



Praktisk tidsrækkeanalyse

Eksamensnummer: 205

Vintereksamen 2023 - Ordinær

Submitted on: 03/12/2023

ECTS: 7.5

Pages: 42



Opgave 1

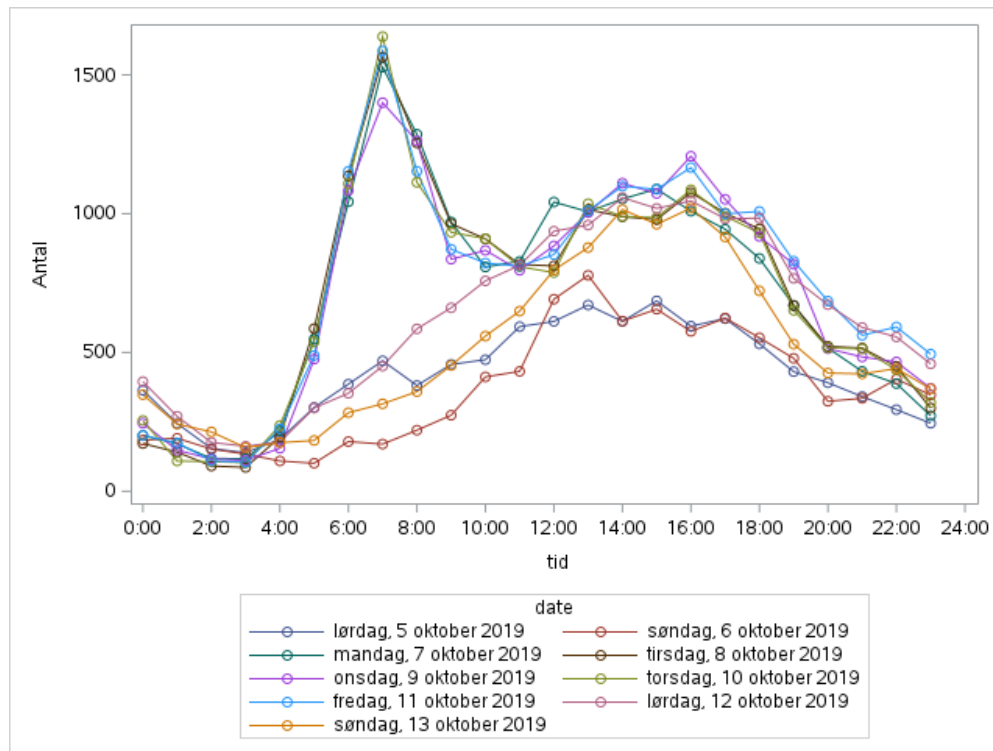
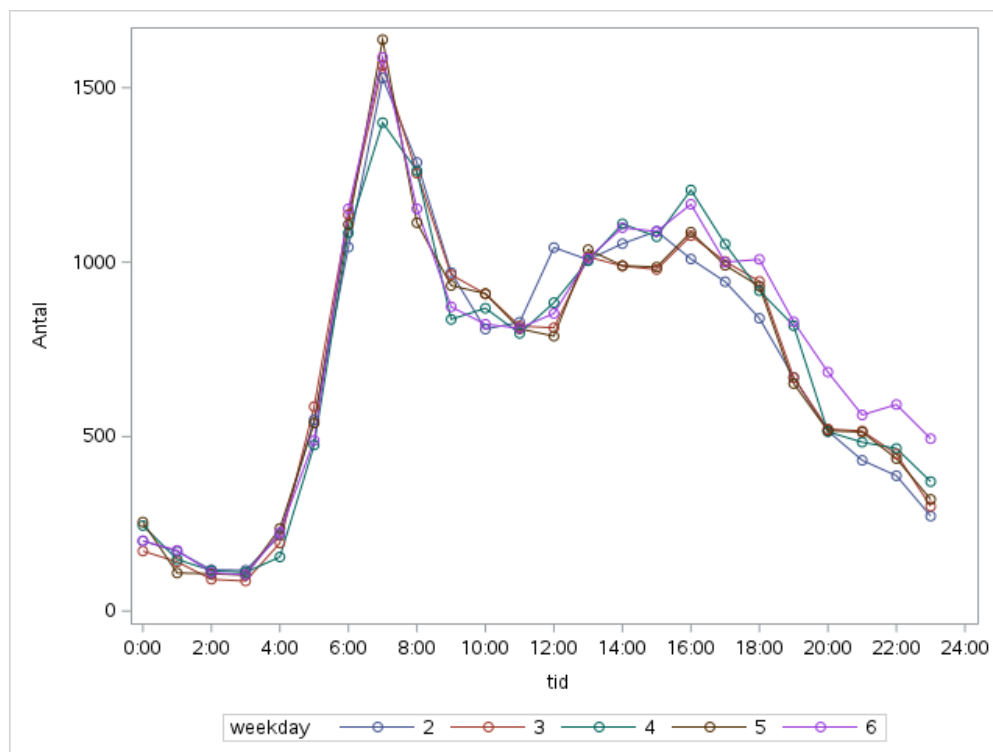
Spørgsmål (a)

Jeg starter med at plotte min tidsserie for antallet af passerede biler for alle dage, fra adressen "West 220 Street" til "9 Avenue" i Figur 1. Jeg observerer, at der generelt er få biler, der kører fra West 220 Street til 9 Avenue klokken 0.00 til 4.00. Fra kl. 4.00 til 7.00 ser jeg en stor stigning i antallet af passerende biler, hvilket kan skyldes, at folk skal ind på arbejde. Denne stigning er størst i hverdagene, som er vist i Figur 2, hvorimod lørdag og søndag ikke har en stigning fra klokken 4.00 til 7.00. Dette skyldes nok, at folk ikke arbejder lige så meget i weekender som i hverdagene. Fra klokken 7.00 til 11.00 ser jeg, at antallet af biler falder lidt i hverdagene, hvilket nok skyldes, at folk allerede er kommet ind til arbejdet, fordi de møder klokken 7 eller 8. Fra klokken 11 til 17 er antallet af biler relativt stabilt med en lille stigning, hvilket kan skyldes, at der også er mange der arbejder på restauranter og detail butikker, som typisk først åbner ved 17-18 tiden. Fra klokken 19.00 til 3.00 om natten ser jeg, at antallet af biler falder markant, hvilket skyldes at der ikke kører mange biler til 9 Avenue på dette tidspunkt. Det kan skyldes, at folk er på vej hjem fra arbejde, og at der altså ikke er så mange, som bor på 9 avenue, hvorfor jeg ikke ser mange biler køre til 9 Avenue på dette tidspunkt.

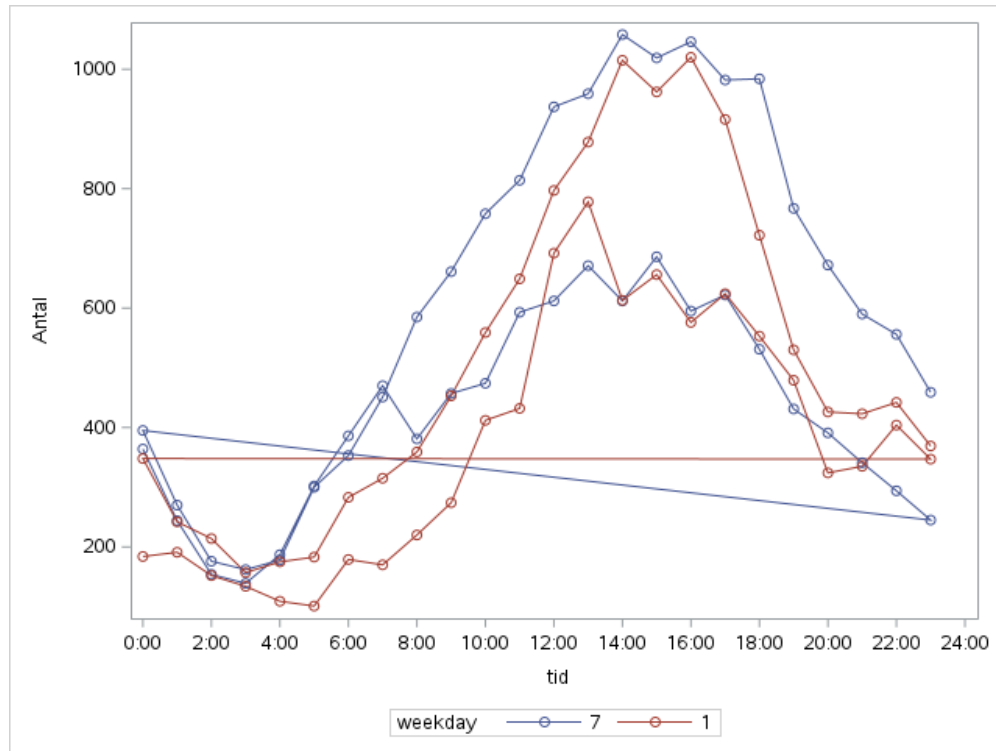
I weekenden, som jeg har illustreret i Figur 3, ser jeg, at antallet af biler ikke stiger så kraftigt om morgenen fra klokken 4.00 til 7.00, som jeg så i hverdagene. Desuden ser jeg, at antallet af biler til 9 Avenue er størst omkring klokken 14 til 19, hvilket kan skyldes, at folk tager til 9 Avenue for at shoppe og gå på café. Fra klokken 19.00 til 0.00 falder antallet af biler, som kører til 9 Avenue, hvilket blot skyldes at butikkerne lukker, og dermed vil folk gerne hjem. Når jeg kigger på y-aksen, er der generelt færre biler i weekenden end i hverdagene.

Spørgsmål (b)

Jeg starter med af anvende PROC ESM på det oprindelige datasæt for antallet af timer, hvor jeg forudsiger 168 timer frem, hvilket svarer til 7 døgn. Jeg vælger at benytte metoden "addseasonal", eftersom at denne metode anvender en additiv sæsonkomponent til modellen. Ud fra graferne visualiseret i opgave (a), antager jeg, at sæsonvariationen er konstant over tid, uafhængigt af

Figure 1: Antal biler alle dage**Figure 2:** Antal biler i hverdagene

niveauet af tidsserien. Jeg så en tydeligt sæson struktur, hvor der var flere biler om dagen end om natten, og færre biler i weekenden end i hverdagene.

Figure 3: Antal biler i weekender

Den grundlæggende formel for opdatering af niveauet i PROC ESM er

$$\bar{X}_t = \alpha \bar{X}_{t-1} + (1 - \alpha) X_t$$

hvor $\bar{X}_t$ er det estimerede niveau og α er en glatningskonstant, $0 < \alpha < 1$, (Milhoj, 2013).

Ved at benytte ovenstående formel og inkludere en sæson komponent, opnår jeg følgende formel:

$$\bar{X}_t = \alpha(y_t - \bar{S}_{t-m}) + (1 - \alpha)\bar{X}_{t-1}$$

Hvor y_t er den observerede værdi på tidspunkt t og $\bar{S}_{t-m}$ er sæson effekten for sæson tidspunktet $t - m$, hvor m er antallet af sæsoner i den pågældende periode.

$\bar{S}_t$ er givet ved følgende formel:

$$\bar{S}_t = \omega(y_t - \bar{X}_t) + (1 - \omega)\bar{S}_{t-m}$$

Hvor ω er sæson vægten, (InformIT, 2015)

Det første jeg observerer er parameter estimerne i Figur 4. Den fortæller, at niveauvægten estimeret er 0.999, hvilket betyder, at modellen lægger næsten al vægt på den seneste observation, når den forudsiger niveauet af tidsserien. Desuden ser jeg, at parameteren er statistisk signifikant med en meget lav p-værdi. Dermed har niveauet af min tidsserie en stærk påvirkning på forudsigelsen.

