

# Arbejdsløshed - Ugeeksamen i Forecasting

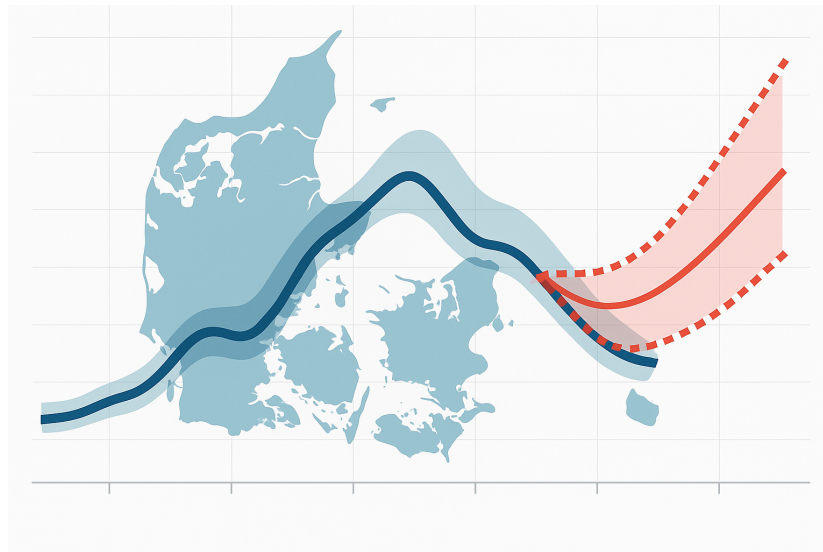
Christine Hegelund

Jing Wei

Marcus Nielsen

2025-06-13

Forecasting Eksamen



**Antal tegn (inkl. mellemrum): 97023**

**Vejledere:**

Bjarne Taulo Sørensen

**Tabel over figurer**

## Indholdsfortegnelse

# 1 Introduktion

Formålet med denne analyse er at modellere og forudsige arbejdsløshedstal i Danmark opdelt på region og køn, baseret på månedlige data fra 2007 til 2019. Ved hjælp af klassiske tidsseriemodeller, herunder ARIMA, ETS og en simpel benchmarkmodel, analyseres trends, sæsonvariation og residualer i data.

Opgaven er struktureret omkring faserne i en klassisk tidsserieanalyse: fra eksplorativ dataanalyse og dekomposition til modelvalg, modelvalidering via time series cross-validation og endelig forecast for året 2020. Den bedste model vælges individuelt for hver af de ti serier (5 regioner  $\times$  2 køn) baseret på performance-metrikker (RMSE og MAPE).

Der lægges vægt på at dokumentere og begrunde de metodiske valg, med udgangspunkt i pensum og best practice fra faget. Prognoserne præsenteres både numerisk og grafisk, og resultaterne sammenlignes på tværs af serier.

Analysen er udelukkende baseret på de udleverede data. Forretningsmæssige forklaringer og COVID-relaterede effekter er bevidst udeladt, i henhold til opgavens afgrænsning.

## 2 Problemformulering

Hvordan kan klassiske tidsseriemodeller anvendes til at analysere og forudsige arbejdsløsheden i Danmark, fordelt på region og køn, baseret på månedlige data fra 2007 til 2019?

Med udgangspunkt i metoder som ARIMA, ETS og en simpel benchmarkmodel er formålet at:

- identificere trends og sæsonmønstre i arbejdsløsheden gennem eksplorativ dataanalyse og dekomposition
- finde og validere den bedst egnede model for hver serie ved brug af time series cross-validation og performance-metrikker som RMSE og MAPE
- forecaste arbejdsløsheden for året 2020 med kvantificeret usikkerhed via prædiktionsintervaller
- sammenligne resultaterne på tværs af regioner og køn for at belyse forskelle i modelperformance og arbejdsløshedsniveauer

### 2.1 Afgrænsning

Analysen er udelukkende baseret på de udleverede data og fokuserer på anvendelse og vurdering af klassiske forecasting-metoder.

Der inddrages ikke eksterne faktorer som COVID-19, konjunkturændringer eller politiske tiltag. Der gives heller ikke forretningsmæssige anbefalinger, da formålet er metodisk – ikke strategisk.

Fokus er således på at demonstrere korrekt og fagligt funderet anvendelse af tidsserieanalyse og forecasting-teknikker i en struktureret, reproducerbar analyse.

### 2.1.1 Anvendelse af AI-værktøj

I forbindelse med udarbejdelsen af denne opgave er ChatGPT (GPT-4o) benyttet som et støtteværktøj. Værktøjet har primært været anvendt til idéudvikling, sproglig sparring, forbedring af formuleringers klarhed, grammatisk gennemgang samt støtte ved udformning af enkelte kodestumper. Anvendelsen har udelukkende omfattet sproglige, strukturelle og tekniske elementer. Alt analytisk, fortolkende og konkluderende indhold er selvstændigt udarbejdet af gruppens medlemmer.

## 2.2 Definitioner/forkortelser

I dette afsnit afklares centrale begreber og forkortelser, der anvendes gennem opgaven, for at sikre en ensartet forståelse.

## 2.3 Struktur

Opgaven er opbygget i overensstemmelse med en klassisk tilgang til tidsserieanalyse. Først gennemføres en eksplorativ dataanalyse (EDA), hvor datasættet undersøges for trend, sæsonvariation og andre karakteristika gennem visualisering, deskriptiv statistik og STL-dekomposition. Herefter følger modelvalgsfasen, hvor tre modeller – ARIMA, ETS og en simpel benchmarkmodel – estimeres for hver tidsserie. Modellerne evalueres gennem time series cross-validation med fokus på RMSE og MAPE, hvorefter den bedst performende model vælges for hver serie. Afslutningsvis foretages forecast for 2020 med tilhørende prædiktionsintervaller, efterfulgt af en sammenlignende vurdering og konklusion.

# 3 Data og forberedelse

Analysen er baseret på et datasæt med månedlige arbejdsløshedstal i Danmark fra januar 2007 til december 2019. Data er opdelt på køn (mænd og kvinder) og region (de fem danske regioner), hvilket resulterer i ti separate tids-serier. Datasættet er udleveret i forbehandlet format som en tsibble med korrekt angivne indeks- og nøglevariabler, hvilket gør det velegnet til modellering i fable/tidyverts-pakken.



## 4 Eksplorativ dataanalyse (EDA)

### 4.1 Visualisering af trends og sæsonmønstre

### 4.2 Deskriptive statistikker

### 4.3 STL-dekomposition

## 5 Modelvalg

### 5.1 ARIMA og SARIMA

### 5.2 ETS (Exponential Smoothing)

### 5.3 Simpel benchmarkmodel

### 5.4 Kort beskrivelse af modellerne

## 6 Modelvalidering

### 6.1 Time series cross-validation

### 6.2 Evaluering med RMSE og MAPE

### 6.3 Valg af bedste model pr. serie

### 6.4 Test for hvid støj

## 7 Forecasting

### 7.1 Træning og forecast for 2020

### 7.2 Prædiktionsintervaller

### 7.3 Visualisering af forecasts

### 7.4 Centrale prognoseværdier

## 8 Sammenligning og fortolkning

### 8.1 Sammenligning på tværs af serier

### 8.2 Tendenser mellem regioner og køn

## 9 Konklusion

## 10 Kildeliste



## 11 Bilagsoversigt

- Bilag 1: Udliveredede Powerpointpræsentationer v. Michael Freundlich: Chefkonsulent, Erhvervsservice og facility