American Journal of Applied Statistics and Economics (AJASE)

ISSN: 2992-927X (ONLINE)

VOLUME 5 ISSUE 2 (2026)



PUBLISHED BY
E-PALLI PUBLISHERS, DELAWARE, USA

American Journal of Applied Statistics and Economics (AJASE)

Volume 5 Issue 2, Year 2026
ISSN: 2992-927X (Online)
DOI: https://doi.org/10.54536/ajase.v5i2.8742
https://journals.e-palli.com/home/index.php/ajase

Impact of Likelihood Approximation Choices on the Performance of Markov-Switching GARCH Models: Evidence from Thirty Global Financial Assets

Joshua Ogundairo[1*], Oluwadare O. Ojo[1]

Article Information

Received: July 19, 2026

Accepted: October 04, 2026

Published: November 06, 2026

Keywords

Gaussian Quadrature, Kurtosis, Likelihood Approximation, Markov-Switching GARCH, Particle Filter, Regime Switching, Volatility Modelling

ABSTRACT

Markov-Switching GARCH (MS-GARCH) models are widely used to capture changing volatility regimes in financial returns, but their exact likelihood is computationally intractable because it requires summing over an exponentially growing number of possible regime histories. Several approximation methods have been developed to solve this problem, yet there is little systematic evidence on how these methods compare when applied to real financial data with diverse distributional properties. This gap is important: choosing the wrong approximation method can lead to materially different parameter estimates, convergence failures, and unreliable risk assessments. This study addresses the gap by comparing four likelihood approximation methods for MS-GARCH(1,1) models across thirty financial assets from five market types: equities, bonds, cryptocurrencies, derivatives, and foreign exchange. The four methods are Gray's moment-matching collapsing procedure, the Plug-in approximation, five-point Gauss-Hermite Quadrature, and a Particle Filter using sequential importance resampling with 1,000 particles. Models are estimated under two specifications: a baseline (internal volatility dynamics only) and an exogenous (cross-market spillover conditioning) version, giving 220 experimental cases in total. Assets are grouped into four kurtosis categories - low ($K<10$), medium ($10\leq K<20$), high ($20\leq K<50$), and extreme ($K\geq 50$) - to examine how heavy-tailed distributions affect method performance. The Plug-in method produced the highest log-likelihood values across all kurtosis groups and market types, often exceeding Gray and Quadrature by more than 2,000 log-likelihood units for equity-index futures. Gray and Quadrature converged more reliably for cryptocurrency assets under exogenous conditioning. The Particle Filter converged most often but always gave the lowest log-likelihood values, falling two to four thousand units below the analytical methods for all asset classes and with parameters stuck near their starting values. The CBOE Volatility Index (^VIX) was the only asset for which all analytical methods failed to converge in every specification, marking a hard estimation limit for the adopted framework. Exogenous conditioning improved model fit for some assets but not all, with bonds and equity markets benefiting most. A practical, kurtosis-based method-selection guide is proposed for practitioners.

INTRODUCTION

Regime-switching models are widely used in financial econometrics to model the shifts between distinct market states seen in asset returns. Among these, the Markov-Switching Generalised Autoregressive Conditional Heteroskedasticity (MS-GARCH) model stands out. It combines the discrete regime-transition structure of Hidden Markov Models with the volatility clustering dynamics of GARCH processes (Haas *et al.*, 2004; Hamilton, 1989). The model is attractive because it captures two key features of financial returns at once: volatility persistence within each regime, and abrupt shifts between very different market states. Studies across many asset classes have shown that MS-GARCH models often outperform single-regime alternatives in volatility forecasting, risk measurement, and regime identification (Ardia *et al.*, 2018; Guidolin & Timmermann, 2007; Klaassen, 2002).

Despite these advantages, estimating MS-GARCH models faces a serious computational problem that has not received enough systematic attention in the methodological literature. The problem comes from the path-dependent nature of the GARCH variance recursion combined with unobserved regime transitions. The conditional variance at any time depends on past variances, which in turn depend on the sequence of regimes that prevailed when they were formed. Exact likelihood evaluation therefore requires summing over all 2^T possible regime histories for a sample of length T. Even for modest sample sizes, this number is far beyond any practical computing capacity (Bauwens *et al.*, 2010; Gray, 1996; Haas *et al.*, 2004). Researchers have developed several approximation strategies to make estimation feasible: Gray's moment-matching collapsing procedure (Gray, 1996), plug-in approximations that substitute filtered probability estimates for regime uncertainty (Durbin and Koopman, 2012; Klaassen, 2002), quadrature-based numerical integration methods (Tanizaki, 1996), and particle filtering via sequential Monte Carlo techniques (Doucet & Johansen, 2009; Pitt & Shephard, 1999).

Each method introduces a different type of error. Gray's procedure and quadrature methods have deterministic errors that decrease with finer discretisation but at higher computational cost (Augustyniak, 2014; Haas *et al.*, 2004). Plug-in methods treat uncertain regime assignments as

[1] Department of Statistics, Federal University of Technology, Akure (FUTA), Nigeria
[*] Corresponding author's e-mail: ogundairosta154470@futa.edu.ng

known quantities, which can introduce bias by ignoring regime uncertainty (Durbin & Koopman, 2012). Particle filters introduce random variability that depends on the number of particles and the resampling scheme, and they can suffer from particle degeneracy and impoverishment (Doucet *et al.*, 2001; Pitt & Shephard, 1999). Despite these differences, the MS-GARCH literature provides very little guidance on how to choose among these methods. Most studies simply use the most convenient method — often Gray's procedure by default — without checking whether the results would hold under a different approach (Ardia *et al.*, 2019; Augustyniak, 2014; Catania *et al.*, 2019).

