Long Memory Diagnostics and Forecast Performance of NASDAQ Volatility Proxies

Ramon Ferrer

ENEB Business School

DOI: https://doi.org/10.47772/IJRISS.2026.101000670

Received: 29 October 2026; Accepted: 03 November 2026; Published: 14 November 2026

ABSTRACT

Fractional dependence, changing volatility levels, or their combination can produce persistent stock-market volatility. Using 9,067 daily returns from January 1990 to December 2025, this study examines their empirical implications for the NASDAQ Composite. Qu's test against spurious long memory was combined with local Whittle and log-periodogram estimates. Calibration of a cumulative-sum diagnostic used a fitted fractional model. Autoregressive fractionally integrated moving-average, autoregressive moving-average, heterogeneous autoregressive, and historical-mean models were compared through rolling forecasts. For log squared returns, the primary local Whittle estimate was 0.371, with a 95% confidence interval of 0.321–0.422. Across bandwidths, subperiods, and transformations of near-zero returns, the estimates remained positive. Little corresponding persistence appeared in signed returns. Under both filtered and unfiltered specifications, however, Qu's test rejected the genuine long-memory null. Only a slight exceedance of the critical value occurred for the primary filtered statistic. With a simulation probability of 0.227, the cumulative-sum diagnostic did not reject the fitted Gaussian fractional model. At five and 22 trading days, the fractional model achieved the lowest root mean squared error across 196 common forecast origins. Relative to the heterogeneous autoregressive model, its reductions were 2.1% and 5.6%. Significance did not persist for these comparisons after Holm adjustment. Persistent loss differences also appeared in comparisons with the historical mean, weakening conventional forecast inference. The evidence supports persistence in the observed volatility proxies. ARFIMA produced smaller forecast errors at some horizons, but its superiority over dynamic alternatives was not established. These rankings apply to average log squared returns in one holdout period. Their relevance to conditional variance forecasting and financial decisions remains untested.

Keywords NASDAQ Composite, volatility persistence, long memory, structural change, forecast evaluation

INTRODUCTION

Even when the direction of daily returns is difficult to predict, stock-market volatility remains elevated after large price movements. Established models for this behaviour include autoregressive conditional heteroskedasticity (ARCH) and generalised ARCH (GARCH) [1, 2]. Long-memory specifications have also been motivated by slowly decaying autocorrelations of absolute returns [3]. Slow volatility decay is also allowed by fractionally integrated exponential GARCH models [4]. The assumed decay rate affects how a model uses distant observations and projects future volatility, making the distinction relevant.

Aggregating heterogeneous dynamic processes can generate fractional dependence [5]. Such dependence has a parsimonious representation through fractional differencing [6]. Fractionally integrated conditional variance models [7] and long-memory stochastic volatility models [8] are among the extensions. Different observable memory properties can result from nonlinear transformations of latent volatility processes, as Robinson showed [9]. Evidence from one volatility proxy therefore need not carry over unchanged to another.

Seasonal or cyclical structure can coexist with persistence. A seasonal long-memory stochastic volatility formulation was developed by Gonzaga [10]. Measured volatility persistence can also be affected by intraday periodicity [11]. When persistence occurs around frequencies beyond zero, this broader setting becomes relevant. The underlying process, however, cannot be identified from slowly declining sample autocorrelations alone. The

difficulty of distinguishing real and spurious memory in stock-market data was documented by Lobato and Savin [12]. Apparent long memory can result from occasional level shifts [13]. Similar behaviour can arise through regime switching [14]. With these alternatives, a positive fractional estimate remains an incomplete basis for volatility model selection.

The estimation problem has attracted substantial attention. Robinson's semiparametric methods allow low-frequency dependence to be studied without specifying the complete short-run dynamics [15]. The lowest Fourier frequencies enter his Gaussian semiparametric estimator [16]. A related route to inference comes from his analysis of log-periodogram regression [17]. Generalised fractional dependence can also be represented through wavelet methods. A Bayesian analysis using wavelet coefficients was developed by Gonzaga and Kawanaka [18]. Because these methods address different aspects of estimation, they should not be read as interchangeable evidence about a single financial mechanism.

A separate question concerns forecast performance. Daily, weekly, and monthly components allow the heterogeneous autoregressive model to approximate persistent volatility [19]. Because both models can summarise dependence across several horizons, it offers a useful benchmark for a fractional model. The measurement of volatility also affects forecast comparisons. Andersen and Bollerslev showed how forecast quality can be obscured by noisy return-based measures [20]. More informative measures are obtained from intraday information in realised-volatility research [21]. Conditions for preserving forecast rankings with imperfect volatility proxies were established by Patton [22]. In the present study, forecasts are evaluated directly on the scale of the daily return transformation used.

This study examines whether evidence of persistent NASDAQ volatility is accompanied by a dependable forecasting advantage for ARFIMA. It combines memory estimates, tests of the fractional explanation, and comparisons with persistent short-memory forecasts. The contribution is an empirical assessment of how these findings agree and where they differ. Positive memory estimates persist across several checks, but the evidence for a stable fractional mechanism remains uncertain. ARFIMA also has the smallest errors at some horizons without a clear advantage over dynamic alternatives. These results show why a positive memory estimate alone is insufficient grounds for choosing a forecasting model.

METHODOLOGY

Data and volatility proxies

The NASDAQ-supplied NASDAQCOM series [24] provided daily NASDAQ Composite closing values through the Federal Reserve Bank of St. Louis database. Data were requested for 1 January 1990 to 31 December 2025. The downloaded file contained 324 date records without a price. These records were removed before returns between consecutive observed prices were calculated. They were not reconciled with the exchange calendar, so the analysis cannot distinguish scheduled market closures from genuine data gaps. From 3 January 1990 to 31 December 2025, the resulting series comprised 9,067 returns. Two returns were exactly zero. Calendar interpolation was not applied. Intervals between observed prices exceeded four calendar days on three occasions.

$$r_t = 100\{\log(P_t) - \log(P_{t-1})\}, \quad a_t = |r_t|, \quad y_t = \log(r_t^2 + 10^{-8}). \qquad (1)$$

The observed closing index level is denoted by P_t. Log squared returns form the primary series y_t. A transformation sensitivity check is provided by absolute returns a_t, with signed returns r_t serving as a comparison. Trading observations determine the time index. Reinvested dividends are excluded from the returns, which measure changes in a price index. As noisy volatility proxies, squared close-to-close returns do not measure intraday realised variance.

