University of Illinois Urbana-Champaign
ECON 503 Econometrics

The Determinants of Price Volatility and Its Consistency Between Crypto and Stock Markets During Past Decade

Hsiang Lee

1. Research Objective

Price volatility plays a central role in financial economics, it shapes investment strategies, market stability, and even broader monetary policy decisions. Understanding the determinants of price volatility is crucial for individual and institutional investors to design lower risk and statistically evidence-based strategies, and for policymakers to regulate and stabilize the financial markets. Based on Random Walk Theory (RWT), since the price is unpredictable by using past data, hence the volatility of price could be inferred that is also unpredictable. However, many researchers showed the opposite evidence to challenge RWT. Jegadeesh & Titman (1993) suggested that there is a "momentum effect" in stock market which keeps price tend to be move in the same way. Similarly, "Volatility Clustering" is high volatility tends to be followed by high volatility and vice versa, which seems to be a common phenomenon among financial markets shown by many researchers such as Thomas Lux (1999). Apart from the lagged terms, there are also various exogenous factors which are correlated with the price volatility. Schwert (1989) indicated that macroeconomic factors like inflation and interest rates have important impacts on price volatility. Besides, trading volume is also relevant for volatility suggested by Timothy J Brailsford (1996). Since there are many opposite evidence of the characteristic of price volatility, this study aims to provide a comprehensive overview by investigating different financial markets and long-term daily data.
More specifically, the following research questions need to be addressed:

- What is the relationship of price volatility to its lagged variable (Is there any "inertia effect") and other exogenous variables?
- Does the characteristic and explanation of price volatility tend to be consistent across different markets and times?

2. Data

This study adopts an empirical research design, as the two main questions listed above, there are two parts of this study. The first part is to build models to research the determinants of price volatility by utilizing time-series data. And the second part is to compare the volatility among different markets and years. The time scope of first part is limited to the period after 2017, when the cryptocurrencies gradually became known by public so that the liquidity was enough to keep the price more stable and accurate, preventing some abnormally large volatility in the first decade of Bitcoin. And the financial assets which will be covered for comparison is S&P 500 from stocks market since stocks is the most widely traded and accessible financial asset by individual investors.

As we all know that the markets characteristics among different financial assets are quite different, some are more volatile than others; however, this study aims to explore an underlying explanation of price volatility and some hidden pattern similarities across different times and markets. If the random walk theory is correct, then price action should be consistent among different times and markets. Secondly, even if we consider the GARCH model and exogenous variables, the explanation will not be highly obvious and effective.

3. Models

3.1. Preliminary Model

The objective of the analysis is to capture the trend and feature of the price volatility. There are two financial assets to be evaluated in this paper, one is BTCUSD and the other is S&P 500. Both in daily data scale, and for the first part, both covers time range from 2017/1/1 to 2023/12/31. GARCH model would be the core econometric framework to be used in this paper due to the objective is volatility. It allows conditional variance to evolve over time as a function of past shocks, making it well suited for this analysis. In the preliminary study, GARCH(1,1), which is the basic form of GARCH, would be implemented. The other variations of GARCH model will be evaluated in the second half of this chapter in order to build a more robust model. GARCH model consists of two components, one is mean equation, the other is variance equation. In the GARCH(1,1), there is only one input required: the difference of log price. The model assumes the mean of the input is zero, and the distribution of residual is normally distributed by default.

The form of GARCH(1,1) is:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

There are three main parameters:

ω (omega) — Long-run average variance (constant term):

Represents the baseline level of variance that the model reverts to in the absence of recent shocks or volatility. A higher omega implies a higher unconditional variance.

α_1 (alpha1) — Short-term shock sensitivity (ARCH effect):

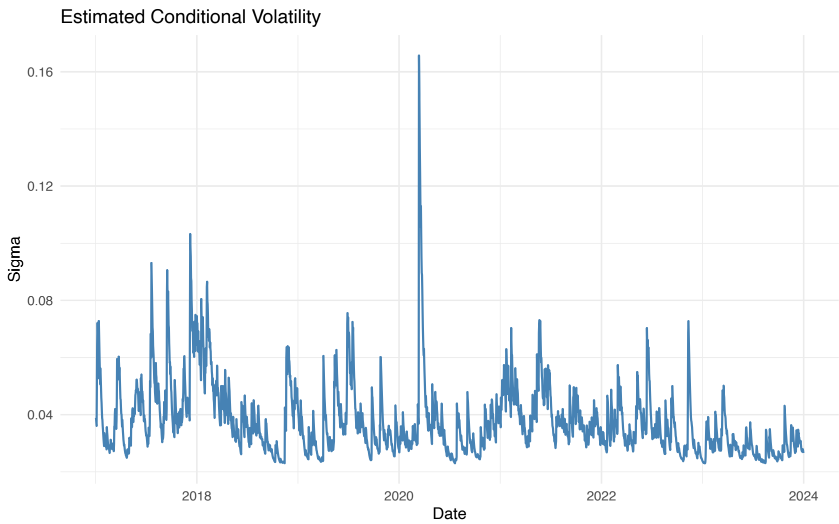
Measures the immediate impact of a new shock (the squared residual from the previous period) on current volatility. A larger alpha1 indicates that volatility reacts more strongly to recent news or shocks.

