

Establishing an ARMA-GARCH model fit for volatility effect of Climatic Variables on Tea Production in Tea Zones in Kenya

Consolata A. Muganda¹, Sewe Stanley², Winnie Mokeira Onsongo³

^{1,2}Department of Mathematics and Actuarial Science, The Catholic University of Eastern Africa, Kenya

³Department of Statistics and Actuarial Science, University of Ghana, Ghana.

Abstract: The influence of many climatic variables such as NDVI, minimum and maximum humidity, rainfall, minimum and maximum temperature, and solar radiation make tea production volatile. The clustering of low and high volatility periods values makes determining a suitable GARCH and ARIMA model difficult. Climatic variables are risk measures useful in understanding tea production data. A proper understanding of variables to help monitor and forecast the volatility in tea production output is paramount in applied statistics. The existence of affirmative consent on the standard performance of GARCH(1,1) can be misleading due to variation in data volatility. The nonlinear nature of tea production and climatic variables creates everlasting interest to scholars to model a forecast of future tea production based on the volatile climatic conditions. We use Box and Jenkins model to outline 63 combinations of ARMA(m,n)-GARCH(p,q) models in tables with m and n are either 0, 1, or 2. We use AIC, BIC, and LogL criteria to select the best model. The results based on the rubric indicated that the ARMA(1,1)-GARCH(2,2) is the suitable model.

Keywords: ARMA, Climatic Variables, GARCH, Model fit, Tea Production, Volatility Effect.

I. INTRODUCTION

Tea production is volatile due to influence from many factors such as climatic variability, among other factors [9]. Tea production yield presents volatile clusters representing high and low periods of volatility. Identification of generalized autoregressive condition heteroscedasticity (GARCH) order of tea production and climatic variability is a challenge due to data fluctuation in tea production and climatic conditions. GARCH model has predominantly been used in finance to model conditional volatility of the stock in time series.

Volatility is a measure of risk [7, 13]. Thus, volatility is a measure of conditional standards of underlying asset return in finance. In line with this notion, we can estimate volatility given the condition of success on some set standards [13, p.~5]. Thus, modeling volatility of given variables requires a proper understanding of yield like tea production given various climatic conditions for a period.

In order to monitor and forecast the volatility in tea production output with interest in climatic variable data requires modeling to help us understand the data. The selected models help us decide on the volatile cluster. Many GARCH models exist, with many giving affirmative consent on the performance of GARCH(1,1), hence making it a standard model. This can be misleading due to variation in data volatility. Thus, the ARIMA model has been proposed by many existing studies based on Box and Jenkin to address the forecasting problems [8]. However, the ARIMA models are limited to identifying linear relationships based on the current and past time-series data. Tea production and climatic variables are nonlinear, which creates scholars' interest, prompting further research to model a forecast of future tea production based on the volatile climatic conditions.

This article therefore proposed the selection and approximation of the GARCH model to predict future tea production. The full text is organized as follows: Section 1 presents a brief introduction. Section 2 summary of the methodology. Section 3 presents model fitting, analysis and results. Section 4 concludes the research.

II. METHODOLOGY

Three-time series models exist, Autoregressive Moving Average (ARMA), GARCH, and Autoregressive Conditional Heteroscedasticity (ARCH). Box and Jenkins [3] in the mid 20th century proposed ARIMA(m,D,n) models where autocorrelation term is m, differencing elements is D and moving average term is D. The *I* in ARMA differentiates when the series is not stationary. An assumed stationary series is model via three classes of time series process: autoregressive (AR), moving-average (MA), and mixed autoregressive and moving average (ARMA). The latter has found usage in the proposed works. AR(m) is expressed as [6];

$$\gamma_t = \mu + \phi_1\gamma_{t-1} + \dots + \phi_m\gamma_{t-m} + \mu_t \quad (1)$$

MA(n) is expressed as

$$\gamma_t = \mu + u_t + \theta_1u_{t-1} + \dots + \theta_nu_{t-n} \quad (2)$$

where $u_t: t = 1,2,3$ is white noise disturbance term such that $E(u_t) = 0$ and $var(u_t) = \sigma^2$.

