Contents

1. Introduction

This paper investigates the SSE Composite Index from the time range 04/Jan/2010 to 28/Oct/2022. The behavior of the SSE Index return series is analyzed with time series econometrics, with the aim of building an adequate model for the return series and its volatility.

2. Data Characteristics

The nature of the data is a longitudinal univariate pure time series for the return of the SSE Index as a function of the time series. From the graph plotted below, there appears to be a long-term consistent trend and an outlier – a spike in 04-06-2015.

Figure 1: Adjusted Closing price of Shanghai Stock Exchange (SSE) Composite Index

In the dataset, the x variable of interest is adj_close (price) against the date. We transform the series into percentage returns. We use the function ***dlog*** here as it creates the first difference of the logarithm. This is equivalent to:

rt = 100*(log(adj_price)-log(adj_price(-1)))

To test the model viability in terms of prediction results, we separate the data into training and test sets:

Training = 1/04/2010 to 6/1/2020

Testing = 6/2/2020 to 10/28/2022

3. Model Selection and Interpretation

This section identifies how a model selected with Box-Jenkins analysis. Autoregressive and moving average can be combined to form an ARMA process. The recommended model will be **ARMA (6, 6)-EGARCH(1, 1).** The return is stationary and thus ARIMA need not be used, which is the next-step model to ARMA with integration that differences the data to achieve stationarity. The white noise, however, has an ARCH effect. The next sections identify the thought process that would lead up to the model selection – including identification with information criterions, estimation, and diagnostic checking.

4. Justifications of the Selected Model

Wold's decomposition theorem indicates the necessity of stationary time series in constructing an ARMA model. To identify stationarity, we investigate with a correlogram. With the correlogram diagram, if there is no serial correlation, the autocorrelation (AC) and partial autocorrelation (PAC) at all lags should be near zero and all Q-statistic should be insignificant. In this case, up until the fifth autocorrelation and partial autocorrelation appear strongly significant. Thus, the test statistic rejects the null hypothesis of autocorrelation for all number of lags up ARMA (6, 6). A variable that is not serially correlated indicates that it is random (i.e., a white noise process). This concludes that a mixed ARMA process could be appropriate up to ARMA (6, 6).

Furthermore, to be stationary and invertible, the inverted AR roots and inverted MA roots must be less than 1 in absolute value. In this case, the MA and AR roots are clearly stationary and invertible.

Further analysis is needed, with the Unit Root test conducted in the next section.

4.1 Correlogram of SSE Index Returns

Correlogram of RT

Date: 12/11/22 Time: 12:53
Sample (adjusted): 1/05/2010 10/28/2022
Included observations: 3114 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.031	0.031	3.0117	0.083
		2	-0.024	-0.025	4.7358	0.094
		3	0.018	0.019	5.7079	0.127
		4	0.033	0.031	9.0071	0.061
		5	-0.001	-0.002	9.0102	0.109
		6	-0.055	-0.054	18.541	0.005
		7	0.038	0.040	23.027	0.002
		8	0.029	0.023	25.599	0.001
		9	0.019	0.022	26.775	0.002
		10	-0.021	-0.019	28.132	0.002
		11	-0.016	-0.017	28.930	0.002
		12	0.004	-0.001	28.993	0.004
		13	0.045	0.049	35.397	0.001
		14	-0.062	-0.063	47.553	0.000
		15	0.005	0.012	47.619	0.000
		16	0.028	0.019	50.140	0.000
		17	0.016	0.013	50.901	0.000
		18	0.005	0.011	50.990	0.000
		19	0.002	0.007	51.002	0.000
		20	0.053	0.041	59.663	0.000
		21	0.024	0.024	61.501	0.000
		22	-0.016	-0.014	62.346	0.000

Figure 2: Correlogram reflecting autocorrelation (AC) and partial autocorrelation (PAC)

4.2 Unit Root Test for Stationarity

By conducting a unit root test, we can determine if the data is stationary, a fundamental requirement for time series modelling. The Augmented Dickey-Fuller (ADF) test is widely used to test if a series contains a unit root, based on a regression of the change in that variable on the lag of the level of that variable. From the ADF test, the null hypothesis of a unit root (i.e., H0 = unit root) is rejected as the test statistic = 0.001 < 0.05. Meanwhile, from the KPSS test, the null hypothesis of stationarity (i.e., H0 = stationary) is not rejected. Thus, the series is stationary.