β_1 (beta1) — Volatility persistence (GARCH effect):
Captures the effect of past volatility on current volatility. A larger beta1 implies that volatility is more persistent and takes longer to die out after a shock.

The first model found is:

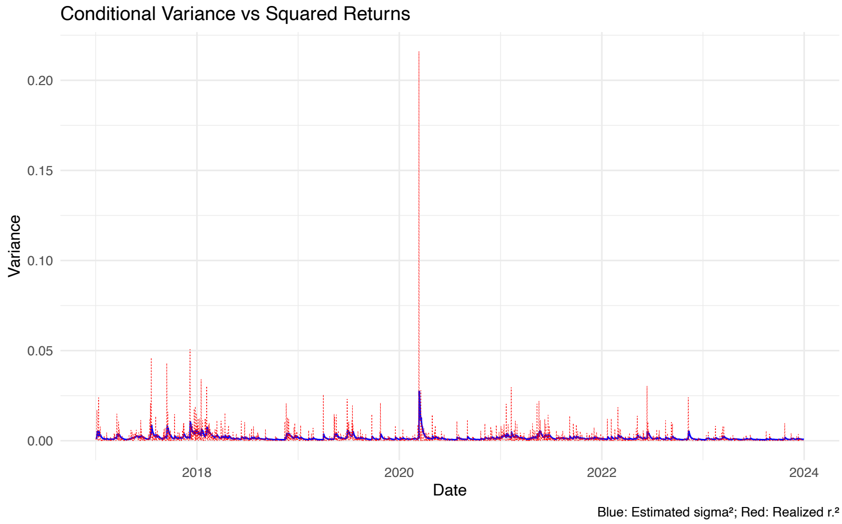
$$\sigma_t^2 = 0.000082 + 0.121521\varepsilon_{t-1}^2 + 0.833846\sigma_{t-1}^2$$

The sum of (alpha1 + beta1) is often used to assess the persistence of volatility:
If (alpha1 + beta1) is close to 1, shocks to volatility decay slowly, indicating long memory in volatility. In our first model, the value of the sum is 0.955 which shows a strong evidence of volatility clustering. We could implement two graphs to see the volatility over time:



The time series plot of the estimated conditional standard deviation (sigma) reveals clear evidence of volatility clustering, a hallmark of financial return series. Significant spikes in volatility are observed around early 2020, which may correspond to major market disruptions resulting from COVID-19 pandemic. Periods of elevated volatility are followed by periods of relative calm, consistent with the GARCH model's ability to capture dynamic changes in volatility over time.

The comparison between the estimated conditional variance ($sigma^2$) and the realized squared returns (r_t^2) demonstrates that the model effectively captures the timing of volatility spikes. While the realized squared returns display sharper and more irregular movements due to their inherent randomness, the conditional variance tracks the underlying volatility dynamics smoothly. Notably, the model responds appropriately to major volatility shocks, reinforcing the adequacy of the GARCH(1,1) specification in modeling conditional heteroskedasticity. (Annotation: the evaluation of this graph is to see the two values are synchronized with each other, not by the value itself because squared return is the proxy of estimated variance, two values are inherently different.)



In conclusion, the sigma plot exhibits typical volatility clustering, while the overlaid plot shows that the model captures the timing of major volatility events effectively, despite the inherent noise in realized returns. Both plots provide strong preliminary validation of the GARCH(1,1) model.

3.1.1. Testing the Model

The evaluation of GARCH model is very different from multiple linear regression because there are no explanatory variables and R-squared to be discussed. There will be five tests: autocorrelation of standardized residuals, the ARCH effect, stability of parameters, the symmetry of positive and negative shocks, and the goodness of fit in standardized residuals.

Significance of Parameters

Parameter	Estimate	Std. Error	t-value	p-value
ω	0.000082	0.000013	6.3494	0.0000
α_1	0.121521	0.016067	7.5633	0.0000
β_1	0.833846	0.018435	45.2317	0.0000

Robust Standard Errors

Parameter	Estimate	Std. Error	t-value	p-value
ω	0.000082	0.000026	3.1095	0.0019
α_1	0.121521	0.036355	3.3427	0.0008
β_1	0.833846	0.033650	24.7796	0.0000

The p-value of three parameters are really low, meaning their estimation are strongly significant and they do deliver contribution to the conditional variance. Even with robust standard errors, which means some allowance for model misspecification, all these three parameters are still significant.

