Національний Технічний Університет України “КПІ”
Навчально-науковий комплекс
«Інститут прикладного системного аналізу»

ЛАБОРАТОРНА РОБОТА № 5

Моделювання та оцінювання гетероскедастчних процесів

Виконавці роботи

студенти гр. КА-41
бригада № 5

Барзій Ілля
Лєсніков Богдан
Шрам Владислав

Прийняла

Кузнєцова Наталія Володимирівна

____
(підпис, дата)

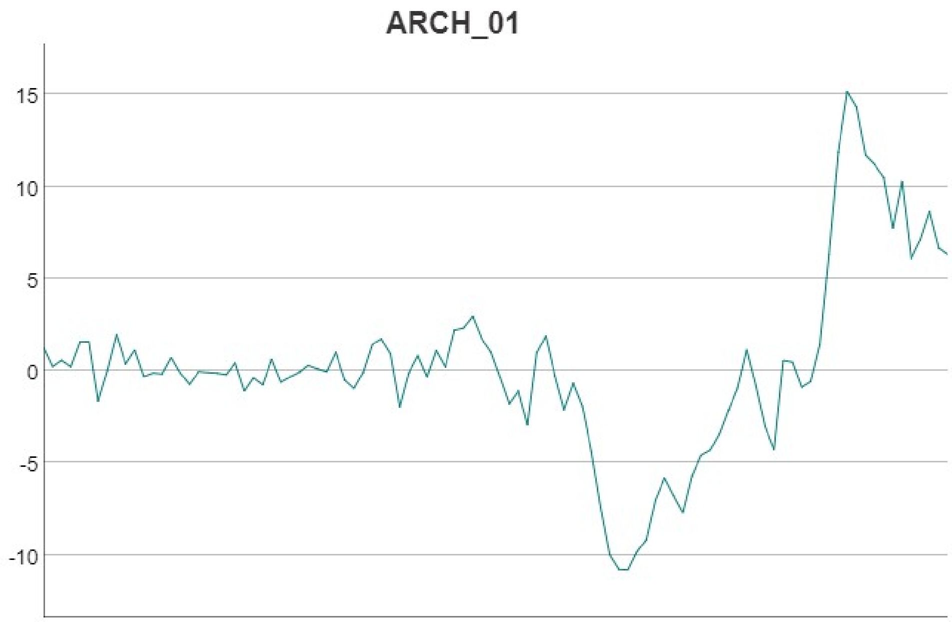
КИЇВ 2017

Мета роботи: *Навчитися аналізувати та моделювати гетероскедастичні процеси (процеси із змінною дисперсією).*

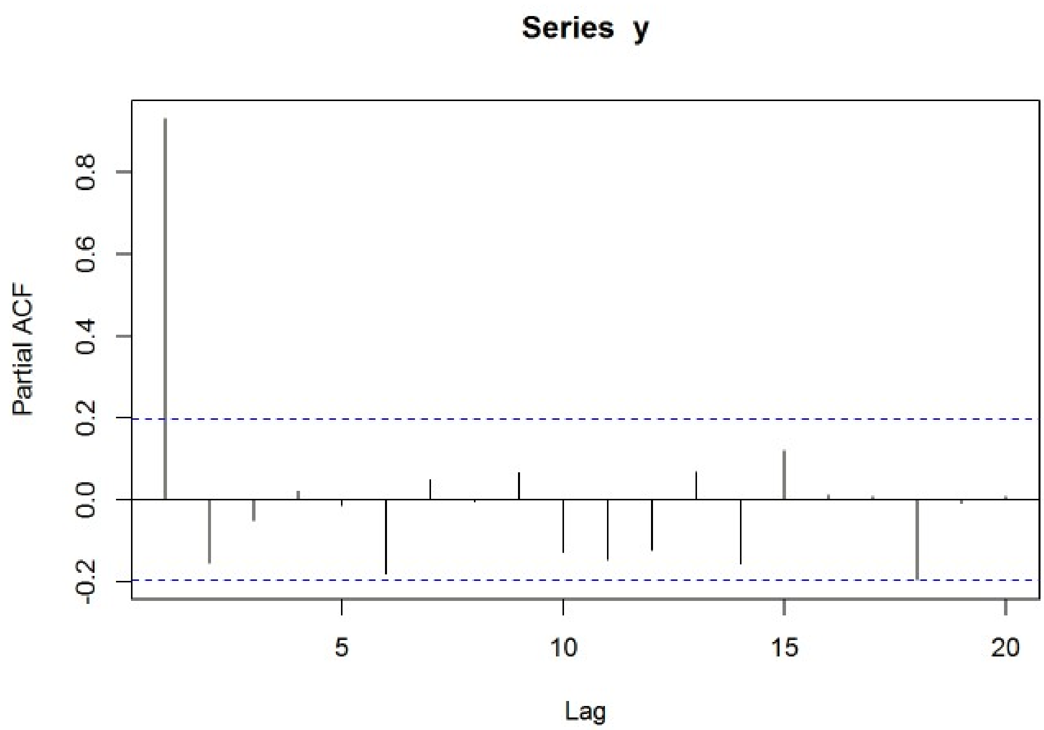
Виконання роботи :

1.3) Введіть (прочитайте) дані з вказаного вище файла з диска, знайдіть середнє значення ряду, стандартне відхилення і побудувати для введених даних модель АР(1).

```
y <- ts(scan("Data/ARCH_01.dat"))
dygraph(y, main = "ARCH_01") %>% dyRangeSelector()
```



