

Regime-Switching Value-at-Risk Using Hidden Markov Models: A Cross-Market Analysis of Emerging and Developed Equity Indices

Staenly¹ and Maria Yus Trinity Irsan²

Abstract— Accurate risk measurement is essential for financial stability, particularly in markets where returns frequently transition between distinct volatility regimes. This paper applies a Hidden Markov Model (HMM) to daily log returns of three equity indices, namely the Jakarta Composite Index (JCI), Nikkei 225, and S&P 500, over the period January 2008 to December 2024, covering approximately 4,100–4,300 observations per index and multiple major stress episodes. Gaussian HMMs with two, three, and four hidden states are estimated, with the Bayesian Information Criterion selecting three states for JCI and Nikkei 225, and four states for S&P 500, indicating that optimal regime granularity is market-specific. A notable structural finding is that the JCI and S&P 500 transition matrices exhibit near-zero probabilities for direct calm-to-stress transitions, implying that crises are preceded by an intermediate regime of elevated uncertainty. Regime probabilities are used to construct a regime-conditional Value-at-Risk (HMM-VaR), benchmarked against Historical Simulation (HS-VaR) and Student-t GARCH(1,1) VaR (t-GARCH-VaR) using the Kupiec unconditional coverage test and the Christoffersen conditional coverage test. Sub-period analysis at the 95% confidence level reveals that during crisis periods, static methods exceed the nominal violation rate by up to 2.4 times the margin of adaptive methods, with HS-VaR violation rates reaching 11.88% (JCI) and 13.55% (S&P 500) against a nominal 5%. During normal periods, HMM-VaR is the only method that consistently passes both coverage tests across all three markets, with violation rates closely approximating the nominal level (4.90%–5.61%). The findings demonstrate that regime-switching VaR provides superior and more consistent risk measurement across diverse market structures, while also revealing that aggregate backtesting can be misleading due to the cancellation of opposing errors between crisis and normal sub-periods.

Keywords: Cross-Market Analysis, Hidden Markov Model, Regime-Switching, Value-at-Risk

I. INTRODUCTION

*This work was not supported by any organization

¹Staenly is with the Department of Actuarial Science, President University, staenly@president.ac.id

²Maria Yus Trinity Irsan is with the Department of Actuarial Science, President University, maria.irsan@president.ac.id

Manuscript received May 18, 2026; accepted September 16, 2026.

VALUE-AT-RISK (VaR) has become the industry standard for measuring market risk and is widely adopted by banks and financial institutions under the regulatory framework of the Basel Committee. Despite its widespread use, a fundamental limitation of VaR lies in its dependence on the assumed distribution of asset returns. In practice, the accuracy of VaR estimates is highly sensitive to this assumption. This issue is particularly pronounced in markets where returns exhibit non-stationarity, volatility clustering, and fat-tailed distributions, violating the assumptions underlying conventional VaR models [1], [2].

These distributional challenges are not confined to any single market. The Jakarta Composite Index (JCI), representing one of Southeast Asia's largest emerging markets, has experienced several major stress episodes over the past fifteen years, including the Global Financial Crisis in 2008, the taper tantrum in 2013, the commodity price downturn in 2015, the COVID-19 pandemic in 2020, and the global monetary tightening cycle during 2022–2023. Developed markets exhibit similar dynamics: the Nikkei 225 was severely affected by the Fukushima disaster in 2011 and the China-led slowdown of 2015, while the S&P 500 experienced the U.S. debt ceiling crisis in 2011 and the sharp Q4 selloff of 2018, in addition to the global crises shared across all markets. In each case, the market disruption was characterized not merely by a decline in average returns but by a structural change in the return distribution, including volatility spikes and amplified tail risk [3], [4]. As a consequence, traditional VaR models tend to underestimate risk precisely during periods when accurate risk measurement is most critical, potentially leading to insufficient capital buffers [5], [6].

The adoption of regime-switching approaches offers a potential solution to this limitation. By intuition, financial returns are unlikely to originate from a single homogeneous distribution; instead, they may be better described as arising from a mixture of distinct regimes. For instance, a low-volatility environment with positive mean returns may be characterized as a "calm" regime, while a high-volatility environment with negative returns corresponds to a "stress" regime [7], [8], [9]. Hidden Markov Models (HMM) provide a probabilistic framework to capture such regime dynamics endogenously, without requiring prior specification of regime boundaries. By allowing the data to determine the underlying state structure, HMM offers a flexible approach to modeling time-varying risk. Importantly, the estimated regime probabilities can be directly incorporated into VaR computation, resulting in a

more adaptive risk measure [10], [11].

These models have been applied in developed financial markets. The example work of regime-switching models in macroeconomic analysis was done by Akbal (2024), while from the other perspective, Tan (2025) extended the framework to equity returns context [12], [13]. Kelendji (2023) applied regime-switching models to international asset allocation, particularly focusing on time-varying correlations [14]. Subsequent studies, including Ding (2025), Li (2025), and Shu (2024) have further explored applications in volatility forecasting and dynamic portfolio allocation [15], [16], [17].

Several studies have demonstrated the advantages of regime-switching VaR models in the context of risk measurement. Assaf (2017) examined international equity markets, while more recent studies such as Luo (2022) and Ma (2017) extended the analysis to futures markets [18], [19], [20]. A consistent finding across these studies is that regime-switching VaR models tend to outperform conventional approaches, particularly during periods of heightened market volatility. However, empirical evidence for Southeast Asian markets, and Indonesia in particular, remains limited.

Within the Indonesian context, existing VaR studies predominantly rely on historical simulation, parametric approaches, or GARCH-type models. While these methods capture certain aspects of time-varying risk, they generally assume a single underlying distribution. Applications of regime-switching models to the JCI are relatively scarce and are often limited to descriptive analyses, such as identifying bull and bear phases, rather than integrating regime detection into practical risk measurement frameworks. Moreover, cross-market studies comparing the behavior of regime-switching VaR across markets with different structural characteristics, such as emerging versus developed equity markets, remain limited. Given the regulatory role of the Financial Services Authority (OJK) in overseeing risk management practices, the development of more accurate and adaptive risk measures carries direct practical relevance.

This paper addresses the gap by applying a HMM to daily returns of three equity indices (JCI, Nikkei 225, and S&P 500) over the period 2008–2024. The estimated regime structures are used to construct regime-switching VaR (HMM-VaR) forecasts for each market. The performance of the HMM-VaR is benchmarked against two conventional approaches: Historical Simulation VaR and Student-t GARCH(1,1) VaR, the latter of which explicitly accommodates heavy-tailed return distributions through the use of Student-t innovations. Model accuracy is evaluated using the Kupiec unconditional coverage test and the Christoffersen conditional coverage test, with particular emphasis on differences in performance between normal and stress periods.

There are four main contributions of this paper. First, it provides one of the first systematic applications of HMM-based VaR that compares performance across emerging and developed equity markets using a common methodological framework, enabling direct assessment of whether the advantages of regime-switching risk measurement are market-specific or generalizable. Second, it demonstrates that the optimal number of hidden states is market-dependent which shown

by the BIC, it selects three regimes for the JCI and Nikkei 225 but four for the S&P 500, suggesting that markets with greater depth and complexity exhibit more granular regime structures. Third, it reveals a structural property of market dynamics whereby direct transitions between calm and stress regimes are essentially impossible in certain markets, requiring passage through an intermediate regime of elevated uncertainty. Fourth, it uncovers a dual failure pattern in static VaR methods that is obscured by aggregate backtesting, simultaneous overestimation during normal periods and underestimation during crises, both with opposing errors canceling which produce misleadingly accurate aggregate violation rates.

The remainder of this paper is organized as follows. Section 2 presents the theoretical framework underlying Hidden Markov Models, Value-at-Risk, and backtesting procedures. Section 3 describes the data, methodology, and research design. Section 4 presents the empirical results and discussion. Section 5 concludes with a summary of findings, practical implications, and directions for future research.

II. PRELIMINARIES

A. Hidden Markov Model and State Probabilities

A Hidden Markov Model (HMM) is a doubly stochastic process consisting of an unobserved state process and an observed emission process. Let $\{r_t\}_{t=1}^T$ denote the observed time series of log returns, and let $\{S_t\}_{t=1}^T$ denote the corresponding sequence of hidden states taking values in $\mathcal{S} = \{1, 2, \dots, K\}$.