High-frequency observations are also used in research on generalised long-memory volatility [25]. A different information set is available in the daily series considered here. Intraday seasonality, in particular, cannot be separated from overnight variation using this series. Persistence and predictability in the observed daily proxies are therefore the focus of the analysis. An intraday latent-volatility model is not estimated.

Semiparametric memory estimation

Near zero, the spectral density of a stationary fractionally integrated process behaves as $G\lambda^{-2d}$, with G positive. Under the model assumptions, stationary long-range dependence is indicated by d values between zero and one-half. Over d from −0.45 to 0.95, the local Whittle estimator minimises the objective below. Estimates can approach the stationarity boundary within this broad numerical range without being forced below it.

$$Q(d) = \log\left\{\frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d} I(\lambda_j)\right\} - \frac{2d}{m}\sum_{j=1}^{m}\log(\lambda_j), \quad \lambda_j = \frac{2\pi j}{n}. \qquad (2)$$

Calculation of the periodogram $I(\lambda_j)$ used the demeaned series without padding. With $m = \text{floor}(n^{0.65})$, the primary bandwidth comprised 373 frequencies. Exponents of 0.55, 0.60, 0.70, and 0.75 were used in the sensitivity analyses. The asymptotic standard error $1/(2\sqrt{m})$ was used to construct pointwise 95% intervals. These intervals require the assumptions that support semiparametric inference. Near or above the stationarity boundary, the intervals become less secure to interpret.

Following Geweke and Porter-Hudak [26], a second estimate for the primary series was obtained by log-periodogram regression. Using the same primary bandwidth, the implementation regressed the log periodogram on $-2\log[2\sin(\lambda_j/2)]$. An asymptotic standard error of $\pi/\sqrt{(24m)}$ was reported. The bandwidth and short-run spectrum affect finite-sample bias [27]. Inference in long-memory stochastic volatility is also affected by measurement noise [28]. Comparing bandwidths and estimators is therefore preferable to treating a single interval as definitive.

For further checks, separate estimates of d were obtained for the first and second sample halves, the initial training period, and the final 3,000 observations. Values from 10^{-10} to 10^{-4} were used for the positive offset in equation (1). The lower and upper 0.5% of y_t were winsorised in a separate sensitivity analysis. All observations were retained for the primary analysis. Whether extreme or near-zero returns drive the result is assessed by these checks. Uncertainty about changes in the generating process remains despite them.

Tests of the fractional explanation

The genuine long-memory null was assessed against spurious-memory alternatives using Qu's test [29]. A bandwidth of $m = \text{floor}(n^{0.7})$, or 589 frequencies, was used for the test. At a 5% significance level, the primary specification used the package's short-memory filter and trimming parameter $\varepsilon = 0.05$. Sensitivity checks comprised trimming at $\varepsilon = 0.02$ and an unfiltered specification. Under the test's assumptions, rejection counted as evidence against the genuine long-memory null. Identification of a particular break process was not inferred from rejection.

Qu's test was implemented with SpuriousMemory version 1.0.0 [30]. Filtered and unfiltered results were compared to assess sensitivity to short-run dynamics. The test decision was interpreted alongside the memory estimates and cumulative-sum diagnostic.

An additional sensitivity check used a trimmed cumulative-sum statistic. Between the 15th and 85th sample percentiles, it measured the largest absolute cumulative deviation from the sample mean. The statistic was scaled by dividing it by the sample standard deviation and $\sqrt{n}$. From the full-sample fitted autoregressive fractionally integrated moving-average model, 999 Gaussian series were simulated for calibration. The analysed series in each simulation was preceded by 5,000 initial observations. The probability was calculated as $(1 + N_e)/1{,}000$, with N_e counting simulated statistics at least as large as the observed value.

Persistent fractional dependence is allowed in the null model under this calibration. It uses the estimated parameters and Gaussian innovations without re-estimating the model within each simulation. The cumulative-sum feature is therefore assessed through a model-based diagnostic. Exclusion of structural change in general is not provided by this test. For multiple-break methods, separate assumptions and selection procedures are required [31, 32]. Neither the number nor the timing of structural breaks is identified by these diagnostics.

Fractional and short-memory models

An autoregressive fractionally integrated moving-average model, abbreviated ARFIMA, was used as the fractional model. Before the remaining autoregressive and moving-average dynamics were modelled, the

demeaned log squared-return series was fractionally differenced.

$$\Phi(L)(1-L)^{d}(y_t-\mu)=\Theta(L)u_t. \qquad (3)$$

The lag operator is L, the fitted mean is μ, and the model innovation is u_t. The software's moving-average sign convention is used for the polynomial Θ(L). With 200 terms in the likelihood approximation and d constrained to [0, 0.49], estimation used the approximate likelihood implemented in fracdiff. Haslett and Raftery provide the basis for this approximation [33]. For the fitted specification, a fractional estimate near the upper constraint was treated as a warning.

The final 3,000 observations available before the holdout period were used to select the initial orders. Each candidate autoregressive and moving-average order could take values from zero to two. Within each model family, selection among the nine candidates used the Bayesian information criterion [34]. The forecast experiment subsequently held these orders fixed. To describe the complete historical series, a separate full-sample fit used the selected fractional order. Estimation of the held-out forecasts did not use this descriptive fit.

An autoregressive moving-average model, abbreviated ARMA, served as the short-memory comparator, with its orders selected from the same grid. A mean was included in each model family. The forecast package generated the forecasts [35]. Descriptive summaries and Ljung–Box statistics at lags 22, 66, and 126 were used to examine model residuals [36]. After fractional estimation, the residual tests were interpreted as approximate diagnostics.

The forecast experiment was restricted to models fitted directly to log squared returns. GARCH-family models were not included. Their conditional variance forecasts do not directly equal the target in equation (5). A comparison on that target would require forecasts of the expected log squared return under a specified return distribution. GARCH models could also enter a separate comparison on a common variance scale. Their omission limits the conclusions about ARFIMA's practical value relative to established volatility models.