This lack of comparative evidence has practical consequences. When different research teams use different methods on similar questions, conflicting findings can emerge that may reflect computational differences rather than genuine economic ones. Financial institutions using MS-GARCH models for risk management, Value-at-Risk estimation, or portfolio decisions also face uncertainty about whether their choice of approximation method significantly affects their results (Ardia *et al.*, 2018; Guidolin & Timmermann, 2007). The few existing comparative studies mostly use simulated data from known processes. The relative performance on real financial data — with heavy tails, asymmetric responses, and unknown true regimes — remains largely unexplored (Augustyniak, 2014). This paper addresses these gaps with a controlled empirical study that keeps the MS-GARCH(1,1) model specification fixed and varies only the approximation method across thirty financial assets from five market types: equities, bonds, cryptocurrencies, derivatives, and foreign exchange. Both baseline and exogenous specifications are estimated for each asset, giving 220 experimental cases. Assets are further grouped into four kurtosis categories to assess whether tail behaviour affects approximation performance. The findings provide the first large-scale, multi-market comparison of likelihood approximation methods for MS-GARCH models on real financial data, offering evidence-based guidance for method selection. The rest of the paper is organised as follows: Section 2 details the methodology; Section 3 presents the market-level findings; Section 4 develops the kurtosis-based analysis; and Section 5 concludes.

MATERIALS AND METHODS

Theoretical Background

The study uses a two-regime MS-GARCH(1,1) model. Observed returns are generated by a latent Markov chain $\{S_t\}^T_{t=1}$, $S_t \in \{1,2\}$, evolving with constant transition probabilities:

$$p_{jk} = Pr(S_t = k \mid S_{t-1} = j), j, k \in \{1,2\}, \quad (1)$$

collected in the transition matrix $P=(p_{jk})$ with $\sum_k p_{jk} = 1$. Conditional on the regime, returns follow:

$$y_{i,t} = \mu_{i,S_t} + \sigma_{i,t}(S_t)\varepsilon_{i,t}, \varepsilon_{i,t} \sim \mathcal{N}(0,1), \quad (2)$$

with regime-dependent conditional mean $\mu_{(i,k)}=c_{(i,k)}+\gamma_{(i,k)} x_t$ (exogenous specification) or $\mu_{(i,k)}=c_{(i,k)}$ (baseline), where x_t is the log-return of the market pivot asset. The conditional variance follows a GARCH(1,1) process within each regime: where $\omega_{(i,k)}>0$, $\alpha_{(i,k)}\geq 0$, $\beta_{(i,k)}\geq 0$, and

$$\sigma^2_{i,t}(k) = \omega_{i,k} + \alpha_{i,k} y^2_{i,t-1} + \beta_{i,k}\sigma^2_{i,t-1}, \quad (3)$$

$\alpha_{(i,k)}+\beta_{(i,k)}<1$ to ensure within-regime stationarity. The full parameter vector is $\Psi_i=\{c_{(i,1)},c_{(i,2)},\gamma_{(i,1)},\gamma_{(i,2)},\omega_{(i,1)},\omega_{(i,2)},\alpha_{(i,1)},\alpha_{(i,2)},\beta_{(i,1)},\beta_{(i,2)},p_{11},p_{22}\}$, comprising twelve parameters. Exact maximum likelihood requires integrating over all 2^T regime sequences through the recursion: which is computationally intractable because of the path-dependence of the GARCH variance recursion.

$$L(\Psi_i) = \Sigma^2_{S_1=1} \cdots \sum^2_{S_T=1} \pi_{S_1} \prod^T_{t=2} p_{S_{t-1},S_t} \prod^T_{t=1} f(y_{i,t} \mid S_t, \mathcal{F}_{t-1}; \Psi_i), \quad (4)$$

Gray's Collapsing Procedure handles path-dependence by collapsing the growing variance distribution to two representative values at each step. At time t, the approximate conditional variance for regime k is (Gray, 1996):

$$\tilde{\sigma}^2_{i,t}(k) = \omega_{i,k} + \alpha_{i,k} y^2_{i,t-1} + \beta_{i,k}\bar{\sigma}^2_{i,t-1}(k), \quad (5)$$

where $\sigma^{-2}_{i,t-1}(k)=\sum_j \Pr(S_{t-1}=j \mid S_t=k)\sigma^{-2}_{i,t-1}(j)$ is the probability-weighted average from the previous period.

Gaussian Quadrature uses N quadrature nodes $\{v_n\}$ and weights $\{w_n\}$ to approximate the variance distribution. Here a five-point Gauss-Hermite scheme (N=5) is used. The algorithm tracks N×2 distinct variance paths and computes likelihood contributions by summing over them (Monfort *et al.*, 2015; Tanizaki, 1996).

Plug-In Approximation replaces integration over regime uncertainty with point estimates based on filtered probabilities. The plug-in variance is (Klaassen, 2002):

$$\hat{\sigma}^2_{i,t} = \sum^2_{k=1} Pr(S_t = k \mid \mathcal{F}_{t-1})\sigma^2_{i,t}(k), \quad (6)$$

with $\sigma^2_{i,t}(k)=\omega_{(i,k)}+\alpha_{(i,k)} y^2_{(i,t-1)}+\beta_{(i,k)} \sigma^2_{(i,t-1)}$.

The Particle Filter represents the variance path distribution using N_p=1,000 weighted random samples (particles). These are propagated forward via sequential importance resampling (SIR) (Doucet and Johansen, 2009; Pitt and Shephard, 1999). Each particle n carries a regime sequence and a corresponding variance path:

$$\sigma^{2(n)}_{i,t} = \omega_{i,S^{(n)}_t} + \alpha_{i,S^{(n)}_t} y^2_{i,t-1} + \beta_{i,S^{(n)}_t}\sigma^{2(n)}_{i,t-1}. \quad (7)$$

A fixed random seed ensures reproducibility across all 220 runs.

