An empirical analysis of the Chinese stock market indexes by ARMA-GARCH models

Qiuyi HUANG

Adviser: Xuejun JIANG

Department of Statistics and Data Science
The Southern University of Science and Technology

May 12, 2022

Content

GARCH Model

GARCH (p, q) [1]

Series $\{a_t\}$ follows a GARCH (p, q) model if

$$a_t = \sigma_{t|t-1}\varepsilon_t,$$
$$\sigma^2_{t|t-1} = \omega + \sum_{i=1}^{q} \alpha_i a^2_{t-i} + \sum_{j=1}^{p} \beta_j \sigma^2_{t-j|t-j-1},$$

where $\{a_t\}$ is a zero mean time series, $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ for $i > 0$, $j > 0$. $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean zero and variance one.

- Note that p is the order of GARCH part, q is the order of ARCH part.
- GARCH(p, q) for a_t implies ARMA$(\max(p, q),\ p)$ for a_t^2.
- $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$, under the assumption of finite variance.

GARCH Model Cont' d

IGARCH$(1,1)$ [2]

Series $\{a_t\}$ follows an integrated GARCH$(1,1)$ model if

$$\begin{gathered} a_t = \sigma_{t|t-1}\varepsilon_t, \\ \sigma^2_{t|t-1} = \alpha_0 + \beta_1\sigma^2_{t-1|t-2} + (1-\beta_1)a^2_{t-1} \end{gathered}$$

The unconditional variance of a_t is not defined.

TGARCH(p,q) [3]

Series $\{a_t\}$ follows an threshold GARCH (p,q) model if

$$\begin{gathered} \sigma^2_{t|t-1} = \alpha_0 + \sum_{i=1}^{q}(\alpha_i + \gamma_i N_{t-i})\, a^2_{t-i} + \sum_{j=1}^{p}\beta_j\sigma^2_{t-j|t-j-1}, \\ N_{t-i} = \begin{cases} 1 & \text{if } a_{t-i} < 0 \\ 0 & \text{if } a_{t-i} \geq 0 \end{cases} \end{gathered}$$

From the model, a positive a_{t-i} contributes $\alpha_i a^2_{t-i}$ to $\sigma^2_{t|t-1}$, whereas a negative a_{t-i} has a larger impact $(\alpha_i+\gamma_i)a^2_{t-i}$ with $\gamma_i > 0$.

ARMA-GARCH model

ARMA(p,q) Model

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

where ε_t is a white noise series and p and q are non-negative integers.

ARMA(p,q)-GARCH(m,n) model

$$r_t = C + \underbrace{\sum_{i=1}^{p} \phi_i r_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}}_{\text{ARMA part}} + \underbrace{\sigma_{t|t-1}\varepsilon_t}_{\text{GARCH part}},$$

$$\sigma^2_{t|t-1} = \omega + \textstyle\sum_{j=1}^{m} \beta_j \sigma^2_{t-j|t-j-1} + \sum_{i=1}^{n} \alpha_i a^2_{t-i}$$

Research background

- For certain stocks:
  - the expected revenue is measured by the conditional expectation of the return rate $E(r_t \mid F_{t-1})$,
  - the risk size is measured by the conditional variance of the return rate $Var(r_t \mid F_{t-1})$.
- The ARMA-GARCH models
  - sketch the conditional mean and conditional variance of time series simultaneously and
  - have good interpretability and extensibility.
- Two problems hinder the application of ARMA-GARCH models in the Chinese stock market:
  - more evidence on the validity of ARMA-GARCH model extensions (later introduced) for the Chinese stock market needs to be collected,
  - the model establishment process is unreasonable .

Research object

In this study we compare:

- the ARMA-GARCH model extensions, including:
  - GARCH family models like: GARCH, IGARCH, TGARCH,
  - Heavy-tailed innovation distributions like: Normal distribution, Student's t distribution, Generalized error distribution and their skewed extensions,
- two model establishment methods, including:
  - the Tsay's method[4]: determines the order of the ARMA-GARCH model sequentially then estimates the parameters jointly,
  - the All possible regression method: directly estimates all the models within the search scope and selects the best model with the lowest AIC.

