한국방송통신대학교

Forecasting Methods

예측방법론
Forecasting Methods

08 강

시계열모형을 이용한 예측(2)

통계·데이터과학과 **이긍희** 교수

Forecasting Methods

한국방송통신대학교

목차

CONTENTS

Forecasting Methods

chapter 01

ARIMA 모형의 추정

1 추정방법

- ARIMA 모형의 추정 : 최대가능도추정법, 조건부. 비조건부 최소제곱(최소자승)추정법 등으로 모수값 추정

» $W_t = \mu + \phi_1 W_{t-1} + \cdots + \phi_p W_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$

$W_t = \Delta^d Y_t$

» 모수 : $\phi_1, \cdots, \phi_p, \theta_1, \cdots, \theta_q, \mu, \sigma^2$

» 최대가능도추정법
: 가능도함수, 즉 결합확률밀도함수 최대화하여 모수 추정

» 최소제곱추정법
: 오차 제곱합을 가장 작게 하는 모수 추정량 구하는 방법

2 추정량의 분포

- 통계량(추정량) 은 근사적으로 t분포

» 추정량 : $\hat{\mu},\ \hat{\phi}_1,\ \cdots,\ \hat{\phi}_p,\ \hat{\theta}_1,\ \cdots,\ \hat{\theta}_q$

$$\widehat{W}_t = \hat{\mu} + \hat{\phi}_1 W_{t-1} + \cdots + \hat{\phi}_p W_{t-p} + \hat{\theta}_1 \varepsilon_{t-1} + \cdots + \hat{\theta}_q \varepsilon_{t-q}$$

» 검정
: 모수들이 각각 0과 다른지 검정, 기각역 및 유의확률

» 추정값
: 모수에 추정값 대입, 오차 제거하여 구체화

3 추정의 예

- **로그변화된 종합주가지수의 ARIMA 모형의 추정**

① ARIMA(1,1,1)

```
          ar1     ma1
      -0.6820  0.8634
s.e.   0.1586  0.1135

sigma^2 = 0.001387:  log likelihood = 281.6
AIC=-557.2   AICc=-557.03   BIC=-548.16
```

② ARIMA(0,1,1)

```
         ma1
      0.2179
s.e.  0.0867

sigma^2 = 0.001407:  log likelihood = 280.15
AIC=-556.3   AICc=-556.21   BIC=-550.27
```

③ ARIMA(1,1,0)

```
         ar1
      0.1846
s.e.  0.0808

sigma^2 = 0.001415:  log likelihood = 279.7
AIC=-555.4   AICc=-555.31   BIC=-549.37
```

4 자동식별과 추정

- **하인드먼과 칸다카르**(Hyndman and Khandakar) : ARIMA **모형의** (p, d, q)(P, D, Q)$_s$**를 자동 식별 → 추정량은 근사적으로** t**분포**

» 차분수 d
: 단위근검정과 계절단위근검정을 반복하여 구함

» p, q
: AIC, AICc, BIC 모형선택기준을 최소로 하는 p, q 정함

Forecasting Methods

chapter 02

ARIMA 모형의 진단

1 진단의 개요

- **진단과정** : ARIMA **모형의 식별 및 추정 단계를 거친 후 얻어진 잠정모형이 타당한지 검토 → 타당하면 예측모형 사용**

» 잠정모형의 타당성
: 잠정모형이 타당 → 예측에 사용
타당하지 않으면 → 식별과정부터 다시 검토

- **진단** : **과대적합진단, 잔차분석**

→ **두 가지 진단에 문제가 없을 때 잠정모형을 최종 예측모형으로 결정**

2 과대적합진단

- **과대적합진단 : 잠정** ARIMA **모형보다 한 단계 복잡한** ARIMA **모형 작성**, **검토**

» MA(1) 모형 → MA(2), ARMA(1,1)
: 잠정모형에 추가된 모수가 통계적으로 유의
→ 새로운 모형이 잠정모형보다 설명력이 있음

2 과대적합진단

- **로그변환된 종합주가지수 모형에 대한 과대적합진단**

• ARIMA(1,1,2) 모형의 추정

```
          ar1     ma1    ma2
      -0.9107  1.1600  0.160
s.e.   0.0434  0.0983  0.095

sigma^2 = 0.001366:  log likelihood = 282.33
AIC=-556.66   AICc=-556.39   BIC=-544.62
```