Data and Sources

Daily return data for thirty financial assets are downloaded

using the yfinance Python library, covering September 2015 through August 2025. This ten-year period includes multiple market cycles, the COVID-19 shock, and the subsequent inflationary episode. The data cover five market types, each with five target assets and one pivot asset used as the exogenous conditioning variable:

Equities: FTSE, N225, IXIC, EEM, VTV; pivot = GSPC

Bonds: IEI, TLT, BNDX, LQD, HYG; pivot = TNX

Cryptocurrencies: ETH-USD, LTC-USD, XRP-USD, DOGE-USD, NMC-USD; pivot = BTC-USD

Derivatives: ES=F, NQ=F, VIX, GC=F, CL=F; pivot = BTC=F

Forex: GBPUSD=X, JPY=X, AUDUSD=X, CHF=X, EURGBP=X; pivot = EURUSD=X

All price series are transformed to continuously compounded percentage returns: $y_{(i,t)}=100\times\log(P_{(i,t)}/P_{(i,t-1)})$. All series pass the Augmented Dickey-Fuller stationarity test and Engle's ARCH-LM test for conditional heteroskedasticity ($p<0.001$ throughout), confirming that GARCH-type modelling is appropriate. Table 1 summarises the distributional properties of all thirty assets.

Experimental Setup and Specifications

Each of the thirty assets is estimated with two model specifications and four approximation methods, giving 220 experimental cases in total. The baseline specification ($\gamma_{(i,k)}=0$) models internal regime-dependent volatility without any external input. The exogenous specification includes the market pivot ($\gamma_{(i,k)}\neq0$) to capture cross-market spillovers. Pivot assets are estimated under the baseline specification only. Estimation uses the L-BFGS-B algorithm via scipy.optimize with box constraints on parameter values ($\omega>0$, $\alpha+\beta<1$, $0<p_{ii}<1$). The estimation engine is implemented in C++ and interfaced with Python via Cython for speed. A penalty mechanism redirects the optimiser from infeasible parameter regions.

The kurtosis-based grouping of assets came from the preliminary distributional analysis. Assets with excess kurtosis $K<10$ (low kurtosis) include stable Forex pairs and short-duration bond instruments. Medium-kurtosis assets ($10\leq K<20$) comprise equity indices, several cryptocurrencies, and the VIX. High-kurtosis assets ($20\leq K<50$) include equity ETFs, corporate bonds, energy derivatives, and the 10-year Treasury yield. Extreme-kurtosis assets ($K\geq50$) comprise DOGE-USD ($K=98.51$) and EURGBP=X ($K=248.50$).

Evaluation Metrics

Approximation performance is measured using four metrics: (i) Log-Likelihood (LogLik) - the primary fitting criterion; (ii) Akaike Information Criterion (AIC =-2"LogLik"+2k, k=12) - penalised fit for model parsimony; (iii) Convergence status (a binary indicator, TRUE when the gradient norm meets the L-BFGS-B stopping criterion) - a measure of numerical stability; and (iv) Regime persistence (p_{11}, p_{22}) - the stability of the identified low- and high-volatility regimes. A method is considered dominant if it achieves the highest LogLik for a given asset and specification, regardless of convergence status. Convergence-adjusted dominance requires the highest LogLik among converged solutions.

Table 1: Descriptive Statistics of Daily Log-Returns (%)

Asset	Market	Mean	StdDev	Skew	Excess Kurt.	Kurt. Class	ADF (p)	ARCH-LM (p)
Bond Market								
IEI	Bond	0.0044	0.2205	0.3795	5.68	Low	0.0000	5.53e-45
TLT	Bond	−0.0071	0.8337	0.0705	8.12	Low	2.17e-20	1.19e-139
BNDX	Bond	0.0045	0.2273	−0.6025	11.68	Medium	0.0000	1.23e-96
LQD	Bond	0.0062	0.4979	0.3778	36.95	High	9.68e-24	6.18e-190
HYG	Bond	0.0128	0.4695	−0.0644	37.30	High	1.69e-20	1.29e-117
TNX	Bond	0.0251	2.7052	0.3490	47.73	High	1.46e-14	1.69e-249
Cryptocurrency								
ETH-USD	Crypto	0.0625	4.5557	−0.8818	10.91	Medium	8.55e-29	1.45e-13
LTC-USD	Crypto	−0.0404	4.8588	−0.5167	7.94	Low	0.0000	1.81e-18
XRP-USD	Crypto	0.0470	5.5694	0.6839	17.33	Medium	0.0000	2.69e-20
DOGE-USD	Crypto	0.1266	6.8135	4.7529	98.51	Extreme	6.99e-16	4.04e-10
NMC-USD	Crypto	−0.0455	7.8798	−0.9870	29.31	High	6.97e-19	6.71e-38
BTC-USD	Crypto	0.0635	3.5104	−0.9417	13.89	Medium	0.0000	1.95e-10
Derivatives								
ES=F	Derivative	0.0329	1.0404	−0.7946	20.95	High	3.68e-20	9.67e-138
NQ=F	Derivative	0.0470	1.2778	−0.3442	10.58	Medium	1.39e-17	7.92e-80
VIX	Derivative	0.0108	6.7940	1.7173	14.28	Medium	0.0000	2.13e-16
GC=F	Derivative	0.0363	0.8025	−0.2476	6.58	Low	0.0000	3.26e-33

