SAR- og CAR-modellerne
I teori og praksis

Helle Jakobsen

Speciale 10. semester

Institut for Matematiske Fag
Aalborg Universitet

Abstract

Regular linear models are usually applied when a data set consists of observations which are influenced by changes in the appurtenant independent variables. In reality many data sets consist of observations which are influenced by changes in the independent variables of neighbor observations in which case the observations are referred to as spatial dependent. Two econometric models can be applied to spatial dependent and approximately normally distributed observations: The Spatial Autoregressive Model also referred to as the SAR-model and the Conditional Autoregressive Model known as the CAR-model. This paper describes the theory behind each model and gives two methods for estimating the parameters of the models. The first method is maximum likelihood estimation and the second is the Bayesian approach where the latter of the two incorporate prior knowledge of the parameters. Most of the theory is translated into practice using data from the real estate agent HOME.

Institut for Matematiske Fag
Aalborg Universitet
Fredrik Bajers Vej 7G
9220 Aalborg Ø
http://www.math.aau.dk

Titel: SAR- og CAR-modellerne
i teori og praksis

Tema:

Rumlige, økonometriske modeller

Projektperiode:
Forårssemestret 2013

Projektgruppe:
G4-109a

Deltager:

Helle Jakobsen

Vejleder:
Jakob Gulddahl Rasmussen

Synopsis:

Almindelige, lineære modeller anvendes til at beskrive fordelinger, hvor en observation påvirkes af ændringer i sine tilhørende baggrundsvariable. Der findes dog mange eksempler på mængder af observationer fra virkeligheden, hvor observationerne afhænger af hverandre og derfor påvirkes af en ændring i en anden observations baggrundsvariable. To modeller, der beskriver rumlige observationer, som er indbyrdes afhængigt, er den rumlige autoregressive model, SAR-modellen, og den betingede autoregressive model, CAR-modellen. I denne rapport vil teorien bag hver model blive beskrevet, hvorefter der gennemgås to metoder til at estimere parametrene i hver model i form af maximum likelihood-estimation og bayesiansk parameterestimering. Til sidst vil der blive givet et praktisk eksempel på, hvordan der kan estimeres en SAR- og CAR-model med henholdsvis maximum likelihood-estimation og den bayesianske metode, hvor der anvendes et datasæt med huspriser fra ejendomsmægleren HOME.

Oplagstal: 4

Sideantal: 139

Afsluttet den 06-06-2013

Forord

Dette speciale er udarbejdet af Helle Jakobsen i foråret 2013 ved Institut for Matematiske Fag på Aalborg Universitet. Specialet bygger videre på rapporten „Rumlig økonometri i teori og praksis“, der blev forfattet af Lea Nørgreen Gustafsson og undertegnede i efteråret 2012, og som udgjorde Lea Nørgreen Gustafssons speciale, der kun var halvårligt. Da vi ønskede at skrive sammen på henholdsvis Leas 11. semester og mit 9. semester, blev rapporten til hendes speciale og mit 9. semesters projekt. Det var naturligt for mig at arbejde videre med teorien fra rapporten „Rumlig økonometri i teori og praksis“ i mit speciale, som derfor indeholder dele af rapporten i omskrevet form, hvilket drejer sig om:

- Indledningen i kapitel 1, der medtager pointer fra indledningen i rapporten „Rumlig økonometri i teori og praksis“.
- Teorien i kapitel 2 om SAR-modellen var også med i rapporten „Rumlig økonometri i teori og praksis“ i omskrevet form.
- Afsnit 4.2 om påvirkninger mellem observationer er udvidet til også at indeholde CAR-modellen i forhold til det lignende afsnit i „Rumlig økonometri i teori og praksis“.
- Afsnit 5.1 og 5.2 indgik også i teorien om maximum likelihood-estimation i „Rumlig økonometri i teori og praksis“ og er medtaget i omskrevet form.
- Afsnit 6.1 var også med i „Rumlig økonometri i teori og praksis“ i omskrevet form, ligesom teorien om Metropolis-Hastings og MCMC-algoritmen i afsnit 6.2. Bayesiansk estimering af modelparametrene i SAR-modellen fra afsnit 6.4 var også en del af teorien i „Rumlig økonometri i teori og praksis“.
- Størstedelen af regneregler og fordelinger i Appendiks er også med i „Rumlig økonometri i teori og praksis“.

Datasættet fra ejendomsmægleren HOME blev også anvendt i databehandlingen i „Rumlig økonometri i teori og praksis“, men i denne rapport er datasættet opdelt på en ny måde og al databehandlingen er derfor „ny“, selvom der også estimeres en SAR-model i denne rapport.

Rapporten er henvendt til den interesserede læser, som forventes at have et grundlæggende kendskab til sandsynlighedsregning og derudover have gennemført statistikkurser på universitetsniveau, da basale, statistiske begreber ikke vil blive gennemgået i detaljer i rapporten.

Der rettes en stor tak til vejleder Jakob Gulddahl Rasmussen for inspirerende vejledning og for konstruktiv kritik gennem hele projektforløbet.

Læsevejledning

Rapporten består af otte kapitler, der tilsammen beskriver teorien bag de to, rumlige modeller SAR- og CAR-modellen, teori til at afbilde rumlighed og teori til at estimere parametrene i de fundne modeller, hvor der til slut gives et praktisk eksempel på at anvende dele af teorien i praksis.

Kapitel 1 indeholder indledningen, der giver et kort, historisk overblik over rumlig økonometri som disciplin og en overordnet beskrivelse af rumlig afhængighed i en mængde observeret data.

Kapitel 2 gennemgår teorien for den rumlige autogregressive model, der forkortes SAR-modellen, og hvad de forskellige parametre i modellen angiver.

Kapitel 3 introducerer den betingede autoregressive model, som forkortes CAR-modellen, hvor der også gennemgås en del grafteori om Markovfelter og -egenskaber, da CAR-modellen bygger på denne teori.

Kapitel 4 undersøger rumlig afhængighed mellem punkter, blandt andet ved at give en praktisk metode til at definere en mængde observationers naborelationer ved brug af et Voronoi-diagram, og der ses også nærmere på, hvordan SAR- og CAR-modellen afbilder påvirkningerne mellem observationerne.

Kapitel 5 introducerer metoden maximum likelihood-estimation til at finde parameterestimater i en givet model, og metoden uddybes for både SAR- og CAR-modellen.

Kapitel 6 introducerer den bayesianske metode til at finde parameterestimater for en model, hvor der i modsætning til maximum likelihood-estimation kan medtages forhåndsviden om modellens parametre. Der ses også på simulering af parametre for henholdsvis SAR- og CAR-modellen ved brug af en Markov Chain Monte Carlo-algoritme, der kan indeholde både Gibbs-sampling og Metropolis Hastings-sampling.

Kapitel 7 giver et eksempel på, hvordan dele af teorien fra rapporten kan anvendes i praksis ved at finde en SAR- og CAR-model for et datasæt fra ejendomsmægleren HOME, hvor der anvendes både maximum likelihood-estimation og bayesiansk parameterestimering til at finde parametrene i de to modeltyper ved brug af programmet R.

Kapitel 8 indeholder konklusionen, der samler op på rapportens indhold.

Notation

Kildehenvisninger i rapporten er angivet på formen [forfatter(e), årstal (, placering i kilden)], eksempelvis [LeSage and Pace, 2009] eller [LeSage and Pace, 2009, kap. 2].

Definitioner, sætninger, lemmaer og eksempler er nummereret ved X.Y, hvor X er nummeret på det aktuelle kapitel, og Y er et fortløbende heltal, der starter med 1 i hvert kapitel. Ligninger er nummereret på samme måde, dog med parenteser om henvisningen, (X.Y). Henvisninger til definitioner, sætninger, eksempler og kode fra R skrives eksempelvis ved „sætning X.Y“, mens henvisninger til ligninger for det meste bare skrives ved „(X.Y)“.

Der benyttes desuden følgende notation i rapporten:

- Beviser afsluttes med ■, og eksempler afsluttes med □.
- Vektorer angives med fed skrift, eksempelvis $\mathbf{x}$, og matricer med stort bogstav som X.
- Modelparametre angives med græske bogstaver som α, β og ρ.

Indholdsfortegnelse

Indledning 1

Økonometri blev fastlagt som en videnskabelig disciplin omkring 1930 og fokuserede i starten mest på makroøkonomiske modeller til at hjælpe regeringer og store firmaer med at træffe langsigtede, økonomiske beslutninger. Økonometri har siden da udviklet sig til et effektivt værktøj til at modellere empiri fra de fleste områder inden for økonomi og erhverv. Økonometriske modeller sammenfatter den relevante information fra store datasæt og hjælper med at forstå sammenhængen mellem de økonomiske variable, så det er muligt at analysere potentielle effekter af en beslutning, inden den tages. Senere kom der fokus på at udforske økonometriske modeller, der tog højde for rumlig afhængighed mellem observationerne i et datasæt, og det er to af disse modeller, der vil blive undersøgt nærmere i denne rapport.

Først vil teorien bag de to rumlige, økonometriske modeller SAR-modellen (Spatial Autoregressive model) og CAR-modellen (Conditional Autoregressive model) blive undersøgt, og der vil også blive set på to forskellige metoder til at estimere parametrene i hver model. Til sidst vil der blive givet et eksempel på, hvordan modellerne kan anvendes i praksis til at afbilde et datasæt med huspriser fra ejendomsmægleren HOME, og her vil der også blive set på, hvor præcist hver model afbilder datasættet.

1.1 Rumlig økonometri

Afsnittet tager udgangspunkt i [Anselin, 2010].

Tilbage i i 1960'erne og 1970'erne var rumlig økonometri kun en marginal disciplin, men siden da er den vokset til at være en del af mainstream i dag og et velanset felt inden for anvendt økonometri, hvilket også kan ses ved, at mængden af litteratur, som udgives om emnet, er vokset. En grund, til rumlige modeller har været længe om at blive en del af økonometri, kan være, at rumlige modeller hovedsagligt har haft fokus på geografisk modellering, som det har været vanskeligt at overføre til brugbare teknikker til anvendt rumlig økonometri. Derudover har en forbedret teknologi og mere tilgængelige beregningsmetoder også øget anvendelsen af rumlige modeller, mens et generelt skift i teoretiske perspektiver har fået flere til at interessere sig for naborelationer mellem observationer.

Rumlig økonometri kan defineres som økonometriske metoder, som indeholder rumlige aspekter, der kan findes i simultane observationer defineret af både tid og rum. Her indgår variable, som viser observationernes placeringer i forhold til hverandre, og disse variable indgår også som en særskilt del af modelspecifikation, -estimation og -test samt forecasting. Anvendelsen af rumlig økonometri til at træffe økonomiske beslutninger givet en mængde kvantitativ data kan inddeles i fire trin, der er givet i det følgende.

1. Modelspecifikation

Modelspecifikation går ud på at finde det formelle, matematiske udtryk for rumlig afhængighed og rumlig ensartethed i regressionsmodeller, som anvendes til at beskrive rumlig data. Rumlighed kan adskilles i to overordnede former for rumlige effekter, der er rumlig afhængighed og rumlig ensartethed.

Rumlig afhængighed kan opfattes som en form for afhængighed mellem simultane observationer ved, at observationernes placering i forhold til hverandre inklusive afstande mellem observationer og grupperinger af samme bestemmer korrelationsstrukturen mellem stokastiske variable tilknyttet observationerne. De teknikker, der anvendes til at modellere rumlig afhængighed, er mere avancerede end bare at udvide teknikkerne fra autoregressive processer til to dimensioner. Modeller for rumlig afhængighed indeholder ofte rumligt laggede variable, som er vægtede gennemsnit af værdier for en observations naboliggende observationer. Afhængigheden mellem de rumlige observationer angives som regel i form af naborelationer i en matrix, og der kan også haves rumligt laggede variable for de tilhørende baggrundsvariable eller fejlled, eller kombinationer af samme.

Rumlig ensartethed er, når naboliggende observationer påvirkes i samme retning af udefrakommende påvirkninger, og at de påvirker hinanden på en sådan måde, at de tilnærmelsesvist ensrettes, hvor begge dele kan udnyttes i en modelspecifikation. Rumlig ensartethed kan enten indgå som direkte led i modellen, eller den kan være til stede i data uden at indgå som eksplicit parameter i modellen. I modsætning til rumlig afhængighed er det ikke altid nødvendigt med separate metoder til at behandle rumlig ensartethed, der matematisk kan udtrykkes på flere måder. Rumlig ensartethed kan klassificeres som enten diskret eller kontinuert, hvor diskret ensartethed viser sig ved en givet mængde af rumligt adskilte enheder, som modelparametre må variere indenfor, mens kontinuert ensartethed præciserer, hvordan regressionsparametrene varierer i rummet, hvor de enten kan være givet på forhånd eller være bestemt ud fra data ved parameterestimering.

2. Modelestimering

Når de rumlige effekter er inddraget og præciseret i en regressionsmodel, estimeres parametrene ud fra metoder, der tager højde for den simultanitet, som den rumlige afhængighed medfører. Her kan eksempelvis anvendes maximum likelihood-estimation eller bayesianske estimeringsmetoder, der inddrager eventuel forhåndsviden om parametrene i modellen.

3. Modeltest

Når der er fundet estimater for modellens parametre, undersøges det, i hvilken grad der er afvigelser mellem det givne data og den estimerede model, for at se om der eventuelt er behov for en anden type model. En model kan testes på mange måder, eksempelvis ved brug af likelihood ratio-test, hvis estimaterne er fundet ved brug af mixaimum likelihood-estimation, og der kan også anvendes residualplot, der afbilder afvigelserne mellem data og den estimerede model.

4. Rumlig forecasting

Når der haves en rumlig, økonometrisk model, der viser sammenhængen mellem de forskellige variable i et datasæt, er det muligt at foretage forecasting, hvilket er at anvende modellen til at komme med et bud på fremtidige observationer. I makroøkonomisk perspektiv kan forecasting give et bud på effekten af ændringer i økonometriske variable som eksempelvis skattetryk.

1.2 Rumlige data

Afsnittet er skrevet ud fra [Wall, 2003] og [Besag, 1974].

Hvis en mængde observationer er spredt over et geografisk område, hvilket eksempelvis kan være huspriser placeret i forskellige postnumre i Danmark, er det ikke altid muligt at inddrage alle de faktorer, som kan påvirke observationerne, som variable i en model. Som regel kan det dog antages, at de udeladte variable ligner hinanden, da eksempelvis huspriser i et område i reglen påvirkes af huspriserne i de naboliggende områder, hvorved fejlleddene vil blive rumligt autokorrelerede.

Observationer kan fordele sig i rummet på mange måder, og systemer af rumlige observationer kan derfor klassificeres i forhold til

1. om systemets punkter er regulære eller irregulære,
2. om hvert enkelt punkt betegner punkter eller regioner,
3. om de tilhørende, stokastiske variable er diskrete eller kontinuerte.

Når et datasæt analyseres, er det vigtigt at være opmærksom på, hvilket af ovenstående tilfælde eller kombinationer af tilfælde, datasættet hører til, da der er forskel på at arbejde med diskrete punkter med præcist angivne placeringer eller om der haves data, som er samlet over et område, og dermed ikke er tilknyttet et enkelt, geografisk punkt. I praksis haves der som regel kombinationer af de tre punkter herover, som for eksempel et regulært gitter med regioner som punkter, der har diskrete variable. Et praktisk eksempel er, når der haves stikprøver fra en irregulært fordelt befolkning, hvor stikprøverne tages ved at inddele et område i et rektangulært gitter og så tælle hvor mange individer, der er i hvert kvadrat.

Det er muligt at opsummere en datamængdes struktur i en generel, rumlig proces $\{Z(s)|s \in \mathcal{D}\}$. To specialtilfælde heraf er rumlige processer for data, der danner et gitter, hvor forskellen for de to metoder til at afbilde data ses i antagelserne om datamængden $\mathcal{D}$.

Den første metode antager, at mængden $\mathcal{D}$, som observationerne stammer fra, er kontinuert. Det er her muligt at inddele planen i områder ved brug af et Voronoi-diagram og summere over observationerne i hvert område, som knyttes til et geografisk punkt midt i området. Her kan der eksempelvis summeres over huspriser i en kommune, og afstanden mellem de forskellige centrer for kommunerne kan anvendes til at definere, om to kommuner er naboer. Et problem ved metoden er, at det er forsimplet i forhold til

virkeligheden at placere alle observationerne i centrum af et område, og derudover er det et problem, at det ikke er praktisk muligt at måle data kontinuert, så en kontinuer model vil ende med at blive en tilnærmet model for diskrete data.

Den anden metode antager, at $\mathcal{D}$ er en diskret mængde, og denne metode vil blive anvendt i kapitel 2 og kapitel 3. Naboer defineres her til at være områder, hvor grænserne rører hinanden, hvorved det er muligt at bestemme en autoregressiv model. To modeller, som tager højde for det diskrete mønster i data, er SAR-modellen (Spatial Autoregressive Model) og CAR-modellen (Conditionally Autoregressive Model). Modellerne blev oprindeligt udviklet til at modellere endelige, regulære gitre, hvorved de er i familie med den velkendte, autoregressive proces. Et gitter er irregulært, når observationerne ikke har samme antal naboer eller samme afstand til hinanden, og hvis SAR- eller CAR-modellen anvendes på et endeligt, irregulært gitter, kan modellerne antyde kovariansens struktur, som viser, at der ikke er konstant varians for observationer, som er samme antal naboer fra hinanden.

SAR-modellen 2

Kapitlet er skrevet med udgangspunkt i [Heij et al., 2004, s. 538-539], [Horvath and Johnston], [LeSage and Pace, 2009, kapitel 2] og [Wall, 2003].

En af de modeller, der kan anvendes til at beskrive et datasæt med rumligt afhængige observationer, er en model for en rumlig, autoregressiv proces, der forkortes SAR-modellen (Spatial Autoregressive Model). Overordnet set er det den rumlige version af den almindelige, autoregressive proces $AR(p)$, der kendes fra økonometri og har formen

$$y_t = \alpha + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots + \varphi_p y_{t-p} + \varepsilon_t.$$

Her er y_t værdien for en tidsserie til tiden t, ε er et fejlled med $\mathrm{E}[\varepsilon y_{t-k}] = 0$ for $k \geq 1$ og α samt φ_i, $i = 1,2,\ldots,p$ er ukendte parametre. Derudover er $t = p+1, p+2, \ldots, p+n$, hvor p er den autoregressive proces' orden, hvilket svarer til antallet af led, processen medtager.

Inden der ses på den rumlige udgave af processen $AR(p)$, er det relevant at se på en definition af en datamængde $\mathcal{D}$, der består af rumlige punkter. Eksempelvis kan en mængde huspriser tilknyttes koordinatsættene for husenes beliggenheder, hvorved der haves et rumligt datasæt med en tilhørende mængde af relevante baggrundsvariable som boligareal, afstand til dagligvarer og lignende. Hver observation fra datamængden kan ses som et diskret punkt i et såkaldt gitter $\mathcal{D}$.

Definition 2.1 (Gitter):
Et gitter er en datamængde $\mathcal{D} = \{A_1, \ldots, A_n\}$, hvor elementerne i $\{Z(A_i) | A_i \in (A_1, \ldots, A_n)\}$ er normalfordelte, og derudover må $\{A_1, \ldots, A_n\}$ opfylde, at

$$A_1 \cup A_2 \cup \ldots \cup A_n = \mathcal{D}$$
$$A_i \cap A_j = 0 \quad \forall \quad j \neq i.$$

Det vil sige, at et gitter er en mængde af observationer, som det er muligt at inddele i disjunkte delmængder. SAR-modellen er en rumlig model for et gitter, da den simultane fordeling for $\mathbf{Z} = \begin{bmatrix} Z(A_1) & Z(A_2) & \ldots & Z(A_n) \end{bmatrix}^T$ er multivariat normalfordelt i SAR-modellen, hvilket uddybes i næste afsnit.

2.1 SAR-modellen

Som nævnt ovenfor er SAR-modellen en rumlig, økonometrisk model for en datamængde, der kan beskrives med en rumlig, autoregressiv proces. SAR-modellen er en mulig

normalfordeling for en endelig mængde $\mathbf{Z} = \begin{bmatrix} Z(A_1) & Z(A_2) & \dots & Z(A_n) \end{bmatrix}^T$, der er defineret i definition 2.1, og kan skrives som

$$Z(A_i) = \mu_i + \sum_{j=1}^{n} \rho w_{ij} \left(Z\left(A_j\right) - \mu_j \right) + \varepsilon_i \tag{2.1}$$
$$\varepsilon_i \sim N\left(0, \sigma^2\right).$$

Her er $\mu_i = \mathrm{E}[Z(A_i)]$ for $i = 1, \dots, n$, mens produktet ρw_{ij} består af kendte eller ukendte konstanter, hvor $w_{ii} = 0$ for alle $i = 1, \dots, n$ observationer. SAR-modellen kaldes også simultan, da fejlleddet ε_i og $Z(A_i)$, $j \neq i\}$, har samme indeks i (2.1), hvilket gør, at de to led er korrelerede. Hvis der haves et endeligt antal observationer $i = 1,2,\dots,n$ i (2.1), er det muligt at definere en matrix W bestående af alle elementerne w_{ij}, således at $W = \left(w_{ij}\right)$. Gitteret $\mathcal{D}$ viser, hvordan observationerne er fordelt, og afgør dermed også strukturen for matricen W, der også kaldes nabomatricen.

Definition 2.2 (Nabomatrix W):
Elementerne $\left(w_{ij}\right)$ i matricen W er givet ved

$w_{ij} = 1$ hvis A_i har en fælles grænse med A_j
$w_{ij} = 0$ hvis $i = j$
$w_{ij} = 0$ ellers.

Her er A_i og A_j delmængder af datamængden $\mathcal{D} = \{A_1, \dots, A_n\}$.

Matricen W går både under navnet nabomatrix og vægtmatrix, og det er altid muligt at konstruere en rækkestokastisk udgave af W, hvor hver række har summen én. Det kan gøres ved hjælp af en invertibel matrix eller ved at definere $W = \left(w_{ij}^*\right)$, hvor

$$w_{ij}^* = \frac{w_{ij}}{w_{i+}} \tag{2.2}$$
$$w_{i+} = \sum_{k=1}^{n} w_{ik}.$$

Her står w_{i+} for, at der summerres over alle ettallerne i j'te række i W, og $W = \left(w_{ij}^*\right)$ vil dermed resultere i en rækkestokastisk matrix, medmindre der haves en række uden ettaller i den oprindelige W. Hvis W indeholder en række kun med nuller, betyder det, at en delmængde A_i er fuldstændig isoleret fra resten af delmængderne A_j, $j = 1,2,\dots,n$, og dermed ikke er en del af det rumligt afhængige datasæt, da den tilknyttede observation $Z(A_i)$ ikke afhænger af andre end sine egne baggrundsvariable. Det kan overvejes at udelade isolerede observationer, da de alligevel ikke har indflydelse på resten af observationerne eller selv påvirkes deraf, og derfor muligvis vil blive bedre afbildet med en individuel fordeling. Derudover vil det også være muligt at konstruere en rækkestokastisk matrix W for de indbyrdes afhængige observationer, som i mange tilfælde er nemmere at regne med end versionen, der ikke er rækkestokastisk.

Der vil nu blive givet et eksempel på konstruktion af en rækkestokastisk nabomatrix W.

Eksempel 2.3:
Eksemplet tager udgangspunkt i [LeSage and Pace, 2009, s. 8-10].

1. januar 2007 blev Danmark inddelt i fem regioner, som er Region Nordjylland, Region Midtjylland, Region Syddanmark, Region Sjælland og Region Hovedstaden, der er vist på figur 2.1.

Figur 2.1: Regioner i Danmark fra [A/S].

For at kunne definere naborelationerne mellem regionerne, antages det først, at naboregioner er defineret som regioner med indbyrdes vej-, bro eller færgeforbindelse. Det er tydeligt, at Region Midtjylland er nabo til Region Nordjylland, og da der går en direkte færgeforbindelse mellem Kalundborg og Århus, som er den eneste færgeforbindelse, der betragtes i dette eksempel, har Region Midtjylland tre naboer i form af Region Nordjylland, Region Syddanmark og Region Sjælland. Disse kaldes førsteordensnaboer. Det er nu muligt at lave en (5×5)-vægtmatrix W over naborelationer mellem de fem regioner, hvilket giver

$$W = \begin{bmatrix} & No & Mi & Sy & Sj & Ho \\ No & 0 & 1 & 0 & 0 & 0 \\ Mi & 1 & 0 & 1 & 1 & 0 \\ Sy & 0 & 1 & 0 & 1 & 0 \\ Sj & 0 & 1 & 1 & 0 & 1 \\ Ho & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Hvis to regioner tilhørende den (i,j)'te plads i matricen er førsteordensnaboer, angives det med $w_{ij} = 1$ og ellers noteres et 0. Her ses det blandt andet, at Region Hovedstaden kun er nabo til Region Sjælland, hvilket kan aflæses både ud for søjlen og rækken Ho. Det bemærkes desuden, at alle indgangene i diagonalen i W er nul, idet ingen region er førsteordensnabo til sig selv. Matricen W kan nu gøres rækkestokastisk ved brug af

ligning (2.2), hvilket giver

$$W = \begin{bmatrix} & No & Mi & Sy & Sj & Ho \\ No & 0 & 1 & 0 & 0 & 0 \\ Mi & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ Sy & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ Sj & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ Ho & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Nabomatricen W kan også anvendes til at finde anden- og højereordensnaboer blandt observationerne. Eksempelvis er andenordensnaboerne til Region Nordjylland givet ved Region Midtjyllands naboer, der er Region Syddanmark, Region Sjælland og Region Nordjylland selv. Andenordensnaboerne findes ved at udregne W^2, der her er

$$W^2 = \begin{bmatrix} & No & Mi & Sy & Sj & Ho \\ No & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ Mi & 0 & \frac{11}{18} & \frac{1}{9} & \frac{1}{6} & \frac{1}{9} \\ Sy & \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ Sj & \frac{1}{9} & \frac{1}{6} & \frac{1}{9} & \frac{11}{18} & 0 \\ Ho & 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \end{bmatrix}.$$

Det ses at diagonalen i W^2 ikke længere er nul, som den var i W, da alle observationer er nabo til deres egne naboer. På samme måde kan tredjeordensnaboer findes som W^3 og så videre. □

Det er nu blevet vist, hvordan parametrene w_{ij}, $i,j = 1,\ldots,n$, fra udtrykket (2.1) kan sammenfattes i en matrix W, og det er også muligt at skrive resten af udtrykket (2.1) på matrix-form. Mængden $\mathbf{Z} = \begin{bmatrix} Z(A_1) & Z(A_2) & \ldots & Z(A_n) \end{bmatrix}^T$ kan skrives som vektoren $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \ldots & y_n \end{bmatrix}^T$, der indeholder det observerede data, og da $Z(A_i)$ jævnfør definition 2.1 er normalfordelt, vil $\mathbf{y}$ følge en multivariat normalfordeling (se appendiks B.1)

$$\begin{aligned} &\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \Sigma) \\ &\boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_2 & \ldots & \mu_n \end{bmatrix}^T. \end{aligned} \tag{2.3}$$

Udtrykket (2.1) kan omskrives til matrixform ved først at sætte $y_i = Z(A_i)$, så

$$y_i = \mu_i + \sum_{j=1}^{n} \rho w_{ij}(y_j - \mu_j) + \varepsilon_i.$$

Det er muligt at skrive summen $\sum_{j=1}^{n} w_{ij}(y_j - \mu_j)$ som en vektor for hver observation $i = 1,\ldots,n$, så

$$\begin{bmatrix} \sum_{j=1}^{n} w_{1j}(y_j - \mu_j) \\ \vdots \\ \sum_{j=1}^{n} w_{nj}(y_j - \mu_j) \end{bmatrix} = \begin{bmatrix} w_{11} & \ldots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \ldots & w_{nn} \end{bmatrix} \begin{bmatrix} y_1 - \mu_1 \\ \vdots \\ y_n - \mu_n \end{bmatrix} = W(\mathbf{y} - \boldsymbol{\mu}).$$

Matrixformen af (2.1) bliver derfor

$$\begin{aligned}\mathbf{y} &= \boldsymbol{\mu} + \rho W\left(\mathbf{y} - \boldsymbol{\mu}\right) + \boldsymbol{\varepsilon} \\ &= \boldsymbol{\mu} + \rho W\mathbf{y} - \rho W\boldsymbol{\mu} + \boldsymbol{\varepsilon} \\ &= \rho W\mathbf{y} + \left(I_n - \rho W\right)\boldsymbol{\mu} + \boldsymbol{\varepsilon}.\end{aligned}$$

Hvis det defineres at $\boldsymbol{\mu} = \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}$, hvor $\boldsymbol{\beta}$ er en parametervektor, og X er en matrix, bliver ovenstående til

$$\begin{aligned}\mathbf{y} &= \rho W\mathbf{y} + \left(I_n - \rho W\right)\left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \rho W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}.\end{aligned}$$

Denne måde at skrive SAR-modellen på er anvendt i definition 2.4, og den vil brugt i resten af rapporten.

