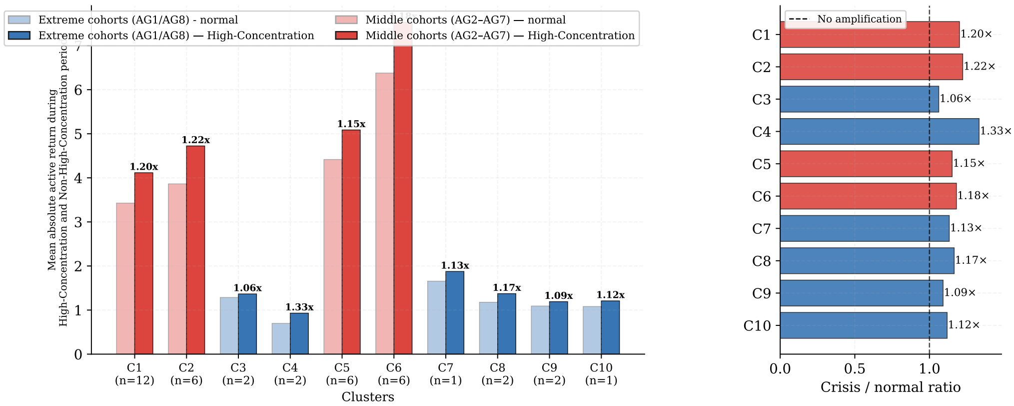

Figure 12. Mean absolute active return by cluster under Non-High and High-Concentration regimes (**left panel**) and the crisis amplification ratio (**right panel**), defined as the mean absolute active return on High-Concentration days divided by that on non-high days. Red bars denote middle-cohort (growth) clusters AG2–AG7; blue bars denote extreme-cohort (conservative) clusters AG1 and AG8/TIPF. Light shading = non-high regime; dark shading = High-Concentration regime.

The cluster-level evidence thus confirms the regime-dependent amplification that the daily-frequency regression of Table 19 could only suggest: mean-absolute-active-return amplification is a genuine, universal feature of the High-Concentration regime, concentrated where it matters most in absolute terms (the growth clusters) but reaching even the capital-preservation funds. This completes the cross-step link from the cluster structure of Section 3.2 through the regime classification of Section 3.3, and substantiates the paper's central reading of pension-network systemic risk as an internal, regulation-driven phenomenon rather than an imported reflection of global markets.

4. Discussion

Read against the pension-economics literature, the result is not that lifecycle glide paths are individually irrational. Age-dependent risk reduction is the standard normative prescription of lifecycle portfolio models. The systemic issue arises when similar lifecycle paths are implemented across all providers in the same funded pillar: such a design can be micro-prudentially attractive and macro-prudentially fragile, protecting each participant from age-inappropriate risk while the funded pillar loses cross-sectional diversification.

The results in Section 3 point to a network whose risk architecture is shaped less by what providers choose to do than by what regulation requires them all to do in similar ways. We organise the discussion around four findings and what they imply for supervision.

4.1. When "Diversification" Stops Being Diversification

Across the seven-year sample, the absorption ratio never falls below 0.728. Even in the calmest 60-day windows we observe, a single common factor explains nearly three-quarters of the variance across 40 funds. By Proposition 3, this corresponds to an average pairwise correlation above 0.72, which is hard to reconcile with the textbook picture of a diversified system.

The clustering results tell us why. The first-order split is between conservative (AG1, AG8) and growth (AG2–AG7) cohorts; within each segment, funds group by provider rather than the reverse. Cluster C1 pulls together the 12 growth-segment funds of Providers 2 and 5, the clearest case of cross-provider merging—while the extreme-cohort funds form their own dedicated, low-deviation clusters. That DTW k-means reproduces the joint provider $\times$ segment design exactly (ARI $=$ 1.00, Table 6), with the hierarchical and k-means partitions in substantial agreement (ARI $=$ 0.78) and DTW agreeing almost exactly with correlation-based clustering (ARI $=$ 0.97, Table 9), confirms that the structure is not an artefact of one algorithm or one distance metric. Because our data are NAV returns rather than holdings, we read this as an associational, institutional pattern rather than a causal one: the returns cannot separately identify lifecycle regulation from voluntary benchmark herding or common provider optimisation, and tracing the full regulation $\rightarrow$ allocation-homogeneity $\rightarrow$ co-movement chain would require holdings data. The cross-provider merging of Providers 2 and 5 (cluster C1, Table 5) is the extreme case: from a systemic-risk perspective, the growth books of these two providers constitute a single functional node, amplifying the network-concentration finding beyond what fund count alone would suggest.

The economic reading of a high systematic risk share therefore inverts in this setting: standard portfolio theory treats it as a sign that idiosyncratic risk has been diversified away, but in a pension network, it signals that the cross-section has been compressed to the point of fragility. This changes the regulator's problem. The standard supervisory framework under IORP treats each fund as the unit of risk; the clustering result implies that the operative unit is the lifecycle segment, not the individual fund. A shock affecting any middle-cohort fund simultaneously stresses all middle-cohort funds across every provider, so a participant who switches providers to reduce exposure is not reducing exposure at all. That this structure is recovered under average linkage (UPGMA), without any squared-Euclidean assumption, and across scale-free distance measures (Section 3.2), confirms it is a property of the data rather than of the linkage criterion. Whether the supervisor should respond by loosening allocation bands or by imposing a network-level diversification requirement is a normative and equilibrium question that a descriptive pipeline cannot settle; we return to it in Section 4.5. What the analysis establishes here is the trade-off itself, which it makes precise and previously invisible.

