

## Практическая работа №1 по эконометрике-2. Петраков Сергей э-301.

### Задание 1.

По заданию мне предстоит работать с временными рядами Y5 Y18 Y51.

(а) Подберём ARIMA модель подробным образом для ряда Y5.

Будем действовать в соответствии с методологией Бокса-Дженкинса. Она подразумевает под собой 4 последовательных шага, на первом из которых мы тестируем ряд на стационарность (смотрим график и проводим формальный тест на стационарность – тест Дики-Фуллера). Необходимо найти стационарный временной ряд, если первоначальный ряд нестационарен, то переходим к его первым разностям и снова проводим тест, если и они не стационарны, то берём вторые разности. В итоге мы должны получить стационарный ряд и именно с ним продолжить работу.

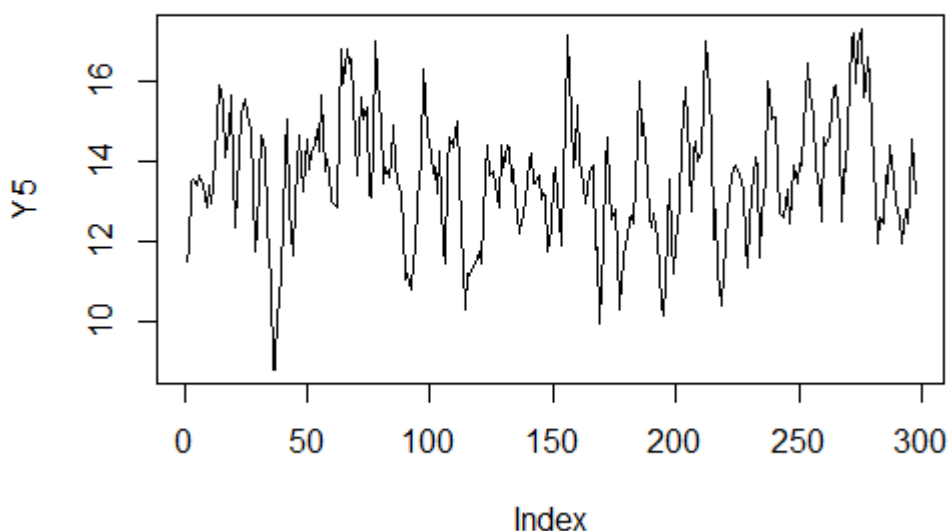
Второй шаг подразумевает анализ автокорреляционной функции и частной автокорреляционной функции для лучшего подбора модели.

На третьем шаге мы оцениваем предполагаемую модель и проверяем её на адекватность, используя информационные критерии и анализируя поведение остатков (информационные критерии чем меньше, тем лучше, а поведение остатков хорошей модели должно быть похожим на белый шум.)

На четвёртом шаге осуществляется прогноз по подобранной модели.

Проведём вышеописанные шаги с временным рядом Y5 в R.

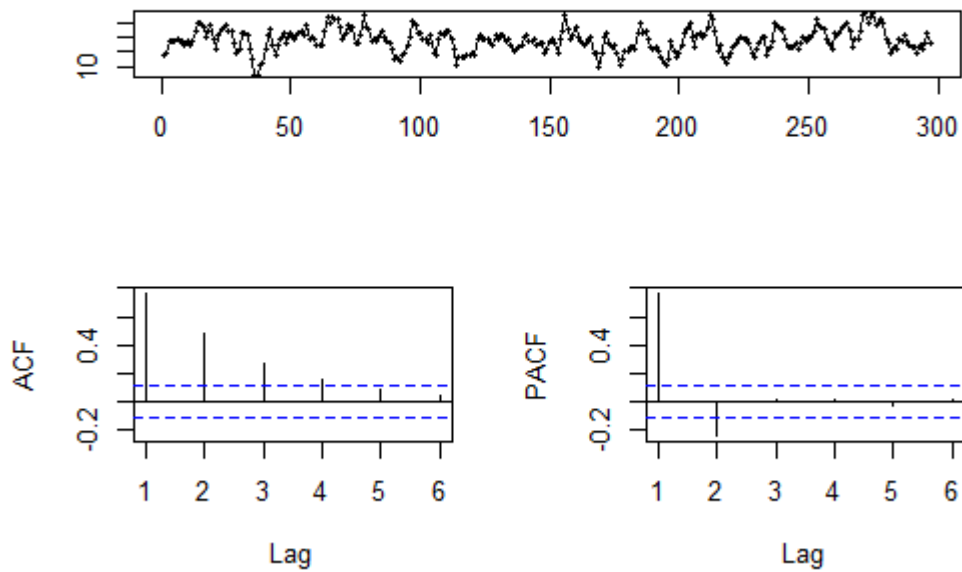
Но для начала, сделаем нулевой шаг и посмотрим, как выглядят наши данные.