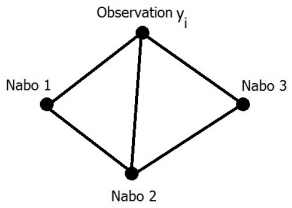
Definition 2.4 (SAR-modellen - Spatial Autoregressiv Model):
Den rumlige, autoregressive model, SAR-modellen, er givet ved

$$\begin{aligned}\mathbf{y} &= \rho W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N_n\left(\mathbf{0}, \sigma^2 I_n\right),\end{aligned}$$

hvor ρ er en parameter, som angiver graden af rumlig afhængighed mellem de n observationer i vektoren $\mathbf{y}$, mens den rækkestokastiske $(n \times n)$-matrix W angiver naborelationerne mellem observationerne i $\mathbf{y}$. Matricen X indeholder observationernes baggrundsvariable, og $\boldsymbol{\beta}$ er den tilhørende parametervektor. Fejlvektoren $\boldsymbol{\varepsilon}$ følger en n-dimensional, multivariat normalfordeling, der har en $(n \times 1)$-nulvektor som middelværdi og en variansmatrix givet ved $\sigma^2 I_n$.

Den rumlige afhængighed i SAR-modellen indgår via leddet $\rho W\mathbf{y}$, der gør $\mathbf{y}$ til en funktion af sine egne naboer, som er defineret i matricen W. Her kaldes faktoren $W\mathbf{y}$ for den rumlige lagvektor, og hver indgang heri består af linearkombinationer af naboliggende observationer fra $\mathbf{y}$. Hvis der haves en mængde observationer $\mathbf{y}$ med baggrundsvariable X, kan en passende SAR-model laves ved at estimere parametrene $\boldsymbol{\beta}$, σ^2 og ρ, hvilket der vil blive set på metoder til senere i rapporten.

Parameteren ρ i definition 2.4 beskriver graden af afhængighed mellem naboobservationer, og sættes ρ sammen med nabomatricen W, haves den reelle afhængighed mellem observationerne. Hvis der haves en mængde normalfordelte observationer $\mathbf{y}$, angiver en positiv værdi for ρ, at der er positiv korrelation mellem naboobservationer, så naboliggende observationer både ligner og påvirker hinanden i samme retning. Hvis parameteren $\rho = 0$, betyder det, at der ingen korrelation er mellem naboobservationer, hvormed naboobservationerne vil være uafhængige af hinanden. Hvis der derimod gælder, at $\rho < 0$, så er der ofte negativ korrelation mellem naboobservationer, men ved komplicerede naborelationer, kan det være vanskeligt at fortolke den egentlige betydning af $\rho < 0$. Et eksempel er, hvis en observation y_i har tre naboer, hvor to af naboerne også er naboer til y_i's tredje nabo, hvilket er illustreret på figur 2.1.



Figur 2.2: En observation y_i med naboer, hvor fortolkning af ρ ikke er entydig.

Negativ korrelation betyder, at hvis observationen y_i stiger, vil det påvirke nabo 1 og nabo 3 negativt, så de tilhørende y-værdier falder. Hvis der også er negativ korrelation mellem nabo 1 og 2 og mellem nabo 2 og 3, vil det få y-værdien til nabo 2 stige, men det gør, der ikke kan være negativ korrelation mellem y_i og nabo 2. I tilfælde som dette, er det nødvendigt med en dybere analyse af ρ's fortolkning i forbindelse med naborelationerne.

I nogle tilfælde kan det være praktisk at tilføje et led, $\alpha\iota_n$, til SAR-modellen i definition 2.4, hvor ι_n er en søjlevektor med n ettaller og α er en parameter. Leddet $\alpha\iota_n$ har det formål at tage højde for den situation, hvor observationerne $\mathbf{y}$ ikke har middelværdi nul, og i definition 2.4 er leddet indeholdt i vektoren $X\boldsymbol{\beta}$. Det kan være en fordel at have $\alpha\iota_n$ som en seperat vektor, hvis SAR-modellen eksempelvis omskrives til at have to X-matricer, X_1 og X_2, og to $\boldsymbol{\beta}$-vektorer, $\boldsymbol{\beta}_1$ og $\boldsymbol{\beta}_2$, så $\alpha\iota_n$ ikke kommer til at indgå to gange i ligningen. Hvis $\alpha\iota_n$ skrives som et seperat led, får SAR-modellen formen

$$\begin{aligned} &\mathbf{y} = \rho W\mathbf{y} + \alpha\iota_n + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &\boldsymbol{\varepsilon} \sim N_n\left(\mathbf{0},\sigma^2 I_n\right). \end{aligned} \tag{2.4}$$

I nogle tilfælde kan det være praktisk at skrive SAR-modellen fra definition 2.4 som en datagenererende proces, hvilket vil sige, at observationerne $\mathbf{y}$ isoleres på den ene side af lighedstegnet, så det er muligt at generere nye data fra en estimeret model. De genererede værdier kan sammenlignes med de givne observationer for at se, hvor præcis modellen er.

Definition 2.5 (DGP - Data Generating Process):
Den datagenererende proces er en model, hvor den afhængige variabel er isoleret, hvorved det er muligt at simulere nye data for modellen.

Hvis der ses på den datagenererende proces for en model som eksempelvis SAR-modellen, kan det nogle gange være lettere at aflæse dataenes fordeling, mens man ud fra modellens form i definition 2.4 lettere kan analysere hver modelparameters betydning. Den datagenererende proces for SAR-modellen tager form som i sætning 2.6.

Sætning 2.6 (Datagenererende proces for SAR-modellen):
Den datagenererende proces for SAR-modellen har formen

$$\begin{aligned}\mathbf{y} &= \left(I_n - \rho W\right)^{-1}\left(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)\\ \boldsymbol{\varepsilon} &\sim N_n\left(\mathbf{0}, \sigma^2 I_n\right),\end{aligned}$$

hvor $\boldsymbol{\beta}$ og ρ er parametre, W er en rækkestokastisk nabomatrix og $\mathbf{y}$ er en vektor med n observationer. Matricen X indeholder baggrundsvariable, og $\boldsymbol{\varepsilon}$ er en multivariat normalfordelt fejlvektor.

Bevis:
SAR-modellen fra definition 2.4 omskrives ved

$$\begin{aligned}\mathbf{y} &= \rho W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} && \Rightarrow\\ \left(I_n - \rho W\right)\mathbf{y} &= X\boldsymbol{\beta} + \boldsymbol{\varepsilon} && \Rightarrow\\ \mathbf{y} &= \left(I_n - \rho W\right)^{-1}\left(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)\\ \boldsymbol{\varepsilon} &\sim N_n\left(\mathbf{0}, \sigma^2 I_n\right).\end{aligned}$$

Omskrivningen er kun gyldig, hvis matricen $I_n - \rho W$ er invertibel, hvilket der vil blive givet beviser for i lemma 2.11 og lemma 2.12 senere i rapporten. ■

Observationerne $\mathbf{y}$ i SAR-modellen er som tidligere nævnt normalfordelte, og i sætning 2.7 vil det præcise udtryk for fordelingen (2.3) blive vist, da fordelingen er vigtig i forbindelse med at finde estimater for modellens parametre.

Sætning 2.7:
Observationerne $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}^T$ i SAR-modellen er multivariat normalfordelt med

$$\mathbf{Y} \sim N_n\left(\left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}, \left(I_n - \rho W\right)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T\right).$$

Bevis:
Ses der på den datagenererende proces for SAR-modellen fra sætning 2.6, kan middelværdien for $\mathbf{y}$ udregnes som

$$\begin{aligned}\mathrm{E}\left[\mathbf{y}\right] &= \mathrm{E}\left[\left(I_n - \rho W\right)^{-1}\left(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)\right]\\ &= \mathrm{E}\left[\left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta} + \left(I_n - \rho W\right)^{-1}\boldsymbol{\varepsilon}\right]\\ &= \mathrm{E}\left[\left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}\right] + \mathrm{E}\left[\left(I_n - \rho W\right)^{-1}\boldsymbol{\varepsilon}\right]\\ &= \mathrm{E}\left[\left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}\right]\\ &= \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}.\end{aligned}$$

Det anden sidste lighedstegn fås fra definitionen af vektoren $\boldsymbol{\varepsilon}$, der er multivariat normalfordelt med en nulvektor som middelværdi, og det sidste lighedstegn fås,

idet $\left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}$ er en talvektor. Resultatet stemmer overens med definitionen af middelværdivektoren $\boldsymbol{\mu} = \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}$, som blev anvendt ved omskrivningen af (2.1).

Variansen for $\mathbf{y}$ i SAR-modellens datagenererende proces kan nu findes som

$$\begin{aligned}\text{Var}\left[\mathbf{y}\right] &= \text{Var}\left[\left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta} + \left(I_n - \rho W\right)^{-1}\boldsymbol{\varepsilon}\right] \\ &= \text{Var}\left[\left(I_n - \rho W\right)^{-1}\boldsymbol{\varepsilon}\right] \\ &= \left(I_n - \rho W\right)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T.\end{aligned}$$

Sidste lighedstegn opnås ved at indsætte variansen for $\boldsymbol{\varepsilon}$, der er givet ved $\sigma^2 I_n$, og sætte konstanten $\left(I_n - \rho W\right)^{-1}$ i anden uden for variansen. ■

Når den præcise fordeling for SAR-modellen er kendt, kan den tilhørende tæthed (se appendiks B.1) anvendes til at finde estimater for modellens parametre, hvilket der vil blive givet to metoder til i henholdsvis kapitel 5 og 6.

Da SAR-modellens funktion er at modellere rumlig afhængighed mellem observationer, vil der nu blive givet et eksempel på, hvordan en simpel SAR-model kan konstrueres og fortolkes for en lille mængde indbyrdes afhængige observationer.

Eksempel 2.8:

Eksemplet tager udgangspunkt i [LeSage and Pace, 2009, s. 2-3 og 17-20], og oplysningerne er hentet fra [Krak] og [Statistik, Opslag: Folketal].

Som i eksempel 2.3 vil de fem regioner i Danmark blive anset for indbyrdes afhængige punkter, der kan knyttes forskellige værdier til. Hvis man eksempelvis ønsker at rejse fra Skagen til København i bil, passerer man gennem alle fem regioner, hvis der ikke medtages færgeruter, hvilket der ikke vil blive gjort her. Dermed vil de fem regioner kunne anses for at ligge på en linje, der repræsenterer ruten gennem Danmark, hvor man passerer gennem hver enkelt region undervejs. Rejsetiden gennem hver region er omtrent

Rejsetider i minutter	**Område**
109	Region Nordjylland (Skagen - Hobro)
80	Region Midtjylland (Hobro-Vejle)
84	Region Syddanmark (Vejle-Korsør)
60	Region Sjælland (Korsør- Roskilde)
34	Region Hovedstaden (Roskilde-København K)

De ovenstående rejsetider kan samles i en vektor $\mathbf{y}$, hvor hver indgang er rejsetiden fra den givne region til Region Hovedstaden, hvilket giver

$$\mathbf{y} = \begin{bmatrix} 367 \\ 258 \\ 178 \\ 94 \\ 34 \end{bmatrix}.$$

Her er tiden målt i minutter, hvor værdien i første indgang er rejsetiden fra Region Nordjylland til Region Hovedstaden, og den sidste værdi er rejsetiden gennem Region Hovedstaden, der ikke er nul, da det også tager tid at rejse fra et sted til et andet i regionen. Matricen X indeholder baggrundsvariable, der påvirker rejsetiden $\mathbf{y}$, hvilket eksempelvis kan være antal indbyggere i hver region, da mængden af mennesker i et område påvirker trafikken og dermed rejsetiden. Et andet element, der har indflydelse på rejsetiden er den afstand, som tilbagelægges i hver region, så matricen X kan eksempelvis have formen

$$X = \begin{array}{c} \begin{matrix} \text{Afstand} & \text{Befolkningsantal} \end{matrix} \\ \begin{bmatrix} 532{,}9 & 543.116 \\ 371{,}5 & 1.159.189 \\ 242{,}1 & 1.093.729 \\ 111{,}9 & 757.257 \\ 37{,}2 & 1.442.536 \end{bmatrix} \end{array},$$

hvor afstande er angivet i kilometer. Den første indgang i $\mathbf{y}$ er den tid, det tager at rejse fra Region Nordjylland til Region Hovedstaden, og den afhænger af rejsetiderne gennem hver af de andre regioner, så der er muligt at anvende SAR-modellen fra definition 2.4 til at beskrive den indbyrdes afhængighed mellem observationerne. Den rækkestokastiske nabomatrix W har formen

$$W = \begin{bmatrix} & No & Mi & Sy & Sj & Ho \\ No & 0 & 1 & 0 & 0 & 0 \\ Mi & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ Sy & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ Sj & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ Ho & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Det er muligt at estimere parametrene $\hat{\boldsymbol{\beta}}$ og $\hat{\rho}$ fra SAR-modellen ved brug af maximum likelihood, som vil blive gennemgået i kapitel 5, hvilket giver estimaterne

$$\hat{\rho} = 0{,}0048$$
$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0{,}65 \\ -0{,}000012 \end{bmatrix},$$

der er udregnet i programmet R. Da $\hat{\rho}$ er positiv og mindre end 1, er der positiv korrelation mellem rejsetiderne i $\mathbf{y}$. Hvis de estimerede parametre anvendes til at generere forudsigelser for nye observationer, udregnes disse ved brug af

$$\hat{\mathbf{y}}^{(1)} = \left(I_n - \hat{\rho} W\right)^{-1} X \hat{\boldsymbol{\beta}}, \tag{2.5}$$

der fås ud fra definition 2.4. Det siges også at SAR-modellen er en model for spillover-effekter, der opstår, når en ændring i i'te række i matricen X påvirker alle indgange i $\mathbf{y}$ mere eller mindre og ikke kun observation y_i. Hvis det for eksempel antages, at den globale opvarmning får vandstanden til at stige så meget, at oversvømmelser får 500.000

hollændere til at immigrere til Region Syddanmark, så bliver den nye matrix X^* til

$$X^* = \begin{bmatrix} 532{,}9 & 543.116 \\ 371{,}5 & 1.159.189 \\ 242{,}1 & \mathbf{1.593.729} \\ 111{,}9 & 757.257 \\ 37{,}2 & 1.442.536 \end{bmatrix}.$$

Der kan nu laves en ny ligning til at generere forudsigelser ud fra den ændrede matrix X^*, der har formen

$$\hat{\mathbf{y}}^{(2)} = \left(I_n - \hat{\rho} W\right)^{-1} X^* \hat{\boldsymbol{\beta}}. \tag{2.6}$$

Udregnes rejsetiderne både ved brug af (2.5) og (2.6) kan forskellen mellem dem udregnes, hvilket er gjort i tabel 2.1. Det ses klart, at hvis en enkelt værdi fra matricen X ændres, bliver rejsetiden for hver region ændret.

Region	$\hat{\mathbf{y}}^{(1)}$	$\hat{\mathbf{y}}^{(2)}$	$\hat{\mathbf{y}}^{(2)} - \hat{\mathbf{y}}^{(1)}$
Nordjylland	340,7	340,7	-0,000066
Midtjylland	228,9	228,9	-0,014
Syddanmark	145,2	139,4	-5,8
Sjælland	64,3	64,2	-0,014
Hovedstaden	7,8	7,8	-0,000066

Tabel 2.1: Estimater for rejsetider ved anvendelse af SAR-modellen.

Summeres der over fjerde kolonne i tabel 2.1, fås den samlede påvirkning på observationerne $\mathbf{y}$. Den tredje indgang i kolonnen $\hat{\mathbf{y}}^{(2)} - \hat{\mathbf{y}}^{(1)}$ viser den såkaldte direkte påvirkning fra ændringen i X, mens resten af indgangene viser den indirekte effekt. Da $\hat{\rho}$ er positiv og $\hat{\beta}_2$ er negativ, bliver rejsetiderne formindsket, når der anvendes den ændrede matrix X^*, hvilket intuitivt ikke giver mening. Der kan muligvis opnås bedre estimater ved at anvende mere signifikante baggrundsvariable. Det ses dog alligevel, at alle rejsetiderne påvirkes af en ændring i en enkelt variabel i X, hvor førsteordensnaboerne til y_3 givet ved y_2 og y_4 påvirkes mere end andenordensnaboerne, og y_3 selv påvirkes mest. □

2.2 Tidsafhængig udgave af SAR-modellen

SAR-modellen repræsenterer normalt en model for en mængde observationer, der er observeret på samme tid, men det er også muligt, at opfatte SAR-modellen som en tidsafhængig model. I det tilfælde erstattes vektoren $\mathbf{y}$ af observationer i SAR-modellen med en tidsafhængig vektor $\mathbf{y}_t$, der eksempelvis kan indeholde skatteniveauer i forskellige områder til tiden t. Her er det muligt, at nogle af områderne har sat skatteniveauet efter naboområdernes skatteniveau til tiden $t-1$, hvorved der kan konstrueres en nabomatrix W, hvor kombinationen $W\mathbf{y}_{t-1}$ beskriver indflydelsen fra

naboområders tidligere skatteniveauer på det nuværende $\mathbf{y}_t$. Det svarer til, at en tidsafhængig version af SAR-modellen har formen

$$\mathbf{y}_t = \rho W \mathbf{y}_{t-1} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t.$$

Det kan vises, at når t vokser, vil middelværdien for ovenstående gå mod en ligevægt, der er giver i sætning 2.9.

Sætning 2.9:
SAR-modellen kan skrives som en langsigtet ligevægtstilstand på formen

$$\lim_{t\to\infty} \mathrm{E}[\mathbf{y}_t] = \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}.$$

Her indeholder vektoren $\mathbf{y}_t$ observationer til tiden t, mens W er en rækkestokastisk nabomatrix, og parameteren $|\rho| < 1$ er graden af afhængighed mellem naboobservationerne. Derudover er X en matrix med baggrundsvariable til tiden t, og $\boldsymbol{\beta}$ er den tilhørende parametervektor.

Bevis:
Lad $\mathbf{y}_t$ være en vektor med observationer til tiden t, hvor nabomatricen W indeholder information om hvilke værdier i $\mathbf{y}_t$, der afhænger af en eller flere værdier fra vektoren $\mathbf{y}_{t-1}$. Matricen X_t indeholder baggrundsvariablene til tiden t, men da baggrundsvariablene kan antages at være forholdsvist fastholdte over tid, betegnes matricen bare X. SAR-modellen kan dermed skrives som

$$\mathbf{y}_t = \rho W \mathbf{y}_{t-1} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t. \tag{2.7}$$

Erstattes t i ovenstående med $t-1$, ses det at $\mathbf{y}_{t-1}$ afhænger af værdierne i $\mathbf{y}_{t-2}$ og på samme måde afhænger $\mathbf{y}_{t-2}$ af $\mathbf{y}_{t-3}$. Dermed haves en Markovkæde, og leddet $\mathbf{y}_{t-1}$ i (2.7) kan erstattes med $\rho W \mathbf{y}_{t-2} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{t-1}$, hvilket giver

$$\begin{aligned}\mathbf{y}_t &= \rho W\left(\rho W \mathbf{y}_{t-2} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{t-1}\right) + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \\ &= X\boldsymbol{\beta} + \rho W X\boldsymbol{\beta} + \rho^2 W^2 \mathbf{y}_{t-2} + \boldsymbol{\varepsilon}_t + \rho W \boldsymbol{\varepsilon}_{t-1}.\end{aligned}$$

Her kan $\mathbf{y}_{t-2}$ erstattes med $\rho W \mathbf{y}_{t-3} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{t-2}$ og så videre, hvilket giver

$$\begin{aligned}\mathbf{y}_t &= \left(I_n + \rho W + \rho^2 W^2 + \ldots + \rho^{s-1} W^{s-1}\right) X\boldsymbol{\beta} + \rho^s W^s \mathbf{y}_{t-s} + u \\ u &= \boldsymbol{\varepsilon}_t + \rho W \boldsymbol{\varepsilon}_{t-1} + \rho^2 W^2 \boldsymbol{\varepsilon}_{t-2} + \ldots + \rho^{s-1} W^{s-1} \boldsymbol{\varepsilon}_{t-(s-1)} \\ &\Downarrow \\ \mathbf{y}_t &= \left(\sum_{j=0}^{s-1} \rho^j W^j\right) X\boldsymbol{\beta} + \rho^s W^s \mathbf{y}_{t-s} + \sum_{j=0}^{s-1} \rho^j W^j \boldsymbol{\varepsilon}_{t-j},\end{aligned} \tag{2.8}$$

hvor $s < t$ er antallet af forudgående tidsperioder, der antages at have indflydelse på $\mathbf{y}_t$. Da fejlvektoren $\boldsymbol{\varepsilon}$ i SAR-modellen har middelværdi nul, er $\mathrm{E}[\boldsymbol{\varepsilon}_{t-r}] = 0$ for $r = 0, \ldots, s-1$, hvilket er ensbetydende med $\mathrm{E}[u] = 0$ i (2.8). Da det i sætning 2.9 er antaget, at $|\rho| < 1$, vil betydningen af $\rho^s W^s \mathbf{y}_{t-s}$ blive lille, når der haves en stor værdi for s, idet den rækkestokastiske nabomatrix W har en maksimal egenværdi på én. Da $|\rho| < 1$ og W er

rækkestokastisk, gælder det jævnfør sætning 2.10, at (2.8) konvergerer, og det er dermed muligt at anvende den geometriske sumformel, så

$$\lim_{s\to\infty}\left(\sum_{j=0}^{s-1}\rho^j W^j\right)X\boldsymbol{\beta} = \left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta},$$

hvorved den forventede værdi for en langsigtet ligevægtstilstand bliver

$$\lim_{t\to\infty}\mathrm{E}[\mathbf{y}_t] = \left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta},$$

hvor det følger af $t\to\infty$ og $s<t$, at $s\to\infty$. ■

I sætning 2.9 blev det antaget, at $|\rho|<1$ for at sikre konvergens, og der vil nu blive set nærmere på denne antagelse.

Sætning 2.10:
Lad W være en rækkestokastisk $(n\times n)$-matrix og $\mathbf{y}_t$ en vektor med n observationer til tiden t, mens matricen X indeholder baggrundsvariable, $\boldsymbol{\beta}$ er en parametervektor og $\boldsymbol{\varepsilon}$ er en fejlvektor. Da konvergerer udtrykket

$$\mathbf{y}_t = \lim_{s\to\infty}\left(\left(\sum_{j=0}^{s-1}\rho^j W^j\right)X\boldsymbol{\beta}+\rho^s W^s\mathbf{y}_{t-s}+\sum_{j=0}^{s-1}\rho^j W^j\boldsymbol{\varepsilon}_{t-j}\right),$$

hvis parameteren ρ opfylder betingelsen $|\rho|<1$.

Bevis:
Grænseværdien for den autoregressive proces med rumlige lags udtrykt i (2.8) findes ved fortsat at medtage trin længere tilbage i tiden, så $\mathbf{y}_{t-s}$ erstattes med $\rho W\mathbf{y}_{t-s-1}+X\boldsymbol{\beta}+\boldsymbol{\varepsilon}_{t-s}$ og så videre, så $s\to\infty$. Det svarer til

$$\mathbf{y}_t = \lim_{s\to\infty}\left(\left(\sum_{j=0}^{s-1}\rho^j W^j\right)X\boldsymbol{\beta}+\rho^s W^s\mathbf{y}_{t-s}+\sum_{j=0}^{s-1}\rho^j W^j\boldsymbol{\varepsilon}_{t-j}\right). \tag{2.9}$$

De værdier, parameteren ρ kan have, kan inddeles i tre forskellige tilfælde, der er $|\rho|>1$, $|\rho|=1$ og $|\rho|<1$. Parameteren ρ indgår i alle tre led i (2.9) via potenser af faktoren ρW, der ved brug af spektral dekomposition kan skrives som

$$\rho W = \rho V\Lambda_W V^{-1} \Rightarrow \rho^s W^s = \underbrace{\left(\rho V\Lambda_W V^{-1}\right)\left(\rho V\Lambda_W V^{-1}\right)\cdots\left(\rho V\Lambda_W V^{-1}\right)}_{s} = \rho^s V\Lambda_W^s V^{-1},$$

hvor det antages at matricen W er diagonaliserbar og at matricen V opfylder $VV^{-1}=I_n$. Matricen Λ_W er en diagonalmatrix med egenværdierne $\lambda_1,\ldots,\lambda_n$ for W på diagonalen, hvorved det er muligt at skrive

$$\left(\rho\Lambda_W\right)^s = \begin{bmatrix}\left(\rho\lambda_1\right)^s & 0 & \cdots & 0\\ 0 & \left(\rho\lambda_2\right)^s & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \left(\rho\lambda_n\right)^s\end{bmatrix}. \tag{2.10}$$

Først ses der på tilfældet $|\rho| > 1$. Hvis en af egenværdierne λ_i for W opfylder $|\rho\lambda_i| > 1$, så er det ensbetydende med $|\rho| > 1$, idet $\lambda_{\max} = 1$. Det vil sige, at for $|\rho| > 1$, vil både ρ^s og Λ_W^s blive større i takt med s bliver større, så for $s \to \infty$ vil $\rho^s W^s$ blive uendeligt stor og matricen (2.10) vil ikke konvergere. Dermed vil udtrykket i (2.9) blive ved med at vokse uden at nærme sig en grænseværdi, når $\left|\rho\right| > 1$. Derudover ses det, at når $|\rho| > 1$ og $s \to \infty$, så får naboer af højere og højere orden mere indflydelse på $\mathbf{y}_t$ gennem $\rho^s W^s$, ligesom observationer $\mathbf{y}_{t-s}$ og fejl $\boldsymbol{\varepsilon}_{t-(s-1)}$ langt tilbage i tiden får mere indflydelse når s vokser, hvilket ikke virker logisk.

Det andet tilfælde er, når $\left|\rho\right| = 1$, hvorved der eksisterer et λ_i, så $|\rho\lambda_i| = 1$, idet $\lambda_{\max} = 1$. Det svarer til, at matricen (2.10) er lig Λ_W^s, der for $s \to \infty$ konvergerer mod en matrix, som ikke er nul-matricen. Derudover vil $\left|\rho\right| = 1$ ændre (2.9) til

$$\mathbf{y}_t = \lim_{s\to\infty}\left(\left(\sum_{j=0}^{s-1} W^j\right) X\boldsymbol{\beta} + W^s\mathbf{y}_{t-s} + \sum_{j=0}^{s-1} W^j\boldsymbol{\varepsilon}_{t-j}\right),$$

hvor det ses, at graden af afhængighed er ens for naboer af alle ordener, så observationer til tiden $t-1$ påvirker $\mathbf{y}_t$ lige så meget som for $t-s$, hvor $s \to \infty$. Det svarer til, processen har uendelig hukommelse, og at alle $\mathbf{y}_{t-s}$ for $s \to \infty$ påvirker $\mathbf{y}_t$ lige meget, hvilket ikke er tilfældet for de fleste af virkelighedens scenarier.