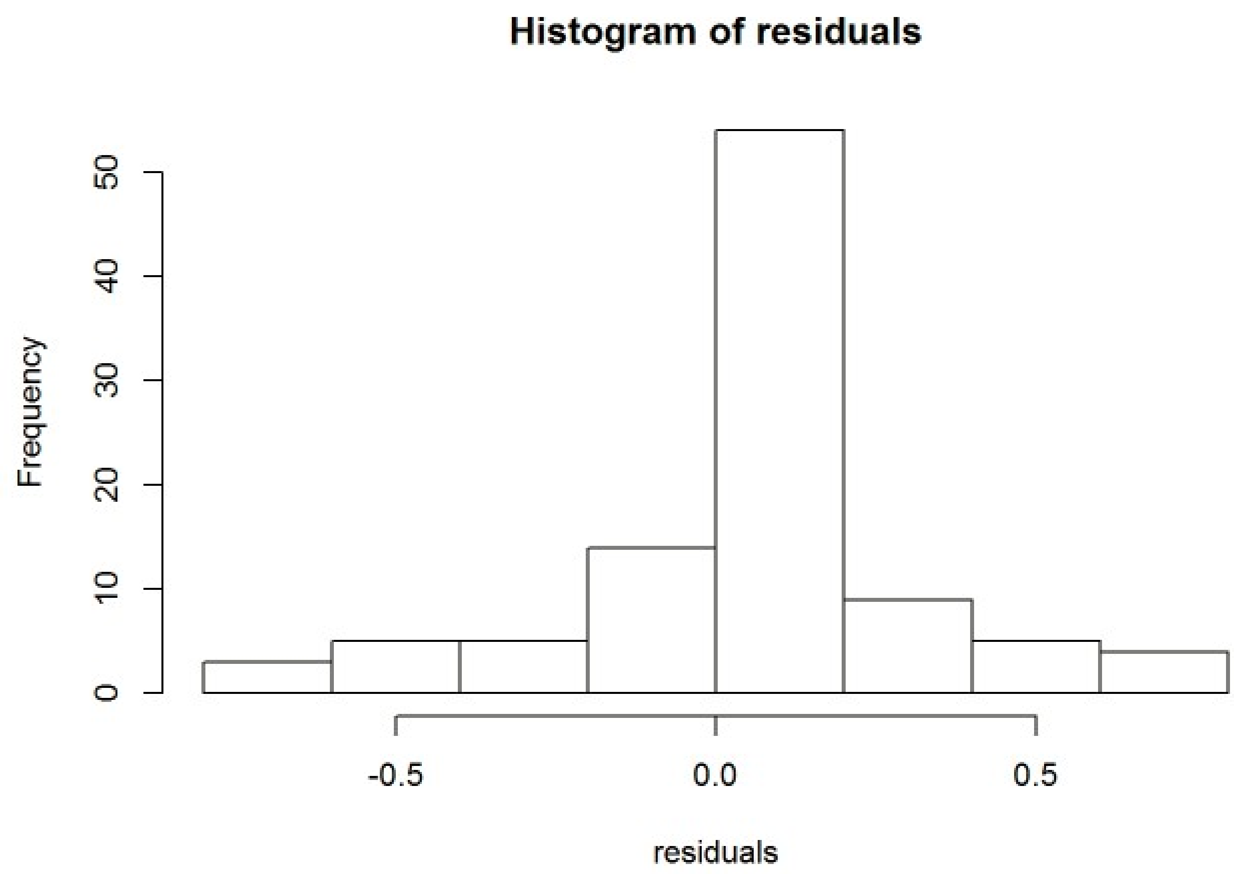

```
pacf(y)
```



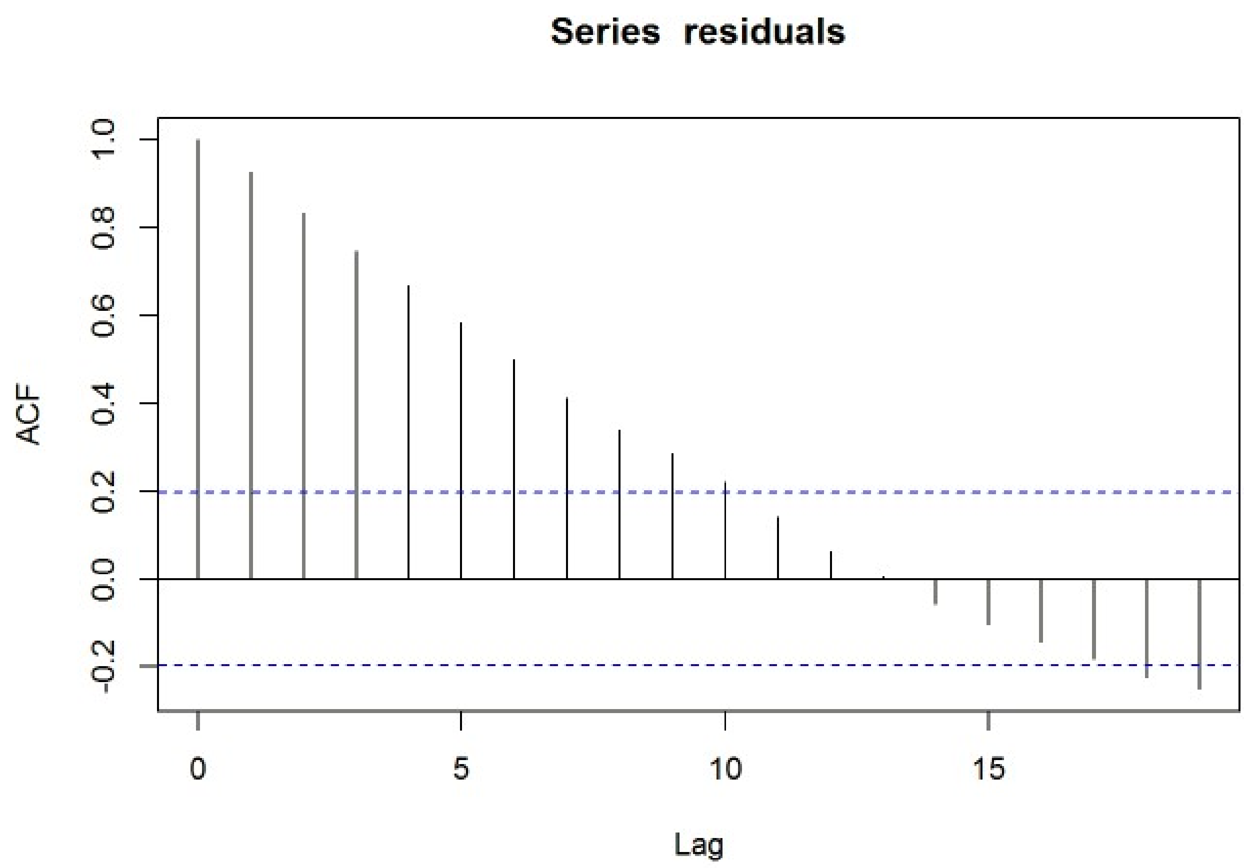
```
paste('mean = ', mean(y))
## [1] "mean =  0.26336948"
paste('sd = ', sd(y))
## [1] "sd =  4.89409139173037"
model1 <- dynlm(y ~ L(y,1))
summary(model1)
## Time series regression with "ts" data:
## Start = 2, End = 100
##
## Call:
## dynlm(formula = y ~ L(y, 1))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.6782 -1.3426 -0.0319  0.8705  5.7974
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.06219    0.17409   0.357    0.722
## L(y, 1)      0.94387    0.03582  26.350   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.731 on 97 degrees of freedom
## Multiple R-squared:  0.8774, Adjusted R-squared:  0.8762
## F-statistic: 694.3 on 1 and 97 DF,  p-value: < 2.2e-16
```

1.4) Побудуйте автокореляційну функцію (АКФ) для отриманого після побудови моделі АР(1) ряду із залишків $\hat{\varepsilon}(k)$.

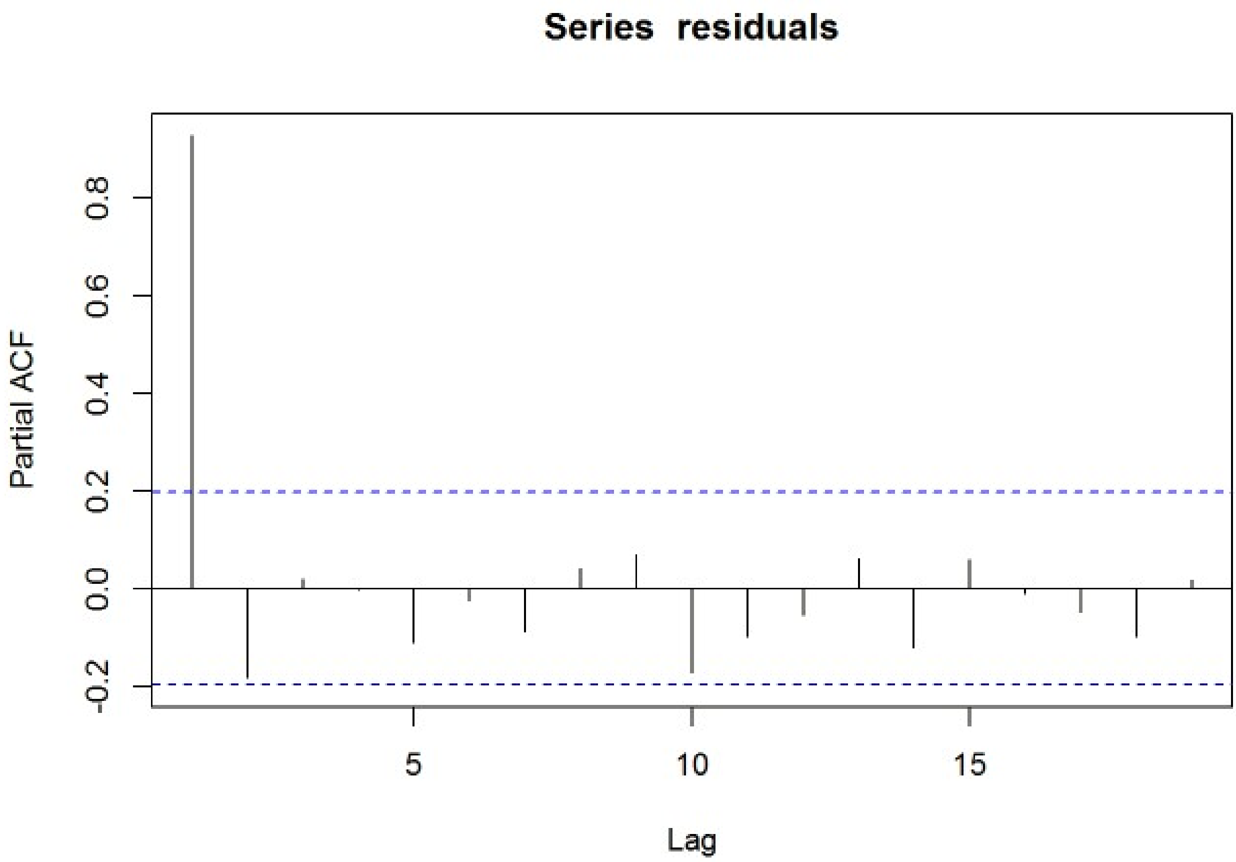
```
residuals <- lag(fitted.values(model1),1) - y
hist(residuals)
```




```
acf(residuals, na.action = na.pass)
```



```
pacf(residuals)
```



