

**Эконометрика,
Домашнее задание 3**

**Выполнил Кондаков Артём,
студент группы БЭК222**

Задание 1

а)

$$\gamma(0) = \text{Var}(y_t) = \text{Var}(2 + \varepsilon_t + 0.1\varepsilon_{t-1} - 0.06\varepsilon_{t-2}) = 16 + 0.01 * 16 + 0,0036 * 16 = 16.2176$$

$$\begin{aligned}\gamma(1) &= \text{cov}(y_t, y_{t-1}) = \text{cov}(2 + \varepsilon_t + 0.1\varepsilon_{t-1} - 0.06\varepsilon_{t-2}, 2 + \varepsilon_{t-1} + 0.1\varepsilon_{t-2} - 0.06\varepsilon_{t-3}) \\ &= \text{cov}(0.1\varepsilon_{t-1}, \varepsilon_{t-1}) + \text{cov}(-0.06\varepsilon_{t-2}, 0.1\varepsilon_{t-2}) = 0.1 * 16 - 0.06 * 0.1 * 16 = 1.504\end{aligned}$$

$$\gamma(2) = \text{cov}(y_t, y_{t-2}) = \dots = \text{cov}(-0.06\varepsilon_{t-2}, \varepsilon_{t-2}) = -0.06 * 16 = -0.96$$

$$\gamma(k) = \text{cov}(y_t, y_{t-k}) = 0, k \geq 3$$

ACF:

$$\rho(1) = \frac{\gamma(1)}{\gamma(0)} = 0.0927$$

$$\rho(2) = \frac{\gamma(2)}{\gamma(0)} = -0.0592$$

$$\rho(k) = 0, k \geq 3$$

б)

$$\varphi_{11} = \rho(1) = 0.0927$$

Для φ_{22} решим систему:

$$\begin{cases} \gamma(1) = \varphi_{21}\gamma(0) + \varphi_{22}\gamma(1) \\ \gamma(2) = \varphi_{21}\gamma(1) + \varphi_{22}\gamma(0) \end{cases}$$

Решая, получим:

$$\varphi_{22} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} = -0.0684$$

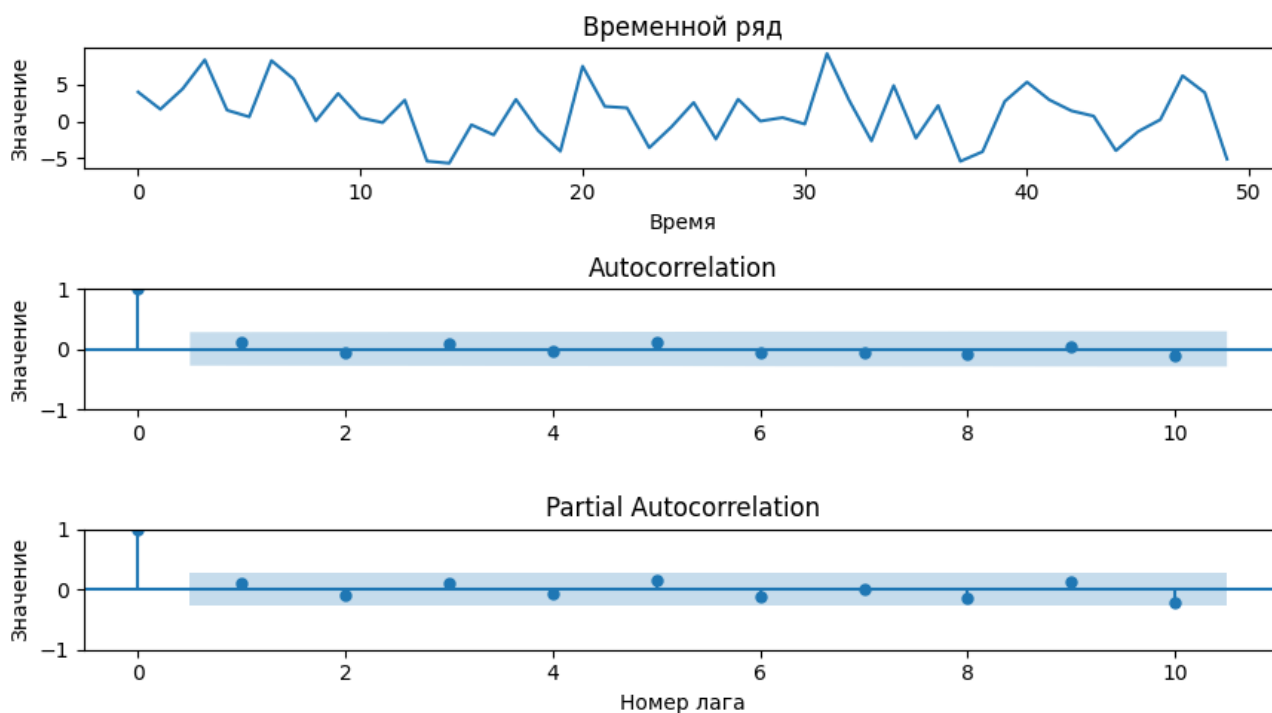
Аналогично получаем:

$$\varphi_{33} = \frac{-\varphi_{21}\rho(2) - \varphi_{22}\rho(1)}{1 - \varphi_{21}\rho(1) - \varphi_{22}\rho(2)} = 0.0124$$

с)