The risk of regulatory design producing precisely this kind of portfolio synchronisation is not confined to lifecycle allocation rules. In the context of the proposed IORP II revision, PensionsEurope has warned that performance-benchmarking requirements may induce uniform portfolio alignment across providers: institutions seeking to avoid being flagged as outliers allocate increasingly toward reference benchmarks, generating concentration risk and reducing the scope for differentiated long-term strategies [63]. Our clustering results give that warning empirical content: the segment-driven structure documented here is the ex post outcome of a regulatory design that already constrains cross-sectional variation, and the framework makes the cost of deepening it measurable. Any change that increases portfolio similarity within cohorts would raise the absorption-ratio floor above its current minimum of 0.728 and compress the already narrow window between the lower threshold $\tau_1 = 0.813$ and the crisis threshold $\tau = 0.845$.

This also links the result to the literature on diversification across pension pillars: if diversification within the funded pillar is weakened by within-segment homogeneity, then diversification between PAYGO and funded schemes becomes more important as a system-level risk-sharing channel.

4.2. Elevated Co-Movement Is the Norm, Not the Exception

The HMM identifies three regimes. The Low-Concentration regime ($\hat{\mu} = 0.773$) accounts for only 13.3% of trading days, the Moderate-Concentration regime ($\hat{\mu} = 0.829$) for 54.2%, and the High-Concentration regime ($\hat{\mu} = 0.874$) for 32.6%. The long-run stationary distribution (14.1%/55.6%/30.3%) is broadly similar, placing if anything slightly less mass on the High regime. The implication is that monitoring frameworks calibrated on the assumption that calm is the default state will systematically underestimate the frequency and persistence of risk in this sector: the network's baseline condition is already elevated, not one that occasionally enters stress.

Two features of the regime structure deserve emphasis beyond the occupancy shares. First, regimes are sticky: the expected duration of the High-Concentration state is approximately 64 trading days (Table 13), so systemic stress episodes do not resolve quickly. Second, the transition matrix is asymmetric: direct escalation from low to high occurs with probability 0.0046, so the system almost always passes through the moderate band before a full episode. This makes the Moderate-Concentration regime not a neutral middle state but the zone in which a supervisor has the most time to act; a sustained rise of AR_t above the lower threshold $\tau_1 = 0.813$ is a structural precursor signal rather than a coincident one. Prior HMM applications to equity markets typically find low-risk states dominating the unconditional distribution; here, the low state accounts for only 13.3% of days, the opposite pattern.

The empirical crisis threshold $\tau = 0.845$ sits substantially above the range commonly used in the systemic-risk literature [49]. This is simply the homogeneity finding in different units: in a network whose average pairwise correlation never drops below 0.72, the line between ordinary stress and crisis has to sit higher, and importing a generic threshold would label a large majority of the sample as "crisis" without being informative.

4.3. The Network Is Not Just Tracking Global Markets

A natural alternative explanation for the elevated absorption ratio is that it simply mirrors global equity-market stress transmitted through fund holdings. The regime-conditional regressions rule this out. Slopes of AR_t on the S&P 500, MSCI Europe, and MSCI World are statistically indistinguishable from zero in both calm and High-Concentration regimes, with R^2 below 0.006, and the same holds for the euro-area 10-year and 2-year yield shocks, so the conclusion is not specific to equities. The decoupling is visible in Figure 8: through late 2022 and into 2023, global benchmarks stabilise and recover while the network remains in the High-Concentration regime.

The cluster-level result reinforces this. The daily-frequency interaction coefficient is positive and economically meaningful (+0.036) but, as expected with HAC adjustment and a small effective sample within the High-Concentration regime, falls short of conventional significance. Aggregating to the cluster level removes the ambiguity: every one of the ten clusters amplifies during High-Concentration periods, from 1.06× (C3, conservative) to 1.33× (C4, conservative, off a very low base), with the four growth clusters in a tighter 1.15×–1.22× band. Whatever initiates a High-Concentration episode, what sustains it is internal: amplification is universal, not confined to the most equity-exposed funds. The practical implication is direct, supervisory tools designed to monitor externally transmitted shocks (global VaR limits, benchmark drawdown alerts, MSCI factor-exposure reports) are largely blind to this risk. A regulator relying exclusively on market-linked surveillance would have observed recovering global benchmarks through late 2022 and early 2023 while the network remained in the High-Concentration regime. The gap is not a calibration problem better parameters would fix; it is structural, because the signal lies in the cross-sectional co-movement of fund returns rather than in any market index, and closing it requires a network-internal indicator of the kind the absorption ratio provides.

4.4. What the Framework Makes Possible

The principal operational advance is that systemic-risk monitoring for a pension-fund network can be conducted using only publicly available net asset value data, with no proprietary holdings, no bilateral exposure records, and no labelled crisis history. The Bank of Lithuania publishes daily NAV data for all second-pillar funds; the absorption ratio and regime probabilities can be computed from these with a one-day lag and updated as a standing dashboard. As discussed in Section 3.3.5, the operational object for this purpose is the filtered/forward representation rather than the smoothed Viterbi path, and a fully validated early-warning deployment would require out-of-sample evaluation beyond the single COVID-19 episode; we therefore describe the framework as a monitoring and measurement tool, with early-warning use as a moderated, prospective claim. This was not possible with previously available unsupervised frameworks, which either required balance-sheet data unavailable at daily frequency or relied on external benchmarks to define regime boundaries.