1.5) Згенеруйте новий ряд із квадратів залишків: $\{\hat{\varepsilon}^2(k)\} = \{e^2(k)\}$, де $e(k)$ – залишки (похибки) моделі АР(1).

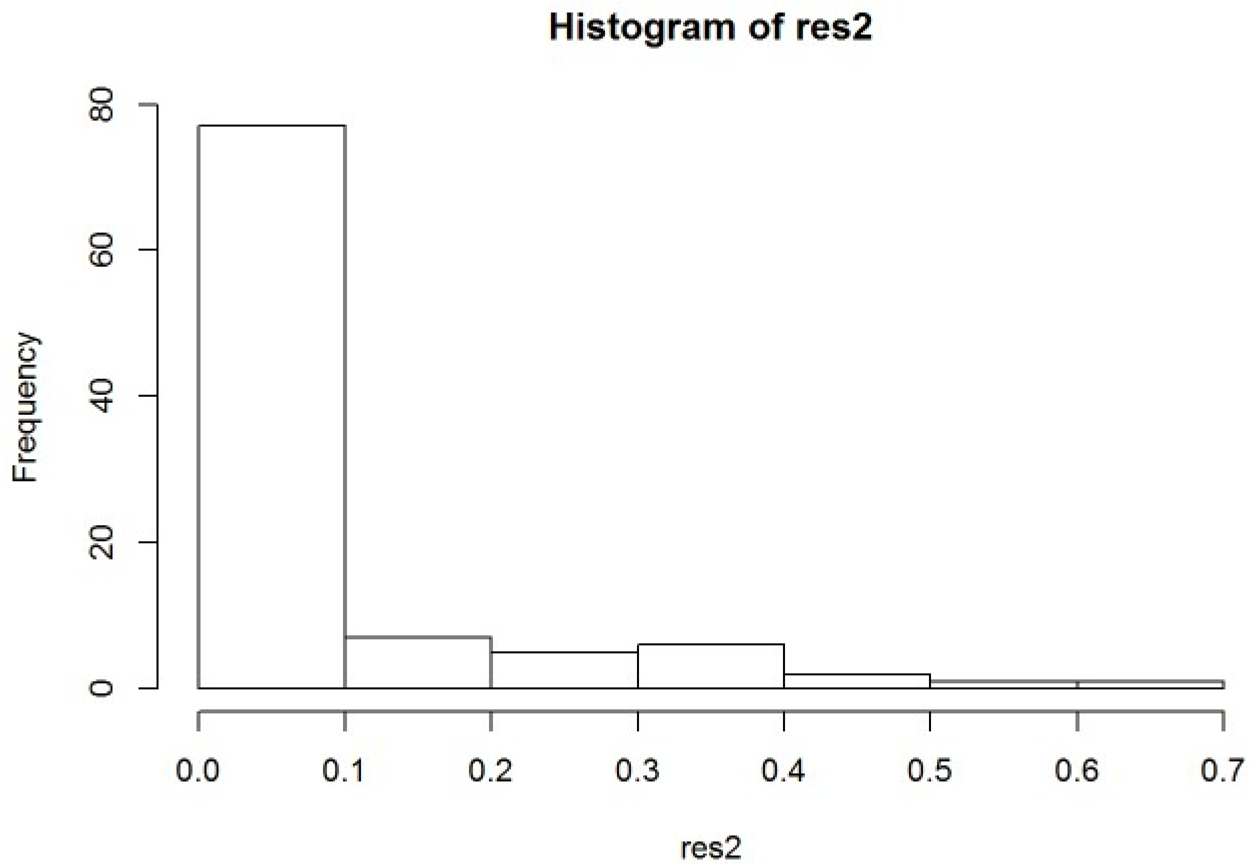
```
res2 <- residuals * residuals
res2
```

```
## Time Series:
## Start = 1
## End = 99
## Frequency = 1
##  [1] 8.583820e-05 2.477600e-03 9.085486e-04 2.461980e-03 6.438119e-04
##  [6] 7.020378e-04 2.356363e-02 3.824139e-03 2.276449e-03 1.729905e-03
## [11] 6.235583e-07 6.364143e-03 4.795020e-03 5.185087e-03 4.824341e-04
## [16] 4.902104e-03 1.068799e-02 4.200525e-03 4.567542e-03 4.830843e-03
## [21] 5.494301e-03 1.476599e-03 1.539926e-02 6.980158e-03 1.104134e-02
## [26] 7.231993e-04 9.395058e-03 6.606405e-03 4.417443e-03 2.086498e-03
## [31] 3.129719e-03 4.259663e-03 3.751851e-05 8.240381e-03 1.336894e-02
```