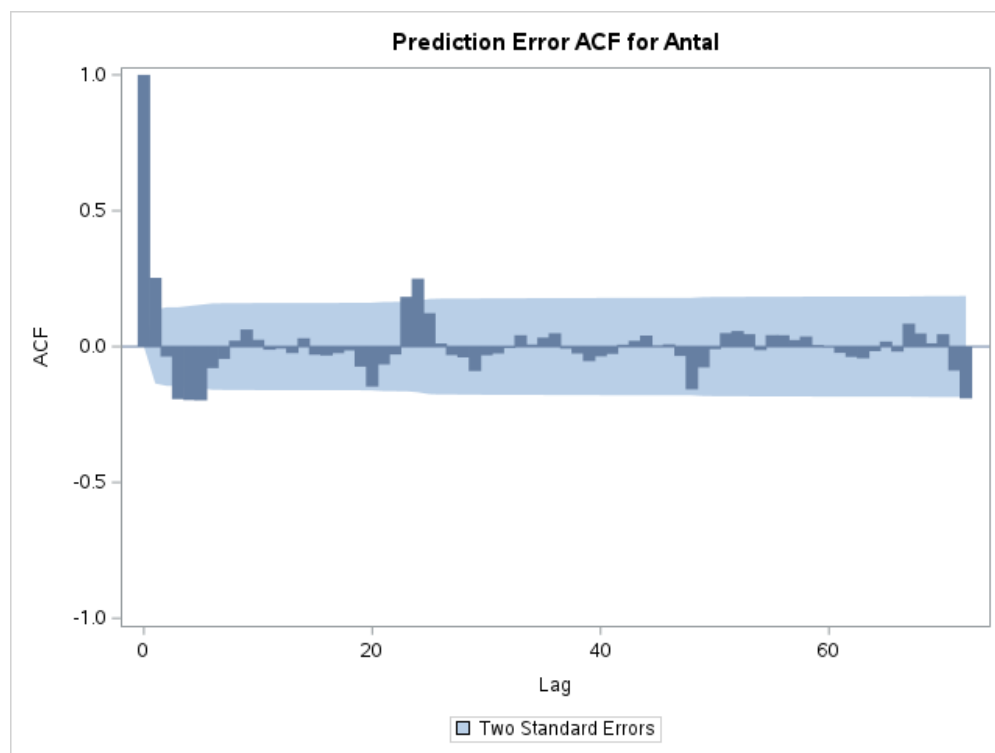
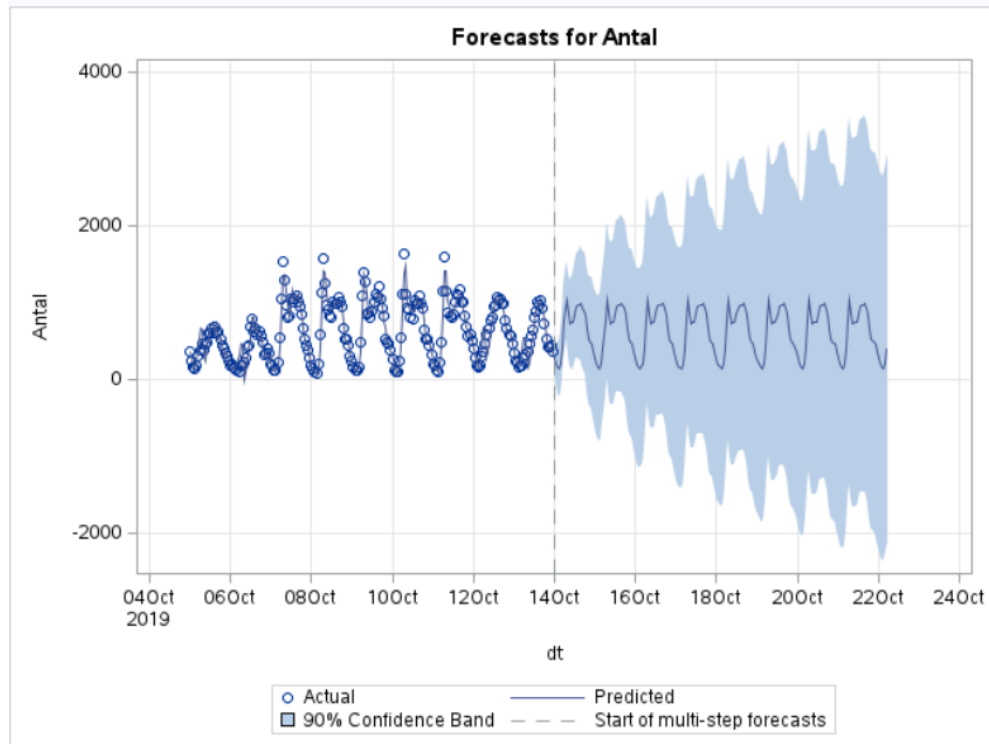
Omvendt for sæsonvægten ser jeg, at sæsonvægt estimeret er 0.001, hvilket betyder, at modellen giver lidt vægt til de nyeste sæsonmæssige ændringer, når den opdaterer den sæsonmæssige komponent. Dette betyder, at modellen i høj grad stoler på den eksisterende sæsonmæssige struktur og ikke hurtigt tilpasser sig eventuelle nye sæsonmæssige mønstre eller ændringer i mit data. Derudover viser parameteren en høj standardfejl (25.19077) og en p-værdi på 1.0000. Dette indikerer, at skønnet for "Seasonal Weight" ikke er statistisk signifikant, hvilket antyder, at den sæsonmæssige komponent måske ikke har en meningsfuld indflydelse i modellen.

Figure 4: Parameter Estimates Opgave 1b

Seasonal Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.99900	0.04934	20.25	<.0001
Seasonal Weight	0.0010000	25.19077	0.00	1.0000

Jeg ser, at der er tegn på autokorrelation i modellen, som er vist i Figur 5. Jeg ser at lag 1 og 2 overgår konfidensintervallerne, hvilket også gælder for lag 4, 5 og 6, som går i modsat retning. Dette indikerer en negativ korrelation mellem den ene periode til den anden periode, hvilket kan skyldes udsvingene jeg så i dagstimerne.

Forecast plottet for antallet af biler til 9 Avenue er vist i Figur 6. Forudsigelsen ser fornuftig ud, hvilket betyder, at jeg ser de samme mønstre udfolde sig over de næste 7 døgn, hvor der er flere biler om dagen end om aftenen og natten. Dog ligner det ikke, at forudsigelsen tager højde for antallet af biler mellem hverdagene og weekender, hvor jeg havde forventet at se færre biler i weekenderne.

Figure 5: PROC ESM ACF opgave 1b**Figure 6:** Forecasts for antal 1b

Nu vælger jeg at aggregere tidsrækken til hele døgn ved at benytte PROC Timeserie, hvor jeg benytter metoden Total for at få det totale antal biler per døgn. Dette giver mig følgende obser-

vationer vist i Figur 7.

Figure 7: broadway day

Obs	date	Antal
1	05OCT2019	10182
2	06OCT2019	8939
3	07OCT2019	17096
4	08OCT2019	17191
5	09OCT2019	17398
6	10OCT2019	17108
7	11OCT2019	18088
8	12OCT2019	15132
9	13OCT2019	12437

Ved kørslen af PROC ESM når jeg har aggregeret til dage, får jeg nu følgende parameter estimater vist i Figur 8. Jeg ser nu, at niveauvægten ikke lægger lige så meget vægt på den nyeste observation, som den gjorde tidligere med data aggregeret til timer. Dog forbliver den stadig signifikant. Omvendt for sæsonvægten, ser jeg at den nu lægger al sin vægt på den nyeste observation for sæsonstrukturen for at forudsige sæsonens næste observation. Denne ændring skyldes nok, at datasættet er blevet markant mindre, hvorfor der ikke er et bredt sæsonmønster at gå ud fra.

I forhold til min metode, vælger jeg stadig at benytte addseasonal, eftersom at jeg observerer, at der er færre biler i weekenden end i hverdagene. Dog kan denne metodes anvendelighed diskuteres, fordi der ikke er meget data. Derfor er det svært at være sikker på, at der altid er færre biler i weekenden end i hverdagene, hvorfor hvis dette ikke er tilfældet, ville metoden "simpel" være en bedre metode, eftersom jeg ville vurdere, at der så ikke vil være nogen trend eller sæson.

Figure 8: Parameter Estimates for dage 1b

Seasonal Exponential Smoothing Parameter Estimates				
Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Level Weight	0.78831	0.23854	3.30	0.0130
Seasonal Weight	0.99900	1.28319	0.78	0.4618

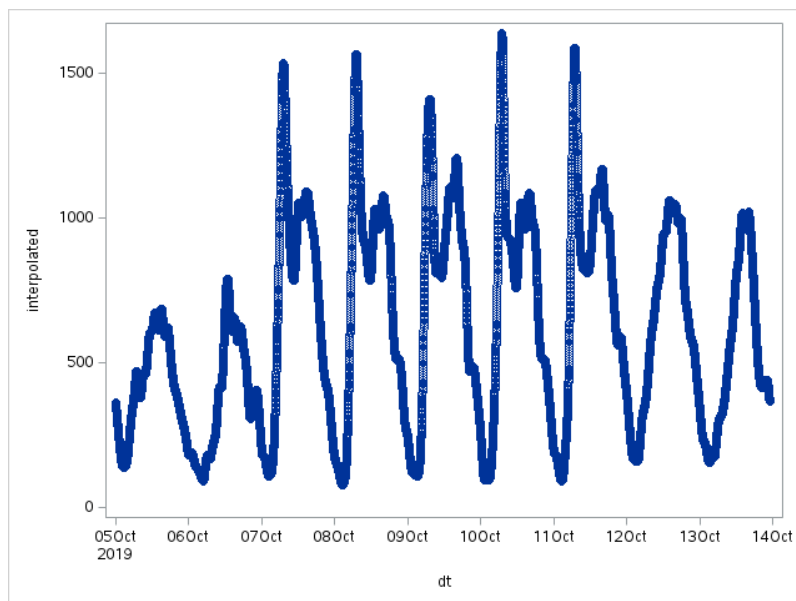
Spørgsmål (c)

PROC EXPAND interpolerer manglende observationer i tidsserier ved brug af enkle metoder til kurvetilpasning. I dette eksempel interpolerer jeg de manglende minutter i datasættet, som er vist i Figur 9. Jeg ser nu, at tidsrækken har fået mange flere observationer, men det overordnede mønster beskrevet i opgave 1 (a) forbliver ens.

Jeg benyttede metoden "spline" i RPOC Expand, som skaber en glat kurve gennem de observerede data i modsætning til join, som forbinder de observerede datapunkter med en lige linje. Spline bruger matematiske funktioner kaldet, splines, som er polynomier til at tilpasse segmenter mellem de observerede punkter.

Dette resulterer i en glat og kontinuerlig kurve, der passer godt til komplekse mønstre i data og kan håndtere store huller mellem observationer.

Figure 9: Broadway interpolerede til minutter



PROC UCM er en nyttig metode, fordi den tillader en fleksibel modellering af tidsserier ved at dekomponere dem i flere komponenter, såsom niveau (level), trend, sæsonmæssighed (seasonality), cyklus (cycle), og støj (irregular component). Hver af disse komponenter repræsenterer en underliggende struktur i mit data, som kan variere over tid.

Den basale model indeholder en niveau-komponent og et rest led.

$$y_t = \mu_t + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

Niveaukomponenten μ_t antages at blive genereret af

$$\mu_t = \mu_{t-1} + \eta_t, \eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$$

som begynder med en initial værdi μ_0 ved tidspunkt $t = 0$, hvilket er lige før den første observation. Begge restkomponenter ε_t og η_t antages at være serier af uafhængige fejl, såkaldte hvid støj-serier. Og de antages yderligere at være gensidigt uafhængige. I ligningen ovenover udtrykket restledet ε_t , en stokastisk komponenten der ikke kan blive forklaret af modellen, (Milhoj, 2013).

Eftersom at jeg tidligere observerede, at der ikke var en trend i tidserien, vælger jeg ikke at inkludere en trend komponent i modellen. Desuden vælger jeg ikke at inkludere en sæsonkomponent, eftersom at jeg i stedet inkluderer 3 forskellige cycles. Den grundlæggende model udvides nu med trigonometriske cykli den grundlæggende form af en cyklisk komponent er trigonometrisk:

$$C_t = a \cos(\lambda t) + b \sin(\lambda t) = \gamma \cos(\lambda t - \varphi)$$

Dette giver den regelmæssige form af oscillationen. Frekvensen λ angiver antallet af oscillationer pr. tidsenhed, (Milhoj, 2013). .

Efter gentagende kørsler er jeg kommet frem til følgende model. Jeg benytter en irregulær komponent, hvor jeg finder, at variansen er signifikant. Dog ser jeg, at niveau komponentens varians ikke er signifikant, hvorfor jeg vælger at sætte den varians lig 0. I den irregulær komponent vælger jeg at inkludere en 3 autoregressiv (AR) komponent. Dermed opnår jeg en AR(3) som hjælper med at forklare den nuværende værdi af min tidsserie som en funktion af sine tidligere værdier. Formlen for en AR(3) ser ud som følgende:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \alpha_3 x_{t-3} + \varepsilon_t, \quad \varepsilon_t \sim NID(0, \sigma^2), \text{ for } t = 1, 2, \dots, n. \quad (1)$$

hvor x_t repræsenterer værdien af tidsserien på tidspunkt t . $\alpha_1, \alpha_2, \alpha_3$ er koefficienterne for tidsse-

riens værdier ved tidspunkterne $t - 1$, $t - 2$, og $t - 3$ henholdsvis. ε_t er fejlleddet ved tidspunkt t , som antages at følge en normalt uafhængigt identisk fordeling (NID) med middelværdi 0 og varians σ^2 .

Jeg indsætter estimaterne fra 10 i ligning (1) og får følgende udtryk for min AR(3) model:

$$x_t = 0.99964x_{t-1} + 0.99990x_{t-2} - 0.99968x_{t-3} + \varepsilon_t \quad (2)$$

Jeg har valgt at inkludere 3 cycles med perioderne 1440 minutter, som svarer til 1 døgn. Dernæst 720 minutter som svarer til et halvt døgn, og til sidst har jeg valgt 480 minutter, som svarer til en tredjedel af et døgn, altså 8 timer.

Parameterestimaterne er vist i Figur 10, som viser at parametrene for AR(1) og AR(3) er signifikante, hvor de indeholder modsat fortegn med et estimat på 0.99. Dernæst observerer jeg, at mine cycles indeholder tre forskellige parameter, "Damping Factor", som er dæmpningsfaktoren for de cykliske komponenter. Værdierne for dem alle er tæt på 1, hvilket indikerer en relativt lav dæmpning. "Period" som viser længden af cyklussen. De meget høje værdier og P-værdier tæt på 1 indikerer, at disse perioder er statistisk insignifikante, hvilket kan tyde på, at de ikke bidrager betydeligt til modellen. Til sidst, "Error Variance" som er fejlvariansen i hver cyklisk komponent. De meget små værdier sammen med høje P-værdier (tæt på 1) indikerer, at disse varianser er statistisk insignifikante, hvorfor det kan være relevant af sætte disse varianser lig 0.