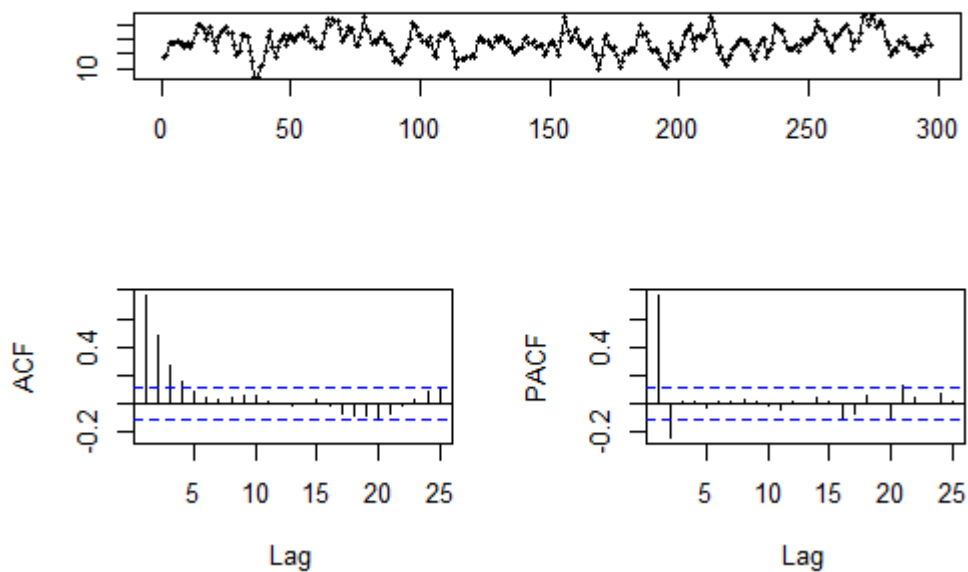
Поведение ряда похоже на стационарное, также можно заметить, что график колеблется примерно у 13,6 в среднем, поэтому для формального теста на стационарность надо использовать модель с дрифтом и, как видно, без тренда.

Графики автокорреляционных функций и частных автокорреляционных функций приведены ниже.

**Y5**



**Y5**



(По заданию требовалось ACF и PACF только для первых шести лагов, но для того, чтобы посмотреть как себя ведут последующие значения функций (чтобы убедиться, что они действительно затухают) построил для 25 лагов.)

Значимы оказались для ACF 1, 2, 3, 4 лаги, а для PACF только 1 и 2.

Анализируя, ACF и PACF можно предположить, что перед нами процесс авторегрессии второго порядка, так как этот процесс характеризуется постепенным убыванием значений функции ACF, и резким уменьшением значений PACF. Второй порядок в силу двух значимых лагов для PACF.

Проведём формальный тест Дики-Фуллера на стационарность. В сессии R я также провёл тест KPSS, который подтвердил результаты.

Если провести тест с дрифтом, то мы действительно получим в результате, что гипотезу о стационарности временного ряда можно принять при любом разумном уровне значимости.

Результат из R.

```
#####  
# Augmented Dickey-Fuller Test Unit Root Test #  
#####
```