• ARIMA(2,1,1) 모형의 추정

```
oefficients:
          ar1     ar2     ma1
      -0.7060  0.0980  0.9337
s.e.   0.1598  0.1139  0.1376

sigma^2 = 0.001388:  log likelihood = 282.05
AIC=-556.1   AICc=-555.83   BIC=-544.06
```

3 잔차분석

- ARIMA(p,d,q) **모형의 오차항** ε_t **백색잡음계열**
  → **모형진단** : **오차항의 추정값인 잔차** $r_t = W_t - \widehat{W}_t$**가 백색잡음계열과 같은지 검토**

3 잔차분석

- **검토방법**

  ① 잔차에 대한 도표

  ② 상관도표와 부분상관도표

  ③ 포트맨토 검정(portmanteau test) : 륭-박스 검정

  ④ 잔차의 스펙트럼

3 잔차분석

- **검토방법**

» 상관도표와 부분상관도표

: 잔차의 표본자기상관계수와 표본부분자기상관계수가 모두 $\left(-\frac{1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}}\right)$의 구간 내에 있을 것임

3 잔차분석

- **검토방법**

» 포트맨토 검정(portmanteau test) : 륭-박스 검정

: 귀무가설 : $H_0 : \rho(1) = \rho(2) = \cdots = \rho(m) = 0$

검정통계량 : $Q = n(n+2)\sum_{h=1}^{m} \frac{\hat{\rho}^2(h)}{n-h}$

귀무가설 기각 : $Q \geq \chi^2_\alpha(m-p-q)$

3 잔차분석

- **검토방법**

» 잔차의 스펙트럼

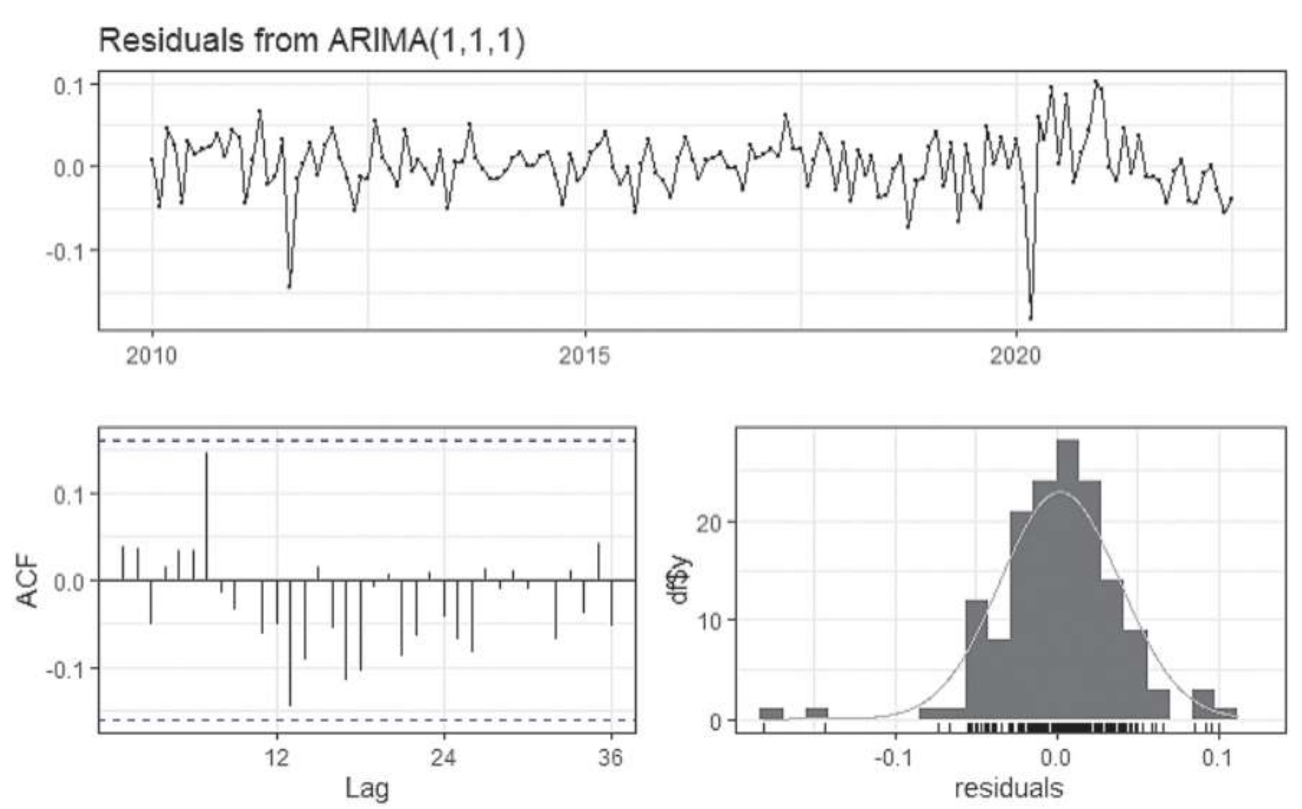
: 잔차의 스펙트럼이 평평하지 않음 → 잔차가 백색잡음이 아님

» 스펙트럼 이용 검정

: 피셔–카파(Fisher–Kappa) 검정,