The framework's credibility for regulatory use rests on a property most HMM applications to financial data cannot claim: the regime structure is identified by the data rather than by modelling choices. The crisis threshold $\tau = 0.845$ is derived from the emission distributions, not imposed; the number of states is selected by combining BIC with a Bayesian identifiability analysis under two data-calibrated priors; and the EM and Bayesian point estimates agree to within 0.001 across all three regimes (Table 14). A supervisor implementing the framework does not need to defend a discretionary tuning decision.

4.5. Implications for Supervision and Regulation

Three operational implications follow. The first two are measurement-anchored and follow directly; the third is a design question the descriptive pipeline frames but cannot resolve.

For **supervisory monitoring**, the framework yields two reference levels. The lower threshold $\tau_1 = 0.813$ separates the Low- and Moderate-Concentration regimes; given the asymmetric transition structure (Table 15), a sustained absorption ratio above it is a structural precursor to escalation and a natural trigger for enhanced monitoring. The crisis threshold $\tau = 0.845$ marks the boundary of the High-Concentration regime. Consistent with the real-time caveat of Section 3.3.5, these are monitoring signals derived from the retrospective regime decomposition and the forward transition forecasts, not validated alarm levels; both can be computed daily from the NAV data published by the Bank of Lithuania.

For **lifecycle regulation**, the segment-driven cluster structure reveals a tension between micro- and macro-prudential objectives that current rules do not resolve. Lifecycle allocation constraints protect individual participants from age-inappropriate risk but simultaneously remove provider choice as a system-level diversification channel. Two instruments could in principle address this: loosening cohort-level allocation bands to widen provider variation within a cohort, or imposing a system-level diversification requirement at the network rather than the fund level, analogous to concentration limits in banking. We present these as questions, not recommendations, because the method is descriptive and contains no behavioural model of how providers choose portfolios. Whether widening bands would in fact increase differentiation is an equilibrium question, providers might re-optimise toward similar allocations under competitive or benchmarking pressure that an unsupervised method cannot settle; resolving it requires a structural model of provider behaviour. What our analysis contributes is the measurement counterpart: it makes the trade-off between individual- and system-level risk precise and quantifies the segment-driven structure on which any such reform would have to act. Which instrument is preferable ultimately depends on whose risk the regulation is meant to manage, a normative rather than empirical question.

For **participant welfare**, the amplification observed for the conservative AG8/TIPF funds, $1.12\times$ for the standalone TIPF cluster C10, and above one for every conservative cluster is the result that most directly contradicts standard risk disclosures for capital-preservation vehicles. Participants drawing down retirement income during a High-Concentration episode face elevated income uncertainty at a life stage where intertemporal recovery is not available. Risk disclosures and guarantee structures for capital-preservation funds should reflect this residual systemic sensitivity, invisible to fund-level volatility metrics but measurable in the network framework developed here.

5. Conclusions

This paper establishes that systemic risk in a regulated pension-fund network is generated substantially within the network's own structure rather than imported from global markets. The finding reframes a regulatory design question: the same lifecycle allocation rules that protect individual participants leave the funded pillar with little cross-sectional diversification, because provider choice no longer differentiates risk at the system level. A monitoring framework that does not account for this internal dynamic will be blind to a dominant source of risk in the sector.

Applied to 40 Lithuanian second-pillar pension funds over January 2019–September 2025, the framework inverts the causal direction assumed by most supervisory frameworks. The implicit prior model is that external shocks propagate into pension funds through their market exposures; the result here is that the network displays its own High-Concentration states, with no statistically significant relationship between the absorption ratio and global equity benchmarks or euro-area yield shocks in either regime, and spends the majority of its time in elevated states. The High-Concentration state persists through periods of global market recovery, indicating that network stress is not resolved by external stabilisation. Mean-absolute-active-return amplification during stress is universal across all ten clusters and reaches even the conservative funds serving retirement-age participants (the standalone TIPF cluster by 1.12×). Because the analysis uses NAV returns rather than holdings, we read the source of the vulnerability as structural homogeneity consistent with lifecycle allocation regulation rather than as a directly documented causal mechanism.

Three actors face distinct implications. The Bank of Lithuania can implement the monitoring framework immediately: the absorption ratio and regime probabilities are computable from the NAV data it already publishes, and the levels $\tau_1 = 0.813$ and $\tau = 0.845$ provide operationally precise reference points. The European Insurance and Occupational Pensions Authority (EIOPA) is positioned to extend it to cross-jurisdictional comparison. If lifecycle regulation is the operative mechanism, segment-driven cluster structures and internally generated High-Concentration regimes should appear in any EU second-pillar system with comparable lifecycle mandates, a testable prediction the pipeline can evaluate using publicly available NAV data. The Lithuanian regulator faces the normative question that the empirical analysis cannot answer: whether to restore provider differentiation by loosening cohort allocation bands or to introduce network-level diversification requirements that operate above the fund. The analysis makes the trade-off visible and measurable for the first time.

By identifying which episodes matter, which threshold separates regimes, and which structural feature drives co-movement, the framework makes several extensions precise rather than merely plausible. Cross-jurisdictional comparison is the most direct: the pipeline requires only publicly available NAV data, so applying it to other EU second-pillar systems would test whether the segment-driven structure and internally generated concentration regimes are specific to Lithuania or a general consequence of lifecycle regulation. Because the four-state non-identification is most consistent with the label-switching behaviour of overlapping finite Gaussian mixtures rather than any feature specific to this network (Section 3.3.6), replication elsewhere would confirm whether this pattern holds generally rather than reopening it as a structural property. Portfolio-level decomposition is now motivated by a precise question: holdings data would reveal which asset classes drive the leading eigenvalue during High-Concentration periods and whether the fire-sale dynamics hypothesised by [5] are operative. Real-time applicability is a specific empirical question: the forward state probabilities in Table 16 make quantitative predictions for the months following September 2025, which subsequent data will allow to be evaluated against the realised regime sequence.

