

Arbejdsløshed - Ugeeksamen i Forecasting

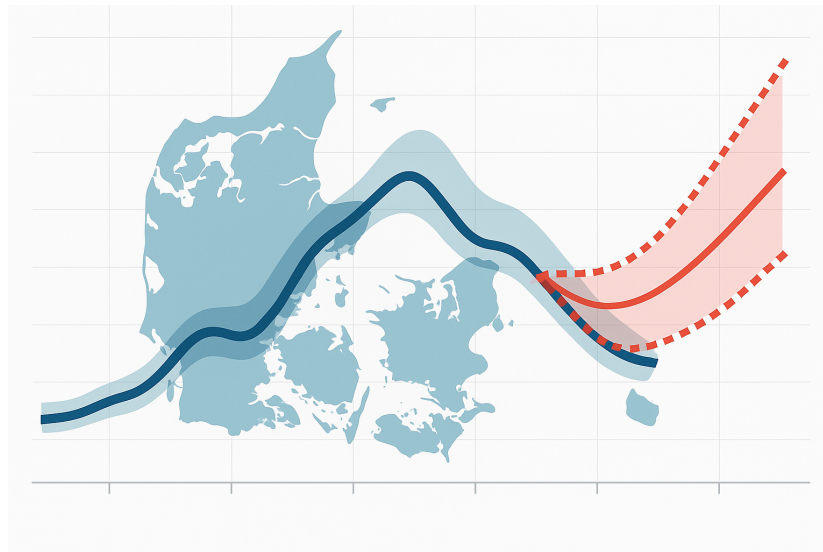
Christine Hegelund

Jing Wei

Marcus Nielsen

2025-06-13

Forecasting Eksamen



Antal tegn (inkl. mellemrum): 97023

Vejledere:

Bjarne Taulo Sørensen

Tabel over figurer

Indholdsfortegnelse

1 Introduktion

Arbejdsløshedstal er en central indikator for et lands økonomiske tilstand og udvikling. De påvirker både den enkelte borger og den nationale økonomi og indgår som en væsentlig faktor i politiske og økonomiske beslutningsprocesser. I takt med øget datatilgængelighed og forbedrede statistiske værktøjer er det blevet muligt at analysere og forudsige sådanne udviklingstræk med større præcision.

Denne opgave har til formål at modellere og forudsige udviklingen i arbejdsløsheden i Danmark, fordelt på køn og region, med afsæt i månedlige data for perioden januar 2007 til december 2019. Datasættet består af ti tidsrækker – én for hver kombination af fem regioner og to køn – hvilket muliggør en detaljeret og differentieret analyse af både strukturelle og sæsonbetingede mønstre i ledigheden.

Metodisk anvender analysen klassiske teknikker til tidsrækkeanalyse med særligt fokus på modellerne ARIMA, ETS og en simpel benchmarkmodel (Seasonal Naive). For hver serie estimeres de tre modeltyper, og deres performance evalueres ved hjælp af forecast-metrikker som RMSE og MAPE. Derudover inddrages elementer som STL-dekomposition, transformation af data (log og Box-Cox), residualanalyse og test for hvid støj (Ljung-Box).

Formålet er ikke blot at fremskrive arbejdsløsheden for året 2020, men også at vurdere styrker og begrænsninger ved klassiske modeller til tidsrækkeanalyse. Med dette udgangspunkt opstår spørgsmålet, hvordan sådanne modeller bedst kan anvendes til at analysere og forudsige arbejdsløsheden i Danmark, fordelt på region og køn.

2 Problemformulering

Hvordan kan klassiske tidsrækemodeller anvendes til at analysere og forudsige udviklingen i arbejdsløsheden i Danmark, fordelt på region og køn?

For at besvare dette spørgsmål undersøges følgende delspørgsmål:

1. Hvordan varierer arbejdsløshedens udvikling og sæsonmønstre på tværs af regioner og køn, og hvordan kan disse identificeres gennem eksplorativ dataanalyse og dekomposition?
2. Hvordan performer modellerne ARIMA, ETS og Seasonal Naive i forhold til hinanden, når det gælder præcision, residualstruktur og prognoseegenskaber?
3. Hvilke modeller er bedst egnede til at fremskrive arbejdsløshedstallene for 2020, og hvordan varierer resultater og usikkerhed på tværs af serierne?

2.1 Afgrænsning

Analysen er udelukkende baseret på de udleverede data og har fokus på metodisk klarhed, reproducerbarhed og faglig konsistens. Eksterne forklarende faktorer som COVID-19, konjunkturændringer og politiske tiltag inddrages ikke.

Formålet er ikke at give strategiske eller forretningsmæssige anbefalinger, men at demonstrere korrekt og velbegrunder anvendelse af klassiske forecasting-metoder. Der lægges vægt på modellering, evaluering og sammenligning af tids-seriemodeller i en struktureret analytisk ramme.

2.1.1 Anvendelse af AI-værktøj

I udarbejdelsen af denne opgave er ChatGPT (GPT-4o) anvendt som et støtte-værktøj. Værktøjet har primært været anvendt til idéudvikling, sproglig spar-ring, forbedring af formuleringers klarhed, grammatisk gennemgang samt støtte ved udformning af enkelte kodestumper. Anvendelsen har udelukkende omfat-tet 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.

- **ARIMA:** AutoRegressive Integrated Moving Average – klassisk tidsseri-
emodel med trend og autokorrelation
- **ETS:** Exponential Smoothing State Space Model – glatningsbaseret
model til tidsserier med trend og sæson
- **SNAÏVE:** Seasonal Naive – benchmarkmodel, hvor forecast er lig med
seneste sæsonværdi
- **RMSE:** Root Mean Squared Error – fejlmål hvor store afvigelser vægtes
højt
- **MAPE:** Mean Absolute Percentage Error – fejl angivet i procent
- **STL:** Seasonal-Trend decomposition using Loess – metode til dekompo-
nering af tidsserier
- **CV:** Cross-validation – validering af modeller via rullende trænings-
/testvinduer

- **Ljung-Box:** Test for autokorrelation i modelresidualer (white noise)

2.3 Struktur

Opgaven er struktureret i overensstemmelse med en klassisk tilgang til analyse af tidsserier og følger fire hovedfaser.