콜모고로프–스미르노프(Kolmogorov–Smirnov) 검정

3 잔차분석

- **로그변환된 종합주가지수 모형의 잔차분석**



Forecasting
Methods

chapter
03

ARIMA 모형의 예측

1 예측의 개요

- **시계열분석의 최종 목적 : (아직 관측되지 않은) 시계열 예측**
- ARIMA **모형 설정 → 시계열이 이 모형의 형태로 움직일 것으로 가정, 이 모형으로 예측**

» 시계열 예측값 표현

: 관측 시계열 $\{Y_1, Y_2, \cdots, Y_n\}$

미래 시계열 $Y_{n+1}, Y_{n+2}, \cdots$

1시차 후 예측값 $Y_n(1)$, 2 시차 후 예측값 $Y_n(2)$

예측오차 $\hat{\varepsilon}_n(l) = Y_{n+l} - Y_n(l)$

1 예측의 개요

- 추정방법 : 점추정, 구간추정

» 95% 예측구간

$$Y_n(l) \pm 1.96\sqrt{Var[\hat{\varepsilon}_n(l)]}$$

- 예측 시점이 멀어지면 예측오차가 커짐

  → 시계열모형 예측은 단기예측에 유용, 장기예측에 부적당

1 예측의 개요

- ‘좋은’ 예측값 선택 → 예측오차 제곱의 기댓값(MSE) 최소화 과정을 통해 예측값 결정

$$Y_n(l) = E(Y_{n+l} | Y_1, \cdots, Y_n)$$

2 AR 모형의 예측

- AR(1) **모형**

$$Y_t = \mu + \phi Y_{t-1} + \varepsilon_t$$

» l 시차 앞 예측값

$$Y_n(1) = \hat{\mu} + \hat{\phi} Y_n$$

$$Y_n(2) = \hat{\mu} + \hat{\phi} Y_n(1) = \hat{\mu} + \hat{\phi}\hat{\mu} + \hat{\phi}^2 Y_n$$

$$\vdots$$

$$Y_n(l) = (1 + \hat{\phi} + \cdots + \hat{\phi}^{l-1})\hat{\mu} + \hat{\phi}^l Y_n$$