As noted in Section 1, the Lithuanian application should be read as a complete single-country case study: the framework applies to funded, multi-fund arrangements that publish regular NAV series rather than PAYGO or single defined-benefit schemes, and cross-national or cross-type validation, requiring harmonised daily NAV data, comparable fund classifications, and institutional information on lifecycle rules, is an important next step for assessing external validity.

The broader question this paper opens is whether the tension between micro-prudential protection and macro-prudential homogeneity is specific to lifecycle-regulated pension systems or a structural feature of any regulatory design that standardises portfolio behaviour across a large institutional population. Sovereign wealth funds, insurance general accounts, and target-date retail funds all operate under mandates that constrain cross-sectional variation. If the mechanism generalises, the implication is that diversification at the participant level and fragility at the network level are not accidental co-occurrences but two consequences of the same regulatory logic, and that improving one without worsening the other requires instruments operating at the network rather than the fund level. That is the design problem this paper makes legible.

Author Contributions: Conceptualisation, M.N.J.S. and A.K.; methodology, M.N.J.S.; software, M.N.J.S.; validation, M.N.J.S. and A.K.; formal analysis, M.N.J.S.; investigation, M.N.J.S.; resources, data curation, A.K.; writing—original draft preparation, M.N.J.S.; writing—review and editing, M.N.J.S. and A.K.; visualisation, M.N.J.S.; supervision, A.K.; project administration, A.K.; funding acquisition, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union's Horizon Europe research and innovation programme under MSCA Industrial Doctoral Network on Digital Finance (DIGITAL), grant agreement No. 101119635. More information is available at https://www.digital-finance-msca.com/ (accessed on 28 May 2026).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The pension-fund net asset value data used in this study are sourced from publicly available records published by the Bank of Lithuania (https://www.lb.lt/en/pf-performance-indicators (accessed on 15 November 2025)). Provider identities have been anonymised to preserve confidentiality. Global benchmark series (MSCI World, MSCI Europe, S&P 500) are publicly available via Yahoo Finance. Replication code, including a synthetic dataset generator that reproduces the qualitative structure of the empirical analysis for end-to-end verification, is openly available on GitHub at https://github.com/Staures92/identifiable-regime-detection (accessed on 1 July 2026) and permanently archived on Zenodo at https://doi.org/10.5281/zenodo.20408012.

Acknowledgments: The first author thanks Swedbank (Vilnius, Lithuania) for hosting her industrial secondment under the MSCA Industrial Doctoral Network on Digital Finance. During the preparation of this manuscript, the authors used Claude (Anthropic, Opus 4.7) and ChatGPT (OpenAI and GPT-4.5) for language refinement and restructuring of expository passages. GitHub Copilot (GPT-4.1-based) was used during the development of the Python analysis pipeline for code completions and refactoring suggestions. All analytical methods, model specifications, empirical results, and substantive interpretations are entirely the authors' own. The authors have reviewed and edited all AI-assisted output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AR	Absorption Ratio
MAAR	Mean Absolute Active Return
BIC	Bayesian Information Criterion
BW	Baum–Welch (algorithm)
CRISP-DM	Cross-Industry Standard Process for Data Mining
DB	Davies–Bouldin (index)
DBA	DTW Barycenter Averaging
DTW	Dynamic Time Warping
EM	Expectation-Maximization
HAC	Heteroskedasticity and Autocorrelation Consistent (estimator)
HMM	Hidden Markov Model
IORP	Institutions for Occupational Retirement Provision
MCD	Minimum Covariance Determinant
ML	Machine Learning
MP	Marchenko–Pastur
MSCA	Marie Skłodowska-Curie Actions
NAV	Net Asset Value
NUTS	No-U-Turn Sampler
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
RMT	Random Matrix Theory
TIPF	Target Income Pension Fund

Appendix A. External Factor Correlations and Lead–Lag Analysis

This appendix reports supplementary evidence on the relationship between the baseline absorption ratio AR_t and external financial factors. The objective is to assess whether fluctuations in systemic co-movement within the pension fund network are directly imported from global equity markets or euro-area interest-rate conditions.

Appendix A.1. Contemporaneous Correlations

Table A1 reports simple contemporaneous correlations between the absorption ratio and the interpretive series used in Figure 3. Correlations are uniformly small in magnitude. The strongest relationship is observed for the MSCI World factor (-0.060), while correlations with euro-area yield shocks are close to zero.

These results indicate that although external market conditions influence the overall level of pension fund returns, they explain relatively little of the variation in the internal covariance structure captured by the absorption ratio.

Table A1. Contemporaneous correlations between the baseline absorption ratio and external factors.

Variable	Correlation with AR_t
MAAR	0.0862
MSCI World	-0.0597
MSCI Europe	-0.0437
S&P 500	-0.0356
ΔY_{10Y}	-0.0026
ΔY_{2Y}	-0.0326

Appendix A.2. Lead–Lag Relationships

To account for non-synchronous market closing times and potential delayed transmission effects, lead–lag correlations were computed over a ± 5 trading-day horizon. Following the convention

$$\mathrm{corr}(AR_t, X_{t-k}),$$

positive values of k indicate that the external factor leads the absorption ratio, while negative values imply that the absorption ratio leads the external factor.