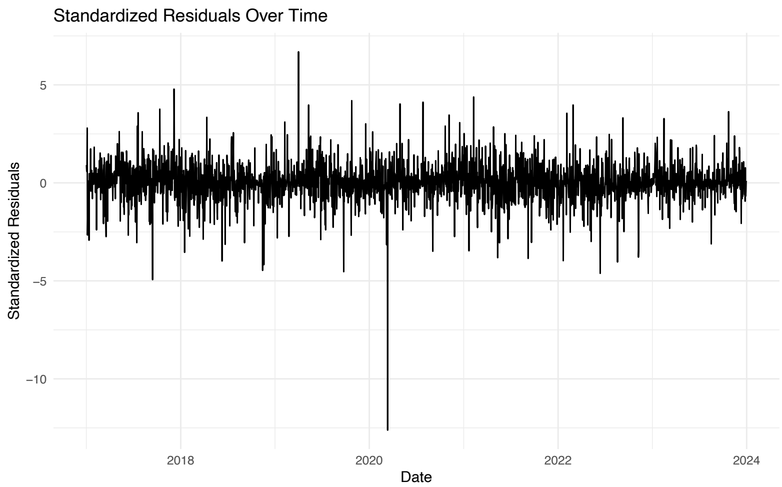
Information Criteria

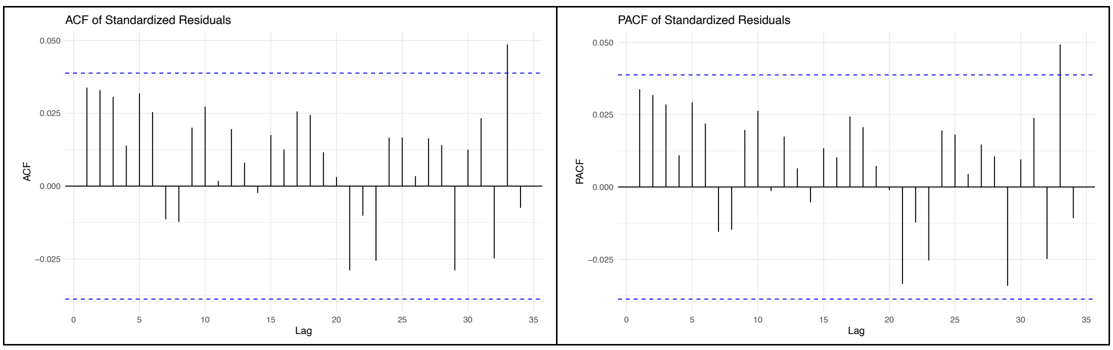
Akaike	-3.7938
Bayes	-3.7870
LogLikelihood	4851.511

We will compare these values with below updated models. The higher the value of LogLikelihood and the lower value of AIC/ BIC, the better fit model is.

Autocorrelation

To test for the autocorrelation of standardized error, first we can plot its sequence along the data time period and we want it to show the characteristic of white noise. It looks pretty close to white noise except for one abnormal peak when the pandemic started in 2020/3. Then, we plot the ACF and PACF graph of the standardized residual to see how correlated the lag terms are. This time we can see almost all the lag terms are within the dash line, which means insignificance of correlation. This could also indicate that our mean equation, which is zero, is suitable for this model.





Weighted Ljung-Box Test on Standardized Residuals

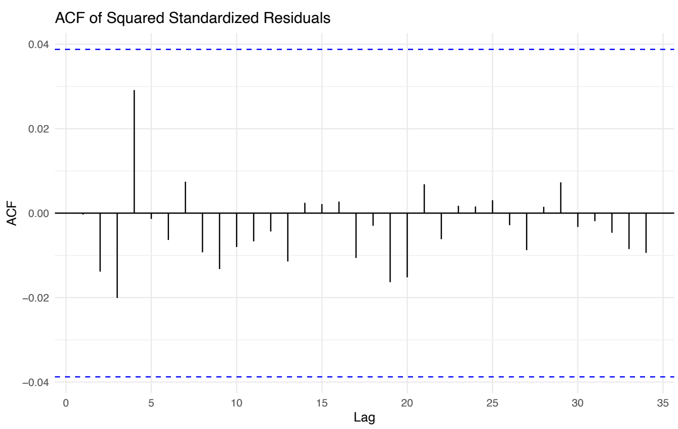
Lag term	statistic	p-value
Lag[1]	2.924	0.08728
Lag[2*(p+q)+(p+q)-1][2]	4.311	0.06221
Lag[4*(p+q)+(p+q)-1][5]	7.302	0.04378

H0: No serial correlation

After the visual evaluation, we will perform the weighted Ljung-Box test to formally determine if the residual meets the criteria or not. The null hypothesis for Ljung-Box test is that the data are independently distributed up to a certain lag h; that is, all autocorrelations up to lag h are zero. Hence, the result we want to get is p-value > 0.05. In the table we could see that Lag 1 and Lag 2 passed the test, however, Lag 5 is slightly below the threshold (p-value≈ 0.44). This suggests that the mean equation is mostly adequate, although there may be minor model misspecification at higher lags. It is a minor problem, however, we will observe if it passes the test in updated models.

ARCH Effect

Aside from the test for mean equation, there is another important assumption about the variance equation. Similarly, to detect the autocorrelation in the conditional variance, we first plot the ACF graph of it, and the outcome shows every lag terms are within up and low bound. We still perform the Ljung-Box test on standardized squared residuals. This test is important because the presence of autocorrelation in squared residuals would indicate that the GARCH model has not fully captured the volatility clustering phenomenon. This time, p-value of three lag terms are much higher than 0.05, not rejecting H0. In GARCH model diagnostics, passing the Ljung-Box test on both residuals and squared residuals is essential to validate the adequacy of the conditional mean and variance equations.



