



Praktisk tidsrækkeanalyse

Eksamensnummer: 205

Vintereksamen 2023-24 - Ordinær

Submitted on: 03/12/2023

ECTS: 7.5

Pages: 20



Contents

1	Proc UCM	1
1.1	Anvendelse af PROC UCM copy paste	1
1.2	Filtered and smoothed level plot	2
1.3	PROC UCM outlier	2
1.4	Panel plot for fit of the model	3
1.5	Smoothed slope plot og estimer eksempel	4
1.6	Fix autocorrelation i PROC UCM	5
1.6.1	Residual white noise plot	5
1.7	Smoothed level plot	6
1.8	Smoothed level plot	6
1.9	Basic Model	6
1.10	ARIMA	7
1.10.1	Udvidelse af modellen med trend	8
1.10.2	Udvidelse af modellen med sæson	8
1.10.3	Udvidelse af modellen med trigonometriske cyklusser	8
2	Nonseasonal ARIMA model	9
2.1	Seasonal ARIMA model	10
2.2	Seasonal ARIMA model SAS Eksempel	10
2.3	PROC UCM SAS Eksempel AR(1)	11
2.4	UCM Sæson metoder	12
3	PROC ESM	13
3.1	SIMPLE exponentiel smoothing	13
3.2	Double exponentiel smoothing	13
3.3	LINEAR exponentiel smoothing	14
3.4	Damptrend exponentiel smoothing	14
3.5	Addseasonal or Seasonal exponentiel smoothing	14
3.6	Multiseasonal exponentiel smoothing	14
3.7	WINTERS exponentiel smoothing	15
3.8	ADDWINTERS exponentiel smoothing	15

4	Proc x12	15
4.1	mode = add	15
4.2	mode = mult	16
4.3	mode = logadd	16
4.4	PlotD11	16
4.5	PlotD12	16
4.6	Plot D10	17
4.7	Plot D13	17
4.8	Plot C17	18
4.9	Ugedagseffekt og påskeeffekt	18
5	Proc varmax	19
6	Q-Q plot	20
7	Sæson justering	20
8	Proc expand	22
8.1	method join	22
8.2	method Spline Interpolation	22
8.3	method step	23

1 Proc UCM

1.1 Anvendelse af PROC UCM copy paste

PROC UCM er en nyttig metode, fordi den tillader en fleksibel modellering af tidsrækker 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 data, som kan variere over tid.

Jeg vælger nu at skrive modellen op for PROC UCM metoden. Først skriver jeg udtrykket op for den basale model, dernæst inkludere jeg en trend og til sidst en sæsonkomponent.

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

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

Niveauelementen 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 restleddet ε_t , en stokastisk komponenten der ikke kan blive forklaret af modellen, (Milhoj, 2013).

Jeg udvider nu modellen til at inkludere en trend og sæson, for at efterligne modellen jeg estimerer ved hjælp af PROC UCM. Først starter jeg med at inkludere en trend i niveau serien u_t ved at bruge følgende definition:

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

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 inkludere 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.

1.2 Filtered and smoothed level plot

Det filtrerede plot, Figur 11.1, er defineret ved kun at bruge tidligere observationer, så kun observationer før og ikke senere end tidspunktet t er inkluderet i beregningen af niveauet på tidspunktet t . I den glattede version, Figur 11.2, bruges alle observerede værdier både før og efter tidspunktet t i beregningen af niveauet på tidspunktet t , hvilket fører til en glattere ex post plot. Plottene ser meget ens ud, men da det glattede plot er baseret på flere observationer, er konfidensgrænserne i observationsperioden smallere.

1.3 PROC UCM outlier

I Figur x er residualerne plottet. De mest dominerende residualer er for årene xxxx og xxxx, men en positiv residual i xxxx og de negative residualer i slutningen af xxxx'erne synes også at være betydningsfulde. OUTLIER-erklæringen frembringer en formel outlier-test, som er vist i Output xx.x. I denne test anvendes en dummy-komponent for hvert år for hver observation, og hvis denne komponent testes som signifikant, udskrives observationen som en additiv outlier. Disse tests udelukker indflydelsen fra potentielle outliers på både residual- og niveaukomponenten, så resultaterne er ikke en simpel gentagelse af det, der vises i residualplottet. Estimerne for de udskrevne outliers estimeres for dummy-komponenten, der er beregnet til at kompensere for outlier. Af den grund er fortegnet for estimatet i Output xx.x det modsatte af fortegnet for den tilsvarende residual i Figur xx.x. Som det bemærkes i Afsnit x.xx, er outliers for årene xxxx–xxxx sandsynligvis på grund af registreringsfejl i årene for den danske genforening med de sydlige dele af Jylland.

1.4 Panel plot for fit of the model

Histogram med Kernel Tætheds Estimation:

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 kerne-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 acceptabel.

Q-Q Plot: 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.

Autokorrelationsfunktion (ACF) Plot:

Nederst til venstre ses et autokorrelationsplot, der viser korrelationen mellem residualerne og deres tidligere værdier (lags). Hver søjle repræsenterer korrelationen ved en specifik lag. Hvis søjlerne ligger indenfor de blå grænser (som repræsenterer to standardfejl), er korrelationen for den forsinkelse ikke statistisk signifikant. Eksempel: Autokorrelationsplottet, som også er inkluderet i panelet i Figur 11.5, peger på nogle positive autokorrelationer for forsinkelser (lags) fra én til fem. Dette betyder, at modellen først fanger et nyt fertilitetsniveau med nogle års forsinkelse. De første to partielle autokorrelationer er positive, hvilket indikerer, at en autoregressiv model kan være passende.

