Qeios

Peer Review

Review of: "Unpacking the Complexities of Bitcoin Volatility: A Time Series Data with Long-term Memory or Long-range Dependence"

Arati Kamat[1]

1. Independent researcher

Conflict of Interest Disclosure

I am an independent researcher in fintech and cryptocurrency market microstructure. I have no personal, financial, or advisory relationship with the author. My prior published research overlaps thematically with this paper but is not cited herein.

Summary

This paper applies fractionally integrated GARCH-family models (FIGARCH, FIEGARCH, FIAPARCH, HYGARCH) to Bitcoin daily returns over four sub-periods spanning 2014 to 2022: pre-COVID-19, COVID-19, between COVID-19 and the Russo-Ukrainian War, and the Russo-Ukrainian War. The author selects models via information criteria, estimates parameters, and conducts diagnostic tests on standardized residuals. The topic (volatility regime dependence during geopolitical crises) is timely and relevant. However, the paper suffers from serious methodological errors, incoherent exposition, data integrity concerns, and conclusions that are not supported by the statistical evidence presented.

Strengths

Appropriate model class for the research question. Fractionally integrated GARCH models are well-suited to capturing long-memory volatility persistence, which is a documented feature of Bitcoin returns. The choice to compare FIGARCH, FIEGARCH, FIAPARCH, and HYGARCH across crisis regimes is analytically sensible.

doi.org/10.32388/ZMMST8

Regime-specific estimation. Estimating separate models for four distinct sub-periods is a defensible approach to testing whether crisis events alter volatility dynamics. The pre-COVID, COVID, inter-crisis, and war partition maps onto genuinely different market conditions.

Diagnostic testing. The inclusion of normality tests (Jarque-Bera), Ljung-Box Q-statistics on standardized residuals, and ARCH-LM tests shows awareness of standard GARCH model validation procedures.

Weaknesses

Critical methodological error: the d parameter exceeds unity. In Table 4, the estimated d parameter for the "Between COVID-19 Crisis and Russo-Ukrainian War" period is 1.186. This value is outside the parameter space required for covariance stationarity in any FIGARCH-family model. Standard FIGARCH requires $0 \leq d \leq 1$, with $d = 1$ collapsing to IGARCH; FIEGARCH requires $d < 0.5$ for stationarity of the log-variance process. A value of 1.186 implies the conditional variance process is non-stationary and explosive. This invalidates the FIEGARCH(2,1) specification for that sub-period. The paper cannot claim this is the "appropriate model" when its key parameter lies outside the feasible parameter space. Re-estimation with constrained optimization is required.

Standard errors are not reported for the fractional integration parameter. Table 4 gives point estimates for d but no standard errors, t-statistics, or confidence intervals. Without these, it is not possible to test formally whether d differs significantly from 0 (short memory) or from 1 (IGARCH), which is the core inferential question a FIGARCH-family paper is meant to answer. Reporting the full set of parameter standard errors has been standard practice since Baillie, Bollerslev, and Mikkelsen (1996). The $d = 1.186$ estimate is doubly problematic in the absence of a standard error, since it is not possible to determine whether the estimate is imprecisely measured (in which case the boundary could plausibly be reconciled by a wide confidence interval) or precisely misestimated (in which case the model is invalid). Every parameter table in the paper should be recompiled with standard errors and significance markers.

The FIGARCH-CHUNG model produces nonsensical information criteria. In Table 3, the FIGARCH-CHUNG(2,2) model during the COVID-19 period reports Akaike = 102.945124, Shibata = 102.944415, Schwarz = 103.019614, Hannan-Quinn = 102.974721. These values are approximately 20 times larger than all other models in the same row and 19 times larger than the same model in other periods. This is not a minor discrepancy. It suggests a convergence failure, numerical overflow, or misspecified likelihood. A responsible analysis would flag this as an estimation failure and exclude the model from comparison.

The absence of any acknowledgment of this anomaly weakens confidence in the estimation procedure and is itself a concern.