Weighted Ljung-Box Test on Standardized Squared Residuals

Lag term	statistic	p-value
Lag[1]	0.0003166	0.9858
Lag[2*(p+q)+(p+q)-1][5]	1.8840212	0.6461
Lag[4*(p+q)+(p+q)-1][9]	2.8854336	0.7777

H0: No serial correlation

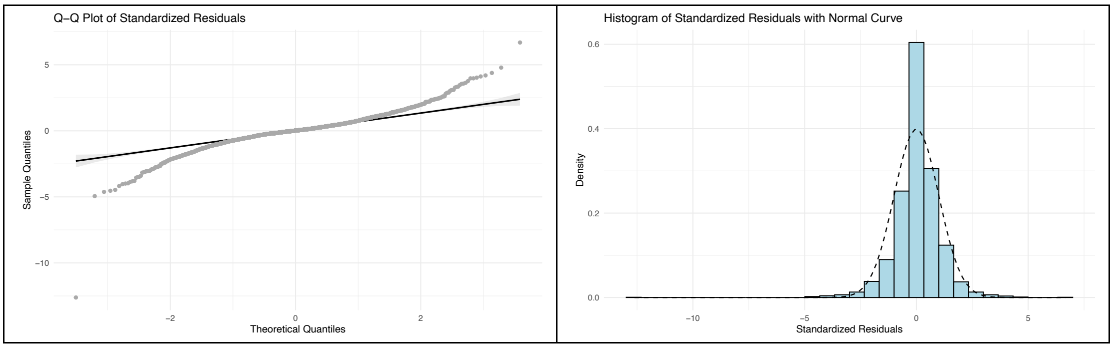
Weighted ARCH LM Tests

Lag term	statistic	shape	scale	p-value
ARCH Lag[3]	1.031	0.500	2.000	0.3099
ARCH Lag[5]	2.770	1.440	1.667	0.3250
ARCH Lag[7]	3.014	2.315	1.543	0.5109

Additionally, we could also use ARCH-LM test, which is designed to detect the presence of autoregressive conditional heteroskedasticity (ARCH effects). The null hypothesis is that no ARCH effects remain in the residuals (the variance is constant after conditioning). If the p-value>0.05, we fail to reject H0, indicating that the conditional variance has been adequately modeled. The p-value of three lag terms are much higher than 0.05. Through both Ljung-Box test and the ARCH-LM test, we could conclude the model has fully captured conditional heteroskedasticity.

Goodness-of-Fit Test

In the default settings for GARCH model, we have to assume a particular distribution of its residual beforehand. After fitting a GARCH model, it is crucial to evaluate whether the standardized residuals meet the assumed distribution, which is normal distribution in the base model. Two graphical diagnostic tools employed are Q-Q plot and histogram of standard residuals.



The Q-Q (quantile-quantile) plot visually assesses whether the standardized residuals align with the theoretical quantiles of the assumed distribution. Deviations from the reference line indicate departures from the hypothesized distribution, such as fat tails or skewness. We can see that only the middle portion of the data is aligned with the reference line. However, to the right and the left, the deviation is abnormally high, indicating the presence of heavy tails and asymmetry. In addition, the histogram of standardized residuals is another way that provides an intuitive visualization of the empirical distribution. Overlaying the theoretical density curve allows for a direct comparison between the observed and expected shapes. The outcome graph exhibits asymmetry, a much higher peak and heavier tails compared to the normal distribution. Hence, we can preliminarily conclude that normally distribution is not an appropriate assumption. Such behavior is commonly observed in financial time series data and indicates that using a distribution with heavy tails such as Student's t-distribution might be more suitable.

Adjusted Pearson Goodness-of-Fit Test:

group	statistic	p-value(g-1)
20	366.0	5.029e-66
30	384.7	9.447e-64
40	392.3	6.621e-60
50	403.2	3.661e-57

Lastly, the formal statistical test we will perform is Adjusted Pearson Goodness-of-Fit test. This test evaluates whether the standardized residuals conform to the assumed distribution by comparing observed frequencies within intervals to expected frequencies under the model. A significant test result (small p-value) suggests that the residuals deviate systematically from the hypothesized distribution. The outcome shows extremely low p-values (10^-60) for each group, indicating significant departure from assumption.

Parameter Stability

To test the stability of parameters in GARCH model, we use Nyblom Stability test to make sure the parameters are constant over time. The test evaluates both joint and individual stability of the estimated parameters. A joint statistic is to test the overall stability of all parameters simultaneously. The null hypothesis is that the parameters are stable over time. And the H1 is at least one parameter exhibits time variation. The computed statistics are compared against asymptotic critical values (at 10%, 5%, and 1% levels). If the statistic exceeds the critical value, the null hypothesis is rejected, indicating instability. If the parameters are stable, the GARCH model is considered reliable across the sample period. If instability is detected, it suggests that the volatility dynamics might have changed over time, and a time-varying model may be more appropriate.

