

ECON3110 - Hjemmeeksamen 2025

Høsten 2025

Denne eksamensoppgaven består av 3 oppgaver. Forsøk å svare på alle oppgavene. For hver oppgave teller alle del-oppgaver likt.

Alle spørsmålene skal besvares med tekst. For noen oppgaver er det relevant å legge med R-koden du har brukt. Du kan selv velge hvilket program du vil skrive besvarelsen i, men merk at besvarelsen må leveres som én PDF-fil. Sørg for at det er et tydelig skille mellom tekst og R-kode i besvarelsen din. Husk å legge ved kommentarer i koden.

Oppgavesettet består av 3 oppgaver fordelt på 3 sider. Følgende datasett brukes:

Oppgave 1: Corruption Perception Index: corruption.csv ([lenke](#))

Oppgave 2: Time Series data: time_series.csv ([lenke](#))

Oppgave 1 (50% vekt)

I denne oppgaven skal du bruke en utvidet versjon av datasettet Corruption Perception Index (2012 - 2024). CPI indeksen går fra 0 (mest korrupsjon) til 10 (minst korrupsjon). Datasettet finner du [her](#). I datasettet finnes det følgende variabler:

country	Navn på landet
iso3c	ISO kode, 3 bokstaver
region	Region
Year	År
CPI	Korrupsjonsindeks
gdp	BNP per kapita, kjøpekraftsjustert
population	Befolkning (tusener)
school_enroll	Andel i videregående skole
area	Størrelse i km ²
nat_rents	Andel av BNP fra naturressurser
armed_forces	Militær som andel av arbeidsstyrken
mammal_threat	Antall truede pattedyr
rail_lines	Antall km med jernbane
real_interest_rate	Reell rente
poverty_gap830	Fattigdomsmål
pop_growth_pct	Befolkningsvekst siste år i prosent
pop_largest_city	Andel av befolkning i største by
pop_density	Befolkningstetthet
democracy	Demokrati-indeks
union_freedom	Fagforeningsfrihet - indeks
rule_of_law_index	Rettstat - indeks
x1	Tilfeldig støy
x2	Tilfeldig støy
x3	Tilfeldig støy
no_variable_control	En kontroll for hver kovariat; indikerer manglende verdi for variabel.

Merk: Alle manglende verdier er erstattet med -1000 , derav en variabel som indikerer manglende verdi.

a)

Standardiser variablene i datasettet, og del datasettet i én treningsdel og en testdel.

b)

Definer en funksjon for K-fold kryssvalidering av en algoritme.

Funksjonen skal ta som input et datasett og antall folder (K). Funksjonen skal returnere et datasett med én ny kolonne som indikerer fold.

c)

Bruk funksjonen fra **b)** på trenings-datasettet til å tune en LASSO modell som skal predikere korrupsjon (CPI). Dersom du ikke får til **b)** kan du bruke en alternativ metode. Du kan søke i domenet 0 til 0.1.

d)

Med hyperparameter-konfigurasjonen du kom frem til i **c)**, tren modellen på hele treningsdataen og rapporter OOS-prediksjonsnøyaktighet.

e)

Tolk resultatene fra LASSO modellen - hvilke variabler later til å predikere korrupsjon best?

f)

Gjenta c) og d) med Neural Net. *Dersom neuralnet ikke konvergerer kan du sette $stepmax = 1e+06$ eller høyere.*

Oppgave 2 (30% vekt)

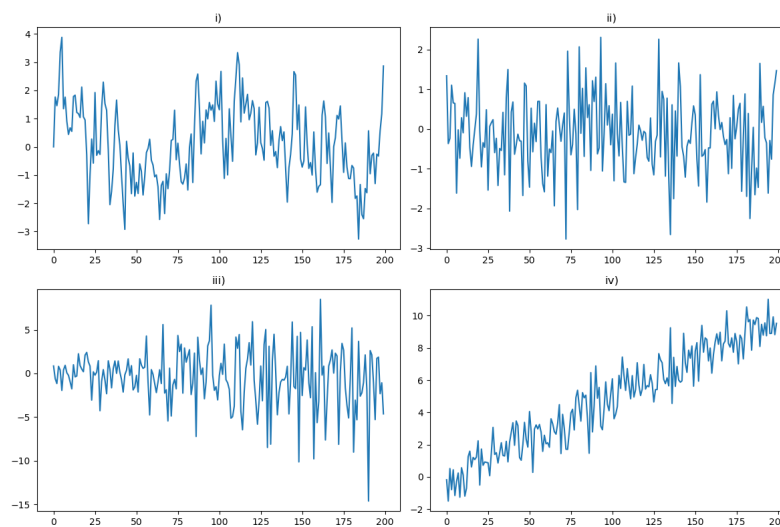


Figure 1: Tidsserier

a)

Figur 1 viser fire tidsserier, i) til iv). Hvilke(n) tidserie(r) er ikke-stasjonær(e)? Begrunn svaret.

b)

Last inn og plot tidsserien *time_series.csv*.

c)

Test om tidsserien er stasjonær, forklar hvilken test du bruker, hvordan du tolker resultatet, og hvordan du løser eventuelle utfordringer med tidsserien.

d)

Del tidsserien opp i en treningsdel (de 80% første observasjonene) og en testdel (de 20% siste observasjonene).

e)

Lag et ACF og PACF plot for tidsserien. Estimer deretter en ARIMA(p, d, q) modell på treningsdatasettet. Du skal selv avgjøre verdiene på p , d , og q . Husk å begrunne valget av verdiene p , d , og q .

f)

Bruk den estimerte modellen fra c) til å gjøre en one-step ahead forecast over hele testdatasettet. Det vil si at dere for hver periode $t + 1$ i testdatasettet skal bruke modellen fra c) til å predikere $E(y_{t+1}|y_t)$. Lag et plot som viser både de faktiske verdiene og de predikerte verdiene. *Hint: Bruk Arima() funksjonen for å predikere.*

Oppgave 3 (20% vekt)

De tre deloppgavene henger ikke sammen.

a)

Du ønsker å estimere effekten av Z på Y , men mistenker at kovariatene $X_1, \dots, X_{5000}$ påvirker både Z og Y . Forklar hvordan du kan bruke DML til å estimere effekten av Z på Y .

b)

Du ønsker å estimere en funksjon $f(X_1, \dots, X_{5000})$ for å predikere Y med $X_1, \dots, X_{5000}$. Uten å vite noe mer enn at interaksjonseffekter er sentrale for prediksjonen, ville du brukt LASSO eller Neural Net for å estimere $f(\cdot)$? Begrunn svaret.

c)

Hva menes med *honesty* i *Causal Forest* algoritmen?