Først gennemføres en eksplorativ dataanalyse (EDA), hvor serierne visualiseres og undersøges for tendenser og sæsonmønstre. Dette sker ved hjælp af grafer, deskriptive statistikker og STL-dekomposition. Data transformeres ved behov. Dernæst estimeres tre modeltyper for hver serie: ARIMA, ETS og Seasonal Naive. Disse modeller valideres gennem krydsvalidering og residualanalyse. På baggrund af modelperformance foretages fremskrivninger af arbejdsløsheden for 2020 med tilhørende prædiktionsintervaller. Afslutningsvis sammenlignes modellerne og arbejdsløshedsniveauerne på tværs af regioner og køn, og der konkluderes på analysens hovedspørgsmål.

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 tidsserier.

Datasættet er udleveret i forbehandlet format som en tsibble med korrekt angivne indeks- og nøglevariabler.

```
# Indlæs datasættet  
data <- read_rds("data/Airidk_long.rds")
```

4 Eksplorativ dataanalyse (EDA)

Dette afsnit har til formål at skabe et overblik over datasættets struktur og variation forud for modelvalget. Gennem visualiseringer, deskriptive statistikker og dekomponering undersøges arbejdsløshedens udvikling på tværs af regioner og køn fra 2007 til 2019. Analysen afdækker tendenser, sæsonmønstre og forskelle i niveau og variation. STL-dekomposition anvendes til at isolere trend, sæson og residualer, mens datatransformation vurderes for at sikre modellérbarhed. Resultaterne danner det metodiske grundlag for den efterfølgende modellering.

4.1 Visualisering af udvikling og mønstre

Formålet med dette afsnit er at identificere centrale mønstre og karakteristika i arbejdsløshedsdataene – herunder trends, sæsonvariationer, regionale forskelle

og forskelle mellem køn. Visualiseringerne danner grundlag for valg af passende modeller til hver tidsserie.

4.1.1 Udvikling over tid

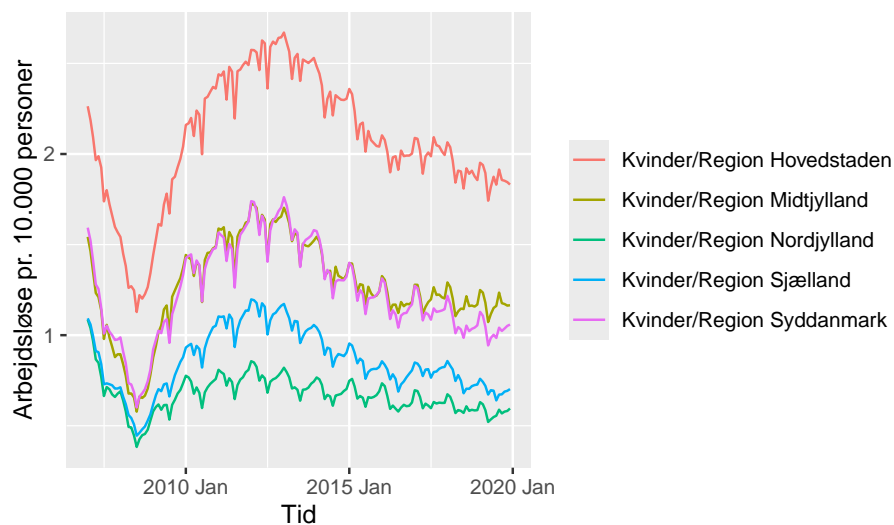
Dette afsnit har til formål at undersøge den overordnede udvikling i arbejdsløsheden over tid – opdelt på køn og region. Gennem visualiseringer vurderes generelle trends og variationer i niveau, og særlige hændelser som fx finanskrisen fremhæves.

Figur 1: Udvikling i kvinders arbejdsløshed pr. region

```
# Figur 1

data |>
  filter(kon == "Kvinder") |>
  autoplot(svalue) +
  labs(
    title = "Figur 1: Udvikling i kvinders arbejdsløshed pr. region",
    x = "Tid", y = "Arbejdsløse pr. 10.000 personer"
  ) +
  theme(legend.title = element_blank())
```

Figur 1: Udvikling i kvinders arbejdsløshed pr. region



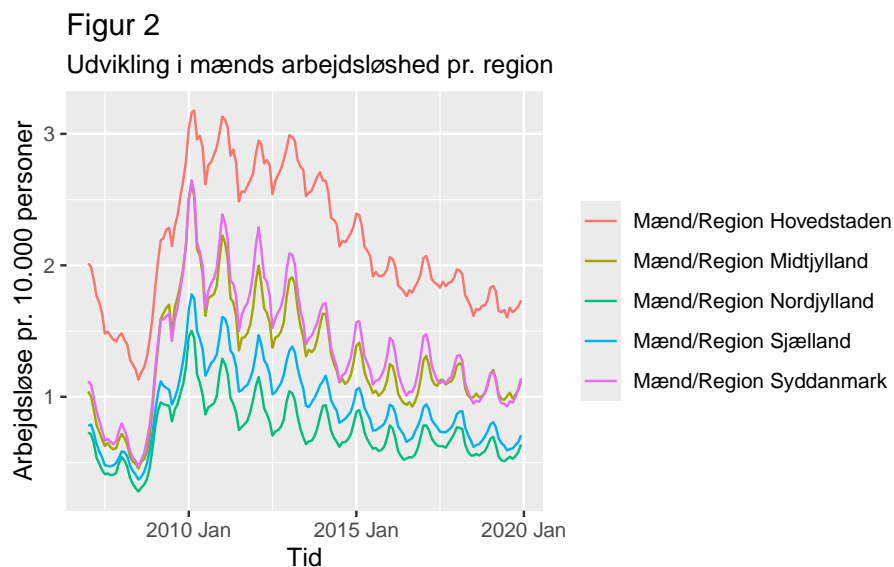
Figuren viser den månedlige arbejdsløshed blandt kvinder i de fem danske regioner fra 2007 til 2019. Der fremgår et tydeligt sæsonmønster med højere ledighed i vintermånederne og lavere i sommerperioden. Niveauet varierer regionalt, hvor

Region Hovedstaden typisk ligger højest og Nordjylland lavest. Samlet ses en faldende tendens efter finanskrisen.

