基于 ARMA-GARCH 族模型的中国股票市场波动率分析
毕业论文中期汇报

黄秋逸

南方科技大学

Mar. 6 2022

Table of Contents

Timeline

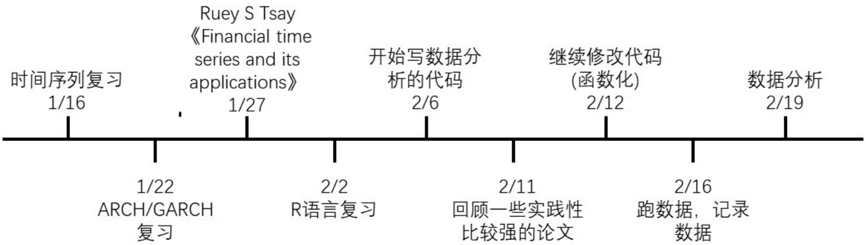


Figure 1: Research timeline

Table of Contents

Objective

Our objective is to introduce an ARMA-GARCH family model under six different innovation distribution assumptions. We will apply two different methods to perform the model selection and compare the results between them. The model is used to estimate the conditional variances of the Chinese stock market. The ARCH effect is found to be significant and this model shows some shared manner between two Chinese major stock indexs.

Table of Contents

ACF & PACF

Let's begin with some basic definitions in the time series analysis.

Autocorrelation Function

Consider a time series $\{r_t\}$, the correlation function of r_t and its past values $r_{t-\ell}$ is generalized as autocorrelation function at lag ℓ:

$$\rho_{\ell}=\frac{\operatorname{Cov}\left(r_t, r_{t-\ell}\right)}{\sqrt{\operatorname{Var}\left(r_t\right) \operatorname{Var}\left(r_{t-\ell}\right)}}$$

Partial autocorrelation function

Further, the correlation between r_t and its past value r_{t-l} after removing the effect of the intervening variables $r_{t-1} \cdots r_{t-l+1}$ is of interest. This coefficient is called the partial autocorrelation at lag l:

$$\phi_{\ell \ell}=\operatorname{Corr}\left(r_t, r_{t-\ell} \mid r_{t-1} \cdots r_{t-\ell+1}\right)$$

ARMA Model

AR(p) Model

$$r_t = \phi_0 + \phi_1 r_{t-1} + \cdots + \phi_p r_{t-p} + \varepsilon_t$$

where p is a non-negative integer and $\{\varepsilon_t\}$ is assumed to be a white noise series with mean zero and constant variance. This model says that the past p values $r_{t-i},\ i = 1, 2, \cdots, p$ jointly determine the conditional expectation of r_t given the past data.

MA(q) Model

$$r_t = c_0 + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

where q is a non-negative integer and $\{\varepsilon_t\}$ is assumed to be a white noise series with mean zero and constant variance.

ARMA Model

ARMA(p,q) Model

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

which is:

$$\left(1 - \phi_1 B^1 - \cdots - \phi_p B^p\right) r_t = \phi_0 + \left(1 - \theta_1 B^1 - \cdots - \theta_q B^q\right) \varepsilon_t$$

where ε_t is a white noise series and p and q are non-negative integers. Backshift notation B ensures $B * r_t = r_{t-1}$.

- The AR(p) and MA(q) models are special cases of the ARMA(p, q) model.
- The roots of $1 - \phi_1 x^1 - \cdots - \phi_p x^p = 0$ should lies out of the unit circle.

ARMA Model Order Determination

- We specify the order of AR and MA by ACF and PACF:

Model	AR(p)	MA(q)	ARMA(p,q)
ACF	Not cut off	Cut off at lag q	Not cut off
PACF	Cut off at lag p	Not cut off	Not cut off

- For ARMA model, we use EACF and Infomation criteria.

AR/MA	0	1	2	3	4	5	6
0	o	o	o	x	o	x	o
1	x	o	o	x	o	o	o
2	x	x	$\boldsymbol{o}$	o	o	o	o
3	x	x	x	o	o	o	o
4	o	x	x	x	o	o	o
5	o	x	x	x	x	o	o
6	x	x	x	x	x	x	o

Characteristics of Finantial Time Series

The most beautiful characteristics of finantial time series are derived from the continously compounded return.

Continously Compounded Return

$$r_t = \ln P_t - \ln P_{t-1}$$

The continuously compounded multiperiod return is simply the sum of continuously compounded one-period returns involved.