Figur 11 viser hvilke parametre, der er signifikante. Jeg ser at "Irregular" og "Level" komponenterne er vigtige og signifikante for modellen, mens de specifikke cykliske komponenter ikke bidrager signifikant til modellen.

Figur 12 viser et panel plot af proc ucm kørslen. Øverst til venstre ses et histogram, der repræsenterer fordelingen af residualer. Histogrammet sammenlignes med en normalfordelingskurve (vist med en glat orange linje), hvilket er en tætheds estimering, der viser, hvor tæt residualernes fordeling ligger på en normalfordeling. Dette bruges til visuelt at vurdere, om residualerne ser ud til at være normalfordelte. Denne figur viser, at hypotesen om normalfordelte fejl virker næsten acceptabel. Dog ser det ud til, at observationerne har en negativ skævhed.

Figure 10: Final Estimates for PROC UCM 1c

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > t
Irregular	Error Variance	0.01491	0.0001856	80.31	<.0001
Irregular	AR_1	0.99964	0.0001524	6561.06	<.0001
Irregular	AR_2	0.99990	.	.	.
Irregular	AR_3	-0.99968	0.0001523	-6563.2	<.0001
Cycle_1	Damping Factor	0.94143	0.70906	1.33	0.1843
Cycle_1	Period	1433.49732	965786.3	0.00	0.9988
Cycle_1	Error Variance	1.530888E-8	2.30084E-7	0.07	0.9470
Cycle_2	Damping Factor	0.94526	0.68604	1.38	0.1682
Cycle_2	Period	462.83986	104152.0	0.00	0.9965
Cycle_2	Error Variance	1.429152E-8	2.22286E-7	0.06	0.9487
Cycle_3	Damping Factor	0.94299	0.70160	1.34	0.1789
Cycle_3	Period	707.57654	239237.5	0.00	0.9976
Cycle_3	Error Variance	1.481584E-8	2.26336E-7	0.07	0.9478

Figure 11: Signifikant analyse

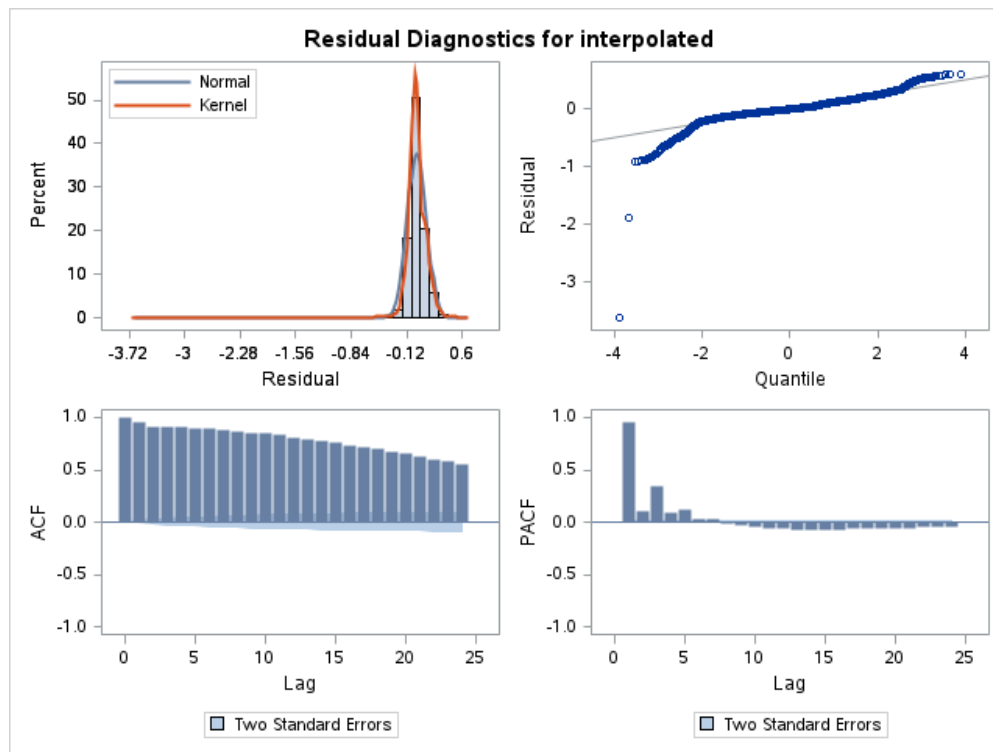
Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Irregular	1	973.54	<.0001
Level	1	5978.40	<.0001
Cycle_1	2	0.00	1.0000
Cycle_2	2	0.00	1.0000
Cycle_3	2	0.00	1.0000

Øverst til højre er et normalt Q-Q plot (kvantil-kvantil plot), som er en grafisk metode til at vurdere, om et datasæt stammer fra en normalfordeling. Punkterne repræsenterer de observerede kvantiler af residualerne mod de forventede kvantiler af en normalfordeling. Hvis punkterne falder på den linje, der repræsenterer den teoretiske fordeling, tyder det på, at residualerne er normalfordelte. I dette plot tyder det på, at datasættet er "heavy tailed", hvilket stemmer overens med histogrammet.

I figurene ACF og PACF nederst til venstre og højre, ser jeg tydelig tegn på autokorrelation. Jeg prøvede at tage hånd om dette ved at inkludere forskellige autoregressive led ind i modellen og benytte forskellige værdier for deplag. Dog var der intet der hjalp betydeligt for autokorrelation-

sproblemet, hvor jeg endte med en AR(3) komponent, som gav det pæneste resultat.

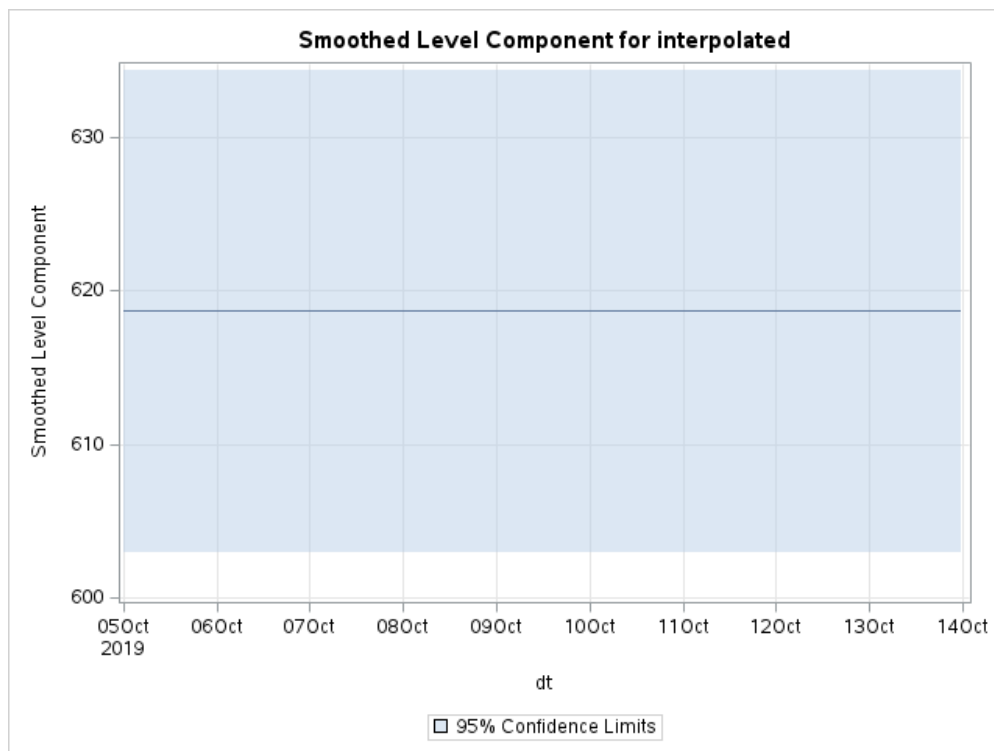
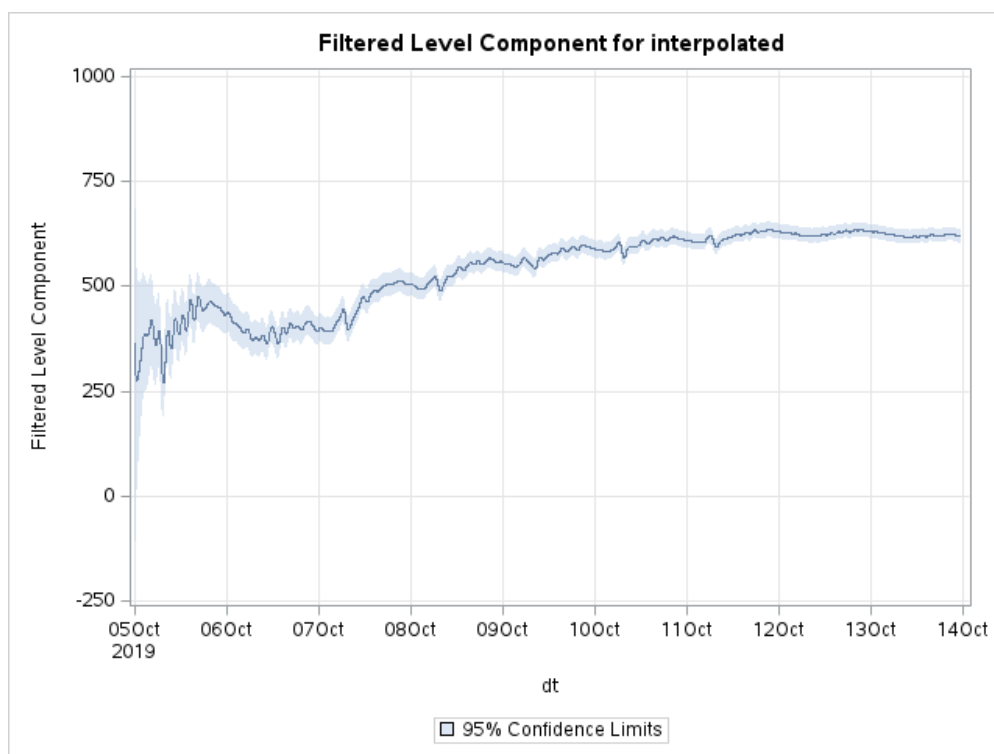
Figure 12: Residual Diagnostics



Dernæst kigger jeg på på Figur 13, som viser et billede af Smoothed Level komponenten. Figuren viser, at niveauet ligger mellem 600 og 630 biler, hvilket antyder, at det udglattede niveau ikke svinger dramatisk i den pågældende periode. Altså niveauet er relativt stabilt i denne periode. Det samme gør sig gældende for det filtrerede level plot, som viser samme relative stabilitet, dog med lidt flere udsving set i Figur 14.

Alle plots for de forskellige cycles både smoothed og filtered er alle stabile omkring 0. At disse værdi ligger omkring 0 indikerer, at der ikke er nogen stærk cyklisk effekt påvist i tidsserien i de givne intervaller nævnt tidligere. Med andre ord, der er ingen regelmæssige udsving, der overstiger det, der ville blive forventet af tilfældig variation i disse specifikke cykli.

I følge Figur 15 ser jeg, at plottet for den filtrerede irregulære komponent næsten er identisk med "smoothed" plottet. Figuren viser det tydelige sæsonmønster, hvor der er flere biler om morgenen og om eftermiddagen end om aften og natten. Desuden viser figuren, at der er færre

Figure 13: Smoothed Level Component**Figure 14: Filtered level plot 1c**

biler i weekenden d. 05-06 oktober end i hverdagene. Dermed viser figuren ikke tilfældige udsving i tidsserien, hvilket kan betyde, at der er en tydelig sæsoneffekt.

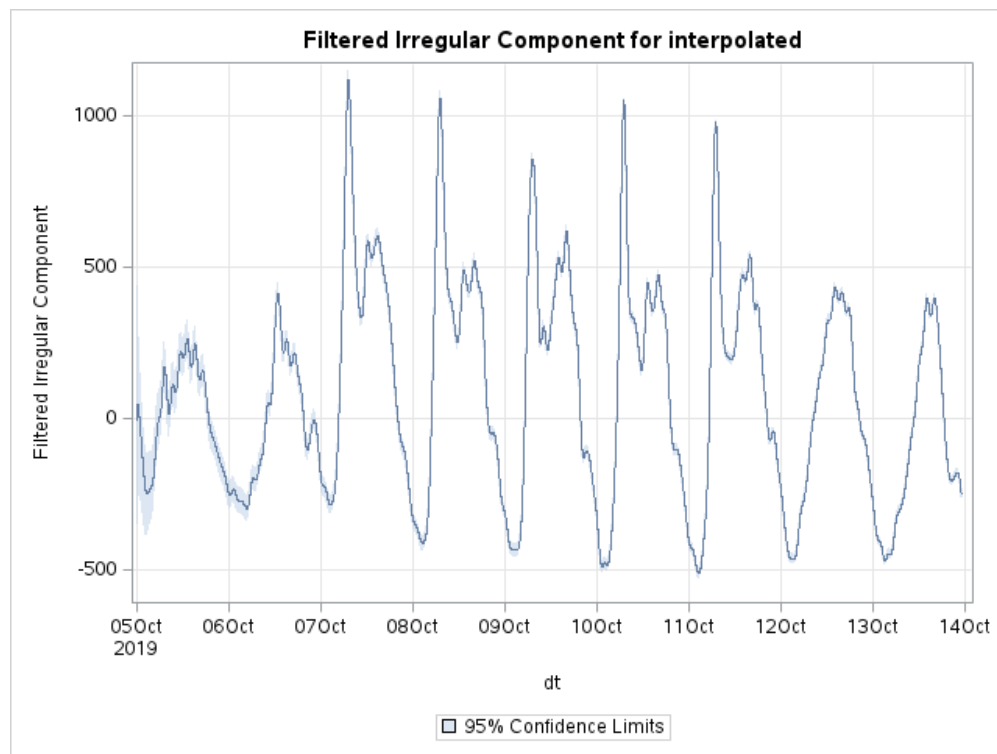


Figure 15: Filtered irregulær komponent

Opgave 2

Spørgsmål (a)

Formålet med sæsonjustering er at udlede værdien af tidsserien i en situation, hvor der ikke er sæsonmæssige påvirkninger. Sæsonmæssighed skyldes ofte meteorologiske forhold såsom temperatur og nedbør, hvilket påvirker næsten al menneskelig aktivitet. Religiøse og nationale festdage kan også påvirke en tidsserie.