Nyblom stability test

Joint Statistic	0.816
ω	0.5031
α_1	0.2203
β_1	0.3882

Asymptotic Critical Values (10% 5% 1%); Joint Statistic: 0.846, 1.01, 1.35; Individual Statistic: 0.35, 0.47, 0.75

We can see the statistic of omega and beta1 are slightly higher than their 10% significance value, but lower than 5% and 1% threshold respectively, indicating a mild evidence of instability. On the other hand, joint statistic and alpha1 are lower than 10% significance value. Given that the magnitude of instability is not substantial and the main dynamics (captured by alpha1 and beta1) remain stable, the overall model specification is considered adequate for describing the conditional variance process.

Asymmetry Effect

The Sign Bias test examines whether positive and negative shocks have asymmetric effects on volatility, something that the standard GARCH(1,1) model does not capture. The null hypothesis is that positive and negative shocks have symmetric effects on volatility. And H1 is asymmetrical volatility impact. If the sign bias is significant (p-value < 0.05), it indicates that the current GARCH model does not adequately capture the asymmetric volatility response to shocks. If asymmetry is detected, it suggests that the model may be mis-specified. More advanced models such as EGARCH which explicitly model asymmetry might be necessary.

Sign Bias Test

	t-value	p-value
Sign Bias	1.533298	0.1253
Negative Sign Bias	0.830641	0.4063
Positive Sign Bias	1.656983	0.0976
Joint Effect	3.659746	0.3006

The results of the Sign Bias Test indicate no significant evidence of asymmetry in the standardized residuals. Both the individual and joint tests produce p-values greater than 0.05, suggesting that the standard GARCH(1,1) model adequately captures the impact of shocks on volatility. Although the positive sign bias is marginally close to significance at the 10% level, it does not provide strong evidence of asymmetric effects.

3.2. Updated Models

GARCH(1,1) with t-distribution

The biggest issue with standard GARCH(1,1) model with normal distribution is the strong evidence that the normal distribution assumption is highly inappropriate for modeling the conditional distribution of returns in this dataset. To address it, the model is re-estimated under the Student's t-distribution which allows for heavier tails. Switching distribution led to substantial improvement in model fit, as evidenced by higher p-values in the goodness-of-fit test (in the order of 10^-5 compared to 10^-60). A higher log-likelihood also indicates that the t-distribution more accurately captures the heavy-tailed nature of Bitcoin returns. In addition, the model still captures ARCH effect perfectly. All of the statistics of sign bias test are also passed.

However, the Ljung-Box test on standardized residuals of lag 2 and lag 5 show the effect of autocorrelation, even the p-value of lag 1 is marginally higher than 0.05. This suggests that autocorrelation remains in the residuals after accounting for conditional heteroskedasticity and fat tails. The residual dependence may stem from misspecified ARMA structure or omitted explanatory variables. The Nyblom stability test revealed significant parameter instability in the GARCH(1,1)-t model, in contrast to the normal specification which exhibited stable parameters. Two most important parameters, alpha and beta, both demonstrate an instability. We will keep conducting different types of GARCH model to see if it could be solved.

ARMA-GARCH

After fixing the main issue with assumed t-distribution, we still have a few minor problems to solve. Because the Ljung-Box test on the standardized residuals indicated mild autocorrelation, we attempted to address this by estimating an ARMA-GARCH model with an ARMA mean equation. The goal was to capture potential autocorrelation in the residuals via the mean equation.

The model selection procedure ultimately identified ARMA(0,0) as the most appropriate mean specification, suggesting that no autoregressive or moving average terms significantly improved the model fit. This implies that the mild autocorrelation at Lag 5 and some spikes stay above or below the dashed line from ACF and PACF plots may have been a result of random noise rather than a persistent structure. Therefore, the mean process is best modeled as white noise, and the original GARCH(1,1) specification remains adequate in this regard.

GARCH-X

To extend the standard GARCH(1,1) model, a GARCH-X model was considered in order to examine whether external factors have explanatory power over the conditional volatility of the asset. The rationale behind using exogenous variables lies in the economic intuition that certain macroeconomic or market-specific indicators may influence volatility beyond what is captured by past squared returns and past variances. This can enhance the model's explanatory power and improve volatility forecasts if the external variables are indeed relevant. The three economic factors we will perform in this study are:

DFF: Federal Funds Effective Rate (DFF) is the interest rate at which depository institutions lend reserve balances to other institutions overnight on an uncollateralized basis. It is a key monetary policy tool used by the Federal Reserve to influence liquidity and overall economic activity. Changes in the federal funds rate often signal shifts in monetary policy stance. This is a daily dataset.

CPI: Consumer Price Index for All Urban Consumers: All Items Less Food and Energy in U.S. City Average (CPILFESL) is the indicator that reflects the core inflation trend and is closely monitored by policymakers for its role in assessing price stability. This is a monthly data, we first calculate the percentage change for each month and expand monthly data to daily frequency by forward-filling values so that we could meet the criterion of GARCH-X which needs the response and explanatory variables to be aligned with.

GDP: Real Gross Domestic Product (GDPC1) is the value of all goods and services produced in the economy, adjusted for inflation. It captures the overall economic output and growth. This is a quarterly data, similarly, we first calculate the percentage change for each quarter and expand quarterly data to daily frequency by forward-filling values.