Figur 2: Udvikling i mænds arbejdsløshed pr. region

```
# Figur 2

data |>
  filter(kon == "Mænd") |>
  autoplot(svalue) +
  labs(
    title = "Figur 2",
    subtitle = "Udvikling i mænds arbejdsløshed pr. region",
    x = "Tid",
    y = "Arbejdsløse pr. 10.000 personer"
  ) +
  theme(
    legend.title = element_blank()
  )
```

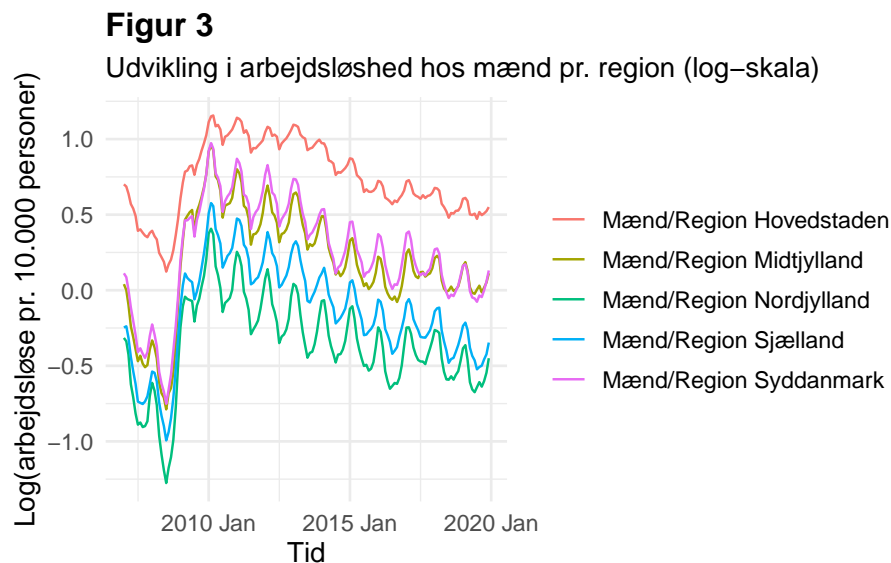


Sammenlignet med kvinder ses større udsving og et generelt højere niveau i mænds arbejdsløshed. Sæsonmønstrene er mere markante, og variationen mellem regionerne er tydeligere – især i Region Syddanmark og Midtjylland, hvor krisen i 2008-2009 medfører et kraftigt midlertidigt løft i ledigheden.

Figur 3: Udvikling i arbejdsløshed hos mænd pr. region (log-skala)


```
# Figur 3

data |>
  filter(kon == "Mænd") |>
  autoplot(log(svalue)) +
  labs(
    title = "Figur 3",
    subtitle = "Udvikling i arbejdsløshed hos mænd pr. region (log-skala)",
    x = "Tid",
    y = "Log(arbejdsløse pr. 10.000 personer)"
  ) +
  theme_minimal(base_size = 12) + # Giver et pænere layout
  theme(
    legend.title = element_blank(),      # Fjerner titel fra farvelegenden
    plot.title = element_text(face = "bold") # Fremhæver titlen lidt
  )
```



Da mændenes tidsserier generelt udviser højere niveauer og større variation end kvindernes, undersøges i figur 3, om en log-transformation kan give et mere stabilt og jævnt udgangspunkt for modellering. Formålet er ikke at sammenligne direkte med det oprindelige plot, men at vurdere, om transformationen reducerer skævhed og fremhæver relative ændringer på en måde, der kan være fordelagtig for videre analyse.

Ud over de langsigtede trends spiller sæsonvariation en vigtig rolle for arbejdsløsheden. Det undersøges nærmere i de følgende visualiseringer.

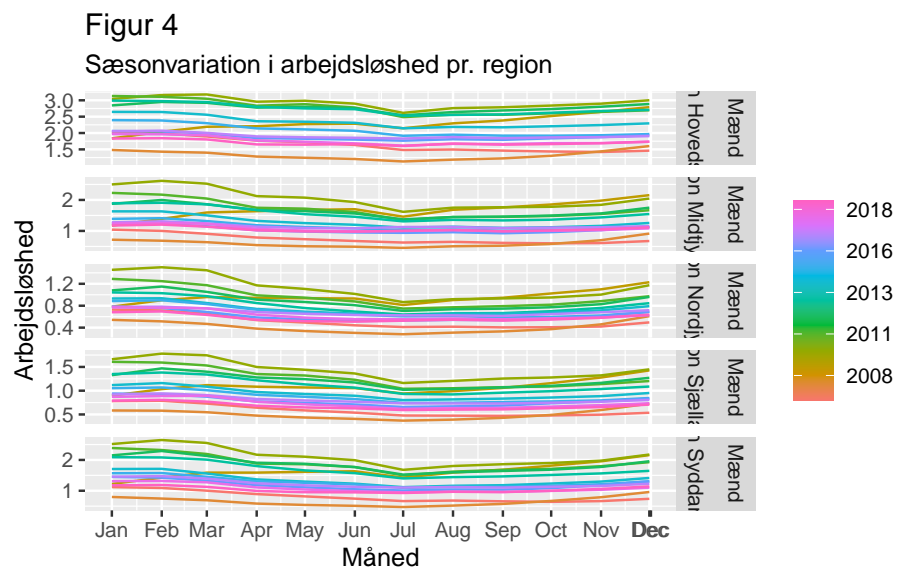
4.1.2 Sæsonmønstre

For bedre at forstå de tilbagevendende variationer i arbejdsløsheden undersøges sæsonmønstre separat for mænd og kvinder. Der anvendes både sæsonplots og subserieplots, som fremhæver gentagne mønstre over måneder og år.