Wavelet representations also offer a way to estimate generalised fractional models. Decorrelation properties of wavelet packet coefficients for generalised long-memory stochastic volatility were derived by Gonzaga [23]. A different estimation approach is addressed by that result. In equation (3), the ARFIMA model imposes a zero-frequency fractional component, with neither wavelet coefficients nor additional cyclical poles estimated.

Rolling forecasts and evaluation

The holdout period comprised the final 1,000 trading observations. Training initially ended on 5 January 2022, with the holdout starting on 6 January 2022. Five trading observations separated successive forecast origins. Each origin in the restricted common set had 22 subsequent observations. At every horizon, this yielded 196 origins. Only the preceding 3,000 observations were used to refit each model.

The previous daily value and the previous five-day and 22-day means of y_t entered the heterogeneous autoregressive comparator, abbreviated HAR. Corsi's lag structure was adapted to log squared returns in this model. The equation fitted to the series appears below.

$$y_t=\beta_0+\beta_d y_{t-1}+\beta_w \bar{y}_{t-1,5}+\beta_m \bar{y}_{t-1,22}+e_t. \qquad (4)$$

Only observations preceding t enter the averages in equation (4). Recursion was used to generate multi-step forecasts. The same rolling estimation window was applied to the historical-mean benchmark. The target for every model was the average of future daily log squared returns over h trading observations.

$$Y_{t,h} = \frac{1}{h}\sum_{j=1}^{h} y_{t+j}, \qquad h \in \{1, 5, 22\}. \qquad (5)$$

Equation (5) defines an average of future daily log squared returns. It differs from both the logarithm of aggregated squared returns and their arithmetic average. Root mean squared error, abbreviated RMSE, and

mean absolute error, abbreviated MAE, were calculated on this log scale. Exponentiating these forecasts would not provide the expected average squared return. No variance-scale forecasts or QLIKE comparisons were obtained. QLIKE would require strictly positive forecasts for a common variance target and an appropriate proxy for that target [22]. The reported rankings therefore concern prediction of the transformed series. Intraday stock volatility has also been examined using seasonal fractional forecasting [37]. Over a later holdout, the present design evaluates a daily zero-frequency approximation without assuming that intraday findings carry over.

Squared-loss differences between ARFIMA and each of ARMA, HAR, and the historical mean formed the paired comparisons. Diebold and Mariano provide the predictive-accuracy framework [38]. When comparing forecasting procedures under parameter estimation, a fixed rolling window is also relevant [39]. ARFIMA is favoured by negative mean differences.

Newey–West heteroskedasticity and autocorrelation consistent (HAC) standard errors were used for confidence intervals and two-sided tests [40]. The larger of the overlap requirement in forecast-origin units and floor[4(N/100)$^{2/9}$] determined the lag. For every reported primary comparison, this gave a lag of four. Across horizons, forecast errors have different overlap, while dependence may extend beyond that mechanical overlap. Bandwidth choice consequently continues to affect covariance estimation [41]. Lags five, ten, and twenty were used for sensitivity checks.

The nine primary model-by-horizon comparisons were adjusted using Holm's procedure [42]. No multiplicity adjustment was applied to the plotted 95% confidence intervals, which remain pointwise. Persistence in the loss differences was assessed through local Whittle estimates with bandwidth floor($N^{0.60}$), or 23 frequencies. Sufficiently weak dependence is required for conventional HAC inference. In a short loss series, a positive memory estimate warns about this assumption without definitively classifying the integration order.

R was used for the analysis with fracdiff, forecast, SpuriousMemory version 1.0.0, and sandwich. The fractional likelihood used 200 terms, with d constrained to [0, 0.49]. Autoregressive and moving-average orders ranged from zero to two. Forecasts used a rolling window of 3,000 observations, origins five trading days apart, and horizons of one, five, and 22 trading days. The cumulative-sum calibration used 999 Gaussian simulations, each preceded by 5,000 discarded observations. Only information available before each origin entered the forecast estimates.

RESULTS

Observed persistence and return characteristics

For daily percentage log returns, the mean was 0.0433 and the standard deviation was 1.4650. Returns ranged from a minimum of −13.1492% to a maximum of 13.2546%. Skewness and excess kurtosis of returns were −0.1775 and 6.6868, respectively. With skewness of 11.3856 and excess kurtosis of 208.442, squared returns were much more skewed. After transformation to log squared returns, these values fell to −1.0375 and 2.2677, respectively. A mean of −1.1234 and a standard deviation of 2.4724 characterised the transformed series.

The daily series and contrasting dependence patterns appear in Figure 1. At the selected lags, signed returns had small autocorrelations. At lags one and 252, these were −0.0365 and 0.0038. For absolute returns, autocorrelations were 0.2818 at lag one, 0.2295 at lag 22, and 0.1041 at lag 252. Positive dependence also remained in log squared returns, whose corresponding values were 0.1393, 0.1175, and 0.0735. Persistence was therefore found mainly in measures of return magnitude.