I den sæsonjusterede tidsserie bør effekterne af alle sådanne begivenheder fjernes, så kun effekterne af niveauet og de specifikke begivenheder i den justerede værdi forbliver.

Jeg benytter SAS funktionen PROC X12 til at udfører en sæsonjustering ved hjælp af den originale X11-algoritme. Desuden vælger jeg, at log transformere datasættet for at stabilisere variansen, reducere skævhed og mindske potentielle outliers.

Dernæst betragt følgende formel, hvor den originale tidsserie X_t er summen af disse tre kompo-

nenter;

$$X_t = TC_t + S_t + I_t$$

Komponenten TC_t er den trend cykliske komponent, hvilket svarer til tidsseriens niveau. Denne komponent er beregnet til at omfatte både trends og konjunkturcykler. (Milhoj, 2013).

Komponenten I_t repræsenterer den irregulære komponent, der indeholder alle elementer, der udelukkende er relateret til den nuværende observation for det specifikke kvartal i serien for det pågældende år.

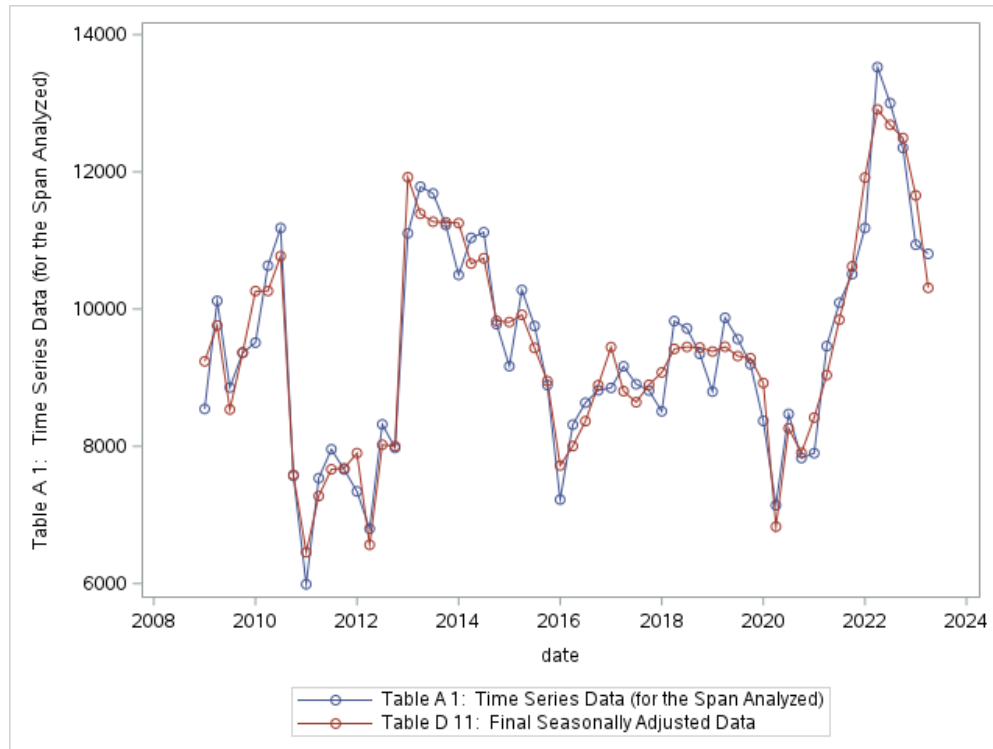
Sæsonkomponenten S_t kan betragtes som en komponent, der afspejler en specifik effekt, der forekommer i det samme kvartal hvert år i tidsserien. I denne notation indikerer t kvartaler.

Funktionen PROC X12 tillader mig blandt andet at vise følgende figurer:

- a1: den oprindelige tidsserie X_t
- d10: den sæsonmæssige komponent S_t
- d11: den justerede serie, defineret som $TC_t + I_t$
- d12: trend-cyklisk komponent TC_t
- d13: irregulær komponent I_t
- c17: vægte for den irregulær komponent I_t

Figur 16 viser, hvad der sker med tidsserien for Servicestationer, når jeg trækker sæsonkomponenten fra for at isolere og undersøge de underliggende mønstre uden sæsonmæssig påvirkning. D11 plottet sammenlignes med den originale tidsserie A1 for at se, om sæsonen har en effekt på tidsserien. Der ses en lille forskel mellem de to grafer, hvor den sæsonrensede tidsserie D11 er mere glat, specielt i perioden 2017-2020. Det tyder på, at der er en lille sæsoneffekt på tidsserien, eftersom de to grafer ikke er helt ens.

Figur 17 viser de sæsonjusterede faktorer for hver periode i tidsserien. Disse faktorer bruges til at justere rådataene for sæsonmæssige udsving, så det er muligt at se dataene uden den variation, der

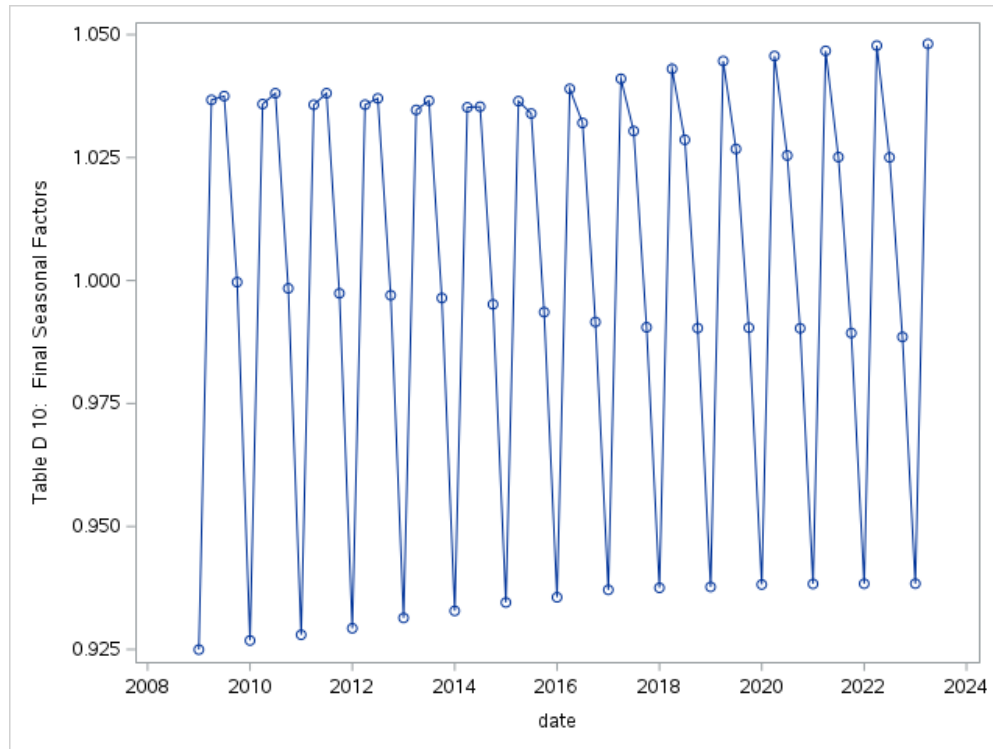
Figure 16: D11 og A1 for Servicestationer

skyldes sæsonmæssighed. Tallet 1 på y-aksen repræsenterer baseline-niveauet uden sæsonmæssig udsving. Punkterne over 1 betyder, at disse værdier har en højere værdi end gennemsnittet på grund af sæsonmæssighed, mens værdier under 1 indikerer det modsatte. Eksempelvis 2009 Q2 ligger på 1.04, hvilket betyder, at værdien i 2009 Q2 var 4% højere end forventet. Det samme mønster gør sig gældende for de andre års samme kvartal. Dog i slutningen 2017 ser jeg, at afvigelsen fra 1 bliver større for 2 kvartal, hvorimod afvigelsen fra baseline-niveauet for 3 kvartal aftager.

For 1 kvartal ser jeg, at det typisk ligger under baseline-niveauet med 7.5% i 2009 Q1, men er svagt stigende de efterfølgende år.

Alt i alt tyder der på, at der er en svag sæsoneffekt på min tidsserie, men det skal dog betragtes at skalaen på y-aksen ikke er stor.

Figur 18 viser den endelige trend-cykliske komponent af min tidsserie. Trendkomponenten repræsenterer den langsigtede bevægelse i tidsserien og giver en glat repræsentation af dataenes overordnede retning.

Figure 17: D10 for servicestationer

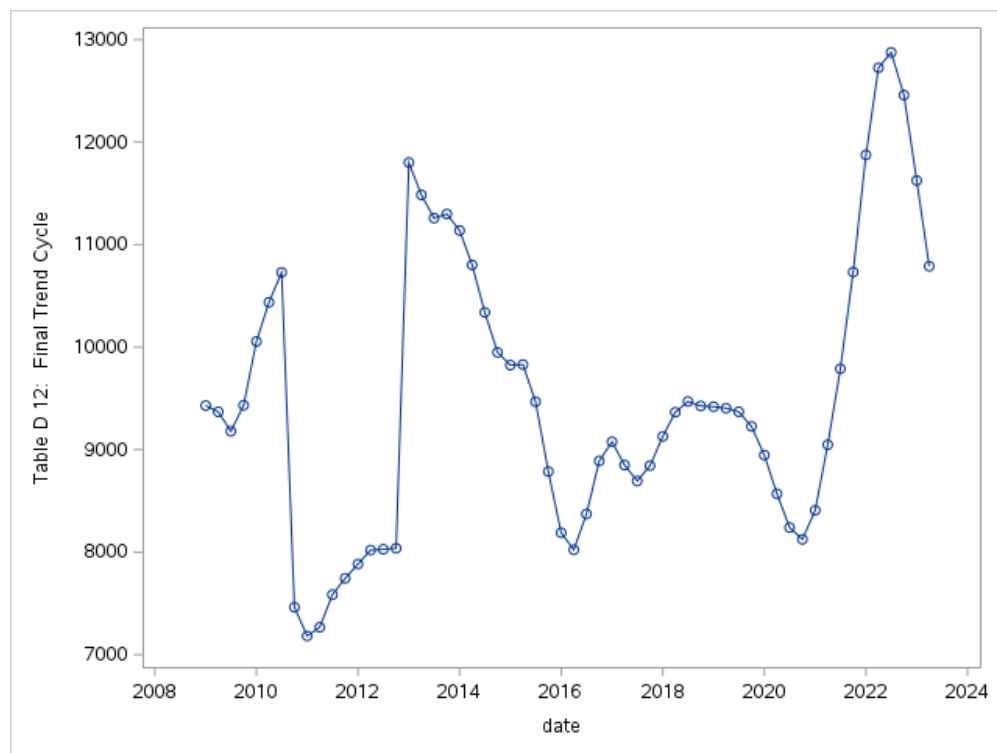
Overordnet set er salget af servicestationer både steget og faldet igennem hele perioden og overordnet set, er der ikke nogen trend. Faldet fra 2010 Q3 til 2011 Q1 er det første signifikante fald i salget. En mulig forklaring på dette kan skyldes, at effekten af finanskrisen i 2007-2008 har en forsinket påvirkning på salget af servicestationer i 2011. Krisen har haft en negativ effekt på økonomien, hvilket betyder at købekraften blev svækket, hvorfor der opleves færre salg i denne periode.

I perioden 2012 Q4 til 2013 Q1, ser jeg en betydelig stigning i antallet af salg af servicestationer. Dette kan skyldes et opsving i økonomien. Efter en økonomisk nedgang, som den der fandt sted i kølvandet på den globale finanskrisen, kan en økonomisk genopretning føre til en stigning i forbruger- og virksomhedstilliden, hvilket kan resultere i flere køb og salg af servicestationer.

En anden faktor kan være en stigning i bil salg eller ændringer i transportvaner, som har ført til højere brændstofforbrug og dermed større efterspørgsel efter servicestationstjenester.

De efterfølgende år efter 2013 ser jeg et langsomt fald af salget af servicestationer. Desuden ser jeg, at i perioden 2020, hvor COVID-19 kom til Danmark, der faldt salget af servicestationer indtil midt år 2021. Dette kan skyldes de strenge nedlukninger og rejserestriktioner for at begrænse spredningen af virussen. Disse restriktioner ville have reduceret mængden af rejser betydeligt, både lokalt og internationalt, hvilket førte til, at det ville være svært at sælge sine servicestationer. Desuden skabte pandemien økonomisk usikkerhed og sandsynligvis en tilbageholdenhed med at foretage større beslutninger, såsom salg af servicestationer.