CL=F	Derivative	0.0254	2.4723	0.0479	39.31	High	6.76e-17	2.33e-95
BTC=F	Derivative	0.0628	3.5341	−0.2230	8.21	Low	0.0000	3.20e-10
Equities								
FTSE	Equity	0.0080	0.8322	−1.3901	25.19	High	2.12e-25	5.11e-121
N225	Equity	0.0233	1.0739	−0.5317	17.57	Medium	0.0000	8.42e-107
IXIC	Equity	0.0415	1.2542	−0.4579	12.87	Medium	1.29e-25	1.25e-92
GSPC	Equity	0.0329	1.0377	−0.7358	22.64	High	6.07e-20	8.80e-138
EEM	Equity	0.0103	1.1082	−0.9329	16.92	Medium	1.31e-28	1.86e-110
VTV	Equity	0.0279	0.9566	−0.9430	27.39	High	3.09e-20	1.95e-182
Forex								
GBPUSD=X	Forex	0.0006	0.4688	−0.2666	8.04	Low	3.94e-30	5.42e-41
JPY=X	Forex	0.0095	0.4643	−0.4158	6.56	Low	0.0000	1.27e-20
AUDUSD=X	Forex	−0.0056	0.5424	−0.3614	6.83	Low	0.0000	3.88e-22
CHF=X	Forex	−0.0075	0.3975	−0.7058	8.14	Low	0.0000	7.51e-22
EURGBP=X	Forex	−0.0009	0.4546	0.2235	248.50	Extreme	0.0000	1.48e-240
EURUSD=X	Forex	−0.0002	0.3865	−0.0128	4.49	Low	0.0000	6.92e-19

Note: Mean and StdDev are in percentage (%). Excess Kurtosis is relative to the normal distribution (K=0). All ADF and ARCH-LM p-values are <0.001, confirming stationarity and ARCH effects throughout.

RESULTS AND DISCUSSION

Figure 1 shows the daily return profiles of all thirty assets across the five market types. Return volatility varies widely - from the calm fluctuations of short-duration bond ETFs to the extreme, spike-dominated dynamics of cryptocurrencies and the VIX - which makes this a genuinely diverse test environment for the approximation methods. Figure 2 summarises the comparative log-likelihood performance across all methods, giving a visual overview of the results described below. Full parameter estimates for all 220 models are provided in a companion reference document; see the Appendix for repository details.

Cryptocurrency Market. The cryptocurrency market

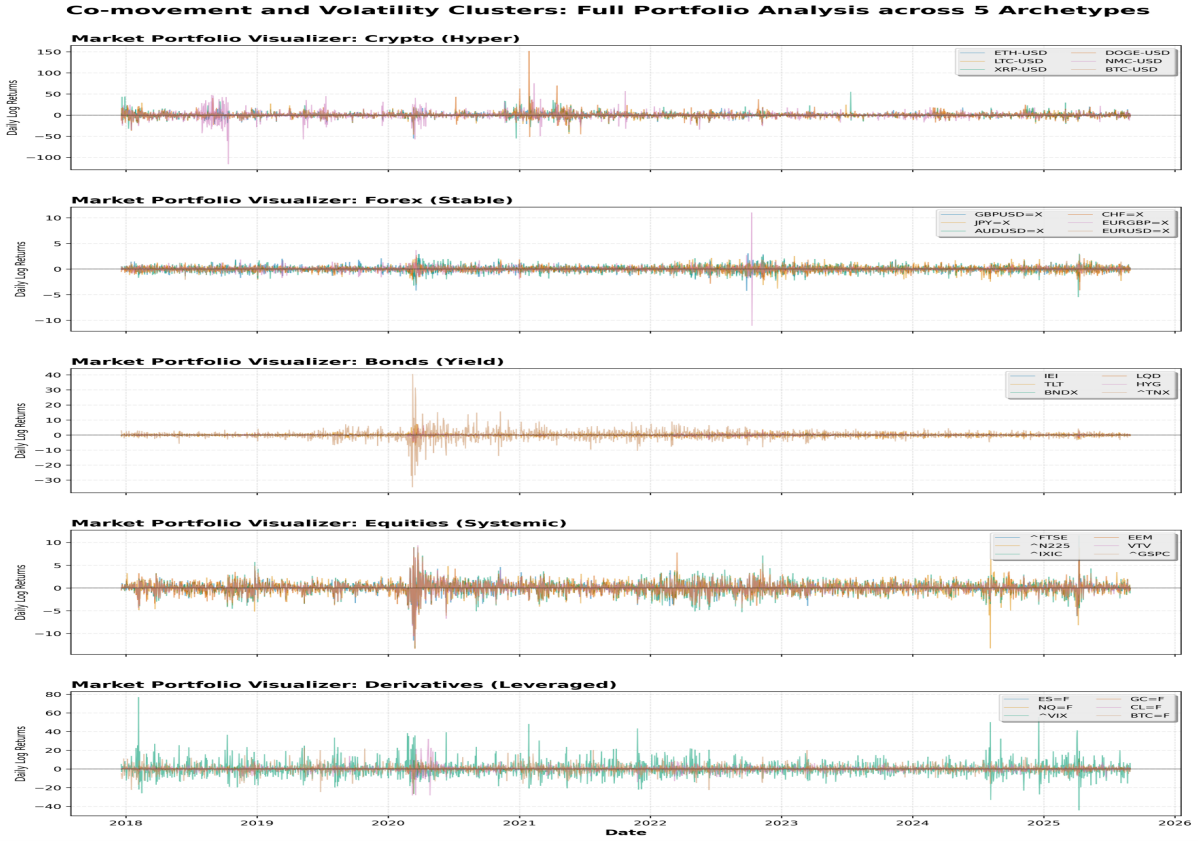


Figure 1: Daily Log-Return Series for All Thirty Assets, Stratified by Market Archetype (Sep 2015 – Aug 2025). Shaded bands indicate the COVID-19 shock period (Mar–Apr 2020).