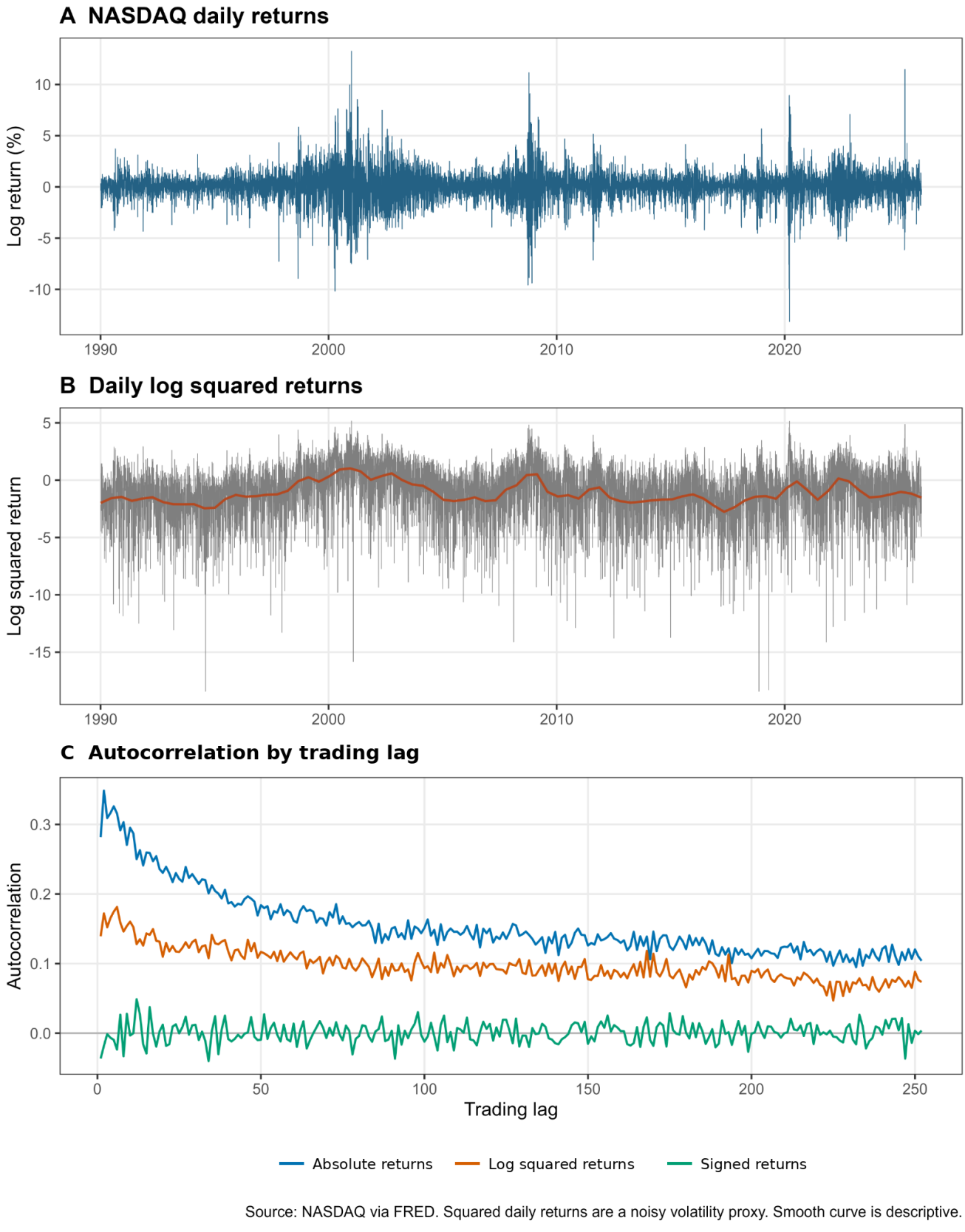

Figure 1. NASDAQ returns and volatility-proxy persistence. Panel A shows daily percentage log returns. Panel B shows log squared returns and a descriptive smooth curve. Panel C shows autocorrelations through 252 trading lags for signed returns, absolute returns, and log squared returns. The smooth curve is descriptive and was not used in estimation.

Memory estimates and sensitivity checks

The volatility-proxy estimates appear in Table 1. For log squared returns, the primary local Whittle estimate

was 0.3713, with a 95% confidence interval of 0.3206–0.4221. For signed returns, the corresponding estimate was 0.0312, with an interval of −0.0195–0.0819 and p = 0.228. The larger absolute-return estimate was 0.4623, with an interval of 0.4115–0.5130. One-half therefore lay within the primary absolute-return interval.

Table 1. Local Whittle memory estimates across bandwidths

Bandwidth exponent	Frequencies	Absolute returns d [95% CI]	Log squared returns d [95% CI]
0.55	150	0.483 [0.403, 0.563]	0.468 [0.388, 0.548]
0.6	236	0.495 [0.432, 0.559]	0.394 [0.330, 0.458]
0.65	373	0.462 [0.412, 0.513]	0.371 [0.321, 0.422]
0.7	589	0.452 [0.412, 0.492]	0.332 [0.292, 0.373]
0.75	929	0.425 [0.393, 0.457]	0.305 [0.273, 0.338]

N = 9,067. CI is short for confidence interval. An exponent of 0.65 defines the primary bandwidth. Pointwise asymptotic 95% intervals require the relevant stationarity and spectral conditions. Each row is a sensitivity estimate from the same series. Crossing 0.5, an interval does not establish stationary fractional dependence.

Log squared returns gave estimates from 0.4681 at the narrowest bandwidth to 0.3055 at the widest. Every pointwise interval omitted zero, but the narrowest-band interval extended above one-half. Absolute returns gave estimates between 0.4251 and 0.4954. The Figure 2A bandwidth pattern indicates that no single d value summarises all low-frequency estimates equally well.

For log-periodogram regression, d was 0.3743, with a 95% interval of 0.3093–0.4394. The primary local Whittle estimate lay close to this value. Local Whittle estimates ranged from 0.3703 to 0.3815 as the offset for near-zero returns varied. After winsorising the outer 0.5% of log squared returns, d was 0.3775, with an interval of 0.3267–0.4282. Neither the two exactly zero returns nor the most extreme transformed observations therefore explained the positive estimates.

Estimates from the first and second sample halves were 0.3743 and 0.3569. Their intervals spanned 0.3108–0.4379 and 0.2934–0.4204. For the initial training sample and final 3,000 observations, estimates were 0.3654 and 0.3161, respectively. Zero lay outside both intervals. Although these estimates show persistence across broad sample portions, their overlapping intervals do not establish parameter constancy.

Evidence against a pure fractional explanation

In all four settings reported, Qu's test rejected the genuine long-memory null. The filtered statistic at $\varepsilon = 0.05$ was 1.15643 against a 5% critical value of 1.155. The critical value was exceeded by only 0.00143. At $\varepsilon = 0.02$, filtering gave a statistic of 1.39991 against a critical value of 1.252. At $\varepsilon = 0.05$ and $\varepsilon = 0.02$, the unfiltered statistics were 2.21789 and 2.43062. Although the primary decision remained rejection, filtering substantially reduced the evidence against the null.