Figure 18: D12 for servicestationer



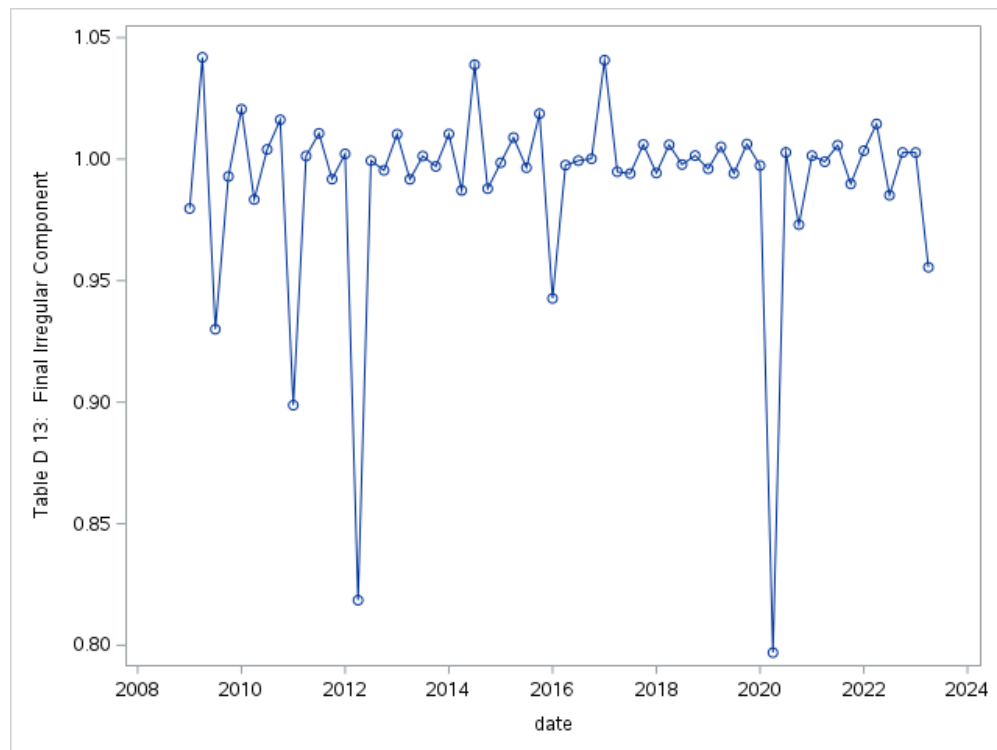
Men i perioden midt 2021 og frem ser jeg en betydelig stigning i salget af servicestationer, hvilket kan skyldes udbredelsen af COVID-19 vacciner og en bedre forståelse af, hvordan man håndterer virussen. Dette førte til et øget behov for transport og dermed en øget brug af servicestationer, hvorfor salget også steg. En anden forklaring kan være indførslen af støttepakker. Den danske regering indførte økonomiske støttepakker for at hjælpe virksomheder og stimulere økonomierne. Disse tiltag kan have forbedret likviditeten og givet virksomhederne midler til at investere i aktiver som servicestationer.

Figur 19 viser den endelige irregulære komponent af tidsserien efter at have fjernet både den sæsonmæssige og den trend-cykliske komponent. Overordnet set er den irregulære komponent relativ stabil omkring 1, men indeholder store udsving i 2012 Q1 og 2020 Q1. Disse punkter under 1, indikerer at de faktiske værdier er lavere end de forventede værdier fra modellen. Eksempelvis kan den lave værdi i 2020 Q1 skyldes COVID-19, hvorfor det kan være en potentiel outlier.

Desuden ser jeg få værdier over 1, hvilket betyder, at disse værdier er højere end modelens forventede forudsigelse.

Alt i alt ser jeg ikke en langsigtet opadgående eller nedadgående trend i den irregulære komponent, hvilket indikerer, at de tilfældige udsving ikke samler sig i en bestemt retning over tid.

Figure 19: D13 for servicestationer



Den sidste figur jeg vil forklare ses i Figur 20. Figuren viser de endelige vægte tildelt til den irregulære komponent i min tidsserie.

Observationer, der har en vægt på 1.0, er fuldt inkluderet i beregningen af den irregulære kompo-

ment. Det betyder, at disse værdier anses for at være pålidelige og ikke påvirket af outlier-effekter.

Observationer med vægte mellem 0% og 100% er kun delvist inkluderet i beregningen. En del af deres vægt kommer fra de faktiske data, mens den resterende del kommer fra udglattede værdier fra de omkringliggende observationer. Dette tyder på, at der kan være en vis grad af usikkerhed eller forstyrrelse i disse data.

Når værdierne er tæt på 0, indikerer det, at de pågældende observationer afviger betydeligt fra det forventede mønster, mere præcist mere end 2,5 standardafvigelser fra forventningen. Sådanne observationer modtager en lavere vægt i den endelige justering, hvorfor disse værdier kan være potentielle outliers. Eksempelvis kan punkterne 2009 Q4, og 2020 Q3 være potentielle outliers i følge figuren, hvor deres værdi ligger meget tæt på 0. Dette kan skyldes forrige betragtning omkring finanskrisen og corona krisen.

Værdier omkring 0.2 betyder, at den aktuelle observation kun bidrager med 20 procent, mens resterende 80 procent af justeringen afhænger af den estimerede sæsoncyklus og trendkomponent, hvilket ses i 2022Q4.

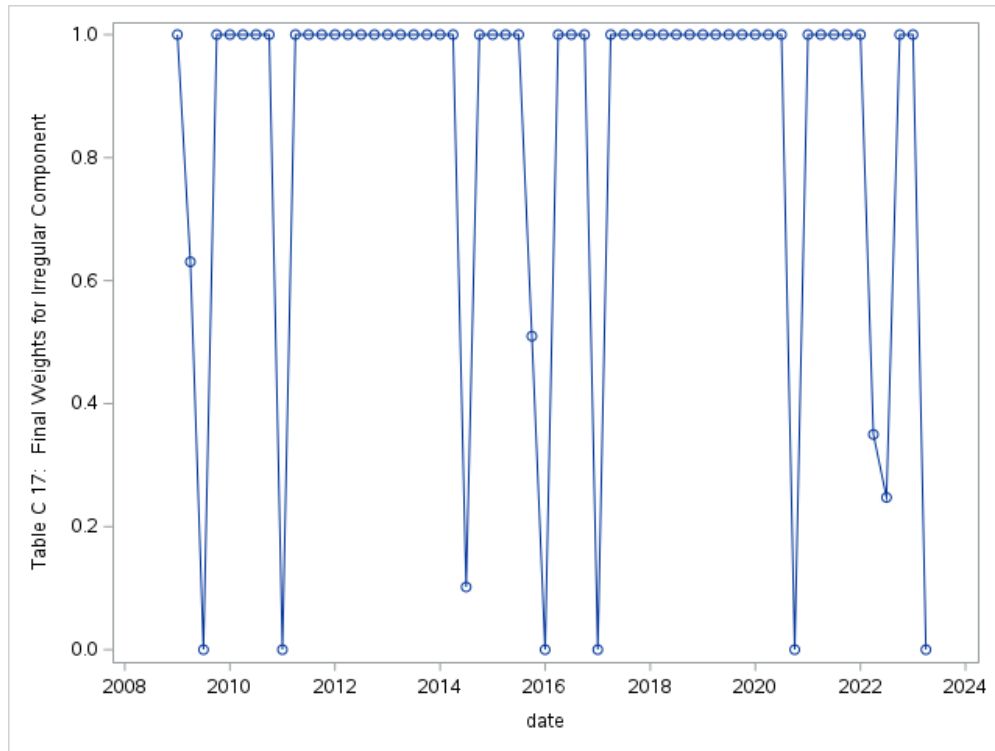
Spørgsmål (b)

Når der er sæsonmæssige faktorer til stede, betegnes modellen $\text{ARIMA}(p,d,q) \times \text{ARIMA}_4(P,D,Q)$, når man antager, at den sæsonmæssige cyklus er 4 perioder.

Den ikke-sæsonmæssig del af ARIMA-modellen er dette led $\text{ARIMA}(p,d,q)$, hvor:

- p er antallet af autoregressive led,
- d er antallet af ikke-sæsonmæssige forskelle, der er nødvendige for stationaritet, og
- q er størrelsen på det glidende gennemsnitsvindue, også kendt som rækkefølgen af det glidende gennemsnit. Dvs. antallet af tid lags.

Den sæsonmæssige del af modellen er angivet ved dette led (P,D,Q) . P angiver antallet af sæson autoregressive komponenter, D er antallet af sæson differencer, og Q er antallet af sæson gli-

Figure 20: C17 for servicestationer

dende gennemsnit (SMA) Seasonal moving Average komponenter.

Jeg ender med at estimere end sæson ARIMA model med formen $ARIMA(1, 0, 1) \times ARIMA_4(0, 1, 1)$ ved at benytte PROC X12. Det betyder at PROC X12 har generet en model som opfanger, at der sæson i min tidserie. I forhold til sæson delen af min ARIMA model, så angiver den, at der 0 sæson autoregressive komponenter, 1 sæson differens og 1 sæson glidende gennemsnit.

Den ikke-sæsonmæssige del af modellen er $ARIMA(1,0,1)$, som kan skrives som:

$$X_t = \phi X_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \quad (3)$$

Den sæsonmæssige del af modellen er $ARIMA_4(0,1,1)$, som, når den kombineres med den ikke-sæsonmæssige del, kan skrives som:

$$(1 - B)(1 - B^4)X_t = (1 - \Theta B^4)\varepsilon_t \quad (4)$$

Derefter kombinerer jeg disse to modeller for at få den samlede $ARIMA(1, 0, 1) \times ARIMA_4(0, 1, 1)$

model:

$$(1 - \phi B)(1 - B)(1 - B^4)X_t = (1 + \theta B)(1 - \Theta B^4)\varepsilon_t \quad (5)$$

hvor X_t er tidsserien på tidspunkt t , B er "backshift operatoren", ϕ er parameteren for AR-delen, θ og Θ er parametrene for henholdsvis ikke-sæson og sæson MA delen, og ε_t er fejllidet.

Indsætter jeg estimaterne i formlen, som er vist i Figur 21 får jeg følgende udtryk:

$$(1 - 0.81373B)(1 - B)(1 - B^4)X_t = (1 - 0.9999B)(1 - 0.9999B^4)\varepsilon_t \quad (6)$$

Figure 21: ARMA Estimation

Exact ARMA Maximum Likelihood Estimation					
For Variable X647					
Parameter	Lag	Estimate	Standard Error	t Value	Pr > t
Nonseasonal AR	1	0.81373	0.06950	11.71	<.0001
Nonseasonal MA	1	-0.99993	0.10547	-9.48	<.0001
Seasonal MA	4	0.99995	0.11086	9.02	<.0001

Estimatet på 0.81373 for den autoregressive parameter ved lag 1 indikerer et stærkt positivt forhold mellem X_t og X_{t-1} hvilket betyder, at den nuværende værdi af tidsserien er stærkt afhængig af sin umiddelbare forrige værdi.

Estimatet på -0.99993 for den ikke-sæsonmæssige moving average parameter ved lag 1 indikerer, at der er en stærk negativ effekt af den foregående periodes observation på den nuværende periodes værdi.

Estimatet på 0.99995 for den sæsonmæssige moving average parameter ved lag 4 indikerer, at der er en næsten perfekt positiv effekt af fejlen fire kvartaler tilbage på den nuværende periodes værdi.

Samlet set antyder disse estimer, at både de ikke-sæsonmæssige og sæsonmæssige komponenter er meget vigtige for at forklare variationen i tidsserien,

Figur 22 viser, om der er en ugedageffekt og en påskeeffekt ved at benytte funktionen "td easter" i RPOC X12, hvormed jeg kan konkludere, at der ingen ugedageffekt er, fordi alle er insignifikante. Dog er der en påskeeffekt, hvor jeg estimerede de 2 dage op til påske søndag, for at se om disse dage vil have en effekt på salget af servicestationer, hvilket de har. Det betyder at salget af servicestationer falder på dagene op til påske søndag med et estimat på -0.031.

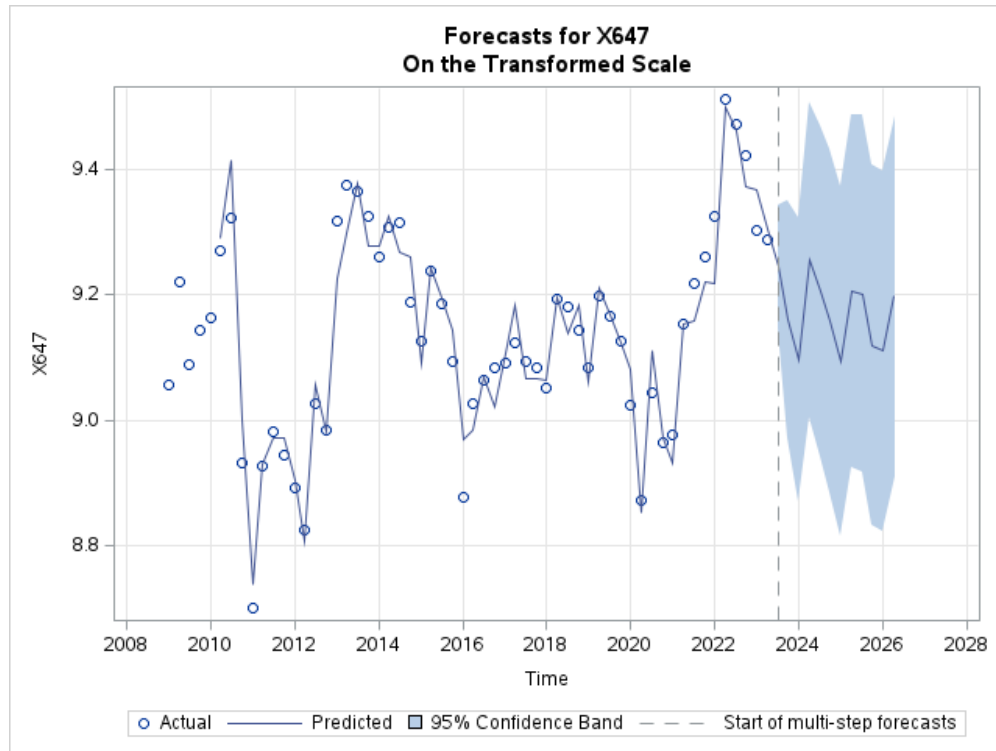
Desuden når jeg kigger på tabellen for den samlede ugedageffekt er denne heller ikke signifikant.