The state process follows a homogeneous first-order Markov chain, governed by a transition matrix $\mathbf{P} = [p_{ij}]$ with entries

$$p_{ij} = P(S_{t+1} = j \mid S_t = i), \quad (1)$$

satisfying

$$\sum_{j=1}^K p_{ij} = 1. \quad (2)$$

The initial state distribution is denoted by $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)^\top$. Conditional on the current state, returns follow a Gaussian emission distribution:

$$r_t \mid S_t = k \sim \mathcal{N}(\mu_k, \sigma_k^2), \quad k = 1, \dots, K. \quad (3)$$

The complete parameter vector is $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \mathbf{P}, \boldsymbol{\pi})$, estimated via the Baum–Welch algorithm, an expectation–maximization procedure that maximises the observed-data log-likelihood [21], [22].

Two conditional distributions over the hidden state are of central interest. The *filtered probability*

$$\xi_t(k) = P(S_t = k \mid r_{1:t}) \quad (4)$$

conditions on information up to time t and is computed via the forward algorithm. The *smoothed probability*

$$\gamma_t(k) = P(S_t = k \mid r_{1:T}) \quad (5)$$

conditions on the full sample via the forward–backward algorithm. The filtered probability is relevant for real-time forecasting, whereas the smoothed probability supports retrospective regime analysis.

B. Value-at-Risk and Its Regime-Conditional Extension

For a loss $L_{t+1} = -r_{t+1}$ and confidence level $1 - \alpha$, Value-at-Risk (VaR) is defined as the α -quantile of the conditional loss distribution:

$$\text{VaR}_{\alpha,t} = \inf \{ \ell \in R : P(L_{t+1} > \ell \mid \mathcal{F}_t) \leq \alpha \}, \quad (6)$$

where $\mathcal{F}_t$ is the information filtration at time t . Equivalently, VaR is the negative of the α -quantile of the return distribution. Common confidence levels in practice are $1 - \alpha = 0.95$ and $1 - \alpha = 0.99$.

The regime-switching extension (HMM-VaR) exploits the filtered probabilities to construct a predictive distribution that adapts to evolving market conditions. Given $\xi_t(k)$, the one-step-ahead predicted state probabilities are

$$\hat{\pi}_{t+1|t}(k) = \sum_{j=1}^K \xi_t(j) p_{jk}, \quad (7)$$

and the one-step-ahead predictive density of the return is the Gaussian mixture

$$f(r_{t+1} \mid \mathcal{F}_t) = \sum_{k=1}^K \hat{\pi}_{t+1|t}(k) \phi(r_{t+1}; \mu_k, \sigma_k^2), \quad (8)$$

where $\phi(\cdot; \mu_k, \sigma_k^2)$ denotes the Gaussian density with mean μ_k and variance σ_k^2 .

The HMM-VaR at level α is defined as

$$\text{VaR}_{\alpha,t}^{\text{HMM}} = -F^{-1}(\alpha \mid \mathcal{F}_t), \quad (9)$$

where F is the cumulative distribution function associated with the mixture density above. Since this quantile lacks a closed form for $K \geq 2$, it is computed numerically via bisection.

The adaptive property of this construction is central. When filtered probabilities shift toward a high-volatility regime, the weight on the corresponding mixture component rises, the tails of the predictive distribution become heavier, and the VaR estimate increases accordingly—without any external intervention.

C. Backtesting Value-at-Risk

VaR accuracy is evaluated via two standard coverage tests based on the violation indicator

$$I_t = 1\{r_t < -\text{VaR}_{\alpha,t}\}. \quad (10)$$

The Kupiec (1995) *unconditional coverage test* examines whether the observed violation proportion matches the nominal level α . Its likelihood ratio statistic is given by

$$LR_{uc} = -2 \ln \left(\frac{(1 - \alpha)^{n_0} \alpha^{n_1}}{(1 - \hat{\pi})^{n_0} \hat{\pi}^{n_1}} \right), \quad (11)$$

follows an asymptotic χ_1^2 distribution under H_0 . The Kupiec test evaluates whether the average frequency of violations is correct, but does not assess whether violations are clustered in time [23].

The Christoffersen (1998) conditional coverage test extends this framework by additionally requiring that violations are independently distributed over time. Define the transition counts n_{ij} for $i, j \in \{0, 1\}$ as the number of times a day

with violation status i is followed by a day with status j . The estimated transition probabilities are

$$\hat{\pi}_{01} = \frac{n_{01}}{n_{00} + n_{01}}, \hat{\pi}_{11} = \frac{n_{11}}{n_{10} + n_{11}} \quad (12)$$

The independence statistic is

$$LR_{ind} = -2 \ln \left(\frac{(1 - \hat{\pi})^{n_{00} + n_{10}} \hat{\pi}^{n_{01} + n_{11}}}{(1 - \hat{\pi}_{01})^{n_{00}} \hat{\pi}_{01}^{n_{01}} (1 - \hat{\pi}_{11})^{n_{10}} \hat{\pi}_{11}^{n_{11}}} \right) \quad (13)$$

and the combined conditional coverage statistics is

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi_2^2, \quad (14)$$

under the joint null of correct coverage and independence [24]. The two tests are complementary. A model that passes the Kupiec test but fails the Christoffersen test produces the correct average violation rate but clusters its violations in time, it is a pattern indicating failure to respond dynamically to changing market conditions. This is precisely the type of deficiency that regime-switching approaches are designed to address.

III. METHODOLOGY

This section describes the research procedure used to apply the framework of Section 2 to three equity indices representing distinct market structures. The methodology is organized sequentially: data preparation, HMM estimation and model selection, VaR computation, and backtesting design.

A. Data

Daily closing prices are obtained for three equity indices over the period from 1 January 2008 to 31 December 2024: the Jakarta Composite Index (JCI; Yahoo Finance ticker $\hat{\text{JKSE}}$), representing an emerging market in Southeast Asia; the Nikkei 225 ($\hat{\text{N225}}$), representing a developed market in Asia; and the S&P 500 ($\hat{\text{GSPC}}$), representing the world's largest developed equity market. The selection of these three indices enables a comparative analysis across markets with different structural characteristics, including differences in depth, sectoral composition, and sensitivity to regional versus global shocks.

After differencing, the sample yields 4,124 daily log returns for the JCI, 4,156 for the Nikkei 225, and 4,278 for the S&P 500. The sample period encompasses several major stress episodes. Three are global in nature and affect all indices: the global financial crisis (2008–2009), the COVID-19 pandemic (2020), and the global monetary tightening cycle (2022). Additional market-specific episodes include the taper tantrum (2013) and commodity price collapse (2015) for the JCI; the Fukushima disaster (2011) and China-led slowdown (2015) for the Nikkei 225; and the U.S. debt ceiling crisis (2011) and the Q4 sell-off (2018) for the S&P 500. This combination of shared and idiosyncratic stress events provides sufficient variation to evaluate VaR performance across diverse market conditions.

Log returns are expressed in percentage points,

$$r_t = 100 \times \ln \left(\frac{P_t}{P_{t-1}} \right), \quad (15)$$

to facilitate numerical stability during HMM estimation.

B. HMM Estimation and Model Selection

For each index, Gaussian HMMs with $K = 2$, $K = 3$, and $K = 4$ hidden states are estimated on the full sample using the Baum-Welch algorithm. To mitigate the risk of convergence to local optima (difficulty of the expectation-maximization (EM) algorithm) each specification is estimated with twenty random initializations, and the solution with the highest log-likelihood is retained. Convergence is assessed using a tolerance of 10^{-3} on the change in log-likelihood between successive EM iterations, with a variance floor of 10^{-3} imposed to prevent degenerate solutions in which one state collapses to a near-zero variance.

The choice of Gaussian emission distributions is motivated by three considerations. First, while individual returns exhibit heavy tails, a mixture of K Gaussian components can approximate heavy-tailed behavior through the combination of components with different variances, particularly when a high-variance stress regime generates substantial probability mass in the tails. Second, the Gaussian HMM is the most widely implemented specification in the applied HMM literature, facilitating comparison with prior studies. Third, the use of Gaussian emissions provides a transparent baseline that clearly reveals the limitations of the normality assumption, as demonstrated in the empirical results. To complement this baseline, a Student- t GARCH(1, 1) model is included among the benchmarks to directly assess the incremental value of fat-tailed innovations.