Null Hypothesis: RT has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic - based on SIC, maxlag=28)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-54.04825	0.0001
Test critical values:	1% level	-3.432261	
	5% level	-2.862270	
	10% level	-2.567203	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RT)
Method: Least Squares
Date: 12/13/22 Time: 15:40
Sample (adjusted): 1/06/2010 10/28/2022
Included observations: 3113 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RT(-1)	-0.968886	0.017926	-54.04825	0.0000
C	-0.003717	0.023565	-0.157735	0.8747

R-squared	0.484269	Mean dependent var	-0.001108
Adjusted R-squared	0.484103	S.D. dependent var	1.830510
S.E. of regression	1.314782	Akaike info criterion	3.385860
Sum squared resid	5377.832	Schwarz criterion	3.389743
Log likelihood	-5268.092	Hannan-Quinn criter.	3.387254
F-statistic	2921.214	Durbin-Watson stat	1.996997
Prob(F-statistic)	0.000000		

Null Hypothesis: RT is stationary
Exogenous: Constant
Bandwidth: 11 (Newey-West automatic) using Bartlett kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.078977
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	1.729105
HAC corrected variance (Bartlett kernel)	1.870942

KPSS Test Equation
Dependent Variable: RT
Method: Least Squares
Date: 12/13/22 Time: 15:59
Sample (adjusted): 1/05/2010 10/28/2022
Included observations: 3114 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.003421	0.023568	-0.145175	0.8846

R-squared	0.000000	Mean dependent var	-0.003421
Adjusted R-squared	0.000000	S.D. dependent var	1.315166
S.E. of regression	1.315166	Akaike info criterion	3.386124
Sum squared resid	5384.435	Schwarz criterion	3.388064
Log likelihood	-5271.194	Hannan-Quinn criter.	3.386820
Durbin-Watson stat	1.936618		

Figures 3 and 4 (Left and Right): ADF and KPSS Unit Root Test

4.3 Information Criteria for ARMA model selection

The model with the lowest AIC and SBIC will be chosen. It is worth noting that AIC has a less strict penalty while SBIC has a strict penalty. As AIC is more appropriate to predict future observations, while BIC is more useful in selecting a correct model, the AIC method will be used.

ARMA (6,6) is the best in terms of AIC.

Click Quick (Top row) > Estimate Equation > rt c ar(5) ma(5) / rt c ar(6) ma(6)

Dependent Variable: RT
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/13/22 Time: 16:45
Sample: 1/05/2010 6/01/2020
Included observations: 2529
Convergence achieved after 64 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004322	0.033403	-0.129382	0.8971
AR(1)	0.476500	0.760305	0.626722	0.5309
AR(2)	-0.177338	0.131620	-1.347343	0.1780
AR(3)	-0.082036	0.187790	-0.436853	0.6623
AR(4)	0.813139	0.165019	4.927535	0.0000
AR(5)	-0.258137	0.522666	-0.493885	0.6214
MA(1)	-0.441074	0.763343	-0.577819	0.5634
MA(2)	0.127778	0.124062	1.029958	0.3031
MA(3)	0.155829	0.162096	0.961337	0.3365
MA(4)	-0.815560	0.203980	-3.998240	0.0001
MA(5)	0.229109	0.513478	0.446191	0.6555
SIGMASQ	1.830654	0.030498	60.02460	0.0000

R-squared	0.019479	Mean dependent var	-0.004220
Adjusted R-squared	0.015194	S.D. dependent var	1.366660
S.E. of regression	1.356238	Akaike info criterion	3.452128
Sum squared resid	4629.724	Schwarz criterion	3.479818
Log likelihood	-4353.216	Hannan-Quinn criter.	3.462175
F-statistic	4.545762	Durbin-Watson stat	2.002499
Prob(F-statistic)	0.000001		

Inverted AR Roots	.90	.33	.07-.97i	.07+.97i
	-.91			
Inverted MA Roots	.89	.30	.09-.95i	.09+.95i
	-.93			

Dependent Variable: RT
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 12/13/22 Time: 16:48
Sample: 1/05/2010 6/01/2020
Included observations: 2529
Convergence achieved after 110 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004342	0.032769	-0.132494	0.8946
AR(1)	-1.125350	0.176428	-6.378540	0.0000
AR(2)	-1.140838	0.301753	-3.780705	0.0002
AR(3)	-0.434158	0.421982	-1.028856	0.3036
AR(4)	0.189399	0.390417	0.485120	0.6276
AR(5)	0.579084	0.282436	2.050320	0.0404
AR(6)	0.627133	0.127804	4.906987	0.0000
MA(1)	1.158139	0.180055	6.432149	0.0000
MA(2)	1.151568	0.316499	3.638459	0.0003
MA(3)	0.482929	0.437616	1.103546	0.2699
MA(4)	-0.124607	0.410767	-0.303351	0.7616
MA(5)	-0.494487	0.300052	-1.648007	0.0995
MA(6)	-0.626068	0.142149	-4.404316	0.0000
SIGMASQ	1.821356	0.030611	59.50031	0.0000