Characteristics of finantial time series

1. The series r_t is either serially uncorrelated or with minor serial correlations. However, the squared series r_t^2 is serially correlated.
2. The volatility is clustered, asymmetric and bounded.
3. Leptokurtosis.

ARCH Model

ARCH(q) Model (Engle 1982)

Series $\{a_t\}$ follows an ARCH(q) model if

$$a_t = \sigma_{t|t-1}\varepsilon_t,$$
$$\sigma^2_{t|t-1} = \alpha_0 + \alpha_1 a^2_{t-1} + \cdots + \alpha_q a^2_{t-q}$$

where $\alpha_0 > 0$, and $\alpha_i \geqslant 0$ for $i > 0$. $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean zero and variance 1, which is often assumed to follow the Normal, Student or GED distribution.

- $\mathrm{E}(a_t) = 0$
- $\sigma^2_{t|t-1} = \mathrm{Var}(a_t \mid F_{t-1}) = \mathrm{E}(a^2_t \mid F_{t-1})$
- ARCH(q) for a_t implies AR(q) for a^2_t.
- $\sum_{i=1}^{q} \alpha_i < 1$, under the assumption of finite variance.

GARCH Model

GARCH(p, q)

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

$$a_t = \sigma_{t|t-1}\varepsilon_t,$$
$$\sigma^2_{t|t-1} = \alpha_0 + \sum_{i=1}^{q} \alpha_i a^2_{t-i} + \sum_{j=1}^{p} \beta_j \sigma^2_{t-j|t-j-1}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_i \geq 0$ for $i > 0$.

- 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.

- IGARCH: An integrated $\text{GARCH}(1,1)$ model can be written as

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

The unconditional variance of a_t, hence that of r_t, is not defined under the above IGARCH(1,1) model.

- TGARCH:

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

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 $\gamma_i > 0$.

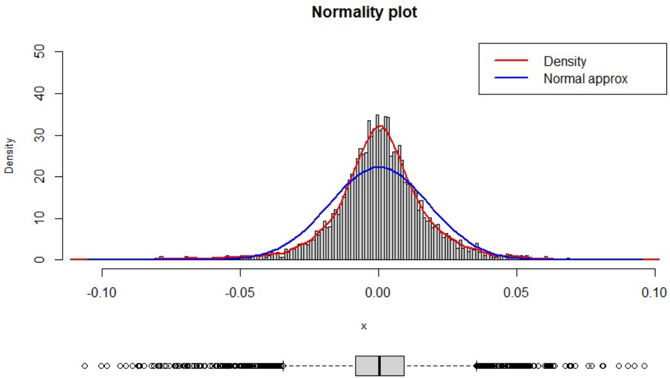
Table of Contents

Data Selection

- *A complete cycle of the stock market, including bull market, bear market and consolidation.*
  —《不同分布假设 GARCH 模型择优及其在股市波动溢出效应中的研究》江西财经大学