Tables A2–A4 report the results for the three global equity benchmarks.

Across all specifications, correlations remain economically small, generally ranging between -0.06 and 0.00. Although several coefficients attain statistical significance owing to the large sample size, their magnitudes remain negligible. Most importantly, there is no evidence of a pronounced positive correlation at $k = 1$, which would be expected if elevated systemic co-movements were primarily driven by delayed transmission from overseas equity markets.

Table A2. Lead–lag correlations between AR_t and MSCI World returns.

Lag k	Correlation	HAC p-Value
−5	−0.050	0.067
−4	−0.047	0.117
−3	−0.043	0.132
−2	−0.051	0.089
−1	−0.038	0.176
0	−0.060	0.055
1	−0.050	0.097
2	−0.060	0.043
3	−0.054	0.057
4	−0.061	0.024
5	−0.049	0.078

Table A3. Lead–lag correlations between AR_t and MSCI Europe returns.

Lag k	Correlation	HAC p-Value
−5	−0.026	0.303
−4	−0.028	0.296
−3	−0.022	0.367
−2	−0.029	0.285
−1	−0.014	0.566
0	−0.044	0.085
1	−0.036	0.112
2	−0.046	0.046
3	−0.037	0.078
4	−0.045	0.025
5	−0.031	0.112

Table A4. Lead–lag correlations between AR_t and S&P 500 returns.

Lag k	Correlation	HAC p-Value
−5	−0.021	0.391
−4	−0.021	0.401
−3	−0.016	0.477
−2	−0.022	0.388
−1	−0.005	0.822
0	−0.036	0.114
1	−0.029	0.144
2	−0.040	0.055
3	−0.032	0.091
4	−0.040	0.029
5	−0.026	0.129

Overall, the evidence suggests that variation in the absorption ratio is only weakly associated with external market factors. While common shocks affect the level of pension fund returns, changes in the degree of systemic co-movement appear to arise primarily from the internal structure of the pension fund network rather than direct transmission from global equity or interest-rate markets.

Appendix B. Covariance Estimator Robustness

This appendix recomputes AR_t using three alternative covariance estimators: Ledoit–Wolf shrinkage [54], the minimum covariance determinant (MCD), and a Marchenko–Pastur (MP) filtered matrix with trace-preserving noise replacement [55,56]. Each targets a different concern about the sample covariance: shrinkage stabilises the spectrum for portfolio optimisation, MCD reduces sensitivity to outliers, and MP filtering separates signal eigenvalues from finite-sample noise.

The MP filter follows [56]. Within each rolling window, returns are standardised to zero mean and unit variance, so the resulting correlation matrix has $\sigma^2 = 1$. The matrix is eigendecomposed; all eigenvalues at or below the upper noise bound

$$\lambda_+ = \sigma^2 \left(1 + \sqrt{N/T}\right)^2$$

are replaced by their average, while eigenvalues above λ_+ are retained. Averaging the bulk eigenvalues preserves the trace, conserving total variance while redistributing noise variance uniformly across the bulk, and the filtered correlation matrix is rescaled to covariance using the original return standard deviations. With $N = 40$, $T = \omega = 60$, and $\sigma^2 = 1$, the bound is $\lambda_+ \approx 3.30$.

Table A5 reports descriptive statistics for the four resulting series, Table A6 their deviations from the sample-covariance baseline, and Figure A1 the series over the full sample.

Table A5. Descriptive statistics of AR_t under alternative covariance estimators, computed from $R^{(f)}$ with rolling window $\omega = 60$.

Estimator	Mean	Std	Min	Max
Sample covariance (baseline)	0.8366	0.0363	0.7280	0.9387
Ledoit–Wolf shrinkage	0.7751	0.0416	0.5970	0.8702
Minimum Covariance Determinant	0.8299	0.0430	0.6735	0.9289
Marchenko–Pastur filtered	0.8255	0.0375	0.7119	0.9334

Table A6. Differences in AR_t relative to the sample-covariance baseline. Mean Diff is the average pointwise difference (alternative − baseline); MAD is the mean absolute deviation; Correlation is the Pearson correlation with the baseline series.

Estimator	Mean Diff	Min Diff	Max Diff	MAD	Correlation
Ledoit–Wolf	−0.061	−0.255	−0.029	0.061	0.68
MCD	−0.007	−0.158	0.089	0.020	0.78
Marchenko–Pastur filtered	−0.011	−0.029	−0.0003	0.011	0.98

The three estimators differ from the baseline in distinct, informative ways.

The **Ledoit–Wolf** estimator produces a systematic downward shift of about 6.1 percentage points (mean 0.775 vs. 0.837), with the largest single reduction reaching −0.255. This is the intended effect of shrinkage, which compresses the eigenvalue spectrum toward a structured target and so reduces the relative dominance of $\lambda_1(t)$. Appropriate for portfolio optimisation, this property works against the purpose of AR_t, which is to measure that dominance directly.