3 MA 모형의 예측

- MA(1) 모형

$$Y_t = \mu + \theta \varepsilon_{t-1} + \varepsilon_t$$

» l 시차 앞 예측값

$$Y_n(1) = \hat{\mu} + \hat{\theta} \cdot \hat{\varepsilon}_n$$

$$Y_n(l) = \hat{\mu}, \quad l \geq 2$$

4 ARMA 모형의 예측

- ARMA(1,1) **모형**

$$Y_t = \mu + \phi Y_{t-1} + \theta \varepsilon_{t-1} + \varepsilon_t$$

» <u>l 시차 앞 예측값</u>

$$Y_n(1) = \hat{\mu} + \hat{\phi} Y_n + \hat{\theta}\, \hat{\varepsilon}_n$$

$$Y_n(2) = \hat{\mu} + \hat{\phi} Y_{n+1} = \hat{\mu} + \hat{\phi}\hat{\mu} + \hat{\phi}^2 Y_n$$

$$\vdots$$

$$Y_n(l) = (1 + \hat{\phi} + \cdots + \hat{\phi}^{l-1})\hat{\mu} + \hat{\phi}^l Y_n$$

5 ARIMA 모형의 예측

- ARIMA(1,1,0) **모형**

$$W_t = \mu + \phi W_{t-1} + \varepsilon_t$$

$$Y_t = \mu + (1+\phi)Y_{t-1} - \phi Y_{t-2} + \varepsilon_t$$

» l 시차 앞 예측값

$$Y_n(1) = \hat{\mu} + (1+\hat{\phi})Y_n - \hat{\phi}Y_{n-1}$$

$$Y_n(2) = \hat{\mu} + (1+\hat{\phi})Y_n(1) - \hat{\phi}Y_n$$

$$\vdots$$

$$Y_n(l) = \hat{\mu} + (1+\hat{\phi})Y_n(l-1) - \hat{\phi}Y_n(l-2),\ l \geq 2$$

5 ARIMA 모형의 예측

- 로그변환된 종합주가지수의 ARIMA 모형을 이용한 예측

Forecasting Methods

chapter 04

변동성 모형의 작성 및 예측

1 변동성

- 자산수익률은 두꺼운 꼬리를 갖는 분포, 시간에 따라 변동성은 밀집하는 경향
- 변동성은 파생상품 가격추정, 위험관리, 자산배분 등에 활용
  → 변동성을 오차항의 제곱을 조건부로 추정

» ARIMA 모형

: 조건부 분산은 일정

2 ARCH 모형

- ARCH **모형** : 1982**년 엥글**(Engle)**이 시간에 따라 변하는 분산을 모형화**

» ARCH(1) 모형

$$r_t = \sigma_{t|t-1}\varepsilon_t$$

$$\sigma^2_{t|t-1} = \alpha_0 + \alpha_1 r^2_{t-1}$$

$\sigma^2_{t|t-1}$은 r_t의 조건부분산

3 GARCH 모형

- GARCH **모형** : ARCH **모형의 일반화**

» GARCH(p,q) 모형

$$r_t = \sigma_{t|t-1}\varepsilon_t$$

$$\sigma^2_{t|t-1} = \omega + \alpha_1 r^2_{t-1} + \alpha_2 r^2_{t-2} + \cdots + \alpha_p r^2_{t-p} + \beta_1 \sigma^2_{t-1|t-2} + \cdots + \beta_q \sigma^2_{t-q|t-q-1}$$