A combination of AR(m) and MA(n) model forms ARMA(m,n) model. AR(m) and MA(n) models require many data structures when modeling separately; thus making these models complex. Box, Jenkins and Reinsel [12] introduced ARMA model as a combination of AR(m) and MA(n) models to handle the dependencies in the series. ARMA(m,n) model is represented as [6]

$$\gamma_t = \mu + \phi_1\gamma_{t-1} + \dots + \phi_m\gamma_{t-m} + \mu_t + u_t + \theta_1u_{t-1} + \dots + \theta_nu_{t-n} \quad (3)$$

We summarize Eq. (3) to Eq. (4) as

$$\gamma_t = C + \sum_{i=1}^m \phi_i\gamma_{t-i} + \sum_{j=1}^n \theta_j\varepsilon_{t-j} \quad (4)$$

γ_t is the dependent variable, which in the proposed work is tea production, ϕ_i are the autoregressive parameter components of order m , θ_j are parameters of moving average component of order n and ε_t is error time at time t . m and n are non-negative integer.

ARCH and GARCH model are popular two time-varying volatility models among researchers. ARCH models predicts the conditional variance of dependent variable return series and is given by

$$\begin{aligned} \gamma_t &= C + \varepsilon_t \\ \varepsilon_t &= z_t\sigma_t \end{aligned} \quad (5)$$

where γ_t is the observed dependent variable data series, C is a constant value, ε_t is residual, z_t is the standardized residual with independently and identically distributed with $\mu = 0; var = 1; \sigma_t = \sqrt{|Var(y_t|x_t)|}$. The ARCH model is not described in detail in this work since it did not inform the analysis of the tea production and climatic variables. The ARCH model's need for estimating a large number of parameters makes it unsuitable in this work [2, 10]. Instead, the GARCH model is used due to its need for few parameters compared to ARCH model [10]. The general form of the GARCH(p,q) model is written as

$$\sigma_t^2 = \eta + \sum_{i=1}^p \beta_i\sigma_{t-i}^2 + \sum_{j=1}^q \alpha_j\varepsilon_{t-j}^2 \quad (6)$$

where σ_i and ε_i are defined in (5), η is the long-run volatility and is such that $\eta > 0$, $\beta_i \geq 0$; $i = 1, \dots, p$ and $\alpha_j \geq 0$; $j = 1, \dots, q$. Suppose $\beta_i + \alpha_j < 1$, the GARCH(p,q) model becomes stationary and yields no results as covariance is stationary.

GARCH models have found numerous usage in determining conditional volatility in financial data. Thus, researchers used GARCH to inform them on financial decisions [1, 2]. Combination of ARMA-GARCH models have also found numerous usage in many areas highlighted in the following existing state-of-the-art scholarly works [4, 5, 11, 14, 15]. However, non of the existing literature has presented ARMA-GARCH model forecasting volatility of tea production based on climatic variables in tea zones in Kenya.

III. EXPERIMENT AND FINDINGS

GARCH processes model the volatility, and ARMA process models series of Tea production and climatic variables. GARCH process alone assumes the series data has no autocorrelation, no seasonal effect, and no drift. ARMA model incorporates all these, that is, assumes data has seasonality, drift, and autocorrelation. We use both GARCH and ARMA since the data contains all the three characteristics based on the previous work [8]. We try different models such as ARMA(0,0), ARMA(1,0), ARMA(0,1) ... ARMA(m,n). The GARCH analysis is performed in the order listed in Table 1 using the R language code. However, we noted that for $(n + m) > 2$, the GARCH order ARMA(m,n) GARCH(p,q) becomes singular (see Test Code 43 \geq in Table 2).

Table 1: Garchorder and armaorder combination for GARCH analysis presented in Tables