The cumulative-sum statistic took the value 5.17192. Fitted Gaussian ARFIMA calibration gave a 95th percentile of 7.20545 and a simulation probability of 0.227. This probability had a Monte Carlo standard error of 0.0132. On the cumulative-sum feature, the fitted fractional model was not rejected by the diagnostic. Qu's result was not overturned, because the statistics assess different features with different null calibrations.

ARFIMA(0, d, 1) and ARMA(1, 1) were selected by the training criterion. The minimum within-family Bayesian information criterion values for these models were 13,833.0 and 13,833.8. Because the likelihood calculations differ, these values were not used as a separate cross-family test. Under the fracdiff sign convention, full-sample ARFIMA estimates were d = 0.2939 and a moving-average coefficient of 0.2866. The estimate stayed below the 0.49 upper bound.

Negative skewness and heavy tails persisted in the fitted fractional residuals. Residual skewness and excess kurtosis were −1.2858 and 3.0264. At lags 22, 66, and 126, approximate Ljung–Box tests rejected residual white

noise, with p-values of 6.85×10^{-10}, 1.43×10^{-7}, and 7.33×10^{-6}. The selected fractional specification did not remove all residual autocorrelation. The order search covered autoregressive and moving-average orders from zero to two, but it did not establish robustness across broader model structures. These diagnostics weaken the interpretation of ARFIMA(0, d, 1) as an adequate description of the full series. They do not identify which alternative specification would perform better. The model remains a forecasting approximation whose adequacy must be judged separately from its descriptive ranking. Autoregressive and moving-average coefficients of 0.966 and −0.896 in the initial ARMA fit indicated that the short-memory comparator could also represent highly persistent dynamics over a finite sample.

Forecast accuracy across horizons

Every model produced successful forecasts for all 196 origins. Common-origin RMSE values and ARFIMA–HAR loss comparisons appear in Table 2. The lowest one-day RMSE, 2.3601, came from ARMA. HAR's 2.3763 was followed by ARFIMA's 2.3813. Historical-mean forecasts produced an RMSE of 2.4329. For ARFIMA–HAR, the squared-loss difference was 0.0239, with a pointwise 95% interval of −0.1124–0.1602.

Table 2. Forecast accuracy for future average log squared returns

Horizon	Model	RMSE	ARFIMA minus HAR ΔMSE [95% CI]	Holm p
1	ARFIMA	2.3813	0.0239	1
			[-0.1124, 0.1602]	
1	ARMA	2.3601	—	—
1	HAR	2.3763	—	—
1	Historical mean	2.4329	—	—
5	ARFIMA	1.101	-0.0515	0.0597
			[-0.0899, -0.0131]	
5	ARMA	1.1023	—	—
5	HAR	1.1242	—	—
5	Historical mean	1.3183	—	—
22	ARFIMA	0.7254	-0.0641	0.1298
			[-0.1187, -0.0094]	
22	ARMA	0.7455	—	—
22	HAR	0.7683	—	—
22	Historical mean	0.9901	—	—

All models shared 196 forecast origins and a rolling window of 3,000 observations. Trading observations define the horizons. Root mean squared error is abbreviated RMSE. Only ARFIMA rows show the paired mean squared-loss difference, ΔMSE. ARFIMA is favoured by negative values. The reported intervals are conventional pointwise 95% HAC intervals using lag four. Their coverage depends on assumptions about dependence in the loss differences. The sensitivity results show that the associated inference requires caution. All nine ARFIMA comparisons across models and horizons, including those reported in the text, receive Holm adjustment. Dashes mark comparisons that are unreported on a given row.

The lowest five-day RMSE was 1.1010 for ARFIMA. Nearly the same value, 1.1023, was obtained for ARMA, compared with 1.1242 for HAR. Relative to HAR, ARFIMA reduced RMSE by 2.1%. At 22 days, RMSE was 0.7254 for ARFIMA, compared with 0.7455 for ARMA and 0.7683 for HAR. Relative to HAR, this represented a reduction of 5.6%. With MAE values of 0.8501 and 0.6119, ARFIMA was also favoured at five and 22 days. The lowest one-day MAE was 1.7540 for ARMA.

As Figure 2B shows, zero lay outside the pointwise ARFIMA–HAR loss intervals at five and 22 days. Differences of −0.0515 and −0.0641 were obtained. The unadjusted tests gave p-values of 0.00853 and 0.02163. Adjustment by Holm's procedure over all nine comparisons gave p-values of 0.0597 and 0.1298. The adjusted 5% threshold was met by neither comparison. After adjustment, significance was not found for any ARFIMA–ARMA comparison. At one, five, and 22 days, the corresponding adjusted p-values were 0.6029, 1.0000, and 0.6338.