Der ses nu på tilfældet $|\rho| < 1$. Idet $\lambda_{\max} = 1$ svarer $\rho \in (-1,1)$ til at $\rho\lambda_i \in (-1,1)$, hvilket gør, at (2.10) konvergerer mod nul-matricen. Dermed må det gælde, at $\left|\rho\right| < 1$ for at (2.9) har en grænseværdi, hvor $|\rho| < 1$ også kaldes den stationære antagelse, da den gør systemet stationært. Der ses nu på, hvad der sker med hvert led i (2.9) når $\left|\rho\right| < 1$.

Første led fra (2.9) kan som tidligere nævnt omskrives ved brug af den geometriske sumformel, da $|\rho| < 1$ og matricen W er rækkestokastisk, hvilket giver

$$\lim_{s\to\infty}\left(\left(\sum_{j=0}^{s-1} \rho^j W^j\right) X\boldsymbol{\beta}\right) = \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta},$$

der resulterer i en konstant vektor. Andet led i (2.9) med $\left|\rho\right| < 1$ resulterer i

$$\rho^s W^s\mathbf{y}_{t-s} \to 0 \quad \text{for} \quad s \to \infty.$$

Det kan ikke siges med sikkerhed, hvordan det tredje og sidste led i (2.9) udvikler sig, når $\left|\rho\right| < 1$ og $s \to \infty$, men hvis det antages, at summen af fejlled konvergerer, fås det at

$$\begin{aligned}
\mathbf{y}_t &= \lim_{s\to\infty}\left(\left(\sum_{j=0}^{s-1} \rho^j W^j\right) X\boldsymbol{\beta} + \rho^s W^s\mathbf{y}_{t-s} + \sum_{j=0}^{s-1} \rho^j W^j\boldsymbol{\varepsilon}_{t-j}\right) \\
&= \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta} + 0 + \sum_{j=0}^{\infty} \rho^j W^j\boldsymbol{\varepsilon}_{t-j} \\
&= \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta} + \sum_{j=0}^{\infty} \rho^j W^j\boldsymbol{\varepsilon}_{t-j}.
\end{aligned}$$

Dermed kan observationerne $\mathbf{y}_t$ til tiden t skrives som en konstant adderet med en uendelig sum af fejlled fra tidligere observationer. Her ses det også, at fejlleddene får

mindre betydning for $\mathbf{y}_t$, des længere tilbage i tiden de stammer fra, da $\rho^j W^j$ går mod nulmatricen når $j \to \infty$. Det kan også siges, at værdien for ρ bestemmer processens hukommelse. ■

Ud fra sætning 2.9 og 2.10 ses det, at SAR-modellen kan opfattes som en tidsafhængig model, der eksempelvis modellere skatteniveauer $\mathbf{y}_t$ i forskellige områder, hvor skatten er fastsat ud fra skatten i naboområderne før tiden t. Hvis der ses bort fra tidsafhængigheden, udgør (2.7) SAR-modellen, hvorved overvejelserne om ρ i ovenstående også gælder for SAR-modellen.

2.3 SAR-modellens kovariansmatrix

Kovariansmatricen til SAR-modellen vil nu blive undersøgt nærmere, da flere betingelser må være opfyldt, for at den er veldefineret. For det første må parameteren ρ tilhøre et passende interval, der afhænger af egenværdierne for den rækkestokastiske vægtmatrix W, og for det andet må det gælde, at matricen $I_n - \rho W$ er invertibel, hvor beviset for dette også afhænger af egenværdierne for vægtmatricen W. Ifølge sætning 2.7 er kovariansmatricen for SAR-modellen givet ved

$$\left(I_n - \rho W\right)^{-1} \sigma^2 \left(\left(I_n - \rho W\right)^{-1}\right)^T. \tag{2.11}$$

Det er nok, at kovariansmatricen er positivt semidefinit og symmetrisk, for at definitionen af normalfordelingen, som SAR-modellen følger, stadig gælder, men ønskes der en veldefineret tæthed for modellen samt en invertibel kovariansmatrix, må kravet skærpes til positivt definit. I sætning 2.13 samles de forudsætninger, der kræves for at have en veldefineret kovariansmatrix, og de følgende to lemmaer indgår i beviset for sætningen.

Lemma 2.11:
Den inverse af matricen $I_n - \rho W$ eksisterer, hvis parameteren ρ tilhører intervallet $\left(\lambda_{\min}^{-1}, 1\right)$, hvor $\lambda_{\min}$ er den mindste egenværdi for den rækkestokastiske vægtmatrix W, der har reelle egenværdier.

Bevis:
Den inverse til matricen $I_n - \rho W$ eksisterer, hvis determinanten $|I_n - \rho W| \neq 0$. For at bevise dette, ses der først på $(n \times n)$-matricen W, der kan skrives på Jordan kanonisk form (se [Perko, 2001, s. 39-40]), hvorved

$$J = V^{-1} W V \Rightarrow W = V J V^{-1}.$$

I det tilfælde W kan diagonaliseres, bliver J til en diagonalmatrix, og ellers fås en øvre triangulær matrix med en både diagonal og en superdiagonal, så

$$J = \begin{bmatrix} \lambda_1 & j_{1,2} & 0 & \dots & 0 \\ 0 & \lambda_2 & j_{2,3} & \ddots & \vdots \\ 0 & 0 & \lambda_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & j_{n-1,n} \\ 0 & \dots & 0 & 0 & \lambda_n \end{bmatrix}.$$

Her angiver $\lambda_1,\lambda_2,\ldots,\lambda_n$ egenværdierne for W, mens elementerne $(J)_{ij}$ er enten 0 eller 1 for $i = 1,\ldots,n-1$ og $j = i+1$, og nul for alle andre indgange i J uden for diagonalen. Ved brug af ovenstående kan determinanten $|I_n - \rho W|$ omskrives til

$$|I_n - \rho W| = |I_n - \rho V J V^{-1}| = |V||I_n - \rho J||V^{-1}| = |I_n - \rho J|,$$

idet $|V||V^{-1}| = |I_n| = 1$. Da matricen J indeholder egenværdierne $\lambda_1,\lambda_2,\ldots,\lambda_n$ for W på diagonalen, kan determinanten skrives

$$|I_n - \rho J| = \begin{vmatrix} 1-\rho\lambda_1 & -\rho j_{1,2} & 0 & \ldots & 0 \\ 0 & 1-\rho\lambda_2 & -\rho j_{2,3} & \ddots & \vdots \\ 0 & 0 & 1-\rho\lambda_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & -\rho j_{n-1,n} \\ 0 & \ldots & 0 & 0 & 1-\rho\lambda_n \end{vmatrix} = \prod_{i=1}^{n} \left(1-\rho\lambda_i\right),$$

idet produktet af diagonalelementerne udgør determinanten til en øvre, triangulær matrix. En matrix er invertibel, hvis determinanten er ulig med nul, hvilket er tilfældet for $|I_n - \rho J|$, hvis

$$\rho \neq \frac{1}{\lambda_i} \quad \text{for} \quad i = 1,2,\ldots,n. \tag{2.12}$$

Det undersøges nu, hvilket interval ρ må tlhøre, for at (2.12) er opfyldt. I lemma 2.11 blev det antaget, at egenværdierne for W er reelle, og derudover er matricen W rækkestokastisk, så

$$W \cdot \iota_n = 1 \cdot \iota_n, \tag{2.13}$$

hvilket gør $\lambda_k = 1$ til en egenværdi for W. For at bevise at den største egenværdi for W er $\lambda_i = 1$, antages det, at der eksisterer en egenværdi $\lambda_l > 1$. Da W er rækkestokastisk, udgør $W\mathbf{x}$ et vægtet gennemsnit af indgangene i vektoren $\mathbf{x}$, og for $i = 1,\ldots,n$ gælder

$$(W\mathbf{x})_i \in [x_{\min}, x_{\max}],$$

hvor $x_{\max}$ og $x_{\min}$ er henholdsvis den største og den mindste værdi i $\mathbf{x}$. Det blev antaget, at $\lambda_l > 1$, og det er derfor muligt, at et element $x_l \in \mathbf{x}$ opfylder $\lambda_l x_l > x_{\max}$, så ligheden $W\mathbf{x} = \lambda_l\mathbf{x}$ ikke længere er gyldig, fordi $\lambda_l x_l \notin [x_{\min}, x_{\max}]$. Det vil sige $\lambda_l > 1$ ikke kan være en egenværdi for W, så det må gælde, at W's egenværdier $\lambda_i \leq 1$ for $i = 1,\ldots,n$. Sammen med resultatet (2.12) betyder det, at $\rho \neq 1$.

Der ses nu på den mindste egenværdi for W, der noteres $\lambda_{\min}$. Indgangene i diagonalen i W er per definition lig nul, og sættes dette sammen med reglen $\operatorname{tr}(W) = \sum_{i=1}^{n} \lambda_i$, fås det at $\sum_{i=1}^{n} \lambda_i = 0$. Fra (2.13) vides det desuden, at der for et $k \in \{1,2,\ldots,n\}$ gælder, at $\lambda_k = 1$, så derfor har W mindst én negativ egenværdi. Det er ikke nødvendigt at præcisere den mindste egenværdi for W yderligere, da kravet $\rho \neq \frac{1}{\lambda_i}$ for $i = 1,2,\ldots,n$ kan opfyldes ved at lade ρ tilhøre det åbne interval $\left(\lambda_{\min}^{-1}, 1\right)$.

Det er hermed blevet bevist, at matricen $I_n - \rho W$ invertibel, når $\rho \in \left(\lambda_{\min}^{-1}, 1\right)$. ■

Det gælder ikke altid, at matricen $I_n - \rho W$ er symmetrisk eller similær med en symmetrisk matrix, og den er heller ikke altid positiv definit. I disse tilfælde har matricen W som regel komplekse egenværdier, og det er her relevant at undersøge, hvilket interval for ρ, der kan gøre $I_n - \rho W$ invertibel.

Lemma 2.12:
Hvis den rækkestokastiske nabomatrix W har komplekse egenværdier, eksisterer den inverse af matricen $I_n - \rho W$, hvis parameteren ρ tilhører intervallet $\left(\lambda_{\min}^{-1}, 1\right)$, hvor $\lambda_{\min}^{-1}$ er den laveste, negative, reelle egenværdi til W.

Bevis:
Som i beviset for lemma 2.11 ses der på determinanten af $I_n - \rho W$, der kan skrives

$$|I_n - \rho W| = \prod_{j=1}^{n} \left(1 - \rho\lambda_j\right) = \left[\prod_{j=3}^{n} \left(1 - \rho\lambda_j\right)\right] \left(1 - \rho\lambda_1\right)\left(1 - \rho\lambda_2\right).$$

Her er egenværdierne til W givet ved λ_i for $i = 1, \ldots, n$. Matricen $I_n - \rho W$ er ikke invertibel, hvis determinanten er nul, hvilket den er når $\left(1 - \rho\lambda_2\right)\left(1 - \rho\lambda_1\right) = 0$ for to egenværdier λ_1 og λ_2. Egenværdierne til W kan være komplekse, så λ_1 og λ_2 noteres som de konjugerede par

$$\begin{aligned} \lambda_1 &= r + ic \\ \lambda_2 &= r - ic \\ i &= \sqrt{-1}. \end{aligned}$$

Her er c og r reelle tal, hvor det antages, at $c \neq 0$. Det er nu muligt at finde de værdier for ρ, som ikke resulterer i en invertibel matrix $I_n - \rho W$, hvilket gøres ved at løse

$$\begin{aligned} 0 &= \left(1 - \rho\lambda_1\right)\left(1 - \rho\lambda_2\right) \\ &= \left(1 - \rho r - \rho ic\right)\left(1 - \rho r + \rho ic\right) \\ &= 1 - 2\rho r + \rho^2 r^2 - \rho^2 i^2 c^2 \\ &= 1 - 2\rho r + \rho^2\left(r^2 + c^2\right). \end{aligned} \tag{2.14}$$

Udtrykket (2.14) har diskriminanten

$$\begin{aligned} d &= (-2r)^2 - 4 \cdot \left(r^2 + c^2\right) \cdot 1 \\ &= 4\left(r^2 - r^2 - c^2\right) \\ &= -4c^2 < 0. \end{aligned}$$

En negativ værdi for diskriminanten resulterer i to komplekse løsninger til ρ, hvor begge løsningerne vil medføre, at matricen $I_n - \rho W$ ikke er invertibel. Antages det derimod på forhånd, at ρ kun må antage reelle værdier, vil invertibiliteten af $I_n - \rho W$ ikke blive påvirket af, om W har komplekse egenværdier eller ej.

I det tilfælde, hvor W har komplekse egenværdier, vil $I_n - \rho W$ være invertibel, når det er opfyldt, at $\rho \in \left(\lambda_{\min}^{-1}, 1\right)$, hvor $\lambda_{\min}^{-1}$ er den laveste, negative, reelle egenværdi for W. ■

Lemma 2.9 og lemma 2.10 viser, at matricen $I_n - \rho W$ er invertibel, både når W har reelle, og når den har komplekse egenværdier, så længe ρ tilhører intervallet $\left(\lambda_{\min}^{-1},1\right)$. I praksis antages det ofte, at $\rho \in [0,1)$ både i tilfældet med reelle egenværdier og i det med komplekse egenværdier, da det som tidligere nævnt kan være vanskeligt at fortolke negative værdier for ρ, og kravet $\rho \in [0,1)$ sikre også, at $I_n - \rho W$ er invertibel. Resultaterne fra lemmaerne 2.9 og 2.10 kan nu anvendes til at finde kravene for, hvornår kovariansmatricen til SAR-modellen er veldefineret.

Sætning 2.13:
Hvis parameteren ρ tilhører intervallet $\left(\lambda_{\min}^{-1},1\right)$, hvor $\lambda_{\min}$ er den mindste egenværdi for den rækkestokastiske nabomatrix W, så er SAR-modellens kovariansmatrix

$$\left(I_n - \rho W\right)^{-1} \sigma^2 \left(\left(I_n - \rho W\right)^{-1}\right)^T$$

positivt definit og symmetrisk og dermed en veldefineret kovariansmatrix.

Bevis:
For overskuelighedens skyld defineres først en matrix A som $A = I_n - \rho W$. For enhver matrix A gælder der ifølge [Axler, 1997, s. 144], at matricen givet ved $A^T A$ er positivt semidefinit, da

$$\mathbf{x}^T A^T A\mathbf{x} = (A\mathbf{x})^T A\mathbf{x} = \|A\mathbf{x}\|^2 \geq 0,$$

hvor $\mathbf{x}$ er en $(n \times 1)$-vektor. Hvis matricen $A^T A$ er invertibel, er determinanten ulig nul, og dermed bliver $\|A\mathbf{x}\|^2 \geq 0$ i ovenstående til $\|A\mathbf{x}\|^2 > 0$, hvorved den også vil være positivt definit. Matricen $A^T A$ er invertibel, hvis matricen A er invertibel, idet

$$\left(A^T A\right)^{-1} = (A)^{-1}\left(A^T\right)^{-1} = (A)^{-1}\left(A^{-1}\right)^T.$$

Anvendes lemma 2.11 og 2.12 ses det, at matricen $A = I_n - \rho W$ er invertibel, hvormed $A^T A$ også er invertibel, så

$$A^T A = \left(I_n - \rho W\right)^T \left(I_n - \rho W\right) \Rightarrow$$
$$\left(A^T A\right)^{-1} = \left(I_n - \rho W\right)^{-1}\left(\left(I_n - \rho W\right)^T\right)^{-1} = \left(I_n - \rho W\right)^{-1}\left(\left(I_n - \rho W\right)^{-1}\right)^T$$

er invertibel og derfor også positivt definit. Kovariansmatricen for SAR-modellen er givet ved

$$\left(I_n - \rho W\right)^{-1} \sigma^2 \left(\left(I_n - \rho W\right)^{-1}\right)^T,$$

hvor konstanten $\sigma^2 > 0$, hvormed kovariansmatricen er invertibel og dermed også positivt definit. Derudover gælder der for enhver matrix A, at matricen givet ved $A^T A$ er en symmetrisk matrix, hvorved den inverse $\left(A^T A\right)^{-1}$ også er en symmetrisk matrix. Det vil sige, $\left(I_n - \rho W\right)^{-1} \sigma^2 \left(\left(I_n - \rho W\right)^{-1}\right)^T$ er en veldefineret kovariansmatrix. ■

Kravet $\rho \in \left(\lambda_{\min}^{-1},1\right)$ medfører, at ρ både kan antage negative værdier, der kan være svære at fortolke, og ρ kan også være lig nul. Som tidligere nævnt kan det i stedet antages, at $\rho \in [0,1)$, selvom muligheden $\rho = 0$ stadig er med og dermed også den mulighed, at W og dermed naborelationerne forsvinder fra modellen. Det vil resultere i en lineær, normal model uden rumlig afhængighed, men nul medtages alligevel i intervallet for ρ, da det også er relevant at undersøge muligheden, at observationerne er ukorrelerede.

CAR-modellen 3

Det følgende afsnit er skrevet på baggrund af [Besag, 1974], [Cressie, 1993, s. 414-421] og [Lauritzen, 1996], medmindre andet er angivet.

Et datasæt, der består af en mængde diskrete punkter, kan opfattes som en graf, hvor forbundne punkter viser hvilke observationer, der er naboer. En model, der kan anvendes på denne form for data, er en *Conditional autoregressive model*, der forkortes som CAR-modellen. CAR-modellen er en rumlig model med en betinget sandsynlighedsfordeling, og den kaldes også en autonormal model, da den tilpasser en multivariat normalfordeling til datasættet. CAR-modellen tager udgangspunkt i Markovfelter, der vil blive gennemgået i afsnit 3.2, og en særlig vigtig egenskab ved modellen er givet i sætning 3.14, der bevises ved brug af redskaber fra grafteori, som også gennemgås i dette kapitel. Allerførst er der dog brug for et grundlæggende resultat om betingede sandsynlighedsfunktioner, hvilket vil blive præsenteret i afsnit 3.1.

3.1 Rumlige processer og betinget sandsynlighed

Der ses nu på et rumligt datasæt, der er defineret på et rektangulært gitter, hvor hvert punkt er nummereret som et talpar (i,j) for $i,j, \in \mathbb{Z}$. Der er knyttet en stokastisk variabel, $X_{i,j}$ til hvert talpar, og den samlede mængde af stokastiske variable er derfor $\{X_{i,j} | i,j, \in V \subseteq \mathbb{Z}\}$, hvor der i dette tilfælde bort fra, om mængden er endelig eller uendelig. Hvis der for eksempel haves et endeligt datasæt bestående af huspriser samt placeringen af hvert hus, vil hvert punkt (i,j) være et hus med en givet pris $x_{i,j}$. Rumlige, stokastiske processer kan præciseres ved brug af en sandsynlighedsfunktion, som for eksempel kan være den simultane sandsynlighedsfordeling givet ved

$$\prod_{i,j} Q_{i,j}(x_{i,j}, x_{i-1,j}, x_{i+1,j}, x_{i,j-1}, x_{i,j+1}),$$

hvor $x_{i,j}$ er angiver en værdi for $X_{i,j}$, og Q er en funktion.

En anden metode til at præcisere en rumlig proces er at definere den betingede sandsynlighedsfordeling, hvor det antages, at der haves et endeligt antal punkter nummeret $i = 1,\ldots,n$ med tilhørende, stokastiske variable $\{X_1,\ldots,X_n\}$. Da er sandsynligheden for hvert punkt X_i betinget med alle de andre punkter givet ved

$$P(x_i | x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n).$$

Det er jævnfør [Besag, 1974, s. 195] muligt at udvide teorien til at omfatte kontinuerte variable, selvom den her er noteret som bestående af diskrete variable.

Det antages nu, at hvis hver af værdierne $x_1,\ldots,x_n$ kan forekomme ved punkterne $1,\ldots,n$ hver for sig, så kan de også være på punkterne samtidig. Det vil sige, at hvis

sandsynligheden $P(x_i) > 0$ for hvert i, så er den simultane sandsynlighed $P(x_1,\ldots,x_n) > 0$. Jævnfør [Besag, 1974, s. 195] vil denne antagelse blive betegnet som *posivitetsbetingelsen.*

Det er muligt at skrive den simultane sandsynlighed som et produkt af betingede sandsynligheder, hvilket vises i sætning 3.2, men inden denne bevises, er det nødvendigt at definere mængden K_P.

Definition 3.1:
Mængden K_P er givet ved $K_P = \{\mathbf{x}|P(\mathbf{x}) > 0\}$ hvor $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \ldots & x_n \end{bmatrix}^T$.

Mængden K_P består derfor af alle de mulige sammensætninger af værdier for $x_1,\ldots,x_n$, der opfylder $P(x_1,\ldots,x_n) > 0$.

Sætning 3.2 (Brooks lemma):
For alle vektorer $\mathbf{x},\mathbf{y} \in K_P$ gælder

$$\frac{P(\mathbf{x})}{P(\mathbf{y})} = \prod_{i=1}^{n} \frac{P(x_i|x_1,\ldots,x_{i-1},y_{i+1},\ldots,y_n)}{P(y_i|x_1,\ldots,x_{i-1},y_{i+1},\ldots,y_n)}.$$

Bevis:
Den simultane sandsynlighed for værdierne $x_1,\ldots,x_n$ kan skrives ved brug af den betingede sandsynlighed $P(x_n|x_1,\ldots,x_{n-1})$, hvilket giver

$$P(\mathbf{x}) = P(x_n|x_1,\ldots,x_{n-1})P(x_1,\ldots,x_{n-1}). \tag{3.1}$$

Det er vanskeligt at finde $P(x_{n-1}|x_1,\ldots,x_{n-2})$ og dermed faktorisere hele højre side af (3.1) med betingede sandsynligheder på denne form, så i stedet anvendes det faktum, at

$$\begin{aligned} P(x_1,\ldots,x_{n-1},y_n) &= P(y_n|x_1,\ldots,x_{n-1})P(x_1,\ldots,x_{n-1}) \Rightarrow \\ P(x_1,\ldots,x_{n-1}) &= \frac{P(x_1,\ldots,x_{n-1},y_n)}{P(y_n|x_1,\ldots,x_{n-1})}, \end{aligned} \tag{3.2}$$

hvor y_n er den n'te værdi fra $\mathbf{y} \in K_P$. Indsættes (3.2) i (3.1) fås

$$P(\mathbf{x}) = P(x_n|x_1,\ldots,x_{n-1})\frac{P(x_1,\ldots,x_{n-1},y_n)}{P(y_n|x_1,\ldots,x_{n-1})}. \tag{3.3}$$

Leddet $P(x_1,\ldots,x_{n-1},y_n)$ kan nu omskrives med hensyn til x_{n-1} på samme måde, som $P(\mathbf{x})$ blev omskrevet med hensyn til x_n. Først ses det, at den simultane sandsynlighed kan skrives

$$P(x_1,\ldots,x_{n-1},y_n) = P(x_{n-1}|x_1,\ldots,x_{n-2},y_n)P(x_1,\ldots,x_{n-2},y_n).$$

Her kan $P(x_1,\ldots,x_{n-2},y_n)$ omskrives som i (3.2), ved

$$\begin{aligned} P(x_1,\ldots,x_{n-2},y_{n-1},y_n) &= P(y_{n-1}|x_1,\ldots,x_{n-2},y_n)P(x_1,\ldots,x_{n-2},y_n) \\ \Rightarrow P(x_1,\ldots,x_{n-2},y_n) &= \frac{P(x_1,\ldots,x_{n-2},y_{n-1},y_n)}{P(y_{n-1}|x_1,\ldots,x_{n-2},y_n)}, \end{aligned}$$

hvilket giver

$$P(x_1,\ldots,x_{n-1},y_n) = P(x_{n-1}|x_1,\ldots,x_{n-2},y_n)\frac{P(x_1,\ldots,x_{n-2},y_{n-1},y_n)}{P(y_{n-1}|x_1,\ldots,x_{n-2},y_n)}.$$

Indsættes dette i (3.3) bliver den simultane sandsynlighed til

$$\begin{aligned} P(\mathbf{x}) &= \frac{P(x_n|x_1,\ldots,x_{n-1})}{P(y_n|x_1,\ldots,x_{n-1})}P(x_1,\ldots,x_{n-1},y_n) \\ &= \frac{P(x_n|x_1,\ldots,x_{n-1})}{P(y_n|x_1,\ldots,x_{n-1})}\cdot\frac{P(x_{n-1}|x_1,\ldots,x_{n-2},y_n)}{P(y_{n-1}|x_1,\ldots,x_{n-2},y_n)}P(x_1,\ldots,x_{n-2},y_{n-1},y_n). \end{aligned}$$

Fortsættes metoden, fås det til sidst, at

$$P(\mathbf{x}) = \left(\prod_{i=1}^{n}\frac{P(x_i|x_1,\ldots,x_{i-1},y_{i+1},\ldots,y_n)}{P(y_i|x_1,\ldots,x_{i-1},y_{i+1},\ldots,y_n)}\right)P(\mathbf{y}). \tag{3.4}$$

hvor vektoren $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \ldots & y_n \end{bmatrix}^T$. Ligningen i sætning 3.2 følger af (3.4). ■

Den antagelse, at $P(x_i) > 0$ for hvert i, gør, at den simultane sandsynlighed $P(x_1,\ldots,x_n)$ også opfylder $P(x_1,\ldots,x_n) > 0$, hvilket i dette tilfælde sikrer, at nævneren i produktet i sætning 3.2 ikke er nul. Det kan også ses ud fra sætningen, at der findes mange måder at faktorisere $\frac{P(\mathbf{x})}{P(\mathbf{y})}$, da navngivningen af de forskellige punkter i systemet er arbitrær, men da resultatet i alle tilfældene er det samme, indikerer det, at det er begrænset hvilke former, den betingede sandsynlighedsfunktion kan have. Selvom den relative sandsynlighedsfunktion $\frac{P(\mathbf{x})}{P(\mathbf{y})}$ fra sætning 3.2 kan være forholdsvist simpel at finde i praksis, så kan det i nogle tilfælde være næsten umuligt at tolke likelihoodfunktionerne, når der haves absolutte sandsynlighedsfunktioner $P(\mathbf{x})$ og $P(\mathbf{y})$, da disse sandsynligheder generelt indeholder en uhåndterlig normaliseringskonstant.

3.2 Markovfelter

Markovfelter er rumlige systemer med definerede naborelationer mellem punkterne i hele eller dele af systemet, og de leder frem til en sætning af Hammersley og Clifford, der er vigtig ved konstruktion af gyldige, rumlige systemer. Denne sætning præsenteres sidst i dette afsnit. Først defineres mængden af naboer til hvert af punkterne, som er nummereret $i = 1,\ldots,n$.

Definition 3.3:
Punktet j er nabo til punktet i for $j \neq i$, hvis og kun hvis sandsynlighedsfordelingen $P(x_i|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$ afhænger af variablen x_j.

Et eksempel på ovenstående nabodefinition kan være præcisering af naboer i en Markovkæde $X_1,\ldots,X_n$, hvor X_i kun afhænger af X_{i-1} og X_{i+1}. Det vil sige, at når $2 \leq i \leq n-1$, så har X_i de to naboer X_{i-1} og X_{i+1}, mens X_1 og X_n kun har en nabo hver. Punkter, der er indbyrdes naboer, kan samles i såkaldte nabolag.

Definition 3.4 (Nabolag):
Et nabolag til punktet i er givet ved mængden $N_i = \{k | k \neq i \text{ er nabo til } i\}$.

Et eksempel er, hvis punkterne i et rumligt system danner et rektangulært gitter, hvor hvert punkt nummereres med (i,j) for heltal $i,j \in \{1,\ldots,n\}$. Hvis der for hvert punkt (i,j) gælder, at $P(x_{i,j}|x_{-(i,j)}) = P(x_i|x_{i,j-1},x_{i,j+1},x_{i-1,j},x_{i+1,j})$, hvor $x_{-(i,j)} = \{x_{k,m}|k,m \in (\{1,\ldots,n\}\backslash\{i,j\})\}$, så har hvert punkt (i,j) fire naboer. Det vil sige, nabolaget for i er givet ved $N_i = \{(i,j-1),(i,j+1),(i-1,j),(i+1,j)\}$.