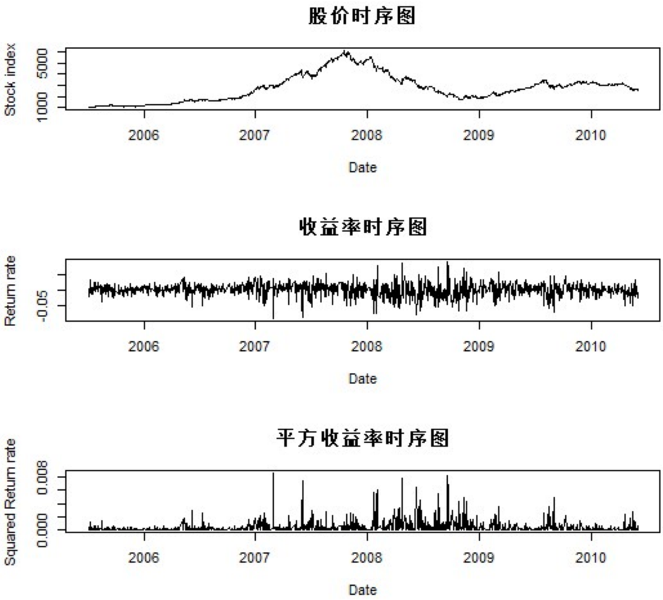
Statistics Characteristics

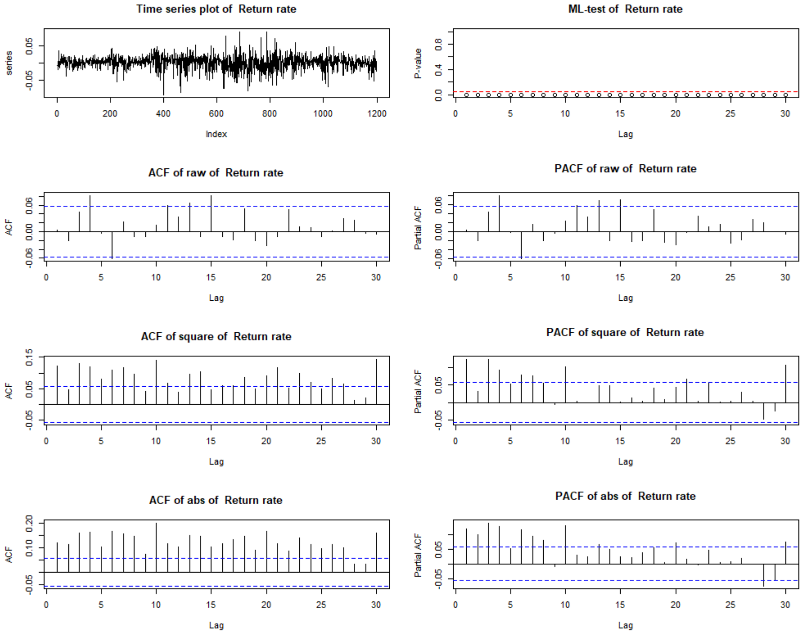
- Stock market, interval, N.
- Transform to return rate.
- Basic statistics.
- Characteristic plots.



Basic Statistics for Four Datasets

		SSEC	SZI
2005/07/01-2010/06/01	n	1199	1199
	min	Date Close Rate 2007-02-27 2771.79 -0.093	Date Close Rate 2007-02-27 7790.82 -0.098
	max	Date Close Rate 2008-09-19 2075.09 0.090	Date Close Rate 2008-04-24 12914.76 0.092
	sd	0.02042495	0.02245218
	skewness	-0.434222	-0.4328163
	kurtosis	2.336674	1.757114
2013/11/01-2016/12/31	n	775	775
	min	Date Close Rate 2015-08-24 3209.91 -0.089	Date Close Rate 2015-06-26 14398.78 -0.086
	max	Date Close Rate 2015-07-09 3709.33 0.056	Date Close Rate 2015-09-16 9890.43 0.063
	sd	0.01740382	0.01978073
	skewness	-1.207604	-0.9666158
	kurtosis	5.448543	3.513512


股价时序图
Stock index
1000
5000
2006
2007
2008
2009
2010
Date
收益率时序图
Return rate
-0.05
2006
2007
2008
2009
2010
Date
平方收益率时序图
Squared Return rate
0.000
0.008
2006
2007
2008
2009
2010
Date

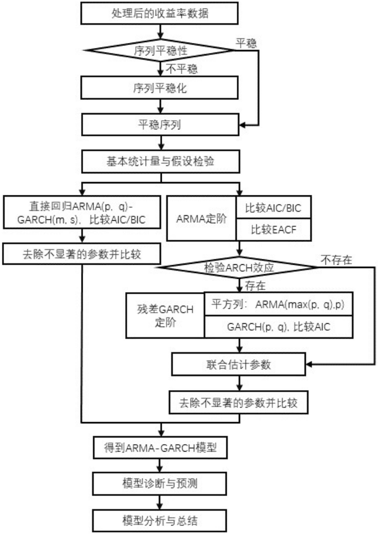

Time series plot of Return rate
series
0.05
-0.05
0
200
400
600
800
1000
1200
Index
ML-test of Return rate
P-value
0.0
0.4
0.8
Lag
ACF of raw of Return rate
ACF
PACF of raw of Return rate
Partial ACF
0.06
0.00
-0.06
ACF of square of Return rate
PACF of square of Return rate
0.15
0.05
-0.05
ACF of abs of Return rate
PACF of abs of Return rate
0.20
0.10
0.05
-0.05
0
5
10
15
20
25
30

Basic Statistical Tests

- Normality: Normality plot, QQ-plot, kstest, jbtest, swtest.
- Unit root: ADF test.
- Serial correlation: ACF & PACF, BDS test(H_o: $i.i.d.$ sereis.)
- ARCH effect: Engle's LM test, McLeod-Li test.