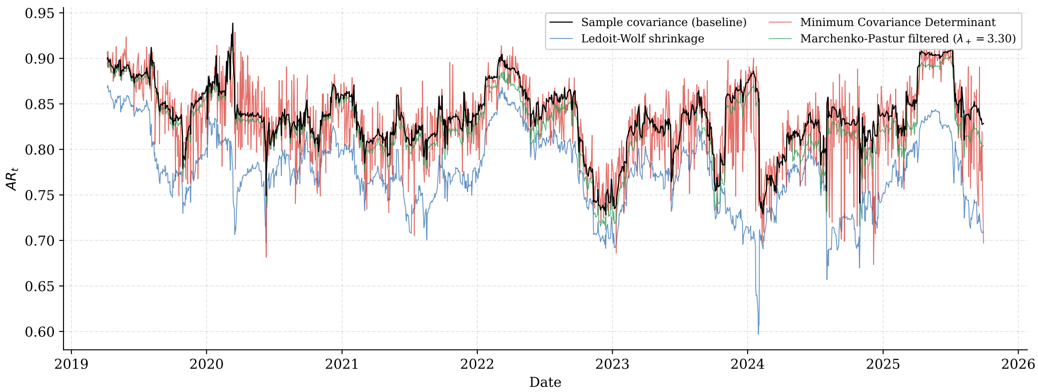

Figure A1. Absorption ratio AR_t under four covariance estimators: sample covariance (black, baseline), Ledoit–Wolf shrinkage (blue), minimum covariance determinant (red), and Marchenko–Pastur filtered (green, $\lambda_+ = 3.30$). The series share similar temporal dynamics. Ledoit–Wolf runs consistently below the baseline owing to spectral shrinkage; MCD tracks the baseline on average but is more volatile day-to-day because outlier flagging is sensitive within each 60-day window; the MP-filtered series tracks the baseline most closely (correlation 0.98), confirming that the leading eigenvalue lies well outside the MP noise bulk.

The **MCD** estimator tracks the baseline closely on average (mean difference -0.007, MAD 0.020) but is locally volatile, with pointwise differences from -0.158 to $+0.089$. This reflects the instability of robust subset selection within short windows: small changes in which observations are flagged as outliers produce visible day-to-day fluctuations in AR_t.

The **Marchenko–Pastur filtered** estimator agrees most closely with the baseline (correlation 0.98, MAD 0.011); all pointwise differences are non-positive (-0.029 to -0.0003), reflecting the trace-preserving construction. This near-equivalence is economically informative: it implies that the leading eigenvalue $\lambda_1(t)$ exceeds the noise bound $\lambda_+ \approx 3.30$ throughout the sample and therefore lies well outside the noise bulk. The filter consequently leaves the dominant signal eigenvalue intact and only redistributes noise variance among the remaining eigenvalues, so AR_t is essentially unchanged. Elevated absorption-ratio values therefore reflect genuine cross-fund co-movement rather than finite-sample noise in the leading principal component, even in the moderately undersampled regime where $N/T \approx 0.67$.

Taken together, the three estimators confirm that the regime structure identified by the HMM is robust to the covariance specification. The sample covariance is retained as the baseline because it preserves the full empirical concentration of variance in $\lambda_1(t)$, the quantity AR_t is designed to measure: MP filtering confirms that this concentration is a genuine signal rather than noise, MCD confirms it is not driven by outliers, and Ledoit–Wolf shows the distortion introduced when an estimator built for a different purpose is applied.

References

1. Billio, M.; Getmansky, M.; Lo, A.W.; Pelizzon, L. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J. Financ. Econ.* **2012**, *104*, 535–559. [CrossRef]
2. Agrawal, S.; Broszeit, T.; Singh, R.; Sugimoto, N.; Surti, J.; Vogelsang, S. *Pension Funds and Financial Stability*; Global Financial Stability Notes; International Monetary Fund: Washington, DC, USA, 2025; Volume 2025. [CrossRef]
3. Greenwood, R.; Landier, A.; Thesmar, D. Vulnerable banks. *J. Financ. Econ.* **2015**, *115*, 471–485. [CrossRef]
4. Delpini, D.; Battiston, S.; Caldarelli, G.; Riccaboni, M. Systemic risk from investment similarities. *PLoS ONE* **2019**, *14*, e0217141. [CrossRef] [PubMed]
5. Cont, R.; Schaanning, E. *Fire Sales, Indirect Contagion and Systemic Stress Testing*; Working Paper 2/2017; Norges Bank: Oslo, Norway, 2017. [CrossRef]
6. Juneja, H. Pension Funds and Systemic Risks. 2025. Available online: https://www.managementstudyguide.com/pension-funds-and-systemic-risks.htm (accessed on 1 July 2026).