Test Code	Armorder		Garchorder		Test Code	Armorder		Garchorder	
1	0	0	0	0	33	0	2	1	2
2	0	0	0	1	34	0	2	2	1
3	0	0	1	0	35	0	2	2	2
4	0	0	1	1	36	2	0	0	0
5	0	0	1	2	37	2	0	0	1
6	0	0	2	1	38	2	0	1	0
7	0	0	2	2	39	2	0	1	1
8	0	1	0	0	40	2	0	1	2
9	0	1	0	1	41	2	0	2	1
10	0	1	1	0	42	2	0	2	2
11	0	1	1	1	43	1	2	0	0
12	0	1	1	2	44	1	2	0	1
13	0	1	2	1	45	1	2	1	0
14	0	1	2	2	46	1	2	1	1
15	1	0	0	0	47	1	2	1	2
16	1	0	0	1	48	1	2	2	1
17	1	0	1	0	49	1	2	2	2
18	1	0	1	1	50	2	1	0	0
19	1	0	1	2	51	2	1	0	1
20	1	0	2	1	52	2	1	1	0
21	1	0	2	2	53	2	1	1	1
22	1	1	0	0	54	2	1	1	2
23	1	1	0	1	55	2	1	2	1
24	1	1	1	0	56	2	1	2	2
25	1	1	1	1	57	2	2	0	0
26	1	1	1	2	58	2	2	0	1
27	1	1	2	1	59	2	2	1	0
28	1	1	2	2	60	2	2	1	1
29	0	2	0	0	61	2	2	1	2
30	0	2	0	1	62	2	2	2	1
31	0	2	1	0	63	2	2	2	2
32	0	2	1	1					

A. Model Fitting

After defining the ARMA-GARCH order, we proceed to specify the model. We generate 63 models based on the conditional variance specification in ARMA(m,n)-GARCH(p,q) models where m and n are either 0,1 or 2 and a similar p and q. We generated and compared the models based on criteria; AIC and BIC, and LogL to select the best model. We highlighted these comparisons in Table 2. We provide a summary of the best models in Table 3. The ARMA(1,1)-GARCH(2,2) model presented the best model based on AIC, and LogL, while ARMA(1,0)-GARCH(1,2) provided the best model based on BIC. Overall, the ARMA(1,1)-GARCH(2,2) is the suitable model based on the rubric investigated in Table 3. The ARMA(1,1)-GARCH(2,2) has the lowest AIC with the highest LogL; its BIC is also very low due to a lack of significant difference between those presented in the same table.

Table 2: Garch analysis for the Armaorder and Garchorder in the test code 1 to 64 in Table 1. Lower AIC and BIC are desirable since it shows the model is closer to the truth. Maximum Log-likelihood (LogL) is desirable since it shows superior performance. ... is Singular

Test Code	IC		LogL	Test Code	IC		LogL
	AIC	BIC			AIC	BIC	
1	-	-	-				
2	31.743	31.770	-10280.68				
3	31.497	31.525	-10201.14	24	29.898	29.939	-9680.954
4	30.864	30.899	-9995.023	25	28.993	29.041	-9386.607
5	30.844	30.886	-9987.523	26	28.973	29.028	-9379.204
6	30.868	30.909	-9995.227	27	28.996	29.051	-9386.563
7	30.847	30.896	-9987.523	28	28.969	29.032	-9377.081
8	-	-	-	29	-	-	-
9	30.788	30.823	-9970.44	30	30.378	30.420	-9836.597
10	31.120	31.154	-10077.85	31	31.013	31.054	-10042.18
11	30.210	30.251	-9782.026	32	29.968	30.016	-9702.563
12	30.214	30.262	-9782.191	33	29.971	30.026	-9702.564
13	30.214	30.262	-9782.192	34	29.971	30.026	-9702.563
14	30.217	30.272	-9782.191	35	29.974	30.036	-9702.563
15	-	-	-	36	-	-	-
16	29.339	29.374	-9500.963	37	29.346	29.388	-9502.265
17	29.973	30.008	-9706.365	38	29.943	29.984	-9695.53
18	28.995	29.036	-9388.288	39	29.001	29.049	-9389.265
19	28.973	29.021	-9380.255	40	28.980	29.035	-9381.561
20	28.998	29.046	-9388.262	41	29.004	29.059	-9389.255
21	28.971	29.027	-9378.711	42	28.978	29.040	-9379.716
22	-	-	-	43	⋮	⋮	⋮
23	29.335	29.376	-9498.525	⋮			
				64			

Table 2 shows the value of the Akaike information criterion (AIC) monotonically decreases when moving from the simpler model (standard GARCH) to the more complicated ones. We investigate three test codes whose results suggest the lowest AIC, BIC, and LogL in the Table 3.

Table 3: Rubric of the desirable AIC and BIC and LogL in red.