was the most challenging estimation environment, as shown by the extreme excess kurtosis values in Table 1. In the baseline specification, Plug-in achieved the best log-likelihood for four of the six assets — ETH-USD (-8,073.57), LTC-USD (-8,269.21), DOGE-USD (-8,613.45), and BTC-USD (-7,232.32) — while Quadrature delivered the best fit for XRP-USD (-8,514.52). For NMC-USD, Quadrature (-9,121.91) marginally outperformed Gray (-9,126.19), with Plug-in inferior at -9,326.54. Plug-in, however, often failed the formal convergence criterion even though it reached higher likelihood values — a pattern consistent with the rough likelihood surfaces of high-kurtosis return series. Gray's method converged for BTC-USD and ETH-USD but achieved consistently lower log-likelihoods than Plug-in. The Particle Filter converged for most assets but was always the weakest estimator, with log-likelihoods ranging from -7,854 (BTC-USD) to -10,526 (NMC-USD) — well below the analytical methods. Its parameter estimates were trapped near the starting values ($\omega \approx 0.30$, $\alpha \approx 0.05$, $\beta \approx 0.85$, $p_{11} \approx 0.95$, $p_{22} \approx 0.80$). In the exogenous specification conditioned on BTC-USD, Gray and Quadrature were the most reliable methods. Both achieved the best converged log-likelihoods for ETH-USD (-8,072.60 and -8,072.25, respectively) and LTC-USD (-8,292.09 and -8,291.69). Plug-in gave lower values and failed convergence more often in the exogenous configuration. The Particle Filter gave nearly identical results regardless of the pivot specification, continuing to produce persistence estimates close to the starting values. High persistence values (>0.90) across most crypto assets confirmed the well-known volatility persistence of digital asset returns. Exogenous conditioning with BTC-USD produced modest rather than large changes in regime dynamics, which suggests that intra-market spillovers are not the main driver of volatility clustering in cryptocurrencies.

Foreign Exchange (Forex) Market. The Forex market showed a consistent Plug-in advantage in log-likelihood across both specifications. In the baseline, Plug-in achieved the highest log-likelihood for every Forex asset: GBPUSD=X (63.64), JPY=X (71.36), AUDUSD=X (-376.14), CHF=X (372.12), EURGBP=X (557.11), and EURUSD=X (399.99). Gray ranked second in most cases, with Quadrature third. The Particle Filter showed the most extreme underperformance of any market type, with log-likelihoods from approximately -1,879 to -2,384 — two to four thousand units below the analytical methods — despite converging for most assets. Moving from baseline to exogenous specifications conditioned on EURUSD=X produced negligible changes in log-likelihood for all currency pairs — shifts of less than two units (e.g., CHF=X Gray: 302.30→303.39; Plug-in: 372.12→372.35). This confirms that Forex volatility regimes are mainly driven by individual currency dynamics rather than cross-currency spillovers from the Euro-Dollar benchmark. Convergence was generally stable for the analytical methods in both specifications. The exception was JPY=X, where Gray, Plug-in, and Quadrature all failed to converge in the exogenous configuration, suggesting that the EURUSD=X factor slightly destabilised the Yen-Dollar likelihood surface. The analytical methods identified two regimes: a highly persistent primary regime ($p_{ii} > 0.95$) and a near-transient secondary regime ($p_{ii} \approx 0.0002$). This pattern reflects the generally low-volatility environment of major developed-market currency pairs over the study period.

Equity Market. Equity results showed a consistent and large advantage for the Plug-in method in log-likelihood. In the baseline specification, Plug-in achieved the best log-likelihood for FTSE (-753.01), N225 (-1,183.10), IXIC (-1,647.97), EEM (-1,451.69), and VTV (-835.80). The Plug-in advantage was especially large for IXIC and EEM. For IXIC, Gray reached only -4,070.86 while Plug-in achieved -1,647.97 — a gap of over 2,400 log-likelihood units. For EEM, Gray reached -3,672.50 (non-converged) against Plug-in's -1,451.69. The GSPC pivot showed a different pattern. Gray was the only converged analytical method in the baseline (-3,333.75), while Plug-in (-1,054.82) and Quadrature (-1,414.35) reached better likelihoods but did not formally converge. In the exogenous configuration conditioned on GSPC, Plug-in remained dominant, with improvements for most assets: FTSE improved from -753.01 to -726.36; N225 from -1,183.10 to -958.26. Gray's method was hurt by the introduction of the exogenous pivot. For N225, Gray fell from -1,523.48 (converged) in the baseline to -3,164.01 (non-converged) in the exogenous model. This suggests that the extra gamma parameters interacted poorly with the moment-matching procedure for several equity assets. The Particle Filter was the weakest method for all equity assets in both specifications, with log-likelihoods two to four thousand units below Plug-in and parameters trapped near their starting values. The analytical methods identified a two-state regime structure for equity returns: a high-persistence low-volatility regime and a near-transient high-volatility state. The allocation of persistence between the two states varied by method and asset.