Definition 3.5 (Klike):
En klike er mængden af et eller flere punkter, hvor alle punkterne i mængden indbyrdes er naboer.

For eksempel gælder det for en Markovkæde $X_1,\ldots,X_n$, at $\{X_1,X_2\}$ er en klike. Det er nu muligt at definere et Markovfelt.

Definition 3.6 (Markovfelt):
Et Markovfelt er et system med n punkter og et sandsynlighedsmål, hvor den betingede fordeling definerer nabostrukturen $\{N_i|i = 1,\ldots,n\}$.

Det kan her bemærkes, at definition 3.3 beskriver Markovegenskaben, så definitionen beskriver afhængighedsstrukturen i Markovfelter. Sætning 3.14, der også kaldes Hammersley og Clifford-sætningen, udtrykker sammenhængen mellem den simultane sandsynlighedsfunktion og det indbyrdes forhold mellem punkterne i det rumlige system, det vil sige, om de tilhører den samme klike. Sætningen kræver, at der først indføres en funktion Q, hvor dennes egenskaber vises i lemma 3.7, samt at der gøres følgende to antagelser:

1. Der er et endeligt antal værdier, som kan tilknyttes hvert af punkterne.
2. Punkterne kan have værdien nul.

Antagelse nummer to gør det muligt, at $P(\mathbf{0}) > 0$ for en nulvektor $\mathbf{0}$, da det er antaget tidligere, at hvis sandsynligheden $P(x_i) > 0$ for hvert i, så er den simultane sandsynlighed $P(x_1,\ldots,x_n) > 0$. Dermed er det muligt at definere funktionen $Q(\mathbf{x})$ som

$$Q(\mathbf{x}) = \ln\left(\frac{P(\mathbf{x})}{P(\mathbf{0})}\right) \tag{3.5}$$

for ethvert $\mathbf{x} \in K_P$. Derudover defineres $\mathbf{x}_i$ for et givet $\mathbf{x}$ som vektoren
$\mathbf{x}_i = \begin{bmatrix} x_1 & \ldots & x_{i-1} & 0 & x_{i+1} & \ldots & x_n \end{bmatrix}^T$.

Lemma 3.7:
Funktionen $Q(\mathbf{x})$ opfylder

$$\exp(Q(\mathbf{x}) - Q(\mathbf{x}_i)) = \frac{P(\mathbf{x})}{P(\mathbf{x}_i)} = \frac{P(x_i|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)}{P(0|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)}, \tag{3.6}$$

givet at naboerne til hvert punkt $i = 1,2,\ldots,n$ er kendte værdier. For enhver sandsynlighedsfordeling $P(\mathbf{x})$, der opfylder (3.6), eksisterer der desuden en entydig udvidelse af $Q(\mathbf{x})$ på mængden K_P givet ved

$$\begin{aligned} Q(\mathbf{x}) = &\sum_{1\le i\le n} x_i G_i(x_i) + \sum_{1\le i< j\le n}\sum x_i x_j G_{i,j}(x_i,x_j) + \sum_{1\le i<j<k\le n}\sum\sum x_i x_j x_k G_{i,j,k}(x_i,x_j,x_k) \\ &+ \ldots + x_1 x_2 \cdots x_n G_{1,2,\ldots,n}(x_1,x_2,\ldots,x_n) \end{aligned} \tag{3.7}$$

for $\mathbf{x} \in K_P$ og en funktion G.

Bevis:
Først bevises ligning (3.6) ved at indsætte (3.5) i udtrykket til venstre for lighedstegnet, hvilket giver

$$\begin{aligned} \exp\left(Q\left(\mathbf{x}\right) - Q\left(\mathbf{x}_i\right)\right) &= \exp\left(\ln\left(\frac{P\left(\mathbf{x}\right)}{P\left(\mathbf{0}\right)}\right) - \ln\left(\frac{P\left(\mathbf{x}_i\right)}{P\left(\mathbf{0}\right)}\right)\right) \\ &= \exp\left(\ln\left(\frac{P\left(\mathbf{x}\right)P\left(\mathbf{0}\right)}{P\left(\mathbf{0}\right)P\left(\mathbf{x}_i\right)}\right)\right) \\ &= \frac{P\left(\mathbf{x}\right)}{P\left(\mathbf{x}_i\right)}. \end{aligned} \tag{3.8}$$

Defineres det nu, at

$$z_j = \begin{cases} x_j \text{ når } j \neq i \\ 0 \text{ når } j = i, \end{cases}$$

så bliver (3.8) jævnfør sætning 3.2 til

$$\begin{aligned} \frac{P\left(\mathbf{x}\right)}{P\left(\mathbf{x}_i\right)} &= \prod_{j=1}^{n} \frac{P(x_j|x_1,\ldots,x_{j-1},z_{j+1},\ldots,z_n)}{P(z_j|x_1,\ldots,x_{j-1},z_{j+1},\ldots,z_n)} \\ &= \frac{P\left(x_j|x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n\right)}{P\left(0|x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n\right)}. \end{aligned}$$

Anden halvdel af lemmaet bevises ved at definere funktionen G. Først defineres $\mathbf{x}_I$ som mængden af x_i, hvor $i \in I$ og $I \subseteq \{1,\ldots,n\}$, det vil sige

$$\mathbf{x}_I = \{x_i | i \in I\}.$$

For mængden $I \subseteq \{1,\ldots,n\}$ defineres G_I som en funktion, der opfylder

$$\left(\prod_{i\in I} x_i\right) G_I\left(\mathbf{x}_I\right) = \sum_{\tilde{I}\subseteq I} (-1)^{|I\setminus\tilde{I}|} Q\left(\mathbf{x}_{\tilde{I}}\right),$$

hvor nullerne i notationen af Q er udeladt. Der summeres her over en delmængde $\tilde{I}$ af I, hvor fortegnet afhænger af antallet af elementer fra I, der er ikke er medtaget i delmængden $\tilde{I}$. Heraf følger det, at

$$x_i G_i(x_i) = Q\left(\underbrace{0,\ldots,0}_{i-1}, x_i, \underbrace{0,\ldots,0}_{n-i}\right) - Q(\mathbf{0}),$$

når $I = \{i\}$. På samme måde gælder der for $I = \{i,j\}$, hvor $i < j$, at

$$\begin{aligned} x_i x_j G_{i,j}(x_i, x_j) = & Q\left(\underbrace{0,\ldots,0}_{i-1}, x_i, \underbrace{0,\ldots,0}_{j-i-1}, x_j, \underbrace{0,\ldots,0}_{n-j}\right) \\ & - Q\left(\underbrace{0,\ldots,0}_{i-1}, x_i, \underbrace{0,\ldots,0}_{n-i}\right) - Q\left(\underbrace{0,\ldots,0}_{j-1}, x_j, \underbrace{0,\ldots,0}_{n-j}\right) + Q(\mathbf{0}), \end{aligned}$$

hvor der fortsættes på samme måde, når I indeholder flere elementer. Ifølge Möbius' inversionslemma 3.13 gælder det, at

$$Q(\mathbf{x}_I) = \sum_{\tilde{I} \subseteq I} \left(\prod_{i \in \tilde{I}} x_i\right) G_{\tilde{I}}(\mathbf{x}_{\tilde{I}}).$$

Resultatet (3.7) følger, når $I = \{1,\ldots,n\}$. ■

Entydighed opnås ved at definere $G_{i,j,\ldots} \equiv 0$ når $x_i = 0$, $x_j = 0$ eller en af de andre x-værdier er nul.

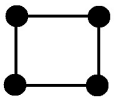
Sættes de to resultater, (3.6) og (3.7) sammen, kan det ses, at funktionen $Q(\mathbf{x})$ består af betingede sandsynligheder. For eksempel er

$$\begin{aligned} x_i G_i(x_i) &= \ln\left(\exp\left(Q(\underbrace{0,\ldots,0}_{i-1}, x_i, \underbrace{0,\ldots,0}_{n-i}) - Q(\mathbf{0})\right)\right) \\ &= \ln\left(\frac{P(x_i|\overbrace{0,\ldots,0}^{n-1})}{P(0|\underbrace{0,\ldots,0}_{n-1})}\right) = 0, \end{aligned}$$

der kan indsættes i (3.7) sammen med samme omskrivning for resten af funktionerne, der indeholder G.

Den simultane sandsynlighed $P(\mathbf{x})$ kan jævnfør sætning 3.2 skrives som produktet af de betingede sandsynligheder for hvert x_i fra vektoren $\mathbf{x}$. Des flere værdier n der er i $\mathbf{x}$, des flere led bliver der i faktoriseringen af den simultane sandsynlighed. Hvis der for eksempel haves fire punkter, der er forbundet som på figur 3.1, så kan disse generelt grupperes i delmængder bestående af den tomme mængde, et punkt (fire måder), to punkter (seks måder), tre punkter (fire måder) eller fire puntker. Det vil sige, at en faktorisering af den simultane sandsynlighed for de fire punkter vil bestå af fire gange så

mange led, som der er punkter. Hvis der gøres den antagelse, at alle punkter i hver gruppe skal være naboer indbyrdes, så vil de fire punkter kun have otte mulige grupperinger bestående af et eller to punkter.



Figur 3.1: Fire punkter forbundet så hvert punkt har to naboer.

Funktionen $Q(\mathbf{x})$ fra lemma 3.7 er en måde at omskrive den simultane sandsynlighed fra et produkt til en sum, og sætningen 3.14 viser, at summen kan reduceres betydeligt ved at sætte alle led med funktionen G lig nul, hvis de punkter, der indgår, ikke er indbyrdes naboer. I praksis reducerer det længden af $Q(\mathbf{x})$ betydeligt, og det kan være en stor fordel, når længden n af $\mathbf{x}$ er meget stor. Beviset for denne egenskab er nemmest at overskue, når datasættet betragtes som punkterne i en graf, og derfor vil næste afsnit introducere nogle relevante begreber fra grafteori, der er nødvendige for at kunne bevise sætning 3.14.

3.3 Grafer og Hammersley Clifford

Følgende afsnit er skrevet på baggrund af [Lauritzen, 1996].

Et datasæt af diskrete punkter, som for eksempel kan være huspriser tilknyttet kommuner, kan opfattes som punkter i en graf. I grafteori er det almindeligt at benytte notationen, der er givet i definition 3.8.

Definition 3.8:
En graf betegnes med $\mathcal{G} = (V,E)$, hvor V er mængden af punkter i grafen, og E er mængden af kanter mellem punkterne.

En graf $\mathcal{G}$ kaldes komplet, hvis alle punkterne i V er indbyrdes forbundet med kanter. På samme måde kan en delmængde af punkter være komplet, hvis de udgør en komplet graf, hvilket for eksempel kan være en klike. Derudover kan der tilknyttes en stokastisk variabel X_i til hvert punkt i i mængden af observationer V, hvor hver stokastisk variabel X_i kan antage en værdi x fra mængden χ, der består af reelle tal. En mængde af værdier fra χ skrives som vektoren $\mathbf{x}_A$, som indeholder de værdier, der hører til punkterne i delmængden $A \subset V$.

En graf kan have tre forskellige slags Markovegenskaber, der er givet i definitionerne herunder, og sammenhængen mellem Markovegenskaberne er givet i lemma 3.12.

Definition 3.9 (Parvis Markovegenskab):
Hvis der for et vilkårligt par af punkter (i,j) i grafen $\mathcal{G} = (V,E)$ gælder, at den stokastiske variabel X_i er uafhængig af X_j, når mængden af stokastiske variable $\{X_k | k \in V \backslash \{i,j\}\}$ er givet, så opfylder sandsynlighedsmålet P, der tager værdier i mængden χ, den parvise Markovegenskab med hensyn til $\mathcal{G}$.

I definition 3.10 anvendes betegnelsen for en aflukning om et punkt i givet ved $\text{cl}(i) = i \cup N_i$, hvilket er mængden af naboer til punktet i inklusive punktet selv.

Definition 3.10 (Lokal Markovegenskab):
Hvis der for ethvert punkt $i \in V$ for grafen $\mathcal{G} = (V,E)$ gælder, at X_i er uafhængig af mængden af stokastiske variable $\{X_k | k \in V \backslash \{\text{cl}(i)\}\}$, givet at $\{X_k | k \in N_i\}$ er kendt, så opfylder sandsynlighedsmålet P på χ den lokale Markovegenskab med hensyn til $\mathcal{G}$.

Definition 3.11 (Global Markovegenskab):
Lad (A,B,S) være tre, valgte delmængder $A,B,S \subseteq V$, hvor S adskiller mængderne A og B i $\mathcal{G}$. Hvis mængden af stokastiske variable $\{X_k | k \in A\}$ er uafhængig af $\{X_k | k \in B\}$ givet mængden $\{X_k | k \in S\}$ kendes, så har sandsynlighedsmålet P på χ den globale Markovegenskab.

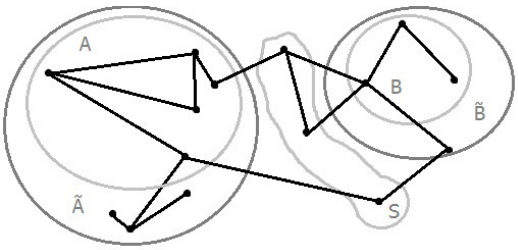
Generelt hører Markovegenskaben for en graf sammen med, om det er muligt at faktorisere sandsynlighedsmålet P på χ, hvilket også vises i lemma 3.12. Sandsynligheden P kan faktoriseres med hensyn til grafen $\mathcal{G}$, hvis der for alle delmængder $a \subseteq V$ eksisterer funktioner $\psi_a : \chi \to [0,\infty]$, hvor ψ_a kun afhænger af de x-værdier, der hører til punkterne i mængden a, hvilket vil sige vektoren $\mathbf{x}_a$. Derudover må der eksisterer et produktmål $\mu = \otimes_{i \in V} \mu_i$ på mængden χ så P har tætheden f med hensyn til μ. Sandsynlighedsmålet P kan da faktoriseres, idet f har formen

$$f(x) = \prod_{a \in C} \psi_a(x), \tag{3.9}$$

hvor C er mængden af kliker i $\mathcal{G}$. Produktmålet μ er en valgt værdi og grupperne af ψ_a kan multipliceres på forskellige måder, hvilket gør, at funktionerne ψ_a ikke er entydigt bestemt. De tre Markovegenskaber hænger sammen med faktoriseringen af P som vist i det følgende lemma.

Lemma 3.12:
Der gælder for enhver sandsynlighedsfordeling P på χ og enhver ikke-orienteret graf $\mathcal{G}$, at hvis P kan faktoriseres ⇒ Global Markovegenskab ⇒ Lokal Markovegenskab ⇒ Parvis Markovegenskab.

Bevis:
Det vises først, at hvis P kan faktoriseres så gælder den globale Markovegenskab, hvilket vil sige, det antages, at P kan faktoriseres. Derudover vælges tre, disjunkte delmængder $A,B,S \subseteq V$ så mængden S adskiller A og B. Herefter defineres en mængde $\tilde{A}$, der består af mængden A samt de punkter, der er forbundet til punkter i A, hvor der gælder, at den nye mængde tilhører $V \setminus S$ i grafen $\mathcal{G} = (V,E)$. Der defineres desuden mængden $\tilde{B} = V \setminus \left(\tilde{A} \cup S\right)$. Et eksempel på definition af sådanne mængder ses på figur 3.2.



Figur 3.2: Eksempel på definition af mængderne A, B, S, $\tilde{A}$ og $\tilde{B}$.

Elementerne i mængderne A og B tilhører hver sin del af forbundne punkter i mængden $V \setminus S$, da mængden S adskiller A og B. Dermed udgør enhver klike i V enten en delmængde i $\tilde{A} \cup S$ eller $\tilde{B} \cup S$. Hvis C_A er mængden af kliker i $\tilde{A} \cup S$, så fås det fra (3.9) at

$$f(x) = \prod_{a \in C} \psi_a(x) = \prod_{a \in C_A} \psi_a(x) \prod_{a \in C \setminus C_A} \psi_a(x) = h\left(\mathbf{x}_{\tilde{A} \cup S}\right) k\left(\mathbf{x}_{\tilde{B} \cup S}\right). \tag{3.10}$$

Ifølge [Lauritzen, 1996, s. 29] gælder, at den stokastiske variabel X er uafhængig af variablen Y givet Z kendes, hvis og kun hvis

$$f\left(x,y,z\right) = h\left(x,z\right) k\left(y,z\right) \tag{3.11}$$

for to funktioner h og k. Da (3.10) har samme form som (3.11), gælder det, at den stokastiske variabel $X_{\tilde{A}}$ er uafhængig af $X_{\tilde{B}}$ givet X_S er kendt. Fra [Lauritzen, 1996, s. 29] vides det, at hvis en stokastisk variabel X er uafhængig af Y givet Z kendes og $U = h(X)$, så er U uafhængig af Y givet Z. Da X_A kan opfattes som en funktion $X_A = h(\tilde{A})$, hvor h er en identitetsfunktion på A, og B kan opfattes som en funktion af $\tilde{B}$ giver resultatet fra [Lauritzen, 1996, s. 29], at X_A er uafhængig af X_B givet X_S er kendt, hvilket er den globale Markovegenskab fra definition 3.11.

For at bevise at den globale Markovegenskab medfører den lokale Markovegenskab, antages det, at den globale Markovegenskab gælder. Jævnfør definition 3.11 betyder, at der eksisterer tre delmængder $A,B,S \subseteq V$, hvor S adskiller A og B i $\mathcal{G}$. Hvis de tre mængder vælges som

$$\begin{aligned} A &= i \\ B &= V \setminus \mathrm{cl}(i) \\ S &= N_i, \end{aligned}$$

så gælder det jævnfør definition 3.11, at den stokastiske variabel X_i er uafhængig af mængden $\{X_k | k \in V \setminus \mathrm{cl}(i)\}$ givet mængden $\{X_k | k \in N_i\}$ kendes, hvilket svarer til den lokale Markovegenskab i definition 3.10.

For at vise at den lokale Markovegenskab medfører den parvise Markovegenskab, antages det, at den lokale Markovegenskab gælder, og at punkterne i og j ikke er nabopunkter. Ifølge [Lauritzen, 1996, s. 29] gælder det for tre stokastiske variable X, Y og Z, at hvis X er uafhængig af Y givet Z er kendt og $U = h(X)$, så er X uafhængig af Y givet (Z,U) er kendt. Den lokale Markovegenskab giver, at X_i er uafhængig af mængden $\{X_k | k \in V \backslash \{\mathrm{cl}(i)\}\}$ givet at $\{X_k | k \in N_i\}$ er kendt. Idet j ikke er nabo til i, hvilket også kan skrives

$$j \in V \backslash \mathrm{cl}(i),$$

så giver resultatet fra [Lauritzen, 1996, s. 29], at X_i er uafhængig af X_j givet de stokastiske variable tilhørende punkterne i mængden

$$N_i \cup \big((V \backslash \mathrm{cl}(i)) \backslash \{j\}\big) = V \backslash \{i,j\}$$

kendes. Dette er præcis den parvise Markovegenskab fra definition 3.9 for punkterne i og j. Dermed er lemmaet bevist. ■

Der er brug for endnu et lemma, inden sætning 3.14 kan bevises.

Lemma 3.13 (Möbius' inversionslemma):
Lad Ψ og Φ være funktioner defineret på mængden af alle delmængder af den endelige mængde V, der tager værdier i gruppen $(\mathbb{R},+)$. Da er udtrykkene

$$\Psi(a) = \sum_{b \subseteq a} \Phi(b) \quad \text{for alle } a \subseteq V \tag{3.12}$$

$$\Phi(a) = \sum_{b \subseteq a} (-1)^{|a \backslash b|} \Psi(b) \quad \text{for alle } a \subseteq V \tag{3.13}$$

ækvivalente.

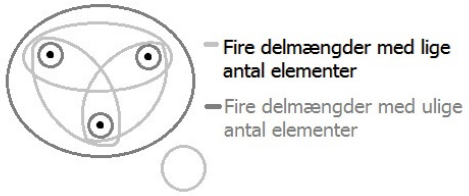
Bevis:
Det vises, at udtrykket (3.13) medfører udtrykket (3.12) ved at omskrive udtrykket (3.13) som

$$\begin{aligned} \sum_{b \subseteq a} \Phi(b) &= \sum_{b \subseteq a} \sum_{c \subseteq b} (-1)^{|b \backslash c|} \Psi(c) \\ &= \sum_{c \subseteq a} \Psi(c) \left(\sum_{c \subseteq b \subseteq a} (-1)^{|b \backslash c|} \right). \end{aligned}$$

Indføres mængden $d = b \backslash c$, fås det, at

$$\sum_{b \subseteq a} \Phi(b) = \sum_{c \subseteq a} \Psi(c) \left(\sum_{d \subseteq a \backslash c} (-1)^{|d|} \right), \tag{3.14}$$

Hvis a og c er ens, vil d være den tomme mængde, så (3.14) bliver til $\sum_{b \subseteq a} \Phi(b) = \sum_{c \subseteq a} \Psi(c)$. Hvis mængden $a \neq c$, er summen (3.14) lig nul, da enhver endelig mængde har samme antal delmængder, som består af et lige antal elementer, som det har delmængder, der består af et ulige antal elementer, hvilket figur 3.3 er et eksempel på.



Figur 3.3: Eksempel på at der til en endelig mænde af punkter eksisterer lige mange delmængder med et lige antal elementer, som der eksisterer delmængder med et ulige antal elementer.

Beviset at (3.12) medfører udtrykket (3.13) foregår på samme måde og udelades derfor. ■

Sætning 3.14 viser, at hvis en fordeling P opfylder den parvise Markovegenskab, så kan P faktoriseres og omvendt, hvilket gør, at egenskaberne nævnt i lemma 3.12 kan opfattes som ækvivalente (se [Lauritzen, 1996, s. 34] for yderligere bevis for ækvivalens).

Sætning 3.14 (Hammersley og Clifford):
Lad P være en sandsynlighedsfordeling, der har positiv og kontinuert tæthed f med hensyn til et produktmål μ. Sandsynlighedsmålet P kan faktoriseres med hensyn til en ikke-orienteret graf $\mathcal{G} = (V,E)$ hvis og kun hvis P opfylder den parvise Markovegenskab med hensyn til $\mathcal{G}$.

Bevis:
I lemma 3.12 blev det vist, at hvis et sandsynlighedsmål P kan faktoriseres, så gælder den parvise Markovegenskab, så det er kun nødvendigt at bevise, at hvis den parvise Markovegenskab gælder, så kan P faktoriseres. Det antages derfor at den parvise Markovegenskab gælder.

Hvis P kan faktoriseres, gælder ligning (3.9), og da det er antaget i sætning 3.14, at tætheden f er positiv, så det er muligt at tage logaritmen på begge sider af ligningen. Det giver

$$\log\big(f(x)\big) = \sum_{a \subseteq V} \varphi_a(x), \tag{3.15}$$

hvor $\varphi_a(x) = \log\big(\psi_a(x)\big)$, og derudover er $\varphi_a(x) = 0$, hvis a ikke er en klike i mængden V. Hvis der vælges en fikseret værdi $x^* \in \chi$, så kan funktionen

$$H_a(x) = \log\big(f\left(\mathbf{x}_a, \mathbf{x}^*_{a^c}\right)\big)$$

defineres for alle $a \subseteq V$. Her er a^c komplementærmængden til a, og $\left(\mathbf{x}_a, \mathbf{x}^*_{a^c}\right)$ er et element $\mathbf{y}$, der opfylder at $y_k = x_k$ for $k \in a$ og $y_k = x^*_k$ når $k \notin a$. Det vil sige, at funktionen $H_a(x)$ kun afhænger af værdierne $\mathbf{x}_a$ tilhørende punkterne i delmængden a, da x^* er konstant. Derudover defineres

$$\varphi_a(x) = \sum_{b \subseteq a} (-1)^{|a \setminus b|} H_b(x) \tag{3.16}$$

for alle $a \subseteq V$, hvilket vil sige, at φ_a også kun afhænger af $\mathbf{x}_a$. Anvendes Möbius' inversionslemma 3.13 medfører ligning (3.16), at

$$H_V(x) = \sum_{a \subseteq V} \varphi_a(x),$$

hvilket er det samme som udtrykket i ligning (3.15), så $\log\big(f(x)\big) = H_V(x)$.

Sætningen er bevist, hvis det kan vises, at $\varphi_a = 0$ når a ikke er en klike i V. Det antages derfor, at der eksisterer to punkter $i, j \in a$, der ikke er naboer, hvilket vil sige, mængden a ikke er en klike. Derudover defineres en mængde $c = a \backslash \{i, j\}$, og det fås fra (3.16), at

$$\begin{aligned} \varphi_a(x) &= \sum_{b \subseteq a} (-1)^{|a \backslash b|} H_b(x) \\ &= \sum_{b \subseteq c \cup \{i,j\}} (-1)^{|c \cup \{i,j\} \backslash b|} H_b(x) \\ &= \sum_{b \subseteq c} (-1)^{|c \backslash b|} \big(H_b(x) - H_{b \cup \{i\}}(x) - H_{b \cup \{j\}}(x) + H_{b \cup \{i,j\}}(x)\big) \end{aligned} \tag{3.17}$$

Der defineres en mængde $d = V \backslash \{i, j\}$, og da det er antaget, at den parvise Markovegenskab gælder, så er den stokastiske variabel X_i uafhængig af X_j, når mængden $\{X_k | k \in d\}$ er givet. Jævnfør [Lauritzen, 1996, s. 29] så gælder det for $\big(x, y, z\big) \neq (0,0,0)$, at

$$f\big(x, y, z\big) = f(x|z)\, f\big(y, z\big),$$

hvis den stokastiske variabel X er betinget uafhængig af Y givet Z er kendt. Dette resultat svarer til, at

$$f\big(x_i, x_j, \mathbf{x}_d\big) = f(x_i|\mathbf{x}_d)\, f\big(x_j, \mathbf{x}_d\big),$$

hvorved det fås, at

$$\begin{aligned} H_{b \cup \{i,j\}}(x) - H_{b \cup \{i\}}(x) &= \log\left(\frac{f\big(\mathbf{x}_b, x_i, x_j, \mathbf{x}^*_{d \backslash b}\big)}{f\Big(\mathbf{x}_b, x_i, x^*_j, \mathbf{x}^*_{d \backslash b}\Big)}\right) \\ &= \log\left(\frac{f\big(x_i|\mathbf{x}_b, \mathbf{x}^*_{d \backslash b}\big)\, f\big(\mathbf{x}_b, x_j, \mathbf{x}^*_{d \backslash b}\big)}{f\big(x_i|\mathbf{x}_b, \mathbf{x}^*_{d \backslash b}\big)\, f\Big(\mathbf{x}_b, x^*_j, \mathbf{x}^*_{d \backslash b}\Big)}\right). \end{aligned} \tag{3.18}$$

Anvendes det nu, at

$$\frac{f\big(x_i|\mathbf{x}_b, \mathbf{x}^*_{d \backslash b}\big)}{f\big(x_i|\mathbf{x}_b, \mathbf{x}^*_{d \backslash b}\big)} = 1 = \frac{f\big(x^*_i|\mathbf{x}_b, \mathbf{x}^*_{d \backslash b}\big)}{f\big(x^*_i|\mathbf{x}_b, \mathbf{x}^*_{d \backslash b}\big)},$$

bliver (3.18) til

$$\begin{aligned} H_{b \cup \{i,j\}}(x) - H_{b \cup \{i\}}(x) &= \log\left(\frac{f\big(x^*_i|\mathbf{x}_b, \mathbf{x}^*_{d \backslash b}\big)\, f\big(\mathbf{x}_b, x_j, \mathbf{x}^*_{d \backslash b}\big)}{f\big(x^*_i|\mathbf{x}_b, \mathbf{x}^*_{d \backslash b}\big)\, f\Big(\mathbf{x}_b, x^*_j, \mathbf{x}^*_{d \backslash b}\Big)}\right) \\ &= \log\left(\frac{f\big(\mathbf{x}_b, x^*_i, x_j, \mathbf{x}^*_{d \backslash b}\big)}{f\Big(\mathbf{x}_b, x^*_i, x^*_j, \mathbf{x}^*_{d \backslash b}\Big)}\right) \\ &= H_{b \cup \{j\}}(x) - H_b(x). \end{aligned}$$

Udtrykket (3.17) bliver dermed til

$$\varphi_a(x) = \sum_{b \subseteq c} (-1)^{|c \setminus b|} \left(H_b(x) - H_b(x) - H_{b \cup \{j\}}(x) + H_{b \cup \{j\}}(x) \right) = 0,$$

og dermed er sætningen bevist. ■

Resultatet fra sætning 3.14 kan også formuleres på en anden måde, der involvere funktionen G fra lemma 3.7. Hvis mængden $K_P = \{\mathbf{x} | P(\mathbf{x}) > 0\}$, hvor $\mathbf{x} = (x_1, \ldots, x_n)$, er et Markovfelt, der opfylder posivitetsbetingelsen, så opfylder funktionen $Q(\mathbf{x})$ fra lemma 3.7, at

$$G_{i,j,\ldots,s}\left(x_i, x_j, \ldots, x_s\right) \equiv 0$$

hvis punkterne $1 \leq i < j < \ldots < s \leq n$ ikke danner en klike. Det vil sige, at den simultane sandsynlighed $P(\mathbf{x})$ kan skrives som produkt af funktioner af kliker og dermed forsimples udregningerne, i stedet for at skulle tage alle G-funktioner fra (3.7) i betragtning.