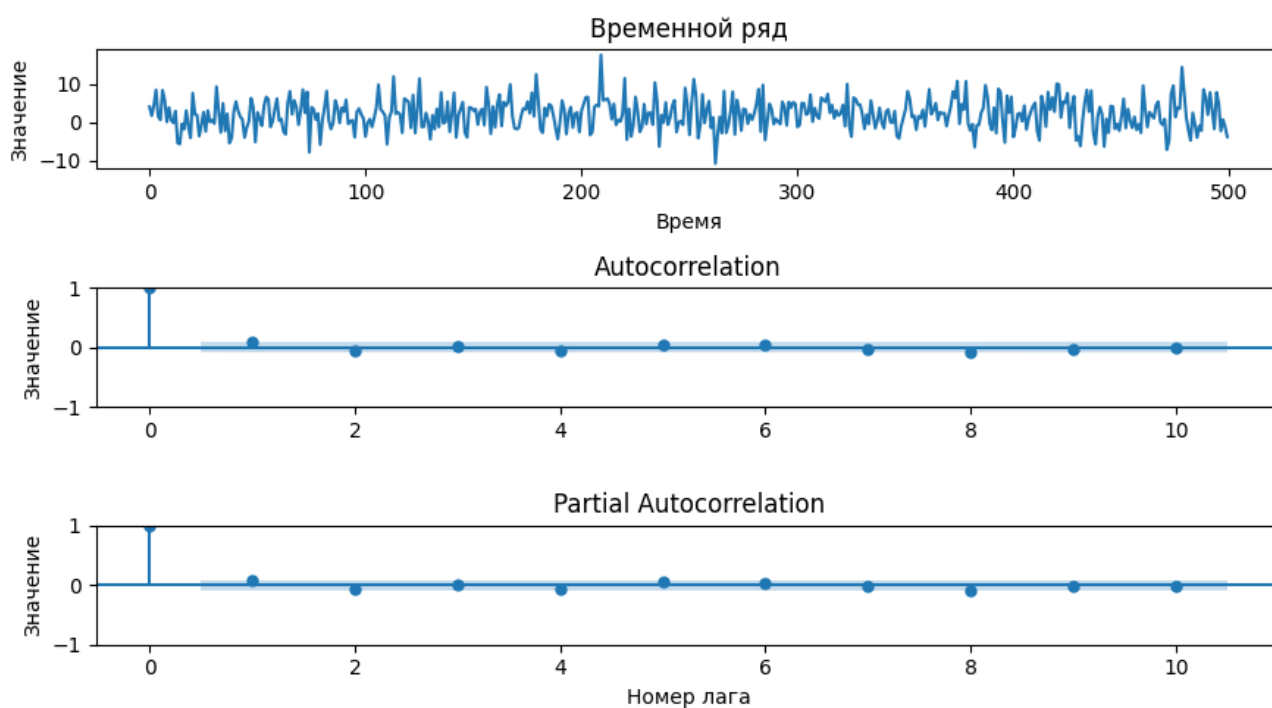
Колебание ряда вокруг одного среднего уровня, отсутствие значимых автокорреляций после 0-го лага и затухание PACF после 2-го лага подтверждают, что сгенерированный временной ряд действительно относится к процессу MA(2). Небольшие отличия выборочных характеристик от

теоретических происходят из-за маленького размера выборки.



d)

На графике временного ряда также видны колебания вокруг одного уровня. Главный эффект увеличения числа наблюдений виден в диаграммах ACF и PACF. Истинный коэффициент ρ_1 почти совпадает с верхней границей интервала и выделяется небольшим ростом на первом лаге. Для ρ_2 модуль оказывается ниже порогового значения, и точка на 2-м лаге лежит внутри доверительной области, а начиная с третьего лага ACF обнуляется.



Доверительный интервал стал более узким.

Задание 2

a)

$$E(y_t) = E(2 + y_{t-1} + \varepsilon_t) = 2 + 0 + E(y_{t-1}) = 2 + 2 + E(y_{t-2}) = \dots = 2t + E(y_0) = 2t + 1$$

$$Var(y_t) = Var(2 + y_{t-1} + \varepsilon_t) = 4 + Var(y_{t-1}) = \dots = 4t$$

$$cov(y_{12}, y_{15}) = cov(y_{12}, 2 + y_{14} + \varepsilon_{15}) = \dots = cov(y_{12}, 6 + y_{12} + \varepsilon_{13} + \varepsilon_{14} + \varepsilon_{15}) = cov(y_{12}, y_{12}) = Var(y_{12}) = 4 * 12 = 48$$