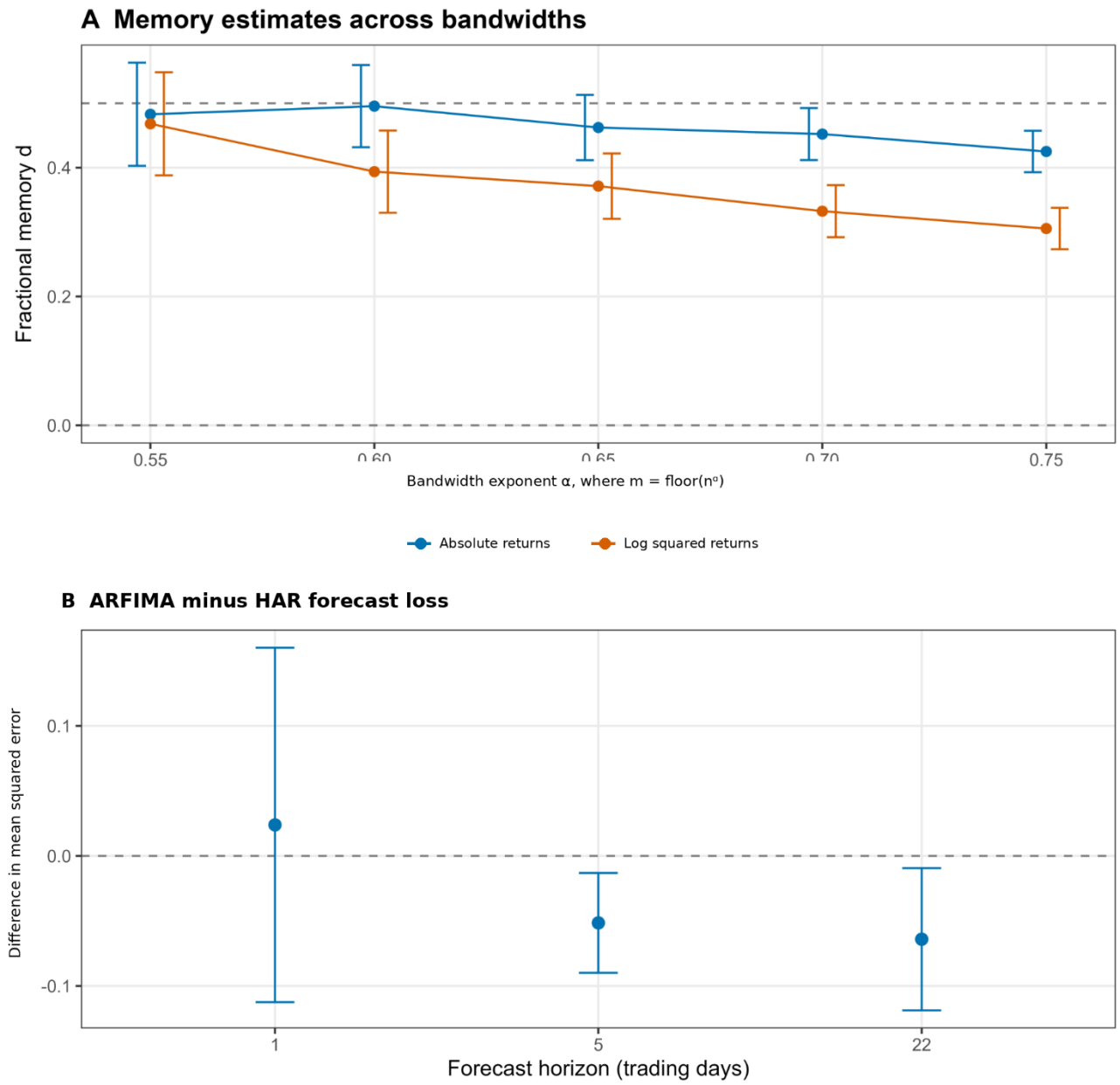

Figure 2. Memory estimates and forecast loss differences. Panel A shows local Whittle estimates with pointwise 95% confidence intervals across bandwidths. Dashed lines mark d = 0 and d = 0.5. Panel B shows ARFIMA minus HAR mean squared-error differences for average future log squared returns. Negative values favour ARFIMA. The conventional pointwise 95% HAC intervals are not adjusted for multiple comparisons. Their validity depends on the loss dependence assumptions. Holm-adjusted p-values are 0.0597 at five days and 0.1298 at 22 days.

At every horizon, RMSE was lower for ARFIMA than for the historical mean. Squared-loss differences at five and 22 days were −0.5258 and −0.4542, with nominal Holm-adjusted p = 0.0176 for both. Substantial persistence nevertheless remained in the loss differences. At one, five, and 22 days, their local Whittle estimates were 0.3547, 0.5515, and 0.6548, respectively. Zero lay outside all three reported intervals. Although based on only 196 loss observations and therefore imprecise, these estimates weakened the assumptions behind conventional significance testing.

Forecast inference changed with the HAC lag. For ARFIMA against HAR at 22 days, the unadjusted p-value rose from 0.0236 at lag five to 0.0611 at lag twenty. At five days, the corresponding p-values ranged from 0.0071 to 0.0170. For ARFIMA against the historical mean at 22 days, the p-value rose from 0.0044 to 0.0735. The smaller forecast errors are descriptive results for this holdout. Their inferential strength depends on the assumed dependence in the loss differences.

DISCUSSION

Persistent proxies and an uncertain fractional mechanism

Evidence of persistence is distinguished from evidence for a stable fractional process in the NASDAQ results. Across transformations and subperiods, positive memory estimates were obtained. Little corresponding dependence appeared in the signed-return comparison. Persistent variation in return magnitude is consistent with these findings. A pure stationary fractional interpretation is nevertheless difficult to sustain without qualification given the Qu rejections and residual diagnostics.

After filtering, the primary Qu statistic lay close to its critical value. Much larger statistics were obtained without filtering. The diagnostic was therefore substantially affected by short-run dynamics. Reporting the actual statistics and filtering choices together with the rejection decision is warranted by this pattern. The pattern also helps explain why agreement between two semiparametric estimates cannot settle the interpretation. The same observed series supplies the low-frequency behaviour on which both estimates depend.

The cumulative-sum test examined one feature of a fitted Gaussian fractional model. Its non-rejection does not show that volatility dynamics were unchanged. Qu's rejection weighs against genuine long memory under that test's assumptions, but it does not date a break. No formal procedure was used to locate structural changes, and persistence was not re-estimated within dated regimes. The sample-half estimates are broad sensitivity checks rather than formal regime estimates. Structural change therefore remains a possible explanation, not an identified source of the observed persistence.

Care is also needed when interpreting the transformation results. Absolute-return estimates lay closer to one-half than log squared-return estimates. Some intervals for absolute returns crossed the stationarity boundary. From these differences alone, a single latent-volatility interpretation cannot be drawn. The role of observational noise is identified in theoretical work on the local Whittle estimator for long-memory stochastic volatility [43]. Short-run spectral structure, changes in persistence, or finite-sample effects may account for the decline in estimates as the bandwidth expands. Their contributions cannot be separated using the present results.

Fractional approximation and practical forecasting

In the descriptive ranking, the fractional model's forecasting value rose with the horizon. Although it was not best for next-day log squared returns, its RMSE and MAE were smallest at five and 22 days. Some daily noise is reduced and the forecast target changes when future log proxies are averaged. Equation (5), rather than direct forecasts of daily variance, therefore defines the longer-horizon error reductions.

The close five-day performance of ARFIMA and ARMA is especially relevant. About 0.0012 separated their RMSE values, while adjusted significance was supported by none of their paired comparisons. Over the horizons considered in this experiment, persistent dependence can be approximated by a finite autoregressive moving-average model. Evaluation of a lower fractional-model error should therefore include a dynamic competitor as well as a constant mean.