Partiel Autokorrelationsfunktion (PACF) Plot:

Nederst til højre er et partielt autokorrelationsplot. Mens ACF måler den totale korrelation mellem to serier ved en given lag, tager PACF kun den lineære forudsigelighed, der ikke er forklaret af tidligere forsinkelser, i betragtning.

Dette plot bruges ofte til at identificere antallet af lags, der skal inkluderes i en autoregressiv model. Ligesom med ACF, indikerer søjler, der ikke overskrider de blå grænser, at der ikke er nogen signifikant partielle autokorrelationer ved disse lags.

Denne artikel er god til pacf:

1.5 Smoothed slope plot og estimerer eksempel

Hædningskomponenten, vist i Figur 11.6, varierer i overensstemmelse med den lokale trendadfærd i den observerede serie. Hædningskomponenten viser et fald i fertiliteten over de første cirka 30 år, men viser derefter fertiliteten som værende ret konstant i omkring 50 år, selvom der er store udsving. Hædningskomponenten er signifikant negativ i de første 30 år, hvor fertiliteten falder med en stabil rate, men siden 1930 er hædningskomponenten mere volatil. Fra et overordnet perspektiv kunne den betragtes som nul, fordi en trend intuitivt ikke kan ændre sig så hurtigt.

De estimerede variancer, Output 11.3, viser, at variansen for hældningen er ubetydelig. På trods af diskussionen ovenfor kan hældningen betragtes som en konstant i stedet for som en tidsvarierende komponent. Hældningen fastsættes ved at fastholde variansen til 0 i hældningsudtalelsen. Muligheden $\text{var}=0$ giver værdien 0 som en startværdi for variansparameteren. Muligheden noest specificerer, at denne parameter ikke estimeres sammen med de andre modelparametre.

I denne situation er hædningsparameteren en konstant. Denne konstante hædningsparameter er signifikant, som det rapporteres i tabellen over tilstande for niveau- og hædningskomponenten for den sidste observation i Output 11.4. Den estimerede koefficient -21.3 er numerisk større end dobbelt så stor som dens standardfejl. Fertiliteten er derfor modelleret som havende en fast negativ trend. Dette kunne naturligvis ses som en acceptabel grov tilpasning til den observerede fertilitet for de mere end 100 år i prøveperioden. Men når det kommer til prognoser, virker det urealistisk, når man ser prognoseplottet, Figur 11.7. En konstant, lineært aftagende fertilitet betyder, at den danske befolkning snart vil uddø. Den tidsvarierende hædningskomponent i den oprindelige kode for Program 11.2, som er plottet i Figur 11.6, ender næsten ved nul for de seneste observationer. Dette svarer til en næsten horisontal trend i prognoseperioden, hvilket er mere rimeligt, selvom variansparameteren kun er ubetydeligt positiv.

1.6 Fix autocorrelation i PROC UCM

Autokorrelation er en situation, hvor fejftermerne (residualer) i en tidsseriemodel korrelerer med sig selv over forskellige tidsintervaller. Dette kan føre til, at de estimer, der bruges i modellen, ikke er effektive, og at standardfejlene bliver skæve, hvilket påvirker både konfidensintervallerne og teststatistikkerne.

I SAS PROC UCM (Unobserved Components Models) kan jeg løse problemet med autokorrelation ved at inkludere autoregressive (AR) komponenter. Dette gøres ved at tilføje en irregular komponent med en p option for at angive antallet af autoregressive termer. $p=1$, angiver en AR(1) model for den irregulær komponent. Det betyder at hver værdi i serien påvirkes af den foregående værdi plus en tilfældig støjterm.

En anden metode er at benytte deplag lag=1, som tilføjer en forsinket (lagged) version af den afhængige variabel som en regressor i modellen. Med lag=1 inkluderer jeg en enkelt tidsforsinkelse, hvilket indikerer, at den nuværende observation påvirkes af sin umiddelbare forrige observation. Denne metode bruges til at tage højde for situationer, hvor den nuværende observation påvirkes af sin egen historie, og den hjælper med at korrigere for autokorrelation ved at inkludere denne tidshistorik i modellen.

Figur x, viser at DepLag parameteren ses at være signifikant

1.6.1 Residual white noise plot

Dette plot er et søjlediagram, der viser p-værdier for en test af hvid støj (white noise) for resterende værdier (residuals) ved forskellige forsinkelser (lags). Hver søjle repræsenterer en p-værdi for et specifikt lag, med lag-nummeret på x-aksen og p-værdien på y-aksen på en logaritmisk skala.

Figuren bruges til at vurdere, om tidsseriens residualer er hvid støj, det vil sige tilfældigt fordelt og uden signifikant autokorrelation. Hvis de blå søjler ligger over signifikanslinjen på 5%, dvs. søjlerne går mod 0 og er under 5%, tyder det på at residualerne har signifikant autokorrelation, og er ikke blot hvid støj. Omvendt de søjler som ligger under signifikanslinjen, dvs. ligger tættere mod 1 og er over 5%, betyder at der at residualerne ikke har signifikant autokorrelation, og opfører sig

som hvid støj.