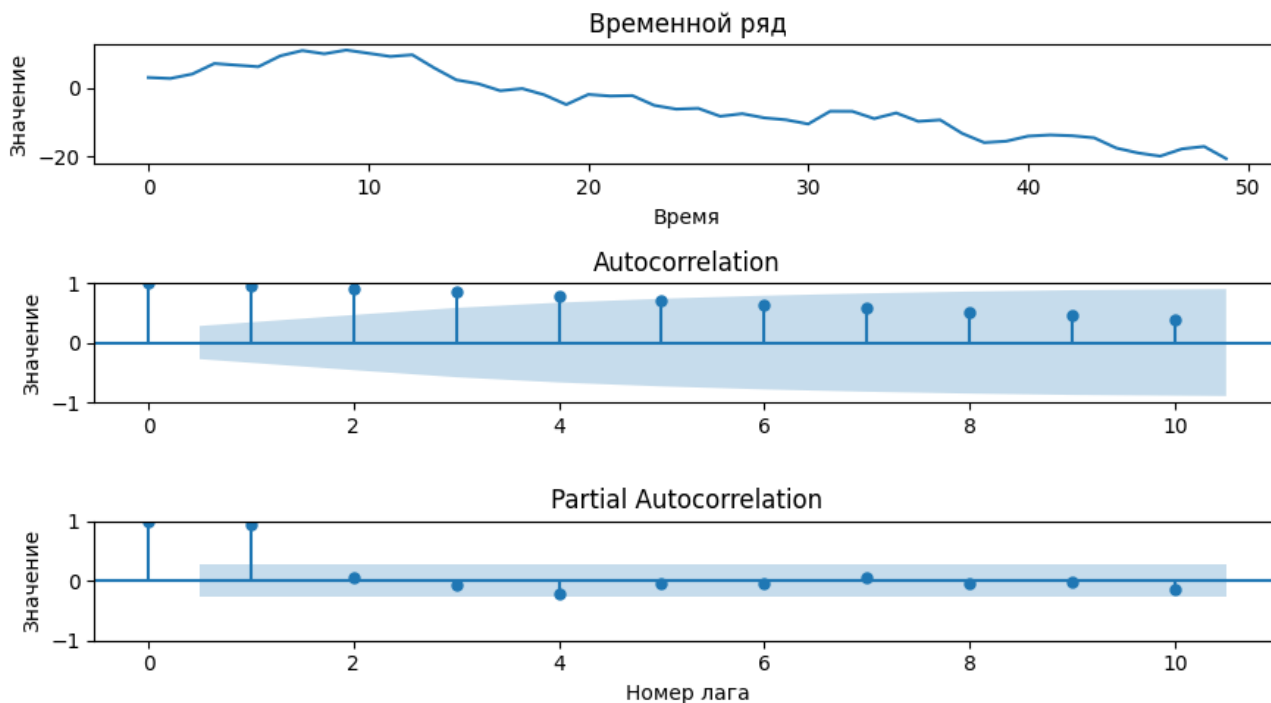
b)

$$corr(y_1, y_5) = \frac{cov(y_1, y_5)}{\sqrt{Var(y_1)Var(y_5)}} = \frac{\sqrt{Var(y_1)}}{\sqrt{Var(y_5)}} = \frac{1}{\sqrt{5}} \approx 0,4472$$

$$corr(y_1, y_5) = \frac{cov(y_{101}, y_{105})}{\sqrt{Var(y_{101})Var(y_{105})}} = \frac{\sqrt{Var(y_{101})}}{\sqrt{Var(y_{105})}} = \frac{\sqrt{101}}{\sqrt{105}} \approx 0.9808$$

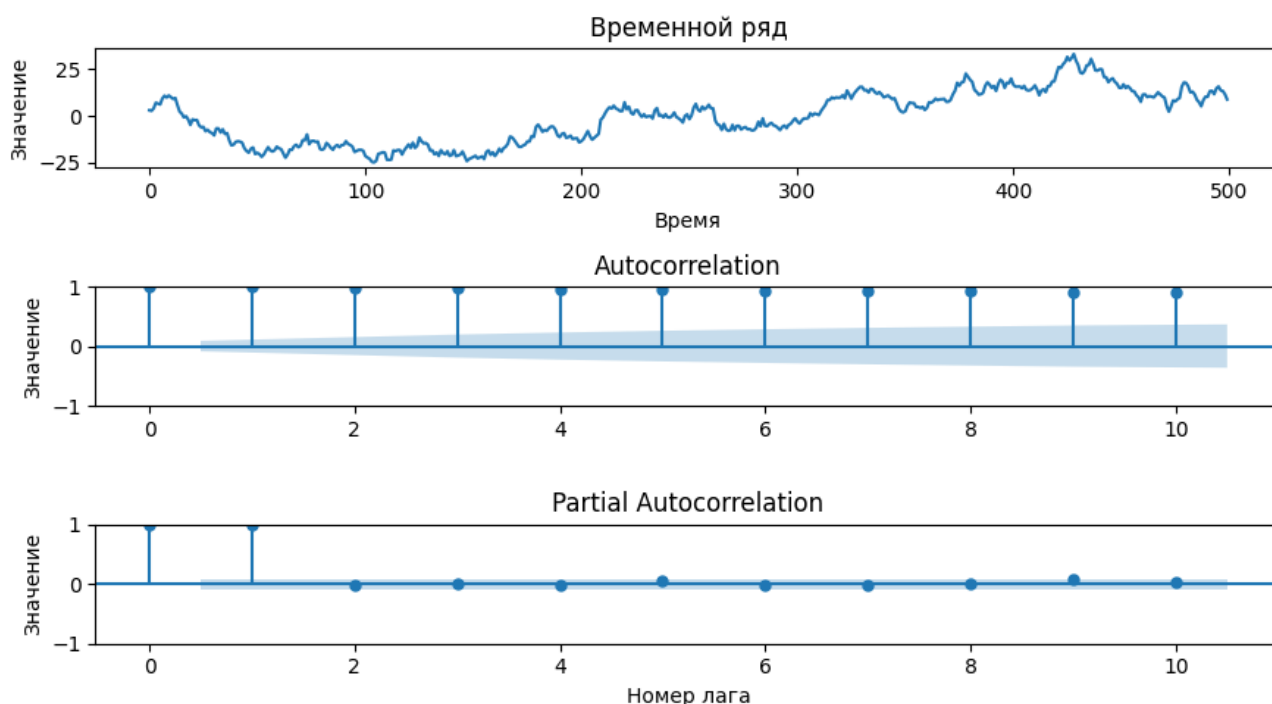
c)

Полученный график показывает выраженное движение: ряд не колеблется вокруг постоянного уровня, а медленно дрейфует.