Model selection is inconsistent and poorly justified across periods. The author selects FIAPARCH-BBM(2,1) for pre-COVID, FIEGARCH(2,2) for COVID, and FIEGARCH(2,1) for both the inter-crisis and war periods. For the Russo-Ukrainian War period, FIEGARCH(2,1) does have the lowest AIC at 5.390 (versus FIAPARCH-BBM(2,1) at 5.404 and HYGARCH(2,1) at 5.486), so the selection is technically correct by information criterion. However, the author never explains why different model families are selected across periods (APARCH versus EGARCH) rather than simply different (p, q) orders within the same family. If the research question is about volatility dynamics changing across crises, holding the model family constant and varying only the orders would isolate the regime effect more cleanly. The current approach confounds model-family differences with regime differences.

The conclusion overstates the evidence. The author claims: "a war will make cryptocurrency more important on the exchange market even than an epidemic in the example of Covid-19." This is not demonstrated. The paper shows that FIEGARCH(2,1) is selected for both the inter-crisis and war periods, while FIEGARCH(2,2) is selected for COVID. The difference between (2,1) and (2,2) is one additional ARCH term, not a finding about "importance on the exchange market." The EGARCH parameters (θ_1, θ_2) reported across the three FIEGARCH specifications are small in magnitude and mixed in sign, and are not compared statistically across regimes. There is no test of parameter equality, no confidence interval for the difference, and no economic metric (trading volume, market capitalization, volatility level) supporting the claim that war exceeds a pandemic in market impact. This is speculation dressed as inference.

The "negative asymmetry effect" is mischaracterized. The author repeatedly states that negative shocks increase volatility more than positive shocks. In FIEGARCH, this is captured by the news impact function g(z_{t-1}) = θ z_{t-1} + γ (|z_{t-1}| - E|z_{t-1}|), so the sign and magnitude of the asymmetry response depend on the interaction of θ and γ, not on θ alone. The paper reports θ_1 = −0.013 and θ_2 = +0.018 for the COVID period, which are mixed in sign and both near zero. For the war period, θ_1 = −0.010 and θ_2 = −0.094. The γ magnitude coefficient does not appear alongside θ in Table 4. In the FIEGARCH specification, γ is the leverage parameter, and its omission means the paper cannot distinguish between a leverage effect (negative shocks amplifying volatility more than positive shocks of equal magnitude) and a mere sign effect (negative shocks acting through θ alone). Without γ, the news impact curve cannot be traced. The author does not test whether θ values are jointly significant, nor whether they differ across regimes.

Claiming "the negative asymmetry effect exists for the Bitcoin series during the four sub-periods" is unsupported by the reported parameter estimates, several of which are positive or economically negligible.

Data description is inadequate. The paper uses daily BTC/USD returns from 31/12/2014 to 15/09/2022 (2,921 observations). The source is not stated. "BTC/USD exchange rate" is ambiguous because Bitcoin trades on hundreds of exchanges with price dispersion. Which exchange, Coinbase, Bitstamp, Kraken, or a composite index? The 2014 starting date is before most institutional price feeds were reliable; early Bitcoin price data are notoriously noisy and subject to exchange-specific anomalies (Mt. Gox and similar). Without source transparency, replication is impossible.

The sub-period definitions are arbitrary and overlap with other crises. The "Pre-COVID-19 Crisis" period (17/09/2014 to 30/12/2019) includes the 2014 ruble crisis, the 2015 Chinese stock market crash, and the 2018 Turkish lira crisis, events the author mentions but does not model separately. The COVID-19 period ends on 31/12/2020, well before the pandemic's financial market effects subsided. The "Between COVID-19 and Russo-Ukrainian War" period (01/01/2021 to 23/02/2022) includes the 2021 crypto bull market and the global energy crisis. Lumping these into a single "inter-crisis" regime obscures more than it reveals. A more rigorous approach would use structural break tests (Bai-Perron, ICSS, Chow) to identify regime changes endogenously rather than imposing them by calendar date.

Exposition is repetitive and poorly structured. The abstract is over 300 words and restates the same findings three times. Section 4.2's table descriptions are copy-pasted from the tables themselves with minimal added interpretation. The conclusion introduces new claims (Bitcoin as "virtual currency and not an electronic currency," references to Tulipomania, unbanked populations) that have no connection to the empirical analysis. The paper reads as if multiple drafts were concatenated without editing.