1.7 Smoothed level plot

Dette "Smoothed Level Component" plot viser den glattede niveaukomponent for en variabel. Den mørke linje, der bevæger sig gennem midten af det lyseblå område, repræsenterer den glattede estimering af niveaukomponenten over tid. Det betyder, at denne linje viser den justerede, gennemsnitlige værdi af den observerede tidsserie, efter at den er blevet justeret for tilfældige fluktuationer og andre komponenter som trend og sæsonmæssighed. Figuren hjælper med at identificere, hvor stabilt eller volatilt "variablen" er over den givne periode. I dette eksempel viser grafen...

1.8 Smoothed level plot

Forecast plottet viser forudsigelserne for variablen .

1.9 Basic Model

Først skriver jeg en tidsserie, kaldet y_t , op i sin enkleste form, som kan benytte til at skrive en ARIMA(0,1,1) model for tidserien y_t . 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 , som også bliver kaldt for residual komponenten, en stokastisk komponenten der ikke kan blive forklaret af modellen. Desuden udtrykker en lille værdi af variansen σ_ε^2 , at der er en tæt tilpasning til den observerede tidsserie ved niveau serien μ_t , (Milhoj, 2013).

Niveau-processen μ_t varierer ifølge variansen σ_η^2 , således at niveauet danner en jævn serie for små værdier af variansen, men niveauet er volatilt for store værdier af σ_η^2 . På denne måde kan en volatil observeret serie modelleres som en jævn niveauserie med en lille værdi af σ_η^2 og en stor værdi af σ_ε^2 , fejltermens varians. Dette kommer til udtryk i plottet ”Smoothed level plot”, som bliver dannet ved PROC UCM kørslen (Milhoj, 2013).

Denne basale model, der kun indeholder én uobserveret komponent, kan generaliseres til modeller, der indeholder mange flere komponenter. For eksempel en model der udtrykker trend og sæson. Den uregelmæssige komponent kan modelleres ikke kun som simple uafhængige fejltermer, men kan også indeholde tidsserieegenskaber som autokorrelation (Milhoj, 2013).

1.10 ARIMA

Ved at kombinere de to ligningerne ovenfor, er det muligt at skrive en ARIMA(0,1,1) model op:

$$y_t - y_{t-1} = \mu_t + \varepsilon_t - (\mu_{t-1} + \varepsilon_{t-1}) = \eta_t + \varepsilon_t - \varepsilon_{t-1}.$$

Hvis jeg nu definerer $w_t = \eta_t + \varepsilon_t - \varepsilon_{t-1}$, finder jeg, at

$$\text{var}(w_t) = 2\sigma_\varepsilon^2 + \sigma_\eta^2$$

og

$$\text{cov}(w_t, w_{t-1}) = -\sigma_\varepsilon^2.$$

Den første-ordens autokorrelation for serien af forskelle $y_t - y_{t-1}$ ses så at være lig med $-\sigma_\varepsilon^2 / (2\sigma_\varepsilon^2 + \sigma_\eta^2)$, og alle højere-ordens autokorrelationer er 0. Dette betyder, at serien w_t har en autokorrelationsfunktion af en glidende gennemsnitsproces

$$w_t = \zeta_t - \theta_1 \zeta_{t-1}.$$

Her betegner ζ_t en hvid støjserie, (Milhoj, 2013).

1.10.1 Udvidelse af modellen med trend

Niveau serien, μ_t kan udvides til at inkludere en trend:

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

Positive værdier af β_t svarer til en opadgående trend i datasættet, men den faktiske trend er tidssafhængig fordi den faktiske hældning β_t ifølge ovenstående formel tillades at variere. Restledende η_t , ε_t og ξ_t antages alle at være gensidigt uafhængige hvid støj-serier, hvilket betyder, at de hver består af identisk fordelte, uafhængige stokastiske termer. Deres varianser giver en idé om komponenternes stabilitet. For eksempel giver værdien $\sigma_\xi^2 = 0$ en model med en konstant trend, men større værdier af σ_ξ^2 tillader trenden at fluktuere, (Milhoj, 2013). .

1.10.2 Udvidelse af modellen med sæson

Det er også muligt at inkorporere en tidsvarierende sæsonkomponent ved, at udvide niveau modellen fra tidligere:

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

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

hvor S_t er den indkluderede sæsonkomponent. Den faktiske definition af sæsonkomponenterne med en sum på næsten 0 bortset fra støjleddet sikrer, at sæsonkomponenten ikke har nogen indflydelse på det estimerede niveau, (Milhoj, 2013). .

1.10.3 Udvidelse af modellen med trigonometriske cyklusser

Modellen kan også udvides med trigonometriske cyklusser. I økonomisk historie ses såkaldte forretningscyklusser ofte i lange økonomiske tidsserier som temmelig regelmæssige oscillationer. 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). .

Et mere intuitivt tal er måske bølgelængden, som er længden af tidsintervallet mellem to efterfølgende toppe. Denne bølgelængde er lig med $2\pi/\lambda$. Dæmpningsfaktoren er $\gamma = \sqrt{a^2 + b^2}$, hvor tallet $\gamma = 1$ svarende til en stabil cyklus. Parameteren φ i den anden formulering af funktionen er en faseforskydning. Det er selvfølgelig muligt at udvide en sådan cyklisk komponent med stokastiske fejlede, der tillader dæmpningsfaktoren og faseforskydningen φ at udvikle sig over tid. Cykluslængden λ antages dog at være fast, fordi den typisk repræsenterer en konstant fra den fysiske verden, såsom en årlig cyklus, (Milhoj, 2013).

2 Nonseasonal ARIMA model

En ikke-sæsonmæssig ARIMA-model klassificeres som en "ARIMA(p,d,q)"-model, 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.