Model selection is based on the Bayesian Information Criterion (BIC) as the primary criterion, with the Akaike Information Criterion (AIC) and log-likelihood as supporting criteria. The BIC is preferred over AIC in this application because of its stronger theoretical foundation for model selection in time series contexts and its more conservative penalization of parameter proliferation. After selection of the preferred K , the regimes are labeled by sorting states in ascending order of variance, and economic interpretation is assigned *post hoc* based on the estimated means and standard deviations.

It is important to clarify that the model selection procedure uses the full sample as a preliminary exploratory step to determine the optimal state dimension K . This full-sample analysis is used solely for architectural selection and is not used for VaR forecasting. Once the optimal K is determined for each index, the rolling out-of-sample VaR estimation proceeds with K fixed, while all model parameters (μ_k , σ_k^2 , $\mathbf{P}$, and $\boldsymbol{\pi}$) are re-estimated within each rolling window using only data available up to time t . The state dimension is not re-selected within each window, as the regime structure is assumed to be a stable architectural feature of each market over the sample period. This approach is standard in the applied HMM literature and balances computational feasibility with out-of-sample integrity.

C. VaR Computation

For each index, three VaR models are constructed at the 95% confidence level ($1 - \alpha = 0.95$): the HMM-VaR using the selected K , and two benchmarks—Historical Simulation VaR (HS-VaR) and Student- t GARCH(1, 1) VaR (t-GARCH-VaR).

All models are implemented with a rolling estimation window of $w = 250$ trading days, corresponding to approximately one calendar year.

Historical Simulation VaR is a non-parametric method that estimates VaR as the empirical α -quantile of the return distribution within the rolling window:

$$\text{VaR}_{\alpha,t}^{\text{HS}} = -\hat{q}_{\alpha}(r_{t-w+1}, \dots, r_t), \quad (16)$$

where $\hat{q}_{\alpha}(\cdot)$ denotes the empirical quantile operator with linear interpolation when the quantile index falls between observations. This method makes no distributional assumptions, which enables it to capture the actual tail weight of the data regardless of distributional form, but at the cost of slow responsiveness to changes in market conditions.

Student- t GARCH(1, 1) VaR models conditional heteroskedasticity while accommodating fat-tailed innovations. The conditional variance follows

$$\sigma_t^2 = \omega + a\varepsilon_{t-1}^2 + b\sigma_{t-1}^2, \quad (17)$$

subject to $\omega > 0$, $a \geq 0$, $b \geq 0$, and $a + b < 1$ for covariance stationarity. Unlike the standard Gaussian GARCH, the standardized residuals are assumed to follow a Student- t distribution with ν degrees of freedom, allowing the model to capture the excess kurtosis observed in the data. Parameters (μ, ω, a, b, ν) are estimated by quasi-maximum likelihood using the `arch` package in Python, with the rolling window re-estimated at each time step. The one-step-ahead VaR is computed as

$$\text{VaR}_{\alpha,t}^{\text{t-GARCH}} = -\left(\hat{\mu} + t_{\alpha}(\nu) \cdot \hat{\sigma}_{t+1|t} \cdot \sqrt{\frac{\nu-2}{\nu}}\right), \quad (18)$$

where $t_{\alpha}(\nu)$ is the α -quantile of the Student- t distribution with ν degrees of freedom, and the scaling factor $\sqrt{(\nu-2)/\nu}$ adjusts the GARCH conditional standard deviation to the standardized Student- t parameterization.

HMM-VaR is computed by forming the Gaussian mixture predictive distribution described in Subsection 2.B, using the filtered probabilities at time t projected forward through the transition matrix. The α -quantile of the mixture is obtained by numerical inversion of the cumulative distribution function via bisection. To avoid look-ahead bias, the HMM is re-estimated on each rolling window using only information available up to time t , with five random initializations per window. This ensures that all VaR forecasts, including those based on the regime-switching approach are genuinely out-of-sample. The computational cost of rolling re-estimation is substantial (3,874 windows for the JCI, 3,906 for the Nikkei 225, and 4,028 for the S&P 500, with five random initializations per window), but is essential for the credibility of the out-of-sample evaluation.

D. Backtesting Design

The out-of-sample period for each index spans from day $w + 1 = 251$ through the end of the sample, yielding 3,874 forecast–realization pairs for the JCI, 3,906 for the Nikkei 225,

and 4,028 for the S&P 500. The Kupiec and Christoffersen tests introduced in Subsection 2.C are applied to each model at the 95% confidence level.

To evaluate whether model performance differs between stress episodes and tranquil conditions, the out-of-sample period for each index is partitioned into crisis and normal sub-periods. Crisis windows are defined *a priori* using calendar dates associated with major market stress events, determined by two criteria: (i) the start date corresponds to the month in which the triggering event occurred or was publicly recognized, and (ii) the end date corresponds to the month in which daily volatility, measured by a 20-day rolling standard deviation, returned to within one standard deviation of its long-run average. Three global crisis windows are applied to all indices: the Global Financial Crisis (September 2008–March 2009), the COVID-19 pandemic (February 2020–June 2020), and the global monetary tightening cycle (January 2022–October 2022). Two additional market-specific windows are defined for each index: the Taper Tantrum (May–September 2013) and Commodity Price Collapse (June–December 2015) for the JCI; the Fukushima disaster (March–June 2011) and China-led slowdown (August–December 2015) for the Nikkei 225; and the U.S. debt ceiling crisis (July–October 2011) and Q4 selloff (October–December 2018) for the S&P 500.

The use of predetermined calendar windows, rather than endogenous regime classifications from the HMM, ensures that the sub-period definitions are independent of the models being evaluated, thereby avoiding circularity in the assessment. Table I summarizes the crisis classifications.

TABLE I
CRISIS WINDOW CLASSIFICATIONS

Index	Crisis Period	Label
All	Sep. 2008–Mar. 2009	Global Financial Crisis (GFC)
All	Feb. 2020–Jun. 2020	COVID-19
All	Jan. 2022–Oct. 2022	Global Tightening
JCI	May 2013–Sep. 2013	Taper Tantrum
JCI	Jun. 2015–Dec. 2015	Commodity Price Collapse
Nikkei 225	Mar. 2011–Jun. 2011	Fukushima Disaster
Nikkei 225	Aug. 2015–Dec. 2015	China-led Slowdown
S&P 500	Jul. 2011–Oct. 2011	U.S. Debt Ceiling Crisis
S&P 500	Oct. 2018–Dec. 2018	Q4 Selloff

Backtesting is performed separately on each sub-period, allowing direct assessment of whether the regime-switching approach delivers its theoretical advantage precisely where it is most needed.

All computations are implemented in Python, using the `hmmlearn` package for HMM estimation, the `arch` package for GARCH estimation, and custom implementations of the forward algorithm and backtesting procedures.

IV. RESULT AND DISCUSSION

This section presents the empirical results obtained from applying the methodology of Section 3 to daily log returns of the JCI, Nikkei 225, and S&P 500 over the period January 2008 to December 2024. The discussion proceeds from preliminary data analysis through HMM estimation and

regime interpretation, and culminates in the comparative VaR backtesting that constitutes the core empirical contribution of this paper.

A. Data and Preliminary Analysis

Table II reports the descriptive statistics for all three indices. All three return series share the stylized facts characteristic of financial data: negative skewness, substantial excess kurtosis, and overwhelming rejection of the normality hypothesis by the Jarque-Bera test ($p < 0.001$ in all cases).

TABLE II
DESCRIPTIVE STATISTICS OF DAILY LOG RETURNS, JANUARY 2008–DECEMBER 2024

Statistic	JCI	Nikkei 225	S&P 500
Observations	4,124	4,156	4,278
Mean (%)	0.023	0.024	0.033
Std. Dev. (%)	1.212	1.498	1.271
Minimum (%)	−11.306	−13.234	−12.765
Maximum (%)	9.704	13.235	10.957
Skewness	−0.593	−0.532	−0.527
Excess Kurtosis	10.088	9.010	12.394
Jarque-Bera	17,729.83	14,253.73	27,579.55
JB p -value	< 0.001	< 0.001	< 0.001

Several cross-market differences are noteworthy. The Nikkei 225 exhibits the highest daily volatility (1.498%), consistent with the well-documented volatility of the Japanese equity market. The S&P 500 displays the highest excess kurtosis (12.394), indicating the heaviest tails among the three indices despite its reputation as one of the world's most liquid and efficient equity markets. The JCI exhibits the most pronounced negative skewness (−0.593), suggesting greater asymmetry in downside risk for the Indonesian market.