R-squared	0.024460	Mean dependent var	-0.004220
Adjusted R-squared	0.019417	S.D. dependent var	1.366660
S.E. of regression	1.353327	Akaike info criterion	3.448715
Sum squared resid	4606.209	Schwarz criterion	3.481020
Log likelihood	-4346.900	Hannan-Quinn criter.	3.460436
F-statistic	4.850639	Durbin-Watson stat	1.990660
Prob(F-statistic)	0.000000		

Inverted AR Roots	.78	.08-.97i	.08+.97i	-.59-.78i
	-.59+.78i	-.88		
Inverted MA Roots	.76	.09-.96i	.09+.96i	-.59+.79i
	-.59-.79i	-.91		

Figures 5 and 6 (Left and Right): ARMA (5, 5) and ARMA (6, 6)

==
Model Selection Criteria Table
Dependent Variable: RT
Date: 12/14/22 Time: 08:02
Sample: 1/04/2010 6/01/2020
Included observations: 2529
==

Model		LogL	AIC*	BIC	HQ
(6	6)(0	-4346.900183	3.448715	3.481020	3.460436
(4	6)(0	-4350.648133	3.450097	3.477787	3.460144
(3	5)(0	-4352.736939	3.450168	3.473242	3.458540
(5	3)(0	-4352.785133	3.450206	3.473280	3.458578
(6	4)(0	-4351.005135	3.450380	3.478069	3.460426
(4	3)(0	-4354.152625	3.450496	3.471264	3.458031
(3	4)(0	-4354.287464	3.450603	3.471370	3.458138
(6	3)(0	-4352.298991	3.450612	3.475994	3.459822
(3	6)(0	-4352.373880	3.450671	3.476053	3.459881
(4	4)(0	-4353.453642	3.450734	3.473809	3.459107
(5	5)(0	-4351.475329	3.450752	3.478441	3.460798

Figure 7: Model Selection Criteria Table indicates ARMA(6,6) as ideal

4.4 Testing for ARCH Effects

Testing for ARCH effects, also known as a test for autocorrelation in the squared residuals, allows us to determine of the assumption of constant variance is valid. We identify ARCH effects by conducting residual diagnostics of the estimated model. According to the test results, the top panel shows the results of the ARCH Lagrange Multiplier (LM) test. Up until RESID^2(-5), the statistics are significant indicating the presence of ARCH effects. We reject the null hypothesis of no ARCH effects, which indicates that the residuals suffer from heteroskedasticity. Thereby, an ARCH model will yield better results as it accounts for the effect of volatility in the time series.

View > Residual Diagnostics > Heteroskedasticity Tests > ARCH = 6

Heteroskedasticity Test: ARCH

F-statistic	54.33849	Prob. F(6,2516)	0.0000
Obs*R-squared	289.4325	Prob. Chi-Square(6)	0.0000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 12/14/22 Time: 08:08
Sample (adjusted): 1/13/2010 6/01/2020
Included observations: 2523 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.812266	0.108860	7.461602	0.0000
RESID^2(-1)	0.093402	0.019937	4.684826	0.0000
RESID^2(-2)	0.146678	0.019935	7.357929	0.0000
RESID^2(-3)	0.136810	0.020082	6.812700	0.0000
RESID^2(-4)	0.081859	0.020081	4.076414	0.0000
RESID^2(-5)	0.094587	0.019934	4.744998	0.0000
RESID^2(-6)	0.000768	0.019937	0.038532	0.9693

R-squared	0.114718	Mean dependent var	1.821950
Adjusted R-squared	0.112606	S.D. dependent var	4.883833
S.E. of regression	4.600647	Akaike info criterion	5.893042
Sum squared resid	53253.54	Schwarz criterion	5.909226
Log likelihood	-7427.072	Hannan-Quinn criter.	5.898914
F-statistic	54.33849	Durbin-Watson stat	1.999487
Prob(F-statistic)	0.000000		