En yderligere generalisering er kombinationen af autoregressive og glidende gennemsnitsmodeller, der tillader generelle autokorrelationsstrukturer. Den grundlæggende form af disse modeller er ARMA(1,1)-modellen:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

Middelværdien i denne notation er

$$\mu = \frac{\phi_0}{1 - \phi_1}$$

Flere led kan tilføjes til både den autoregressive og den glidende gennemsnitsdel. Flere led ville føre til den generelle model betegnet som ARMA(p,q). Hvis der også anvendes differensiering, betegnes klassen af modeller ARIMA(p,d,q), hvor d er antallet af anvendte differenser, som regel $d = 0$ eller $d = 1$, (Milhoj, 2013).

I denne model er:

- X_t : Værdien af tidsserien på tidspunkt t .

- ϕ_0 : En konstant, der repræsenterer interceptet.
- ϕ_1 : Autoregressiv koefficient for første orden.
- θ_1 : Koefficient for det første glidende gennemsnit.
- ε_t : Fejltermen på tidspunkt t , antaget at være hvid støj med en middelværdi på nul og en konstant varians.

2.1 Seasonal ARIMA model

Hvis der er sæsonmæssige faktorer til stede, betegnes modellen $ARIMA(p,d,q) \times ARIMAS(P,D,Q)$, når man antager, at den sæsonmæssige cyklus er S perioder. 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 differenter, og Q er antallet af sæson glidende gennemsnit (SMA) Seasonal moving Average komponenter. Airline-modellen skrives som $ARIMA(0,1,1) \times ARIMA_{12}(0,1,1)$, (Milhoj, 2013).

Airline-model for kvartalsdata, $ARIMA(0,1,1) \times ARIMA_4(0,1,1)$ -modellen kan skrives på følgende måde:

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

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

2.2 Seasonal ARIMA model SAS Eksempel

Den generelle form af $ARMA(2,0,0) \times ARMA(0,1,1)$ modellen er:

$$\phi_1 X_{t-1} + \phi_2 X_{t-2} + (1 - B^S)X_t = (1 + \Theta B^S)\varepsilon_t$$

Her er:

- X_t : Værdien af tidsserien på tidspunkt t .
- ϕ_1, ϕ_2 : Autoregressive koefficienter for lag 1 og 2.

- Θ : Sæsonmæssig glidende gennemsnit.
- ε_t : Fejltermen ved tid t .
- S : Er antallet af perioder i sæson cyklusen.

ARMA Model Med Parametre Indsat:

ARMA(2,0,0)xARMA(0,1,1) modellen med parametrene er:

$$1.20555X_{t-1} - 0.35713X_{t-2} + (1 - B^{12})X_t = (1 - 0.99971B^{12})\varepsilon_t$$

I denne model er nonseasonal AR koefficienten $\phi_1 = 1.20555$ and $\phi_2 = -0.35713$, og sæson MA koefficienten for en 12-periode cyklus er $\Theta = -0.99971$.

Backshift-operatoren B er defineret ved:

$$B^1 X_t = X_{t-1}$$

$$B^{12} X_t = X_{t-12}$$

2.3 PROC UCM SAS Eksempel AR(1)

Den estimerede model i SAS med PROC UCM, baseret på de slutlige estimater af de frie parametre, kan udtrykkes som følger:

$$X_t = \text{Level}_t + \text{Slope}_t + \text{Irregular}_t + \text{Season}_t$$

hvor:

- Level_t er niveau-komponenten, estimeret til 67,348.51478 med en standardfejl på 969.3609.
- Slope_t er hældningskomponenten, estimeret til -8.22814503 med en standardfejl på 28.49237.
- (Irregular_t) er den uregelmæssige komponent, repræsenteret ved en autoregressiv model af første orden (AR_1) med en koefficient på -0.01924 og en fejlvarians på 14,263,778.
- (Season_t) er den sæsonmæssige komponent, hvis fejlvarians er estimeret til 22,983.

2.4 UCM Sæson metoder

Sæsonmæssige Dummyvariabler:

Denne metode bruger dummyvariabler (også kendt som indikatorvariabler) til at repræsentere forskellige sæsoner eller perioder i dataene. Hver sæson har sin egen dummyvariabel, som tager værdien 1 for observationer i den sæson og 0 ellers. Sæsonmæssige dummyvariabler er især nyttige til at fange faste sæsonmæssige effekter og er nemme at fortolke. Denne tilgang fungerer godt med data, der har stabile, konsekvente sæsonmønstre, som ikke varierer over tid. Block Seasonal:

Block seasonal-metoden modellerer sæsonmæssighed som en serie af block-faktorer, der kan ændre sig fra en sæson til den næste. Denne metode er nyttig for tidsserier med flere observationer pr. sæson og hvor det sæsonmæssige mønster kan variere på tværs af sæsonerne. Block seasonal-modeller kan fange komplekse sæsonmæssige strukturer og er mere fleksible end sæsonmæssige dummyvariabler til at repræsentere forskellige typer af sæsonmæssige variationer.

Sæsonmæssige Harmonikker (Trigonometrisk Sæsonmæssighed):

Denne tilgang bruger sinus- og cosinusfunktioner til at modellere sæsonmønstre, og fanger den cykliske natur af sæsonmæssighed. Harmoniske metoder er særligt effektive til at håndtere glatte og regelmæssige sæsonmønstre, der kan repræsenteres godt med trigonometriske funktioner. Denne metode kan være mere effektiv end sæsonmæssige dummyvariabler, når man håndterer lange sæsonperioder, da den kræver færre parametre.