The extreme observations are also informative. The largest single-day losses range from −11.31% (JCI) to −13.23% (Nikkei 225), all occurring during or near the COVID-19 market shock of March 2020. Such extreme realizations, present across all three market structures, provide a stringent test for any risk measurement framework and underscore the need for models capable of accommodating non-Gaussian distributional behavior.

B. HMM Estimation and Model Selection

Table III summarizes the model selection results across the three indices.

TABLE III
MODEL SELECTION ACROSS CANDIDATE STATE DIMENSIONS

K	JCI BIC	Nikkei BIC	S&P 500 BIC
2	11,723.28	14,018.70	12,145.65
3	11,566.19	13,856.80	11,735.78
4	11,576.74	13,877.05	11,656.09
Selected K	3	3	4

The BIC selects $K = 3$ for the JCI and Nikkei 225, but $K = 4$ for the S&P 500. This finding suggests that the optimal number of regimes is market-specific rather than

universal. The S&P 500's selection of four states is consistent with the hypothesis that markets with greater depth, liquidity, and sectoral diversity exhibit more granular regime structures, requiring an additional intermediate state to adequately capture the range of market conditions. For the JCI, the choice of $K = 3$ over $K = 4$ is further supported by examination of the four-state solution, in which States 0 and 1 have highly similar means (0.072 and 0.051) and standard deviations (0.55 and 0.98), suggesting that the fourth state captures variation that is empirically difficult to distinguish from a neighboring regime. The $K = 3$ specification produces three regimes that are clearly separated in both mean and volatility. An analogous pattern holds for the Nikkei 225.

C. Regime Interpretation

Table IV reports the estimated parameters for the selected HMM specification for each index.

TABLE IV
ESTIMATED HMM PARAMETERS

Panel A: JCI ($K = 3$)				
Regime	Label	Mean (%)	Std. Dev. (%)	Duration (days)
0	Calm	0.077	0.591	25.5
1	Normal	0.028	1.117	18.2
2	Stress	-0.259	2.777	18.4

Panel B: Nikkei 225 ($K = 3$)				
Regime	Label	Mean (%)	Std. Dev. (%)	Duration (days)
0	Calm	0.122	0.855	46.1
1	Normal	-0.015	1.467	31.6
2	Stress	-0.453	3.843	13.8

Panel C: S&P 500 ($K = 4$)				
Regime	Label	Mean (%)	Std. Dev. (%)	Duration (days)
0	Calm	0.126	0.429	13.8
1	Moderate	0.053	0.909	14.3
2	Elevated	-0.055	1.551	46.1
3	Stress	-0.413	4.030	36.1

All three indices share a common qualitative structure: a low-volatility regime with positive mean returns (calm), one or more intermediate regimes, and a high-volatility regime with negative mean returns (stress). However, the quantitative characteristics differ meaningfully across markets.

The Nikkei 225 calm regime is the most persistent, with an average duration of 46.1 days, whereas its stress regime is the least persistent, lasting only 13.8 days on average. This pattern suggests that the Japanese market experiences intense but relatively short-lived stress episodes. In contrast, the S&P 500 stress regime persists for an average of 36.1 days, indicating more prolonged crisis dynamics in the U.S. market. The JCI exhibits the most balanced duration structure, with all three regimes lasting approximately 18–26 days.

The stress-regime volatilities also differ across markets: 2.78% for the JCI, 3.84% for the Nikkei 225, and 4.03% for the S&P 500. In absolute terms, the developed markets exhibit higher stress-regime volatility. However, relative to their respective calm regimes, the JCI displays a volatility ratio of approximately $4.7 \times$ ($2.78/0.59$), compared with $4.5 \times$ for the Nikkei 225 ($3.84/0.86$) and $9.4 \times$ for the S&P 500

($4.03/0.43$). The exceptionally large ratio for the S&P 500 primarily reflects its remarkably low calm-regime volatility (0.43%) rather than an unusually high level of stress-regime volatility.

The estimated transition matrices reveal a striking structural property that is broadly consistent across the three markets while exhibiting important market-specific nuances.

The estimated transition matrix for the JCI is

$$\mathbf{P}_{\text{JCI}} = \begin{bmatrix} 0.9607 & 0.0393 & \approx 0 \\ 0.0431 & 0.9451 & 0.0118 \\ \approx 0 & 0.0542 & 0.9458 \end{bmatrix}. \quad (19)$$

For the Nikkei 225, the estimated transition matrix is

$$\mathbf{P}_{\text{Nikkei}} = \begin{bmatrix} 0.9783 & 0.0217 & \approx 0 \\ 0.0223 & 0.9684 & 0.0093 \\ 0.0022 & 0.0704 & 0.9274 \end{bmatrix}. \quad (20)$$

For the S&P 500, the estimated transition matrix is

$$\mathbf{P}_{\text{S\&P500}} = \begin{bmatrix} 0.9275 & 0.0666 & 0.0059 & \approx 0 \\ 0.0633 & 0.9302 & 0.0052 & 0.0013 \\ \approx 0 & 0.0192 & 0.9783 & 0.0025 \\ \approx 0 & \approx 0 & 0.0277 & 0.9723 \end{bmatrix}. \quad (21)$$

The transition matrix $\mathbf{P}$ for each index is the maximum likelihood estimate of the transition probabilities defined in (1) and (2), obtained using the Baum–Welch algorithm applied to the corresponding full sample. Each entry $\hat{p}_{ij}$ represents the estimated probability of transitioning from regime i to regime j on the following trading day. By construction, every row sums to one, satisfying the stochastic constraint given in (2).

The JCI and S&P 500 exhibit a sequential transition structure in which direct transitions between the calm and stress regimes are essentially absent ($\hat{p}_{02} \approx 10^{-10}$ for the JCI and $\hat{p}_{03} \approx 10^{-32}$ for the S&P 500). Consequently, transitions between tranquil and crisis conditions occur through one or more intermediate regimes. This finding has an intuitive economic interpretation: periods of market stress do not emerge abruptly from calm conditions but are typically preceded by an interval of elevated uncertainty during which the market occupies an intermediate state. Likewise, recoveries from crisis conditions proceed through transitional regimes before returning to the calm state.

The Nikkei 225 exhibits a slightly different transition pattern. While $\hat{p}_{02} \approx 0$, indicating that direct transitions from the calm regime to the stress regime are effectively absent, the reverse transition probability is estimated as $\hat{p}_{20} = 0.0022$, which is small but non-zero. This finding suggests that the Japanese market occasionally experiences rapid recoveries from stress directly to calm conditions without passing through an extended transitional regime. Such an asymmetry may reflect the influence of policy interventions by the Bank of Japan during periods of market stress.

The high diagonal entries across all estimated transition matrices (all ≥ 0.92) indicate strong regime persistence, consistent with the well-established empirical phenomenon of volatility clustering in financial markets.