3.4 Automodeller og CAR-modellen

Afsnittet er skrevet på baggrund af [Besag, 1974] og [Besag and Kooperberg, 1995].

Automodeller er en delmængde af Markovfelter og er defineret ved, at funktionen $Q(\mathbf{x})$ fra lemma 3.7 har formen

$$Q(\mathbf{x}) = \sum_{1 \leq i \leq n} x_i G_i(x_i) + \sum_{1 \leq i < j \leq n} \sum \beta_{i,j} x_i x_j,$$

hvor $\beta_{i,j} = 0$ når punkterne i og j ikke er naboer. Den betingede sandsynlighed for automodellller findes ved at tage udgangspunkt i to antagelser.

Det antages først, at sandsynlighedsstrukturen for et system af punkter kun afhænger af kliker, der hver indeholder maksimalt to punkter. Dermed bliver ligning (3.7) til

$$Q(\mathbf{x}) = \sum_{1 \leq i \leq n} x_i G_i(x_i) + \sum_{1 \leq i < j \leq n} \sum x_i x_j G_{i,j}\left(x_i, x_j\right), \tag{3.19}$$

hvor $G_{i,j}\left(x_i, x_j\right) \equiv 0$, når punkterne i og j ikke er naboer.

Den anden antagelse er, at den betingede sandsynlighedsfordeling for hvert punkt tilhører den eksponentielle familie. Skrives den eksponentielle familie så den gælder for hvert punkt $i \in \{1, \ldots, n\}$ og den tilhørende værdi x_i, kan tætheden skrives på formen

$$\ln p_i\left(x_i | x_{N_i}\right) = f_i\left(x_{N_i}\right) \cdot h_i(x_i) + g_i(x_i) + s_i\left(x_{N_i}\right),$$

hvor x_{N_i} er mængden $x_{N_i} = \{x_j | j \in N_i\}$, mens h_i og g_i er givne funktioner, og funktionen $p_i(\ldots)$ er notation for den betingede sandsynlighedsfordeling af X_i givet værdierne for alle andre punkter end i. Derudover er f_i og s_i funktioner af de værdier, der hører til nabopunkterne til i, og funktionen s_i er en normaliseringskonstant. Funktionen f_i bestemmer typen af afhængighed mellem x_i og naboværdierne.

Når $Q(\mathbf{x})$ har formen (3.19), har automodellerne betinget sandsynlighed givet ved

$$\frac{p_i\left(x_i|x_{N_i}\right)}{p_i\left(0|x_{N_i}\right)} = \prod_{i=1}^{n} \frac{P(x_i|x_1,\ldots,x_{i-1},y_{i+1},\ldots,y_n)}{P(y_i|x_1,\ldots,x_{i-1},y_{i+1},\ldots,y_n)}, \tag{3.20}$$

hvor $x_{N_i} = \{x_j | j \in N_i\}$, mens $\mathbf{y} = \begin{bmatrix} x_1 & \ldots & x_{i-1} & 0 & x_{i+1} & \ldots & x_n \end{bmatrix}^T$ og $\mathbf{x} = \begin{bmatrix} x_1 & \ldots & x_n \end{bmatrix}^T$. Anvendes sætning 3.2 og resultatet (3.6) fra lemma 3.7 bliver ovenstående til

$$\begin{aligned}
\frac{p_i\left(x_i|x_{N_i}\right)}{p_i\left(0|x_{N_i}\right)} &= \frac{P(\mathbf{x})}{P(\mathbf{y})} \\
&= \exp(Q(\mathbf{x}) - Q(\mathbf{y})) \\
&= \exp\Bigg(\sum_{1\le i\le n} x_i G_i\left(x_i\right) + \sum_{1\le i< j\le n} \sum x_i x_j G_{i,j}\left(x_i,x_j\right) - \sum_{1\le i\le n} y_i G_i\left(y_i\right) \\
&\quad - \sum_{1\le i< j\le n} \sum y_i y_j G_{i,j}\left(y_i,y_j\right) \Bigg) \\
&= \exp\left(x_i \left(G_i\left(x_i\right) + \sum_{j=1}^{n} \beta_{i,j} x_j \right) \right),
\end{aligned}$$

hvor $\beta_{i,j} = \beta_{j,i}$, og $\beta_{i,j} = 0$ hvis punkterne i og j ikke er naboer. Sidste lighedstegn opnås ved, at de fleste led går ud.

CAR-modellen

Det er nu muligt at definere CAR-modellen, der er en model for n rumlige observationer $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \ldots & y_n \end{bmatrix}^T$, der knytter værdierne y_i, $i = 1,\ldots,n$, til n punkter i rummet. CAR-modellen for en mængde observationer $\mathbf{y}$ fremkommer, når sandsynlighedsfunktionen $p_i\left(y_i|y_{N_i}\right)$ fra (3.20) er normalfordelt, hvilket vil sige, at hvert y_i er fordelt med

$$p_i\left(y_i|y_{N_i}\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\sigma^{-2} \left(y_i - \mu_i - \sum_{j=1}^{n} \beta_{i,j}\left(y_j - \mu_j\right) \right)^2 \right).$$

Det leder frem til definition 3.15.

Definition 3.15:
CAR-modellen har multivariat normalfordelte observationer $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \ldots & y_n \end{bmatrix}^T$ med sandsynlighedsfordeling

$$p\left(\mathbf{y}\right) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}n} |B|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sigma^{-2} \left(\mathbf{y}-\boldsymbol{\mu}\right)^T B \left(\mathbf{y}-\boldsymbol{\mu}\right) \right),$$

hvor $\boldsymbol{\mu}$ er en $(n \times 1)$-vektor af endelige middelværdier μ_i og B er en $(n \times n)$-matrix, der har nuller på diagonalen, og hvor alle andre elementer er $-\beta_{ij}$ for $i, j = 1,\ldots,n$.

Matricen B er defineret til at være symmetrisk, hvilket vil sige, at $\beta_{ij} = \beta_{ji}$, og det kræves også, at den er positiv definit. Derudover gælder det, at hvis to af de n punkter, i og j, er

naboer, så er elementet $\beta_{ij} \neq 0$. Det kan tjekkes, at matricen B er positivt definit ved at se, om udtrykket $\mathbf{x}^T B\mathbf{x}$ er positivt for en $(n \times 1)$-vektor $\mathbf{x}$, hvilket giver

$$\mathbf{x}^T B\mathbf{x} = \sum_{i=1}^{n} B_{i+} x_i^2 - \sum_{i<j} B_{ij} \left(x_i - x_j\right)^2, \tag{3.21}$$

hvor

$$B_{i+} = \sum_{j=1}^{n} B_{ij}.$$

Det antages nu, at der for matricen B gælder, at $\beta_{i+} \leq 1$ for alle i, hvor der for mindst et i gælder, at $\beta_{i+} < 1$ og at mindst et $\beta_{ij} \neq 0$. Elementerne fra matricen B kan nu indsættes i (3.21), hvor $\beta_{ij} \neq 0$ når punkterne i og j er naboer, mens $\beta_{ij} = 0$ når de to punkter i og j ikke er naboer, hvor $i \neq j$. Derudover er $\beta_{ii} = 0$. Det giver

$$\begin{aligned}
\mathbf{x}^T B\mathbf{x} &= \sum_{i=1}^{n} \left(\sum_{j} B_{ij} \right) x_i^2 - \sum_{i<j} B_{ij} (x_i - x_j)^2 \\
&= \sum_{i=1}^{n} \left(\left(\sum_{j \neq i} -\beta_{ij} \right) + 0 \right) x_i^2 - \sum_{i<j} \left(-\beta_{ij}\right) \left(x_i - x_j\right)^2 \\
&= \underbrace{\sum_{i=1}^{n} \left(-\beta_{i+} x_i^2\right)}_{>0} + \underbrace{\sum_{i<j} \beta_{ij} \left(x_i - x_j\right)^2}_{\geq 0}.
\end{aligned}$$

Da det er antaget, at $\beta_{i+} \leq 1$ for alle i og at alle $\beta_{ij} \geq 0$, hvor der for mindst et i gælder at $\beta_{i+} < 1$, så er B positiv definit. Matricen B indgår i kovariansmatricen

$$\left(\frac{1}{\sigma^2} B \right)^{-1},$$

og derudover er middelværdien givet ved

$$\mathrm{E}[Y_i | y_j \text{ for } j \neq i] = \mu_i + \sum \beta_{i,j} \left(y_j - \mu_j\right).$$

Stationær CAR-model

Hvis de observerede punkter danner et regulært gitter, så er det oplagt at vælge CAR-modellen som model for punkterne. Uendelige systemer af punkter kan betragtes som endelige systemer ved hjælp af stationære autoregressioner, og et godt redskab i den forbindelse er den autokovariansgenererende funktion. Sammenhængen mellem den stationære CAR-model og den tilhørende autokovariansgenererende funktion er forholdsvis simpel.

En autokovariansgenererende funktion tilhører samme kategori af funktioner som en såkaldt momentgenererende funktion, der kan anvendes til at finde middelværdien for en funktion. Hvis der eksempelvis haves en mængde stokastiske variable $Y_1, \ldots, Y_n$ med fordelingen $\exp(tY)$ og den momentgenererende funktion $M = \mathrm{E}[\exp(tY)]$, så kan middelværdien findes ved

$$\left. \frac{\partial M(t)}{\partial t} \right|_{t=0} = \left(\mathrm{E}[Y \cdot \exp(tY)] \right)_{t=0} = \mathrm{E}[Y].$$

På samme måde findes der en autokovariansgenererende funktion for CAR-modellen givet ved

$$G(z_1,z_2) = \sum_k \sum_l \gamma_{k,l} z_1^k z_2^l$$

hvor $\gamma_{k,l} = \text{Cov}\left(Y_{i,j}, Y_{i+k,j+l}\right)$. Den stationære udgave af CAR-modellen er defineret ud fra den autokovariansgenererende funktion.

Definition 3.16:
De stokastiske variable $\{Y_{i,j} | i,j = 0, \pm 1, \pm 2, \ldots\}$ udgør en stationær, gaussisk proces, hvis tætheden for $Y_{i,j}$ opfylder definition 3.15, og processen er af endelig orden. Derudover tilhører de observerede punkter et uendeligt, rektangulært gitter, og processen har autokovariansgenererende funktion givet ved

$$c\left(1 - \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \beta_{k,l} z_1^k z_2^l\right)^{-1}, \tag{3.22}$$

hvor c er en konstant, der er forbundet med variansen af $Y_{i,j}$. Derudover gælder det at

- kun et endeligt antal af de reelle koefficienter $\beta_{k,l}$ er ulig nul,
- $\beta_{0,0} = 0$,
- $\beta_{-k,-l} \equiv \beta_{k,l}$ og
- $\sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \beta_{k,l} z_1^k z_2^l < 1$ når $|z_1| = |z_2| = 1$.

Det er nødvendigt med en ny definition af naboer, når der ses på en stationær CAR-model. Punktet $(i-k, j-l)$ er nabo til punktet (i,j) hvis og kun hvis (i,j) tilhører en stationær, gaussisk proces med autokovariansgenererende funktion givet ved (3.22) og $\beta_{i,j} \neq 0$.

Derudover gælder det, at hvis $|r| + |s| > 0$, så medfører ligning (3.22), at

$$\rho_{r,s} = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \beta_{k,l} \rho_{r-k,s-l}. \tag{3.23}$$

Her er $\rho_{r,s}$ autokorrelationen af lag nummer r og s i henholdsvis punkt i og j. Herefter defineres mængden $\{\varepsilon_{i,j} | i,j \in \mathbb{Z}\}$, der kan vise sig at være en endelig mængde, ud fra processen

$$\begin{aligned} Y_{i,j} &= \alpha + \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \beta_{k,l} Y_{i-k,j-l} + \varepsilon_{i,j} \qquad (3.24) \\ \alpha &= \left(\sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} \beta_{k,l}\right) \mu \\ \mu &= \text{E}[Y_{i,j}]. \end{aligned}$$

Dermed er $\varepsilon_{i,j}$ gaussiske variable med middelværdi nul og varians σ^2. Ligning (3.23) antyder desuden, at hvis $|i-i'|+|j-j'|>0$, så er $\varepsilon_{i,j}$ og $Y_{i',j'}$ ikke korrelerede. Sammen med (3.24) gør det, at givet værdierne for enhver, endelig mængde af punkter, der indeholder naboerne til (i,j), så har $Y_{i,j}$ betinget middelværdi

$$\mathrm{E}[Y_{kl}|Y_{-(k,l)}] = \alpha + \sum_k \sum_l \beta_{k,l} y_{i-k,j-l}$$

hvor $Y_{-(u,v)}$ er mængden af alle variable $\{Y_{ij}|(i,j)\neq(u,v)\}$. Variansen er givet ved σ^2, der er uafhængig af de omkringliggende værdier.

3.5 Sammenhæng mellem CAR- og SAR-modellen

Jævnfør definition 3.15 og [Wall, 2003] kan CAR-modellen skrives som en lokalt betinget model ved

$$\begin{aligned} Y_i|\mathbf{Y}_{-i} &\sim N\left(\mu_i + \sum_{j=1}^{n} \beta_{ij}\left(Y_j-\mu_j\right), \tau_i^2\right) \qquad (3.25)\\ \mathbf{Y}_{-i} &= \{Y_j|\, j\neq i\}\\ E[Y_i] &= \mu_i\\ B &= [-\beta_{ij}]\\ \Sigma &= [\tau_i^2], \end{aligned}$$

hvor Y_i er den stokastiske variabel tilknyttet punktet i, τ_i^2 er den betingede varians og β_{ij} er kendte eller ukendte konstanter med $\beta_{ii}=0$ for $i=1,...,n$. Her kan variansen også skrives som $\Sigma=\mathrm{diag}\{\tau_1^2,\tau_2^2,\ldots,\tau_n^2\}$. I det tilfælde at der haves et endeligt antal observationer n, kan der ligesom ved SAR-modellen opskrives en nabomatrix W med elementerne w_{ij}, der viser naborelationerne mellem observationerne $\mathbf{y}$, og som i SAR-modellen er defineret ved

$$\begin{aligned} w_{ij} &= 1 \quad \text{hvis } i \text{ er nabo til } j\\ w_{ij} &= 0 \quad \text{hvis } i=j\\ w_{ij} &= 0 \quad \text{ellers.} \end{aligned}$$

Som beskrevet efter definition 2.2 er det muligt at gøre vægtmatricen W rækkestokastisk, så summen af hver række er en. Nabomatricen W er indeholdt i matricen B for CAR-modellen, hvilket uddybes i beviset for sætning 3.18.

Ved at anvende sætning 3.2 fås den globale version af fordelingen for $\mathbf{Y}$ til

$$\mathbf{Y} \sim N_n\left(\boldsymbol{\mu}, (I_n-B)^{-1}\Sigma\right). \qquad (3.26)$$

CAR-modellen fra definition 3.15 er veldefineret, hvis kovariansmatricen er symmetrisk, hvilket bevises i følgende lemma.

Lemma 3.17:
Kovariansmatricen for CAR-modellen fra definition 3.15 er symmetrisk, hvis matricen $(I_n - B)^{-1}\Sigma$ er invertibel og

$$w_{ij}\tau_j^2 = w_{ji}\tau_i^2,$$

hvor w_{ij} er et element fra den rækkestokastiske, symmetriske nabomatrix W, mens $\tau_i^2 = \frac{\sigma^2}{w_{i+}}$ er den skalerede, betingede varians for observation y_i.

Bevis:
Det gælder generelt, at $(I_n - B)^{-1}\Sigma$ er symmetrisk, hvis og kun hvis den inverse $\Sigma^{-1}(I_n - B)$ også er symmetrisk. Det antages, at

$$w_{ij}\tau_j^2 = w_{ji}\tau_i^2,$$

hvilket medfører, at der for $i \neq j$ og $B = \rho W$ gælder

$$\begin{aligned}[\Sigma^{-1}(I_n - B)]_{ij} &= \tau_i^{-2}\left(0 - w_{ij}\right) = \tau_j^{-2}\left(0 - w_{ji}\right) \\ &= [\Sigma^{-1}(I_n - B)]_{ji}. \end{aligned} \tag{3.27}$$

Ligning (3.27) medfører, at $\Sigma^{-1}(I_n - B)$ er symmetrisk. ■

Lemma 3.17 gælder også i det tilfælde hvor W er symmetrisk og binær, så w_{ij} er enten 0 eller 1, og der haves konstant varians $\tau^2 = \sigma^2$.

Der ses nu på sammenhængen mellem CAR- og SAR-modellen. SAR-modellen har formen som defineret i definition 2.7, og ses der på skrivemåden (2.1), kan CAR-modellen skrives på lignende form og endda omskrives til SAR-modellen. Formen for CAR-modellen er angivet i ligning (3.25) og (3.26), og SAR-modellen kendes fra definition 2.4. Hvis SAR- og CAR-modellen er to måder at skrive den samme model på, må der gælde, at kovariansmatricerne er ens, så

$$\left(I_n - \rho_s W\right)^{-1}\sigma^2\left(\left(I_n - \rho_s W\right)^{-1}\right)^T = (I_n - B)^{-1}\Sigma \tag{3.28}$$

Kovariansmatricen for SAR-modellen ses i ligning (2.11). Isoleres B i ovenstående, fås

$$\begin{aligned}(I_n - B)^{-1}\Sigma = \left(I_n - \rho W\right)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T &\Rightarrow \\ I_n - B = \left(\left(I_n - \rho W\right)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T\Sigma^{-1}\right)^{-1} & \\ = \Sigma\left(I_n - \rho W\right)^T\sigma^{-2}\left(I_n - \rho W\right) &\Rightarrow \\ B = I_n - \Sigma\left(I_n - \rho W\right)^T\sigma^{-2}\left(I_n - \rho W\right). &\end{aligned}$$

Indsættes dette i CAR-modellen fås SAR-modellen. Det er også muligt at isolere ρW i kovariansmatricen for SAR-modellen for at finde en ligning at indsætte i SAR-modellen, hvorved CAR-modellen fås. Resultatet er den datagenererende funktion for CAR-modellen, der ses i sætning 3.18.

Sætning 3.18 (Datagenererende funktion):
Den datagenererende funktion for CAR-modellen er givet ved

$$\mathbf{y} = \left(I - \rho W\right)^{-1} X\boldsymbol{\beta} + \left(\left(I_n - \rho W\right)^{-1}\right)^{\frac{1}{2}} \boldsymbol{\varepsilon},$$

hvor den rækkestokastiske vægtmatrix W opfylder $w_{ij}\tau_j^2 = w_{ji}\tau_i^2$, X er en $(n \times k)$-matrix over baggrundsvariable og $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}^T$ er en vektor over n observationer. Derudover er ρ en parameter, $\boldsymbol{\beta}$ er en parametervektor og $\boldsymbol{\varepsilon} \sim N(0,\Sigma)$ er en fejlvektor.

Bevis:
Det antages, at SAR- og CAR-modellen er to udtryk for samme model, så kovarianserne er ens som i (3.28). Herefter defineres matricerne $D_1 = (I_n - B)^{-1}\Sigma$ og $D_2 = \left(I_n - \rho W\right)^{-1}\sigma$, hvilket medfører, at (3.28) bliver til

$$(I_n - B)^{-1}\Sigma = \left(I_n - \rho W\right)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T \Leftrightarrow \tag{3.29}$$

$$D_1 = D_2 D_2^T. \tag{3.30}$$

Udtrykket (3.30) er opfyldt, hvis $D_2 = D_1^{\frac{1}{2}}$, hvor kvadratroden af D_1 for eksempel kan findes ved hjælp af Cholesky dekomposition. Dermed er (3.29) ækvivalent med

$$\left(I_n - \rho W\right)^{-1}\sigma = \left((I_n - B)^{-1}\Sigma\right)^{\frac{1}{2}} \Rightarrow$$

$$\rho W = I_n - \left(\left((I_n - B)^{-1}\Sigma\right)^{\frac{1}{2}}\sigma^{-1}\right)^{-1}. \tag{3.31}$$

CAR-modellen fremkommer ved at indsætte (3.31) i SAR-modellen $\mathbf{y} = \boldsymbol{\mu} + \left(I_n - \rho W\right)^{-1}\boldsymbol{\varepsilon}$, hvorved

$$\begin{aligned}
\mathbf{y} &= \boldsymbol{\mu} + \left(I_n - \rho W\right)^{-1}\boldsymbol{\varepsilon} \qquad (3.32)\\
&= \boldsymbol{\mu} + \left(I_n - \left(I_n - \left(\left((I_n - B)^{-1}\Sigma\right)^{\frac{1}{2}}\sigma^{-1}\right)^{-1}\right)\right)^{-1}\boldsymbol{\varepsilon}\\
&= \boldsymbol{\mu} + \left((I_n - B)^{-1}\Sigma\right)^{-\frac{1}{2}}\sigma^{-1}\boldsymbol{\varepsilon}.
\end{aligned}$$

Her er

$$\begin{aligned}
\varepsilon_i &\sim N(0,\sigma) \text{ for alle } i \in V\\
\sigma^{-1}\boldsymbol{\varepsilon} &\sim N(0,I_n)\\
\Sigma^{\frac{1}{2}}\sigma^{-1}\boldsymbol{\varepsilon} &\sim N(0,\Sigma).
\end{aligned}$$

Antages det, at

$$\begin{aligned}
\boldsymbol{\mu} &= \left(I - \rho W\right)^{-1} X\boldsymbol{\beta}\\
B &= \rho W
\end{aligned}$$

giver ligning (3.32), at

$$\mathbf{y} = \left(I - \rho W\right)^{-1} X\boldsymbol{\beta} + \left(\left(I_n - \rho W\right)^{-1}\right)^{\frac{1}{2}} \boldsymbol{\varepsilon}.$$

Da det er antaget $w_{ij}\tau_j^2 = w_{ji}\tau_i^2$, så er matricen $(I_n - B)^{-1}\Sigma$ ifølge lemma 3.17 symmetrisk, og dermed er CAR-modellen veldefineret. ■

CAR-modellen kan skrives på en form der ligner SAR-modellen i definition 2.4, idet

$$\begin{aligned}\mathbf{y} &= \left(I-\rho W\right)^{-1} X\boldsymbol{\beta} + \left(\left(I_n-\rho W\right)^{-1}\right)^{\frac{1}{2}}\boldsymbol{\varepsilon} \Rightarrow \\ \mathbf{y} &= \rho W\mathbf{y} + X\boldsymbol{\beta} + \left(I_n-\rho W\right)^{\frac{1}{2}}\boldsymbol{\varepsilon},\end{aligned}$$

hvor det ses, at forskellen mellem de to modeller er leddet $\left(I_n-\rho W\right)^{\frac{1}{2}}$ foran fejlvektoren.

Rumlig afhængighed 4

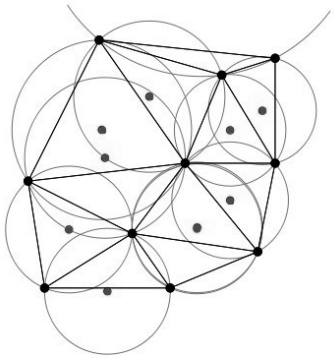
Både SAR- og CAR-modellen er rumlige regressionsmodeller, der forsøger at afbilde den i mange tilfælde komplicerede afhængighedsstruktur mellem observationer. Her spiller vægtmatricen W en stor rolle for begge modeller, da den definerer hvilke observationer, der er naboer og dermed kan påvirke hinanden, også gennem anden- og tredjegradsnaboer, mens parameteren ρ beskriver graden af afhængighed mellem naboerne. Hvis der haves mange observationer, er det en næsten uoverskuelig opgave at sammenholde alle par af punkter manuelt og derefter via matricen W definere, om de er naboer eller ej. Derfor vil der i afsnit 4.1 blive gennemgået en praktisk metode til at afgøre, om et punktpar er naboer eller ej og dermed definere matricen W. Når naborelationerne er defineret, er det interessant at se på, hvordan observationerne præcist påvirker hinanden gennem naborelationerne. Det vil der blive set nærmere på i afsnittet 4.2, hvor påvirkningerne vil blive kategoriseret i direkte, indirekte og samlede påvirkninger. Det er eksempelvis interessant at betragte de indirekte påvirkninger, der betegner, at en ændring i en baggrundsvariabel for en observation $y_i \in \mathbf{y}$ kan påvirke nabopunktet $y_j \in \mathbf{y}$, hvilket jo netop er beskrivelsen af rumlig afhængighed.

4.1 Nabomatrix ved brug af Voronoi-diagram

Afsnittet er skrevet ud fra [LeSage and Pace, 2009, s. 118-120] og [Chen et al., 2012, s. 24-31].

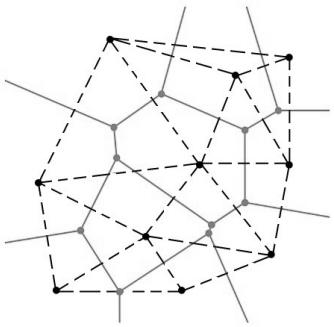
Naborelationerne i matricen W for SAR- og CAR-modellen afhænger af, hvordan nabopunkter defineres. En metode er at udregne afstanden mellem diskrete punkter og fastsætte en maksimal afstand for nabopunkter, så punkter, der ligger længere fra hinanden end denne afstand, ikke defineres som naboer. En anden metode er at definere et område omkring hvert af de diskrete punkter, så hele det plan, punkterne er observeret i, er inddelt i polygoner. Dermed kan polygonernes indbyrdes forhold anvendes til at definere naborelationerne. I dette afsnit ses der nærmere på sidstnævnte metode, og hvordan det er muligt at anvende Voronoi-diagrammer til at definere naboer.

Delaunay-trekanter anvendes til at finde Voronoi-diagrammet for en mængde observationer og er en metode til at inddele planen i trekanter. Hver trekant består af tre observationer og konstrueres således, at den omskrevne cirkel for hver trekant kun indeholder de tre observationer, trekanten består af. Der er ikke en entydig måde at inddele planen i Delaunay-trekanter på. Et eksempel på en Delaunay-tesselation i to dimensioner ses på figur 4.1.



Figur 4.1: Delaunay-trakanter over 10 punkter og de tilhørende, omskrevne cirkler, hvor centrum for hver cirkel er markeret med grå. [Encyclopedia, Opslag: Delaunay triangulation]

Når de n observationer er inddelt i Delaunay-trekanter er det muligt at lave et Voronoi-diagram, hvilket gøres ved at forbinde centrene for alle de omskrevne cirkler til trekanterne. Hvis der for eksempel laves et Voronoi-diagram over figur 4.1, bliver resultatet som på figur 4.2.