Test regression drift

Call:

```
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.56181	-0.64297	0.03117	0.61919	2.57671

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	4.05027	0.53312	7.597	4.13e-13	***
z.lag.1	-0.29742	0.03896	-7.634	3.25e-13	***
z.diff.lag	0.24722	0.05668	4.362	1.79e-05	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9799 on 292 degrees of freedom

Multiple R-squared: 0.1747, Adjusted R-squared: 0.169

F-statistic: 30.9 on 2 and 292 DF, **p-value: 6.699e-13**

Value of test-statistic is: -7.6342 29.1424

Critical values for test statistics:

	1pct	5pct	10pct
tau2	-3.44	-2.87	-2.57
phi1	6.47	4.61	3.79

Больше всего тут нас интересует P-значение и оно очень мало.

В силу проведённого анализа мы проделали первые два шага по методологии Бокса-Дженкинса. Порядок здесь оправдан, поскольку это позволяет нам быстрее подобрать модель.

Попробуем модель ARIMA(2, 0, 0).

Series: Y5

ARIMA(2,0,0) with non-zero mean

**Coefficients:**

	ar1	ar2	mean
	0.9518	-0.2471	13.5844
s.e.	0.0562	0.0564	0.1906

**Информационные критерии**

sigma^2 estimated as 0.9597: log likelihood=-414.3

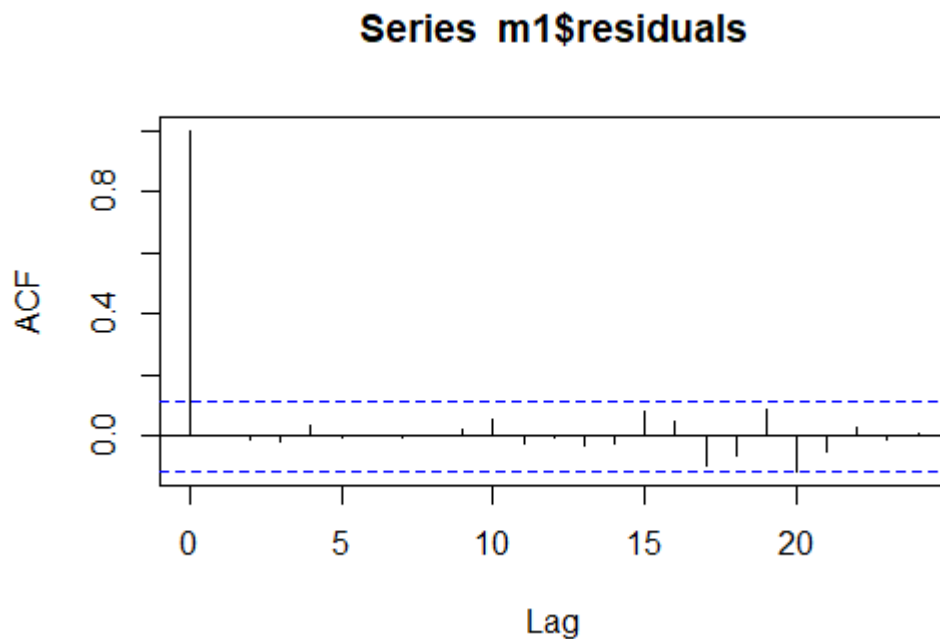
AIC=836.61 AICC=836.74 BIC=851.38

Training set error measures:

	ME	RMSE	MAE	MPE	MAPE	MASE
ACF1						

Training set 0.004843735 0.9746665 0.7791033 -0.5029369 5.823212 0.9158057 0.002405703

Проанализируем остатки модели:



Только нулевой лаг значим и равен 1, что, по определению, верно, а остальные лаги незначимы. График напоминает белый шум. Проведём формальный тест Льюнга-Бокса для анализа поведения остатков:

В качестве результатов получаем

Box-Ljung test

data: resid\_m1  
X-squared = 1.6365, df = 8, p-value = 0.9902

Это значит, что остатки действительно ведут себя как белый шум.

Для сравнения можно посмотреть на модель ARIMA(1,0,0), и обратить внимание на их информационные критерии.

Они получились равными AIC=853.23 AICc=853.32 BIC=864.32

Это больше, чем для модели ARIMA(2,0,0) и даже критерий Шварца (BIC), который предпочитает более простую модель оказался больше.