Bond Market. The bond market results showed Plug-in superiority for most assets and also revealed a unique estimation difficulty posed by the 10-year Treasury yield (TNX). In the baseline specification, Plug-in achieved the best log-likelihood for IEI (2,031.30 vs. Gray's 1,896.91), BNDX (1,992.60 vs. Gray's 1,854.41), LQD (482.71, non-converged, vs. Gray's 310.69), HYG (697.38 vs. Quadrature's 579.36), and TLT (-934.93 vs. Gray's -1,173.63). TNX was the most problematic bond instrument. Gray and Plug-in both failed to converge in the baseline, and only Quadrature converged at a log-likelihood of -5,708.41. This shows that standard moment-matching and probability-weighting approaches have difficulty capturing the non-standard volatility of short-rate instruments. The Particle Filter converged for all bond assets but gave log-likelihoods thousands of units below the analytical alternatives (-1,455 to -5,822), reflecting parameter trapping. In the exogenous

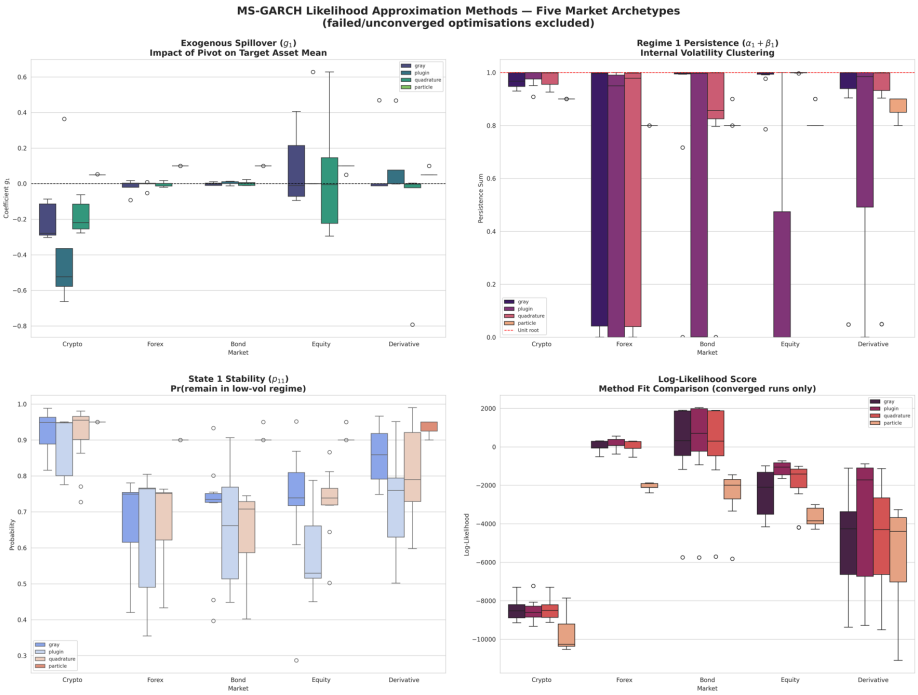


Figure 2: Comparative Log-Likelihood Performance of the Four Approximation Methods Across Market Archetypes. Higher values indicate superior model fit. Particle Filter results are plotted on a secondary scale for readability.

configuration conditioned on TNX, Plug-in retained the best log-likelihood for TLT (-926.61), BNDX (1,965.75), LQD (467.55), and HYG (745.60, non-converged). The TNX pivot produced small changes in log-likelihood for most assets — for example, IEI Gray moved from 1,896.91 to 1,891.71 — confirming that the Treasury yield factor does not substantially change bond volatility dynamics beyond what the internal GARCH parameters already capture. The bond market showed a distinctive regime structure: most assets had near-zero persistence in one state ($p_{ii} \approx 0.0002$) and near-unit persistence in the other ($p_{ii} > 0.99$), consistent with a persistent low-volatility regime interrupted by brief high-volatility episodes. HYG was an exception, with non-trivial second-state persistence (0.31 under the Gray exogenous model), which reflects the greater credit-cycle sensitivity of high-yield instruments compared with investment-grade alternatives.

Derivatives Market. The derivatives market showed the largest divergence in method performance among all five market types. In the baseline, Plug-in achieved the best log-likelihood for ES=F (-1,095.35 vs. Gray's -3,363.57, a gap of over 2,200 units), NQ=F (-1,708.53 vs. Gray's -4,168.20, a gap exceeding 2,400 units), GC=F (-884.73 vs. Gray's -1,105.49), and CL=F (-5,767.89, non-converged, marginally superior to Quadrature's -5,819.91). For BTC=F, Gray and Quadrature converged to nearly identical results (-7,441.63 and -7,441.91) and outperformed Plug-in (-7,692.15, non-converged). This suggests that Bitcoin futures behave more like cryptocurrency assets, where Gray-type approximations are better suited. The VIX was the most problematic case in the entire 220-case study. All three analytical methods — Gray, Plug-in, and Quadrature — failed to converge in both baseline and exogenous configurations, with log-likelihoods ranging from -9,248 to -9,503. The Particle Filter was the only method to formally converge for VIX in all specifications, but its log-likelihood (-11,097 to -11,098) was substantially worse than the non-converged analytical solutions. This confirms that the VIX's extreme asymmetric volatility is beyond the reach of all current MS-GARCH(1,1) approximation strategies. In the exogenous configuration conditioned on BTC=F, Plug-in retained its performance advantage for ES=F (-1,108.91) and NQ=F (-1,717.33). A notable result appeared for CL=F: the Plug-in exogenous model reached -3,123.80 (non-converged) — far better than Gray or Quadrature (both near -5,823) — suggesting a structurally different regime decomposition when Bitcoin futures are included as a conditioning variable for crude oil volatility. The VIX remained intractable for all analytical methods in the exogenous specification, further confirming the numerical limits of deterministic approximation methods for this fear-gauge instrument.

Discussion

The results are now examined through the lens of excess kurtosis, which is the main observable indicator of tail risk and likelihood surface complexity. This kurtosis-based grouping organises all 30 assets regardless of market type, and provides a data-driven guide for method selection that addresses the gap identified in the introduction (Ardia *et al.*, 2018; Augustyniak, 2014).