Nederst på Figur 22 observeres de additive outliers (AO). Disse er særlige punkter i tidsserien, hvor der har været usædvanlige spring i niveauet på salget af servicestationer. Eksempelvis ser jeg en outlier i 2. kvartal af 2020 med en negativ afvigelse på -0.22941, og er statistisk signifikant med en p-værdi mindre end 0.0001. Denne outlier kan være påvirket af COVID-19 pandemien, som har en effekt på salget servicestationer.

Figure 22: Traiding Days og Outliers

Regression Model Parameter Estimates						
For Variable X647						
Type	Parameter	NoEst	Estimate	Standard Error	t Value	Pr > t
Trading Day	MON	Est	-0.00140	0.00840	-0.17	0.8688
	TUE	Est	0.01004	0.01005	1.00	0.3238
	WED	Est	-0.01595	0.00869	-1.83	0.0738
	THU	Est	0.00900	0.00797	1.13	0.2655
	FRI	Est	0.00895	0.00905	0.99	0.3281
	SAT	Est	0.00391	0.00777	0.50	0.6170
	SUN(derived)*	Est	-0.01456	0.01158	-1.26	0.2160
Easter	Easter[2]	Est	-0.03136	0.01092	-2.87	0.0064
Automatically Identified	AO 2009Q2	Est	0.11401	0.02999	3.80	0.0005
	AO 2010Q3	Est	0.18255	0.02595	7.04	<.0001
	AO 2011Q1	Est	-0.10316	0.02612	-3.95	0.0003
	AO 2012Q2	Est	-0.21328	0.02375	-8.98	<.0001
	AO 2013Q1	Est	0.28182	0.02702	10.43	<.0001
	AO 2020Q2	Est	-0.22941	0.02588	-8.86	<.0001

Til sidst har jeg valgt at lave en forecast 3 år frem med PROC X12 for at beskrive effekten af den nylige inflation. Figur 23 viser, at den nylige inflation har haft en positiv indvirkning på salget af servicestationer fra perioden 2021 til 2022Q1, hvor jeg derefter observerer et fald i salget af servicestationer. Forudsigelsen viser, at salget vil fortsat aftage det første år, hvorefter det vil stabilisere.

Figure 23: PROC X12 Forecast

Jeg vælger nu at estimere en ARMA(5,1) igennem PROC VARMAX med metoden Maximum Likelihood. Denne model inkluderer 5 autoregressive terme og 1 moving average term. Dvs denne model indeholder modellen for en AR(5) og en MA(1) model.

En ARMA(5,1) model har følgende form:

$$X_t = c + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \phi_4 X_{t-4} + \phi_5 X_{t-5} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (7)$$

hvor:

- X_t er værdien af tidsserien på tidspunkt t ,
- c er en konstant,
- $\phi_1, \phi_2, \phi_3, \phi_4, \phi_5$ er koefficienterne for de autoregressive terme,
- θ_1 er koefficienten for moving average termet term,
- ε_t er fejltarmen på tidspunkt t ,

- ε_{t-1} er fejltermen fra den foregående periode.

Ved at benytte samme notation som for ARIMA modellerne og bruge backshift-operatoren kan ARMA modellen skrives som:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4 - \phi_5 B^5)X_t = (1 + \theta_1 B)\varepsilon_t \quad (8)$$

Indsætter nu estimerne fra Figur 24 i formelen ovenover:

$$(1 - 1.26988B + 0.53320B^2 - 0.05043B^3 - 0.30885B^4 + 0.43811B^5)X_t = (1 + 0.52267B)\varepsilon_t \quad (9)$$

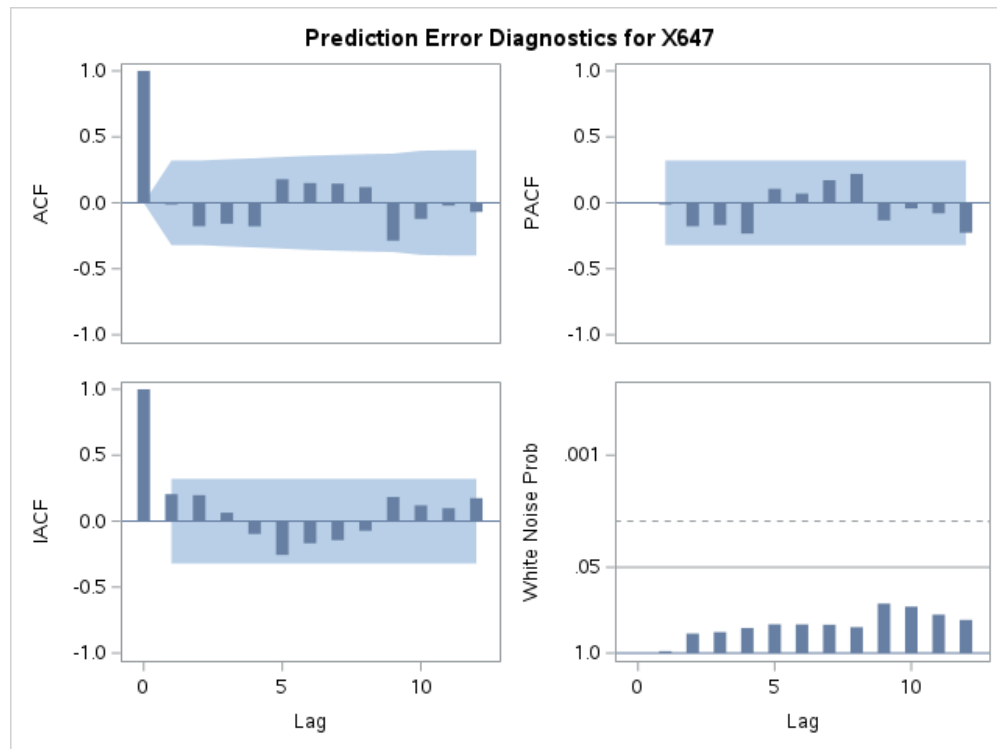
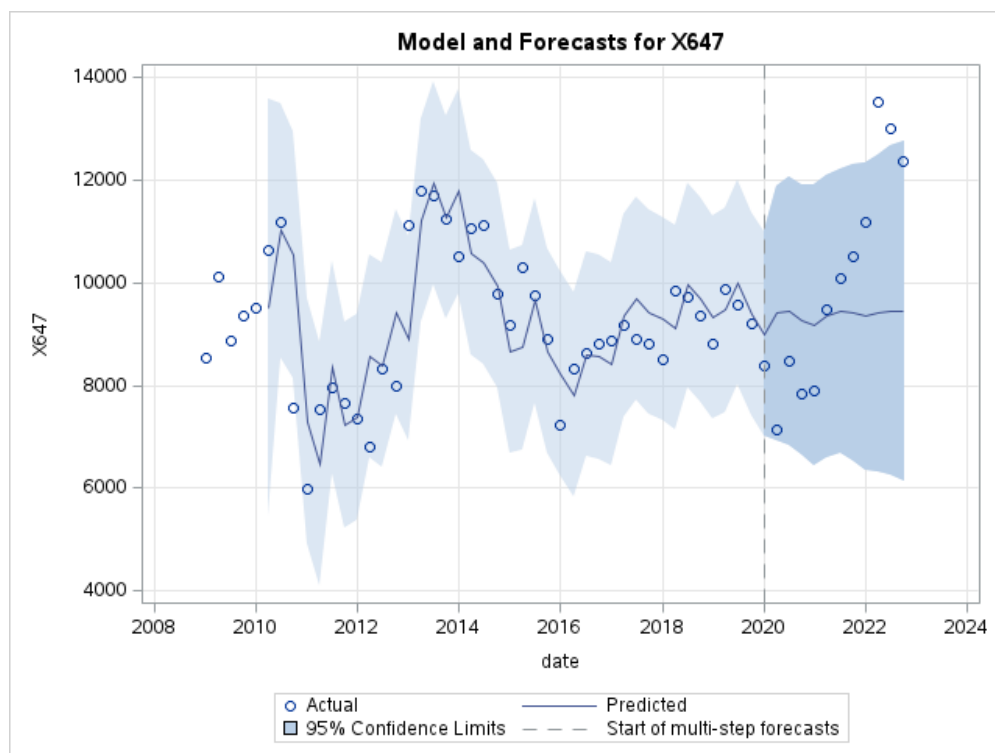
Figure 24: PROC Varmax ARIMA Estimates

Model Parameter Estimates						
Equation	Parameter	Estimate	Standard Error	t Value	Pr > t	Variable
X647	CONST1	3201.82654	1277.05847	2.51	0.0153	1
	AR1_1_1	1.26988	0.28889	4.40	0.0001	X647(t-1)
	AR2_1_1	-0.53320	0.33706	-1.58	0.1196	X647(t-2)
	AR3_1_1	0.05043	0.26753	0.19	0.8512	X647(t-3)
	AR4_1_1	0.30885	0.24029	1.29	0.2043	X647(t-4)
	AR5_1_1	-0.43811	0.15376	-2.85	0.0062	X647(t-5)
	MA1_1_1	0.52267	0.33902	1.54	0.1291	e1(t-1)

For at vurderer om der er sæson i modellen, har jeg plottet residualautokorrelationer i Figur 25

Hvis der er signifikante autokorrelationer ved lag 4, 8, 12 osv. i mit kvartal data, kan det indikere årlig sæsonmæssighed. Men ud fra Figur 25 er der ingen tegn på signifikante autokorrelationer ved lag 4, 8, og 12, hvorfor jeg vurderer at der ikke er sæson.

For at fange effekten af corona og den nylige inflation i PROC Varmax, har jeg benyttet forecast funktionen med back=14 og lead =12, for at se på forudsigelsen før COVID-19 og sammenligne den med de faktisk observation. Dette er visualiseret i Figur 26. Figuren viser at COVID-19, samt den nylige inflation har haft en betydelig effekt på tidsserien, eftersom at forudsigelsen vurderer en relativ stabil periode, hvis COVID-19 og inflationen ikke havde fundet sted.

Figure 25: PROC Varmax residualautokorrelationer**Figure 26:** PROC Varmax Forecast

Spørgsmål (c)

I PROC UCM kørslen finder jeg, at variansen på "Slope" komponenten ikke er er signifikant. Desuden observerede jeg heller ikke en trend i min tidsserie, hvorfor jeg vælger at sætte variansen

lig 0 og benytte "noest" funktionen.

Ved at benytte formelen defineret i opgave 1 (c), kan jeg udvide denne model for at inkludere en trend og sæson komponent, hvor jeg i PROC UCM har valgt en sæson længde på 4, eftersom der er 4 kvartaler på et år.

Først starter jeg med at inkludere en trend i niveau serien u_t ved at bruge følgende definition:

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \eta_t \sim N(0, \sigma_\eta^2) \beta_t = \beta_{t-1} + \xi_t, \xi_t \sim N(0, \sigma_\xi^2)$$

Positive værdier af trenden β_t svarer til en opadgående trend i datasættet, men den faktiske trend er tidsafhængig, fordi den faktiske hældning β_t ifølge ovenstående formel tillades at variere. Res-tledende η_t , ε_t og ξ_t antages alle at være gensidigt uafhængige hvid støj-serier, (Milhoj, 2013).

Til sidst inkluderer jeg også en sæsonkomponent, ved at udvide ovenstående model:

$$y_t = \mu_t + S_t + \varepsilon_t$$

$$S_t = -(S_{t-11} + \dots + S_{t-1}) + \omega_t, \omega_t \sim N(0, \sigma_\omega^2).$$

hvor S_t er den inkluderede sæsonkomponent.

Figur 27 viser at variansen på niveauet er statistisk signifikant, dog fandt jeg at hverken slope eller sæson komponentens varians er signifikant, hvorfor deres varians er sat lig nul.

Figure 27: PROC UCM Final Estimates Opgave 2c

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > t
Level	Error Variance	955063	185527.9	5.15	<.0001

Figur 28, viser at både niveau og sæson komponenten er statisk signifikant, dog er slope parameteren ikke signifikant.

Figure 28: PROC UCM Signifikans analyse af komponenterne

Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Level	1	5339.26	<.0001
Slope	1	0.04	0.8431
Season	3	16.69	0.0008

I min sæson komponent har jeg anvendt trigonometrisk harmonics. Denne tilgang bruger sinus- og cosinusfunktioner til at modellere sæsonmønstre. Figur 29 viser at harmonic 1, som har en periode på 4, hvilket svarer til 1 år er statistisk signifikant. Det betyder at der er en signifikant årlig sæson cyklus som gentager sig.