The Q-statistic interpretation is incorrect. In Table 6, the author reports Q(5) = 11.5482 [0.04153]* for the pre-COVID FIAPARCH model and states this "suggests that there is evidence of autocorrelation in the residuals." The Q-statistic here tests standardized residuals, which should be i.i.d. under a correctly specified GARCH model. Across the four models and four reported lag orders, sixteen Q-tests are shown; under the null of no misspecification with $\alpha = 0.05$, roughly one false rejection would be expected by chance alone. A single significant Q(5), with no matching significance at lags 10, 20, or 50, is therefore consistent with the null rather than evidence against it. If the author suspects residual autocorrelation, the appropriate diagnostic is inspection of the standardized-residual autocorrelation function across all lags up to a practical maximum (approximately $2\sqrt{T} \approx 108$ for $T = 2{,}921$), not lag-by-lag reporting at hand-

picked orders. More importantly, the FIAPARCH model is selected for the pre-COVID period despite this residual autocorrelation, while the FIEGARCH models for later periods show no such issue. This undermines the claim that FIAPARCH is "appropriate" for that period.

Log-likelihood values are not reported for the selected models. Tables report information criteria (AIC, BIC, Shibata, Hannan-Quinn) but not the underlying log-likelihood values. Since information criteria are simple transformations of the log-likelihood and the parameter count, reporting only the criteria makes independent verification of the calculations impossible. More importantly, the absence of log-likelihoods precludes formal likelihood-ratio tests between nested specifications. Testing, for example, whether FIEGARCH(2,2) is a statistically meaningful improvement over FIEGARCH(2,1) requires the LR statistic $2 \times (\log L_{(2,2)} - \log L_{(2,1)})$, distributed $\chi^2(1)$ under the null that the additional parameter is zero, and this test cannot be constructed from information criteria alone. Standard practice is to report log-likelihood, parameter count, and sample size alongside the information criteria.

The conclusion's policy implications are unsupported. The author advises that "policymakers need to consider the impact of their actions on market sentiment and volatility" and that "regulatory actions that result in negative market sentiment could lead to significant volatility." No policy variable is included in the analysis. No regulatory event is modeled. These are generic assertions with no empirical basis in the paper.

Note for the Platform

This paper has accumulated 22 approving open peer reviews on Qeios prior to this assessment. Several of the issues identified above (in particular the $d = 1.186$ parameter estimate outside the feasible space, the absence of any reported standard errors for the fractional integration parameter, the incomplete FIEGARCH parameter reporting, and the unaddressed FIGARCH-CHUNG anomaly of magnitude approximately 20 times the neighboring information criteria) are verifiable from the paper's own tables in isolation and do not require independent replication. The platform may wish to consider whether a correction notice or editorial addendum is warranted alongside the existing review record.

Recommendation

Major revisions required. The paper addresses a relevant topic but contains critical methodological errors that invalidate key results. Before resubmission, the author should:

Address the $d = 1.186$ impossibility. Re-estimate the FIEGARCH(2,1) model for the inter-crisis period with constrained optimization on d. If the constraint binds at the upper boundary, report this as evidence of

IGARCH-like behavior, not fractional long memory. If the model fails to converge, report the failure and select the next-best model from Table 3 with a valid d estimate.

Report parameter standard errors throughout. Recompile every parameter table (Tables 4 through the end) with standard errors, t-statistics, and significance markers.

Report full FIEGARCH parameter sets. Include both θ (sign effect) and γ (leverage effect) in the parameter tables so that the news impact curve can be constructed and the asymmetry claim can be evaluated.

Investigate and explain the FIGARCH-CHUNG anomaly. Report whether the outlier information criteria are due to convergence failure, numerical overflow, or a coding error. If unresolvable, exclude the model from comparison rather than reporting the values.

Report log-likelihood values alongside information criteria. This enables independent verification and formal likelihood-ratio testing between nested specifications.