Low Kurtosis ($K<10$). For assets with low excess kurtosis — mainly Forex pairs (EURUSD=X: K=4.49; JPY=X: K=6.56; AUDUSD=X: K=6.83; GBPUSD=X: K=8.04; CHF=X: K=8.14), short-duration bond instruments (IEI: K=5.68; TLT: K=8.12), gold futures (GC=F:

K=6.58), LTC-USD (K=7.94), and BTC=F (K=8.21) — the likelihood surface is relatively well-behaved. The Plug-in method consistently achieved the best log-likelihood in both baseline and exogenous specifications. Gray and Quadrature were competitive and converged reliably for these assets, but were consistently outperformed by Plug-in in log-likelihood (Gray, 1996; Klaassen, 2002). The Particle Filter was numerically stable but gave inferior fit by margins of two to four thousand log-likelihood units in all cases. This confirms that stochastic simulation is unnecessarily costly and imprecise for assets with well-behaved distributions (Ardia *et al.*, 2019). The near-zero improvement from exogenous conditioning for Forex assets — shown by changes of less than two log-likelihood units when EURUSD=X was added — confirmed that major currency market volatility is mainly driven by internal dynamics over the study period (Bollen *et al.*, 2000). For practitioners, Plug-in is the clear first choice for low-kurtosis assets. Gray is a reliable fallback for convergence. The Particle Filter is not advisable.

Medium Kurtosis (10≤K<20). The medium-kurtosis group includes equity indices (IXIC: K=12.87; N225: K=17.57; EEM: K=16.92), bond ETFs (BNDX: K=11.68), cryptocurrencies (ETH-USD: K=10.91; XRP-USD: K=17.33; BTC-USD: K=13.89), and derivatives (NQ=F: K=10.58; VIX: K=14.28). Plug-in retained its advantage for most assets in this range, particularly for equity assets (FTSE, N225, IXIC, EEM) and most derivative instruments. The clearest departure from the low-kurtosis pattern appeared for cryptocurrency assets under exogenous conditioning. For ETH-USD and LTC-USD conditioned on BTC-USD, Gray and Quadrature jointly achieved the best converged log-likelihoods while Plug-in gave lower values and failed convergence more often. This suggests that the extra gamma parameters interfered with Plug-in's probability-weighting in the crypto likelihood environment (Ardia *et al.*, 2018; Augustyniak, 2014). Quadrature showed its clearest comparative advantage for exogenous cryptocurrency models in the medium kurtosis range. The VIX is classified as medium kurtosis (K=14.28) but has uniquely asymmetric, spike-dominated volatility. It was the only asset for which all three analytical methods failed to converge in every specification (Catania and Grassi, 2017). The Particle Filter converged for VIX but at a log-likelihood well below the non-converged analytical solutions, showing that convergence alone does not mean adequate model fit. For practitioners in the medium-kurtosis range, Plug-in is the preferred baseline estimator, with Quadrature as the preferred choice for exogenous cryptocurrency specifications. The VIX marks a known estimation boundary for the standard 12-parameter MS-GARCH(1,1) framework.

High Kurtosis (20≤K<50). High-kurtosis assets include equity ETFs (VTV: K=27.39; FTSE: K=25.19; GSPC: K=22.64), corporate bond ETFs (LQD: K=36.95; HYG: K=37.30), derivatives (ES=F: K=20.95; CL=F: K=39.31; TNX: K=47.73), and NMC-USD (K=29.31). Plug-in continued to lead in log-likelihood for equity assets and most bond instruments, but convergence became markedly more variable at these kurtosis levels. EEM Plug-in converged in the baseline while Gray failed; N225 Gray converged in the baseline but fell sharply in the exogenous model; and Quadrature failed to converge for NQ=F and IXIC. The large Plug-in advantages for ES=F and NQ=F — exceeding 2,200 and 2,400 log-likelihood units over Gray and Quadrature, respectively — were the largest method-performance gaps in the entire study. This shows that the multi-modal likelihood surfaces of high-kurtosis financial futures are particularly difficult for moment-matching procedures. For BTC=F and TNX, Gray and Quadrature proved more reliable than Plug-in in achieving convergent solutions with competitive likelihoods. This suggests that for certain high-kurtosis instruments with structural properties different from equity-index futures, moment-matching methods remain useful (Haas *et al.*, 2004). The Particle Filter remained the weakest performer in log-likelihood across all high-kurtosis assets, though it converged broadly. For practitioners in the high-kurtosis range, Plug-in should be tried first with careful convergence checking. Gray or Quadrature should be used as fallbacks when convergence fails. The Particle Filter is appropriate only as a last resort when all analytical methods fail.

Extreme Kurtosis (K≥50). Only two assets fall into the extreme-kurtosis category: DOGE-USD (K=98.51) and EURGBP=X (K=248.50). Despite what the literature might suggest — that simulation-based methods should become preferable at very high kurtosis levels — the Particle Filter did not prove to be the better method for either asset. For DOGE-USD, Plug-in achieved the best baseline log-likelihood (-8,613.45, non-converged) while the Particle Filter converged to a much lower -10,260.15 — a deficit of over 1,600 units. For EURGBP=X, Plug-in fully converged to a log-likelihood of 557.11 while the Particle Filter converged to -1,878.74 — a gap of nearly 2,500 units — the largest absolute gap between an analytical and simulation-based method in the entire study. These findings show that even at kurtosis levels between 50 and 248, the Particle Filter is only a last resort when all analytical methods fail completely — it is not the preferred estimator for heavy-tailed data (Doucet *et al.*, 2001; Pitt and Shephard, 1999). For practitioners working with extreme-kurtosis assets, Plug-in should still be tried first. The Particle Filter should be used only when all analytical methods fail entirely, and expectations about parameter estimate quality should be adjusted accordingly.