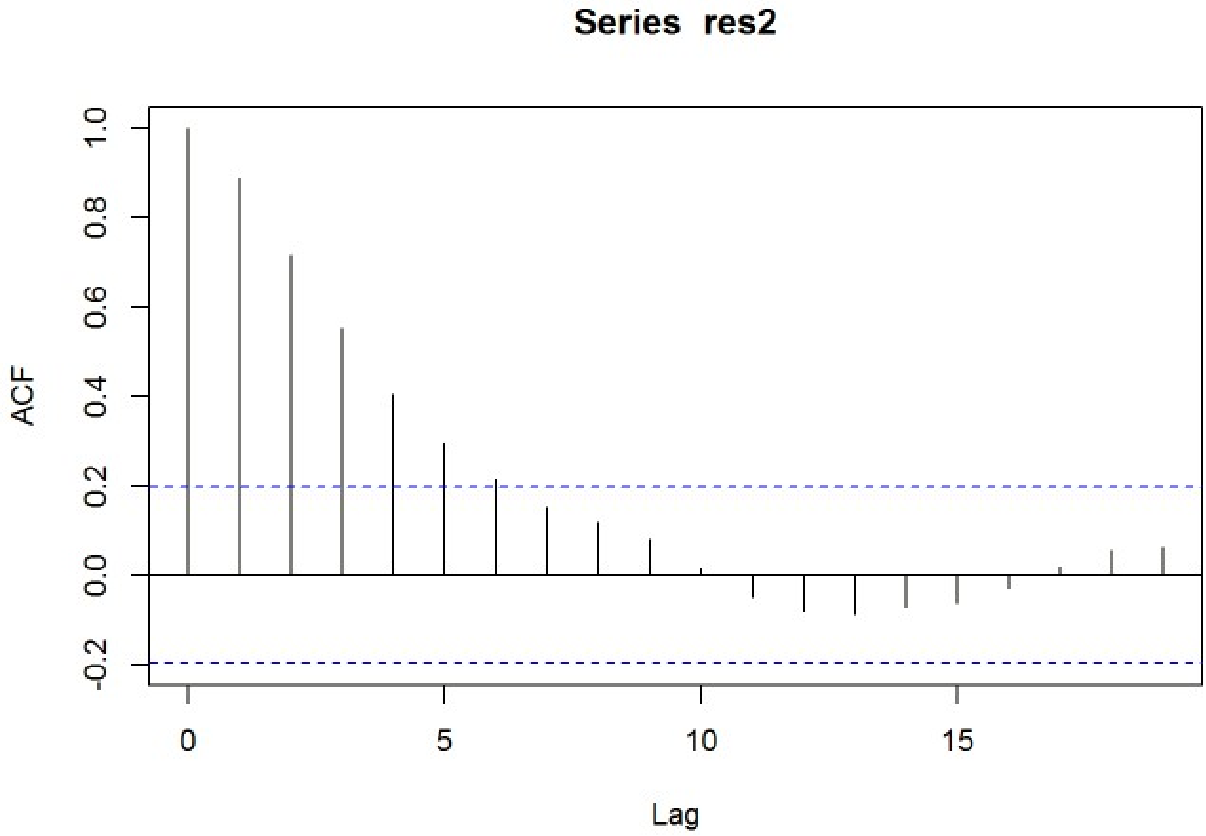
```
## [36] 4.521305e-03 3.267739e-04 1.187911e-03 1.177477e-04 2.936185e-02
## [41] 4.545340e-03 2.764461e-04 6.463934e-03 7.995297e-07 2.543972e-03
## [46] 3.791555e-03 4.597619e-03 1.059695e-02 1.226916e-03 3.209213e-05
## [51] 6.796622e-03 2.629201e-02 1.517718e-02 5.187508e-02 5.837991e-05
## [56] 1.892475e-03 5.963204e-03 3.241899e-02 1.013820e-02 2.894020e-02
## [61] 9.522436e-02 2.272830e-01 3.882814e-01 4.444584e-01 4.467063e-01
## [66] 3.748299e-01 3.333011e-01 2.074016e-01 1.514200e-01 1.966555e-01
## [71] 2.442313e-01 1.458375e-01 1.022425e-01 9.243918e-02 6.487319e-02
## [76] 3.290589e-02 1.333262e-02 2.482012e-06 1.179940e-02 5.388849e-02
## [81] 9.104516e-02 9.337598e-04 1.281624e-03 1.260782e-02 9.151838e-03
## [86] 3.028770e-04 8.610574e-02 3.628410e-01 6.211717e-01 5.507422e-01
## [91] 3.534363e-01 3.208558e-01 2.761170e-01 1.378120e-01 2.650233e-01
## [96] 7.809906e-02 1.136019e-01 1.779162e-01 9.710894e-02
```

1.6) Обчисліть автокореляційну функцію для ряду із квадратів залишків.

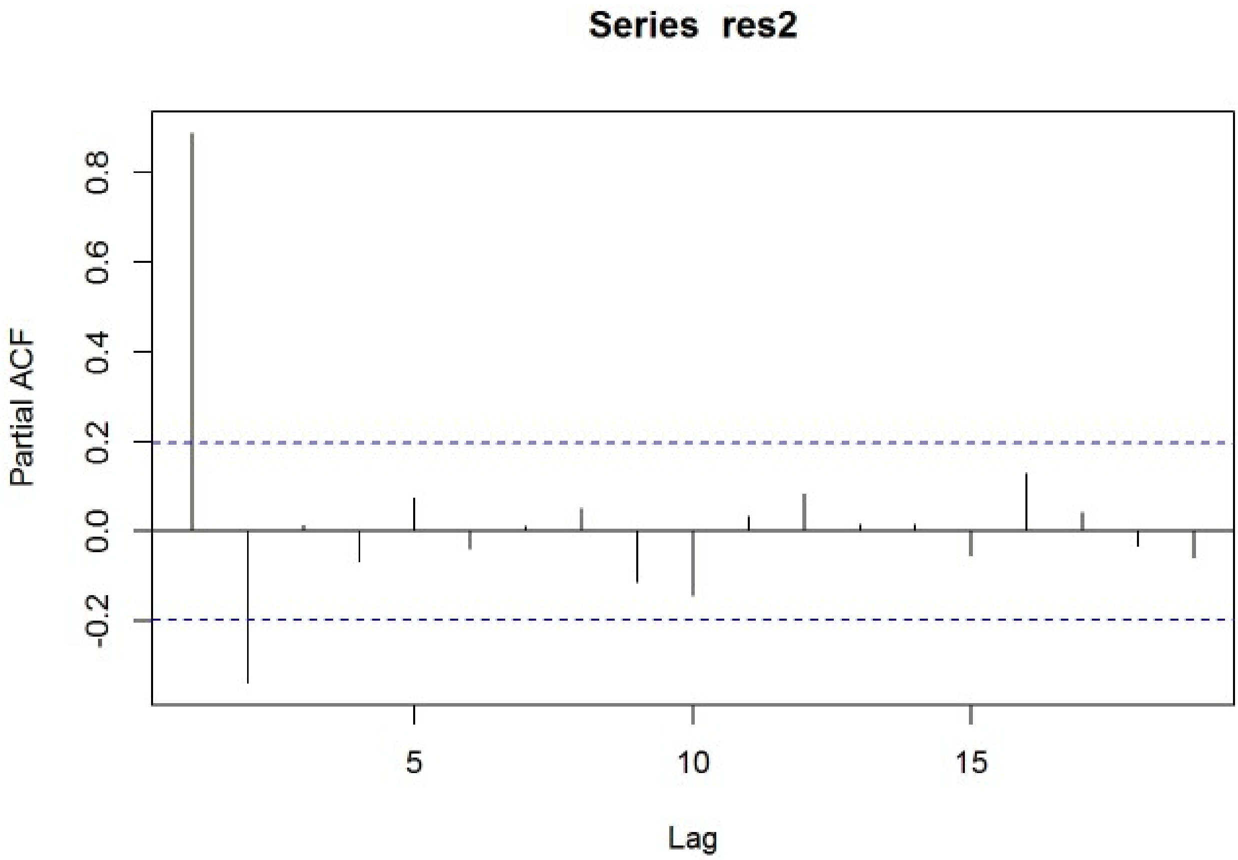
```
hist(res2)
```



```
acf(res2, na.action = na.pass)
```




```
pacf(res2, na.action = na.pass)
```



1.7) За допомогою МНК обчисліть оцінки коефіцієнтів рівняння першого порядку для дисперсії залишків:

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1 \hat{\varepsilon}^2(k-1) + v(k),$$

де $v(k)$ – похибки моделі. Таку модель називають авторегресійною умовно гетероскедастичною або скорочено АРУГ (в даному випадку АРУГ(1)).