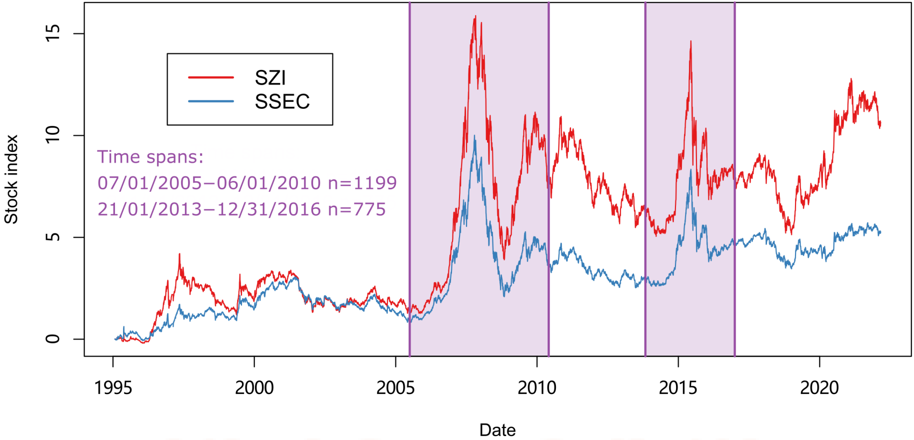
Abbreviation

To save space, in this paper we use the following abbreviations:

SSEC	is the Shanghai Stock Exchange Composite Index,
SZI	is the Shenzhen Component Index,
N	is the Normal distribution,
T	is the Student's t distribution,
GED	is the Generalized error distribution,
SN	is the skewed Normal distribution,
ST	is the skewed Student's t distribution,
SGED	is the skewed Generalized error distribution.

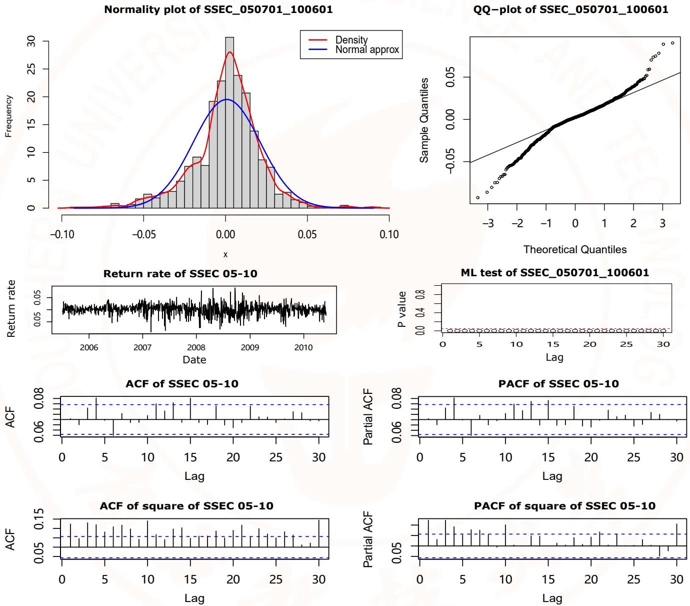
Dataset

Figure 1: Closing price series of SSEC and SZI and selected time spans



Time spans contain the bull, bear and consolidation period are chosen to capture the complete dynamic.

Figure 2: Exploration data analysis on one dataset



Research scheme

Table 1: Scopes of the model establishment for different schemes

Scheme	ARMA	Conditional Var model		Innov. dist.	Model selection method
		Model	Scope		
1	(0,0)-(4,4)	GARCH	(0,0)-(3,3)	N	Tsay's method
2	(0,0)-(4,4)	GARCH	(0,0)-(3,3)	N	All possible regression
3	(0,0)-(4,4)	GARCH TGARCH IGARCH	(0,0)-(3,3)	N	All possible regression
4	(0,0)-(4,4)	GARCH TGARCH IGARCH	(0,0)-(3,3)	N, T, GED	All possible regression
5	(0,0)-(4,4)	GARCH TGARCH IGARCH	(0,0)-(3,3)	N, T, GED, SN, ST, SGED	All possible regression

- Scheme 3 and 4, provide the local best models rest on the Normal and non-skewed assumption respectively. Thus we can find the right balance between model complexity and computational efficiency for the All possible regression method.

Estimation result

Table 2: The best models and corresponding AIC

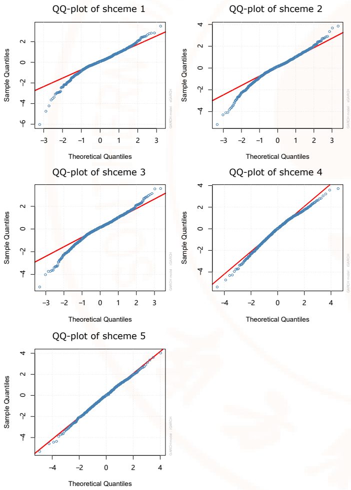
Dataset	Scheme	05-10					13-16				
		ARMA-GARCH	μ	ω	Innov. Dist.	AIC	ARMA-GARCH	μ	ω	Innov. Dist.	AIC
SSEC	1	(2,2)-(1,1)	0	0	N	-5.109	(3,3)-(1,1)	0	0	N	-5.736
	2	(4,4)-(1,1)	0	0	N	-5.112	(3,3)-(1,1)	0	0	N	-5.736
	3	(2,2)-I(1,1)	0	0	N	-5.113	(5,4)-I(1,1)	0	0	N	-5.746
	4	(4,3)-I(1,1)		0	GED	-5.192	(3,3)-I(1,1)		0	GED	-5.847
	5	(4,3)-I(1,1)	0	0	SGED	-5.205	(3,3)-I(1,1)	0	0	SGED	-5.847
SZI	1	(2,2)-(1,1)	0	0	N	-4.891	(3,3)-(1,1)	0	0	N	-5.346
	2	(4,3)-(1,1)	0	0	N	-4.894	(2,2)-(1,1)	0	0	N	-5.353
	3	(4,3)-I(1,1)	0	0	N	-4.898	(2,2)-I(1,1)	0	0	N	-5.357
	4	(1,1)-I(1,1)	0	0	GED	-4.949	(2,2)-I(1,1)	0	0	GED	-5.456
	5	(3,3)-I(1,1)	0	0	SGED	-4.964	(2,2)-I(1,1)	0	0	SGED	-5.466