- GARCH **모형 : '식별-추정-진단'의 절차를 거쳐 작성**

3 GARCH 모형

- **GARCH 모형의 적합성 판단**

  ① 시계열은 두꺼운 꼬리를 가지는 경향

  → 정규성 검정 실시

  ② 조건부 이분산성 검정 : 시계열의 제곱에 자기상관이 존재 여부 검정

  → ARCH-LM 검정

  ③ 시계열에 ARCH 형태의 이분산성 존재

  → GARCH 모형 작성

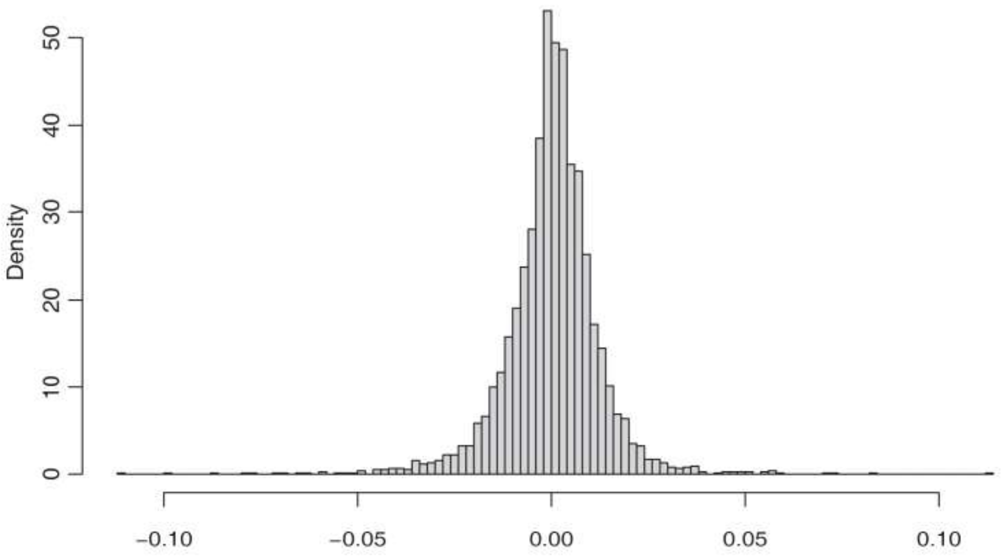
3 GARCH 모형

- 후보 GARCH 모형 추정 후 모형선택기준 비교 → 최종 모형 선정
  - GARCH 모형은 최대가능도추정법 추정
- 진단 : GARCH 모형 잔차 제곱의 상관도표, 륭-박스 검정
- 예측 : 진단 결과 GARCH 모형 잔차가 이분산성 없고 독립적
  → GARCH 모형으로 조건부분산을 추정, 예측