Test Code	AIC	BIC	LogL	Desirable
19 [(1,0)(1,2)]	28.973	29.021	-9380.255	No
21 [(1,0)(2,2)]	28.971	29.027	-9378.711	No
28 [(1,1)(2,2)]	28.969	29.032	-9377.081	Yes

We found that the ARMA(1,1)-GARCH(2,2) is the best model. Now we present the optimal parameters for the ARMA(1,1)-GARCH(2,2) model for the tea production and climatic variables as summarized in Fig. 1. We present these optimal parameters in Table 4.

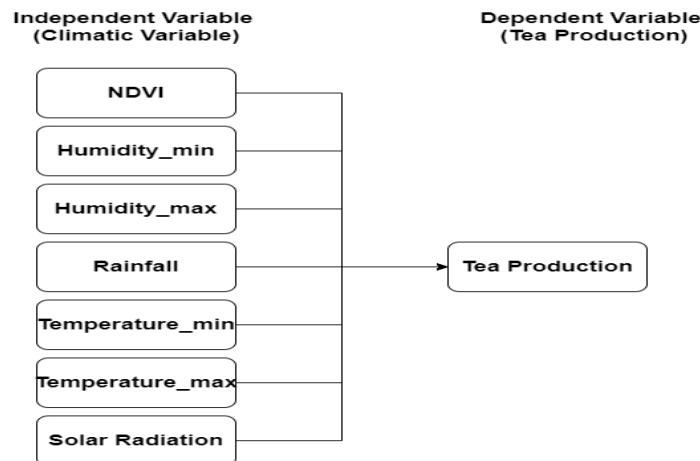


Figure 1: The conceptual framework showing independent variables (climatic variables) and dependent variable (Tea Production)

B. GARCH analysis for tea production and climatic variables

Table 4: Summary of GARCH analysis based on the ARMA(1,1)-GARCH(2,2) selected as the best model fit presented in Table 3 for climatic variables and tea production

Item	NDVI	H_{min}	H_{max}	Rainfall	T_{min}	T_{max}	Solar R
μ				2.129e+06			
ω				4.122e+09			
α_1				2.417e-01			
α_2				1.523e-01			
β_1				1.000e-06			
β_2				6.050e-01			
LogL				-9377.081			
AIC				28.969			
BIC				29.032			

Table 4 indicate that the optimal parameters, LogL and AIC and BIC based on model fit the ARMA(1,1)-GARCH(2,2) remain constant for all the GARCH analysis of Tea production and independent variables. We conducted the analysis independently, that is, Tea production & NDVI, Tea production & Humidity $_{min}$, Tea production & Humidity $_{max}$, Tea production & Rainfall, Tea production & Temperature $_{min}$, Tea production & Temperature $_{max}$, and Tea production & Solar radiation. The observation indicates a lack of change in the effect of climatic variables on the volatility of tea production. Therefore, Eq. (7) presents the volatility relationship between tea production and climatic variables.

$$\gamma_t = \mu + \omega_i + \sum_{i=1} \alpha_i \varepsilon_{t-1}^2 + \sum_{i=1} \beta_i \gamma_{t-1} \quad \forall \quad i = 1,2. \tag{7}$$

C. GARCH analysis based on counties

Table 5: Summary of GARCH analysis based on the ARMA(1,1)-GARCH(2,2) model presented in Table 3 for climatic variables and tea production for Counties.

Counties	LogL	IC	
		AIC	BIC
Embu	-1518.919	28.295	28.518
Kakamega	-1362.474	25.398	25.621
Kericho	-1673.635	31.160	31.383
Kisii	-1539.018	28.667	28.891
Meru	-1573.647	29.308	29.532
Nyeri	-1573.647	29.308	29.532

Table 5 indicate Kakamega county has the desirable change effect of climatic variables and tea production on the volatility. Kericho has the most undesirable change effect of climatic variables and tea production on the volatility.