Figure 29: PROC UCM Harmonics

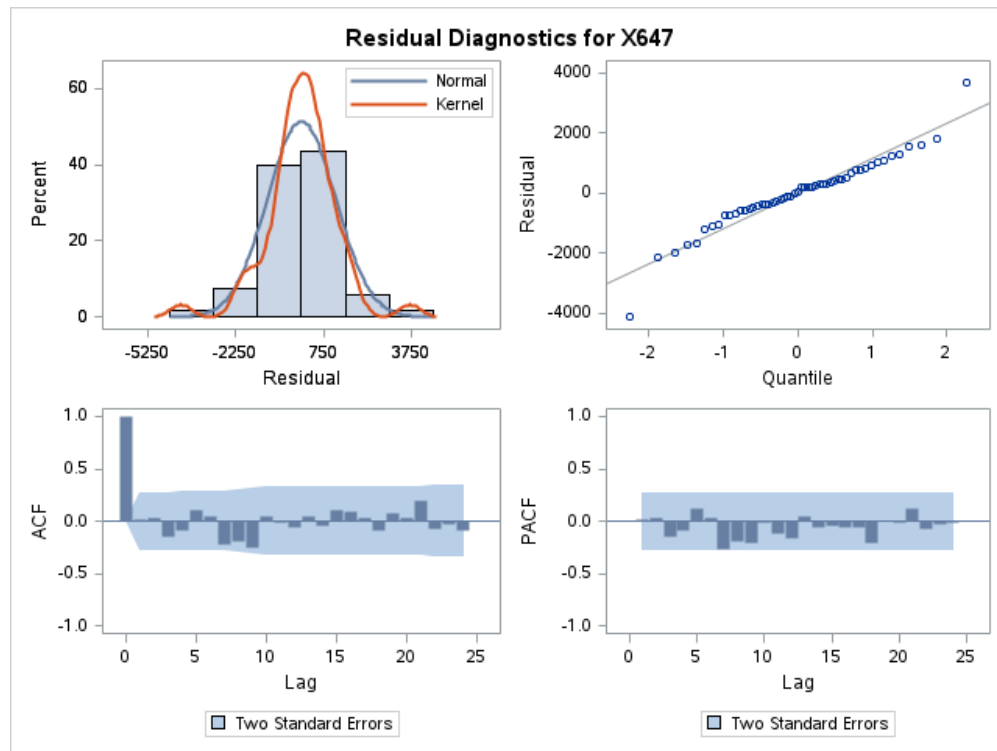
Harmonic Analysis of Trigonometric Seasons (Based on the Final State)						
Name	Season Length	Harmonic	Period	Chi-Square	DF	Pr > ChiSq
Season	4	1	4.00000	15.90	2	0.0004
Season	4	2	2.00000	0.64	1	0.4253

Figur 30 viser at min tidserie er normalfordelt, dog viser QQ-plottet øverst til højre, at der tidsse-rien kan være "heavy tailed", grundet de to observationer i hhv. øverste og nederste hjørne.

I forhold til autokorrelation, ser jeg ingen signifikante spikes i hverken ACF eller PACF figuren. Derudover er der heller ikke nogen signifikante spikes i Figuren 31. Dermed vurderer jeg, at der ikke er problemer med autokorrelation i min model.

Figur 32 viser den udglatte tidserie for niveau komponenten. Ud fra denne, kan jeg vurdere, at niveauet har en tendens til stige og falde i nogle af perioderne.

Figur 33 viser den udglattede tidsserie for slope komponenten. Jeg kan vurdere, at trenden for min tidserie er relativt konstant i hele perioden.

Figure 30: PROC UCM Residual Diagnostics

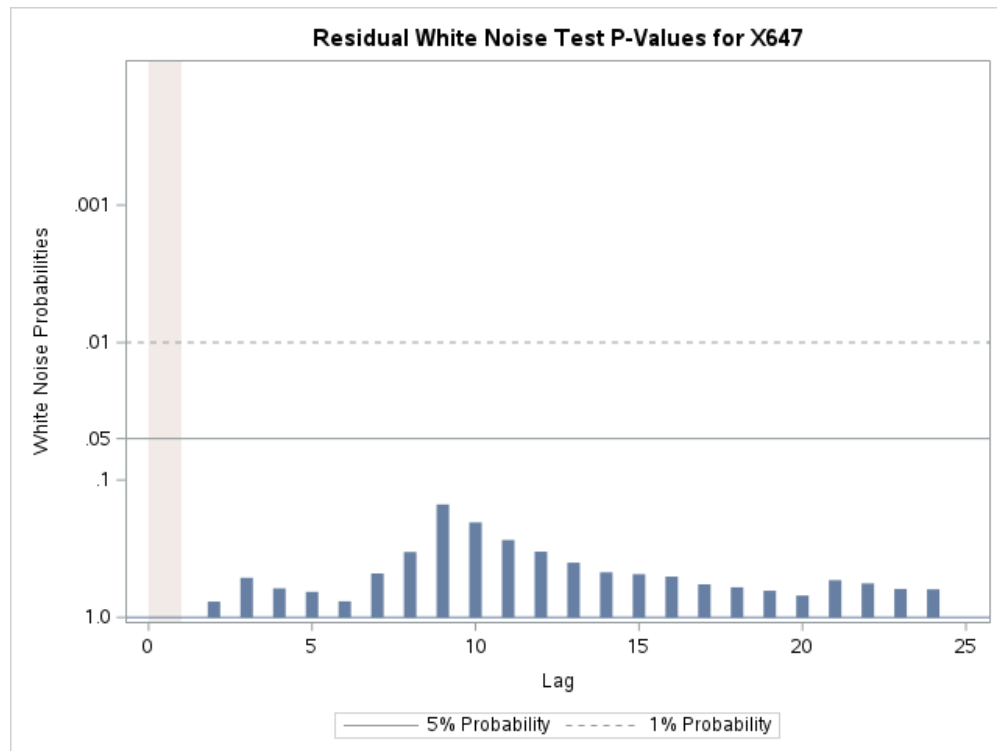
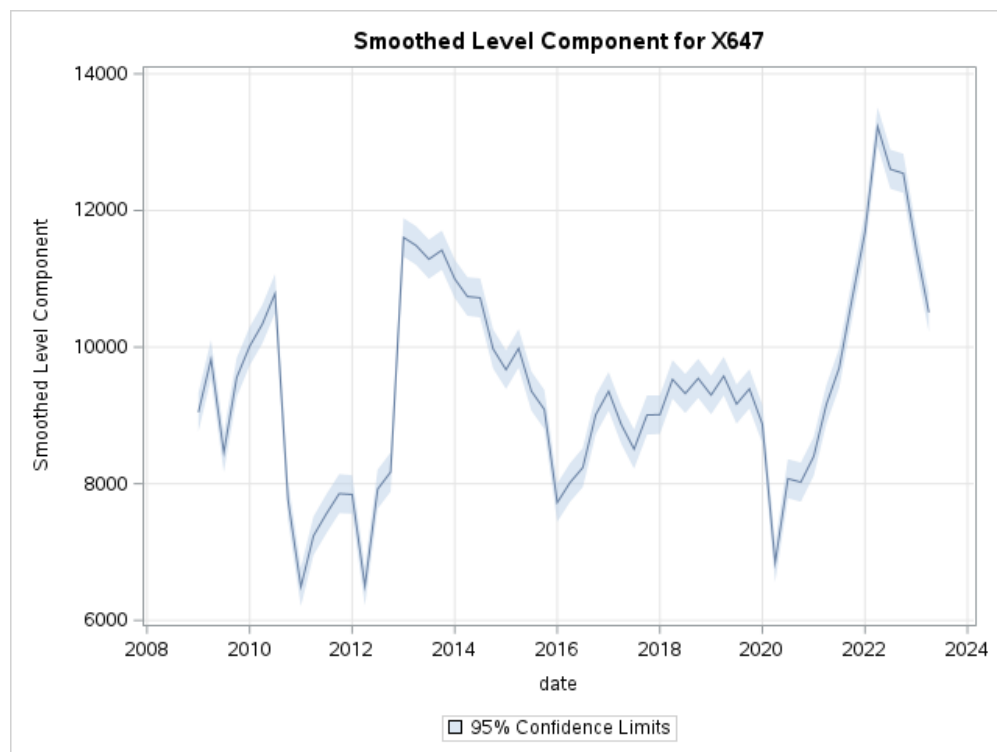
Figur 34, viser den udglattede sæson komponent for min tidserie. Ud fra figuren kan jeg se, at der er en lille sæson effekt i mit datasæt. Men eftersom at y-aksen er relativ lille i forhold til y-aksen på min faktiske tidsserie, som spænder mellem 5000-15000, vurderer jeg, at sæson effekten kun har en lille effekt på min tidsserie.

Spørgsmål (d)

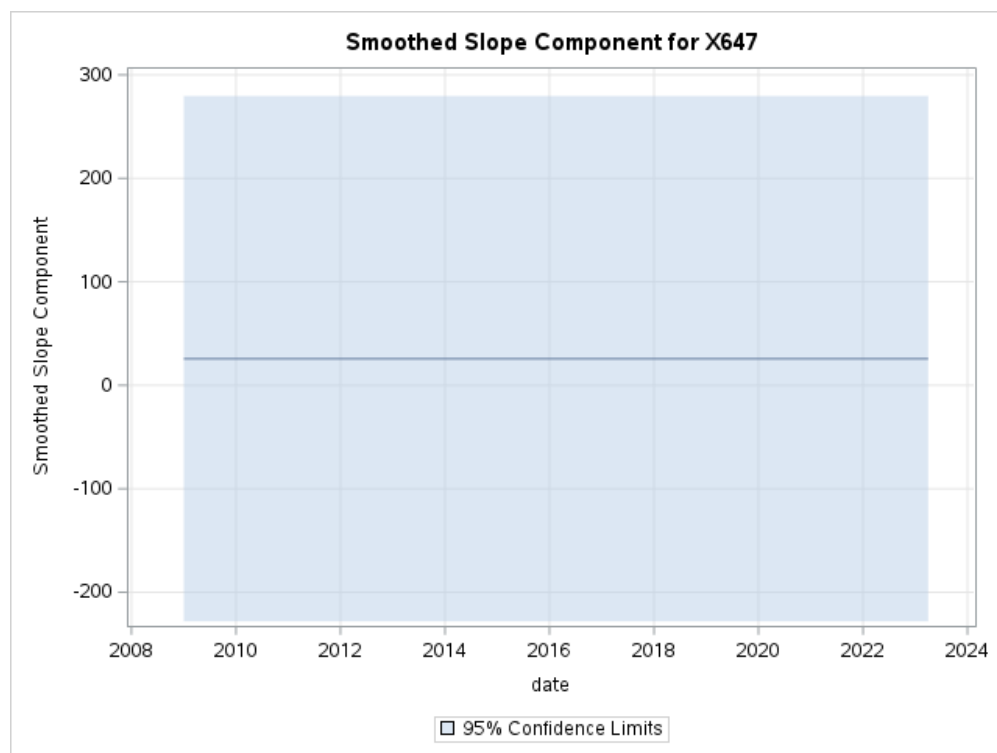
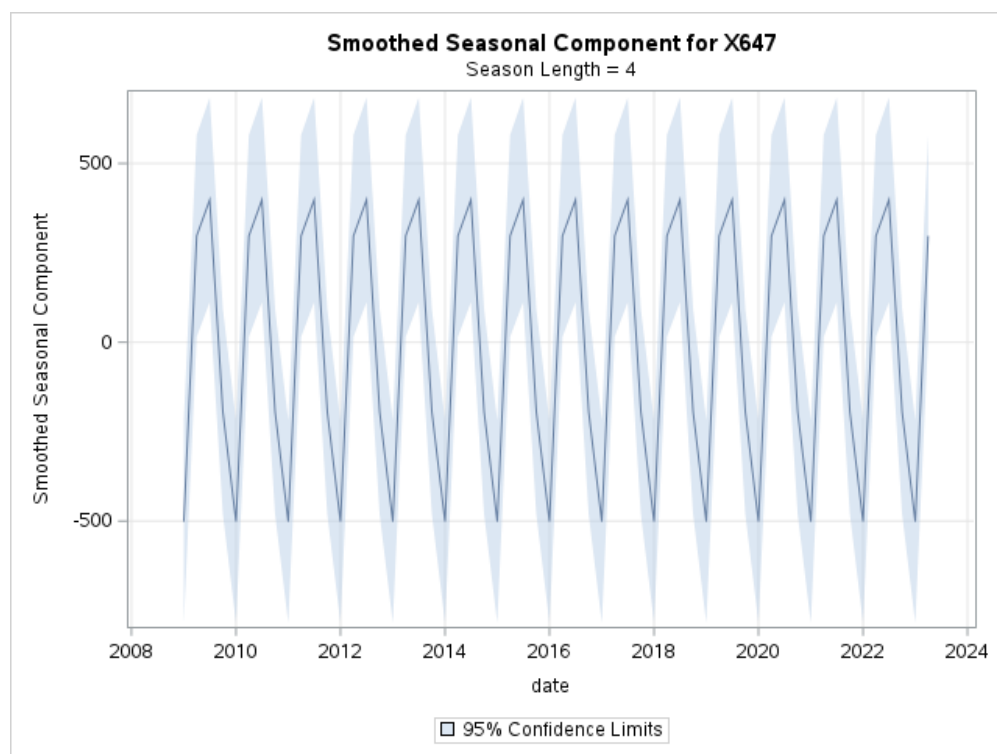
Jeg starter med at vise resultaterne for variabelen X244, som en konstant regressionskoefficient. Figur 35, viser at salget for 'Erhverv i alt', som konstant regressionskoefficient ikke er signifikant på et 5% signifikans niveau. Derudover er estimatet tæt på nul, hvilket betyder at salget for 'Erhverv i alt' ikke er i stand til at forklare salget af servicestationer for hele perioden.