Figur 4.2: Gråt Voronoi-diagram over Delaunay-trekanterne fra figur 4.1, der er angivet med stiplede streger [Encyclopedia, Opslag: Delaunay triangulation]

Et Voronoi-diagram har den egenskab, at alle de punkter, der er indeholdt i polygonen om hver observation, er tættere på observationen i polygonen end på nogen af de andre observationer. Generelt har hvert punkt gennemsnitligt seks naboer, hvor det antages, at to punkter er naboer, hvis de har en fælles polygonside. Hvis der haves n observationer i alt, hvorpå der først foretages en Delaunay-tesselation og derefter laves et Voronoi-diagram over, så vil der være omkring $6n$ Delaunay-trekanter. Når Delaunay-trekanterne er defineret, er det kun nødvendigt med $O(n)$ operationer (tidskompleksitet uddybes i appendiks A) for at konstruere vægtmatricen W, der oftest er tyndt besat. Samlet er det muligt at inddele planen i Delaunay-trekanter og derefter konstruere et Voronoi-diagram med $O(n\ln(n))$ operationer, når der haves data i to dimensioner.

4.2 Påvirkninger mellem observationer

Afsnittet er skrevet på baggrund af [LeSage and Pace, 2009, s. 33-41].

Der vil nu blive set nærmere på, hvordan observationer påvirker hinanden indbyrdes, og hvordan påvirkningerne kan beskrives teoretisk for SAR- og CAR-modellen. Påvirkninger kan inddeles i direkte, indirekte og samlede påvirkninger, hvor de direkte eksempelvis er, når en ændring i en baggrundsvariabel til observation $y_i \in \mathbf{y}$ påvirker y_i selv, mens påvirkningen på nabopunktet y_j fra samme ændring kaldes indirekte. Inden de forskellige kategorier af påvirkninger defineres, er det interessant at se nærmere på, hvor påvirkningerne stammer fra, og hvordan de kan beskrives. Her kan der eksempelvis tages udgangspunkt i en generel regressionsmodel, som beskriver en mængde lineære og uafhængige observationerne $\mathbf{y}$ ved

$$\mathbf{y} = \sum_{r=1}^{k} \mathbf{x}_r \beta_r + \boldsymbol{\varepsilon}, \tag{4.1}$$

hvor vektoren $\mathbf{x}_r$ er den r'te søjle i matricen med baggrundsvariable X, og parameteren β_r er den r'te indgang i vektoren $\boldsymbol{\beta}$. Udregnes de partielt afledede for modellen, fås

$$\frac{\partial y_i}{\partial x_{ir}} = \frac{\partial}{\partial x_{ir}} \left(x_{ir}\beta_r + \varepsilon_i \right) = \beta_r \qquad \forall \quad i,r$$
$$\frac{\partial y_i}{\partial x_{jr}} = 0 \qquad \forall r \text{ når } j \neq i.$$

En måde at betragte de partielt afledede på er, at den mængde af information, som haves om observation y_i i regressionsmodellen, udelukkende består af baggrundsvariable fra X, der hører til y_i. Det svarer til, at hvis der foretages en ændring i baggrundsvariablene i matricen X, så x_{ir} ændres, vil det kun påvirke observation y_i og det i et omfang af β_r.

Indflydelsen fra en ændring i baggrundsmatricen X bliver selvsagt mere indviklet, når der medtages rumlige lags i modellen (4.1), hvorved mængden af information også vil omfatte naboobservationernes information. Hvis der ses på en rumligt afhængig model som SAR- eller CAR-modellen vil en observation $y_i \in \mathbf{y}$ ikke kun blive påvirket af en ændring i en variabel x_{ir}, men også af ændringer i baggrundsvariable for andre observationer som x_{jr} for $j \neq i$. For at præcisere hvordan enkelte observationer i henholdsvis SAR- og CAR-modellen påvirkes af en ændring i matricen X, findes de partielt afledede, der viser sig at være ens for de to modeller.

Sætning 4.1:
Både SAR- og CAR-modellen har partielt afledede givet ved

$$\frac{\partial y_i}{\partial x_{ir}} = S_r(W)_{ii} \quad \text{og} \quad \frac{\partial y_i}{\partial x_{jr}} = S_r(W)_{ij}, \tag{4.2}$$

hvor parameteren β_r er den r'te indgang i vektoren $\boldsymbol{\beta}$, og $S_r(W) = \left(I_n - \rho W\right)^{-1} \beta_r$. Derudover er det antaget, at parameteren $|\rho| < 1$.

Bevis:
Først findes de partielt aflede for SAR-modellen, der kan omskrives ved

$$\begin{aligned}\mathbf{y} &= \left(I_n - \rho W\right)^{-1}\left(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right)\\ &= \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta} + \left(I_n - \rho W\right)^{-1}\boldsymbol{\varepsilon}.\end{aligned}$$

Derudover gælder det, at

$$\begin{aligned}X\boldsymbol{\beta} = \sum_{r=1}^{k}\mathbf{x}_r\beta_r = \sum_{r=1}^{k}\beta_r\mathbf{x}_r \quad \Rightarrow\\ \mathbf{y} = \sum_{r=1}^{k}\left(S_r\left(W\right)\mathbf{x}_r\right) + V\left(W\right)\boldsymbol{\varepsilon}.\end{aligned} \tag{4.3}$$

Her er $\mathbf{x}_r$ den r'te søjle i matricen X med baggrundsvariable, mens k er antallet af søjler i X, og

$$\begin{aligned}S_r\left(W\right) &= V\left(W\right)\beta_r\\ V\left(W\right) &= \left(I_n - \rho W\right)^{-1} = I_n + \rho W + \rho^2 W^2 + \ldots.\end{aligned} \tag{4.4}$$

Den geometriske sumformel er anvendt til at omskrive funktionen $V\left(W\right)$ i (4.4), hvilket er muligt, idet det antages ρ opfylder $|\rho| < 1$. Ud fra omskrivningen af $V\left(W\right)$ ses det også, at påvirkninger fra naboer af alle grader medregnes i SAR-modellen, da matricen W indgår med stigende potesner. Den datagenererende proces for SAR-modellen kan nu skrives på matrixform som

$$\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_n\end{bmatrix} = \sum_{r=1}^{k}\begin{bmatrix}S_r\left(W\right)_{11} & S_r\left(W\right)_{12} & \cdots & S_r\left(W\right)_{1n}\\ S_r\left(W\right)_{21} & S_r\left(W\right)_{22} & \cdots & S_r\left(W\right)_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ S_r\left(W\right)_{n1} & S_r\left(W\right)_{n2} & \cdots & S_r\left(W\right)_{nn}\end{bmatrix}\begin{bmatrix}x_{1r}\\ x_{2r}\\ \vdots\\ x_{nr}\end{bmatrix} + V\left(W\right)\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{bmatrix}. \tag{4.5}$$

Udtages ligningen for en enkelt observation y_i fra ligningssystemet, fås

$$y_i = \sum_{r=1}^{k}\left(S_r\left(W\right)_{i1}x_{1r} + S_r\left(W\right)_{i2}x_{2r} + \ldots + S_r\left(W\right)_{in}x_{nr}\right) + V\left(W\right)_i\boldsymbol{\varepsilon}.$$

Den partielt afledede kan dermed udregnes som

$$\frac{\partial y_i}{\partial x_{ir}} = S_r\left(W\right)_{ii} \quad \text{og} \quad \frac{\partial y_i}{\partial x_{jr}} = S_r\left(W\right)_{ij}. \tag{4.6}$$

Nu ses der på de samme udregninger for CAR-modellen, hvor den datagenererende funktion ifølge sætning 3.18 er givet ved

$$\mathbf{y} = \left(I - \rho W\right)^{-1} X\boldsymbol{\beta} + \left(\left(I_n - \rho W\right)^{-1}\right)^{\frac{1}{2}}\boldsymbol{\varepsilon}.$$

Anvendes omskrivningen (4.3) på CAR-modellen, kan den datagenererende funktion skrives som

$$\begin{aligned}\mathbf{y} &= \left(I - \rho W\right)^{-1}\sum_{r=1}^{k}\left(\mathbf{x}_r\beta_r\right) + \left(\left(I_n - \rho W\right)^{-1}\right)^{\frac{1}{2}}\boldsymbol{\varepsilon}\\ &= \sum_{r=1}^{k}\left(S_r\left(W\right)\mathbf{x}_r\right) + \left(\left(I_n - \rho W\right)^{-1}\right)^{\frac{1}{2}}\boldsymbol{\varepsilon},\end{aligned}$$

hvor de partielt afledede giver samme resultat som (4.6). ■

Sætning 4.1 viser, at når der haves en rumlig model med indbyrdes afhængige observationer, bestemmes de partielt afledede også af naborelationerne i W og graden af afhængighed ρ, i stedet for kun at bestemmes af parameterværdien β_r for $i = j$. Den afledede for $i \neq j$ er ikke nødvendigvis nul, som den var i det simple, lineære tilfælde (4.1), hvilket afspejler, at en ændring i en forklarende variabel x_{jr} for observation y_j kan påvirke observation y_i.

Den partielt afledede $\frac{\partial y_i}{\partial x_{ir}} = S_r(W)_{ii}$ fra (4.2) beskriver hvordan observation y_i påvirkes af en ændring i en af observationens tilhørende baggrundsvariable fra matricen X, og her medregnes også de påvirkninger, der stammer fra, at ændringen i observation y_i kan påvirke observationen y_j, som igen vil påvirke y_i. Påvirkninger på en observation, der oprindeligt stammer fra en ændring i observationen selv, kaldes feedback og kan foregå via mere end to observationer. For eksempel kan en ændring i y_i påvirke observation y_j, som påvirker y_k, der så igen påvirker y_i.

De partielt afledede for både SAR- og CAR-modellen indeholder matricen $S_r(W)$, der som tidligere nævnt kan omskrives ved brug af den geometriske sumformel, hvilket giver

$$S_r(W) = \left(I_n - \rho W\right)^{-1} \beta_r = \left(I_n + \rho W + \rho^2 W^2 + \ldots\right) \beta_r \tag{4.7}$$

Betragtes matricen med andenordensnaboer W^2, ses det, at diagonalen ikke er nul, som den var i W. Det skyldes feedback, da observation y_i er andenordensnabo til sig selv, så en ændring i y_i vil påvirke naboobservationen y_j, der igen påvirker y_i. Det vil sige, at det er muligt at måle, hvor meget hver grad af naboer påvirker en observation, inklusive feedback, hvor naboer af lav orden som første- og andengradsnaboer påvirker en observation mere end højereordensnaboer på grund af potenserne af ρ, hvor $|\rho| < 1$. Påvirkninger af alle ordener er med i matricen $S_r(W)$ via summen af potenser af W i (4.7).

Hvor meget, en ændring i observation y_i påvirker observationen selv gennem naborelationer og feedback, afhænger af:

1. Hvordan observationerne er fordelt i forhold til hverandre i rummet.
2. Nabomatricen W, som afbilder naborelationerne mellem observationerne og ved højereordensnaboer også graden af naborelationerne, da eksempelvis W^4 som regel har forskellige brøker ved indgange ulig nul.
3. Graden af rumlig afhængighed som angives af parameteren ρ.
4. Parametervektoren $\boldsymbol{\beta}$.

Påvirkninger mellem observationer kan adskilles i direkte og indirekte påvirkninger, der tilsammen giver de samlede påvirkninger. De direkte påvirkninger mellem observationer kan aflæses på diagonalen i matricen $S_r(W)$, og er et udtryk for, hvor meget en ændring i en baggrundsvariabel x_{ir} påvirker den tilhørende observation y_i. Her medregnes også feedback fra påvirkninger gennem naboers naboer, det vil sige, når ændringen i y_i påvirker naboen y_j, der igen påvirker sine naboer inklusive y_i. Resten af elementerne i matricen $S_r(W)$ kaldes indirekte påvirkninger, og er påvirkningerne fra ændringen i y_i på sine naboer og derigennem naboernes naboer og så videre. De samlede påvirkninger fås ved at summere over enten rækker eller søjler i matricen $S_r(W)$.

Der er forskel på hvor meget en ændring i en enkelt baggrundsvariabel til en observation y_i påvirker hver af observationerne i **y**, da det er muligt observation y_i påvirkes mere end naboobservationen y_j, og at femtegradsnaboen y_k måske næsten ikke påvirkes. Derfor kan det være en fordel at se på den gennemsnitlige påvirkning på hver observation, når der ses på, hvor meget en ændring i en baggrundsvariabel i X påvirker **y**. Gennemsnittet for henholdsvis de direkte, de indirekte og de samlede påvirkninger vil blive defineret i det følgende, hvor det også vises, at den gennemsnitlige, samlede påvirkning kan udregnes på to måder, der dog giver samme numeriske værdi.

Definition 4.2 (Gennemsnitlig, direkte påvirkning):
Den gennemsnitlige, direkte påvirkning på en observation $y_i \in \mathbf{y}$ fra en ændring i baggrundsvariablen $x_{ir} \in X$ udregnes for alle $i = 1,\ldots,n$ som

$$\bar{M}(r)_{direkte} = \frac{tr\left(S_r(W)\right)}{n}.$$

Diagonalen i matricen $S_r(W)$ består som tidligere nævnt af de direkte påvirkninger, så de gennemsnitlige, direkte påvirkninger er givet ved middelværdien af diagonalen fra $S_r(W)$, hvilket for SAR- og CAR-modellen svarer til middelværdien af de egenafledede $\frac{\partial y_i}{\partial x_{ir}}$ for alle $i = 1,2,\ldots,n$ jævnfør sætning 4.1. Indsættes (4.4) i definition 4.2, kan udtrykket omskrives ved

$$\begin{aligned}
\bar{M}(r)_{direkte} &= \frac{tr\left(\left(I_n - \rho W\right)^{-1}\right)\beta_r}{n} \Rightarrow \\
\bar{M}_{direkte} &= \begin{bmatrix} \bar{M}(1)_{direkte} \\ \vdots \\ \bar{M}(k)_{direkte} \end{bmatrix} \\
&= \frac{tr\left(\left(I_n - \rho W\right)^{-1}\right)\boldsymbol{\beta}}{n} \\
&= \frac{\mathrm{tr}\left(I_n + \rho W + \rho^2 W^2 + \ldots + \rho^q W^q\right)\boldsymbol{\beta}}{n} \\
&= n^{-1}\sum_{j=0}^{q}\rho^j \,\mathrm{tr}\left(W^j\right)\boldsymbol{\beta} \quad \text{for} \quad q \to \infty.
\end{aligned}$$

Gennemsnittet af de samlede påvirkninger er givet i definition 4.3.

Definition 4.3 (Gennemsnitlig, samlede påvirkning):
Den gennemsnitlige, samlede påvirkning på en observation $y_i \in \mathbf{y}$ fra en ændring i en baggrundsvariabel $x_{ir} \in X$ er for alle $i = 1,\ldots,n$ givet ved

$$\bar{M}(r)_{total} = \frac{\iota_n^T S_r(W)\,\iota_n}{n}.$$

Indsættes (4.4) i udtrykket fra definition 4.3, fås resultatet

$$\bar{M}(r)_{total} = \frac{\iota_n^T \left(I_n - \rho W\right)^{-1} \beta_r \iota_n}{n}.$$

Den gennemsnitlige, samlede påvirkning kan udregnes på to forskellige måder, alt efter om der summeres over rækker eller søjler i matricen $S_r(W)$. Resultatet fra hver af de to regnemetoder kan beskrives som

1. Den gennemsnitlige, samlede påvirkning *på* en observation.
2. Den gennemsnitlige, samlede påvirkning *fra* en observation.

Hvis der tages gennemsnittet for rækkesummerne i matricen $S_r(W)$, fås den gennemsnitlige, samlede påvirkning *på* en observation. Påvirkningen siges at være *på* en observation, da den beskriver det tilfælde, hvor r'te søjle i matricen X ændres med samme værdi δ for alle n observationer, så summen af i'te række i matricen $S_r(W)$ angiver den samlede påvirkning på hver enkelt observation y_i, $i = 1,2,\ldots,n$. Mere præcist kan det siges, at hvis søjlevektoren $\mathbf{x}_r$ med baggrundsvariable, der også er vist i ligning (4.5), ændres til $\mathbf{x}_r + \delta\iota_n$, så vil hver observation y_i blive påvirket med rækkesummen

$$\delta \sum_{k=1}^{n} S_r(W)_{ik}.$$

Da der er n observationer i alt, bliver den gennemsnitlige, samlede påvirkning for hvert y_i til

$$\delta \frac{\iota_n^T S_r(W)\,\iota_n}{n} = \delta n^{-1} \iota_n^T \mathbf{c}_r,$$

hvor ι_n er en søjlevektor bestående af n ettaller, og $\mathbf{c}_r := S_r(W)\,\iota_n$ er en søjlevektor bestående af rækkesummer fra matricen $S_r(W)$. Det vil sige, at den gennemsnitlige, samlede påvirkning *på* en observation er den gennemsnitlige ændring for hvert y_i, når en søjle i matricen X ændres med δ, og ændringen udregnes som et gennemsnit af rækkesummer.

En anden måde at udregne den gennemsnitlige, samlede påvirkning på er ved at benytte søjlesummer fra matricen $S_r(W)$, hvilket vil resultere i den gennemsnitlige, samlede påvirkning *fra* en observation. At påvirkningen kaldes *fra* en observation skyldes, at der her ændres på en enkelt baggrundsvariabel x_{jr} tilhørende observation y_j, og derefter tages gennemsnittet af påvirkningerne på alle observationer y_i, der har uændrede baggrundsvariable. Mere præcist kan det siges, at baggrundsvariablen x_{jr} tilhørende observation y_j ændres til $x_{jr} + \delta$, hvilket vil ændre hele den j'te søjle i matricen $S_r(W)$. Dermed bliver den samlede påvirkning på alle observationer y_i summen over j'te søjle i $S_r(W)$, hvilket er

$$\delta \sum_{k=1}^{n} S_r(W)_{kj}.$$

Da der er n observationer i alt, bliver den gennemsnitlige påvirkning på hver observation

$$\delta \frac{\iota_n^T S_r(W)\,\iota_n}{n} = \delta n^{-1} \mathbf{r}_r \iota_n,$$

hvor vektoren $\mathbf{r}_r := \iota_n^T S_r(W)$ er en rækkevektor over søjlesummerne fra $S_r(W)$. Det svarer til, at den gennemsnitlige, samlede påvirkning *fra* en observation er gennemsnittet af påvirkningerne på hvert y_i, $i = 1,2,\ldots,n$, der stammer fra en ændring i baggrundsvariablen x_{jr} for observation y_j.

Den gennemsnitlige, samlede påvirkning har numerisk samme værdi om den udregnes som et gennemsnit af påvirkninger på eller fra en observation, hvilket er det samme som, at den er middelværdien af $\frac{\partial y_i}{\partial x_{jr}}$ for alle $j,i = 1,2,\ldots,n$. Den gennemsnitlige, indirekte påvirkning er nem at udregne, når den direkte og den samlede påvirkning er kendt, og den er defineret i definition 4.4.

Definition 4.4 (Gennemsnitlig, indirekte påvirkning):
Den gennemsnitlige, indirekte påvirkning på observation $y_i \in \mathbf{y}$ der stammer fra en ændring i baggrundsvariablen $x_{ir} \in X$ er for alle $i = 1,\ldots,n$ givet ved

$$\bar{M}(r)_{indirekte} = \bar{M}(r)_{total} - \bar{M}(r)_{direkte}.$$

I praksis er det ikke altid muligt at beregne en af de tre gennemsnitlige påvirkninger $\bar{M}$ ved brug af definitionerne (4.2), (4.3) og (4.4), idet alle tre definitioner kræver den inverse af $(n \times n)$-matricen $I_n - \rho W$ udregnes, hvilket kan være svært for en stor værdi af n. Det er dog muligt at anvende et tilnærmet udtryk for den inverse matrix ved brug af den geometriske sumformel i form af

$$\left(I_n - \rho W\right)^{-1} \approx I_n + \rho W + \rho^2 W^2 + \ldots + \rho^q W^q,$$

der konvergerer tilnærmelsesvist for en tilstrækkelig høj værdi for q. Det vil sige, at matricen $S_r(W)$, der indeholder værdier for påvirkninger mellem observationerne $\mathbf{y}$, kan tilnærmes med en linearkombination af potenser af W ved

$$S_r(W) \approx \left(I_n + \rho W + \rho^2 W^2 + \ldots + \rho^q W^q\right)\beta_r.$$

Der gælder desuden, at en ændring i en baggrundsvariabel til en observation y_i påvirker naboer af højere orden mindre end naboer af lav orden.

Ses der på SAR- og CAR-modellen, kan begge modellers gennemsnitlige, samlede påvirkninger skrives som i sætning (4.5), da det vides fra sætning 4.1, at de har samme partielt afledede.

Sætning 4.5:
Den gennemsnitlige, samlede påvirkning for både SAR- og CAR-modellen er givet ved

$$\bar{M}(r)_{total} = \left(1 - \rho\right)^{-1} \beta_r,$$

for $|\rho| < 1$. Her betegner β_r den r'te ingang i parametervektoren $\boldsymbol{\beta}$.

Bevis:
Det gælder for både SAR- og CAR-modellen, at $S_r(W) = \left(I_n - \rho W\right)^{-1}\beta_r$, og indsættes dette i definition 4.3, fås

$$\bar{M}(r)_{total} = n^{-1}\iota_n^T S_r(W)\,\iota_n = n^{-1}\iota_n^T\left(I_n - \rho W\right)^{-1}\beta_r\iota_n.$$

Benyttes den geometriske sumformel giver ovenstående

$$\begin{aligned} n^{-1}\iota_n^T\left(I_n - \rho W\right)^{-1}\beta_r\iota_n &= n^{-1}\iota_n^T\left(\sum_{i=0}^{\infty}\rho^j W^j\right)\beta_r\iota_n \\ &= n^{-1}\left(\sum_{i=0}^{\infty}\rho^j\iota_n^T W^j\iota_n\right)\beta_r \\ &= n^{-1}\left(\sum_{i=0}^{\infty}\rho^j n\right)\beta_r, \end{aligned}$$

da $|\rho < 1|$ og nabomatricen W er rækkestokastisk. Den geometriske sumformel kan anvendes en gang til, hvilket giver

$$n^{-1}\left(\sum_{i=0}^{\infty}\rho^j n\right)\beta_r = \left(1-\rho\right)^{-1}\beta_r.$$

■

Det ses ud fra sætning 4.5, at når $\rho < 1$ og positiv, vil faktoren $\left(1-\rho\right)^{-1} > 1$, så $\bar{M}(r)_{total}$ er større end β_r. Det vil sige, at den gennemsnitlige, samlede påvirkning på en observation $y_i \in \mathbf{y}$ fra en ændring i en baggrundsvariabel i X er større end den indgang β_r fra parametervektoren $\boldsymbol{\beta}$, som hører til den ændrede baggrundsvariabel. Præcist hvor meget større, påvirkningen $\bar{M}(r)_{total}$ er i forhold til β_r, afhænger af parameteren for rumlig afhængighed givet ved ρ, hvor en værdi for ρ tæt på én giver en stor samlet påvirkning $\bar{M}(r)_{total}$, mens en ρ-værdi tæt på nul ikke giver nogen stor ændring i forhold til parameteren β_r. Dette resultat stemmer godt overens med, at stor rumlig afhængighed ρ mellem observationerne i et datasæt medfører, at en ændring i en observation vil have stor indflydelse på naboobservationerne.

Det ses også ud fra sætning 4.5, at når ρ er positiv, vil $\bar{M}(r)_{total}$ have samme fortegn som parameteren β_r. Parameterværdi β_r er tilknyttet en bestemt søjle $\mathbf{x}_r$ i matricen X over baggrundsvariable, og afgør dermed om disse variable skal have positiv eller negativ indflydelse på observationerne $\mathbf{y}$. Ændres der på en eller flere baggrundsvariable i den r'te søjle, afgør fortegnet for β_r og størrelsen af ρ, hvor meget den nye værdi for y_i stiger eller falder i forhold til den oprindelige, hvilket sker gennem påvirkningen $\bar{M}(r)_{total}$.

Maximum Likelihood-estimation 5

Medmindre andet er angivet, er teksten i dette kapitel skrevet med udgangspunkt i [Azzalini, 1996, s. 22-29], [LeSage and Pace, 2009, kapitel 3] og [Olofsson, 2005, afsnit 6.4.2].

Ofte anvendes mindste kvadraters metoder til estimering af parametrene i en model, men når der haves et rumligt datasæt med observationer, der indbyrdes afhænger af hverandre, kan det give mindre afvigelser mellem data og model at anvende maximum likelihood-estimation i stedet. Derfor vil der i dette kapitel blive set på maximum likelihood-estimation.

5.1 Maximum Likelihood-estimation

Maximum likelihood-estimation anvendes til at finde en model for et givet datasæt, der for eksempel kan være en endelig mængde af huspriser med tilhørende koordinater. Metoden udvælger de parametre, der maksimerer likelihoodfunktionen, som er defineret i definition 5.1, idet en model med disse parametre har det givne datasæt som det mest sandsynlige resultat. Generelt er maximum likelihood-estimation en veldefineret metode til at estimere parametre, når der haves en normalfordeling, hvilket er tilfældet med både SAR- og CAR-modellen.

Definition 5.1 (Likelihoodfunktionen):
For en datavektor $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}^T$ med n observationer er likelihoodfunktionen $l(\boldsymbol{\theta})$ er givet ved

$$l = l(\boldsymbol{\theta}|\mathbf{y}) = f_{\boldsymbol{\theta}}(\mathbf{y}),$$

hvor $\boldsymbol{\theta} = \theta_1, \theta_2, \dots$ er modelparametre, og funktionen $f_{\boldsymbol{\theta}}(\mathbf{y})$ er sandsynlighedsfunktion eller tæthedsfunktion $f_{\boldsymbol{\theta}}$ til fordelingen for $\mathbf{y}$. Det maksimale likelihood-estimat (MLE) er den vektor $\hat{\boldsymbol{\theta}}$, som indsat i likelihoodfunktionen giver den maksimale værdi af $l(\boldsymbol{\theta}|\mathbf{y})$.

Definition 5.1 viser, at parametrene er givet i tætheden $f_{\boldsymbol{\theta}}(\mathbf{y})$, mens data ikke kendes, og ved likelihoodfunktionen $l(\boldsymbol{\theta}|\mathbf{y})$ er data de givne værdier, mens modelparametrene ikke kendes. Når der haves et datasæt, er det dermed muligt at finde modelparametrene ud fra likelihoodfunktionen og dernæst simulere nye data ud fra tætheden $f_{\boldsymbol{\theta}}(\mathbf{y})$. I tilfældet med SAR-modellen er parametervektoren $\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\beta}^T & \rho & \sigma^2 \end{bmatrix}^T$, hvor parameteren α også vil indgå, hvis vektoren $\alpha\iota_n$ ikke er en del af matricen X.

I praksis ses der ofte på loglikelihoodfunktionen i stedet for likelihoodfunktionen, da likelihoodfunktionen ifølge [Olofsson, 2005, s. 329] som regel består af produkter, der bliver til bliver til summer, når der tages logaritmen. Dermed kan funktionen være nemmere at håndtere, da maksimum ofte findes ved at differentiere funktionen og derefter sætte den lig med nul. Det er muligt, at se på loglikelihoodfunktionens maksimum i stedet for likelihoodfunktionens maksimum, da ln er en kontinuert og stigende funktion.

Definition 5.2 (Loglikelihoodfunktionen):
Loglikelihoodfunktionen defineres ud fra likelihoodfunktionen $l(\boldsymbol{\theta})$ ved

$$L = L(\boldsymbol{\theta}|\mathbf{y}) = \ln l(\boldsymbol{\theta}|\mathbf{y}) = \ln\left(f_{\boldsymbol{\theta}}(\mathbf{y})\right),$$

hvor $\boldsymbol{\theta}$ er en parametervektor.