The estimation results show that all three exogenous regressors in the GARCH-X specification fail to achieve statistical significance (p-values of three variables are far larger than 0.05), indicating that they do not contribute meaningfully to the conditional variance dynamics. The loglikelihood exhibits a minor improvement from 5167.5 to 5176.1, which is barely negligible. What's worse, all three lag terms of standardized residuals show the effect of autocorrelation. While Ljung-Box test on standardized squared residuals and ARCH-LM test are still easily meet the expected results. About the parameter stability test, the joint statistic (5.8241) is much higher than its critical value (1.89). All statistics of sign bias test are greater than 0.05, indicating there is no significance of sign bias. The Goodness-of-Fit test demonstrates the most well suited to assumed t-distribution. The p-value is much higher again, even surpassed 0.05 in group of 30, 40, and 50. While the inclusion of exogenous variables enhances the model's fit, it appears to compromise both the stability of the parameters and the ability to fully capture serial dependence. These results highlight the trade-off between improving distributional assumptions and maintaining model robustness.

Model Comparison

Because the autocorrelation of residuals in GARCH-X may lead to biased forecasts of volatility, which is more intolerable than the mild imperfection in Goodness-of-Fit test. Additionally, the complexity and increased number of parameters in GARCH-X make it difficult to be the best candidate. While the GARCH(1,1) with t-distribution substantially improves the fit to the empirical distribution, it fails to eliminate residual autocorrelation and exhibits parameter instability. In contrast, the standard normal GARCH(1,1) is more stable and passes most diagnostic tests, despite poorer goodness-of-fit.

BTC GARCH Model Comparison

	GARCH	GARCH t	GARCH-X
LogLikelihood	4851.511	5167.536	5176.119
Residuals	p-value= 0.04378	p-value= 0.02244	p-value= 0.01047
ARCH LM	p-value= 0.5109	p-value= 0.6301	p-value= 0.6753
Goodness-of-Fit	p-value= 9.447e-64	p-value= 1.084e-05	p-value= 0.14491
Joint Statistic	0.816	4.9842	5.8241
Joint Effect	p-value= 0.30063	p-value= 0.38507	p-value= 0.75902

*Ljung-Box test on standard residuals and ARCH LM will choose value of last lag.

*Goodness-of-fit will choose the value of group=30.

*Nyblom Stability test will choose joint statistic.

10% asymptotic critical values are: 0.846, 1.07, 1.89

*Sign Bias test will choose joint effect.

4. Comparison between Different Assets and Time Periods

4.1. GARCH Model for S&P 500

As for the volatility capture of S&P 500, we use its daily return in the same time period and conduct the same procedures of model development and statistical tests. In the three models we built, which are GARCH(1,1) with normal distribution, with t-distribution, and ARMA-GARCH, all three perfectly demonstrated strong evidence that there is no autocorrelation and ARCH effect.

However, all three models presented strong joint effect of sign bias test, indicating asymmetry in this particular financial asset. As for goodness of fit, the t-distribution unsurprisingly performed the best but at a cost of extremely unstable parameters. As for the mean equation, its optimization is ARMA(4,4). This suggests that the return series contains short-term autocorrelation and moving average dynamics, possibly arising from market microstructure effects, delayed information processing, or short-term trading patterns. However, due to the extremely unstable parameters and over complex model setting like the previous GARCH-X model, we still choose the GARCH(1,1) with normal distribution which delivers the least degree of parameters instability as the most balanced model for S&P 500.

S&P 500

	GARCH	GARCH t	ARMA-GARCH
LogLikelihood	5777.758	5829.38	5838.593
Residuals	p-value= 0.9838	p-value= 0.9876	p-value= 1.0000
ARCH LM	p-value= 0.6943	p-value= 0.6901	p-value= 0.6743
Goodness-of-Fit	p-value= 4.052e-11	p-value= 1.797e-06	p-value= 1.350e-06
Joint Statistic	1.7841	44.4451	50.0962
Joint Effect	p-value= 0.0001637	p-value= 0.000278	p-value= 0.000127

*Ljung-Box test on standard residuals and ARCH LM will choose value of last lag.

*Goodness-of-fit will choose the value of group=30.

*Nyblom Stability test will choose joint statistic.

10% asymptotic critical values are: 0.846, 1.07, 2.69

*Sign Bias test will choose joint effect.