Modelling Methods

- Method 1:
  Building a volatility model usually consists five steps:
  1.Specify a mean equation.
  2.Use the residuals of the mean equation to test for ARCH effects.
  3.Specify a volatility model.
  4.Perform a joint estimation.
  5.Check the fitted model.
  — Ruey. S. Tsay, Analysis of Financial Time Series
- Method 2:
  Compare the AIC/BIC among all possible ARMA-GARCH models.



Innovation distribution

As what we mentioned earlier, the finantial time series is leptokurtosis, negative biased and failed to pass the normality tests. Moreover, the adjusted pearson goodness-of-fit test also suggests normal distribution is not good enough.

Common used innovation distributions

- Symmetric distributions:
  Normal, Student, GED (generalized error distribution).
- Asymmetric distributions:
  Skewed normal, Skewed Student, Skewed GED.

Scope of potential models for model selection

Scope of potential models for model selection		
	Tsay 方法	全部回归
ARMA	√	√
GARCH family: including GARCH, IGARCH, TGARCH model	GARCH only	√
Innovation distribution: including Normal, Student, GED, Skewed normal, Skewed Student, Skewed GED.	Normal only	√
Parameter restrictions: Omega=0, mu=0	√	√

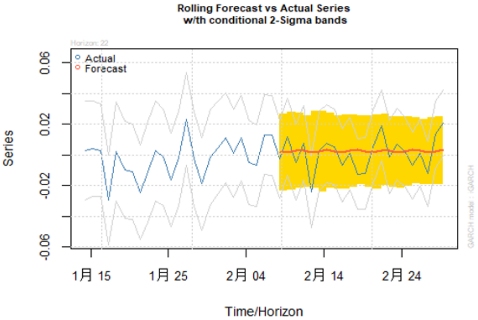
Conclusion

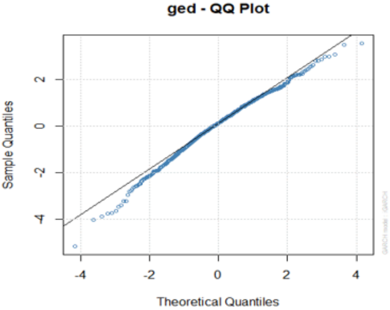
数据集		只包含正态分布 Tsay 方法	只包含正态分布 全部回归	不包含偏分布	包含偏分布
05-10	SSEC	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-NORM	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-NORM	ARMA(2,2)-IGARCH(1,1)-omega=0-GED	ARMA(2,2)-IGARCH(1,1)-omega=0-SGED
	SZI	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-NORM	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-NORM	ARMA(2,2)-IGARCH(1,1)-omega=0-GED	ARMA(2,2)-IGARCH(1,1)-omega=0-SGED
13-16	SSEC	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-ar1=0-ma1=0-NORM	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-ar1=0-ma1=0-NORM	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-ar1=0-ma1=0-GED	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-ar1=0-ma1=0-SGED
	SZI	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-NORM	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-NORM	ARMA(2,2)-IGARCH(1,1)-omega=0-GED	ARMA(2,2)-IGARCH(1,1)-omega=0-mu=0-SGED

- With skewed distribution, the SGED is optimal among all datasets.
- Without skewed distribution, the GED is the optimal one.
- Generally, the optimal model is ARMA(2,2)-IGARCH(1,1)-omega=0.
- The TGARCH model is not significant among all datasets.
- Mu is not significant under normal innovation distribution, but significant under GED innovation distribution.