Den maksimale værdi for loglikelihoodfunktionen kan findes ved at differentiere funktionen partielt med hensyn til hver af parametrene i modellen, hvorefter hver ligning sættes lig nul og den ukendte parameter isoleres. Dermed fås den maksimale værdi for hver parameter i modellen. For overskuelighedens skyld ses der som regel på en enkelt parameter ad gangen i stedet for at estimere alle de ukendte parametre simultant, hvilket svarer til at differentiere profilloglikelihoodfunktionen for hver parameter. Profilloglikelihoodfunktionen, der er defineret i definition 5.3, er loglikelihoodfunktionen hvor alle parametre på nær én anses som konstanter.

Definition 5.3 (Profilloglikelihoodfunktionen):
Profilloglikelihoodfunktionen med hensyn til parameteren θ_j er givet ved

$$L_p(\theta_j) = \ln\left(f_{\theta_j}(y_i)\right),$$

hvor $\theta_j \in \boldsymbol{\theta}$, og $\boldsymbol{\theta}$ er en vektor over modelparametre, hvor alle parametre på nær θ_j antages at være konstanter i loglikelihoodfunktionen $\ln\left(f_{\boldsymbol{\theta}}(\mathbf{y})\right)$.

Hvis der for eksempel ses på SAR-modellen, kan profilloglikelihoodfunktionen tages med hensyn til parameteren ρ, hvilket skrives $L_p(\rho)$, hvor de resterende parametre σ^2 og $\boldsymbol{\beta}$ antages at være konstanter. Løses ligningen $\frac{\partial}{\partial\rho}L_p(\rho) = 0$ fås den maksimale værdi for ρ.

Hvis der haves en model med mere end én ukendt parameter, foretages maximum likelihood-estimation ved at gennemgå følgende punkter for hver af de ukendte parametre:

1. Find likelihoodfunktionen $l(\boldsymbol{\theta}|\mathbf{y})$ og tag logaritmen for at få loglikelihoodfunktionen $L(\boldsymbol{\theta}|\mathbf{y}) = \log l(\boldsymbol{\theta}|\mathbf{y})$.
2. Find profilloglikelihoodfunktionen $L_p(\theta_i)$ med hensyn til parameteren $\theta_i \in \boldsymbol{\theta}$.
3. Differentier funktionen $L_p(\theta_i)$ ved $\frac{\partial}{\partial\theta_i}L_p(\theta_i)$.

4. Løs ligningen $\frac{\partial}{\partial\theta_i}L_p(\theta_i)=0$ for at finde den maksimale værdi $\theta_i=\hat{\theta}_i$.
5. Undersøg om den andenordensafledede $\frac{\partial}{\partial\theta_i}\left(\frac{\partial}{\partial\theta_i}L_p(\theta_i)\right)<0$, så det er sikkert, der er tale om et maksimum.

5.2 Parameterestimering for SAR-modellen

For at kunne anvende maximum likelihood-estimation på SAR-modellen er det nødvendigt at kende likelihoodfunktionen og også loglikelihoodfunktionen, der begge har form som vist i sætning 5.4.

Sætning 5.4:
Likelihoodfunktionen for SAR-modellen er givet ved

$$l=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\cdot\left|I_n-\rho W\right|\cdot\exp\left\{-\frac{\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)}{2\sigma^2}\right\},$$

mens loglikelihoodfunktionen for SAR-modellen er givet ved

$$L=-\frac{n}{2}\ln\left(2\pi\sigma^2\right)+\ln\left|I_n-\rho W\right|-\frac{e^Te}{2\sigma^2}, \tag{5.1}$$

hvor $e=\mathbf{y}-\rho W\mathbf{y}-X\boldsymbol{\beta}$ og $\rho\in\left(\lambda_{\min}^{-1},1\right)$.

Bevis:
Jævnfør sætning 2.7 følger SAR-modellen den multivariate normalfordeling

$$N_n\left(\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta},\left(I_n-\rho W\right)^{-1}\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right).$$

Anvendes appendiks B.1 ses det, at SAR-modellen har tætheden

$$p(\mathbf{y})=\left(\sqrt{(2\pi)^n\left|\left(I_n-\rho W\right)^{-1}\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right|}\right)^{-1} \tag{5.2}$$
$$\cdot\exp\left\{-\frac{1}{2}\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)^{-1}\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right)^{-1}\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)\right\}.$$

For en konstant r og en $(n\times n)$-matrix A gælder det generelt, at $\det(rA)=r^n\det A$ og $\det A=\det A^T$, så den første faktor i ligning (5.2) kan omskrives ved

$$\begin{aligned}\left(\sqrt{(2\pi)^n\left|\left(I_n-\rho W\right)^{-1}\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right|}\right)^{-1}&=\left((2\pi)^n\left|\left(I_n-\rho W\right)^{-1}\right|\left|\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right|\right)^{-\frac{1}{2}}\\&=(2\pi)^{-\frac{n}{2}}\left|I_n-\rho W\right|^{\frac{1}{2}}\left(\left(\sigma^2\right)^n\left|\left(\left(I_n-\rho W\right)^{-1}\right)^T\right|\right)^{-\frac{1}{2}}\\&=(2\pi)^{-\frac{n}{2}}\left|I_n-\rho W\right|^{\frac{1}{2}}\sigma^{-n}\left|I_n-\rho W\right|^{\frac{1}{2}}\\&=(2\pi)^{-\frac{n}{2}}\left|I_n-\rho W\right|\sigma^{-n}\\&=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\left|I_n-\rho W\right|.\end{aligned}$$

Det er også muligt at omskrive udtrykket i eksponenten i ligning (5.2), hvorved det fås, at

$$\begin{aligned}
&-\frac{1}{2}\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)^{-1}\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right)^{-1}\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)\\
&\quad=-\frac{1}{2}\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)^T\left(I_n-\rho W\right)^T\left(\sigma^2\right)^{-1}\left(I_n-\rho W\right)\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)\\
&\quad=-\frac{1}{2}\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\sigma^{-2}\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)\\
&\quad=-\frac{\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)}{2\sigma^2}.
\end{aligned}$$

SAR-modellens likelihoodfunktion har dermed formen

$$l=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\cdot\left|I_n-\rho W\right|\cdot\exp\left\{-\frac{\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)}{2\sigma^2}\right\}.$$

Tages logaritmen til ovenstående, fås SAR-modellens loglikelihoodfunktion $L=\ln l$ som

$$L=-\frac{n}{2}\ln\left(2\pi\sigma^2\right)+\ln\left|I_n-\rho W\right|-\frac{e^Te}{2\sigma^2}$$

hvor

$$\begin{aligned}
e&=\mathbf{y}-\rho W\mathbf{y}-X\boldsymbol{\beta}\\
\rho&\in\left(\lambda_{\min}^{-1},1\right),
\end{aligned}$$

og $\lambda_{\min}^{-1}$ betegner vægtmatricen W's mindste egenværdi. ■

Det er et krav at kovariansmatricen er symmetrisk og positiv definit, for at tætheden og dermed også likelihoodfunktionen for en normalfordeling er veldefineret, og ifølge sætning 2.13 er dette tilfældet for SAR-modellen, når $\rho\in\left(\lambda_{\min}^{-1},1\right)$. Som beskrevet tidligere kan det være vanskeligt at tolke negativ korrelation mellem observationerne, så i praksis vælges det ofte, at $\rho\in[0,1)$.

Estimaterne for parametrene kan findes ved at maksimere loglikelihoodfunktionen for hver parameter, hvilket gøres ved at differentiere loglikelihoodfunktionen for hver parameter, sætte differentialerne lig nul og løse ligningerne simultant. En alternativ metode er at anvende profilloglikelihoodfunktionen $L_p(\theta_i)$, der er loglikelihoodfunktionen L, hvor alle parametrene $\boldsymbol{\theta}\backslash\{\theta_i\}$ antages at være konstante værdier. Estimatet for parameteren θ_i er den værdi, der maksimerer $L_p(\theta_i)$. I SAR-modellens tilfælde er det muligt at skrive parametrene σ^2 og $\boldsymbol{\beta}$ som funktioner af parameteren ρ, hvorved det kun er nødvendigt at maksimere profilloglikelihoodfunktionen $L_p\left(\rho\right)$ for at finde alle parametrene. Parameterestimaterne er ens, hvad enten estimaterne findes ved brug af maximum likelihood-estimation for hver parameter gennem loglikelihoodfunktionen eller maksimering af profilloglikelihoodfunktionen $L_p\left(\rho\right)$.

Mindste kvadraters metode kan anvendes til at finde ligninger, der giver henholdsvis σ^2 og $\boldsymbol{\beta}$ som funktion af ρ. Maximum likelihood-estimation anvendes dermed til at finde et estimat for ρ givet ved $\hat{\rho}$, og derefter anvendes mindste kvadraters metode til at finde de resterende estimater for parametrene σ^2 og $\boldsymbol{\beta}$ som funktioner af ρ. Estimaterne der

findes gennem mindste kvadraters metode er vist i sætning 5.6, men inden sætningen bevises er det nødvendigt med et lemma.

Lemma 5.5:
Lad $(n \times k)$-matricen X have fuld søjlerang. Da eksisterer matricen $\left(X^T X\right)^{-1}$, mens $(k \times k)$-matricen $X^T X$ er positiv definit.

Bevis:
Da det er antaget, at matricen X har fuld søjlerang, er søjlerne i X indbyrdes uafhængige. Dermed gælder det for alle vektorer $\mathbf{v} \in \mathbb{R}^k$, hvor $\mathbf{v} \neq \mathbf{0}$, at

$$\begin{aligned}\mathbf{v}^T \left(X^T X\right)\mathbf{v} &= \left(\mathbf{v}^T X^T\right)(X\mathbf{v}) \\ &= (X\mathbf{v})^T (X\mathbf{v}) \\ &= \|X\mathbf{v}\|^2 > 0.\end{aligned}$$

Jævnfør [Axler, 1997, s. 144] er det dermed bevist, at matricen $X^T X$ er positiv definit, og derfor eksisterer den inverse matrix $\left(X^T X\right)^{-1}$ også. ■

Det er nu muligt, at finde udtryk for estimaterne for parametrene σ^2 og $\boldsymbol{\beta}$ samt kovariansmatricen Σ i SAR-modellen som funktioner af parameteren ρ, hvor estimaterne findes ved brug af mindste kvadraters metode.

Sætning 5.6:
Mindste kvadraters metode giver estimaterne

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(X^T X\right)^{-1} X^T \left(I_n - \rho^* W\right)\mathbf{y} \\ \hat{\sigma}^2 &= \frac{\left\|e\left(\rho^*\right)\right\|^2}{n} = n^{-1} e\left(\rho^*\right)^T e\left(\rho^*\right) \\ \hat{\Sigma} &= \left(I_n - \rho^* W\right)^{-1} \hat{\sigma}^2 \left(\left(I_n - \rho^* W\right)^{-1}\right)^T\end{aligned}$$

i SAR-modellen, hvor $\boldsymbol{\beta}$ er parametervektoren, σ^2 er variansen og Σ er kovariansmatricen, mens ρ^* angiver den ægte værdi for parameteren ρ. Derudover er

$$e\left(\rho^*\right) = \mathbf{y} - \rho^* W\mathbf{y} - X\hat{\boldsymbol{\beta}}.$$

Bevis:
Den ægte værdi for ρ betegnes ρ^* og det antages, at denne værdi er kendt, hvormed SAR-modellen kan skrives som

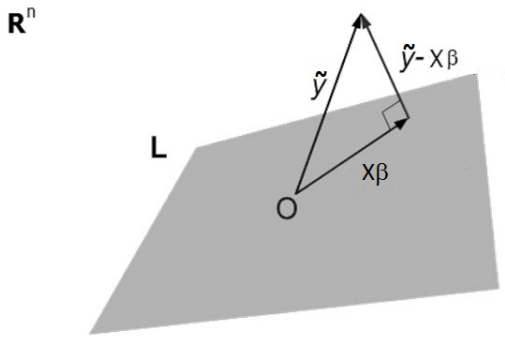
$$\begin{aligned}\mathbf{y} &= \rho^* W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \Rightarrow \\ \left(I_n - \rho^* W\right)\mathbf{y} &= X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N_n\left(\mathbf{0}, \sigma^2 I_n\right).\end{aligned}$$

For at gøre udregningerne simplere defineres det, at $\tilde{\mathbf{y}} := \left(I_n - \rho^* W\right)\mathbf{y}$, hvordved SAR-modellen kan skrives som

$$\tilde{\mathbf{y}} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{5.3}$$

Den nye vektor $\tilde{\mathbf{y}}$ har middelværdivektor $X\boldsymbol{\beta}$, der tilhører et underrum L udspændt af søjlerne i $(n \times k)$-matricen X, hvor underrummet $L \subset \mathbb{R}^n$ har k dimensioner. Derudover er parametervektoren $\boldsymbol{\beta} \in \mathbb{R}^k$. Det ønskes dermed at finde parametrene i en lineær model for observationerne $\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{y}_1 & \tilde{y}_2 & \dots & \tilde{y}_n \end{bmatrix}^T$, så fejlen bliver mindst mulig. Mindste kvadraters metode bestemmer værdien for parameteren $\boldsymbol{\beta}$ så forskellen mellem observationerne og middelværdien for estimaterne $\hat{\boldsymbol{\mu}}$ bliver mindst mulig, hvilket svarer til $\left\|\tilde{\mathbf{y}} - \hat{\boldsymbol{\mu}}\right\|^2$ er så tæt på nul som muligt.

Middelværdien $X\boldsymbol{\beta}$ findes ved at projicere vektoren $\tilde{\mathbf{y}}$ ned på underrummet L, hvorved der også fås vektoren $\left(\tilde{\mathbf{y}} - X\boldsymbol{\beta}\right) \in L^{\perp}$, som vist på figur 5.1.



Figur 5.1: Vektoren $\tilde{\mathbf{y}}$ projiceres ned på underrummet L.

Da SAR-modellen følger en normalfordeling og $\tilde{\mathbf{y}} = X\boldsymbol{\beta} + \varepsilon$, er $\tilde{y} \sim N(X\boldsymbol{\beta}, \sigma^2 I_n)$, mens vektoren $\tilde{\mathbf{y}} - X\boldsymbol{\beta} = \varepsilon \sim N(\mathbf{0}, \sigma^2 I_n)$, hvor $X\boldsymbol{\beta}$ er en konstant.

Underrummet L er defineret til at være udspændt af søjlerne fra X, hvilket er ensbetydende med, at alle vektorer i L, kan skrives som $X\mathbf{v}$ hvor vektoren $\mathbf{v} \in \mathbb{R}^k$. Da vektoren $\left(\tilde{\mathbf{y}} - X\boldsymbol{\beta}\right) \in L^{\perp}$, gælder det for alle $\mathbf{v} \in \mathbb{R}^k$, at

$$\begin{aligned} 0 &= (X\mathbf{v})^T \left(\tilde{\mathbf{y}} - X\boldsymbol{\beta}\right) \\ &= \mathbf{v}^T X^T \left(\tilde{\mathbf{y}} - X\boldsymbol{\beta}\right) \\ &= \mathbf{v}^T \left(X^T \tilde{\mathbf{y}} - X^T X\boldsymbol{\beta}\right). \end{aligned}$$

Det antages, at $\mathbf{v} \neq 0$ for at undgå den trivielle løsning, og da det i lemma 5.5 blev vist, at den inverse af $X^T X$ eksisterer, må det gælde at

$$\begin{aligned} 0 = X^T \tilde{\mathbf{y}} - X^T X\boldsymbol{\beta} \Rightarrow \\ X^T X\boldsymbol{\beta} = X^T \tilde{\mathbf{y}} \Rightarrow \\ \boldsymbol{\beta} = \left(X^T X\right)^{-1} X^T \tilde{\mathbf{y}}, \end{aligned}$$

Nu kan $\tilde{\mathbf{y}} := \left(I_n - \rho^* W\right)\mathbf{y}$ indsættes i ovenstående for at få estimatet for parametervektoren $\boldsymbol{\beta}$ som en funktion af ρ^*, hvilket giver

$$\hat{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T \left(I_n - \rho^* W\right)\mathbf{y}.$$

Afvigelsen mellem data og de estimerede værdier er givet ved en fejlvektor

$$e\left(\rho^*\right) = \mathbf{y} - \rho^* W\mathbf{y} - X\hat{\boldsymbol{\beta}},$$

der er fundet ved at trække den estimerede model fra de givne data $\mathbf{y}$. Variansen for fejlene fås ved at dividere ovenstående fejlvektor med n, der er antallet af observationer, hvorved estimatet for σ^2 bliver

$$\hat{\sigma}^2 = \frac{\left\| e\left(\rho^*\right)\right\|^2}{n} = n^{-1} e\left(\rho^*\right)^T e\left(\rho^*\right).$$

Kovariansmatricens estimat fås ved at indsætte $\hat{\sigma}^2$ og ρ^* i ligningen fra sætning 2.13, hvilket giver

$$\hat{\Sigma} = \left(I_n - \rho^* W\right)^{-1} \hat{\sigma}^2 \left(\left(I_n - \rho^* W\right)^{-1}\right)^T.$$

■

SAR-modellen kan også skrives med parameteren $\alpha\iota_n$ som seperat parameter, hvilket er vist i ligning (2.4). I dette tilfælde defineres en matrix $Z := \begin{bmatrix}\iota_n & X\end{bmatrix}$ samt en parametervektor $\boldsymbol{\delta} := \begin{bmatrix}\alpha & \boldsymbol{\beta}^T\end{bmatrix}^T$, hvorefter parametervektoren $\boldsymbol{\delta}$ kan estimeres ved

$$\hat{\boldsymbol{\delta}} = \left(Z^T Z\right)^{-1} Z^T \left(I_n - \rho^* W\right)\mathbf{y}.$$

Når loglikelihoodfunktionen kendes for SAR-modellen, og parametrene $\boldsymbol{\beta}$ og σ^2 kan skrives som funktioner af parameteren ρ, er det kun nødvendigt at finde profilloglikelihoodfunktionen $L_p\left(\rho\right)$ med hensyn til ρ for at finde alle parameterestimaterne i SAR-modellen.

Profilloglikelihoodfunktionen med hensyn til ρ findes ved at omskrive ligning (5.1), og indsætte ligningerne for parametrene σ^2 og $\boldsymbol{\beta}$ fra sætning 5.6. Dermed haves en funktion med ρ som eneste variabel, der er

$$L_p\left(\rho\right) = \kappa + \ln\left|I_n - \rho W\right| - \frac{n}{2} S\left(\rho\right),$$

hvor

$$\begin{aligned} S\left(\rho\right) &= e(\rho)^T e(\rho) \\ e(\rho) &= \mathbf{y} - \rho W\mathbf{y} - X\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\beta}} &= \left(X^T X\right)^{-1} X^T \left(I_n - \rho W\right)\mathbf{y} \\ \kappa &= -\frac{n}{2}\ln\left(2\pi\hat{\sigma}^2\right). \end{aligned}$$

Konstanten κ er uafhængig af ρ, mens ligningerne for $\hat{\sigma}^2$ og $\hat{\boldsymbol{\beta}}$ fra sætning 5.6 er indsat, da de er funktioner af ρ. Det er muligt at forsimple maksimeringen af profilloglikelihoodfunktionen med hensyn til parameteren ρ ved først at vælge en $\left(q \times 1\right)$-vektor $\boldsymbol{\rho} = \begin{bmatrix}\rho_1 & \dots & \rho_q\end{bmatrix}^T$ af mulige værdier for parameteren ρ. Vektorens værdier tages

fra intervallet $[\rho_{\min},\rho_{\max}]$, der i praksis ofte antages at være $[0,1)$, så hele intervallet er dækket. Dermed fås ligningssystemet

$$\begin{bmatrix} L_p(\rho_1) \\ L_p(\rho_2) \\ \vdots \\ L_p(\rho_q) \end{bmatrix} = \kappa\iota_n + \begin{bmatrix} \ln|I_n - \rho_1 W| \\ \ln|I_n - \rho_2 W| \\ \vdots \\ \ln|I_n - \rho_q W| \end{bmatrix} - \begin{bmatrix} S(\rho_1) \\ S(\rho_2) \\ \vdots \\ S(\rho_q) \end{bmatrix}.$$

Den værdi fra vektoren $\boldsymbol{\rho} = \begin{bmatrix} \rho_1 & \dots & \rho_q \end{bmatrix}^T$, der giver den maksimale værdi i ovenstående, vælges som estimatet $\hat{\rho}$. Herefter er det muligt at udregne estimaterne for resten af parametrene i SAR-modellen ved at anvende formlerne fra sætning 5.6. Leddet $|I_n - \rho W|$ i loglikelihoodfunktionen sikrer, at estimatet for parameteren ρ ikke kun er til for at minimere $S(\rho)$, der består af kvadrerede fejlled. Når vektoren $\boldsymbol{\rho} = \begin{bmatrix} \rho_1 & \dots & \rho_q \end{bmatrix}^T$ anvendes, hvor værdierne i vektoren repræsenterer hele intervallet, sikrer maximum likelihood-estimation et globalt maksimum og ikke kun et lokalt maksimum.

I stedet for at finde maximum likelihood-estimatet ved brug af vektoren $\boldsymbol{\rho} = \begin{bmatrix} \rho_1 & \dots & \rho_q \end{bmatrix}^T$ er det også muligt at anvende Newtons metode til optimering. Dette er en iterativ metode til at finde værdien for ρ, hvor $\frac{\partial}{\partial\rho} L_p(\rho) = 0$, hvor der startes med en valgt værdi $\rho_0 \in [0,1)$ og derefter udregnes

$$\rho_{m+1} = \rho_m - \frac{L_p'(\rho_m)}{L_p''(\rho_m)}$$

hvor $m = 0,1,\dots$, og når $m \to \infty$, vil $\rho_m \to \rho^*$ så $\frac{\partial}{\partial\rho} L_p(\rho^*) = 0$. Det kan være praktisk at anvende Newtons metode, da den giver et estimat for ρ samtidig med den andenordensafledede findes, så det kan sikres, at den fundne værdi giver maksimum.

Maximum likelihood er dermed blevet anvendt til at finde et estimat for parameteren ρ i SAR-modellen, og dette estimat er blevet anvendt til at finde resten af parametrene i modellen ved brug af mindste kvadraters metode. I afsnit 5.3 vil der blive set på, hvordan samme metoder kan anvendes til at finde parametrene i CAR-modellen, men først ses der på en praktisk approksimationsmetode.

Monte Carlo-approksimation af logdeterminanten

Afsnittet tager udgangspunkt i [LeSage and Pace, 2009, kapitel 4] samt nogle velkendte regneregler.

Når der haves et stort antal observationer n i et datasæt, kan det være besværligt at udregne logaritmen til determinanten af $I_n - \rho W$, der indgår i loglikelihoodfunktionen for SAR-modellen, så i nogle tilfælde kan det være en fordel at benytte Monte Carlo-approksimation. Monte Carlo-approksimation kræver ikke særlig meget computerhukommelse at udføre, så den tager ikke lang tid at beregne, og den resulterende fejl i estimatet for ρ er relativt lille. Logaritmen til determinanten af matricen $I_n - \rho W$ er vist i sætning 5.9, der anvender følgende to lemmaer i beviset.

Lemma 5.7:
For en $(n \times n)$-matrix A er determinanten for eksponentialmatricen e^A givet ved

$$\left|e^A\right| = e^{\mathrm{tr}(A)}.$$

Bevis:
Matricen A kan opskrives på jordan kanonisk form ved brug af en invertibel matrix V, så

$$J = V^{-1}AV \Rightarrow A = VJV^{-1},$$

hvor matricen J har egenværdierne for A på diagonalen og en superdiagonal, hvor indgangene er enten nul eller en, mens resten af indgangene i matricen J har værdien nul. Determinanten for en øvre triangulær matrix er ligesom determinanten for en diagonalmatrix givet ved produktet af diagonalelementerne, så det er tilstrækkeligt at se på det tilfælde, hvor matricen A er diagonaliserbar, hvilket giver

$$A = VJV^{-1} = V\Lambda_A V^{-1}.$$

Her er Λ_A en diagonalmatrix med egenværdierne for A på diagonalen, mens V består af de ortonormale egenvektorer til A. Matricen e^A er defineret som en taylorudvikling, hvilket giver

$$\begin{aligned}
e^A &= \sum_{k=0}^{\infty} \frac{A^k}{k!} \\
&= \sum_{k=0}^{\infty} \frac{\left(V\Lambda_A V^{-1}\right)^k}{k!} \\
&= \sum_{k=0}^{\infty} \frac{V\Lambda_A^k V^{-1}}{k!} \\
&= V\left(\sum_{k=0}^{\infty} \frac{\Lambda_A^k}{k!}\right)V^{-1} \\
&= Ve^{\Lambda_A}V^{-1} \Rightarrow \\
\left|e^A\right| &= \left|Ve^{\Lambda_A}V^{-1}\right| = |V|\left|e^{\Lambda_A}\right|\left|V^{-1}\right| = \left|e^{\Lambda_A}\right|. \qquad (5.4)
\end{aligned}$$

Determinanten for en diagonalmatrix er produktet af diagonalelementerne, der for Λ_A er givet ved egenværdierne for A, hvilket giver

$$\left|e^{\Lambda_A}\right| = e^{\lambda_1} \cdot e^{\lambda_2} \cdots e^{\lambda_n} = e^{\lambda_1+\lambda_2+\ldots+\lambda_n} = e^{\mathrm{tr}(\Lambda_A)}, \qquad (5.5)$$

da $\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$. Sættes udtrykkene (5.4) og (5.5) sammen, fås resultatet

$$\left|e^A\right| = e^{\mathrm{tr}(A)}.$$

■

Lemma 5.8:
Når $(n \times n)$-matricen W er rækkestokastisk og parameteren $|\rho| < 1$, kan logdeterminanten af matricen $I_n - \rho W$ udtrykkes ved

$$\ln|I_n - \rho W| = -\sum_{j=1}^{\infty} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j}.$$

Bevis:
Hvis der haves en $(n \times n)$-matrix A vides det fra lemma 5.7, at den eksponentielle matricen e^A opfylder

$$\left|e^A\right| = e^{\operatorname{tr}(A)} \Rightarrow \operatorname{tr}(A) = \ln\left|e^A\right|.$$

Erstattes matricen A med $\ln\left(I_n - \rho W\right)$, bliver ovenstående til

$$\operatorname{tr}\left(\ln\left(I_n - \rho W\right)\right) = \ln\left|e^{\ln(I_n - \rho W)}\right| = \ln\left|I_n - \rho W\right|. \tag{5.6}$$

Der ses nu på en Taylorudvikling for logaritmen til matricen $I_n - B$, hvor B er en $(n \times n)$-matrix, der giver

$$\ln(I_n - B) = -\sum_{j=1}^{\infty} \frac{B^j}{j}. \tag{5.7}$$

Da der tages logaritmen til matricen $I_n - B$, er det nødvendigt at $I_n - B$ er positivt definit. Det blev bevist i lemma 2.11 og lemma 2.12, at matricen $I_n - \rho W$ er invertibel, både når W har komplekse og når den har reelle egenværdier, og for at vise, at matricen $I_n - \rho W$ er positiv definit, ses der først på egenværdierne for matricen.

Antag λ_k er en egenværdi for $(n \times n)$-matricen B, så

$$\begin{aligned} B\mathbf{v} &= \lambda_k \mathbf{v} \Rightarrow \\ (I_n - B)\,\mathbf{v} &= I_n\mathbf{v} - B\mathbf{v} = \mathbf{v} - \lambda_k\mathbf{v} = (1 - \lambda_k)\mathbf{v}, \end{aligned}$$

for en vektor $\mathbf{v}$. Det vil sige, $1-\lambda_k$ er en egenværdi for matricen $I_n - B$. Sættes $B = \rho W$, hvor W har egenværdierne λ_i for $i = 1, \ldots, n$, så ses det, at matricen $I_n - \rho W$ har egenværdier på formen $1 - \rho\lambda_i$. Da W er rækkestokastisk, har den højst en egenværdi lig 1, og da $|\rho| < 1$, vil alle egenværdierne for $I_n - \rho W$ være positive, hvilket sammen med den egenskab, at $I_n - \rho W$ er invertibel, gør matricen positiv definit.