The uncertainty in the forecast comparisons is central to their interpretation. Holm adjustment removed the apparent ARFIMA advantage over HAR, and no advantage over ARMA was established. The reported memory estimates for loss differences apply only to comparisons with the historical mean. They show that conventional inference can be unreliable when loss differences are persistent. The separate HAC-lag checks give a similar warning for some HAR and historical-mean comparisons. Multiplicity adjustment cannot repair p-values that rely on an unsuitable covariance estimate. No inference method designed for persistent loss differences was

implemented. The analysis therefore establishes an observed ranking in this holdout, not a dependable difference in expected forecast accuracy.

The broader long-memory literature also gains context from these results. Seasonal conditional-volatility structure is represented by periodic long-memory GARCH-in-mean models [44]. When the data support persistent cyclical components, such models may be relevant. Their forecasting advantage was not directly tested in the present experiment. The evidence therefore suggests evaluating any extension against persistent short-memory alternatives, with an explicit statement of the target scale.

Scope and limitations

The forecast comparison used one holdout beginning in January 2022 and one rolling estimation window of 3,000 observations. Alternative evaluation periods and window lengths were not tested. The 196 forecast origins describe performance within this design but do not establish stability across market periods. Model orders were fixed after the initial selection, so the rankings may also depend on that choice. Daily close-to-close returns combine overnight and trading-session movements and provide a noisy volatility proxy. The analysis did not evaluate other markets, conditional variance, value at risk, or economic value.

Gaussian innovations are imposed by the fractional likelihood and simulation calibration, despite residual asymmetry and heavy tails. The cumulative-sum calibration uses fitted parameters as fixed values and omits parameter-estimation uncertainty. Regularity and stationarity assumptions underpin the semiparametric intervals. As broad sensitivity analyses, the sample-half checks do not provide a formal test of parameter stability. Fixing forecast orders after initial selection limited their adaptation to later changes. Based on only 196 observations, the loss-memory checks warrant interpretation as diagnostic warnings.

Misleading linear persistence measures can arise from nonlinear short-memory models [45]. Parameter uncertainty in a more explicit latent model can be addressed through Bayesian estimation of generalised long-memory stochastic volatility [46]. The failure of the available diagnostics to identify a unique mechanism is underscored by these alternatives. Evidence that additional structure improves fit and forecasts would be required for a larger model, beyond greater complexity alone.

Long-memory volatility and explicit seasonal adjustment have also been combined in forecasting work [47]. Another possible extension involves external covariates and cyclical poles. Persistent cyclical dynamics are accommodated in Gonzaga's wavelet-based Bayesian analysis of a Gegenbauer autoregressive moving-average model with exogenous inputs (GARMAX) [48]. Exogenous predictors, wavelet estimation, and estimated cyclical poles were absent from the current analysis. Those additional structures remain unresolved by its conclusions, which concern the simpler daily forecasting comparison.

CONCLUSION

Over 1990–2025, NASDAQ volatility proxies displayed persistent low-frequency dependence, with little corresponding memory in signed returns. The primary log squared-return series gave a positive estimate, which stayed positive under several sensitivity checks. The genuine long-memory null was nevertheless rejected by Qu's test, while residual dependence remained in the fitted fractional model. A persistent empirical pattern is supported by these findings, but a stable fractional volatility mechanism is not established.

ARFIMA had the smallest errors for average log squared returns at five and 22 trading days in the 2022–2025 holdout. Its advantage over HAR did not survive Holm adjustment, and an advantage over ARMA was not established. Persistent loss differences also weakened inference against the historical mean. These results support ARFIMA as a candidate approximation for this transformed series. They do not establish a reliable advantage over GARCH-family models, on a variance scale, or across alternative evaluation periods and estimation windows. Broader claims about its practical value remain unsupported by the present comparison.

ACKNOWLEDGEMENTS

Access to the NASDAQ Composite data was provided through the Federal Reserve Bank of St. Louis. Publicly documented R software was used for the analysis.

Data availability

FRED provides the index series as NASDAQCOM at https://fred.stlouisfed.org/series/NASDAQCOM. The methodology specifies the sample transformations and evaluation design.

REFERENCES

1. Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987–1007. https://doi.org/10.2307/1912773
2. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1
3. Ding, Z.; Granger, C. W. J.; Engle, R. F. (1993). A long memory property of stock market returns and a new model. Journal of Empirical Finance, 1(1), 83–106. https://doi.org/10.1016/0927-5398(93)90006-d
4. Bollerslev, T.; Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility. Journal of Econometrics, 73(1), 151–184. https://doi.org/10.1016/0304-4076(95)01736-4
5. Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. Journal of Econometrics, 14(2), 227–238. https://doi.org/10.1016/0304-4076(80)90092-5
6. Hosking, J. R. M. (1981). Fractional differencing. Biometrika, 68(1), 165–176. https://doi.org/10.1093/biomet/68.1.165
7. Baillie, R. T.; Bollerslev, T.; Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 74(1), 3–30. https://doi.org/10.1016/s0304-4076(95)01749-6
8. Breidt, F. J.; Crato, N.; de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility. Journal of Econometrics, 83(1-2), 325–348. https://doi.org/10.1016/s0304-4076(97)00072-9
9. Robinson, P. M. (2001). The memory of stochastic volatility models. Journal of Econometrics, 101(2), 195–218. https://doi.org/10.1016/s0304-4076(00)00079-8
10. Gonzaga, A. C. (2013). Seasonal long-memory stochastic volatility. AIP Conference Proceedings, 1558(1), 2410–2413. https://doi.org/10.1063/1.4826027
11. Andersen, T. G.; Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. Journal of Empirical Finance, 4(2–3), 115–158. https://doi.org/10.1016/S0927-5398(97)00004-2
12. Lobato, I. N.; Savin, N. E. (1998). Real and spurious long-memory properties of stock-market data. Journal of Business & Economic Statistics, 16(3), 261–268. https://doi.org/10.1080/07350015.1998.10524760
13. Granger, C. W. J.; Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. Journal of Empirical Finance, 11(3), 399–421. https://doi.org/10.1016/j.jempfin.2003.03.001
14. Diebold, F. X.; Inoue, A. (2001). Long memory and regime switching. Journal of Econometrics, 105(1), 131–159. https://doi.org/10.1016/s0304-4076(01)00073-2
15. Robinson, P. M. (1994). Semiparametric Analysis of Long-Memory Time Series. The Annals of Statistics, 22(1), 515–539. https://doi.org/10.1214/aos/1176325382
16. Robinson, P. M. (1995). Gaussian Semiparametric Estimation of Long Range Dependence. The Annals of Statistics, 23(5), 1630–1661. https://doi.org/10.1214/aos/1176324317
17. Robinson, P. M. (1995). Log-Periodogram Regression of Time Series with Long Range Dependence. The Annals of Statistics, 23(3), 1048–1072. https://doi.org/10.1214/aos/1176324636
18. Gonzaga, A.; Kawanaka, A. (2006). Wavelet-based Bayesian analysis of generalized fractional process. Journal of Signal Processing, 10(3), 199–205.
19. Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. Journal of Financial Econometrics, 7(2), 174–196. https://doi.org/10.1093/jjfinec/nbp001