7. Financial Stability Board; International Organization of Securities Commissions. *Assessment Methodologies for Identifying Non-Bank Non-Insurer Global Systemically Important Financial Institutions: Proposed High-Level Framework and Specific Methodologies*; Consultative document; Financial Stability Board: Basel, Switzerland, 2014.
8. Bank of England. *Financial Stability in Focus: The FPC's Macroprudential Approach to Operational Resilience*; Bank of England: London, UK, 2024.
9. TIIP The Investment Integration Project. *How Do Asset Owners Address Systemic Risk?* Technical Report; White Paper; Investors for Purpose: Norwich, UK, 2024.
10. Micocci, M.; Gregoriou, G.N.; Masala, G.B. *Pension Fund Risk Management: Financial and Actuarial Modeling*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2010.
11. Lin, Y.; MacMinn, R.D.; Tian, R.; Yu, J. Pension Risk Management in the Enterprise Risk Management Framework. *J. Risk Insur.* **2017**, *84*, 345–365. [CrossRef]
12. Wever, M.; Shah, M.; O'Leary, N. Designing early warning systems for detecting systemic risk: A case study and discussion. *Futures* **2022**, *136*, 102882. [CrossRef]
13. Mashrur, A.; Luo, W.; Zaidi, N.A.; Robles-Kelly, A. Machine Learning for Financial Risk Management: A Survey. *IEEE Access* **2020**, *8*, 203434–203456. [CrossRef]
14. Leo, M.; Sharma, S.; Maddulety, K. Machine learning in banking risk management: A literature review. *Risks* **2019**, *7*, 29. [CrossRef]
15. Caccioli, F.; Barucca, P.; Kobayashi, T. Network models of financial systemic risk: A review. *J. Comput. Soc. Sci.* **2017**, *1*, 81–114. [CrossRef]
16. Eisenberg, L.; Noe, T.H. Systemic Risk in Financial Systems. *Manag. Sci.* **2001**, *47*, 236–249. [CrossRef]
17. Li, B.; Zhang, X. Systemic risk and financial networks. *Q. Rev. Econ. Financ.* **2024**, *94*, 25–36. [CrossRef]
18. Biagini, F.; Fouque, J.P.; Frittelli, M.; Meyer-Brandis, T. A unified approach to systemic risk measures via acceptance sets. *Math. Financ.* **2018**, *29*, 329–367. [CrossRef]
19. Gonon, L.; Meyer-Brandis, T.; Weber, N. Computing Systemic Risk Measures with Graph Neural Networks. *SIAM J. Financ. Math.* **2026**, *17*, 565–619. [CrossRef]
20. Feng, Y.; Min, M.; Fouque, J.P. Deep Learning for Systemic Risk Measures. In *Proceedings of the Third ACM International Conference on AI in Finance (ICAIF '22), New York, NY, USA, 2–4 November 2022*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 62–69. [CrossRef]
21. Balmaseda, V.; Coronado, M.; de Cadenas-Santiago, G. Predicting systemic risk in financial systems using Deep Graph Learning. *Intell. Syst. Appl.* **2023**, *19*, 200240. [CrossRef]
22. Chen, Y.; Wang, G.J.; Zhu, Y.; Xie, C.; Uddin, G.S. Identifying systemic risk drivers of FinTech and traditional financial institutions: Machine learning-based prediction and interpretation. *Eur. J. Financ.* **2024**, *30*, 2157–2190. [CrossRef]
23. Cheng, Y.; Xu, Z.; Chen, Y.; Wang, Y.; Lin, Z.; Liu, J. A Deep Learning Framework Integrating CNN and BiLSTM for Financial Systemic Risk Analysis and Prediction. In *Proceedings of the 2025 International Conference on Economic Management and Big Data Application (ICEMBDA '25), Shenzhen, China, 25–27 July 2025*; Association for Computing Machinery: New York, NY, USA, 2025; pp. 270–274. [CrossRef]
24. Srour, Z.; Hammoud, J.; Tarabay, M. Forecasting Systemic Risk in the European Banking Industry: A Machine Learning Approach. *J. Risk Financ. Manag.* **2025**, *18*, 335. [CrossRef]
25. Tang, P.; Tang, T.; Lu, C. Predicting systemic financial risk with interpretable machine learning. *N. Am. J. Econ. Financ.* **2024**, *71*, 102088. [CrossRef]
26. Li, W. Design of Financial Risk Warning System Based on AI Algorithm and Deep Learning. In Proceedings of the 2024 IEEE 2nd International Conference on Image Processing and Computer Applications (ICIPCA), Shenyang, China, 28–30 June 2024; pp. 1264–1268. [CrossRef]
27. Qiu, Y.; Xu, Y.; Zhu, Y. Design of financial risk warning system in digital economy based on deep neural network. *KSII Trans. Internet Inf. Syst.* **2025**, *19*, 3110–3129. [CrossRef]
28. An, S.; Chen, Y. Systematic financial risk identification and dynamic evolution based on deep learning. *Mob. Inf. Syst.* **2022**, *2022*, 9419248. [CrossRef]
29. Sun, J.; Sun, W.; Sun, L. Design of a Financial Risk Identification and Early Warning Model Based on Machine Learning. In *Recent Developments in Computational Finance and Business Analytics, Proceedings of the 3rd International Conference on Computational Finance and Business Analytics*; Learning and Analytics in Intelligent Systems; Springer: Cham, Switzerland, 2025; Volume 53, pp. 42–51. [CrossRef]
30. Wang, T.; Zhao, S.; Zhu, G.; Zheng, H. A machine learning-based early warning system for systemic banking crises. *Appl. Econ.* **2021**, *53*, 2974–2992. [CrossRef]
31. Laeven, L.; Valencia, F. *Systemic Banking Crises Revisited*; IMF Working Paper WP/18/206; International Monetary Fund: Washington, DC, USA, 2018.