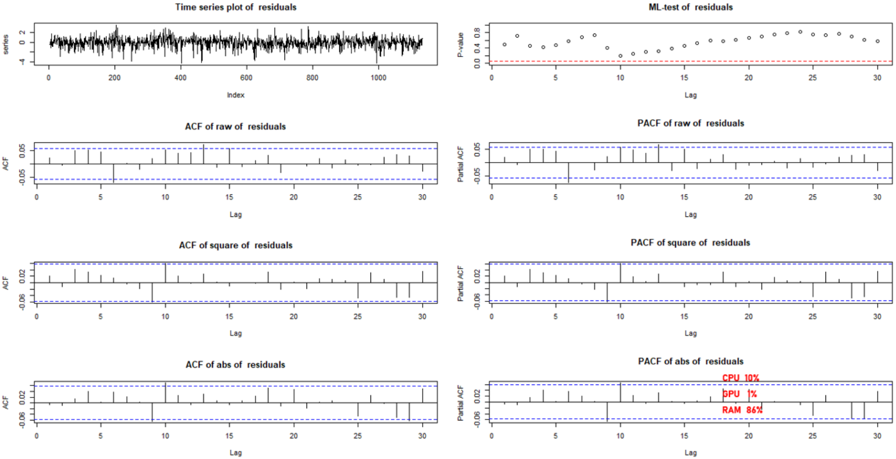
Residuals Analysis

- QQ-plot
- Serieial correlation: ACF & PACF, BDS test
- Prediction





Residuals Analysis



Discussion

Discussion 1: Robustness of two model selection methods

两种模型选择方法在 Normal innovation 假设下的结果比较

数据集		Compare among all possible models		Tsay's method	
		ARMA	GARCH	ARMA	GARCH
050701-100601	SSEC	(2, 2)	(1, 1)	(2, 2)	(1, 1)
	SZI	(2, 2)	(1, 1)	(2, 2)	(1, 1)
131101-161231	SSEC	(2, 2)	(1, 1)	EACF/AIC 提示过大的模型,BIC 提示 (2, 2)	(1, 1)
	SZI	(2, 2)	(1, 1)	EACF/AIC 提示过大的模型,BIC 提示 (3, 2)	EACF/AIC 提示过大的模型,BIC 提示 ARMA(3, 2) 即 (3, 2) [a]

[a] Note: 在 Joint estimation 中，一些参数不显著，消去之后得到 **ARMA(2,2)-GARCH(1,1)** 模型。

Discussion

Discussion 1: Robustness of two model selection methods

回顾：2015 年后半年打击融资融券交易：

- 总体来看融资融券交易还是可以有效地平抑市场波动。
  ——《融资融券对中国股票市场波动影响的研究》江西财经大学
- 深圳市场受两融交易的影响比上海市场更大。
  ——《融资融券对中国 A 股市场波动的影响—基于事件研究法下沪深两市的比较研究》

提示：深圳市场受两融交易限制，救市政策影响，熔断政策影响较大，市场发生超出正常运行的波动,Tsay 方法鲁棒性不足，估计出现较大偏差。

The ARCH model only takes finite past realizations into consideration, thus its memory is not long enough, which result in a larger parameters. Bollerslev(1986) proposed a strong extension known as the generalized ARCH model. The introduction of past conditional

Discussion

Discussion 2: TGARCH Not Significant

其次, 从表 2-7 和表 2-8 中我们可以看出 EGARCH(1,1) 模型和 GJR-GARCH(1,1) 模型中用来刻画波动不对称性的参数均不显著, 这说明在所研究时间段中上证 180 指数的波动并不存在明显的杠杆效应, 这与我们最常接触的假定条件是有所出入的。在此笔者推断可能是由于本次选取的样本数据跨越时间较长, 且包含了一段完整的牛市熊市的转换, 不同市场状态的转换可能会一定程度上抵消波动的不对称特征, 从而使得对于整体波动特征的研究呈现出杠杆效应不显著的结果。另外还可能是由于本论文选择的样本数据是成分指数上证 180 收益率序列, 其中包含的样本均是规模大、行业代表性强的公司股票, 受内部负面消息及外部冲击的影响程度相对较小。同时随着投资知识的不断普及, 投资者在进行投资行为时也逐渐趋于理性, 从而使得上海股票市场的不对称效应没有之前那样显著。

— 《基于 GARCH 模型的上证 180 指数收益率波动研究》中南财经政法大学

还有可能是由于我国股票市场受政策影响较大, 在大的下跌之后可能伴随着救市政策带来的大的上涨, 从而在一定程度上抵消了市场的不对称性。

以上, 我们虽然能够观察到收益率序列的负偏态, 但是无法捕捉到显著的不对称性。提示我们: 更换收益率数据, 讨论描述不对称性的模型的有效性。

Discussion

Discussion 3: Further Research

- 更换数据时间段, 区分牛市, 熊市, 盘整期进行研究, 寻找 TGARCH 效应。
- 更换高频数据 (10min 数据)。
- 拓展两个讨论, 即: 讨论两种估计方法; 讨论 15 年股灾对 Tsay 估计方法的影响。