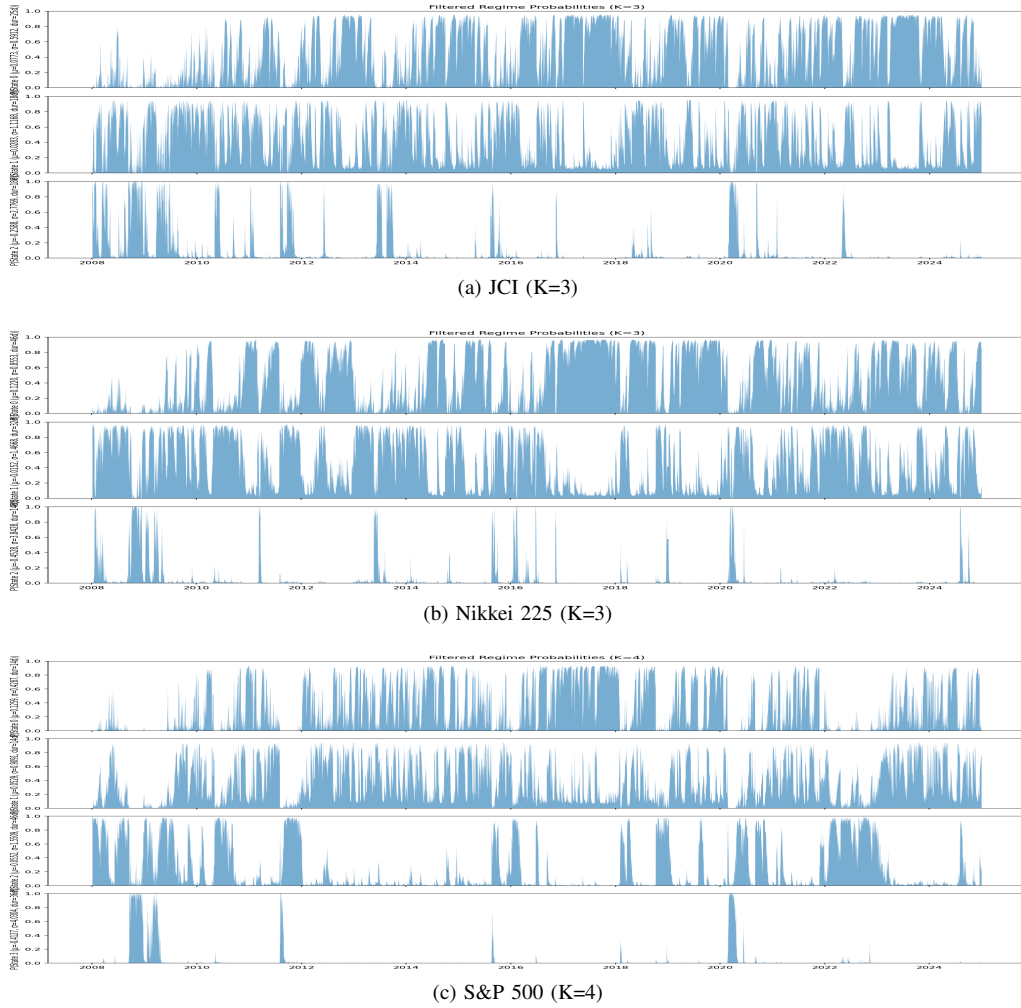


Fig. 1. Filtered regime probabilities estimated on daily log returns, January 2008 – December 2024. Top to bottom within each panel: calm regime (State 0), intermediate regime(s), and stress regime (highest-numbered state). The S&P 500 panel contains four states reflecting its BIC-optimal $K = 4$ specification.

Examination of the smoothed state probability paths (Figures 1(a)-1(c)) further confirms that the estimated regimes correspond closely to identifiable historical market events across all three indices. The stress regime is activated prominently during the Global Financial Crisis of late 2008 and early 2009, the COVID-19 market shock of March–April 2020, and the global monetary tightening cycle of 2022. Market-specific episodes are likewise captured by the model, including the Taper Tantrum of mid-2013 and the commodity price collapse of late 2015 for the JCI, the Fukushima disaster of 2011 for the Nikkei 225, and the U.S. debt ceiling crisis of 2011 together with the fourth-quarter selloff of 2018 for the S&P 500. These correspondences provide external validation that the estimated HMM identifies economically meaningful market regimes rather than merely statistical artifacts.

D. VaR Computation and Full-Sample Backtesting

Following the methodology developed in Sections II and III, three VaR models were constructed for each index: the regime-switching HMM-VaR, Historical Simulation VaR (HS-VaR), and Student- t GARCH(1, 1) VaR (t-GARCH-VaR). All models were implemented using a rolling estimation window

of $w = 250$ trading days at the 95% confidence level ($\alpha = 0.05$). The out-of-sample evaluation comprises 3,874 observations for the JCI, 3,906 for the Nikkei 225, and 4,028 for the S&P 500.

Table V presents the full-sample VaR backtesting results.

At the aggregate level, HS-VaR is the only method that consistently passes the Kupiec unconditional coverage test across all three markets, with violation rates closely matching the nominal 5% level (5.09% for the JCI, 4.97% for the Nikkei 225, and 4.87% for the S&P 500). In contrast, HMM-VaR passes the Kupiec test only for the JCI (5.34%) but fails for the Nikkei 225 (5.89%) and the S&P 500 (6.11%), where the observed violation rates exceed the nominal level. The Student- t GARCH-VaR performs least favorably, failing the Kupiec test across all three indices and producing the highest violation rates in every market (5.81%, 6.27%, and 6.73%, respectively), indicating a systematic tendency to underestimate downside risk.

All three methods fail the Christoffersen conditional coverage test for every index, indicating that VaR violations remain temporally clustered regardless of the forecasting approach. Nevertheless, the magnitude of the clustering differs across

TABLE V
-SAMPLE VaR BACKTESTING RESULTS AT $\alpha = 0.05$

Index	Method	Viol. Rate	LR_{uc}	p_{uc}	Kupiec	LR_{cc}	p_{cc}	Christoffersen
JCI	HMM-VaR	0.0534	0.94	0.332	Pass	7.60	0.022	Fail
	HS-VaR	0.0509	0.06	0.808	Pass	35.93	< 0.001	Fail
	t-GARCH	0.0581	5.07	0.024	Fail	12.18	0.002	Fail
Nikkei 225	HMM-VaR	0.0589	6.16	0.013	Fail	9.23	0.010	Fail
	HS-VaR	0.0497	0.01	0.924	Pass	15.33	< 0.001	Fail
	t-GARCH	0.0627	12.36	< 0.001	Fail	12.55	0.002	Fail
S&P 500	HMM-VaR	0.0611	9.74	0.002	Fail	10.37	0.006	Fail
	HS-VaR	0.0487	0.15	0.695	Pass	27.41	< 0.001	Fail
	t-GARCH	0.0673	22.95	< 0.001	Fail	22.96	< 0.001	Fail

models. The HMM-VaR consistently produces the smallest LR_{cc} statistics (7.60 for the JCI, 9.23 for the Nikkei 225, and 10.37 for the S&P 500), suggesting that the regime-switching specification adapts to changes in market conditions more effectively than either HS-VaR or Student- t GARCH-VaR, even though the improvement is insufficient to satisfy the conditional coverage criterion.

To provide additional insight into the magnitude of the risk estimates produced by each method, Table VI reports the average one-day-ahead VaR forecasts over the full sample as well as separately for crisis and normal market conditions.

The crisis–normal spread, defined as the difference between the average VaR during crisis and normal periods, provides a direct measure of each model’s responsiveness to changing market conditions. HMM-VaR exhibits the largest spread across all three markets (0.59 percentage points for the JCI, 0.67 percentage points for the Nikkei 225, and 1.73 percentage points for the S&P 500), demonstrating the strongest sensitivity to shifts in market regimes. Student- t GARCH-VaR produces comparable spreads, reflecting the adaptive nature of its autoregressive volatility process. In contrast, HS-VaR exhibits the smallest spreads across all markets. For the JCI, the average crisis-period VaR (1.743%) is even slightly lower than the average normal-period VaR (1.762%), indicating that the historical simulation approach fails to increase its risk estimates during periods of heightened market stress.

Figure 2 illustrates the out-of-sample VaR forecasts together with the realized daily returns for all three indices. Across all three markets, both HMM-VaR and Student- t GARCH-VaR respond rapidly to changes in market volatility, particularly during the COVID-19 market shock of March 2020, when both methods sharply increase their VaR forecasts within only a few trading days after the initial sell-off. By contrast, HS-VaR adjusts much more gradually because it relies exclusively on a rolling historical window. After market conditions normalize, HS-VaR remains elevated for an extended period, with its VaR estimates returning to pre-crisis levels only after the crisis observations leave the 250-day estimation window.

While these full-sample results provide useful evidence regarding overall forecasting performance, they may conceal important differences between crisis and tranquil market conditions. The following subsection therefore disaggregates the backtesting analysis into crisis and normal sub-periods.