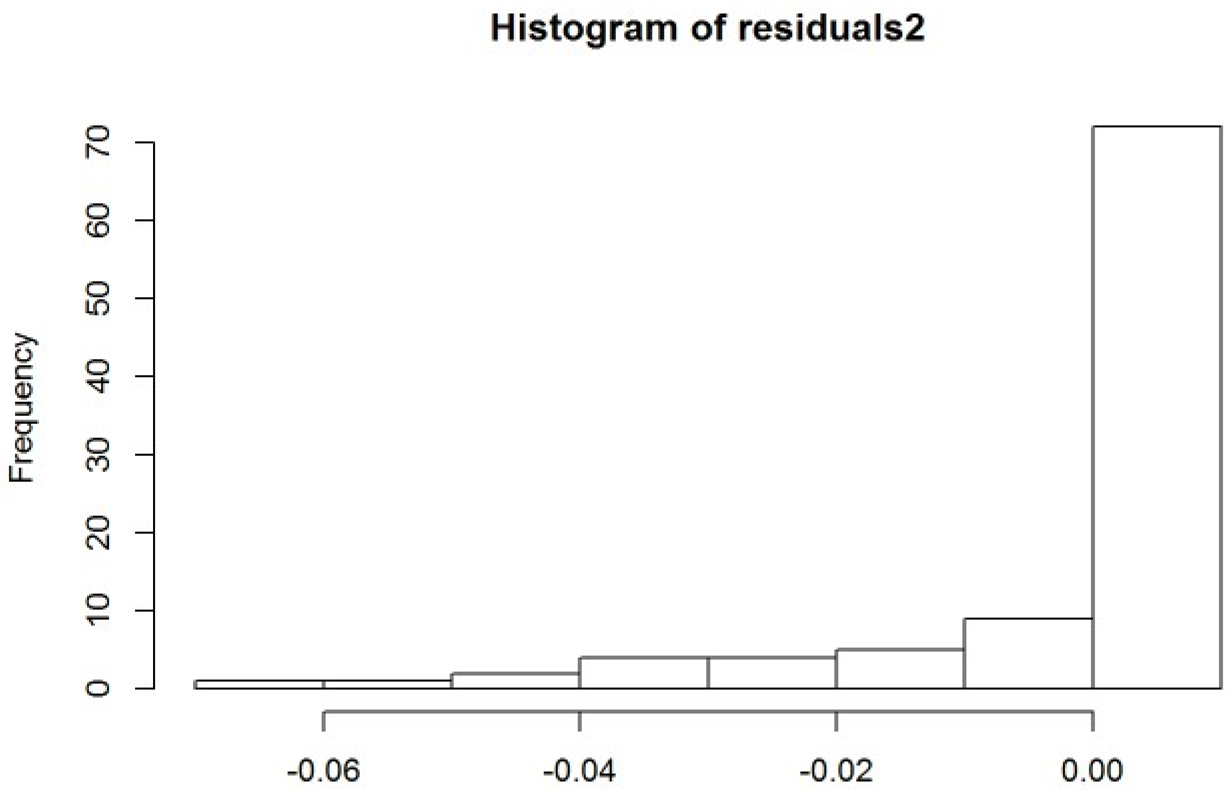
```
model2 <-  dynlm(res2 ~ L(res2,1))
summary(model2)
```

```
## Time series regression with "ts" data:
## Start = 2, End = 99
##
## Call:
## dynlm(formula = res2 ~ L(res2, 1))
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.166741 -0.016057 -0.008799 -0.000991  0.289510
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.009607   0.007246   1.326    0.188
## L(res2, 1)  0.887591   0.046668  19.019   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.06238 on 96 degrees of freedom
## Multiple R-squared:  0.7903, Adjusted R-squared:  0.7881
## F-statistic: 361.7 on 1 and 96 DF,  p-value: < 2.2e-16
```

1.8) Обчисліть коефіцієнти рівняння АРУГ(4):

$$\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1\hat{\varepsilon}^2(k-1) + \alpha_2\hat{\varepsilon}^2(k-2) + \alpha_3\hat{\varepsilon}^2(k-3) + \alpha_4\hat{\varepsilon}^2(k-4).$$

```
residuals2 <-  lag(fitted.values(model2),1) - res2
hist(residuals2)
```




```
model3 <-  dynlm(res2 ~ L(res2,1) + L(res2,2) + L(res2,3) + L(res2,4))
summary(model3)
## Time series regression with "ts" data:
## Start = 5, End = 99
##
## Call:
## dynlm(formula = res2 ~ L(res2, 1) + L(res2, 2) + L(res2, 3) +
##     L(res2, 4))
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.205131 -0.017207 -0.010517  0.009777  0.246700
##
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.012840   0.007371   1.742   0.0850 .
## L(res2, 1)   1.200128   0.106447  11.274   <2e-16 ***
## L(res2, 2)  -0.393228   0.167796  -2.343   0.0213 *
## L(res2, 3)   0.102536   0.168016   0.610   0.5432
## L(res2, 4)  -0.067918   0.106934  -0.635   0.5269
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0604 on 90 degrees of freedom
## Multiple R-squared:  0.8138, Adjusted R-squared:  0.8055
## F-statistic: 98.33 on 4 and 90 DF,  p-value: < 2.2e-16
```

1.9) Побудуйте ряд умовних дисперсій для *ARCH_01.DAT* і побудуйте узагальнену авторегресійну умовно гетероскедастичну (УАРУГ) модель процесу у вигляді:

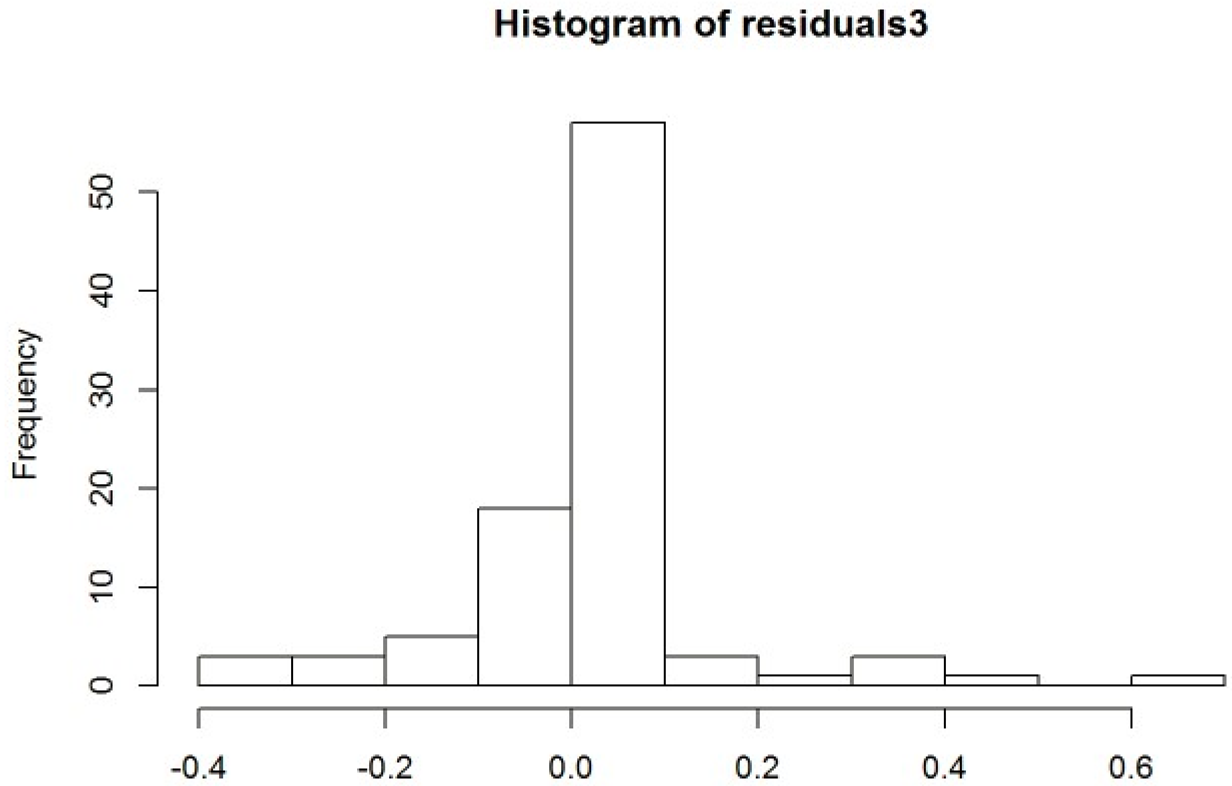
$$h(k) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \hat{\varepsilon}^2(k-i) + \sum_{i=1}^{p} \beta_i h(k-i) + v(k),$$