Jeg observerer stadig, at min model ikke har nogen problemer med autokorrelation, hvorfor jeg ikke indfører nogle lags eller autoregressive terme på min irregulære komponent.

Vælger jeg i stedet kun at se på perioden, hvor der er højere inflation, ved kun at se på årene 2020 og frem, finder jeg igen, at salget for 'Erhverv i alt', som konstant regressionskoefficient ikke

Figure 31: PROC UCM Residual White Noise**Figure 32:** PROC UCM Smoothed level component

er signifikant på et 5%. I Figur 36, ser jeg at komponenten er blevet mindre signifikant grundet COVID-19 og den nylige inflation.

Figure 33: PROC UCM Smoothed Slope Component**Figure 34: Smoothed Seasonal Component**

Nu vælger jeg at inkorporere variablen X244, som en tidsvarierende regressionskoefficient med

Figure 35: PROC UCM med x244 som fast koefficient

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > t
Level	Error Variance	924488	181307.1	5.10	<.0001
X244	Coefficient	0.00197	0.0011880	1.66	0.0971

Figure 36: For år 2020 og 2023 PROC UCm Estimates

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > t
Level	Error Variance	1366174	683086.9	2.00	0.0455
X244	Coefficient	0.00189	0.0016836	1.12	0.2606

funktionen "randomreg". Først finder jeg at dens varians er insignifikant, med en p-værdi på 1 set i Figur 37.

Figure 37: PROC UCM X244 Tidsvarierende

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > t
Level	Error Variance	924487	181306.9	5.10	<.0001
X244	Error Variance	4.97391E-32	1.2461E-19	0.00	1.0000

Kigger jeg på Figur 38, ser jeg at komponenten forstat er insignifikant på et 5% signifikans niveau.

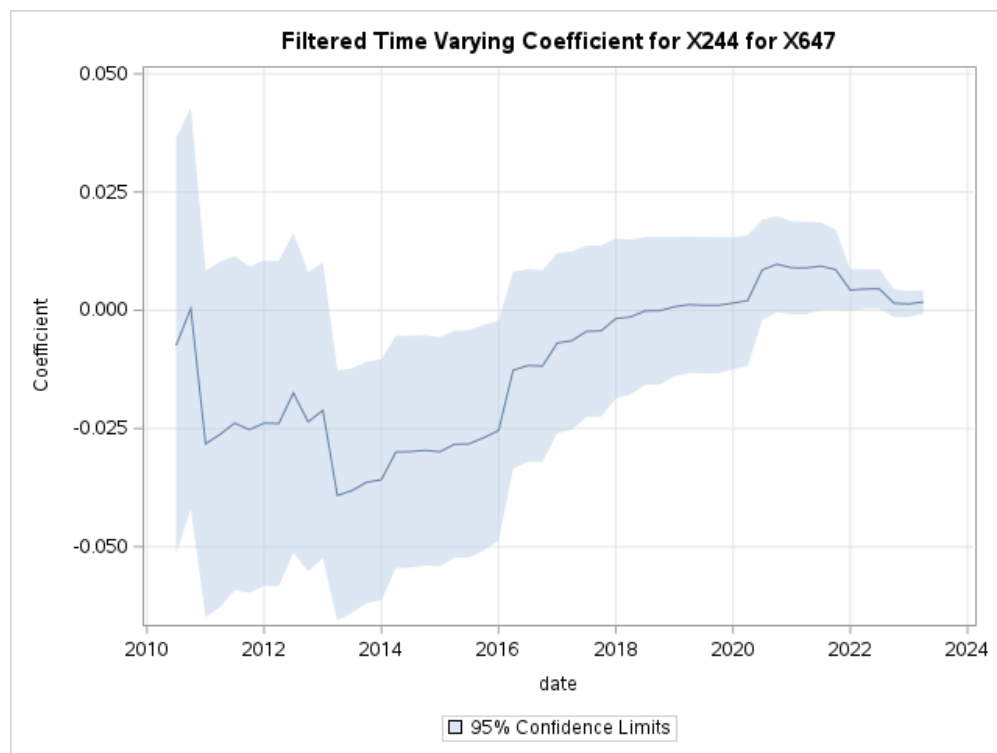
Figure 38: PROC UCM X244 Signifikant Analyse Tidsvarierende

Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Level	1	13.66	0.0002
Slope	1	0.00	0.9709
Season	3	18.70	0.0003
X244	1	2.75	0.0971

Figur 39, viser den filtrerede tidsvarierende koefficient X244, som indikerer at komponenten er relativt stabil igennem hele perioden, bemærk y-aksen er lille. Dette understøttet blot min konklusion om, at variabelen X244, ikke er i stand til at forklare salget af service stationer. Desuden observerer

jeg, at plottet ligger meget tæt på nul i årene 2020 og frem, hvilket indikere, at variabelen ikke forklarer specielt godt i den sidste tid med højere inflation.

Figure 39: Filtered Time Varying Coefficient for X244



Ser jeg igen på den sidste tid med højere inflation for årene 2020 og frem, men nu med tidsvarierende X244 variabel, ser jeg at variansen på X244 nu er signifikant, men omvendt er niveau variansen ikke længere signifikant. Dog forbliver selve X244 komponenten insignifikant, se Figur 40.

Dermed kan jeg konkludere, at salget for 'Erhverv i alt' som forklarende variabel, med både konstant og tidsvarierende regressionskoefficient, ikke er i stand til at forklare udviklingen i salg af servicestationer. Desuden kan variabelen X244 heller ikke hjælpe med at forklare i den sidste tid med højere inflation.

Figure 40: PROC UCM Final Estimates for X244 tidsvarierende for År 2020 og frem

Final Estimates of the Free Parameters					
Component	Parameter	Estimate	Approx Std Error	t Value	Approx Pr > t
Level	Error Variance	0.64621	.	.	.
X244	Error Variance	6.462083E-7	3.23104E-7	2.00	0.0455

Fit Statistics Based on Residuals	
Mean Squared Error	5110517
Root Mean Squared Error	2260.64523
Mean Absolute Percentage Error	14.36024
Maximum Percent Error	14.93785
R-Square	-2.74146
Adjusted R-Square	-3.36504
Random Walk R-Square	-3.53988
Amemiya's Adjusted R-Square	-5.23577
Number of non-missing residuals used for computing the fit statistics = 8	

Significance Analysis of Components (Based on the Final State)			
Component	DF	Chi-Square	Pr > ChiSq
Level	1	2.24	0.1341
Slope	1	0.36	0.5504
Season	3	2.43	0.4883
X244	1	0.96	0.3270

References

- InformIT. (2015). *Simple seasonal exponential smoothing* [Chapter from the book "Predictive Analytics with Microsoft Excel: Working with Seasonal Time Series"]. Retrieved December 1, 2023, from <https://www.informit.com/articles/article.aspx?p=2433607&seqNum=4>
- Milhoj, A. (2013). *Practical time series analysis using sas* (1st and 2nd printing, March and June 2013) [Electronic book: ISBN 978-1-61290-624-9]. SAS Institute Inc.

SAS Kode

```
libname sasts '/courses/d284cd65ba27fe300/PTS/Data';
```

```
*Opgave 1;
```

```
*a);
```

```
data broadway;
```

```
set sasts.broadway_70;
```

```
hour = hour(dt);
```

```
date= datepart(dt);
```

```
format date dandfwkx31.;
```

```
month=month(date);
```

```
year=year(date);
```

```
day=day(date);
```

```
weeknumber = week(date);
```

```
weekday = weekday(date);
```

```
log_antal_v2=log(antal);
```

```
tid=timepart(dt);
```

```
format tid time8.;
```

```
run;
```

```
PROC SGPLOT data=broadway;
```

```
series x=tid y=antal/markers group=date;
```

```
*where month=10 and day>=1 and day<=2;
```

```
run;
```

```
PROC SGPLOT data=broadway;
```

```
series x=tid y=antal/markers group=weekday;
```

```
where weekday>=2 and weekday<=6;
```

```
run;
```

```
PROC SGPLOT data=broadway;
    series x=tid y=antal/markers group=weekday;
    where weekday = 7 or weekday = 1;
run;

*b);

*Syv døgn;

PROC ESM data=broadway print=estimates plot=all lead=168;
id dt interval=hour ;
forecast antal/method=addseasonal alpha=0.1;
run;

PROC TIMESERIES data=broadway out=broadway_day;
id date interval=day accumulate=total setmissing=0;
var antal;
run;
proc print data=broadway_day;
run;

PROC ESM data=broadway_day print=estimates plot=all lead=7;
id date interval=day ;
forecast antal/method=addseasonal alpha=0.1;
run;
```

```
*c);
```

```
PROC EXPAND data=broadway out=broadway_minutter to=minute;  
convert antal=interpolated / method=spline;  
id dt;  
run;
```

```
proc SGPLOT data=broadway_minutter;  
series x=dt y=interpolated/markers;  
run;
```

```
PROC UCM data = Broadway_minutter;  
id dt interval = minute;  
model interpolated;  
level plot=smooth;  
slope;  
cycle plot=smooth;  
outlier;  
estimate plot=all;  
*forecast lead=100 plot=forecasts alpha=0.1;  
run;
```

```
PROC UCM data=broadway_minutter plots=all;  
id dt interval=minute;  
model interpolated;  
irregular p=3;  
level variance=0 noest;
```

```
cycle period=1440 ;  
cycle period=480 ;  
cycle period=720 ;  
estimate;  
run;
```

**Opgave 2;*

**benytte X244 i d;*

**a);*

**x647 er Servicestationer;*

```
data salg;  
set sasts.salg;  
year = year(date);  
run;  
  
PROC X12 data=salg date=date plots=all;  
var x647;  
transform function=log;  
automdl ;  
regression;  
outlier;  
x11;  
output out=out a1 c17 d10 d11 d12 d13;  
run;
```

```
PROC SGPLOT data=out;
series x=date y=x647_A1/markers;
series x=date y=x647_D11/markers;
*xaxis values=('1jan11'd to '1feb13'd by month);
run;

*D11 viser den sæson jsutures, men den ligger meget tæt op af den originale
↪ tidsserie, fordi tidsserien ikke indeholder sæson.;
```

```
PROC SGPLOT data=out;
series x=date y=x647_D10/markers;
*xaxis values=('1jan09'd to '1jan15'd by month);
run;
```

```
PROC SGPLOT data=out;
series x=date y=x647_D12/markers;
*xaxis values=('1jan99'd to '1jan11'd by year);
run;
```

**Beskriv de periode hvor trenden stiger rigtig meget for guld. ;*

```
PROC SGPLOT data=out;
series x=date y=x647_d13/markers;
*xaxis values=('1jan99'd to '1jan11'd by year);
run;
```

```
PROC SGPLOT data=out;
series x=date y=x647_c17/markers;
```

```
*xaxis values=('1jan99'd to '1jan11'd by year);
run;

*b);

PROC X12 data=salg date=date plots=all ;
var x647;
transform function=log;
outlier type=ao;
x11 force=totals;
regression predefined=(td easter(2));
forecast lead=12 ;
ods output ForecastCL=predicted;
output out=out a1 c17 d10 d11 d12 d13 d11f;
run;

PROC VARMAX data=salg plot=forecasts(all) plots=all;
id date interval=qtr;
model x647/method=ml;
output out=out lead=12 back=14;
run;

*c);
PROC UCM data=salg plots=all;
id date interval=qtr;
model x647;
irregular variance=0 noest;
level ;
slope variance=0 noest;
season length=4 type=trig print=harmonics variance=0 noest;
```

```
estimate;
```

```
run;
```

```
*d);
```

```
PROC UCM data=salg plots=all;
```

```
id date interval=qtr;
```

```
model x647=X244;
```

```
irregular variance=0 noest;
```

```
level ;
```

```
slope variance=0 noest;
```

```
season length=4 type=trig print=harmonics variance=0 noest;
```

```
estimate;
```

```
run;
```

```
PROC UCM data=salg plots=all;
```

```
id date interval=qtr;
```

```
model x647=X244;
```

```
irregular variance=0 noest;
```

```
level ;
```

```
slope variance=0 noest;
```

```
season length=4 type=trig print=harmonics variance=0 noest;
```

```
estimate;
```

```
where year>=2020;
```

```
run;
```

```
PROC UCM data=salg plots=all;
```

```
id date interval=qtr;
```

```
model x647;
```

```
irregular variance=0 noest;  
level ;  
slope variance=0 noest;  
season length=4 type=trig print=harmonics variance=0 noest;  
randomreg X244 /plot=smooth;  
estimate;  
run;
```

```
PROC UCM data=salg plots=all;  
id date interval=qtr;  
model x647;  
irregular variance=0 noest;  
level ;  
slope variance=0 noest;  
season length=4 type=trig print=harmonics variance=0 noest;  
randomreg X244 /plot=smooth;  
estimate;  
where year>=2020;  
run;
```