E. Crisis-Window Analysis

The out-of-sample period for each index was partitioned into crisis and normal sub-periods according to the classifications presented in Table I. Table VII reports the VaR backtesting results for the crisis periods.

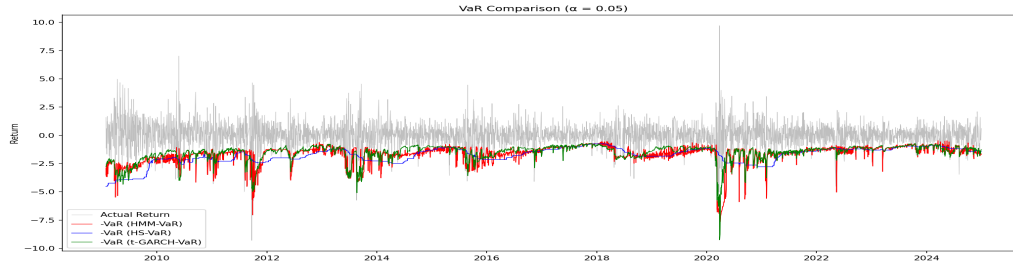
During crisis periods, all three methods fail the Kupiec unconditional coverage test across every index, indicating that none fully captures the elevated level of downside risk associated with severe market stress. Nevertheless, two important patterns emerge from the cross-market comparison.

First, the magnitude of risk underestimation varies substantially across both markets and forecasting methods. For the JCI, HS-VaR produces the highest violation rate (11.88%), which is substantially larger than those of HMM-VaR (7.81%) and Student- t GARCH-VaR (8.32%). A similar pattern is observed for the S&P 500, where HS-VaR yields the highest violation rate (13.55%), compared with identical rates of 10.69% for HMM-VaR and Student- t GARCH-VaR. In contrast, the Nikkei 225 exhibits only modest differences across methods, with violation rates ranging from 7.05% to 7.61%, and HS-VaR marginally outperforming the adaptive approaches. These findings suggest that the benefit of adaptive risk models during crisis periods is market-dependent and is most pronounced in markets where crises are characterized by abrupt volatility spikes, such as the JCI and the S&P 500, rather than in markets such as the Nikkei 225, where stress episodes evolve more gradually.

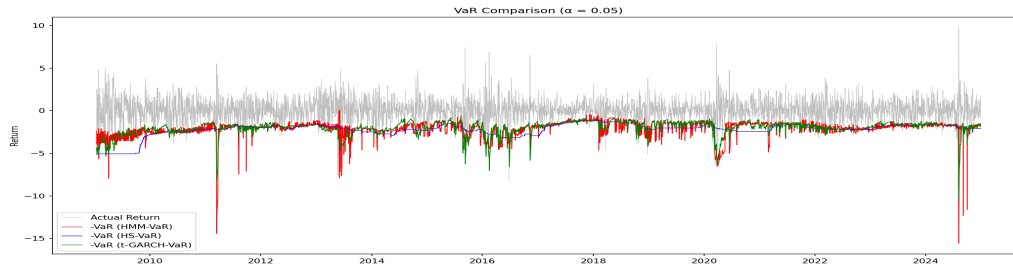
Second, the underlying reason for the relatively poor performance of HS-VaR during crisis periods is evident from the average VaR estimates reported in Table VI. For the JCI, the average HS-VaR during crisis periods (1.743%) is virtually identical to its average during normal periods (1.762%), indicating that the rolling historical window fails to increase the estimated risk despite a marked deterioration in market conditions. By comparison, the adaptive models respond much more aggressively. The average HMM-VaR for the JCI increases from 1.577% during normal periods to 2.162% during crisis periods, representing an increase of approximately 37%. Similarly, Student- t GARCH-VaR adjusts immediately through its conditional variance process in response to large squared returns, while HMM-VaR reacts through an increase in the filtered probability assigned to the stress regime following sequences of unusually adverse returns. Although neither adaptive approach fully captures the

TABLE VI
AVERAGE VaR ESTIMATES BY METHOD AND MARKET CONDITION ($\alpha = 0.05$)

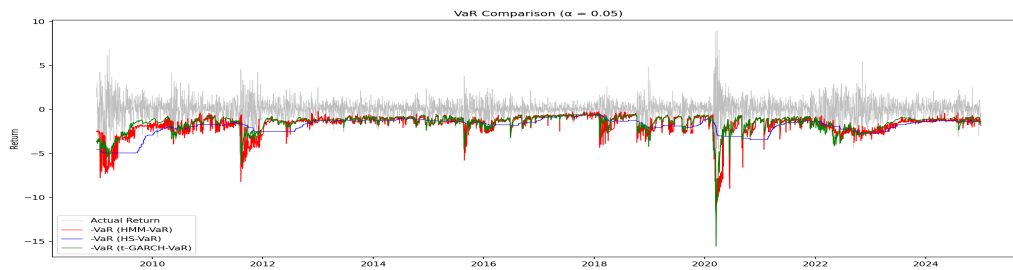
Index	Method	Full (%)	Crisis (%)	Normal (%)	Spread (pp)
JCI	HMM-VaR	1.666	2.162	1.577	0.585
	HS-VaR	1.759	1.743	1.762	-0.019
	t-GARCH	1.535	1.972	1.457	0.515
Nikkei 225	HMM-VaR	2.099	2.677	2.006	0.671
	HS-VaR	2.218	2.434	2.183	0.251
	t-GARCH	2.004	2.575	1.913	0.662
S&P 500	HMM-VaR	1.726	3.227	1.502	1.725
	HS-VaR	1.882	2.360	1.810	0.550
	t-GARCH	1.520	2.910	1.312	1.598



(a) JCI



(b) Nikkei 225



(c) S&P 500

Fig. 2. Out-of-sample VaR comparison at the 95% confidence level ($\alpha = 0.05$). Grey: realized daily log returns. Red: HMM-VaR. Blue: HS-VaR. Green: Student-t GARCH-VaR. A violation occurs when the realized return falls below a method's VaR line.

TABLE VII
VaR BACKTESTING RESULTS DURING CRISIS PERIODS ($\alpha = 0.05$)

Index	Days	Method	Viol. Rate	LR_{uc}	p_{uc}	Kupiec
JCI	589	HMM-VaR	0.0781	8.42	0.004	Fail
		HS-VaR	0.1188	43.12	< 0.001	Fail
		t-GARCH	0.0832	11.49	< 0.001	Fail
Nikkei 225	539	HMM-VaR	0.0761	6.70	0.010	Fail
		HS-VaR	0.0705	4.25	0.039	Fail
		t-GARCH	0.0724	5.01	0.025	Fail
S&P 500	524	HMM-VaR	0.1069	27.30	< 0.001	Fail
		HS-VaR	0.1355	56.12	< 0.001	Fail
		t-GARCH	0.1069	27.30	< 0.001	Fail

severity of market stress—as evidenced by their continued failure of the Kupiec test—they adjust considerably faster than the static historical simulation method.

Table VIII reports the corresponding backtesting results for the normal market periods.

The normal-period results reveal the strongest finding of this study: HMM-VaR is the only method that satisfies both the Kupiec unconditional coverage test and the Christoffersen conditional coverage test across all three equity markets during tranquil market conditions. Its violation rates (4.90% for the JCI, 5.61% for the Nikkei 225, and 5.42% for the S&P 500) remain close to the nominal 5% level, while the corresponding Christoffersen test results indicate that VaR violations occur independently over time rather than exhibiting statistically significant clustering. These findings suggest that the regime-switching framework produces well-calibrated risk forecasts when financial markets operate under relatively stable conditions.

It is important to emphasize that a violation rate below the nominal level should not be interpreted as superior forecasting performance. The Kupiec test is inherently two-sided and rejects the null hypothesis whenever the observed violation rate deviates significantly from the target level α , regardless of whether the deviation is above or below the nominal probability. Consequently, excessively conservative VaR forecasts are penalized in the same manner as overly aggressive forecasts. HS-VaR illustrates this phenomenon clearly. The observed violation rates of 3.87% for the JCI and 3.57% for the S&P 500 are significantly below the nominal 5% level, leading to rejection of the Kupiec test in both markets. This systematic overestimation of risk arises because the rolling 250-day estimation window continues to include observations from previous crisis periods for up to one calendar year after market conditions have normalized. For example, the extreme negative returns recorded during the COVID-19 market collapse of March 2020 remain within the historical sample until approximately March 2021, causing the historical simulation approach to produce excessively conservative VaR estimates long after the period of elevated volatility has ended.