де q – визначається за допомогою АКФ процесу $\{\varepsilon^2(k)\}$; p – визначається за допомогою АКФ для $\{h(k)\}$; $h(k)$ – умовна дисперсія процесу, яка розраховується послідовно для досліджуваного ряду за виразом:

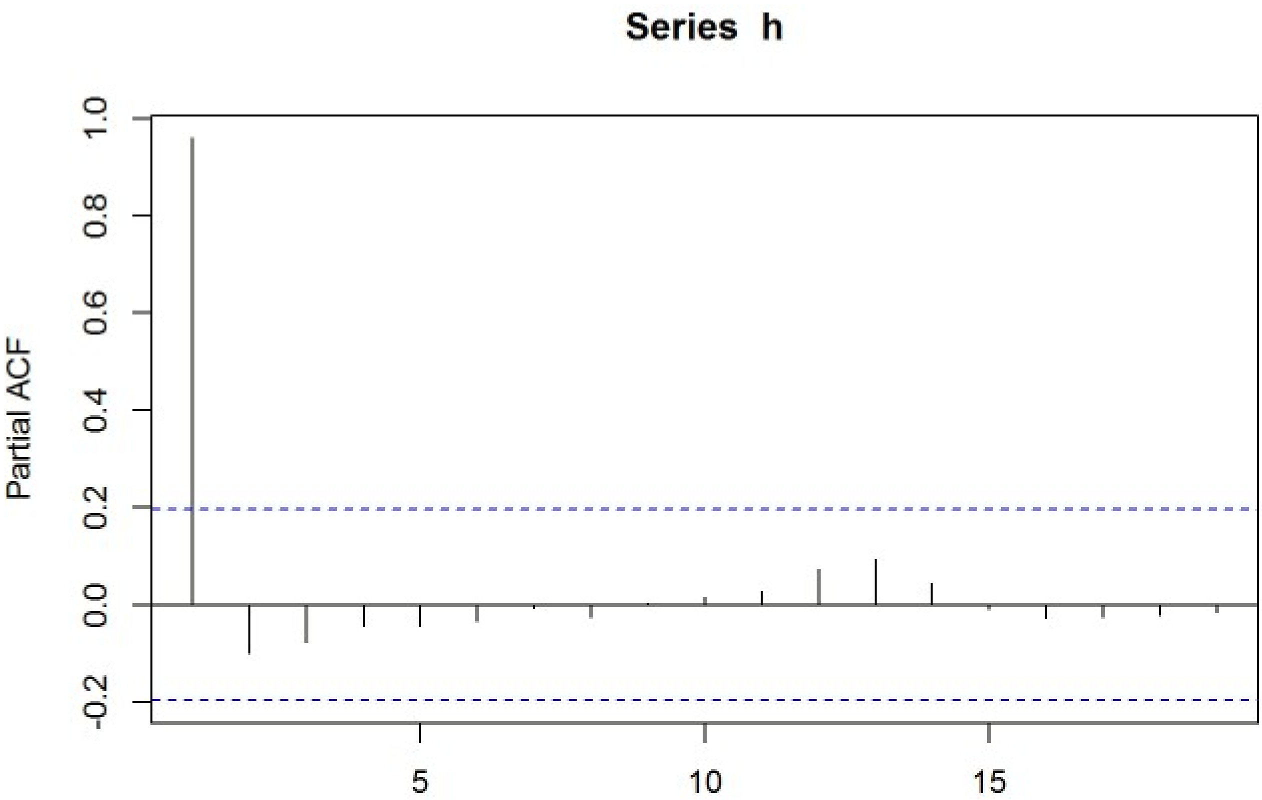
$$h(k) = E_k[y^2(k)] = \frac{1}{k-1}\sum_{i=1}^{k}[y(i) - \bar{y}_k]^2, \ \ k = 2,3,...,N,$$

де N – довжина ряду; $\bar{y}_k = \frac{1}{k-1}\sum_{i=1}^{k} y(i), \ k = 2,...,N$ – умовне математичне сподівання.

```
residuals3 <-  lag(fitted.values(model3),4) - res2
hist(residuals3)
```




```
h <- c()
for (i in 2:length(y)){
  h[i-1] <-  sd(y[1:i])^2
}
h <- ts(h)
pacf(h)
```



```
model4 <-  dynlm(h ~ res2 + L(res2,1) + L(h,1))
summary(model4)

## Time series regression with "ts" data:
## Start = 2, End = 99
##
## Call:
## dynlm(formula = h ~ res2 + L(res2, 1) + L(h, 1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.91320 -0.10847 -0.03230  0.05427  1.35619
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.004379   0.040808   0.107    0.915
## res2         5.313322   0.510347  10.411  < 2e-16 ***
## L(res2, 1)  -2.209045   0.541218  -4.082 9.38e-05 ***
## L(h, 1)      0.998401   0.005927 168.450  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3119 on 94 degrees of freedom
## Multiple R-squared:  0.9981, Adjusted R-squared:  0.9981
## F-statistic: 1.682e+04 on 3 and 94 DF,  p-value: < 2.2e-16
```

1.10) Побудуйте модель гетероскедастичного процесу

$$h(k) = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon^2 (k-i) + \sum_{i=1}^{p} \beta_i h(k-i),$$