Cykler:

Cykler bruges til at modellere mere komplekse og mindre regelmæssige fluktuationer i data, der ikke nødvendigvis er bundet til et fast sæsonmønster. Disse bruges ofte til at fange forretningscykler, økonomiske cykler eller andre cykliske fænomener, der ikke har en fast længde eller som varierer over tid. Cykler modelleres ved hjælp af stokastiske komponenter og kan variere i amplitude og frekvens over tid. Valget mellem disse metoder afhænger af arten af det sæsonmæssige mønster i dine data og det kompleksitetsniveau, du har brug for at modellere. For eksempel, hvis din sæsonmæssighed er meget regelmæssig og forudsigelig, kan sæsonmæssige harmonikker være mest passende. Hvis du har uregelmæssige eller skiftende sæsonmønstre, kan block seasonal eller cykler være bedre egnet. Sæsonmæssige dummyvariabler er ofte et godt udgangspunkt på grund af deres

enkelhed og lette fortolkning, især i tilfælde, hvor den sæsonmæssige effekt er klar og konsekvent.

3 PROC ESM

3.1 SIMPLE exponential smoothing

Jeg benytter metoden simple, fordi dataserien ikke indeholder nogen trend eller sæson. Ideen bag eksponentiel glatning er, at serien varierer omkring en glat kurve, der kunne betragtes som det sande, men uobserverede, niveau, som varierer over tid. De faktiske observationer bortset fra dette sande niveau består af uregelmæssigheder for hvert bestemt tidspunkt. Når det estimerede niveau betegnes med $\bar{X}_t$, er den grundlæggende formel for opdatering af niveauet:

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

hvor α er en glatningskonstant, $0 < \alpha < 1$, (Milhoj, 2013).

Forudsigelsen ud over den sidste observation X_T udføres ved at definere forudsigelsen som:

$$\hat{X}_{T+i} = X_T$$

for alle forudsigelser i . Dette producerer en konstant forudsigelse af alle fremtidige værdier, fordi der ikke er inkluderet ekstern information i algoritmen, (Milhoj, 2013).

3.2 Double exponential smoothing

Jeg benytter metoden double fordi mit dataset indeholder en trend men ingen sæson. Metoden er som følger: Værdierne udregnet fra den sidst observerede værdi af tidsserien, danner grundlag for forudsigelse af fremtidige værdier ved:

$$\hat{X}_{T+i} = \bar{X} + \left(\bar{X}_T - \bar{\bar{X}}_T \right) + i \times \frac{\alpha}{1 - \alpha} \left(\bar{X}_T - \bar{\bar{X}}_T \right)$$

hvor $\bar{\bar{X}}$ er den double udglattede tidsserie og $\bar{X}$ er blot den udglattede tidsserie. i er forudsigelses horisonten, (Milhoj, 2013).

3.3 LINEAR exponentiel smoothing

Den lineare metode er brugbar når min tidsserie indeholder en linear trend men ingen sæson. Den udvider den simple metode til at inkludere en trend.

3.4 Damptrend exponentiel smoothing

Jeg benytter damptrend når min tidsserie har en linear trend som jeg forventer ikke vil fortsætte i fremtiden, men at niveauet i fremtiden aftager.

3.5 Addseasonal or Seasonal exponentiel smoothing

Jeg benytter denne metode fordi mit dataset indeholder sæson men ingen trend, eller når sæson variationerne er relative konstante over tid. Det betyder at min data indeholder en additiv sæsonstruktur.

Ved at benytte ligningen ovenover for den simple metode af ESM, og inkludere en sæsonkomponent, 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)

3.6 Multiseasonal exponentiel smoothing

Jeg benytter denne metode fordi mit dataset indeholder sæson men ingen trend, og fordi jeg ser enten et fald eller stigning i sæson niveauet. Det betyder at min tidsserie indeholder en multipilkativ sæsonstruktur.

3.7 WINTERS exponentiel smoothing

Jeg benytter denne metode fordi mit dataset indeholder multiplikativ sæson og en trend.

3.8 ADDWINTERS exponentiel smoothing

Jeg benytter denne metode fordi mit dataset indeholder additiv sæson og en trend. Det er hvor sæson variationerne er konstante og ikke variere med niveauet.

4 Proc x12

Jeg benytter proc x12 til at lave sæson rensning. Formålet med sæsonkorrektion er at udlede værdien af tidsserien i en situation, hvor sæsonmæssighed ikke påvirker serien. Justerede værdier kan derefter bruges til at sammenligne den reelle markedssituation i f.eks. tredje og fjerde kvartal, selvom det er tydeligt, at salget altid er højere i fjerde kvartal end i tredje kvartal på grund af sæsonmæssighed. Ligeledes kan specifikke begivenheder i blot en enkelt observation opdages ved at kigge på den justerede serie, fordi de sædvanlige sæsonmæssige effekter er korrigeret væk.

4.1 mode = add

Ingen trend men sæson benyt denne

I SAS PROC X12, som er proceduren til sæsonjustering af tidsseriedata, refererer optionen mode=add til tilføjelsesmodellen (additive model), der anvendes, når der udføres sæsonjustering.

Når mode=add anvendes, antages det, at sæsonvariationerne er additive i stedet for multiplikative. Det betyder, at sæsonkomponenten lægges til (eller trækkes fra) tidsserien snarere end at blive multipliceret med den. Dette er passende, når sæsonvariationens størrelse ikke ændrer sig med niveauet af tidsserien. For eksempel, hvis en tidsserie repræsenterer antallet af parkerede biler, og der altid er 100 færre biler om vinteren uanset totalen, ville en additiv model være passende.