Date: 12/14/22 Time: 08:10
Sample (adjusted): 1/05/2010 6/01/2020
Included observations: 2529 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.188	0.188	89.274	0.000
		2	0.230	0.202	223.13	0.000
		3	0.223	0.163	349.35	0.000
		4	0.180	0.092	431.48	0.000
		5	0.191	0.095	524.06	0.000
		6	0.112	0.001	555.96	0.000
		7	0.134	0.035	601.83	0.000
		8	0.099	0.006	626.81	0.000
		9	0.117	0.040	661.77	0.000
		10	0.135	0.062	707.89	0.000
		11	0.102	0.025	734.14	0.000
		12	0.096	0.011	757.70	0.000
		13	0.151	0.081	816.04	0.000
		14	0.102	0.015	842.68	0.000
		15	0.095	0.003	865.52	0.000
		16	0.168	0.094	937.78	0.000
		17	0.117	0.028	972.52	0.000
		18	0.104	0.001	1000.2	0.000
		19	0.124	0.033	1039.7	0.000
		20	0.172	0.087	1115.1	0.000
		21	0.161	0.063	1180.9	0.000
		22	0.086	-0.031	1199.6	0.000
		23	0.095	-0.025	1222.6	0.000
		24	0.082	-0.016	1239.7	0.000
		25	0.149	0.074	1296.3	0.000
		26	0.088	-0.006	1315.9	0.000
		27	0.094	0.013	1338.5	0.000
		28	0.150	0.080	1395.8	0.000
		29	0.098	-0.000	1420.6	0.000
		30	0.140	0.034	1471.0	0.000
		31	0.079	-0.020	1486.9	0.000
		32	0.121	0.025	1524.5	0.000
		33	0.114	0.015	1557.9	0.000
		34	0.112	0.023	1590.0	0.000
		35	0.123	0.029	1628.6	0.000
		36	0.122	0.032	1666.7	0.000

Figures 8 and 9: Heteroskedasticity Test and Correlogram of Squared Residuals

4.5 GARCH Model

To check if the goodness-of-fit for the model, we scrutinize the residuals once again. With the histogram, as the standardized residuals are not normally distributed, the quasi-maximum likelihood (QML) model with Bollerslev-Wooldridge coefficient covariance will be used in specifying the GARCH(1,1) model in EViews.

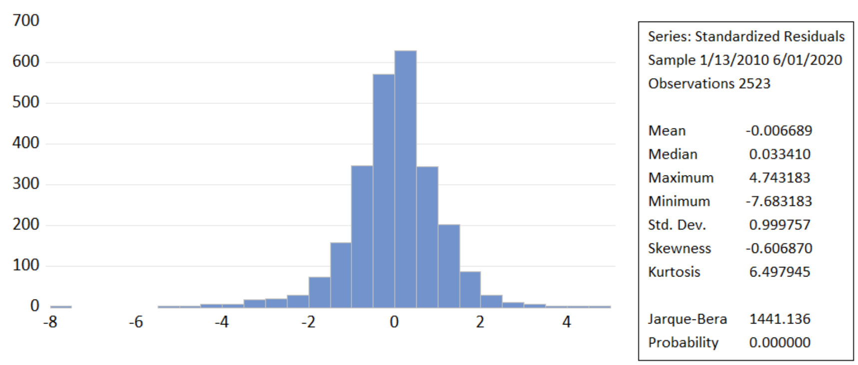


Figure 10: GARCH (1, 1) Standardized Residuals indicate non-normality

Dependent Variable: RT
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 12/14/22 Time: 08:14
Sample (adjusted): 1/13/2010 6/01/2020
Included observations: 2523 after adjustments
Convergence achieved after 48 iterations
Coefficient covariance computed using Bollerslev-Wooldridge QML
sandwich with expected Hessian
MA Backcast: 1/05/2010 1/12/2010
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.005987	0.020793	0.287930	0.7734
AR(6)	-0.848717	0.051215	-16.57174	0.0000
MA(6)	0.871582	0.046666	18.67687	0.0000
		Variance Equation		
C	0.010454	0.005230	1.998725	0.0456
RESID(-1)^2	0.063094	0.018016	3.502167	0.0005
GARCH(-1)	0.934053	0.017187	54.34682	0.0000

R-squared	0.000546	Mean dependent var	-0.004597
Adjusted R-squared	-0.000247	S.D. dependent var	1.366893
S.E. of regression	1.367062	Akaike info criterion	3.175036
Sum squared resid	4709.520	Schwarz criterion	3.188908
Log likelihood	-3999.308	Hannan-Quinn criter.	3.180070
Durbin-Watson stat	1.931609		