ACF показывает нестационарность. На первом лаге она практически близка к 1, после чего медленно убывает и остаётся положительной даже на десятом шаге, как в стационарных AR-процессах. PACF показывает единственный значимый всплеск на первом лаге и не показывает пиков далее, что также типично для случайного блуждания.

d)



Результаты повторяют прошлый пункт, но в более приближённом к теории выводе.

При десятикратном увеличении ряда наблюдаемое поведение стало ещё более показательным: выборочная ACF почти не убывает, а PACF имеет пик только на первом лаге, что указывает на наличие единичного корня.

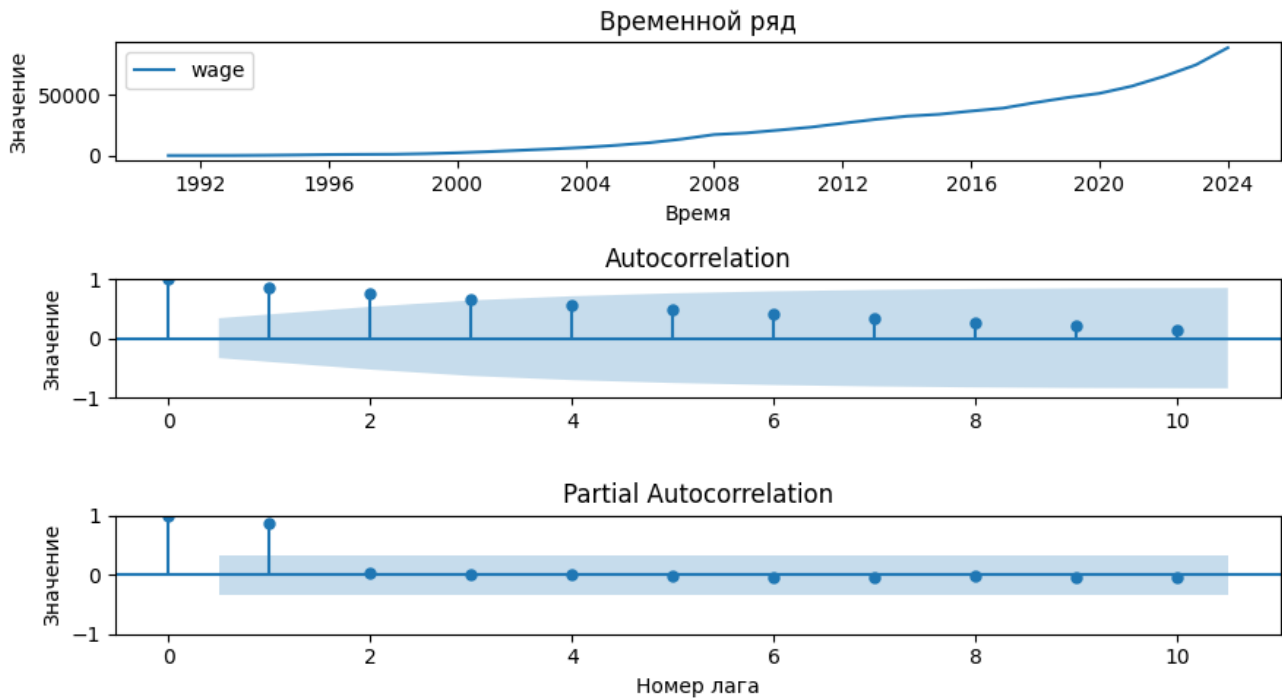
Задание 3

Для выполнения задания я взял данные по средним за год месячным номинальным зарплатам в России за период с 1991 по 2024 годы.

a)

Видно, что колебаний вокруг постоянного уровня нет: ожидаемое значение зарплаты растёт. Выборочная ACF подтверждает это наблюдение. Коэффициент ρ_1 составляет примерно 0.8, а далее корреляции убывают очень медленно, оставаясь положительными даже на десятом лаге. В PACF наблюдается большая корреляция на первом лаге, после чего все частные автокорреляции оказываются незначимыми и находятся внутри доверительного интервала.

Особенность данных состоит в номинальности зарплат. Тренд во многом обусловлен инфляцией.



b)

Проверим гипотезу о стационарности временного ряда:

H_0 : Ряд нестационарен

H_1 : Ряд стационарен

ADF Statistic: 3.242

p-value: 1.000

Полученное значение статистики 3.242 оказалось выше критических уровней – оно положительно. p-value равно 1, что означает полное отсутствие оснований отвергнуть нулевую гипотезу. Следовательно, ряд имеет единичный корень и не является стационарным.

c)

В качестве базовой спецификации для прогноза номинальной заработной платы была выбрана экспоненциальная модель сглаживания без сезонности, но с аддитивным трендом — ETS(AAN). Модель имеет следующий вид:

$$y_t = L_{t-1} + b_{t-1} + \varepsilon_t,$$

$$L_t = L_{t-1} + b_{t-1} + \alpha \varepsilon_t,$$

$$b_t = b_{t-1} + \beta \varepsilon_t,$$

$$\varepsilon_t \sim N(0, \sigma^2) \text{ и независимы,}$$

где L_t и b_t - уровень и тренд, α и β - коэффициенты сглаживания уровня и тренда.

Оба коэффициента сглаживания при оценке оказались почти равными единице. Это типично для рядов со быстро растущим уровнем и подтверждает вывод о нестационарности.