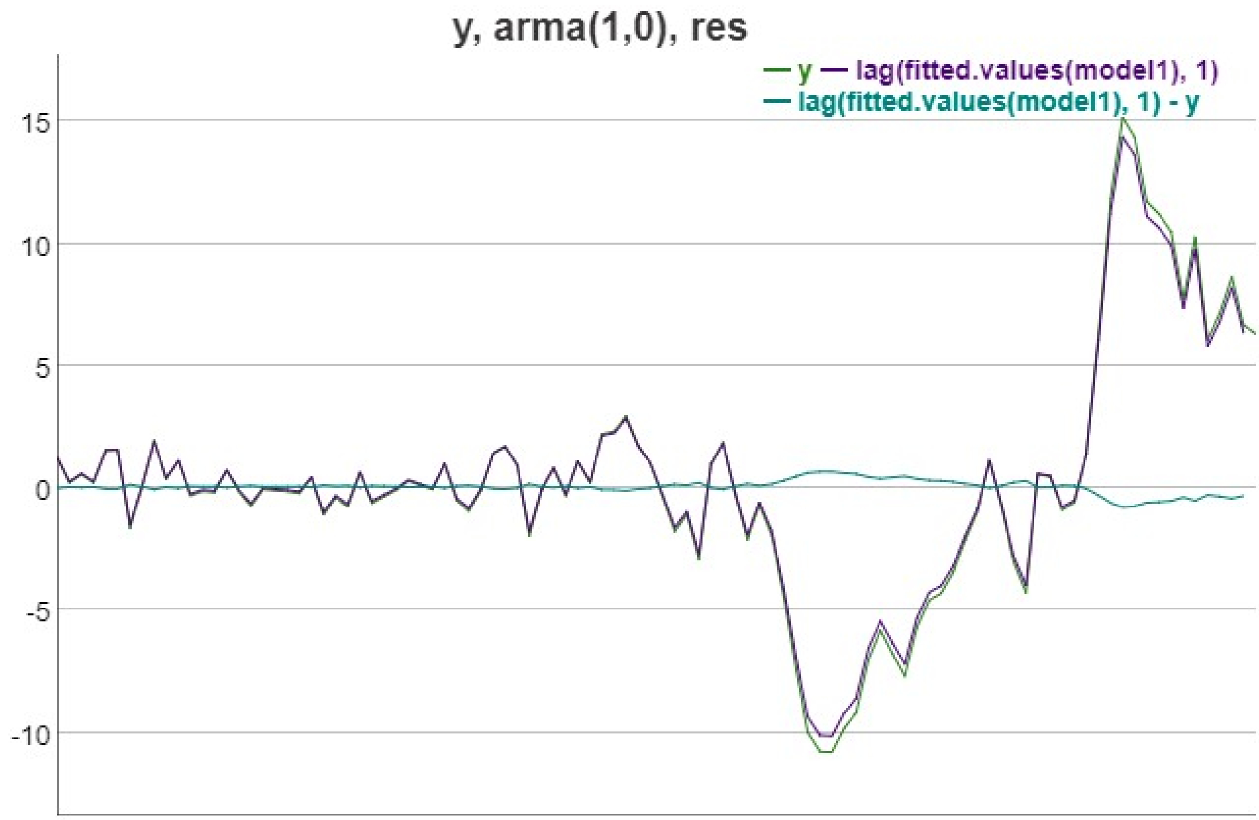
де

$$h(k) = E_k[\varepsilon^2(k)] = \frac{1}{k-1} \sum_{i=1}^{k} [\varepsilon(i) - \overline{\varepsilon}_k]^2, \; k = 2,3,...,N,$$

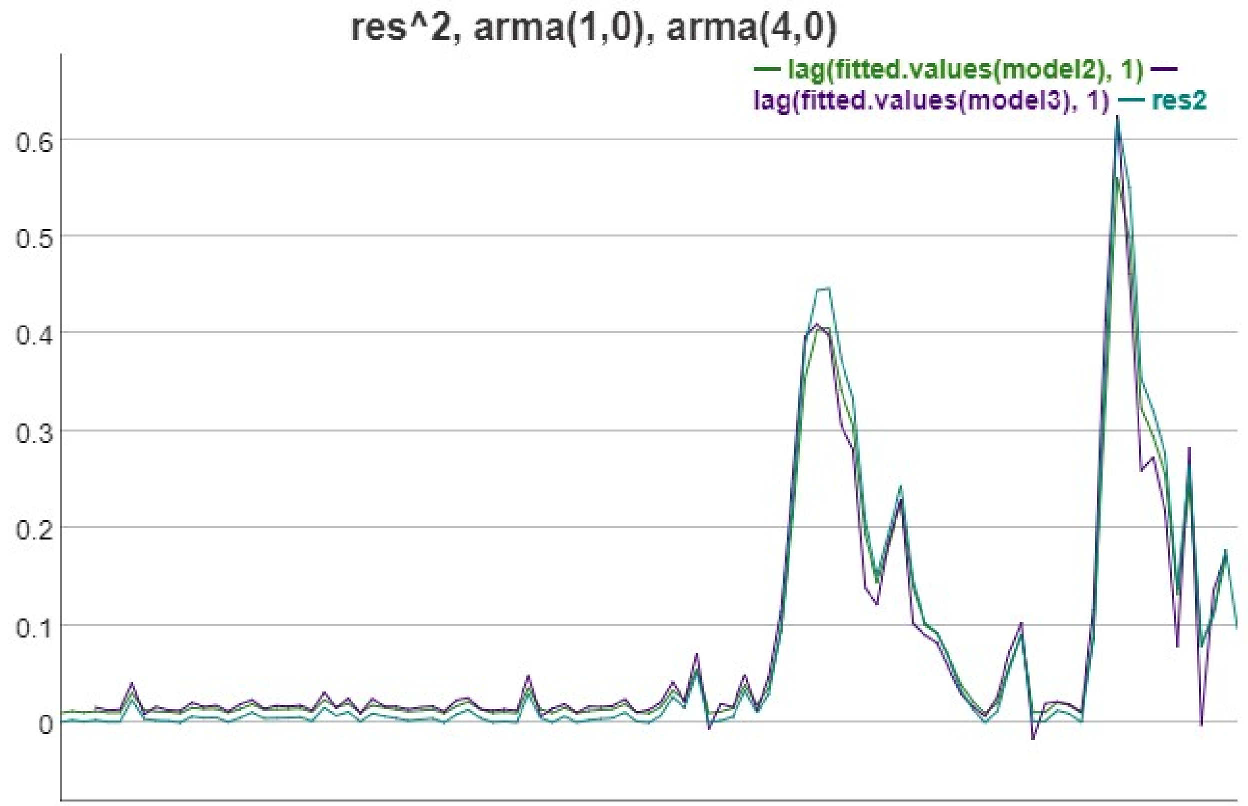
$$\overline{\varepsilon}_k = \frac{1}{k-1} \sum_{i=1}^{k} \varepsilon(i), \; k = 2,...,N.$$

Наведіть графіки процесів: $y(k)$, $h(k)$, $\varepsilon(k)$, $\varepsilon^2(k)$ та автокореляційні функції, використані при побудові моделей.

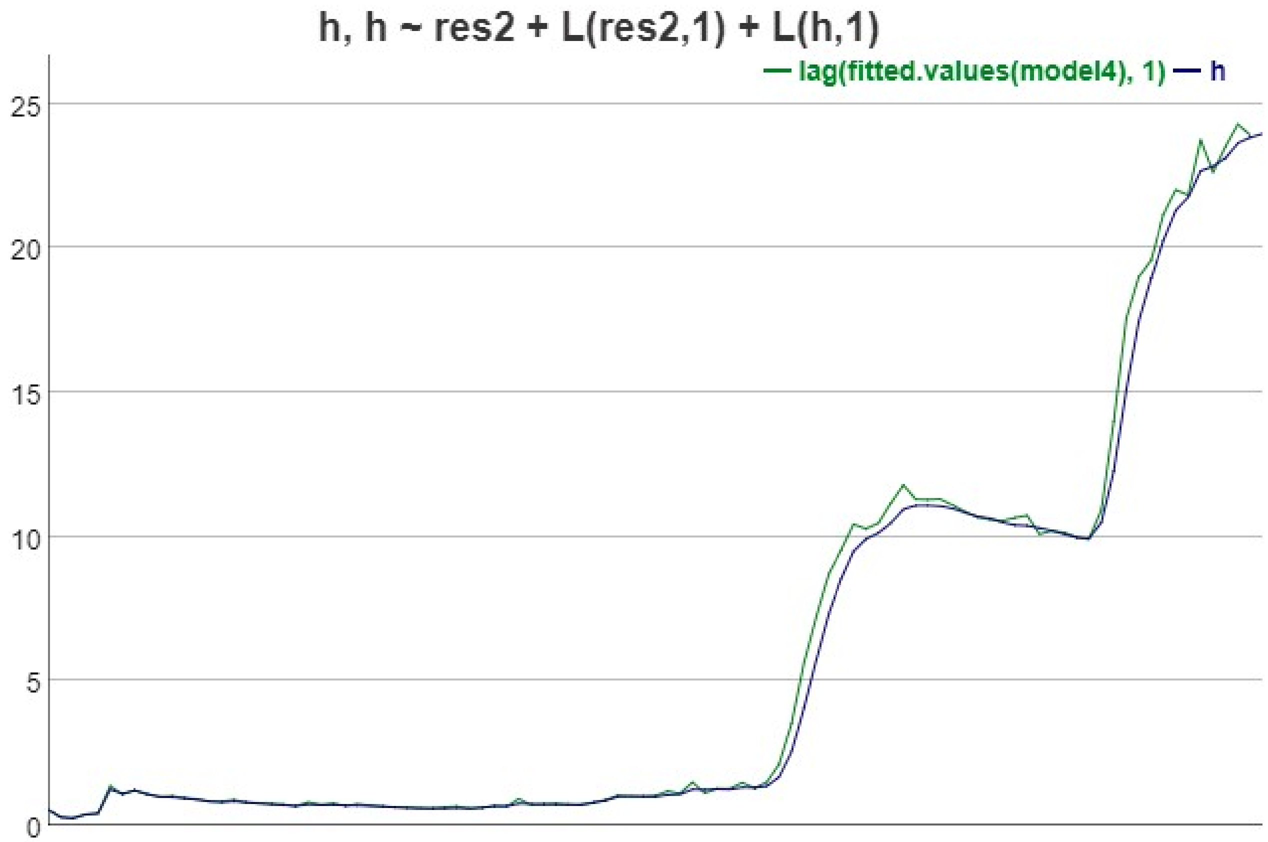
```
dygraph(cbind(y, lag(fitted.values(model1),1), lag(fitted.values(model1),1)-y) , main = "y, arma(1,0), res") %>% dyRangeSelector
```




```
dygraph(cbind(lag(fitted.values(model2),1),lag(fitted.values(model3),1),res2) , main = 'res^2, arma(1,0), arma(4,0)') %>% dyRangeSelector()
```




```
dygraph(cbind(lag(fitted.values(model4),1),h) , main = 'h, h ~ res2 + L(res2,1) + L(h,1)') %>% dyRangeSelector()
```