32. Xu, L.; Kinkyo, T.; Hamori, S. Predicting Currency Crises: A Novel Approach Combining Random Forests and Wavelet Transform. *J. Risk Financ. Manag.* **2018**, *11*, 86. [CrossRef]
33. Kritzman, M.; Li, Y.; Page, S.; Rigobon, R. Principal Components as a Measure of Systemic Risk. *J. Portf. Manag.* **2011**, *37*, 112–126. [CrossRef]
34. So, M.K.P.; Mak, A.S.W.; Chu, A.M.Y. Assessing systemic risk in financial markets using dynamic topic networks. *Sci. Rep.* **2022**, *12*, 2668. [CrossRef] [PubMed]
35. Alonso-Robisco, A.; Azqueta-Gavaldón, A.; Carbó, J.M.; González, J.L.; Hernáez, A.I.; Herrera, J.L.; Quintana, J.; Tarancón, J. *Empowering Financial Supervision: A SupTech Experiment Using Machine Learning in an Early Warning System*; Occasional Paper 2504; Banco de España: Madrid, Spain, 2025.
36. Kou, G.; Chao, X.; Peng, Y.; Alsaadi, F.E.; Herrera-Viedma, E. Machine learning methods for systemic risk analysis in financial sectors. *Technol. Econ. Dev. Econ.* **2019**, *25*, 716–742. [CrossRef]
37. Alexandre, M.; Silva, T.C.; Connaughton, C.; Rodrigues, F.A. The drivers of systemic risk in financial networks: A data-driven machine learning analysis. *Chaos Solitons Fractals* **2021**, *153*, 111588. [CrossRef]
38. Aghabozorgi, S.; Seyed Shirkhorshidi, A.; Ying Wah, T. Time-series clustering—A decade review. *Inf. Syst.* **2015**, *53*, 16–38. [CrossRef]
39. Hamilton, J.D. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* **1989**, *57*, 357–384. [CrossRef]
40. Wirth, R.; Hipp, J. CRISP-DM: Towards a Standard Process Model for Data Mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, Manchester, UK, 11–13 April 2000.
41. Kabašinskas, A.; Kopa, M.; Šutienė, K.; Lakštutienė, A.; Malakauskas, A. Stress testing for IInd pillar life-cycle pension funds using hidden Markov model. *Ann. Oper. Res.* **2024**, 1–53. [CrossRef]
42. Kabašinskas, A. Systemic risk assessment of Lithuanian second-pillar pension funds through connectedness and spillover. *J. Math. Ind.* **2024**, *14*, 5. [CrossRef]
43. Auerbach, A.J.; Kotlikoff, L.J. *Dynamic Fiscal Policy*; Cambridge University Press: Cambridge, UK, 1987.
44. Alonso-García, J.; Devolder, P. Optimal mix between pay-as-you-go and funding for DC pension schemes in an overlapping generations model. *Insur. Math. Econ.* **2016**, *70*, 224–236. [CrossRef]
45. Smets, F.; Wouters, R. Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *Am. Econ. Rev.* **2007**, *97*, 586–606. [CrossRef]
46. Cocco, J.F.; Gomes, F.J.; Maenhout, P.J. Consumption and Portfolio Choice over the Life Cycle. *Rev. Financ. Stud.* **2005**, *18*, 491–533. [CrossRef]
47. Fox, E.B.; Sudderth, E.B.; Jordan, M.I.; Willsky, A.S. A sticky HDP-HMM with application to speaker diarization. *Ann. Appl. Stat.* **2011**, *5*, 1020–1056. [CrossRef]
48. Halim, S.; Miller, T.T.; Dupont, D.C. How Pension Funds Manage Investment Risks: A Global Survey. *Rotman Int. J. Pension Manag.* **2010**, *3*, 30–38.
49. Pereda, A.I.C. Systemic Risk and Default Cascades in Global Equity Markets: A Network and Tail-Risk Approach Based on the Gai Kapadia Framework. *arXiv* **2026**, arXiv:2604.19796. [CrossRef]
50. Jolliffe, I.T.; Cadima, J. Principal component analysis: A review and recent developments. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2016**, *374*, 20150202. [CrossRef] [PubMed]
51. Ortobelli, S.; Tichý, T. On the impact of semidefinite positive correlation measures in portfolio theory. *Ann. Oper. Res.* **2015**, *235*, 625–652. [CrossRef]
52. Neděla, D.; Kabašinskas, A.; Staures, M.N.J. Quantile regression–PCA framework in portfolio selection process. *Cent. Eur. J. Oper. Res.* **2026**, 1–25. [CrossRef]
53. Scherer, B. *Portfolio Construction and Risk Budgeting*; Risk Waters Group Ltd.: London, UK, 2002.
54. Ledoit, O.; Wolf, M. A well-conditioned estimator for large-dimensional covariance matrices. *J. Multivar. Anal.* **2004**, *88*, 365–411. [CrossRef]
55. Laloux, L.; Cizeau, P.; Bouchaud, J.P.; Potters, M. Noise Dressing of Financial Correlation Matrices. *Phys. Rev. Lett.* **1999**, *83*, 1467–1470. [CrossRef]
56. Laloux, L.; Cizeau, P.; Potters, M.; Bouchaud, J.P. Random Matrix Theory and Financial Correlations. *Int. J. Theor. Appl. Financ.* **2000**, *3*, 391–397. [CrossRef]
57. Sakoe, H.; Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* **1978**, *26*, 43–49. [CrossRef]
58. Vidal Ruiz, E.; Casacuberta Nolla, F.; Rulot Segovia, H. Is the DTW “distance” really a metric? An algorithm reducing the number of DTW comparisons in isolated word recognition. *Speech Commun.* **1985**, *4*, 333–344. [CrossRef]
59. Ward, J.H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244. [CrossRef]

60. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. R. Stat. Soc. Ser. B (Methodol.)* **1977**, *39*, 1–22. [CrossRef]
61. Keribin, C. Consistent Estimation of the Order of Mixture Models. *Sankhyā Indian J. Stat. Ser. A (1961–2002)* **2000**, *62*, 49–66.
62. Baudry, J.; Raftery, A.E.; Celeux, G.; Lo, K.; Gottardo, R. Combining Mixture Components for Clustering. *J. Comput. Graph. Stat.* **2010**, *19*, 332–353. [CrossRef] [PubMed]
63. PensionsEurope. *PensionsEurope's Position Paper on the IORP II Review: Pension Package of Supplementary Pensions*; Technical Report; Position Paper; PensionsEurope: Brussels, Belgium, 2025.