Student- t GARCH-VaR exhibits the opposite pattern. Rather than overestimating risk, it tends to underestimate downside risk during normal market conditions, producing violation rates above the nominal level (5.36% for the JCI, 6.12% for the Nikkei 225, and 6.14% for the S&P 500). Although the deviation for the JCI is sufficiently small to satisfy the Kupiec test, the larger deviations observed for the Nikkei 225 and the S&P 500 lead to rejection of the null hypothesis. This evidence suggests that, despite explicitly modeling conditional heteroskedasticity and heavy-tailed innovations, the Student- t GARCH specification generates VaR forecasts that are systematically too low during tranquil market conditions. One possible explanation is that, for the estimated degrees of freedom obtained in the rolling estimation, the Student- t innovation distribution produces quantiles that are insufficiently conservative at the 5% significance level once volatility has reverted to its normal state.

Table IX summarizes the relative performance of the three VaR models across all evaluation settings.

Table IX reveals several important findings. First, HS-VaR ranks first in the full-sample analysis for all three indices. However, this apparent superiority is misleading. As demonstrated by the sub-period analysis, the favorable aggregate performance of HS-VaR results from the cancellation of two opposing sources of error: systematic overestimation of risk during normal market conditions offsets systematic underestimation during crisis periods, yielding an overall violation rate that happens to lie close to the nominal 5% level. Consequently, the full-sample ranking masks substantial deficiencies in the model's responsiveness to changing market environments.

Second, when the evaluation is conducted separately for crisis and normal periods, HMM-VaR ranks first or second in five of the six sub-period evaluations, and it is the only method that achieves full regulatory compliance (passing both the Kupiec and Christoffersen tests) during normal market conditions for every index. The single exception is the Nikkei 225 crisis window, where the three methods are separated by only 0.56 percentage points (7.05% to 7.61%) and the ordering carries little economic content; this contrasts sharply with the JCI and S&P 500 crisis windows, where the gap between HS-VaR and the adaptive methods reaches 4.07 and 2.86 percentage points respectively.

Third, Student- t GARCH-VaR, despite explicitly accommodating heavy-tailed innovations, does not consistently outperform either HMM-VaR or HS-VaR. During normal periods, it tends to underestimate downside risk, resulting in violation rates above the nominal level for both the Nikkei 225 and the S&P 500. During crisis periods, its performance is generally comparable to, but not systematically better than, that of HMM-VaR. These findings suggest that modeling heavy-tailed innovations alone is insufficient for reliable VaR estimation. Accurate risk measurement also requires the ability to recognize changes in the underlying market regime and adjust forecasts accordingly. This additional dimension of adaptiveness is naturally incorporated within the HMM framework but is not fully captured by the conventional GARCH model, even when Student- t innovations are employed.

V. CONCLUSION

This paper has investigated the application of Hidden Markov Models to regime detection and Value-at-Risk estimation across three equity indices representing distinct market structures: the Jakarta Composite Index (emerging market), the Nikkei 225 (developed Asian market), and the S&P 500 (developed Western market), over the period January 2008 to December 2024. For each index, a Gaussian HMM was estimated and used to construct regime-conditional VaR forecasts. The HMM-VaR was benchmarked against Historical Simulation VaR and Student- t GARCH(1,1) VaR using the Kupiec unconditional coverage test and the Christoffersen conditional coverage test at the 95% confidence level.

The empirical analysis yields several key findings. First, the BIC-optimal number of regimes is market-specific: three states for the JCI and Nikkei 225, and four states for the S&P 500. All indices share a common qualitative structure

TABLE VIII
VAR BACKTESTING RESULTS DURING NORMAL PERIODS ($\alpha = 0.05$)

Index	Days	Method	Viol. Rate	LR_{uc}	p_{uc}	Kupiec	LR_{cc}	p_{cc}	Christoffersen
JCI	3,285	HMM-VaR	0.0490	0.07	0.794	Pass	1.29	0.526	Pass
		HS-VaR	0.0387	9.61	0.002	Fail	17.73	< 0.001	Fail
		t-GARCH	0.0536	0.87	0.352	Pass	3.05	0.218	Pass
Nikkei 225	3,367	HMM-VaR	0.0561	2.57	0.109	Pass	4.40	0.111	Pass
		HS-VaR	0.0463	0.98	0.323	Pass	6.58	0.037	Fail
		t-GARCH	0.0612	8.30	0.004	Fail	8.54	0.014	Fail
S&P 500	3,504	HMM-VaR	0.0542	1.28	0.257	Pass	2.63	0.269	Pass
		HS-VaR	0.0357	16.75	< 0.001	Fail	28.74	< 0.001	Fail
		t-GARCH	0.0614	8.90	0.003	Fail	8.91	0.012	Fail

TABLE IX
SUMMARY OF METHOD RANKINGS BY VIOLATION RATE ACCURACY

Index	Full Sample	Crisis Period	Normal Period
JCI	1. HS-VaR (5.09%) * 2. HMM-VaR (5.34%) * 3. t-GARCH-VaR (5.81%) –	1. HMM-VaR (7.81%) – 2. t-GARCH-VaR (8.32%) – 3. HS-VaR (11.88%) –	1. HMM-VaR (4.90%) ** 2. t-GARCH-VaR (5.36%) ** 3. HS-VaR (3.87%) –
Nikkei 225	1. HS-VaR (4.97%) * 2. HMM-VaR (5.89%) – 3. t-GARCH-VaR (6.27%) –	1. HS-VaR (7.05%) – 2. t-GARCH-VaR (7.24%) – 3. HMM-VaR (7.61%) –	1. HS-VaR (4.63%) * 2. HMM-VaR (5.61%) ** 3. t-GARCH-VaR (6.12%) –
S&P 500	1. HS-VaR (4.87%) * 2. HMM-VaR (6.11%) – 3. t-GARCH-VaR (6.73%) –	1. HMM-VaR (10.69%) – 1. t-GARCH-VaR (10.69%) – 3. HS-VaR (13.55%) –	1. HMM-VaR (5.42%) ** 2. HS-VaR (3.57%) – 3. t-GARCH-VaR (6.14%) –
** = Pass Kupiec and Christoffersen; * = Pass Kupiec only; – = Fail Kupiec.			

consisting of calm, intermediate, and stress regimes, but differ in their quantitative characteristics. The JCI exhibits the most balanced regime duration structure (18–26 days), the Nikkei 225 has the most persistent calm regime (46.1 days) but the shortest stress regime (13.8 days), and the S&P 500 requires an additional intermediate state to capture its more granular market dynamics. These differences suggest that risk models should not assume a universal regime architecture but rather allow the data to determine the appropriate number of states for each market.

Second, the estimated transition matrices reveal a sequential transition property in the JCI and S&P 500: the probability of direct transitions between the calm and stress regimes is essentially zero, requiring passage through one or more intermediate regimes. This structural finding implies that crises do not emerge abruptly from tranquil conditions but are preceded by periods of elevated uncertainty. The Nikkei 225 exhibits a slightly different pattern, with a small but non-zero probability of direct stress-to-calm transitions ($p_{20} = 0.002$), possibly reflecting the influence of policy interventions. This cross-market comparison of transition dynamics constitutes a novel contribution to the regime-switching literature.

Third, at the full-sample level, the HS-VaR achieves the most accurate aggregate violation rates across all three markets (5.09%, 4.97%, 4.87%). However, the sub-period analysis reveals that this accuracy is misleading: it arises from the cancellation of opposing errors between crisis and normal conditions. During crisis periods, all methods fail the Kupiec test, but the magnitude of failure varies substantially. In the JCI, the HS-VaR exceedance above the nominal level (6.88 percentage points) is approximately 2.4 times that of the

HMM-VaR (2.81 percentage points). In the S&P 500, the HS-VaR (13.55%) again substantially exceeds both the HMM-VaR and Student-*t* GARCH-VaR (both 10.69%). The average VaR analysis confirms the mechanism: the HS-VaR for the JCI produces a crisis-period average (1.743%) that is marginally lower than its normal-period average (1.762%), indicating a complete failure to raise risk estimates during stress.

