

Домашнее задание №3.

Домашнее задание сдается в электронном виде в SmartLMS.

Необходимо загрузить два файла: отчет (pdf) и отдельно код. Домашние задания без отчета оцениваются 0 баллов.

Дедлайн: 11 июня 2025 года, 23:59.

1. (20 баллов) Рассмотрим $MA(2)$ процесс $y_t = 2 + \varepsilon_t + 0.1\varepsilon_{t-1} - 0.06\varepsilon_{t-2}$, где величины ε_t независимы и нормально распределены $N(0; 16)$.
 - (a) (5 баллов) Рассчитайте теоретическую автокорреляционную функцию процесса.
 - (b) (5 баллов) Рассчитайте первые три значения теоретической частной автокорреляционной функции $\phi_{11}, \phi_{22}, \phi_{33}$.
 - (c) (5 баллов) Сгенерируйте траекторию данного процесса длиной 50 наблюдений. Постройте график ряда, график первых десяти значений выборочной ACF и $PACF$. Прокомментируйте поведение выборочных ACF и $PACF$.
 - (d) (5 баллов) Повторите предыдущий пункт для 500 наблюдений. Верно ли, что с ростом числа наблюдений выборочная ACF сходится к истинной ACF , а выборочная $PACF$ к истинной $PACF$?
2. (20 баллов) Рассмотрим процесс случайного блуждания $y_t = 2 + y_{t-1} + \varepsilon_t$, где величины ε_t независимы и нормально распределены $N(0; 4)$, а $y_0 = 1$.
 - (a) (5 баллов) Рассчитайте $E(y_t), \text{Var}(y_t), \text{Cov}(y_{12}, y_{15})$.
 - (b) (5 баллов) Сравните $\text{Corr}(y_1, y_5)$ и $\text{Corr}(y_{101}, y_{105})$.
 - (c) (5 баллов) Сгенерируйте траекторию данного процесса длиной 50 наблюдений. Постройте график ряда, график первых десяти значений выборочной ACF и $PACF$. Прокомментируйте поведение выборочных ACF и $PACF$.
 - (d) (5 баллов) Повторите предыдущий пункт для 500 наблюдений. Верно ли, что с ростом числа наблюдений выборочная ACF сходится к истинной ACF , а выборочная $PACF$ к истинной $PACF$?
3. (20 баллов) Возьмите любой свежий несезонный ряд годовой периодичности. Можно взять ряд с <https://fedstat.ru/>, <http://sophist.hse.ru/> или других источников.
 - (a) (3 балла) Постройте график ряда, графики выборочных ACF и $PACF$.
 - (b) (3 балла) Визуально оцените, есть ли тренд? Похож ли процесс на стационарный? Проверьте ряд на стационарность с помощью ADF теста (сформулируйте нулевую и альтернативную гипотезы, укажите статистику и ее распределение, обоснуйте выбор спецификации ADF теста).

- (c) (3 балла) Оцените для ряда оптимальную ETS модель. Выпишите полученные уравнение, используя оценённые значения параметров вместо параметров.
 - (d) (3 балла) Получите 80%-й доверительный интервал на один и два шага вперёд «руками», исходя из выписанных в предыдущем пункте уравнений.
 - (e) (2 балла) Получите 80%-й доверительный интервал на один и два шага вперёд встроенными функциями. Сравните с результатами предыдущего пункта.
 - (f) (3 балла) С помощью процедуры Бокса-Дженкинсона подберите оптимальную $ARIMA(p, d, q)$ модель для исследуемого ряда, описав шаги алгоритма.
 - (g) (1 балл) Оцените для ряда оптимальную $ARIMA(p, d, q)$ модель, используя процедуру автоматического подбора. Сравните полученные результаты.
 - (h) (2 балла) Постройте график исходного ряда и прогнозов, полученных по моделям из пунктов (c), (f) и (g).
4. (15 баллов) Возьмите любой свежий сезонный ряд квартальной или месячной периодичности.
- (a) (3 балла) Осуществите разложение ряда на составляющие, используя $ETS(AAA)$ модель. Постройте графики составляющих ряда. Прокомментируйте результаты разложения.
 - (b) (1 балл) Разделите данные на обучающую и тестовую выборку, выделив на тестовую выборку два года наблюдений.
 - (c) (5 баллов) Оцените $ETS(AAA)$ и сезонную наивную модель для исходного ряда и $ETS(AAA)$ для логарифма ряда. Выпишите уравнения оцененных моделей.
 - (d) (3 балла) Для каждого подхода найдите $RMSE$, MAE , $MAPE$, $MASE$ на тестовой выборке.
 - (e) (3 балла) Постройте прогноз, усредняющий прогнозы двух лидирующих по $RMSE$ подходов. Удалось ли обыграть два усредняемых подхода?
5. (25 баллов) В файле `data_hw_03.xlsx` содержатся данные из исследования Vella, Verbick (1998) о мужчинах, которые работали каждый год с 1980 по 1987 год. Рассмотрим уравнение для заработной платы:

$$\ln(wage_{it}) = \theta_t + \beta_1 educ_i + \beta_2 black_i + \beta_3 hispan_i + \beta_4 exper_{it} + \beta_5 exper_{it}^2 + \beta_6 married_{it} + \beta_7 married_{it} + \beta_7 union_{it} + \alpha_i + \varepsilon_{it},$$

где nr — номер индивида

$lwage$ — логарифм заработной платы

$educ$ — образование, в годах

$exper$ — стаж, в годах

$exper^2$ — квадрат стажа

$married = 1$, если состоит в браке

$black = 1$, если афроамериканец

$hispan = 1$, если латиноамериканец

$union = 1$, если является членом профсоюза.

- (a) (5 баллов) С помощью МНК оцените *pooled* модель для заработной платы. Надежны ли стандартные ошибки оценок коэффициентов, полученных с помощью МНК, даже если индивидуальные эффекты α_i не коррелирует со всеми объясняющими переменными? Аргументируйте свой ответ. Вычислите соответствующие стандартные ошибки оценок коэффициентов регрессии.
- (b) (5 баллов) Оцените уравнение заработной платы, используя модель со случайными эффектами. Сравните полученные оценки с результатами предыдущего пункта.
- (c) (5 баллов) Оцените модель с фиксированными эффектами. Почему $exper_{it}$ является лишним фактором в модели, хотя он меняется со временем? Как изменились оценки коэффициентов при переменных $married_{it}$ и $union_{it}$ по сравнению с оценками в модели со случайными эффектами?
- (d) (5 баллов) Теперь добавьте взаимодействия в виде $d81 \cdot educ$, $d82 \cdot educ$;...; $d87 \cdot educ$ и оцените модель с фиксированными эффектами. Увеличивается ли отдача от образования с течением времени?
- (e) (5 баллов) Вернитесь к модели с фиксированными эффектами из пункта (c). Добавьте в число регрессоров переменную $union_{i,t+1}$ и оцените модель с фиксированными эффектами заново (обратите внимание, что вы потеряете данные за 1987 год). Значим ли коэффициент при переменной $union_{i,t+1}$? Что этот результат говорит об экзогенности членства в профсоюзе?