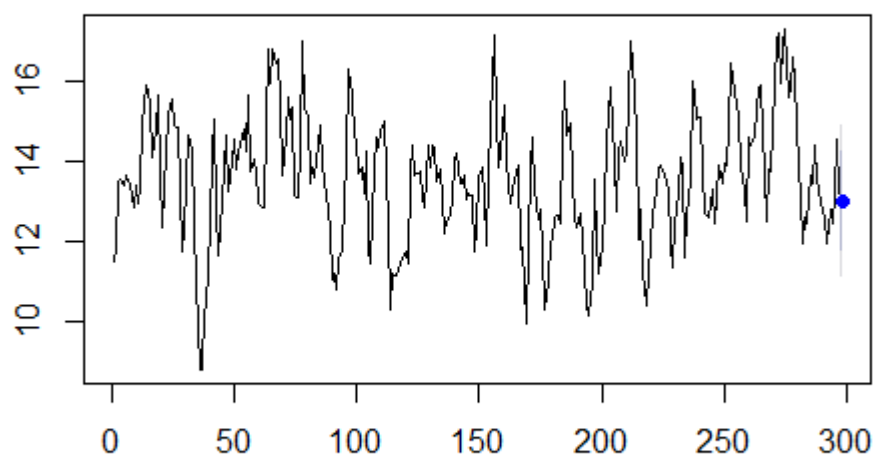
Победная модель:

$$Y_t = 13.58 + 0.95Y_{t-1} - 0.25Y_{t-2} \\ (0.1906)(0.0562) \quad (0.0564)$$

Все коэффициенты значимы

(б) Прогноз по ARIMA(2,0,0)

## Forecasts from ARIMA(2,0,0) with non-zero mean



Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
298	12.99646	11.74102	14.25191	11.07643 14.9165

Таким образом 95% доверительный интервал:

(11.07643, 14.9165)

(в) Остальные две модели оценим с помощью автоподбора модели (auto arima).

Оказалось, что для ряда Y18 наилучшей моделью является ARIMA(1,0,0) (со средним ноль)

$$Y_t = 0,27Y_{t-1}$$

(0,056)

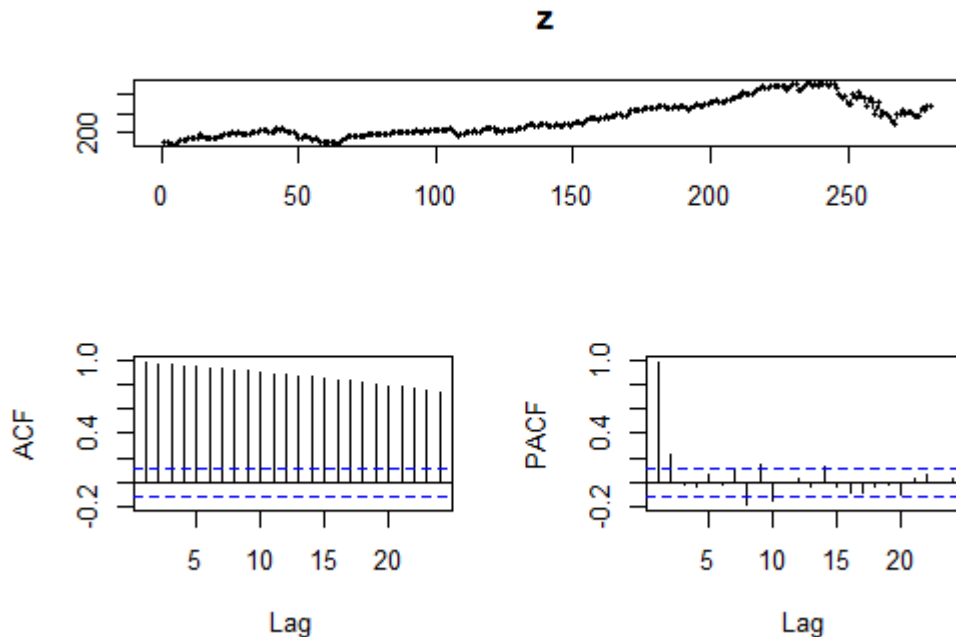
, а для ряда Y51 наилучшей моделью является ARIMA(1,1,0) (с дрифтом)

$$\Delta Y_t = 0,71Y_{t-1} + 3,26$$

(0,0406) (0,0986)

## Задание 2.

Проанализируем AAPL по ARIMA модели



Ряд нестационарный, как можно заметить на графике, поэтому необходимо сделать переход к стационарному ряду. Подбирая модель автоматически получаем, что наилучшей моделью является модель ARIMA(1,1,0).

Провалидируем результаты на действительности. Мы с помощью функции прогноза (forecast) получили:

Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
281	-0.61329489	-8.142488	6.915898	-12.12820	10.90161
282	0.19590310	-7.708077	8.099884	-11.89219	12.28400
283	-0.06257679	-8.003804	7.878650	-12.20764	12.08248

Заметим, что в данном случае мы работаем с первыми разностями, поэтому совершать проверку будем также в разностях.

В действительности мы в эти дни имели следующие результаты

В действительности оказалось так

```
AAPL.Close
2020-04-07 -3.040008
2020-04-08  6.640014
2020-04-09  1.919983
```

То есть модель не очень хорошо описывает происходящее сейчас на фондовом рынке, что вполне закономерно, учитывая, текущую волатильность. Как можно заметить, наша модель прогнозирует на такую сильную волатильность, какая она в действительности есть. Проблемность также в том, что не совпадают знаки предсказания и реального изменения курса акции Apple.