Figur 4: Sæsonmønster for mænd pr. region

```
#Figur 4: Sæsonmønster for mænd pr. region
```

```
data |>
  filter(kon == "Mænd") |>
  gg_season(svalue) +
  labs(
    title = "Figur 4",
    subtitle = "Sæsonvariation i arbejdsløshed pr. region",
    x = "Måned",
    y = "Arbejdsløshed"
  )
```

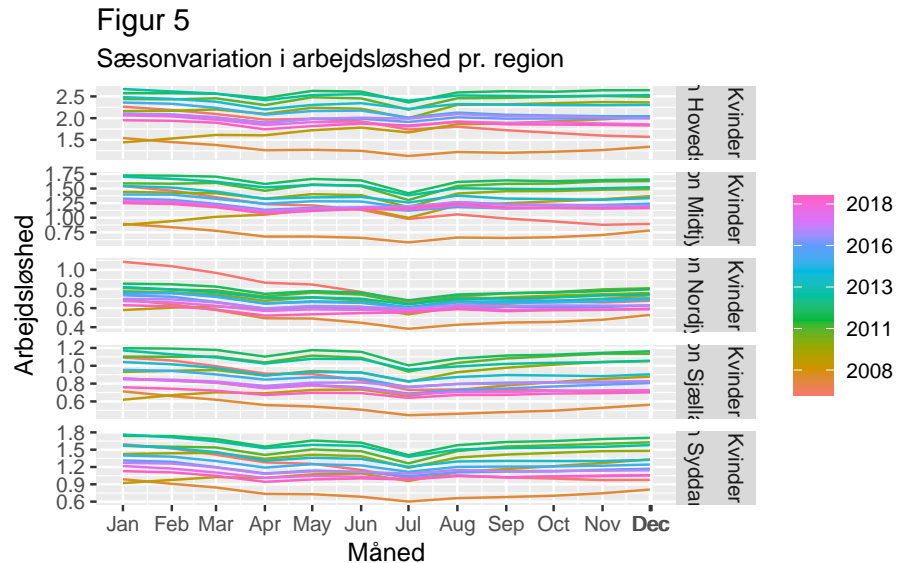


Figur 4 viser sæsonmønstre i mænds arbejdsløshed fordelt på regioner. Hver farvet linje repræsenterer et år, og linjernes forløb illustrerer, hvordan arbejdsløsheden typisk er højest i årets begyndelse (januar-marts) og lavest i sommermånederne (juni-august). Mønsteret er relativt konsistent på tværs af år og regioner, hvilket understøtter tilstedeværelsen af en stabil sæsonkomponent i mænds ledighed. De højeste niveauer ses i Region Hovedstaden, mens Region Sjælland og Region Nordjylland ligger lavere. Bemærk især det markante fald fra marts til juni, som gentager sig på tværs af årstal.

Figur 5: Sæsonmønster for kvinder pr. region

```
#Figur 5: Sæsonmønster for kvinder pr. region

data |>
  filter(kon == "Kvinder") |>
  gg_season(svalue) +
  labs(
    title = "Figur 5",
    subtitle = "Sæsonvariation i arbejdsløshed pr. region",
    x = "Måned",
    y = "Arbejdsløshed"
  )
```



Sammenlignet med mænd er sæsonudsvingene blandt kvinder mindre udtalte, og udviklingen fremstår mere jævn og stabil på tværs af årene. Niveauforskellene mellem regionerne er bevaret, men udsvingene over måneder er mere afdæmpede. Det indikerer en lavere volatilitet i kvinders arbejdsløshed og en mere ensartet sæsonprofil.

Figur 6: Subserieplot for mænd pr. måned og region

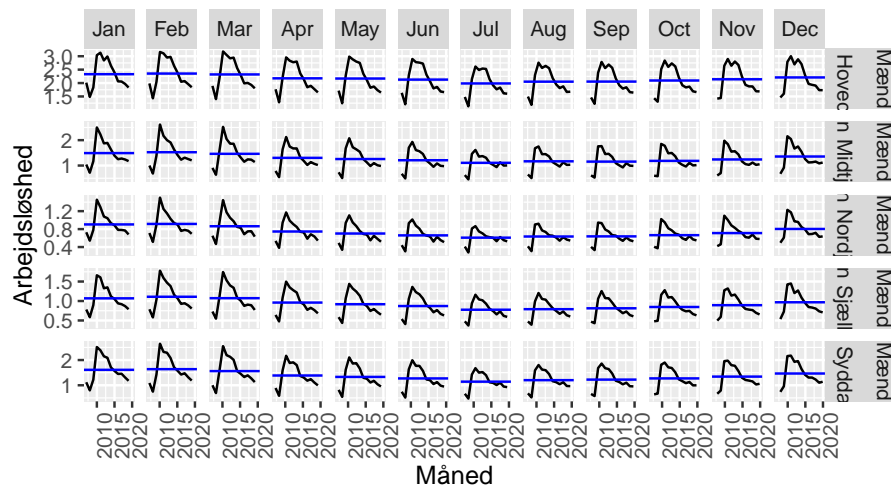
```
#Figur 6

data |>
  filter(kon == "Mænd") |>
  gg_subseries(svalue) +
```

```
labs(
  title = "Figur 6",
  subtitle = "Subserieplot for mænd pr. måned og region",
  x = "Måned",
  y = "Arbejdsløshed"
)
```

Figur 6

Subserieplot for mænd pr. måned og region



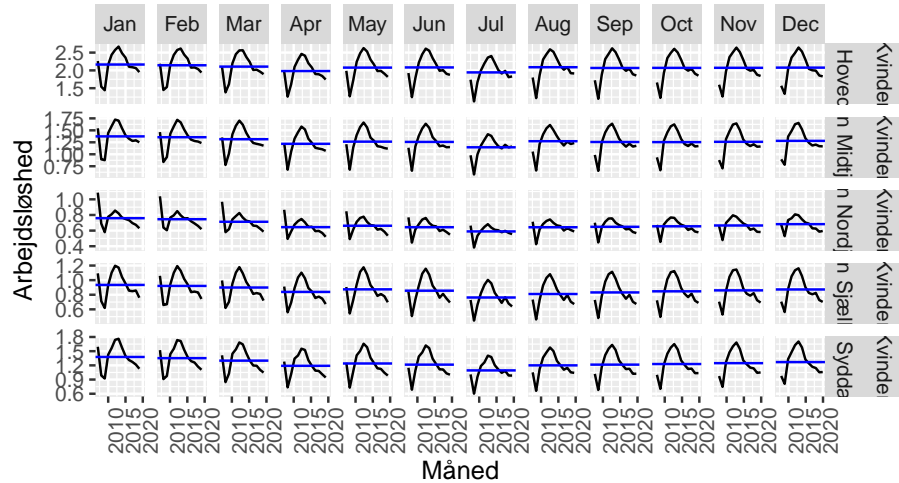
Figur 6 viser udviklingen i mænds arbejdsløshed fordelt på måneder og regioner, hvor hver facetteret celle repræsenterer en måned. Den sorte linje viser udviklingen over tid, mens den blå streg repræsenterer gennemsnittet. Det ses, at arbejdsløsheden konsekvent toppe i årets første måneder og falder hen over sommeren. Desuden er udsvingene mellem år særligt store i vinterhalvåret, hvilket understreger en høj sæsonafhængighed og volatilitet i mænds ledighedsniveau. Mønstret er gennemgående i alle regioner, om end niveauet varierer.