Table 6: Summary of GARCH analysis based on the ARMA(1,1)-GARCH(2,2) model presented in Table 3 for climatic variables and tea production for Embu county

Item	NDVI	H_{min}	H_{max}	Rainfall	T_{min}	T_{max}	Solar R
μ				8.769e+05			
ω				1.278e+08			
α_1				0			
α_2				0			
β_1				9.989e-01			
β_2				6.700e-05			
LogL				-1518.919			
AIC				28.295			
BIC				28.518			

Table 6 suggest that we transform Eq. (7) to (8) to show the volatility relationship between tea production and climatic variables in Embu as

$$\gamma_t = \mu + \omega_i + \sum_{i=1} \beta_i \gamma_{t-1} \quad \forall \quad i = 1, 2, \dots, 7 \quad (8)$$

Table 7: Summary of GARCH analysis based on the ARMA(1,1)-GARCH(2,2) model presented in Table 3 for climatic variables and tea production for Kakamega county

Item	NDVI	H_{min}	H_{max}	Rainfall	T_{min}	T_{max}	Solar R
μ				1.073e+06			
ω				3.895e+08			
α_1				3.014e-01			
α_2				1.000e-06			
β_1				6.976e-01			
β_2				0			
LogL				-1362.474			
AIC				25.398			
BIC				25.398			

Table 7 suggest that we transform Eq. (7) to Eq. (9) to show the volatility relationship between tea production and climatic variables in Embu as

$$\gamma_t = \mu + \omega_i + \sum_{i=1} \alpha_i \varepsilon_{t-1}^2 + \beta_1 \gamma_{t-1} \quad \forall \quad i = 1, 2. \quad (9)$$

Table 8: Summary of GARCH analysis based on the ARMA(1,1)-GARCH(2,2) model presented in Table 3 for climatic variables and tea production for Kericho county

Item	NDVI	H_{min}	H_{max}	Rainfall	T_{min}	T_{max}	Solar R
μ				6.021e+06			
ω				2.622e+09			
α_1				0			
α_2				0			
β_1				9.99e-01			
β_2				5.200e-05			
LogL				-1673.635			
AIC				31.160			
BIC				31.383			

Table 8 suggest that we transform Eq. (7) to Eq. (10) to show the volatility relationship between tea production and climatic variables in Kericho as

$$\gamma_t = \mu + \omega_i + \sum_{i=1} \beta_i \gamma_{t-1} \quad \forall \quad i = 1, 2. \quad (10)$$

Table 9: Summary of GARCH analysis based on the ARMA(1,1)-GARCH(2,2) model presented in Table 3 for climatic variables and tea production for Kisii county

Item	NDVI	H_{min}	H_{max}	Rainfall	T_{min}	T_{max}	Solar R
μ				1.965e+06			
ω				1.909e+08			
α_1				1.953e-02			
α_2				1.076e-01			
β_1				0			
β_2				8.719e-01			
LogL				-1539.018			
AIC				28.667			
BIC				28.891			

Table 9 suggest that we transform Eq. (7) to Eq. (11) to show the volatility relationship between tea production and climatic variables in Kisii as

$$\gamma_t = \mu + \omega_i + \sum_{i=1} \alpha_i \varepsilon_{t-1}^2 + \beta_2 \gamma_{t-1} \quad \forall \quad i = 1,2. \quad (11)$$

Table 10: Summary of GARCH analysis based on the ARMA(1,1)-GARCH(2,2) model presented in Table 3 for climatic variables and tea production for Meru county

Item	NDVI	H_{min}	H_{max}	Rainfall	T_{min}	T_{max}	Solar R
μ				1.9582e+06			
ω				4.7431e+08			
α_1				6.5291e-0			
α_2				0			
β_1				0			
β_2				9.337e-01			
LogL				-1573.647			
AIC				29.308			
BIC				29.532			

Table 10 suggest that we transform Eq. (7) to Eq. (12) to show the volatility relationship between tea production and climatic variables in meru as

$$\gamma_t = \mu + \omega_i + \alpha_1 \varepsilon_{t-1}^2 + \beta_2 \gamma_{t-1}. \quad (12)$$

Table 11: Summary of GARCH analysis based on the ARMA(1,1)-GARCH(2,2) model presented in Table 3 for climatic variables and tea production for Nyeri county

Item	NDVI	H_{min}	H_{max}	Rainfall	T_{min}	T_{max}	Solar R
μ				1.9582e+06			
ω				4.743e+08			
α_1				6.5291e-0			
α_2				0			
β_1				0			
β_2				9.337e-01			
LogL				-1573.647			
AIC				29.308			
BIC				29.532			

Tables 10-11 post similar results, suggesting the GARCH volatility effect of data were unchanged between the two counties (Meru and Nyeri). Table 11 suggest that we transform Eq. (7) to Eq. (13) to show the volatility relationship between tea production and climatic variables in Nyeri as

$$\gamma_t = \mu + \omega_i + \alpha_1 \varepsilon_{t-1}^2 + \beta_2 \gamma_{t-1}. \quad (13)$$