ETS Results

Dep. Variable:	wage	No. Observations:	34			
Model:	ETS(AAN)	Log Likelihood	-290.354			
Date:	Wed, 11 Jun 2025	AIC	590.709			
Time:	14:54:56	BIC	598.340			
Sample:	01-01-1991	HQIC	593.311			
	- 01-01-2024	Scale	1531542.013			
Covariance Type:	approx					
=====						
	coef	std err	z	P> z	[0.025	0.975]

smoothing_level	0.9999	0.310	3.221	0.001	0.391	1.608
smoothing_trend	0.9998	0.268	3.730	0.000	0.474	1.525
initial_level	-206.5654	1240.267	-0.167	0.868	-2637.445	2224.314
initial_trend	351.8043	1756.111	0.200	0.841	-3090.110	3793.719
=====						
Ljung-Box (Q):	1.61	Jarque-Bera (JB):	32.17			
Prob(Q):	0.45	Prob(JB):	0.00			
Heteroskedasticity (H):	21.00	Skew:	1.31			
Prob(H) (two-sided):	0.00	Kurtosis:	6.98			

С оценёнными параметрами модель выглядит так:

$$y_t = L_{t-1} + b_{t-1} + \varepsilon_t,$$

$$L_t = L_{t-1} + b_{t-1} + 0.999\varepsilon_t,$$

$$b_t = b_{t-1} + 0.998\varepsilon_t,$$

$\varepsilon_t \sim N(0, \sigma^2)$ и независимы,

$L_0 = -206.565$, $b_0 = 351.804$ – начальные уровни.

d)

80%-интервал для одного шага: [101696.74 ; 104868.72]

80%-интервал для двух шагов: [113950.98 ; 121042.89]

Для горизонта $h = 1$ год 80-процентный интервал составил от 101 696.74 до 104 868.72 руб. в среднем за месяц. Центральная точка прогноза, таким образом, находится примерно посередине указанного диапазона, то есть около 103 300 руб. При переходе к горизонту $h = 2$ года интервал расширяется до [113 950,98 ; 121 042,89] руб.

e)

Результаты расчёта доверительного интервала встроенными методами показали схожий результат:

	pi_lower	pi_upper
2025-01-01	101696.74	104868.72
2026-01-01	113950.98	121042.89

f)

На основе коррелограмм я выяснил, что оптимальные значения p – от 0 до 1, q – от 0 до 3. Параметр дифференцирования подбирался в диапазоне от 0 до 2. Затем оценивались модели ARIMA для всех комбинаций (p, d, q) . В результате минимальным AIC обладала модель ARIMA(1, 2, 3), минимальным BIC – ARIMA(0, 2, 0):

Минимальный AIC: 547.961, Параметры: (1, 2, 3)

Минимальный BIC: 551.915, Параметры: (0, 2, 0)

Параметры: (1, 2, 3),

AIC: 547.961,

BIC: 555.29

Для дальнейшей работы была выбрана модель ARIMA(1, 2, 3).

g)

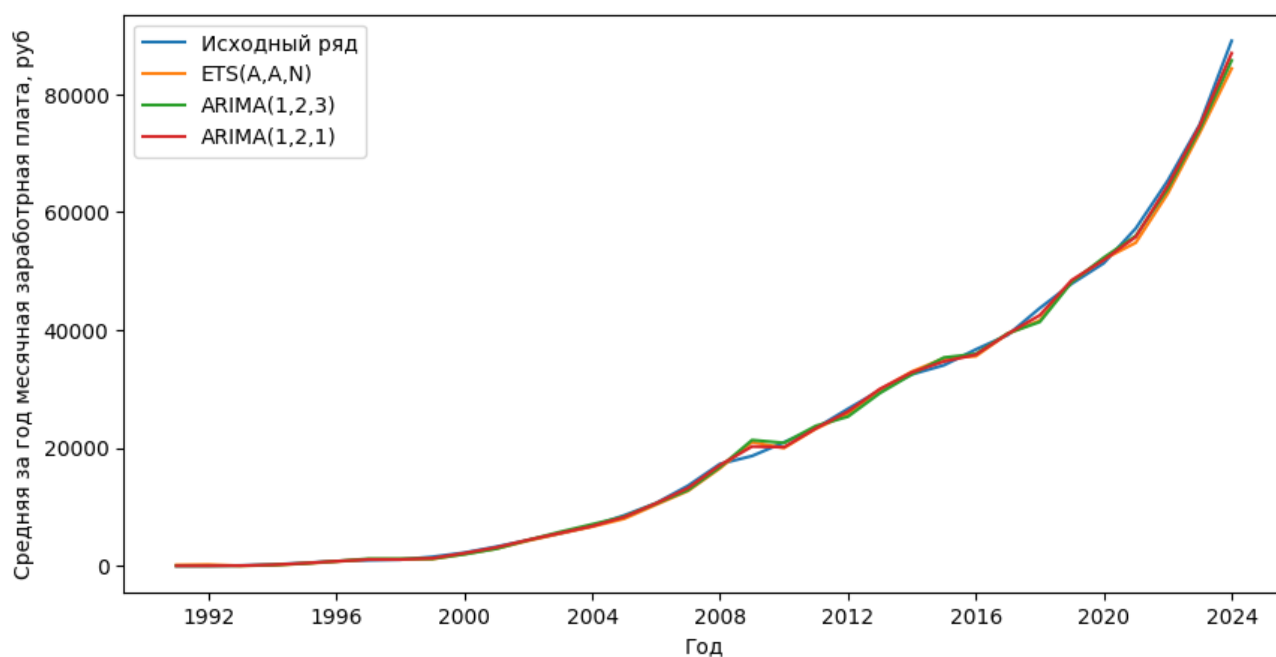
Автоматический перебор моделей показал, что минимальным AIC обладает модель ARIMA(1, 2, 1):