4.2 mode = mult

Hvis der er trend og sæson benytte denne

Alternativet til mode=add er mode=mult, som står for multiplikativ model. Denne model bruges, når sæsonvariationen ændrer sig proportionalt med tidsseriens niveau. For eksempel, hvis der er 10% færre parkerede biler om vinteren sammenlignet med sommeren, så ville en multiplikativ model være mere passende.

4.3 mode = logadd

Dataene transformeres først ved hjælp af den naturlige logaritme, hvilket ofte gøres for at stabilisere variansen og gøre en multiplikativ relation mellem komponenterne til en additiv relation i den log-transformerede skala.

Ved at anvende mode=logadd, kan man udnytte fordelene ved log-transformation til at stabilisere variansen samtidig med, at man bevare en additiv tilgang til de sæsonmæssige variationer. Dette kan være særligt nyttigt for tidsserier, hvor der er en multiplikativ effekt, men hvor størrelsen af sæsonvariationen ikke skalerer med niveauet af tidsserien.

4.4 PlotD11

"D11" plot viser, hvad der sker med tidsserien, når du trækker sæsonkomponenten fra for at isolere og undersøge de underliggende mønstre uden sæsonmæssig påvirkning. D11 plottet sammenlignes med den originale tidsserie, for at se om sæsonen har en effekt på tidsserien.

4.5 PlotD12

Plottet mærket som "Table D12" i PROC X12 viser den endelige trend-cykliske komponent af en tidsserie. I tidsserieanalyse er det vigtigt at adskille en tidsserie i dens forskellige komponenter for at forstå og forudsige fremtidige værdier mere præcist. "D12" viser trenden og cykliske variationer uden påvirkning af sæsonmæssige og uregelmæssige komponenter.

Trendkomponenten repræsenterer den langsigtede bevægelse i tidsserien og giver en glattet repræsentation af dataenes overordnede retning

Cykliske komponenter er svingninger i tidsserien, der ikke er af fast frekvens og som ofte er forbundet med økonomiske forhold, såsom konjunkturcyklusser. Disse kan vise udvidelser og sammentrækninger, der ikke nødvendigvis er bundet til en bestemt sæson.

4.6 Plot D10

I PROC X12-udgangen refererer "Table D10" til plottet, der 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 skyldes sæsonmæssighed.

Sæsonjusteringsfaktorerne er ofte præsenteret som et indeks, hvor tallet 1 repræsenterer baseline-niveauet uden nogen sæsonmæssig påvirkning. Værdier over 1 indikerer, at perioden typisk har en højere værdi end gennemsnittet på grund af sæsonmæssighed, mens værdier under 1 indikerer det modsatte.

4.7 Plot D13

Plottet markeret som "Table D13" i PROC X12-udgangen viser den endelige uregelmæssige eller "irregulære" komponent af tidsserien efter at have fjernet både den sæsonmæssige og den trend-cykliske komponent. Denne komponent repræsenterer de tilfældige fluktuationer, som ikke kan forklares af modellens systematiske dele (sæsonmæssighed og trend/cyklus).

Den irregulære komponent fanger de resterende variationer i tidsserien efter at sæson- og trend-komponenterne er blevet fjernet. Disse variationer kan skyldes tilfældige eller uforudsigelige begivenheder, såsom usædvanlige økonomiske chok, naturkatastrofer, eller andre ikke-repetitive påvirkninger.

I "D13"-plottet kan man forvente at se en vis grad af støj, som vil vise sig som op- og nedsvingninger omkring en central linje (typisk omkring 1, hvis de er normaliserede). Disse udsving er typisk uden et genkendeligt mønster eller systematik og er ofte mere volatile end de systematiske komponenter. Hvis plottet viser en tendens over tid eller gentagne mønstre, kan det indikere, at modellen ikke har fanget al sæsonmæssighed eller der er andre underliggende strukturer, som ikke er blevet modelleret.

4.8 Plot C17

"C17" plot viser de endelige vægte tildelt til den irregulære komponent i en tidsserieanalyse udført af PROC X12.

Observationer, der har en vægt på 1.0, er fuldt inkluderet i beregningen af den irregulære komponent. Det betyder, at disse data 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 c17-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. For eksempel, når en c17-værdi er omkring 0.2, betyder det, at den aktuelle observation kun bidrager med 20 procent, mens resterende 80 procent af justeringen afhænger af den estimerede sæsoncyklus og trendkomponent. Dette sker som en del af processen for at udglatte og justere data for den sidste gang, hvilket sikrer en mere præcis sæsonjustering.

Disse er observationer, der er helt udeladt fra beregningerne og kan indikere outliers eller fejlagtige værdier.

4.9 Ugedageeffekt og påskeeffekt

For at tage højde for ugedageeffekter i min tidsserieanalyse, har jeg anvendt 'td' (trading days) funktionen i PROC X12. Denne funktion genererer seks dummy-variable, som repræsenterer hver enkelt af ugens dage - mandag til lørdag. Søndagens effekt indregnes indirekte, da den er konstrueret således, at summen af koefficienterne for alle syv dage (mandag til søndag) samlet set bliver nul. Dette sikrer, at den samlede ugedageeffekt over en hel uge neutraliseres, hvilket gør det muligt at isolere og vurdere effekten af hver enkelt ugedag i analysen.