3 GARCH 모형

- **일별 종합주가지수 로그차분계열의** GARCH **모형 작성**

(a) 히스토그램

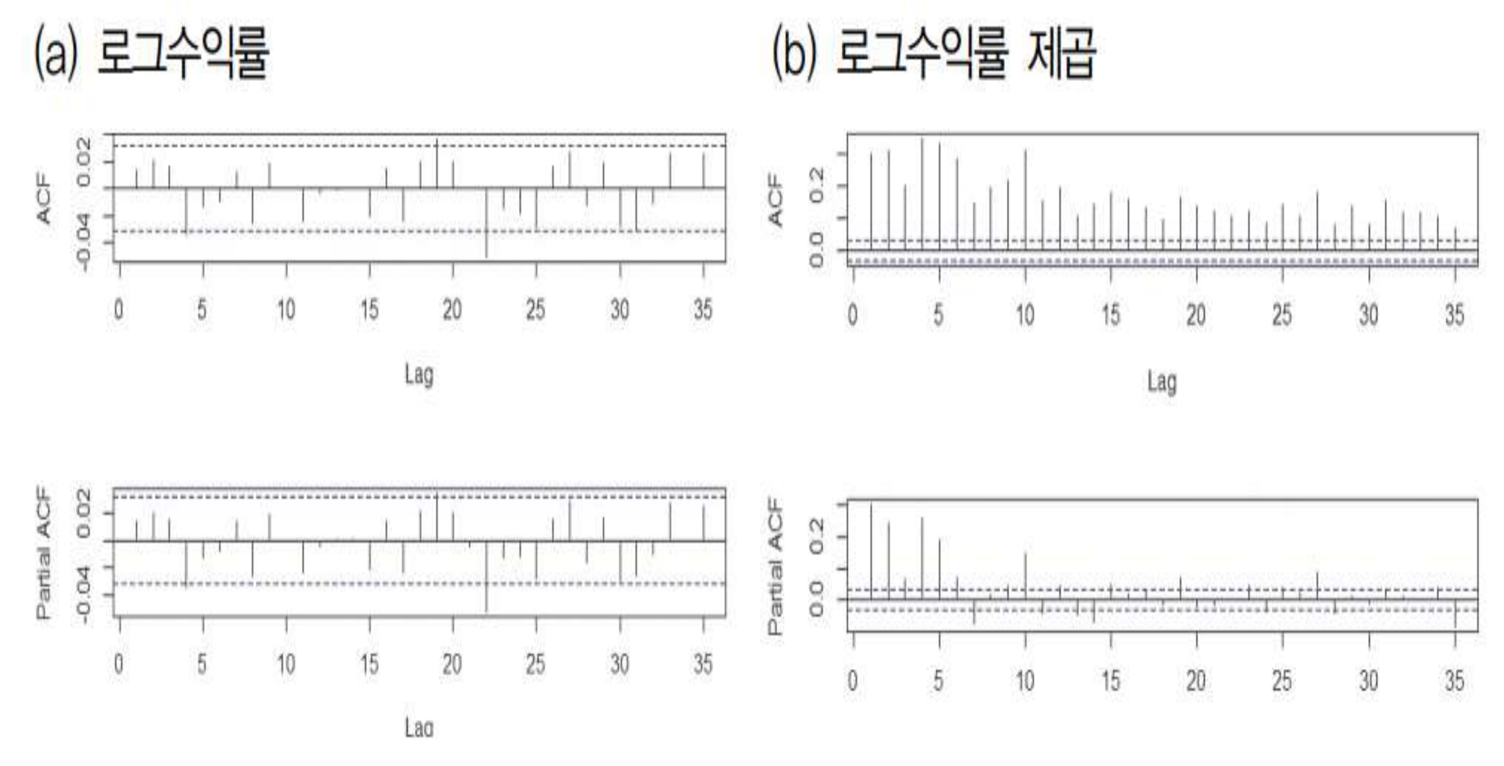


(b) 정규성 검정

```
Jarque Bera Test
        X-squared = 14864, df = 2, p-value < 2.2e-16
```

3 GARCH 모형

- **일별 종합주가지수 로그차분계열의** GARCH **모형 작성**



3 GARCH 모형

- **일별 종합주가지수 로그차분계열의** GARCH **모형 작성**

$$r_t = 0.00033$$

$$\hat{\sigma}^2_{t|t-1} = 0.000002 + 0.0885 v^2_{t-1} + 0.898 \sigma^2_{t-1|t-2}$$

	Estimate	Std. Error	t value	Pr(>|t|)
mu	0.000331	0.000143	2.3206	0.020311
omega	0.000002	0.000001	1.6075	0.107949
alpha1	0.088496	0.014265	6.2035	0.000000
beta1	0.898345	0.015186	59.1579	0.000000