Summary Framework. Table 2 summarises the kurtosis-based method-selection recommendations from the empirical analysis. The main finding is that Plug-in is the dominant log-likelihood optimiser across all four kurtosis groups, with its advantage largest for high-kurtosis equity-index futures. The Particle Filter is consistently the weakest likelihood optimiser and should be used only when all analytical alternatives fail completely. Quadrature shows its clearest relative advantage for medium-kurtosis

cryptocurrency assets in the exogenous specification. Exogenous conditioning improves model fit for some markets but not all — most consistently for bonds and equities — and should not be assumed universally beneficial without prior empirical testing.

Table 2: Kurtosis-Based Method-Selection Framework for MS-GARCH Estimation

Kurtosis Class	Preferred Method	Fallback / Notes
Low ($K < 10$)	Plug-in (baseline & exog.)	Gray for convergence guarantee; Particle Filter inadvisable
Medium ($10 \leq K < 20$)	Plug-in (baseline); Quadrature (exog. crypto)	Gray for convergence; VIX intractable for all analytical methods
High ($20 \leq K < 50$)	Plug-in (verify convergence)	Gray / Quadrature as fallback; Particle Filter last resort
Extreme ($K \geq 50$)	Plug-in (attempt first)	Particle Filter only if all analytical methods yield no solution

CONCLUSION

This study conducted the first large-scale, controlled comparison of four MS-GARCH likelihood approximation methods — Gray's collapsing procedure, Plug-in, Gaussian Quadrature, and Particle Filtering — across thirty financial assets from five market types and 220 experimental cases. The central finding is clear: the choice of likelihood approximation method significantly affects estimation outcomes. This choice should be guided by the observable statistical properties of the data, not by software defaults or disciplinary convention.

Plug-in emerged as the dominant method across all kurtosis groups and all five market types in terms of log-likelihood. For low-kurtosis assets, Plug-in consistently outperformed all alternatives. For high-kurtosis equity-index futures (ES=F, NQ=F), the Plug-in advantage over Gray and Quadrature exceeded 2,000 to 2,400 log-likelihood units — a difference with significant practical implications for model comparison, inference, and risk management. The Particle Filter was, without exception, the weakest log-likelihood optimiser across all kurtosis groups and market types, consistently falling two to four thousand units below the alternatives and showing parameter degeneracy near starting values. The Particle Filter provided value only as a convergence fallback for the VIX and for cases where all analytical methods failed entirely. Quadrature's clearest relative advantage was for cryptocurrency assets in exogenous specifications within the medium-kurtosis range, where its better convergence stability made it the preferred choice among converged solutions.

Exogenous conditioning improved model fit for some markets but not all. Bond market assets benefited most consistently from the Treasury yield pivot. Forex assets showed negligible improvement from the EURUSD=X factor, confirming that major currency volatility is mainly idiosyncratic. The VIX presents a fundamental estimation challenge. No analytical MS-GARCH approximation method converged for this instrument under the 12-parameter specification. This points to the need for extended model frameworks — potentially asymmetric or three-regime specifications — to handle the non-standard dynamics of fear-gauge instruments.

Practitioners should follow the kurtosis-based framework in Table 2: compute excess kurtosis before choosing an approximation method; use Plug-in as the default for most assets with Gray or Quadrature as convergence fallbacks; avoid the Particle Filter except as a last resort; and test exogenous conditioning variables empirically before including them in the model. Researchers should report convergence status alongside parameter estimates, document the approximation method used, and perform at least a two-method robustness comparison when publishing MS-GARCH results.

REFERENCES

Ardia, D., Bluteau, K., Boudt, K., and Catania, L. (2018). Forecasting risk with Markov-switching GARCH models: A large-scale performance study. *International Journal of Forecasting, 34*(4), 733–747.

Ardia, D., Bluteau, K., and Rüede, M. (2019). Regime changes in Bitcoin GARCH volatility dynamics. *Finance Research Letters, 29*,266–271.

Augustyniak, M. (2014). Maximum likelihood estimation of the Markov-switching GARCH model. *Computational Statistics & Data Analysis, 76*,61–75.

Bauwens, L., Preminger, A., and Rombouts, J. V. K. (2010). Theory and inference for a Markov switching GARCH model. *The Econometrics Journal, 13*(2), 218–244.

Bollen, N. P. B., Gray, S. F., and Whaley, R. E. (2000). Regime switching in foreign exchange rates: Evidence from currency option prices. *Journal of Econometrics, 94*(1–2), 239–276.

Catania, L. and Grassi, S. (2017). Dynamic adaptive mixture models for precious metal returns. *Journal of Empirical Finance, 44*, 212–227.

Catania, L., Grassi, S., and Ravazzolo, F. (2019). Forecasting cryptocurrencies under model and parameter instability. *International Journal of Forecasting, 35*(2), 485–501.

Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice. Springer.*

Doucet, A. and Johansen, A. M. (2009). A tutorial on

particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering, 12*(3), 656–704.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford University Press, 2 edition.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics, 42*(1), 27–62.

Guidolin, M. and Timmermann, A. (2007). Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control, 31*(11), 3503–3544.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics, 2*(4), 493–530.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica, 57*(2), 357–384.

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics, 27*(2), 363–394.

Monfort, A., Renne, J.-P., and Roussellet, G. (2015). A quadratic Kalman filter. *Journal of Econometrics, 187*(1), 43–56.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association, 94*(446), 590–599.

Tanizaki, H. (1996). *Nonlinear Filters: Estimation and Applications*. Springer, 2 edition.

Appendix: Repository and Supplementary Materials

The complete project repository — containing the full model parameter reference document (ms_garch_full_models.pdf, 112 pages), the estimation notebook (code/Likelihood_Approximation_for_MS_GARCH.ipynb), all CSV results for the 220 experimental cases, and the result visualisations — is publicly available at:
https://github.com/oluwadamilarejoshua/ms-garch-likelihood-approximation