20. Andersen, T. G.; Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. International Economic Review, 39(4), 885–905. https://doi.org/10.2307/2527343
21. Andersen, T. G.; Bollerslev, T.; Diebold, F. X.; Labys, P. (2003). Modeling and Forecasting Realized Volatility. Econometrica, 71(2), 579–625. https://doi.org/10.1111/1468-0262.00418
22. Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics, 160(1), 246–256. https://doi.org/10.1016/j.jeconom.2010.03.034
23. Gonzaga, A. C. (2017). Asymptotic decorrelation of discrete wavelet packet transform of generalized long-memory stochastic volatility. The Philippine Statistician, 66(2), 1–14. https://psai.ph/docs/publications/tps/tps_2017_66_2_1.pdf
24. NASDAQ, Inc. (2026). NASDAQ Composite [NASDAQCOM]. FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/NASDAQCOM. Accessed 29 October 2026.
25. Gonzaga, A. C. (2017). Estimation of Generalized Long-Memory Stochastic Volatility for High-Frequency Data. International Journal of Applied Engineering Research, 12(9), 1888–1892. https://doi.org/10.37622/ijaer/12.9.2017.1888-1892
26. Geweke, J.; Porter-Hudak, S. (1983). The estimation and application of long memory time series models. Journal of Time Series Analysis, 4(4), 221–238. https://doi.org/10.1111/j.1467-9892.1983.tb00371.x
27. Hurvich, C. M.; Deo, R. S.; Brodsky, J. (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. Journal of Time Series Analysis, 19(1), 19–46. https://doi.org/10.1111/1467-9892.00075
28. Deo, R. S.; Hurvich, C. M. (2001). On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. Econometric Theory, 17(4), 686–710. https://doi.org/10.1017/s0266466601174025
29. Qu, Z. (2011). A Test Against Spurious Long Memory. Journal of Business & Economic Statistics, 29(3), 423–438. https://doi.org/10.1198/jbes.2010.09153
30. Qu, Z.; Kim, C. (2025). SpuriousMemory: Testing True Long Memory Against Spurious Long Memory. R package version 1.0.0. https://doi.org/10.32614/CRAN.package.SpuriousMemory
31. Bai, J.; Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. Econometrica, 66(1), 47–78. https://doi.org/10.2307/2998540
32. Bai, J.; Perron, P. (2003). Computation and analysis of multiple structural change models. Journal of Applied Econometrics, 18(1), 1–22. https://doi.org/10.1002/jae.659
33. Haslett, J.; Raftery, A. E. (1989). Space-Time Modelling with Long-Memory Dependence: Assessing Ireland's Wind Power Resource. Applied Statistics, 38(1), 1–50. https://doi.org/10.2307/2347679
34. Schwarz, G. (1978). Estimating the Dimension of a Model. The Annals of Statistics, 6(2), 461–464. https://doi.org/10.1214/aos/1176344136
35. Hyndman, R. J.; Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software, 27(3), 1–22. https://doi.org/10.18637/jss.v027.i03
36. Ljung, G. M.; Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303. https://doi.org/10.1093/biomet/65.2.297
37. Gonzaga, A. C. (2013). Forecasting seasonal long-memory stochastic volatility. International Journal of Applied Mathematics and Statistics, 51(21), 195–204.
38. Diebold, F. X.; Mariano, R. S. (1995). Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 13(3), 253–263. https://doi.org/10.1080/07350015.1995.10524599
39. Giacomini, R.; White, H. (2006). Tests of Conditional Predictive Ability. Econometrica, 74(6), 1545–1578. https://doi.org/10.1111/j.1468-0262.2006.00718.x
40. Newey, W. K.; West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55(3), 703–708. https://doi.org/10.2307/1913610
41. Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. Econometrica, 59(3), 817–858. https://doi.org/10.2307/2938229
42. Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics, 6(2), 65–70. https://www.jstor.org/stable/4615733
43. Hurvich, C. M.; Ray, B. K. (2003). The Local Whittle Estimator of Long-Memory Stochastic Volati-

lity. Journal of Financial Econometrics, 1(3), 445–470. https://doi.org/10.1093/jjfinec/nbg018
44. Gonzaga, A. C. (2022). Estimation of periodic long-memory GARCH-in-mean model. AIP Conference Proceedings, 2425(1), 420022. https://doi.org/10.1063/5.0081335
45. Granger, C. W. J.; Teräsvirta, T. (1999). A simple nonlinear time series model with misleading linear properties. Economics Letters, 62(2), 161–165. https://doi.org/10.1016/s0165-1765(98)00228-6
46. Gonzaga, A. C. (2024). Bayesian estimation of generalized long-memory stochastic volatility. E3S Web of Conferences, 508, 04014. https://doi.org/10.1051/e3sconf/202450804014
47. Deo, R. S.; Hurvich, C. M.; Lu, Y. (2006). Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. Journal of Econometrics, 131(1–2), 29–58. https://doi.org/10.1016/j.jeconom.2005.01.003
48. Gonzaga, A. (2025). Wavelet-based Bayesian Analysis of GARMAX Model. In 2025 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), pp. 93–98. IEEE. https://doi.org/10.1109/i2cacis65476.2025.11100998