1.11) Дайте відповіді на контрольні запитання, які даються в кінці опису лабораторної роботи.

1. Поясніть стандартні статистичні параметри, що характеризують оцінку авторегресії.

Автокореляційна функція дозволяє встановити наявність залежності ряду від самого себе. Кореляція може бути лінійною або нелінійною в залежності від типу залежності, яка фактично існує між змінними.

Часткова автокореляційна функція дозволяє уточнити порядок авторегресійної складової. ЧАКФ чіткіше відображає порядок АР-моделі завдяки відсутності впливу проміжних коефіцієнтів кореляції на вибрані значення змінної.

Статистика Л'юнга-Бокса, визначається за формулою:

$$Q(r_k) = N(N+2)\sum_{k=1}^{s} r_k^2 /(N-k),$$

використовується для тестування відмінності від нуля групи авторегресивних коефіцієнтів часового ряду.

2. Як може вплинути позитивна кореляція величин $\{v(k)\}$ моделі

$$y(k) = a_0 + a_1 y(k-1) + v(k)$$

на оцінку коефіцієнта a_1 при короткій вибірці даних?

Його вага зменьшиться, так як адекватна модель скоріше усього матиме ненульву частину МА

3. Поясніть відмінність між безумовними та умовними статистичними параметрами: *математичне сподівання та дисперсія*.

Нехай є модель AR(n): $y(k) = a_0 + a_1 y(k-1) + \dots + a_n y(k-n) + \varepsilon(k)$

Умовне математичне сподівання y(k) відносно інформації, що знаходиться в минулих значеннях ряду, є випадковою величиною і дорівнює:

$$E(y_k / y_{k-1}, y_{k-2} \dots , y_{k-n}) = a_0 + a_1 y(k-1) + \dots + a_n y(k-n)$$

Звичайне **математичне сподівання** ряду не змінюється з часом і дорівнює:

$$E(y_k) = \frac{a_0}{1 - a_1 - a_2 - \dots - a_n}$$

Аналогічно, **умовна дисперсія** в моделі ARCH(q) є випадковою велечиною:

$$\varepsilon_t = \sqrt{h_t} v_t,\ v_t \sim i.i.d.N(0,1) \qquad E(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \sqrt{h_t} E(v_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = 0,$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_q \varepsilon_{t-q}^2 \text{ - } ARCH(q) \qquad D(\varepsilon_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = h_t D(v_t \mid \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = h_t;$$

При тому, що **звичайна дисперсія** є сталою величиною:

$$D(\varepsilon_t) = \sigma_\varepsilon^2$$

4. Використовуючи АКФ та ЧАКФ для залишків та квадратів залишків зробіть висновок про наявність процесу АРУГ в процесі, що розглядається (п. 1.6)?

АКФ та ЧАКФ для квадратів залишків допомагає оцінити кількість параметрів в моделі АРУГ/УАРУГ

5. Порівняйте моделі першого та четвертого порядків, знайдені в пп. 1.7 і 1.10.

Модель АРУГ(4) адекватніша, ніж модель АРУГ(1), але дивно, що в АРУГ(4) є від’ємні коефіцієнти.

6. Поясність на прикладі реального процесу природу гетероскедастичності (тобто звідки вона береться)?

   Heteroscedasticity often occurs when there is a large difference among the sizes of the observations.
   A classic example of heteroscedasticity is that of income versus expenditure on meals. As one's income increases, the variability of food consumption will increase. A poorer person will spend a rather constant amount by always eating inexpensive food; a wealthier person may occasionally buy inexpensive food and at other times eat expensive meals. Those with higher incomes display a greater variability of food consumption.
   Imagine you are watching a rocket take off nearby and measuring the distance it has traveled once each second. In the first couple of seconds your measurements may be accurate to the nearest centimeter, say. However, 5 minutes later as the rocket recedes into space, the accuracy of your measurements may only be good to 100 m, because of the increased distance, atmospheric distortion and a variety of other factors. The data you collect would exhibit heteroscedasticity.

7. Поясніть методику визначення гетероскедастичності.

   У першому наближенні наявність гетероскедастичності можна помітити на графіках залишків регресії (або їх квадратів) по деяким змінним, за оціненою залежною змінною або за номером спостереження. На цих графіках розкид точок може змінюватися в залежності від значення цих змінних.
   Для більш суворої перевірки застосовують, наприклад, статистичні тести Уайта, Голдфелда - Куандт, Бройша - Пагана, Парка, Глейзера, Спірмена.

8. В чому полягає різниця між моделями АРУГ та УАРУГ?

   АРУГ складається тільки з авторегресії:
   $\hat{\varepsilon}^2(k) = \alpha_0 + \alpha_1\hat{\varepsilon}^2(k-1) + \alpha_2\hat{\varepsilon}^2(k-2) + \ldots + \alpha_q\hat{\varepsilon}^2(k-q) + v(k),$
   Узагальнена модель АРУГ, яку називають УАРУГ(p,q), складається із двох компонент – авторегресії та ковзного середнього відносно дисперсії гетероскедастичного процесу.
   $$E_{k-1}[\varepsilon^2(k)] = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon^2(k-i) + \sum_{i=1}^{p}\beta_i\, h(k-i).$$

9. Наведіть приклад процесу (ряду), де може бути присутня гетероскедастичність?

Див. питання #6.

10. Наведіть альтернативний вираз для обчислення умовної дисперсії.

$$\begin{aligned}\mathrm{Var}(X|Y=y) &= E\left[(X-\mu_{X|Y}(y))^2 | Y=y\right] \\ &= \sum_{x_i \in R_X} \left(x_i - \mu_{X|Y}(y)\right)^2 P_{X|Y}(x_i) \\ &= E\left[X^2 | Y=y\right] - \mu_{X|Y}(y)^2.\end{aligned}$$

Висновки:

Під час виконання даної лабораторної роботи ми навчилися моделювати та аналізувати гетероскедастичні процеси. Навчилися будучати модель авторегресійу умовно гетероскедастичну АГУГ(1) та АГУГ(4), а також узагальнену авторегресійну умовно гетероскедастичну (УАРУГ).