IV. CONCLUSION

The influence of many factors such as climatic variables (NDVI, minimum and maximum humidity, rainfall, minimum and maximum temperature, and solar radiation) make tea production volatile. The clustering of low and high values of volatility periods makes a suitable GARCH and ARIMA model difficult. Predominantly, volatility is used to measure risk in financial time series data. However, the underlying factors of tea production, like climatic variables, are risk measures useful in understanding production data. A model of these variables requires a proper understanding of variables to help monitor and forecast the volatility in tea production output. The existence of many GARCH models with affirmative consent on the standard performance of GARCH(1,1) can be misleading due to variation in data volatility. The ARIMA model, which is limited to identifying linear relationships, has been proposed by many existing studies. The nonlinear nature of tea production and climatic variables creates everlasting interest to scholars to model a forecast of future tea production based on the volatile climatic conditions. We use Box and Jenkins model to outline 63 combinations of ARMA(m,n)-GARCH(p,q) models in tables with m and n are either 0, 1, or 2. AIC, BIC, and LogL criteria select the best model. The results based on the rubric indicated that the ARMA(1,1)-GARCH(2,2) is the suitable model since it has the lowest AIC and BIC with the highest LogL.

REFERENCES

- [1] Tim Bollerslev et al. Glossary to arch (garch). *CREATES Research paper*, 49:1–46, 2008.
- [2] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.
- [3] GEP Box, GM Jenkins, and GC Reinsel. Autocorrelation function and spectrum of stationary processes and analysis of seasonal time series. *Time Series Analysis: Forecasting and Control. 2nd ed. San Francisco: Holden-Day*, pages 21–43, 1976.
- [4] Liu-jie Chen and Ling Yu. Structural nonlinear damage identification algorithm based on time series arma/garch model. *Advances in Structural Engineering*, 16(9):1597–1609, 2013.
- [5] Huruta Andrian Dolfriandra, Andreas Hans Hananto, Roberto Louis Forestal, Anboli Elangovan, and John Francis Diaz. Revisiting spillover effect: An empirical evidence from garch-arma approach. *Industrija*, 49(1):67–80, 2021.
- [6] IM Md Ghani and HA Rahim. Modeling and forecasting of volatility using arma-garch: Case study on malaysia natural rubber prices. In *IOP Conference Series: Materials Science and Engineering*, volume 548, page 012023. IOP Publishing, 2019.
- [7] Jimmie Goode, Young Shin Kim, and Frank J Fabozzi. Full versus quasi mle for arma-garch models with infinitely divisible innovations. *Applied Economics*, 47(48):5147–5158, 2015.
- [8] Consolata A Muganda, Sewe Stanley, and Winnie Onsongo. Modeling effects of climatic variables on tea production in kenya using linear regression model with serially correlated errors. *Asian Journal of Probability and Statistics*, pages 56–75, 2021.
- [9] Nagaraj Naik, Biju R Mohan, and Rajat Aayush Jha. Garch-model identification based on performance of information criteria. *Procedia Computer Science*, 171:1935–1942, 2020.
- [10] Ser-huang Poon and Clive WJ Granger. Forecasting volatility in financial markets: a review. In *Journal of Economic Literature*. Citeseer, 2003.
- [11] Piotr Ptak, Matylda Jabłonska, Dominique Habimana, and Tuomo Kauranne. Reliability of arma and garch models of electricity spot market prices. In *European Symposium on Time Series Prediction, Porvoo, Finland, September*, volume 17, pages 21–35, 2008.
- [12] GC Reinsel et al. Time series analysis: forecasting and control. *Journal of Marketing Research*, 14(2):5561569, 1994.
- [13] Gabriella Sawma. *Performance of ARMA-GARCH models in Value at Risk estimation*. PhD thesis, Stockholm University, 2019.
- [14] Junmo Song and Jiwon Kang. Sequential change point detection in arma-garch models. *Journal of Statistical Computation and Simulation*, 90(8):1520–1538, 2020.
- [15] Shuairu Tian and Shigeyuki Hamori. Modeling interest rate volatility: a realized garch approach. *Journal of Banking & Finance*, 61:158–171, 2015.