For at vurdere påskeeffekten i min tidsseriedata, benytter jeg metoden `td (easter(1))` i PROC X12. Denne metode bruges til at bestemme, om den måned, som påskesøndag falder i, har en signifikant indvirkning på min analyserede variabel. Det gør den ved at introducere en særlig variabel for påskeperioden, hvilket gør det muligt at isolere og evaluere påskens potentielle indflydelse på tidsserien. Derved kan jeg fastslå, om der er en statistisk signifikant forskel i dataene i den måned, hvor påsken finder sted.

5 Proc varmax

Equation	Parameter	Estimate	Standard Error	t Value	Pr > t	Variable
BEL	CONST1	0.09168	0.23154	0.40	0.6923	1
	AR1_1_1	1.45367	0.04647	31.28	< 0.0001	BEL(t-1)
	AR2_1_1	-0.45381	0.04661	-9.74	< 0.0001	BEL(t-2)

Table 1: Model Parameter Estimates

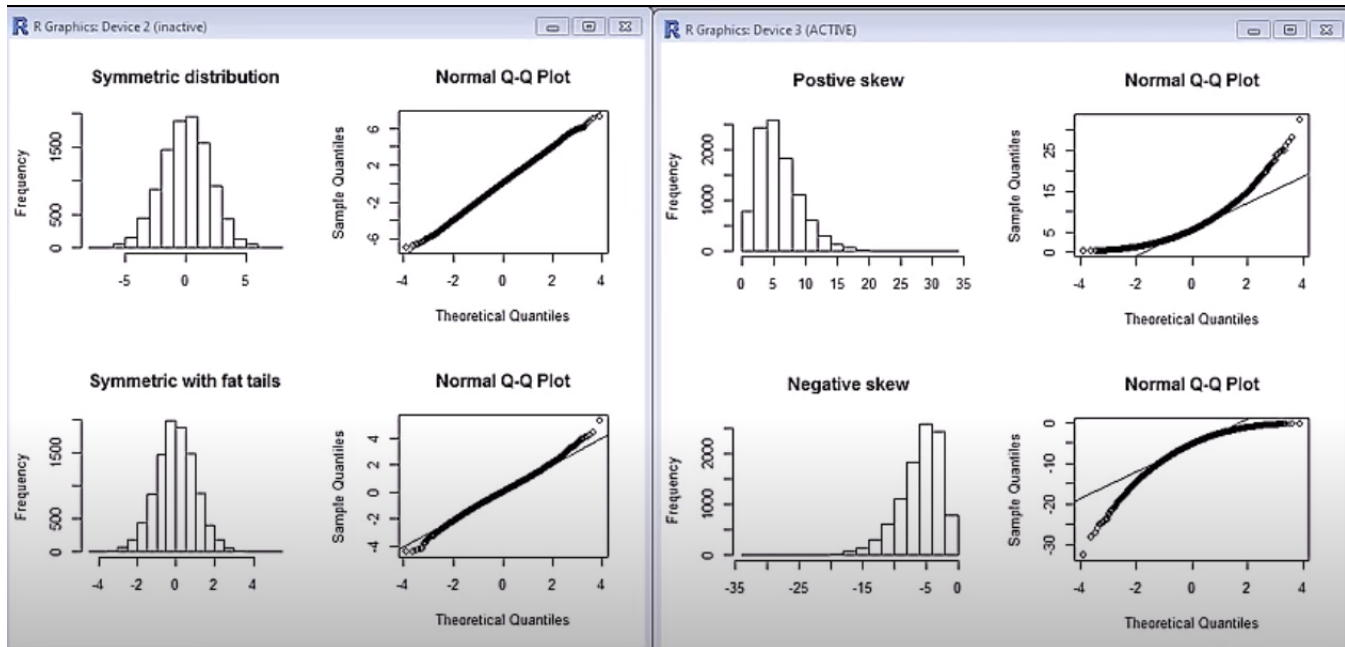
Benytter proc varmax som generer en AR(2,1,1) model med følgende estimator

$$\Delta BEL_t = 0.09168 + 1.45367 \cdot BEL_{t-1} - 0.45381 \cdot BEL_{t-2} + \epsilon_t \quad (1)$$

where:

- ΔBEL_t er den differentierede BEL serie ved tidspunkt t .
- 0.09168 er estimatet for konstant komponenten.
- 1.45367 er estimated for den første autoregressive koefficient som har en effekt på BEL ved tidspunkt $t-1$.
- -0.45381 er estimated for den anden autoregressive koefficient som har en effekt på BEL ved tidspunkt $t-2$.
- ϵ_t is the error term at time t .

Figure 1: QQ plot



Source: Own visualisation

Note: A more detailed version can be found on slide 14.