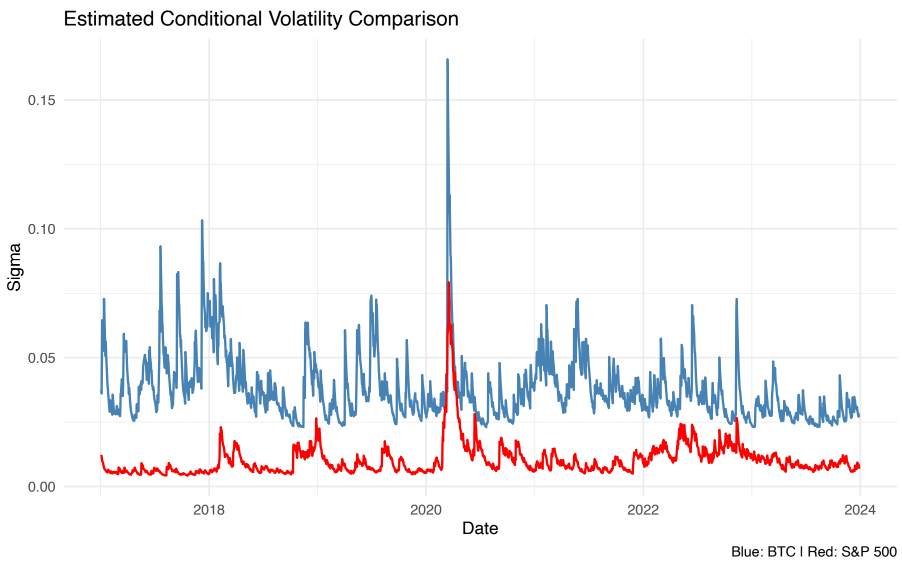
4.2. Comparison Between Two Assets

Since the most balanced GARCH models for BTC and S&P 500 are the same, which are GARCH(1,1) with normal distribution, it is feasible that we could directly compare their parameters.

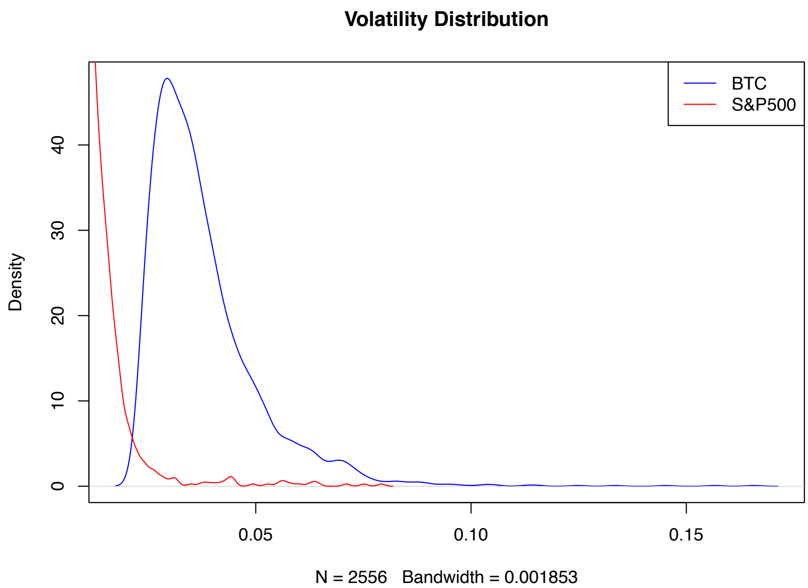
GARCH(1,1) Comparison

	α_1	β_1	$\alpha_1+\beta_1$	Mean(sigma(fit))
BTC	0.121521	0.833846	0.955367	0.0378981
S&P 500	0.180704	0.802443	0.983147	0.0102997

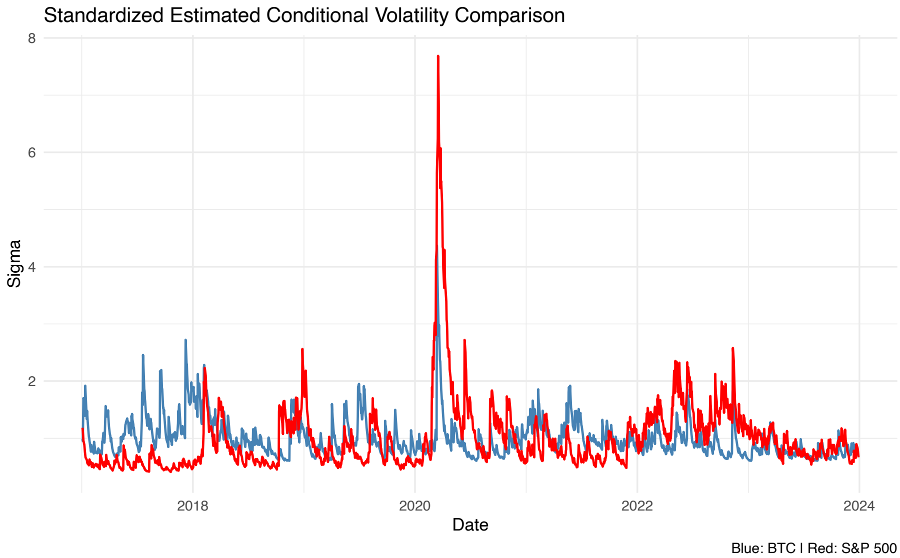
S&P500 exhibits a higher alpha1, indicating a greater sensitivity to recent shocks. In contrast, BTC has a marginally higher beta1, implying more persistent volatility. The sum of alpha1 and beta1 is close to 1 for both financial assets, confirming the long memory of volatility. Lastly, the average conditional volatility of BTC is significantly higher than S&P 500, which could also shown on the below graphs.