Inverted AR Roots	.84-.49i	.84+.49i	.00+.97i	-.00-.97i
	-.84-.49i	-.84+.49i		
Inverted MA Roots	.85+.49i	.85-.49i	.00-.98i	-.00+.98i
	-.85+.49i	-.85-.49i		

Date: 12/14/22 Time: 08:14
Sample (adjusted): 1/13/2010 6/01/2020
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.035	0.035	3.1710	
		2	0.009	0.008	3.3724	
		3	0.040	0.039	7.3638	0.007
		4	0.002	-0.000	7.3793	0.025
		5	0.002	0.001	7.3886	0.060
		6	-0.041	-0.043	11.711	0.020
		7	0.050	0.053	17.990	0.003
		8	0.010	0.007	18.249	0.006
		9	0.029	0.031	20.336	0.005
		10	0.018	0.011	21.124	0.007
		11	0.003	0.001	21.144	0.012
		12	0.002	-0.003	21.154	0.020
		13	-0.000	0.003	21.154	0.032
		14	-0.017	-0.020	21.904	0.039
		15	0.002	0.006	21.919	0.057
		16	0.020	0.018	22.911	0.062
		17	-0.014	-0.016	23.404	0.076
		18	0.006	0.005	23.481	0.101
		19	0.014	0.012	24.000	0.119
		20	0.018	0.016	24.841	0.129
		21	-0.023	-0.023	26.203	0.125
		22	-0.016	-0.014	26.858	0.139
		23	-0.024	-0.027	28.312	0.132
		24	-0.012	-0.006	28.652	0.155
		25	0.031	0.033	31.043	0.122
		26	-0.017	-0.017	31.778	0.133
		27	-0.009	-0.011	31.977	0.159
		28	0.031	0.030	34.458	0.124
		29	-0.030	-0.033	36.716	0.100
		30	-0.008	-0.002	36.862	0.122
		31	-0.004	-0.000	36.894	0.149
		32	-0.015	-0.015	37.441	0.165
		33	0.008	0.012	37.603	0.192
		34	-0.013	-0.011	38.061	0.213
		35	0.016	0.012	38.753	0.226
		36	-0.011	-0.010	39.062	0.253

*Probabilities may not be valid for this equation specification.

Figures 11 and 12: Output of ARMA (6, 6)-GARCH (1,1) and Correlogram of Standardized Residuals

The coefficients of both RESID(-1)^2 and GARCH(-1) adjusted for coefficient variance using the Bollerslev-Wooldridge QML is statistically significant. Thus, shocks impact conditional volatility and there is a presence of persistent volatility clustering. As a conventional GARCH model is unable to enforce an asymmetric formulation to positive/negative shocks, alternative formulations will be explored, namely TARCH, EGARCH, and GARCH-M. Meanwhile, the correlogram reflect white noise.

4.6 Information Criteria for ARMA and TARCH, EGARCH, GARCH-M model selection

With ARMA(6,6) as the baseline, according to the Schwarz's Bayesian Information Criterion (SBIC) and the Akaike Information Criterion (AIC), the ARMA(6,6)-EGARCH(1,1) model works most effectively. This paper proposes ARMA(6,6)-EGARCH(1,1) as the optimal model in predicting SSE Composite Index returns.

	SBIC	**AIC**
TGARCH	3.191995	3.175811
EGARCH	**3.190015**	**3.173831**
GARCH-M	3.191912	3.175728

Table 1: SBIC values for ARMA(6,6)-ARCH models

Dependent Variable: RT
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 12/14/22 Time: 08:19
Sample (adjusted): 1/13/2010 6/01/2020
Included observations: 2523 after adjustments
Convergence achieved after 75 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1/05/2010 1/12/2010
Presample variance: backcast (parameter = 0.7)
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*RESID(-1)^2*(RESID(-1)<0) +
C(7)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.005386	0.021342	0.252353	0.8008
AR(6)	-0.848341	0.056581	-14.99351	0.0000
MA(6)	0.871323	0.052949	16.45604	0.0000
Variance Equation				
C	0.010563	0.002074	5.093705	0.0000
RESID(-1)^2	0.062138	0.006008	10.34235	0.0000
RESID(-1)^2*(RESID(-1)<0)	0.002227	0.006476	0.343878	0.7309
GARCH(-1)	0.933775	0.003665	254.7558	0.0000