Figur 7: Subserieplot for kvinder pr. måned og region

```
#Figur 7
data |>
  filter(kon == "Kvinder") |>
  gg_subseries(svalue) +
  labs(
    title = "Figur 7",
    subtitle = "Subserieplot for kvinder pr. måned og region",
    x = "Måned",
    y = "Arbejdsløshed"
  )
```

Figur 7

Subserieplot for kvinder pr. måned og region



Sammenlignet med mænd viser kvinder tilsvarende sæsonmønstre med højere ledighed i årets begyndelse og lavere i sommermånederne. Udsvingene er dog mindre markante, og variationen mellem år er mindre tydelig. Mønstret fremstår mere stabilt og jævnt fordelt på tværs af regioner, hvilket tyder på lavere volatilitet i kvindernes arbejdsløshed.

Udover sæsonmæssige udsving varierer arbejdsløsheden betydeligt på tværs af både geografi og køn. Dette undersøges nærmere i de følgende visualiseringer.

4.1.3 Regionale og kønsmæssige forskelle

I dette afsnit rettes fokus mod samspillet mellem køn og region. Der undersøges, hvordan forskelle i arbejdsløshedsniveau og -mønstre varierer både mellem regioner og mellem mænd og kvinder. Visualiseringerne præsenteres i forskellige facetteringer for at fremhæve de vigtigste forskelle.

Figur 8: Arbejdsløshed fordelt på regioner – opdelt efter køn

```
#"Figur 8: Arbejdsløshed fordelt på regioner - opdelt efter køn"

data |>
  autoplot(svalue) +
  aes(colour = region) +
  facet_wrap(~ kon) +
  labs(
    title = "Figur 8",
    subtitle = "Arbejdsløshed fordelt på regioner - opdelt efter køn",
```

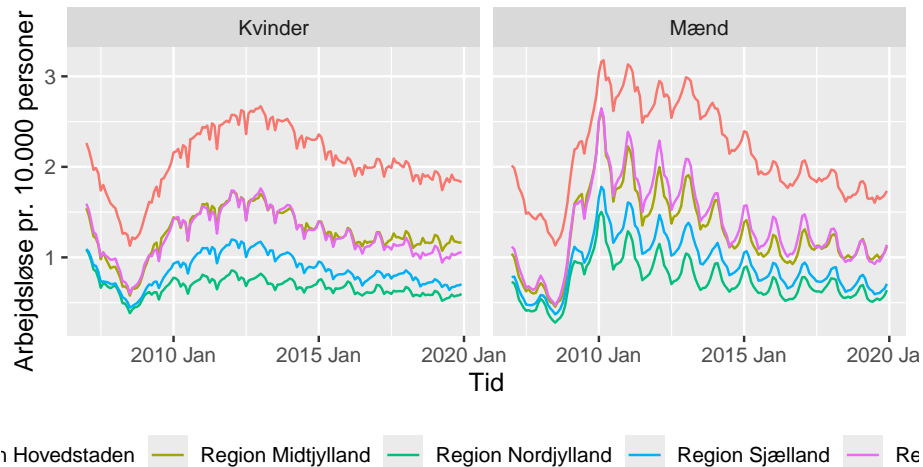
```

y = "Arbejdsløse pr. 10.000 personer", x = "Tid"
) +
theme(legend.position = "bottom", legend.title = element_blank())

```

Figur 8

Arbejdsløshed fordelt på regioner – opdelt efter køn



Figur 8 viser udviklingen i arbejdsløsheden for mænd og kvinder i separate paneller, hvor farverne angiver de fem danske regioner. Denne visning gør det muligt at sammenligne geografiske forskelle inden for hvert køn. Region Hovedstaden har konsekvent det højeste ledighedsniveau, mens Nordjylland ligger lavest. Blandt mænd ses både højere niveauer og større udsving, særligt under og efter finanskrisen. Kvindernes udvikling er generelt mere stabil og glidende. Bemærk, at forskellene mellem regionerne fremstår tydeligere blandt kvinder, hvilket kan tyde på mere vedvarende strukturelle forskelle.

Figur 9: Arbejdsløshed fordelt på køn – opdelt efter region

Figur 9: Arbejdsløshed fordelt på køn – opdelt efter region

```

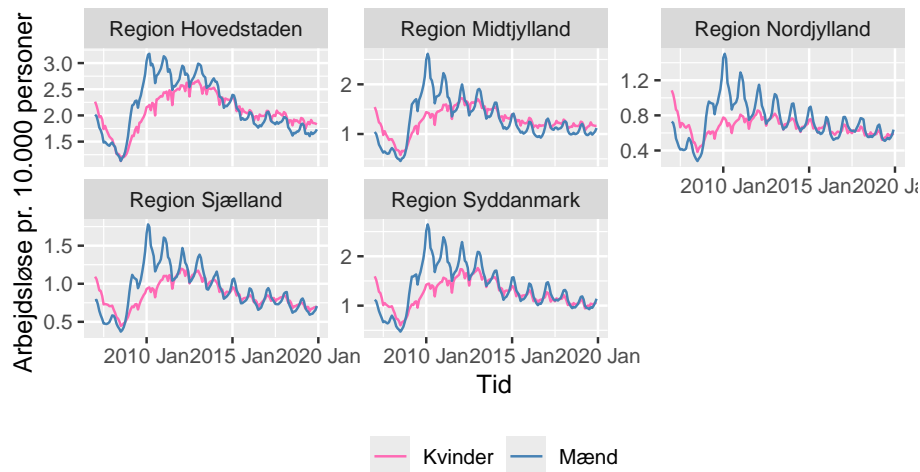
data |>
  autoplot(svalue) +
  aes(colour = kon, group = kon) +
  facet_wrap(~ region, scales = "free_y") +
  labs(
    title = "Figur 9",
    subtitle = "Arbejdsløshed fordelt på køn – opdelt efter region",
    y = "Arbejdsløse pr. 10.000 personer", x = "Tid",
    colour = "Køn"
  ) +