Weighted Ljung-Box Test on Standardized Residuals

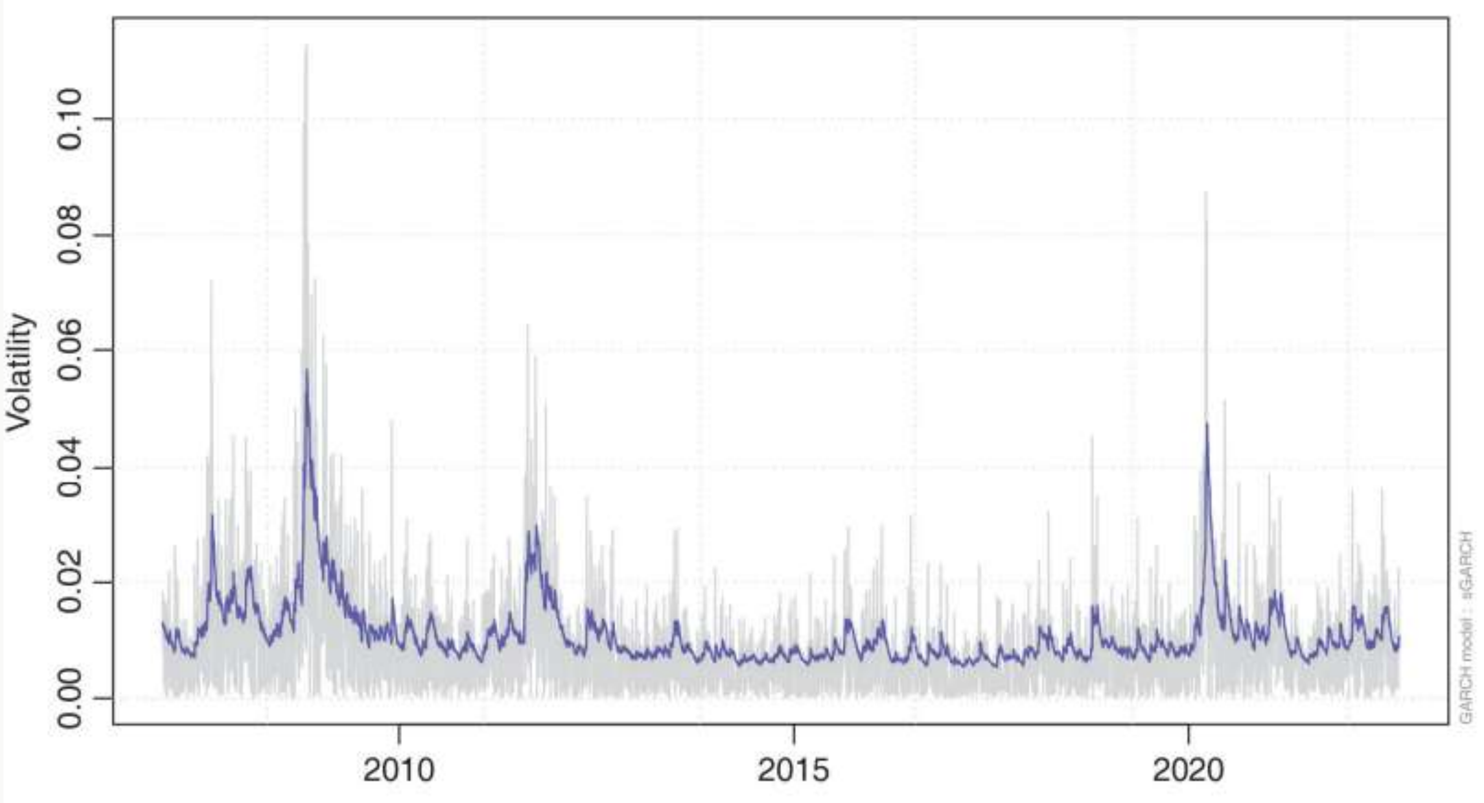
	statistic	p-value
Lag[1]	2.227	0.1357
Lag[2*(p+q)+(p+q)-1][2]	3.072	0.1342
Lag[4*(p+q)+(p+q)-1][5]	4.395	0.2088

Weighted Ljung-Box Test on Standardized Squared Residuals

	statistic	p-value
Lag[1]	0.3244	0.5690
Lag[2*(p+q)+(p+q)-1][5]	2.7034	0.4639
Lag[4*(p+q)+(p+q)-1][9]	6.0505	0.2924