R-squared	0.000528	Mean dependent var	-0.004597
Adjusted R-squared	-0.000266	S.D. dependent var	1.366893
S.E. of regression	1.367074	Akaike info criterion	3.175811
Sum squared resid	4709.608	Schwarz criterion	3.191995
Log likelihood	-3999.286	Hannan-Quinn criter.	3.181684
Durbin-Watson stat	1.931626		

Inverted AR Roots	.84-.49i	.84+.49i	.00+.97i	-.00-.97i
	-.84-.49i	-.84+.49i		
Inverted MA Roots	.85+.49i	.85-.49i	.00-.98i	-.00+.98i
	-.85+.49i	-.85-.49i		

Date: 12/14/22 Time: 08:20
Sample (adjusted): 1/13/2010 6/01/2020
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.036	0.036	3.1853	
		2	0.009	0.008	3.3969	
		3	0.040	0.039	7.3923	0.007
		4	0.003	-0.000	7.4120	0.025
		5	0.002	0.001	7.4223	0.060
		6	-0.041	-0.043	11.751	0.019
		7	0.050	0.053	18.047	0.003
		8	0.010	0.007	18.303	0.006
		9	0.029	0.031	20.375	0.005
		10	0.018	0.012	21.179	0.007
		11	0.003	0.001	21.201	0.012
		12	0.002	-0.003	21.213	0.020
		13	-0.000	0.003	21.213	0.031
		14	-0.017	-0.020	21.957	0.038
		15	0.002	0.006	21.972	0.056
		16	0.020	0.018	22.974	0.061
		17	-0.014	-0.016	23.466	0.075
		18	0.006	0.005	23.546	0.100
		19	0.014	0.012	24.062	0.118
		20	0.018	0.016	24.910	0.127
		21	-0.023	-0.023	26.266	0.123
		22	-0.016	-0.014	26.911	0.138
		23	-0.024	-0.026	28.338	0.131
		24	-0.011	-0.006	28.670	0.155
		25	0.031	0.033	31.055	0.121
		26	-0.017	-0.017	31.784	0.132
		27	-0.009	-0.011	31.973	0.159
		28	0.031	0.030	34.456	0.124
		29	-0.030	-0.033	36.720	0.100
		30	-0.008	-0.002	36.873	0.122
		31	-0.003	-0.000	36.902	0.149
		32	-0.015	-0.015	37.451	0.164
		33	0.008	0.012	37.617	0.192
		34	-0.013	-0.011	38.072	0.212
		35	0.016	0.012	38.768	0.226
		36	-0.011	-0.010	39.085	0.252

*Probabilities may not be valid for this equation specification.

Figures 13 and 14: ARMA(6,6)-TARCH(1,1) Estimation Output and Correlogram of Residuals

Dependent Variable: RT
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 12/14/22 Time: 08:21
Sample (adjusted): 1/13/2010 6/01/2020
Included observations: 2523 after adjustments
Failure to improve likelihood (singular hessian) after 48 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1/05/2010 1/12/2010
Presample variance: backcast (parameter = 0.7)
LOG(GARCH) = C(4) + C(5)*ABS(RESID(-1)/@SQRT(GARCH(-1))) +
C(6)*RESID(-1)/@SQRT(GARCH(-1)) + C(7)*LOG(GARCH(-1))

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.007294	0.020339	0.358621	0.7199
AR(6)	-0.846198	0.056621	-14.94502	0.0000
MA(6)	0.870523	0.052734	16.50796	0.0000
Variance Equation				
C(4)	-0.104437	0.006488	-16.09634	0.0000
C(5)	0.148225	0.008419	17.60547	0.0000
C(6)	-0.008432	0.005335	-1.580416	0.1140
C(7)	0.990387	0.001759	562.9358	0.0000

R-squared	0.000195	Mean dependent var	-0.004597
Adjusted R-squared	-0.000598	S.D. dependent var	1.366893
S.E. of regression	1.367301	Akaike info criterion	3.173831
Sum squared resid	4711.174	Schwarz criterion	3.190015
Log likelihood	-3996.788	Hannan-Quinn criter.	3.179704
Durbin-Watson stat	1.931620		

Inverted AR Roots	.84-.49i	.84+.49i	.00+.97i	-.00-.97i
	-.84-.49i	-.84+.49i		
Inverted MA Roots	.85+.49i	.85-.49i	.00-.98i	-.00+.98i
	-.85+.49i	-.85-.49i		