ARIMA(1, 2, 1)

AIC: 547.59, BIC: 551.987

Тем самым ARIMA(1, 2, 1) улучшила как AIC, так и BIC по сравнению с ARIMA(1, 2, 3). Фактически ARIMA(1, 2, 1) выглядит оптимальным компромиссом между качеством краткосрочного прогноза и устойчивостью параметров.

h)



Кривые практически совпадают на всём интервале. Важная особенность данных -

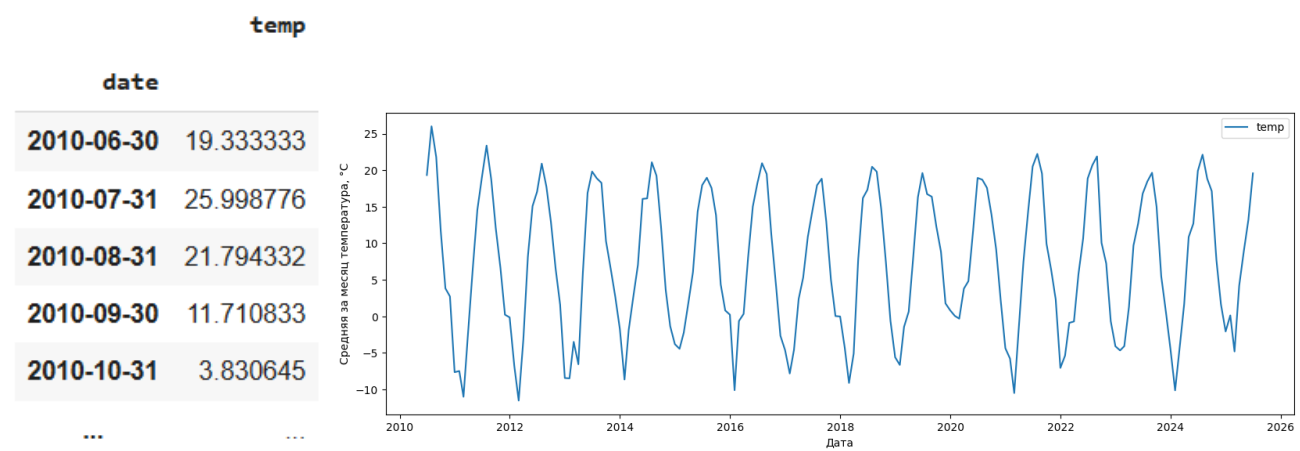
экспоненциально ускоряющийся тренд, и эту особенность изобразили все модели.
Незначительные расхождения проявляются лишь в локальных участках.

Задание 4

Данные для задания 4 получены с московской метеостанции ВДНХ: датасет содержит измерения за каждые три часа, начиная с 1991 года и заканчивая июнем 2025-го.

	Местное время в Москве (ВДНХ)	T	Po	P	Pa		U	DD	Ff	ff10	ff3 ...		Cm	Ch	VV	Td	RRR		tR	E	Tg	E'	sss
10.06.2025 21:00	16.1	748.5	761.8	0.7	50.0	Ветер, дующий с западо-северо- запада	1.0	NaN	NaN	90 или более, но не 100%	...	Перистые нитевидные, иногда когтевидные, не ра...	20.0	5.7	Осадков нет	12.0			NaN	NaN	NaN	NaN	NaN
10.06.2025 18:00	18.8	747.8	761.0	0.2	42.0	Ветер, дующий с северо-запада	1.0	NaN	NaN	60%	...	Перистые (часто в виде полос) и перисто-слоист...	20.0	5.6	Осадков нет	12.0			NaN	NaN	NaN	NaN	NaN
10.06.2025 15:00	19.9	747.6	760.6	-0.2	41.0	Ветер, дующий с западо-северо- запада	1.0	NaN	NaN	20–30%	...	Перистых, перисто-кучевых или перисто-слоистых...	20.0	6.2	NaN	NaN			NaN	NaN	NaN	NaN	NaN
10.06.2025 12:00	18.8	747.8	761.0	0.1	47.0	Ветер, дующий с северо-запада	1.0	NaN	NaN	70 – 80%	...	Перистые нитевидные, иногда когтевидные, не ра...	20.0	7.1	NaN	NaN			NaN	NaN	NaN	NaN	NaN
10.06.2025 09:00	16.3	747.7	761.0	0.6	62.0	Ветер, дующий с западо-северо- запада	2.0	NaN	NaN	70 – 80%	...	Перистых, перисто-кучевых или перисто-слоистых...	20.0	9.0	Осадков нет	12.0	Поверхность почвы влажная.	14.0	NaN	NaN	NaN	NaN	

Из исходного массива выделена температура воздуха в градусах Цельсия, после чего данные с частотой в 3 часа были преобразованы в месячные с помощью усреднения температуры по месяцам. Полученные данные выглядят следующим образом:

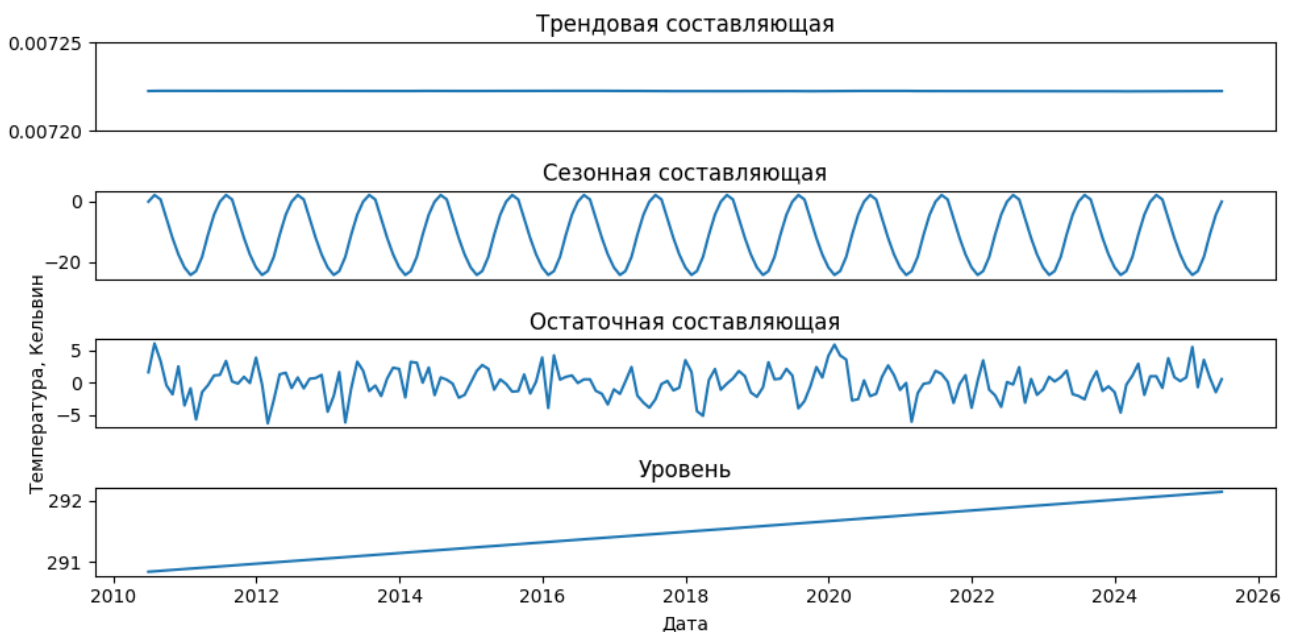


Поскольку температура по шкале Цельсия может принимать отрицательные значения, что затрудняет применение моделей, где предполагаются только положительные величины или, например, планируется логарифмирование, я решил сменить шкалу Цельсия на абсолютную шкалу Кельвина путём прибавления константы 273.15 к каждому наблюдению. После преобразования диапазон смещается в положительную область, и те же пять наблюдений принимают следующий вид:

temp	
date	
2010-06-30	292.483333
2010-07-31	299.148776
2010-08-31	294.944332
2010-09-30	284.860833
2010-10-31	276.980645
...	...

a)

Разложение с помощью ETS(AAA) показывает, что ряд имеет чёткую сезонность: амплитуда колебаний около 20 К и имеет постоянную форму, что позволяет говорить о стабильности годового цикла континентального климата Москвы. Трендовая составляющая, напротив, почти горизонтальна.



Изменение уровня показывает плавный линейный рост «базовой» температуры примерно на 1 градус за 15 лет. Оставшийся шум колеблется в пределах 5 К. Таким образом, ряд неплохо описывается моделью ETS(AAA): сезонная составляющая объясняет большую часть дисперсии, но тренд выражен слабо.

b)

Выборка была разделена в соответствии с заданием.

c)

Для ETS(AAA) на данных критерии оказались следующими: AIC = 744,9 и BIC = 799,9. Параметры сглаживания практически нулевые: $\alpha \approx 0,0001$, $\beta \approx 0,0$, $\gamma \approx 0,0001$. Это значит, что новые

наблюдения мало влияют на состояние модели: уровень сохраняется вблизи 291 К, линейный тренд остаётся около 0,004 К в месяц (около 0,05 К в год), а сезонные индексы формируют амплитуду в 20 К, сопоставимую с реальным годовым ходом температуры.

ETS Results						
=====						
Dep. Variable:	temp	No. Observations:	157			
Model:	ETS(AAA)	Log Likelihood	-354.426			
Date:	Wed, 11 Jun 2025	AIC	744.851			
Time:	17:40:38	BIC	799.864			
Sample:	06-30-2010	HQIC	767.194			
	- 06-30-2023	Scale	5.350			
Covariance Type:	approx					
=====						
	coef	std err	z	P> z	[0.025	0.975]

smoothing_level	0.0001	nan	nan	nan	nan	nan
smoothing_trend	1e-08	0.002	4.42e-06	1.000	-0.004	0.004
smoothing_seasonal	9.999e-05	nan	nan	nan	nan	nan
initial_level	290.9477	nan	nan	nan	nan	nan
initial_trend	0.0043	0.006	0.755	0.450	-0.007	0.016
initial_seasonal.0	-3.9489	nan	nan	nan	nan	nan
initial_seasonal.1	-11.1564	nan	nan	nan	nan	nan
initial_seasonal.2	-18.6036	nan	nan	nan	nan	nan
initial_seasonal.3	-22.8855	nan	nan	nan	nan	nan
initial_seasonal.4	-24.3176	nan	nan	nan	nan	nan
initial_seasonal.5	-21.7010	nan	nan	nan	nan	nan
initial_seasonal.6	-17.4146	nan	nan	nan	nan	nan
initial_seasonal.7	-11.9584	nan	nan	nan	nan	nan
initial_seasonal.8	-5.9522	nan	nan	nan	nan	nan
initial_seasonal.9	0.8538	nan	nan	nan	nan	nan
initial_seasonal.10	2.4145	nan	nan	nan	nan	nan
initial_seasonal.11	0	nan	nan	nan	nan	nan
=====						
Ljung-Box (Q):	41.84	Jarque-Bera (JB):	2.12			
Prob(Q):	0.01	Prob(JB):	0.35			
Heteroskedasticity (H):	0.92	Skew:	-0.26			
Prob(H) (two-sided):	0.77	Kurtosis:	3.24			

$$y_t = L_{t-1} + b_{t-1} + s_{t-12} + \varepsilon_t,$$

$$L_t = L_{t-1} + b_{t-1} + \alpha \varepsilon_t,$$

$$b_t = b_{t-1} + \beta \varepsilon_t,$$

$$s_t = s_{t-12} + \gamma u_t,$$

где L_t и b_t - уровень и тренд, α и β - коэффициенты сглаживания уровня и тренда.