[a] I stands for the IGARCH model.

- The All possible regression method shows better performance in the ARMA-GARCH model establishment.
- For the conditional variance model, the IGARCH (1,1) is recommended.
- For heavy-tailed innovation distributions, the GED and SGED are recommended based on non-skewed and skewed innovation distribution assumptions respectively.

Residual analysis

Figure 3: QQ-plots of five schemes on dataset SSEC 05-10



From the BDS-test, residuals are independent under 95% of confidence level for all schemes.

- Scheme 1, 2 and 3 [Normal]: scatter points deviate from qq-lines on both sides, especially on the left-hand side.
- Scheme 4 [GED]: the scatter points converge towards the qq-lines but shows a left bias.
- Scheme 5 [SGED]: the scatter points fit the qq-lines quite well even for outliers.

Prediction

Figure 4: ten-days ahead predictions and corresponding confidence intervals

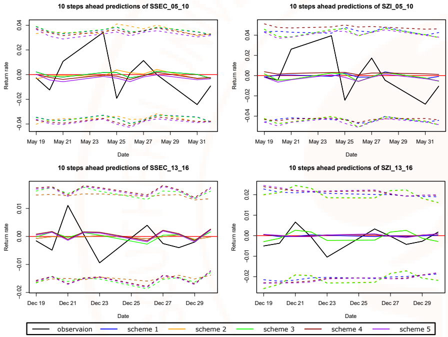


Table 3: Loss functions ($\times 1000$) of predictions

	Scheme 1	Scheme 2	Scheme 3	Scheme 4	Scheme 5
			SSEC 05-10		
MSE	0.261	0.260	0.261	0.253	0.252
MAE	13.188	12.517	13.184	12.742	12.184
			SZI 05-10		
MSE	0.416	0.406	0.406	0.410	0.400
MAE	16.076	15.808	15.805	15.580	15.409
			SSEC 13-16		
MSE	0.030	0.030	0.035	0.039	0.039
MAE	4.282	4.282	4.567	4.889	4.737
			SZI 13-16		
MSE	0.024	0.020	0.020	0.024	0.024
MAE	3.967	3.943	3.949	3.769	3.700

- Scheme 1 with 2: the best models of the All possible regression method yield better prediction performance in three datasets.
- Scheme 2-5, the IGARCH-SGED achieves less loss in time span 05-10 but the GARCH-Normal are better in time span 13-16.
  Because the sample size in time span 13-16 (775) is much less than that of time span 05-10 (1199).

Discussion

Table 4: Comparison of two model selection methods

	Tsay's method	All possible regression method
Computational requirement	Low	High
Expansibility	Only applicable to determine the orders of certain ARMA-GARCH models	Applicable for different GARCH models and innovation distributions
Model establishing procedure	Determine the orders of ARMA and GARCH sequentially	Determine the orders of ARMA and GARCH simutaneously
User requirement	Specify orders from the EACF plots twice at least	No statistical discrimination required from the user
Object	Find a significant model but the optimality isn't checked	Report the significant model outperforming other models under AIC

Conclusion

Conclusion:

- The All possible regression method is recommended.
- The IGARCH (1,1) model is recommended to be the conditional variance model to eliminate the computational intensity.
- The Generalized error distribution and skewed Generalized error distribution are suitable for large datasets ($n>1000$ in our case) under non-skewed and skewed assumption respectively.
- The GARCH (1,1) model and Normal distribution is preferred for small datasets ($n<1000$ in our case) to avoid overfitting.
- The choice of the searching scope of the All possible regression method can follow the rules above to simplify computation.

Further study:

- To capture the leverage property, TGARCH models should be built for different market states individually.
- For the All possible regression method, new criteria are needed to moderate overfitting in small datasets.

References

[1] Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

[2] Robert F. Engle and Tim Bollerslev. Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1):1–50, 1986.

[3] Jean-Michel Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955, 1994.

[4] Ruey S. Tsay. *Analysis of financial time series*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, NJ, 2. ed. edition, 2005.