```

```
scale_colour_manual(
  values = c("Mænd" = "steelblue", "Kvinder" = "hotpink")
) +
theme(legend.position = "bottom", legend.title = element_blank())
```

Figur 9

Arbejdsløshed fordelt på køn – opdelt efter region



Hvor Figur 8 fokuserede på at sammenligne regioner inden for hvert køn, giver Figur 9 et omvendt perspektiv: her sammenlignes køn i hver af de fem danske regioner. Mænd har generelt højere ledighed og mere markante udsving end kvinder – særligt i perioden omkring finanskrisen. Forskellene mellem kønnene er relativt ensartede på tværs af regioner, men mønstret bekræfter, at mænds ledighed både topper højere og varierer mere over tid. Kvinders ledighed er lavere og følger en mere stabil udvikling. Figuren understøtter dermed beslutningen om at analysere serierne individuelt.

Samlet set viser figurene, at arbejdsløsheden varierer tydeligt mellem regioner og køn. Denne variation understøtter behovet for at behandle hver tidsserie separat i den videre analyse. Som næste skridt anvendes deskriptive statistikker til at opsummere centrale forskelle i niveau og udsving mellem serierne.

4.2 Deskriptive statistikker

For at supplere graferne anvendes deskriptive statistikker som gennemsnit, spredning, minimum og maksimum. De opsummerer niveau og variation i arbejdsløsheden pr. region og køn (Hyndman & Athanasopoulos, 2021). Statistikkerne giver et hurtigt overblik over forskelle mellem grupper og hjælper

med at identificere serier med høj volatilitet. Det bidrager til at vurdere, hvor der kan være behov for særlig opmærksomhed i den videre modellering.

```
# Beregning
deskriptiv <- data |>
  as_tibble() |>
  group_by(region, kon) |>
  summarise(
    Gennemsnit = mean(svalue, na.rm = TRUE),
    `Standardafvigelse` = sd(svalue, na.rm = TRUE),
    Minimum = min(svalue, na.rm = TRUE),
    Maksimum = max(svalue, na.rm = TRUE),
    .groups = "drop"
  )

# Visning
deskriptiv |>
  kable(format = "markdown", digits = 2, caption = "Tabel 1: Deskriptiv statistik per region",
  kable_styling(full_width = FALSE, position = "center", bootstrap_options = c("striped", "border"),
  row_spec(0, bold = TRUE, background = "#f0f0f0"))
```

Tabel 1: Tabel 1: Deskriptiv statistik per region og køn

region	kon	Gennemsnit	Standardafvigelse	Minimum	Maksimum
Region Hovedstaden	Kvinder	2.07	0.36	1.13	2.67
Region Hovedstaden	Mænd	2.17	0.52	1.13	3.18
Region Midtjylland	Kvinder	1.27	0.25	0.58	1.72
Region Midtjylland	Mænd	1.29	0.44	0.45	2.62
Region Nordjylland	Kvinder	0.67	0.10	0.38	1.09
Region Nordjylland	Mænd	0.74	0.23	0.28	1.50
Region Sjælland	Kvinder	0.86	0.17	0.45	1.20
Region Sjælland	Mænd	0.92	0.30	0.37	1.78
Region Syddanmark	Kvinder	1.24	0.26	0.60	1.76
Region Syddanmark	Mænd	1.37	0.47	0.47	2.65

Tabel 1 opsummerer arbejdsløsheden på tværs af regioner og køn. Gennemsnittet viser det generelle niveau, mens standardafvigelsen giver et mål for variation over tid. Mænd har generelt højere gennemsnit og større udsving end kvinder, særligt i Region Hovedstaden og Syddanmark. Omvendt ses lavere og mere stabile niveauer i Nordjylland, især blandt kvinder. Minimums- og maksimumsværdier understøtter indtrykket af høj volatilitet i visse serier – hvilket kan få betydning for modelvalget i næste fase.

4.3 STL-dekomposition

For at opnå en bedre forståelse af de strukturelle komponenter i arbejdsløshedsserierne anvendes STL-dekomposition (Seasonal-Trend decomposition using Lo-ess). Metoden adskiller tidsserier i tre komponenter: trend, sæson og remainder (Hyndman & Athanasopoulos, 2021). Det giver indblik i, hvor meget af variationen der kan forklares af langsigtede bevægelser, gentagne sæsonmønstre eller kortsigtede, uregelmæssige udsving.

Da særligt mændenes serier udviser høje niveauer og større variation, er alle serier log-transformeret forud for dekompositionen. Denne transformation stabiliserer variansen og gør det lettere at sammenligne komponenterne på tværs af regioner og køn.

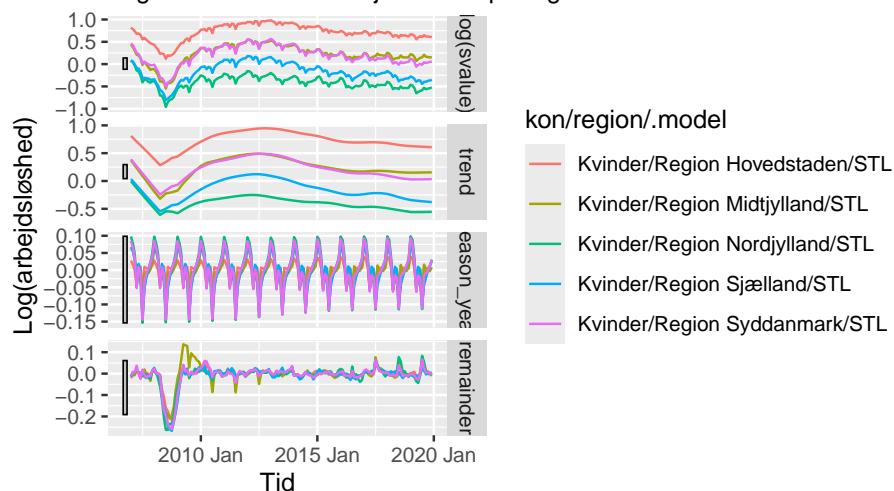
STL anvendes her på de log-transformerede serier for henholdsvis kvinder og mænd i hver region. Resultaterne præsenteres i figur 10 og 11.

```
kvinder_STL <- data |>
  filter(kon == "Kvinder") |>
  model(STL = STL(log(svalue), robust = TRUE))

kvinder_STL |>
  components() |>
  autoplot() +
  labs(
    title = "Figur 10: STL-dekomposition for kvinder",
    subtitle = "Log-transformeret arbejdsløshed pr. region",
    x = "Tid",
    y = "Log(arbejdsløshed)"
  )
```