После оценки модель приняла следующий вид:

$$y_t = L_{t-1} + b_{t-1} + s_{t-12} + \varepsilon_t,$$

$$L_t = L_{t-1} + b_{t-1} + 0.0001 \varepsilon_t,$$

$$b_t = b_{t-1} + 10^{-8} \varepsilon_t,$$

$$s_t = s_{t-12} + 9.999 * 10^{-5} u_t$$

$$L_0 = 290.9477, b_0 = 0.0043, s_0 = -3.9489 \dots, s_{11} = 0$$

Сезонная наивная модель выглядит следующим образом:

$$\hat{y}_t = y_{t-12}$$

ETS(AAA) для логарифмированных данных показала следующие критерии: AIC = -1018,9 и BIC = -963,9. Сглаживающие коэффициенты снова почти нулевые, тренд фактически исчезает, а сезонные эффекты сохраняют прежнюю амплитуду.

$$\ln(y_t) = L_{t-1} + b_{t-1} + s_{t-12} + \varepsilon_t,$$

$$L_t = L_{t-1} + b_{t-1} + \alpha \varepsilon_t,$$

$$b_t = b_{t-1} + \beta \varepsilon_t,$$

$$s_t = s_{t-12} + \gamma u_t,$$

где L_t и b_t - уровень и тренд, α и β - коэффициенты сглаживания уровня и тренда.

ETS Results						
=====						
Dep. Variable:	temp	No. Observations:	157			
Model:	ETS(AAA)	Log Likelihood	527.442			
Date:	Wed, 11 Jun 2025	AIC	-1018.884			
Time:	17:40:40	BIC	-963.872			
Sample:	06-30-2010	HQIC	-996.542			
	- 06-30-2023	Scale	0.000			
Covariance Type:	approx					
=====						
	coef	std err	z	P> z	[0.025	0.975]

smoothing_level	0.0001	nan	nan	nan	nan	nan
smoothing_trend	1e-08	0.002	4.46e-06	1.000	-0.004	0.004
smoothing_seasonal	9.999e-05	nan	nan	nan	nan	nan
initial_level	5.6730	nan	nan	nan	nan	nan
initial_trend	1.699e-05	2e-05	0.848	0.396	-2.23e-05	5.63e-05
initial_seasonal.0	-0.0137	nan	nan	nan	nan	nan
initial_seasonal.1	-0.0391	nan	nan	nan	nan	nan
initial_seasonal.2	-0.0660	nan	nan	nan	nan	nan
initial_seasonal.3	-0.0819	nan	nan	nan	nan	nan
initial_seasonal.4	-0.0873	nan	nan	nan	nan	nan
initial_seasonal.5	-0.0775	nan	nan	nan	nan	nan
initial_seasonal.6	-0.0617	nan	nan	nan	nan	nan
initial_seasonal.7	-0.0420	nan	nan	nan	nan	nan
initial_seasonal.8	-0.0206	nan	nan	nan	nan	nan
initial_seasonal.9	0.0030	nan	nan	nan	nan	nan
initial_seasonal.10	0.0082	nan	nan	nan	nan	nan
initial_seasonal.11	0	nan	nan	nan	nan	nan
=====						
Ljung-Box (Q):	40.67	Jarque-Bera (JB):	3.29			
Prob(Q):	0.02	Prob(JB):	0.19			
Heteroskedasticity (H):	0.91	Skew:	-0.30			
Prob(H) (two-sided):	0.73	Kurtosis:	3.38			

После оценки уравнение модели выглядит так:

$$\ln(y_t) = L_{t-1} + b_{t-1} + s_{t-12} + \varepsilon_t,$$

$$L_t = L_{t-1} + b_{t-1} + 0.0001\varepsilon_t,$$

$$b_t = b_{t-1} + 10^{-8}\varepsilon_t,$$

$$s_t = s_{t-12} + 9.999 * 10^{-5}u_t,$$

$$L_0 = 5.673, b_0 = 1.699 * 10^{-5}, s_0 = -0.0137 \dots, s_{11} = 0$$

d)

RMSE модели без логарифмирования: 2.39

RMSE snaiive: 2.83

RMSE модели с логарифмированием: 2.38

MAE модели без логарифмирования: 1.84
MAE snaiѵe модели: 2.18
MAE модели с логарифмированием: 1.83

MAPE модели без логарифмирования: 0.0066
MAPE snaiѵe модели: 0.0077
MAPE модели с логарифмированием: 0.0065

MASE модели без логарифмирования: 0.36
MASE snaiѵe: 0.42
MASE модели с логарифмированием: 0.35

По результатам сравнения ETS с логарифмированием оказался наилучшей моделью, но лишь незначительно уступая обычному ETS. Наивный прогноз оказался наименее точным.

е)

RMSE усредненного прогноза: 2.38

После усреднения RMSE усреднённого прогноза, ожидаемо, оказался между RMSE обычного ETS и ETS по логарифму.