6 Q-Q plot

7 Sæson justering

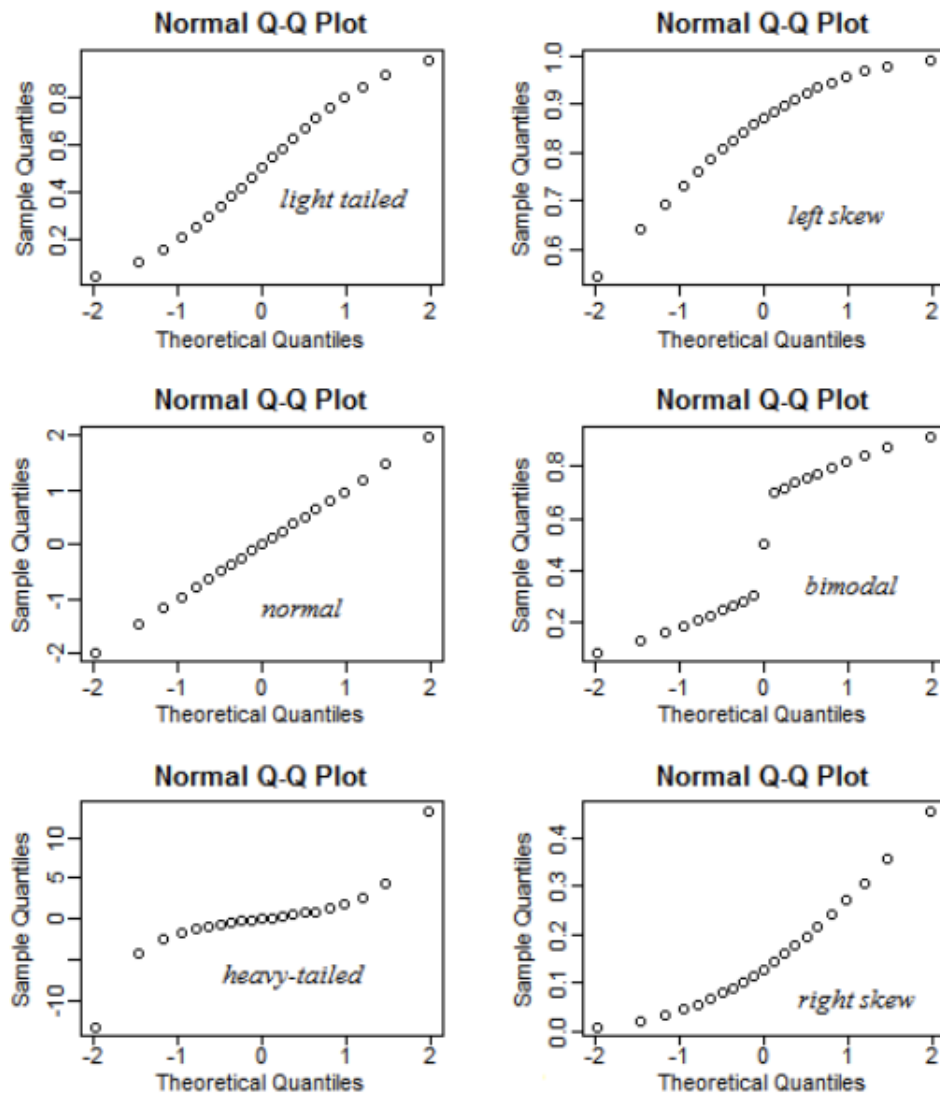
Formålet med sæsonjustering er at udlede værdien af tidsserien i en situation, hvor der ikke er sæsonmæssige påvirkninger. Sæsonmæssighed skyldes ofte meteorologiske forhold såsom temperatur og nedbør, hvilket påvirker næsten al menneskelig aktivitet. Religiøse og nationale festdage kan også påvirke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

I den sæsonjusterede serie 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. Dernæst betragt følgende formel, hvor den originale tidsserie X_t er summen af disse tre komponenter;

$$X_t = TC_t + S_t + I_t$$

Komponenten TC_t er den trend cykliske komponent, hvilket svarer til tidsseriens niveau. Denne

Figure 2: Impact Implementation Plan

Source: Own visualisation

Note: A more detailed version can be found on slide 14.

komponent er beregnet til at omfatte både trends og konjunkturcykler, hvilket er nødvendigt eftersom at mange tidsserier indeholder trends, (Milhoj, 2013).

Komponenten I_t repræsenterer resttermen, der indeholder alle elementer, der udelukkende er relateret til den nuværende observation for den specifikke måned i serien for det pågældende år. Denne komponent kan ses som den unormale del af observationen. Den er ikke en integreret del af seriens faktiske niveau og er ikke til at forklare ved seriens sæsonbetonede opførsel.

Sæsonkomponenten S_t kan betragtes som en komponent, der afspejler en specifik effekt, der forekommer i den samme måned hvert år i tidsserien. I denne notation indikerer t primært måneden, hvilket indebærer, at S_t kun bør have tolv forskellige værdier for månedlige observationer, og den bør gentage den samme værdi (som for eksempel for måneden april) år efter år.

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

8 Proc expand

PROC EXPAND interpolerer manglende observationer i tidsserier ved brug af enkle metoder til kurvetilpasning. En simpel mulighed er at forbinde punkterne i et diagram af de tilgængelige observerede værdier med lige linjer og derefter finde de manglende værdier på disse forbindelseslinjer.

8.1 method join

Denne metode forbinder observerede data med lige linjer. Det betyder, at manglende værdier mellem to kendte punkter estimeres ved at tegne en lige linje mellem disse punkter og vælge punktet på linjen, der svarer til det manglende tidspunkt.

Den er hurtig og enkel men kan skabe skarpe hjørner i data, som ikke altid er realistiske i glatte tidsserier.

8.2 method Spline Interpolation

Spline-interpolation er en mere avanceret teknik, der skaber en glat kurve gennem de observerede data. Den bruger matematiske funktioner, kaldet splines, som typisk 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.

8.3 method step

Denne metode skaber en trappetrinsfunktion, hvor data holdes konstant indtil næste observation.

Det er nyttigt, når data repræsenterer niveauer, der skifter på specifikke tidspunkter.

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.