Figur 10: STL-dekomposition for kvinder

Log-transformeret arbejdsløshed pr. region



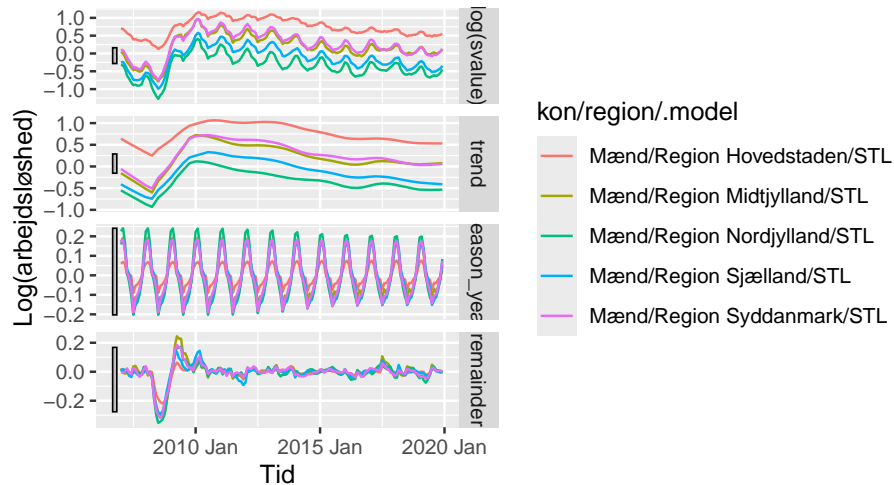
Figur 10: STL-dekomposition for kvinder – log-transformeret arbejdsløshed pr. region Figur 10 viser STL-dekompositionen for kvinder i de fem danske regioner. Der ses tydelige sæsonmønstre med tilbagevendende lavpunkter i sommermånederne og højere ledighed i vinterhalvåret. Trends varierer mellem regionerne: Region Hovedstaden udviser generelt højere ledighed, mens Nordjylland og Sjælland ligger lavere. Residualerne ligger stabilt omkring nul, hvilket indikerer, at modellen formår at fange de dominerende strukturer i data.

```
mænd_STL <- data |>
  filter(kon == "Mænd") |>
  model(STL = STL(log(svalue), robust = TRUE))

mænd_STL |>
  components() |>
  autoplot() +
  labs(
    title = "Figur 11: STL-dekomposition for mænd",
    subtitle = "Log-transformeret arbejdsløshed pr. region",
    x = "Tid",
    y = "Log(arbejdsløshed)"
  )
```

Figur 11: STL-dekomposition for mænd

Log-transformeret arbejdsløshed pr. region



Figur 11: STL-dekomposition for mænd – log-transformeret arbejdsløshed pr. region Figur 11 viser tilsvarende dekomposition for mænd. Her ses en tilsvarende stærk sæsonkomponent, men med mere markante udsving og højere ledighedsniveauer – især efter finanskrisen. Trends følger samme overordnede forløb som for kvinder, men afvigelserne er større. Dette bekræfter tidligere observationer om højere volatilitet i mænds ledighed. Residualkomponenten er mere varierende, hvilket kan indikere, at nogle udsving ikke fanges fuldt ud af modellen.

Disse observationer fra STL-dekompositionen bekræfter, at både trend og sæsonmønstre varierer betydeligt på tværs af serierne. Det understreger behovet for modeller, der kan tilpasses hver series struktur, hvilket vil blive præsenteret i det følgende afsnit.

5 Modelvalg

I dette afsnit estimeres og sammenlignes tre klassiske modeltyper, ARIMA, ETS og benchmarkmodellen SNaive, for hver tidsserie (region \times køn). Hver modeltype har forskellige antagelser og fordele og kan fange forskellige karakteristika i data, såsom trend, sæson og autokorrelation. Modelvalget foretages automatisk for hver serie via fable-pakken, og resultaterne anvendes i den efterfølgende validering og forecast.

5.1 Valgte modeltyper

ARIMA (AutoRegressive Integrated Moving Average)

ARIMA-modeller anvendes til at modellere tidsserier med autokorrelation og ikke-stationaritet. De kombinerer autoregressive (AR) led, differens (I) for at opnå stationaritet, og glidende gennemsnit (MA) led til at modellere fejl. Ved at inkludere sæsonkomponenter (SARIMA) kan modellen håndtere årligt gentagende mønstre. ARIMA er særligt velegnet til serier med langsigtede trends og strukturelle skift, hvor tidligere værdier har stor indflydelse på den nuværende tilstand (jf. Hyndman & Athanasopoulos, kap. 9).

ETS (Exponential Smoothing State Space Model)

ETS-modeller benytter eksponentiel glatning til at vægte nyere observationer højere og beskriver tidsserier ud fra tre elementer: fejl, trend og sæson. Hver komponent kan være additiv eller multiplicativ afhængigt af dataens karakter. ETS egner sig godt til serier med tydelige strukturer og sæsonrytmer, især når der ikke er behov for at modellere autokorrelation direkte (jf. kap. 8).

SNAÏVE (Seasonal Naive)

SNaive er en enkel benchmarkmodel, der baserer hvert forecast på den tilsvarende værdi fra samme sæson i den foregående periode. Modellen fanger sæsonmønstre, men tager ikke højde for trend eller afhængighed i data. Den anvendes til at vurdere, om mere komplekse modeller giver en mærkbart bedre forecast-præcision (jf. kap. 3 og 8).