S&P 500 maintains relatively lower and more stable levels of conditional volatility, with fewer extreme spikes. This reflects the fundamental difference in market behavior: BTC, as a less mature and more speculative asset, is subject to greater uncertainty than the more established equity market represented by the S&P 500.

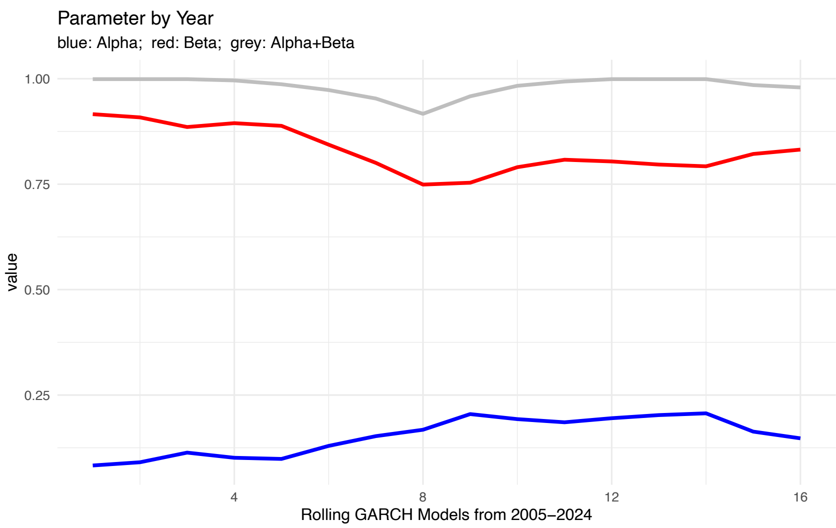


This kernel density plot compares the distribution of estimated sigma values for BTC and the S&P 500. The BTC distribution is wider, meaning the occurrence of higher volatility. In contrast, S&P 500 has a much sharper and more concentrated distribution near zero, reflecting more stable market conditions. The greater dispersion of BTC's volatility distribution once again highlights its higher risk profile.



However, in this plot, the sigma for both assets have been standardized by dividing each series by its own mean. The aim is to compare the relative volatility dynamics instead of their absolute levels. Interestingly, after standardization, the gap between BTC and S&P 500 becomes narrower, and the patterns of spikes are more aligned, especially during turbulent periods like early 2020. The relative spike in S&P 500 volatility exceeds that of Bitcoin, indicating a more severe deviation from its historical norm. This suggests that while Bitcoin is inherently more volatile, the equity market exhibited greater relative instability during periods of systemic shocks.

4.3. Comparison Among Different Periods



To analyze the difference among different time periods, we conduct GARCH models through rolling window approach. The dataset for a model is 1,250 days, and refit every 250 days. In the end, this approach could generate 16 GARCH models by using overlapping time periods through 20 years (2005-2024). This figure presents the evolution of the GARCH(1,1) model parameters estimated. The grey line representing alpha1 + beta1 consistently remains close to 1 throughout the entire sample period, suggesting that volatility shocks are highly persistent over time. This is consistent with financial time series behavior, especially in developed markets. Across all periods, the value of beta1 is significantly higher than alpha1, indicating that volatility is primarily driven by its past values rather than by immediate shocks. While the general pattern is stable, there is a noticeable dip in both beta1 and alpha1 + beta1 around the 8th window (likely around 2015–2017 depending on my starting point). This suggests a brief period of reduced volatility clustering or market turbulence. To sum up, the parameter evolution plot confirms that the volatility behavior of the asset is structurally consistent over time, with high persistence and low sensitivity to recent shocks. The results support the hypothesis that while market conditions fluctuate, the underlying GARCH dynamics remain stable, reinforcing the appropriateness of GARCH-type models for long-term volatility modeling.

5. Conclusion

After establishing various versions GARCH models and a series of thorough testing for both BTC and S&P 500, the GARCH(1,1) with normal distribution turns out to be the most balanced model for each asset. To our surprise, riskier financial assets such as BTC did not demonstrate any obvious asymmetry in positive and negative shocks in all the models. The significance of two main parameters in all models, alpha1 and beta1, is the main evidence that the volatility clustering do exist. And the sum values are always close to 1 in every case, indicating persistent volatility. Hence, at this time, we can conclude that volatility could be forecasted based on its own past behavior and short-term shocks. However, aside from its lag terms, the GARCH-X model showed that macroeconomic exogenous variables are not able to deliver any explanatory power. On the other hand, while the magnitude of price volatility naturally varies across different time periods and financial products, a different picture emerges when we standardize the volatility series to compare their relative movements. After standardization, we observe a striking similarity between different products within the same period. Moreover, even though the absolute level of volatility changes over time for the same product, the GARCH model parameters remain remarkably consistent across different time segments. These two comparisons might suggest underlying similarities beneath the surface of changing dynamics.

References

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1), 65–91.

Lux, T. (1999). The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. *Journal of Economic Behavior & Organization*, 33(2), 143–165.

Schwert, G. W. (1989). Why does stock market volatility change over time? *The Journal of Finance*, 44(5), 1115–1153.

Brailsford, T. J. (1996). Empirical relationships between trading volume, returns and volatility. *Accounting & Finance*, 36(1), 89–111.