State the estimation software and package. Report whether the analysis used R (rugarch, fGarch), Python (arch), MATLAB, EViews, OxMetrics G@RCH, or another platform, and cite the version. Different implementations parameterize the fractional differencing operator differently, and the $d > 1$ estimate may be traceable to a software-specific convention that was not verified against theoretical constraints.

Hold the model family constant across regimes. Estimate FIEGARCH(p, q) for all four periods, varying only (p, q) by information criteria, to isolate regime effects from model-family effects. Alternatively, justify why different families (APARCH versus EGARCH) are necessary and test whether parameter differences across regimes are statistically significant.

State the data source explicitly. Identify the exchange, API, or data vendor (Coinbase Pro, Bitstamp, CryptoCompare, or similar). Report whether prices are closing, VWAP, or median. Document any data cleaning (outlier removal, exchange downtime handling).

Use endogenous structural break tests. Replace the calendar-date regime splits with Bai-Perron, ICSS, or similar tests to identify volatility regimes empirically. This also addresses the classic concern that apparent long memory can arise spuriously from structural breaks in short-memory processes (Diebold and Inoue, 2001).

Test parameter equality across regimes. Use Wald tests or likelihood ratio tests to determine whether the EGARCH asymmetry parameters (θ, γ) differ significantly across the COVID and war periods. Without such tests, claims about war exceeding the pandemic in market impact are unsupported.

Tighten the conclusion. Remove the digressions on Tulipomania, unbanked populations, and the definition of "virtual currency." Focus on what the models actually show: which volatility specifications fit best in which regimes, and what the parameter estimates imply about persistence and asymmetry.

Proofread thoroughly. The abstract is roughly three times longer than typical journal guidelines. Table descriptions should interpret, not recapitulate. The phrase "make you dizzy" (Section 2) is informal and out of place in a technical paper.

Guidance for Authors: Literature Review

The paper's literature engagement would benefit from citing the following, which speak directly to its methodological choices:

Baillie, R.T., Bollerslev, T., and Mikkelsen, H.O. (1996). "Fractionally integrated generalized autoregressive conditional heteroskedasticity." Journal of Econometrics. The original FIGARCH paper; an essential foundation for any FIGARCH application.

Diebold, F.X. and Inoue, A. (2001). "Long memory and regime switching." Journal of Econometrics. Shows that structural breaks in short-memory processes can produce apparent long memory. Directly relevant given the paper's regime-partitioning approach.

Bai, J. and Perron, P. (2003). "Computation and analysis of multiple structural change models." Journal of Applied Econometrics. Reference for the endogenous break-test recommendation above.

Katsiampa, P. (2017). "Volatility estimation for Bitcoin: A comparison of GARCH models." Economics Letters. Prior GARCH-family model comparison for Bitcoin; a useful benchmark.

Ardia, D., Bluteau, K., and Rüede, M. (2019). "Regime changes in Bitcoin GARCH volatility dynamics." Finance Research Letters. Directly addresses regime shifts in Bitcoin volatility using a related methodology.

Chkili, W., Hammoudeh, S., and Nguyen, D.K. (2014). "Volatility forecasting and risk management for commodity markets in the presence of asymmetry and long memory." Energy Economics. FIAPARCH application in a comparable asymmetry-plus-long-memory setting.

Corbet, S., Lucey, B., Urquhart, A., and Yarovaya, L. (2019). "Cryptocurrencies as a financial asset: A systematic analysis." International Review of Financial Analysis. Comprehensive survey of the cryptocurrency risk-return literature.

A thematic reorganization of the literature review would improve coherence. Three natural clusters emerge from the sources above: (a) long-memory volatility modeling (Baillie, Bollerslev, and Mikkelsen, 1996; Chkili, Hammoudeh, and Nguyen, 2014), (b) Bitcoin-specific GARCH-family applications (Katsiampa, 2017; Ardia, Bluteau, and Rüede, 2019; Corbet et al., 2019), and (c) regime-switching and structural breaks (Diebold and Inoue, 2001; Bai and Perron, 2003). Aligning the review with these clusters would tighten the narrative and match citations to the paper's methodological choices.

Declarations

Potential competing interests: No potential competing interests to declare.