5.2 Modellering i praksis

Ved hjælp af fable-pakken estimeres de tre modeller automatisk for hver af de ti tidsserier på baggrund af data fra 2007 til 2019. Herefter genereres forecasts for 2020 med tilhørende prædiktionsintervaller. Prognoserne visualiseres, så man kan sammenligne modellernes adfærd på tværs af regioner og køn.

```
# Estimér modeller for hele perioden 2007-2019
models <- data |>
  model(
    ARIMA = ARIMA(svalue),
    ETS    = ETS(svalue),
    SNaive = SNAIVE(svalue)
  )

# Udtræk modelinformation med glance()
# Brug TIDY til at få modelstruktur
modelinfo <- models |>
  glance() |>
  select(kon, region, .model, AICc)
```

```
# Vis tabellen
modelinfo |>
  arrange(kon, region) |>
  kable(format = "markdown",
        caption = "Tabel 2: Automatisk valgte modelstrukturer for hver serie (region × køn)",
        align = "l") |>
  kable_styling(full_width = FALSE, position = "center", bootstrap_options = c("striped", "r"),
  row_spec(0, bold = TRUE, background = "#f0f0f0"))
```

Tabel 2: Tabel 2: Automatisk valgte modelstrukturer for hver serie (region × køn)

kon	region	.model	AICc
Kvinder	Region Hovedstaden	ARIMA	-519.59277
Kvinder	Region Hovedstaden	ETS	-106.04957
Kvinder	Region Hovedstaden	SNaive	NA
Kvinder	Region Midtjylland	ARIMA	-559.62392
Kvinder	Region Midtjylland	ETS	-186.73947
Kvinder	Region Midtjylland	SNaive	NA
Kvinder	Region Nordjylland	ARIMA	-794.48456
Kvinder	Region Nordjylland	ETS	-420.44010
Kvinder	Region Nordjylland	SNaive	NA
Kvinder	Region Sjælland	ARIMA	-714.46022
Kvinder	Region Sjælland	ETS	-355.53652
Kvinder	Region Sjælland	SNaive	NA
Kvinder	Region Syddanmark	ARIMA	-615.90506
Kvinder	Region Syddanmark	ETS	-194.25088
Kvinder	Region Syddanmark	SNaive	NA
Mænd	Region Hovedstaden	ARIMA	-451.76084
Mænd	Region Hovedstaden	ETS	-94.42293
Mænd	Region Hovedstaden	SNaive	NA
Mænd	Region Midtjylland	ARIMA	-365.22171
Mænd	Region Midtjylland	ETS	24.67626
Mænd	Region Midtjylland	SNaive	NA
Mænd	Region Nordjylland	ARIMA	-536.74821
Mænd	Region Nordjylland	ETS	-202.87509
Mænd	Region Nordjylland	SNaive	NA
Mænd	Region Sjælland	ARIMA	-524.52351
Mænd	Region Sjælland	ETS	-271.66140
Mænd	Region Sjælland	SNaive	NA
Mænd	Region Syddanmark	ARIMA	-387.29279

Mænd	Region Syddanmark	ETS	-55.07382
Mænd	Region Syddanmark	SNaive	NA

Tabel 2 viser de modelstrukturer og AICc-værdier, som fable automatisk har estimeret for hver tidsserie. Modelvalg inden for ARIMA og ETS sker med udgangspunkt i AICc, som balancerer modeltilpasning og kompleksitet og forebygger overfitting, særligt ved kortere tidsrækker. ARIMA-modellerne inkluderer ofte sæsonled og har i flere tilfælde høje AR- eller MA-ordener, hvilket peger på både sæsonvariation og autokorrelation i data. ETS-modellerne spænder over forskellige kombinationer af additive og multiplicative komponenter afhængigt af serien.

SNaive-modellen er medtaget som benchmark og har ingen tilhørende AICc-værdier, da den ikke bygger på en statistisk sandsynlighedsstruktur. Den anvendes i stedet som reference i den efterfølgende forecast-validering.

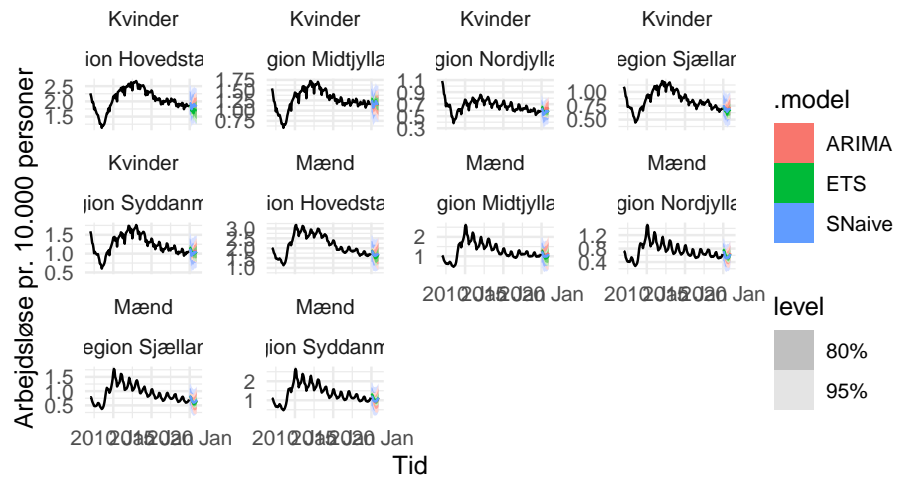
At modelvalgene varierer på tværs af regioner og køn bekræfter behovet for en individualiseret tilgang og underbygger den differentierede modelstrategi, der anvendes i analysen.

```
# Forecast for 12 måneder frem
fc <- models |> forecast(h = "12 months")

# Plot forecasts pr. serie
fc |>
  autoplot(data) +
  labs(
    title = "Figur 12",
    subtitle = "Forecasts for 2020 - pr. region og køn",
    y = "Arbejdsløse pr. 10.000 personer",
    x = "Tid"
  ) +
  facet_wrap(~ kon + region, scales = "free_y") +
  theme_minimal()
```

Figur 12

Forecasts for 2020 – pr. region og køn



Figur 12 viser forecasts og historiske data for hver kombination af region og køn. Visualiseringen giver et overblik over, hvordan de tre modeller tilpasser sig de forskellige seriers niveau og variation. Dette danner udgangspunkt for den efterfølgende vurdering af modellernes præcision.

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

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice* (3rd ed.). OTexts. <https://otexts.com/fpp3/>

11 Bilagsoversigt

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