Date: 12/14/22 Time: 08:21
Sample (adjusted): 1/13/2010 6/01/2020
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.032	0.032	2.5925	
		2	0.006	0.005	2.6804	
		3	0.040	0.039	6.6633	0.010
		4	0.003	0.000	6.6844	0.035
		5	-0.001	-0.001	6.6860	0.083
		6	-0.044	-0.045	11.531	0.021
		7	0.050	0.053	17.753	0.003
		8	0.010	0.007	18.023	0.006
		9	0.032	0.034	20.558	0.004
		10	0.018	0.012	21.410	0.006
		11	0.003	0.001	21.431	0.011
		12	0.004	-0.001	21.479	0.018
		13	-0.001	0.002	21.482	0.029
		14	-0.020	-0.022	22.448	0.033
		15	0.002	0.005	22.455	0.049
		16	0.018	0.016	23.320	0.055
		17	-0.014	-0.015	23.805	0.068
		18	0.004	0.003	23.844	0.093
		19	0.019	0.016	24.734	0.101
		20	0.021	0.019	25.857	0.103
		21	-0.023	-0.023	27.224	0.100
		22	-0.016	-0.015	27.901	0.112
		23	-0.024	-0.026	29.338	0.106
		24	-0.009	-0.004	29.552	0.130
		25	0.031	0.033	31.954	0.101
		26	-0.017	-0.017	32.716	0.110
		27	-0.011	-0.014	33.018	0.131
		28	0.035	0.033	36.162	0.089
		29	-0.028	-0.031	38.141	0.076
		30	-0.008	-0.001	38.288	0.093
		31	-0.005	-0.002	38.340	0.115
		32	-0.014	-0.015	38.878	0.128
		33	0.008	0.012	39.027	0.152
		34	-0.012	-0.009	39.406	0.172
		35	0.018	0.012	40.212	0.181
		36	-0.012	-0.011	40.561	0.203

*Probabilities may not be valid for this equation specification.

Figures 15 and 16: ARMA(6,6)-EGARCH(1,1) Estimation Output and Correlogram of Residuals

Dependent Variable: RT
Method: ML ARCH - Normal distribution (BFGS / Marquardt steps)
Date: 12/14/22 Time: 08:24
Sample (adjusted): 1/13/2010 6/01/2020
Included observations: 2523 after adjustments
Convergence achieved after 53 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1/05/2010 1/12/2010
Presample variance: backcast (parameter = 0.7)
GARCH = C(5) + C(6)*RESID(-1)^2 + C(7)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
@SQRT(GARCH)	-0.028624	0.063129	-0.453416	0.6502
C	0.033980	0.065761	0.516725	0.6053
AR(6)	-0.849876	0.056062	-15.15963	0.0000
MA(6)	0.872564	0.052457	16.63378	0.0000
Variance Equation				
C	0.010443	0.002174	4.804262	0.0000
RESID(-1)^2	0.063003	0.003946	15.96622	0.0000
GARCH(-1)	0.934116	0.003471	269.1256	0.0000

R-squared	0.000817	Mean dependent var	-0.004597
Adjusted R-squared	-0.000373	S.D. dependent var	1.366893
S.E. of regression	1.367148	Akaike info criterion	3.175728
Sum squared resid	4708.245	Schwarz criterion	3.191912
Log likelihood	-3999.181	Hannan-Quinn criter.	3.181601
Durbin-Watson stat	1.933139		

Inverted AR Roots	.84-.49i	.84+.49i	.00+.97i	-.00-.97i
	-.84-.49i	-.84+.49i		
Inverted MA Roots	.85-.49i	.85+.49i	-.00+.98i	-.00-.98i
	-.85-.49i	-.85+.49i		