Fourth, during normal periods, the HMM-VaR is the only method that passes both the Kupiec and Christoffersen tests across all three markets, with violation rates of 4.90% (JCI), 5.61% (Nikkei), and 5.42% (S&P 500). The HS-VaR overestimates risk during normal periods in the JCI (3.87%) and S&P 500 (3.57%), producing violation rates significantly below the nominal level, because the rolling window retains crisis-period observations for up to one year after stress subsides. The Student-*t* GARCH-VaR exhibits the opposite tendency, producing violation rates above 5% in the Nikkei (6.12%) and S&P 500 (6.14%), suggesting that fat-tailed innovations alone do not guarantee accurate VaR estimation. The HMM-VaR's consistent accuracy across all three markets—emerging and developed alike—provides strong evidence that the regime-switching framework offers genuinely generalizable advantages for risk measurement during normal market conditions.

These findings carry practical implications for risk management across different market contexts. For financial institutions and regulators—including the Otoritas Jasa Keuangan (OJK) in Indonesia, the Financial Services Agency (FSA) in Japan, and the Securities and Exchange Commission (SEC) in the United States—the results suggest that continued reliance on Historical Simulation VaR entails a dual cost: insufficient protection during stress periods and excessive capital alloca-

tion during calm periods. The adoption of regime-switching approaches would improve accuracy across both conditions, particularly for day-to-day internal risk monitoring at the 95% confidence level.

Several limitations should be acknowledged. First, the HMM specification employs Gaussian emission distributions. While the mixture structure captures substantial non-normality, the adoption of heavier-tailed emission distributions such as Student- t within the HMM framework could further improve tail risk estimation. Second, the analysis considers only one-day-ahead VaR at the 95% confidence level; multi-period horizons and the 99% level used in regulatory capital calculations raise additional considerations. Third, the rolling re-estimation approach treats each window independently, without exploiting potential structural continuity across adjacent windows. Fourth, the crisis window boundaries, while based on explicit criteria, involve a degree of judgment that may influence the sub-period results.

These limitations suggest several directions for future research. The adoption of Student- t or other heavy-tailed emission distributions within the HMM framework would preserve the regime-switching adaptiveness while better approximating extreme tail behavior. Extension to multi-asset portfolio VaR, potentially using coupled or factorial HMMs to model cross-market regime dependencies, would broaden the practical applicability of the framework. The integration of Markov-switching GARCH (MS-GARCH) models, which combine regime-switching dynamics with conditional heteroskedasticity within a unified framework, represents a particularly promising direction that could capture both the regime identification capabilities of the HMM and the intra-regime volatility dynamics of the GARCH approach. Finally, the sequential transition property identified in the JCI and S&P 500 transition matrices warrants formal investigation as a potential early-warning mechanism for emerging market stress.

REFERENCES

- [1] S. Dlamini and S. C. Shongwe, "Comparative analysis of tail risk in emerging and developed equity markets: An extreme value theory perspective," *International Journal of Financial Studies*, vol. 14, p. 11, 1 Jan. 2026, ISSN: 22277072. DOI: 10.3390/ijfs14010011
- [2] P. Jakšić, M. Pjanić, M. Milošević, D. Marjanović, B. Dudić, and L. Stošić, "Impact of macroeconomic factors on stock indices volatility: A garch analysis for usa, germany, serbia and croatia," *Computational Economics*, 2025, ISSN: 15729974. DOI: 10.1007/s10614-025-11170-1
- [3] S. Peng and Y. Shi, "Information transmission across markets: Tail risk spillovers and cross-market volatility forecasting," *Mathematics*, vol. 14, 4 Feb. 2026, ISSN: 22277390. DOI: 10.3390/math14040686
- [4] P. Gurrola-Perez, D. Murphy, and P. G. Perez, "The impulsive approach to procyclicality measuring the reactivity of risk-based initial margin models to changes in market conditions using impulse response functions," Tech. Rep., 2024.
- [5] C. Confalonieri and P. D. Vincentiis, "Forecasting the worst: Is implied volatility forward-looking enough?" *Journal of Banking Regulation*, vol. 27, 1 Mar. 2026, ISSN: 17502071. DOI: 10.1057/s41261-025-00306-w
- [6] F. Soleymani, Q. Ma, and T. Liu, "Managing the risk via the chi-squared distribution in var and cvar with the use in generalized autoregressive conditional heteroskedasticity model," *Mathematics*, vol. 13, 9 May 2025, ISSN: 22277390. DOI: 10.3390/math13091410
- [7] M. Watorek, J. Kwapien, and S. Drożdż, "Financial return distributions: Past, present, and covid-19," *Entropy*, vol. 23, 7 Jul. 2021, ISSN: 10994300. DOI: 10.3390/e23070884
- [8] S. R. Luwang et al., "Regime discovery and intra-regime return dynamics in global equity markets," Jan. 2026. [Online]. Available: <http://arxiv.org/abs/2601.08571>
- [9] A. Gerunov, "Stock returns under different market regimes: An application of markov switching models to 24 european indices," Tech. Rep., 2023.
- [10] N. S. Putri, T. Takabika, and R. Gupta, "Risk timing and regime awareness in hybrid machine learning-based investment strategies," Tech. Rep., 2025.
- [11] A. Tampouris and C. Dritsaki, "Adaptive hierarchical hidden markov models for structural market change," *Journal of Risk and Financial Management*, vol. 19, p. 15, 1 Dec. 2025, ISSN: 19118074. DOI: 10.3390/jrfm19010015
- [12] O. F. Akbal, "Regime-switching factor models and nowcasting with big data, wp/24/190, september 2024," Tech. Rep., 2024.
- [13] Z. Tan and Y. Wu, *On regime switching models*, Apr. 2025. DOI: 10.3390/math13071128
- [14] P. Kelendji, "Regime-switching correlations with exogenous economic variables," Tech. Rep., 2023.
- [15] Y. Ding, D. Kambouroudis, and D. G. McMillan, "Forecasting realised volatility using regime-switching models," *International Review of Economics and Finance*, vol. 101, Jul. 2025, ISSN: 10590560. DOI: 10.1016/j.iref.2025.104171
- [16] X. Li, J. Chen, X. Tao, and Y. Ji, "Regime-switching asset allocation using a framework combining a jump model and model predictive control," *Mathematics*, vol. 13, 17 Sep. 2025, ISSN: 22277390. DOI: 10.3390/math13172837
- [17] Y. Shu and J. M. Mulvey, "Dynamic factor allocation leveraging regime-switching signals," Oct. 2024. [Online]. Available: <http://arxiv.org/abs/2410.14841>
- [18] A. Assaf, "The stochastic volatility model, regime switching and value-at-risk (var) in international equity markets," *Journal of Mathematical Finance*, vol. 07, pp. 491–512, 02 2017, ISSN: 2162-2434. DOI: 10.4236/jmf.2017.72026
- [19] J. Luo, T. Klein, Q. Ji, and C. Hou, "Forecasting realized volatility of agricultural commodity futures with infinite hidden markov har models," *International Journal of Forecasting*, vol. 38, pp. 51–73, 1 Jan. 2022, ISSN: 0169-2070. DOI: 10.1016/J.IJFORECAST.2019.

- 08.007 [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0169207019302316>
- [20] F. Ma, M. I. Wahab, D. Huang, and W. Xu, "Forecasting the realized volatility of the oil futures market: A regime switching approach," *Energy Economics*, vol. 67, pp. 136–145, Sep. 2017, ISSN: 0140-9883. DOI: 10.1016/J.ENERCO.2017.08.004 [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0140988317302608>
 - [21] L. R. Rabiner, "A tutorial on hidden markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 77 1989. DOI: 10.1109/5.18626
 - [22] R. L. Walter Zucchini Iain L. MacDonald, *Hidden Markov Models for Time Series*. Chapman and Hall, 2016. DOI: 10.1201/b20790
 - [23] P. Kupiec, "Techniques for verifying the accuracy of risk measurement models," *Journal of Derivatives*, 1995. DOI: 10.3905/jod.1995.407942
 - [24] P. Christoffersen, "Evaluating interval forecasts," *International Economic Review*, 1998. DOI: 10.2307/2527341