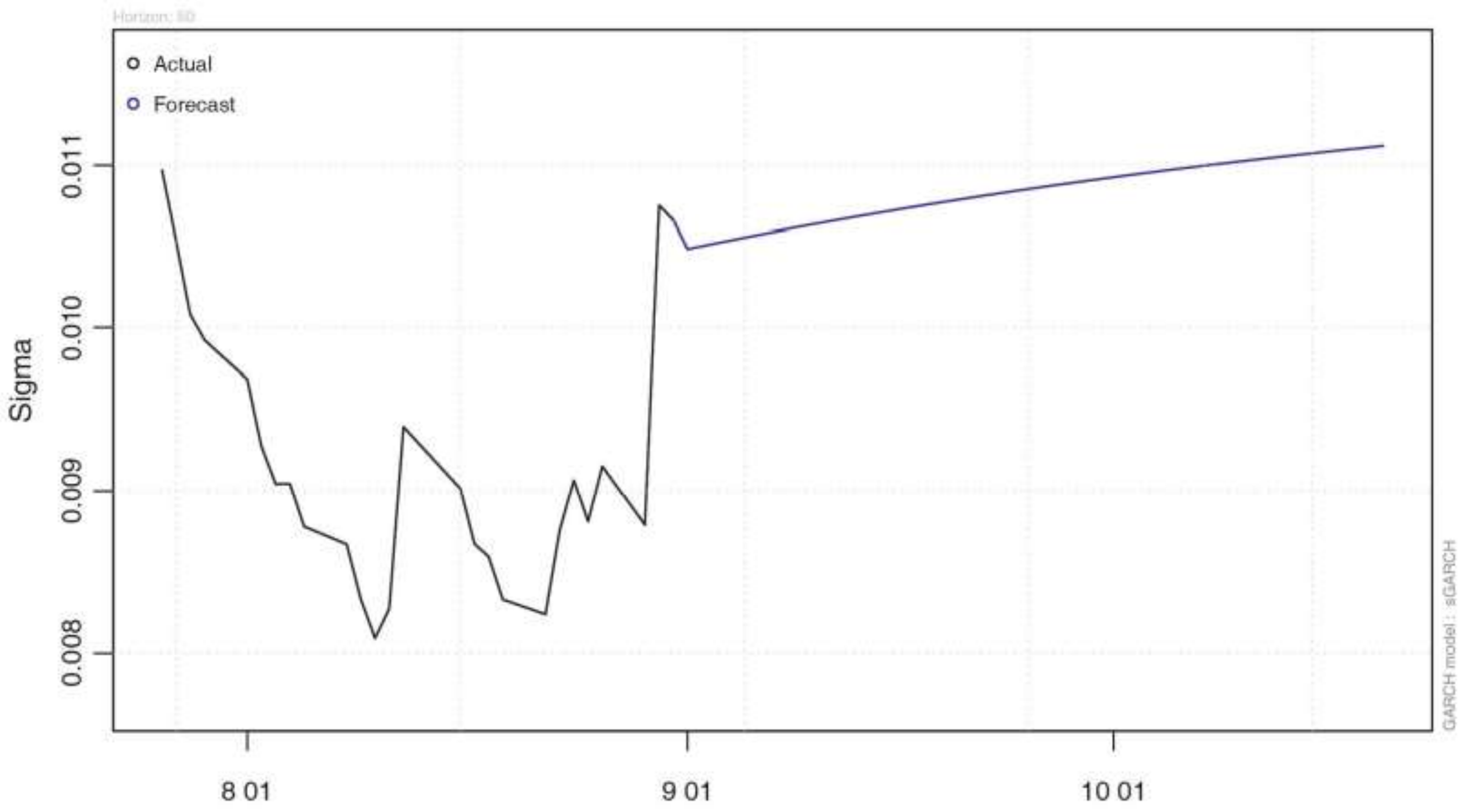
3 GARCH 모형

- 일별 종합주가지수 로그차분계열의 GARCH 모형 추정 결과



3 GARCH 모형

- GARCH **모형을 이용한 변동성 예측**



4 ARMA+GARCH 모형

- GARCH **모형은 평균함수** : ARMA **모형 또는 회귀모형으로 확장**
- **평균함수가** ARMA **모형인** GARCH **모형**(ARMA+GARCH **모형**)

$$
\begin{aligned}
r_t &= \mu + \phi_1 r_{t-1} + \cdots + \phi_u r_{t-u} + \theta_1 v_{t-1} + \cdots + \theta_v v_v + v_t \\
v_t &= \sigma_{t|t-1}\varepsilon_t \\
\sigma^2_{t|t-1} &= \omega + \alpha_1 v^2_{t-1} + \alpha_2 v^2_{t-2} + \cdots + \alpha_p v^2_{t-p} \\
&\quad + \beta_1 \sigma^2_{t-1|t-2} + \cdots + \beta_q \sigma^2_{t-q|t-q-1}
\end{aligned}
$$

Forecasting Methods

chapter 05

R을 이용한 실습

Forecasting Methods

다음시간안내

09 | 회귀모형을 이용한 예측(1)

Forecasting Methods