Date: 12/14/22 Time: 08:24
Sample (adjusted): 1/13/2010 6/01/2020
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob*
		1	0.035	0.035	3.1153	
		2	0.009	0.007	3.3013	
		3	0.039	0.039	7.2131	0.007
		4	0.002	-0.001	7.2224	0.027
		5	0.001	0.001	7.2280	0.065
		6	-0.042	-0.043	11.617	0.020
		7	0.049	0.052	17.776	0.003
		8	0.010	0.007	18.011	0.006
		9	0.028	0.031	20.048	0.005
		10	0.017	0.011	20.791	0.008
		11	0.002	0.001	20.806	0.014
		12	0.001	-0.003	20.812	0.022
		13	-0.000	0.003	20.812	0.035
		14	-0.018	-0.020	21.596	0.042
		15	0.002	0.005	21.606	0.062
		16	0.019	0.018	22.559	0.068
		17	-0.014	-0.016	23.080	0.082
		18	0.005	0.005	23.150	0.110
		19	0.014	0.012	23.644	0.129
		20	0.018	0.016	24.457	0.141
		21	-0.023	-0.023	25.854	0.134
		22	-0.016	-0.014	26.518	0.149
		23	-0.024	-0.027	28.006	0.140
		24	-0.012	-0.006	28.365	0.164
		25	0.030	0.033	30.735	0.129
		26	-0.017	-0.017	31.487	0.140
		27	-0.009	-0.011	31.698	0.167
		28	0.031	0.030	34.165	0.131
		29	-0.030	-0.033	36.433	0.106
		30	-0.008	-0.002	36.579	0.128
		31	-0.004	-0.000	36.612	0.156
		32	-0.015	-0.015	37.164	0.172
		33	0.008	0.012	37.326	0.201
		34	-0.013	-0.011	37.783	0.222
		35	0.016	0.012	38.477	0.235
		36	-0.011	-0.010	38.784	0.263

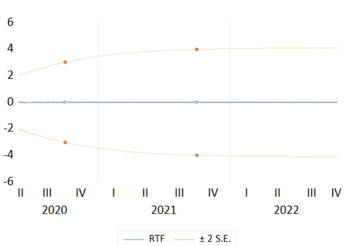
*Probabilities may not be valid for this equation specification.

Figures 17 and 18: ARMA(6,6)-GARCH-M(1,1) Estimation Output and Correlogram of Residuals

4.7 ARMA(6,6)-EGARCH(1,1) Dynamic and Static Forecast

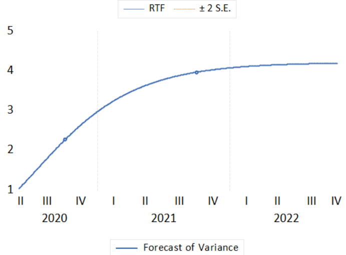
Based on the selected optimal model of ARMA(6,6)-EGARCH(1,1), the RMSE, MAE and Theil Inequality coefficient presents a sufficiently optimal outcome.

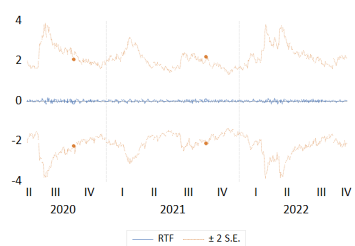
The forecast for the test period 6/2/2020 to 10/28/2022 are as follows:



Forecast: RTF
Actual: RT
Forecast sample: 6/02/2020 10/28/2022
Included observations: 585

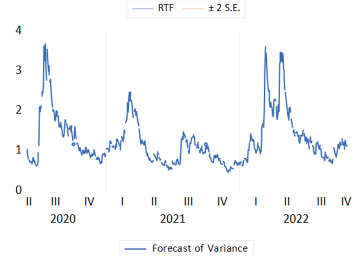
Root Mean Squared Error	1.064285
Mean Absolute Error	0.774931
Mean Abs. Percent Error	104.9320
Theil Inequality Coef.	0.989684
Bias Proportion	0.000047
Variance Proportion	0.984761
Covariance Proportion	0.015191
Theil U2 Coefficient	1.025919
Symmetric MAPE	193.5183





Forecast: RTF
Actual: RT
Forecast sample: 6/02/2020 10/28/2022
Included observations: 585

Root Mean Squared Error	1.064603
Mean Absolute Error	0.775768
Mean Abs. Percent Error	136.0079
Theil Inequality Coef.	0.954328
Bias Proportion	0.000048
Variance Proportion	0.906683
Covariance Proportion	0.093268
Theil U2 Coefficient	1.088081
Symmetric MAPE	178.4342



Figures 19 and 20: Dynamic and Static Forecast

5. Limitations

The sample may include structural breaks. Structural breaks may best be modelled with other time-varying parameter models and Markov-switching models.

Using the GARCH model, we may not be accounting for asymmetricity. Thus, the EGARCH model was used to account for negativity constraints. However, it may not be the best model if one is looking at overcoming leverage effects or incorporating the “higher risk, higher returns” concept.

6. Conclusion

This paper proposes the use of ARMA(6,6)-EGARCH(1,1) model in measuring and predicting returns of the SSE Composite Index.