Sættes $B = \rho W$ i (5.7), ses det, at

$$\ln\left(I_n - \rho W\right) = -\sum_{j=1}^{\infty} \frac{\left(\rho W\right)^j}{j} = -\sum_{j=1}^{\infty} \frac{\rho^j W^j}{j}. \tag{5.8}$$

Sættes resultaterne fra (5.6) og (5.8) sammen, fås

$$\ln\left|I_n - \rho W\right| = \operatorname{tr}\left(\ln\left(I_n - \rho W\right)\right) = \operatorname{tr}\left(-\sum_{j=1}^{\infty} \frac{\rho^j W^j}{j}\right) = -\sum_{j=1}^{\infty} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j}.$$

■

Det kan nu bevises, at logaritmen til determinanten for matricen $I_n - \rho W$ kan approksimeres med en sum bestående af spor af potenser for nabomatricen W.

Sætning 5.9:
Når nabomatricen W er rækkestokastisk og parameteren $|\rho| < 1$, kan logaritmen til determinanten $\left|I_n - \rho W\right|$ approksimeres ved

$$\ln\left|I_n - \rho W\right| \approx -\sum_{j=1}^{m} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j}. \tag{5.9}$$

Bevis:
Det blev bevist i lemma 5.8, at logaritmen til determinanten $|I_n - \rho W|$ kan udtrykkes ved

$$\ln\left|I_n - \rho W\right| = -\sum_{j=1}^{\infty} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j}. \tag{5.10}$$

Det vil sige, at logdeterminanten består af en vægtet sum af spor af potenser for matricen W. Ligning (5.10) gælder også for komplekse værdier af ρ, og hvis matricen W har komplekse egenværdier, idet logaritmefunktionen ln også er defineret for det komplekse plan, dog med undtagelse af værdien nul.

Det er muligt at dele summen fra (5.10) i en endelig sum af orden m adderet med et restled R, der indeholder summen for $j = m+1$ til uendelig, hvilket giver

$$\ln\left|I_n - \rho W\right| = -\sum_{j=1}^{m} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j} - \sum_{j=m+1}^{\infty} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j} = -\sum_{j=1}^{m} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j} - R. \tag{5.11}$$

Værdien m er en valgt, øvre grænse for approksimationen. Da $|\rho| < 1$ og W er rækkestokastisk med største egenværdi 1, vil $\frac{\rho^j \operatorname{tr}(W^j)}{j} \to 0$ når $j \to \infty$, hvilket bevises ved først at se nærmere på den rækkestokastiske matrix W. Matricen W er rækkestokastisk hvis og kun hvis $W \cdot \iota_n = \iota_n$, hvor ι_n er en vektor med n ettaller. Produktet af to rækkestokastiske matricer er også rækkestokastisk, da

$$(W \cdot W) \cdot \iota_n = W \cdot \iota_n = \iota_n,$$

så W^2 er per definition rækkestokastisk. Derfor er $\operatorname{tr}(W^j) \leq n$, som er antaller af rækker og søjler i W, så

$$\frac{\rho^j \operatorname{tr}(W^j)}{j} \leq \frac{\rho^j n}{j} \to 0 \quad \text{for} \quad j \to \infty,$$

da $|\rho| < 1$. Det vil sige, at des større værdien m er i (5.11), des mindre er resteleddet R, så for en lille værdi for R, kan logaritmen til $|I_n - \rho W|$ approksimeres med summen

$$\ln\left|I_n - \rho W\right| \approx -\sum_{j=1}^{m} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j}.$$

■

Hvis der haves meget store datasæt, kan det ofte spare tid at approksimere logdeterminanten, selvom det kan kræve helt op til $O\left(n^3\right)$ operationer at udregne W^j for en stor værdi for j, når W er tæt besat, men der findes også andre metoder til at udregne eller approksimere sporet af W^j. Eksempelvis kan sporet for potenser af W fra (5.9) udregnes ved at bruge egenværdierne λ_i for W, da $\text{tr}\,(W) = \sum_{i=1}^{m} \lambda_i \Rightarrow \text{tr}\left(W^j\right) = \sum_{i=1}^{m} \lambda_i^j$, hvor der er langt simplere at udregne λ_i^j end W^j.

Monte Carlo-approksimation har den fordel, at det er nemt at opdatere udregningerne for enhver værdi af ρ, da W^j i (5.9) forbliver uændret, når ρ ændres. Det kan også ses ved, at summen i (5.9) kan udtrykkes som prikprodukt $\mathbf{a}^T\mathbf{b}$, hvor

$$\mathbf{a} = \begin{bmatrix} -\frac{\text{tr}(W)^{(1)}}{1} & -\frac{\text{tr}(W)^{(2)}}{2} & \cdots & -\frac{\text{tr}(W)^{(m)}}{m} \end{bmatrix}^T$$
$$\mathbf{b} = \begin{bmatrix} \rho & \rho^2 & \ldots & \rho^m \end{bmatrix}^T.$$

Hvis parameteren ρ ændres, er det kun vektoren $\mathbf{b}$, som det er nødvendigt at udregne igen.

Monte Carlo-approksimation påvirker ikke det endelige estimat for parameteren ρ betydeligt, hvilket er uddybet i [LeSage and Pace, 2009, s. 100-105], og kan derfor betragtes som en relativt præcis approksimationsmetode.

5.3 Parameterestimering for CAR-modellen

I afsnit 5.2 blev der anvendt maximum likelihood-estimation til at finde parametrene i SAR-modellen, og det er muligt at udføre lignende udregninger for CAR-modellen. For at kunne anvende maximum likelihood-estimation til at finde parametrene i CAR-modellen er det nødvendigt først at kende likelihoodfunktionen. Definition 3.15 giver tætheden for CAR-modellen, hvilket også er likelihoodfunktionen.

Sætning 5.10:
Likelihoodfunktionen for CAR-modellen er givet ved

$$l = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} |\rho W|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sigma^{-2}\left(\mathbf{y} - \left(I-\rho W\right)^{-1} X\boldsymbol{\beta}\right)^T \rho W \left(\mathbf{y} - \left(I-\rho W\right)^{-1} X\boldsymbol{\beta}\right)\right),$$

hvor W er nabomatricen, ρ er en parameter og $\boldsymbol{\beta}$ er en parametervektor. Derudover er X en matrix med baggrundsvariable, σ^2 er variansen og kovariansmatricen for CAR-modellen givet ved $\sigma^2\left(\rho W\right)^{-1}$. Loglikelihoodfunktionen har formen

$$L = \ln(l) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) + \frac{1}{2}\ln|\rho W| - \frac{1}{2}\sigma^{-2}\left(\mathbf{y} - \left(I-\rho W\right)^{-1} X\boldsymbol{\beta}\right)^T \rho W \left(\mathbf{y} - \left(I-\rho W\right)^{-1} X\boldsymbol{\beta}\right).$$

Bevis:
Ifølge definition 3.15 er CAR-modellen multivariat normalfordelt med sandsynlighedsfordelingen

$$p(\mathbf{y}) = l = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} |B|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sigma^{-2}\left(\mathbf{y}-\boldsymbol{\mu}\right)^T B\left(\mathbf{y}-\boldsymbol{\mu}\right)\right), \tag{5.12}$$

hvor $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}^T$ er en vektor af observationer, $\boldsymbol{\mu}$ er en $(n \times 1)$-vektor af endelige middelværdier μ_i og B er en $(n \times n)$-matrix, der har nuller på diagonalen, og hvor alle andre elementer er $-\beta_{i,j}$ for $i, j = 1, \dots, n$. I beviset for sætning 3.18 blev det antaget, at

$$\begin{aligned} \boldsymbol{\mu} &= \left(I - \rho W\right)^{-1} X\boldsymbol{\beta} \\ B &= \rho W, \end{aligned}$$

hvormed (5.12) får formen

$$l = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} |\rho W|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sigma^{-2}\left(\mathbf{y} - \left(I - \rho W\right)^{-1} X\boldsymbol{\beta}\right)^T \rho W \left(\mathbf{y} - \left(I - \rho W\right)^{-1} X\boldsymbol{\beta}\right)\right).$$

Det vil sige, CAR-modellen har kovariansmatricen $\Sigma_c = \sigma^2 \left(\rho W\right)^{-1}$. Loglikelihoodfunktionen fås ved at tage logaritmen af likelihoodfunktionen, hvorved det fås, at

$$L = \ln(l) = -\frac{n}{2}\ln\left(2\pi\sigma^2\right) + \frac{1}{2}\ln|\rho W| - \frac{1}{2}\sigma^{-2}\left(\mathbf{y} - \left(I - \rho W\right)^{-1} X\boldsymbol{\beta}\right)^T \rho W \left(\mathbf{y} - \left(I - \rho W\right)^{-1} X\boldsymbol{\beta}\right).$$

■

Det er nu muligt at anvende mindste kvadraters metode til at finde ligninger for parametrene σ^2 og $\boldsymbol{\beta}$ i CAR-modellen, på samme måde som i beviset for sætning 5.6.

Sætning 5.11:
Mindste kvadraters metode anvendt på CAR-modellen giver estimaterne

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(X^T X\right)^{-1} X^T \left(I_n - \rho^* W\right)\mathbf{y} \\ \hat{\sigma}^2 &= n^{-1}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}} \\ \hat{\Sigma}_c &= n^{-1}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}\left(\rho^* W\right)^{-1}, \end{aligned}$$

hvor σ^2 er variansen, $\boldsymbol{\beta}$ er parametervektoren og Σ er kovariansmatricen. Den ægte værdi for parameteren ρ noteres ρ^* og derudover er fejlvektoren

$$\hat{\boldsymbol{\varepsilon}} = \left(I_n - \rho^* W\right)^{-\frac{1}{2}}\left(\mathbf{y} - \rho^* W\mathbf{y} - X\hat{\boldsymbol{\beta}}\right).$$

Bevis:
Det antages, at ρ^* er kendt, så CAR-modellen kan jævnfør sætning 3.18 skrives

$$\begin{aligned} \mathbf{y} &= \left(I - \rho^* W\right)^{-1} X\boldsymbol{\beta} + \left(\left(I_n - \rho^* W\right)^{-1}\right)^{\frac{1}{2}} \boldsymbol{\varepsilon} \Rightarrow \\ \left(I - \rho^* W\right)\mathbf{y} &= X\boldsymbol{\beta} + \left(I_n - \rho^* W\right)^{\frac{1}{2}} \boldsymbol{\varepsilon}. \end{aligned}$$

Defineres $\tilde{\mathbf{y}} := \left(I_n - \rho^* W\right)\mathbf{y}$ og $\tilde{\boldsymbol{\varepsilon}} = \left(I_n - \rho^* W\right)^{\frac{1}{2}} \boldsymbol{\varepsilon}$, fås formlen $\tilde{\mathbf{y}} = X\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$, hvilket svarer til udtrykket (5.3) i beviset for sætning 5.6. Dermed kan resultatet fra sætning 5.6 anvendes, så estimatet for $\boldsymbol{\beta}$ i CAR-modellen kan skrives

$$\hat{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T \left(I_n - \rho^* W\right)\mathbf{y},$$

mens variansen σ^2 estimeres ud fra fejlvektoren

$$\begin{aligned}\left(I_n-\rho^* W\right)^{\frac{1}{2}}\hat{\boldsymbol{\varepsilon}} &= \mathbf{y}-\rho^* W\mathbf{y}-X\hat{\boldsymbol{\beta}} \Rightarrow \\ \hat{\boldsymbol{\varepsilon}} &= \left(I_n-\rho^* W\right)^{-\frac{1}{2}}\left(\mathbf{y}-\rho^* W\mathbf{y}-X\hat{\boldsymbol{\beta}}\right).\end{aligned}$$

Dermed bliver den estimerede varians for n observationer

$$\hat{\sigma}^2 = \frac{\|\hat{\boldsymbol{\varepsilon}}\|^2}{n} = n^{-1}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}.$$

Hvis estimatet for variansen indsættes i udtrykket for kovariansmatricen, der for CAR-modellen er givet ved $\Sigma_c = \sigma^2\left(\rho^* W\right)^{-1}$, fås estimatet

$$\begin{aligned}\hat{\Sigma}_c &= n^{-1}\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}\left(\rho^* W\right)^{-1} \\ &= n^{-1}\left(\left(I_n-\rho^* W\right)^{-\frac{1}{2}}\left(\mathbf{y}-\rho^* W\mathbf{y}-X\hat{\boldsymbol{\beta}}\right)\right)^T\left(I_n-\rho^* W\right)^{-\frac{1}{2}}\left(\mathbf{y}-\rho^* W\mathbf{y}-X\hat{\boldsymbol{\beta}}\right)\left(\rho^* W\right)^{-1}. \blacksquare\end{aligned}$$

Det er tilstrækkeligt at finde profilloglikelihoodfunktionen med hensyn til parameteren ρ for at finde estimater for alle parametrene i CAR-modellen, da sætning 5.11 kan anvendes til at finde de resterende parameterestimater, ligesom det blev gjort ved SAR-modellen. Profilloglikelihoodfunktionen med hensyn til parameteren ρ for CAR-modellen har formen

$$L_p(\rho) = \kappa_c + \frac{1}{2}\ln|\rho W| - \nu_c\left(S_c(\rho)\right)^T\rho W\cdot S_c(\rho), \tag{5.13}$$

hvor $\kappa_c = -\frac{n}{2}\ln\left(2\pi\sigma^2\right)$ og $\nu_c = \frac{1}{2\sigma^2}$ er konstanter, mens

$$S_c(\rho) = \mathbf{y}-\left(I-\rho W\right)^{-1}X\left(\left(X^TX\right)^{-1}X^T\left(I_n-\rho W\right)\mathbf{y}\right).$$

For at finde estimatet for ρ differentieres profilloglikelihoodfunktionen, hvilket er nemmest at overskue, hvis det gøres led for led. Det første led i (5.13) er en konstant og udgår derfor ved differentiation, mens det andet led bliver til

$$\begin{aligned}\frac{\partial}{\partial\rho}\left(\frac{1}{2}\ln|\rho W|\right) &= \frac{1}{2}|\rho W|^{-1}|\rho W|\cdot\operatorname{tr}\left(\left(\rho W\right)^{-1}W\right) \\ &= \frac{1}{2}\rho^{-1}\cdot\operatorname{tr}(I_n) \\ &= \frac{n}{2\rho}\end{aligned}$$

ved brug af regneregel C.1. Kædereglen kan anvendes til at differentiere tredje led i profilloglikelihoodfunktionen (5.13), hvilket giver

$$\frac{\partial}{\partial\rho}\nu_c\left(S_c(\rho)\right)^T\rho W\cdot S_c(\rho) = \frac{\partial f}{\partial g}\frac{\partial g}{\partial\rho}+\frac{\partial f}{\partial h}\frac{\partial h}{\partial\rho} \tag{5.14}$$

hvor $f(g,h) = \nu_c\cdot g(\rho)\cdot h(\rho)$, mens $g(\rho) = \left(S_c(\rho)\right)^T$ og $h(\rho) = \rho W S_c(\rho)$. Differentieres funktionen f med hensyn til henholdsvis g og h, fås

$$\begin{aligned}\frac{\partial f}{\partial g} &= \nu_c\rho W S_c(\rho) \\ \frac{\partial f}{\partial h} &= \nu_c\left(S_c(\rho)\right)^T.\end{aligned}$$

Differentialet af funktionen $S_c(\rho)$ anvendes flere steder og er

$$\begin{aligned}\frac{\partial S_c(\rho)}{\partial \rho} &= -\left(\frac{\partial}{\partial \rho}(I-\rho W)^{-1}\right)\cdot X\left(\left(X^TX\right)^{-1}X^T(I_n-\rho W)\mathbf{y}\right)\\ &\quad -(I-\rho W)^{-1}\cdot\left(\frac{\partial}{\partial \rho}X\left(\left(X^TX\right)^{-1}X^T(I_n-\rho W)\mathbf{y}\right)\right)\\ &= (I-\rho W)^{-1}W(I-\rho W)^{-1}\cdot X\left(\left(X^TX\right)^{-1}X^T(I_n-\rho W)\mathbf{y}\right)\\ &\quad +(I-\rho W)^{-1}\cdot X\left(\left(X^TX\right)^{-1}X^TW\mathbf{y}\right),\end{aligned}$$

hvor regneregel C.2 anvendes til at opnå andet lighedstegn. Det vil sige, ligning (5.14) bliver til

$$\begin{aligned}\frac{\partial}{\partial \rho}L_p(\rho) &= \frac{n}{2\rho}-\left(\frac{\partial f}{\partial g}\frac{\partial g}{\partial \rho}+\frac{\partial f}{\partial h}\frac{\partial h}{\partial \rho}\right)\\ &= \frac{n}{2\rho}-\nu_c\rho WS_c(\rho)\cdot\left((I-\rho W)^{-1}W(I-\rho W)^{-1}\cdot X\left(\left(X^TX\right)^{-1}X^T(I_n-\rho W)\mathbf{y}\right)\right.\\ &\quad \left.+(I-\rho W)^{-1}\cdot X\left(\left(X^TX\right)^{-1}X^TW\mathbf{y}\right)\right)^T\\ &\quad -\nu_c\left(S_c(\rho)\right)^T\cdot WS_c(\rho)-\rho W\cdot\left((I-\rho W)^{-1}W(I-\rho W)^{-1}\cdot X\left(\left(X^TX\right)^{-1}X^T(I_n-\rho W)\mathbf{y}\right)\right.\\ &\quad \left.+(I-\rho W)^{-1}\cdot X\left(\left(X^TX\right)^{-1}X^TW\mathbf{y}\right)\right).\end{aligned}$$

Sættes differentialet lig nul kan ρ isoleres og dermed give maximum likelihood-estimatet. Den analytiske løsning af ovenstående ligning kan være mindre overskuelig end den numeriske løsning, hvor der indsættes konkrete værdier for X, W og $\mathbf{y}$, inden ρ isoleres. Det er også muligt at overlade beregingen til programmet R, hvilket er tilfældet i databehandlingen i kapitel 7.

Bayesiansk tilgang til rumligt afhængige modeller 6

Kapitlet er udfærdiget med udgangspunkt i [Olofsson, 2005, s. 373-380], [LeSage and Pace, 2009, kapitel 5] og [Lee, 2004, kapitel 2 og 3], medmindre andet er angivet.

Et alternativ til at anvende maximum likelihood-estimation er at anvende den bayesianske tilgang til at finde estimater for hver parameter i en model. Bayesiansk statistik anvender fordelinger for modelparametrene dannet ud fra eventuelle oplysninger, der haves om hver parameter på forhånd, hvilket kaldes en priorfordeling. Bayes' sætning anvendes til at opdatere priorfordelingen, når der afdækkes flere oplysninger, ved at der fremkommer en posteriorfordeling. Posteriorfordelingen bliver da den nye priorfordeling, hvis eksperimentet gentages.

Bayesiansk statistik adskiller sig fra klassisk statistik ved at medtage forhåndsviden om parametrene i estimeringen af samme, og også ved at en hypotese kan tillægges en sandsynlighed i stedet for kun at kunne af- eller bekræftes som i klassisk statistik.

Når der er fundet en posteriorfordeling for et givet datasæt og de tilhørende, ukendte parametre, kan fordelingen inddeles i marginale fordelinger til hver parameter givet resten af parametrene. Herefter kan Markov Chain Monte Carlo-estimering (MCMC) anvendes til at generere estimater for hver parameter ud fra de marginale fordelinger, hvilket er mere overskueligt end at forsøge sig med en analytisk løsning af posteriorfordelingen, hvor alle parametrene fra modellen indgår.

6.1 Den bayesianske metode

Den bayesianske metode fokuserer både på at finde en fordeling for et givet datasæt, og en fordeling for parametrene i den model, der anvendes på datasættet. Et datasæts fordeling er givet ved likelihoodfunktionen, der blev fundet for SAR- og CAR-modellen i kapitel 5, og fordelingerne for modellens paramtre er til at begynde med givet ved nogle valgte priorfordelinger, som afbilder det, der vides om parametrene på forhånd. Når parametrenes priorfordelinger og observationernes likelihoodfunktion kendes, kan posteriorfordelingen findes. Posteriorfordelingen indeholder al den information, der er relevant med hensyn til at finde estimater for modellens parametre.

Den bayesianske metode tager udgangspunkt i Bayes sætning, som er vist herunder og er bevist ud fra [Olofsson, 2005, s. 49-50].

Sætning 6.1 (Bayes sætning):
Lad loven om total sandsynlighed være gældende, mens hændelsen A opfylder $P(A) > 0$. Da gælder for enhver hændelse B_j, hvor $j = 1,2,\ldots$, at

$$P\left(B_j|A\right) = \frac{P\left(A|B_j\right)P\left(B_j\right)}{\sum_{k=1}^{n} P(A|B_k)\,P(B_k)}. \tag{6.1}$$

Bevis:
Loven om total sandsynlighed (se [Olofsson, 2005, s. 43]) medfører, at ligning (6.1) kan skrives

$$P\left(B_j|A\right) = \frac{P\left(A|B_j\right)P\left(B_j\right)}{P(A)},$$

så det kun er nødvendigt at bevise dette udtryk. Den betingede sandsynlighed for to hændelser A og B_j er jævnfør [Olofsson, 2005, s. 29] defineret som

$$P\left(A|B_j\right) = \frac{P\left(A \cap B_j\right)}{P\left(B_j\right)}$$

og omvendt for $P\left(B_j|A\right)$, hvilket giver

$$\begin{aligned} P\left(A|B_j\right)P\left(B_j\right) = P\left(A \cap B_j\right) = P\left(B_j|A\right)P(A) \Rightarrow \\ P\left(B_j|A\right) = \frac{P\left(A|B_j\right)P\left(B_j\right)}{P(A)}. \end{aligned}$$

■

Givet $P(A) \neq 0$ ses det også, at udtrykket 6.1 kan skrives

$$P\left(B_j|A\right) \propto P\left(A|B_j\right)P\left(B_j\right).$$

Resultatet fra Bayes sætning kan anvendes til at skrive sandsynlighedsfunktionen for en models parametre $\boldsymbol{\theta} = \left[\theta_1,\theta_2,\ldots,\theta_n\right]^T$ givet en mængde data $\mathcal{D}$, som i SAR-modellens tilfælde er $\mathcal{D} = \{\mathbf{y}, X, W\}$, hvilket giver

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{p(\mathcal{D})}. \tag{6.2}$$

Her er $p(\boldsymbol{\theta})$ priorfordelingen for modelparametrene $\boldsymbol{\theta}$ og er et udtryk for, hvad der vides om modelparametrene, inden der observeres data. For eksempel kan det vides fra tidligere, lignende observationer, at en modelparameter vil ligge i nærheden af en bestemt værdi, hvormed der kan konstrueres en priorfordeling, som er centreret om denne værdi. Hvis der haves begrænset eller ingen forhåndsviden, anvendes ofte en uniform fordeling, der tillægger alle eksperimentets udfald samme sandsynlighed. I resten af rapporten vil priorfordelinger for overskuelighedens skyld blive noteret med π, så $\pi(\boldsymbol{\theta})$ er priorfordelingen for modelparametrene. I (6.2) svarer $p(\mathcal{D}|\boldsymbol{\theta})$ til likelihoodfunktionen, så ligningen kan også skrives

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{l(\boldsymbol{\theta}|\mathcal{D})\,\pi(\boldsymbol{\theta})}{p(\mathcal{D})} \propto l(\boldsymbol{\theta}|\mathcal{D})\,\pi(\boldsymbol{\theta}), \tag{6.3}$$

givet $p(\mathcal{D}) \neq 0$. Det er muligt at se bort fra $p(\mathcal{D})$, da parametervektoren $\boldsymbol{\theta}$ ikke indgår i $p(\mathcal{D})$, som dermed er konstant i forhold til $\boldsymbol{\theta}$. Fordelingen $p(\boldsymbol{\theta}|\mathcal{D})$ i ovenstående er posteriorfordelingen for parametrene $\boldsymbol{\theta}$ givet datamængden $\mathcal{D}$, hvilket svarer til, at priorfordelingen $\pi(\boldsymbol{\theta})$ er blevet opdateret med den nye viden, der er opnået gennem de observerede data $\mathcal{D}$.

Det kan have stor indflydelse på den endelige posteriorfordeling, hvis det tilgængelige datasæt kun består af en afgrænset del af en stor mængde af rumlige punkter, der varierer meget fra område til område, så datasættet kun indeholder begrænset information. Da vil posteriorfordelingen afhænge mere af priorfordelingen end af likelihoodfunktionen. Hvis det modsatte er tilfældet, nemlig at der ikke haves ret meget forhåndsviden, så vil posteriorfordelingen lægge større betydning i likelihoodfunktionen end i priorfordelingen.

Overordnet kan den bayesianske metode ligesom den klassiske inddeles i tre trin, som vil blive gennemgået i det følgende.

1. Estimering og inferens af modelparametre

Første trin i den bayesianske metode er at se på modelestimering og inferens for modelparametrene i en bestemt modeltype som SAR eller CAR-modellen. Hvis der haves et datasæt $\mathcal{D}$, tager estimering af parametrene $\boldsymbol{\theta}$ udgangspunkt i posteriorfordelingen $p(\boldsymbol{\theta}|\mathcal{D})$, der er vist i ligning (6.3). Hvis modelparametrene

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \dots \end{bmatrix}^T$$

er indbyrdes uafhængige, kan priorfordelingen $p(\boldsymbol{\theta})$ skrives som produktet $p(\theta_1)\, p(\theta_2) \cdots$, hvilket kan forsimple estimeringen af parametrene. I den bayesianske metode findes estimaterne ud fra Gibbs-sampling kombineret med Metropolis-Hastings i form af en MCMC-algortime, der forklares i afsnit 6.2 og vises i både SAR- og CAR-modellens tilfælde henholdsvis i afsnit 6.4 og 6.5.

2. Modelspecifikation

Ved modelspecifikation og også modeltest ses der på forskellige typer af samme SAR- eller CAR-model. Her er særligt to tilfælde interessante, nemlig tilfældet hvor der haves flere mulige vægtmatricer W_i for $i = 1,\dots,m$, og tilfældet hvor der haves k mulige baggrundsvariable til matricen X. I begge scenarier haves der flere mulige modeller M_i at vælge i mellem, og i afsnit 6.3 vil det blive uddybet, hvordan den mest nøjagtige model vælges, både ved flere alternative vægtmatricer og ved flere mulige baggrundsvariable.

3. Forecasting

Tredje trin i den bayesianske metode er at anvende de fundne estimater til at komme med et bud på hvordan fremtidige observationer vil se ud, hvilket også kaldes forecasting. De simulerede værdier fra forecasting noteres $\hat{\mathbf{y}}$ og er fundet ud fra det givne datasæt $\mathcal{D}$, hvor posteriorfordelingen kan skrives

$$p\left(\hat{\mathbf{y}}|\mathcal{D}\right) = \int_{\Theta} p\left(\hat{\mathbf{y}},\boldsymbol{\theta}|\mathcal{D}\right) d\boldsymbol{\theta} = \int_{\Theta} p\left(\hat{\mathbf{y}}|\mathcal{D},\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}|\mathcal{D}\right) d\boldsymbol{\theta}.$$

Her er $\boldsymbol{\theta}$ modelparametrene og Θ er mængden af mulige værdier, som vektoren $\boldsymbol{\theta}$ kan bestå af. Udtrykket er fundet ved at omskrive den simultane fordeling for de simulerede data $\hat{\mathbf{y}}$ og parametrene $\boldsymbol{\theta}$ til produktet af en marginal og en betinget fordeling.

De tre ovenstående trin gennemgås for samme modeltype, hvorved det også kan være relevant at gennemgå de tre trin igen for en anden modeltype, så der til sidst haves to estimerede modeller at sammenligne, for eksempel en SAR- og en CAR-model. Her kan der eventuelt sammenlignes residualplot for den mest nøjagtige model af hver type og dermed afgøre, hvilken af de to modeltyper, der beskriver datasættet bedst. Derudover er det værd at bemærke, at hvis der haves en stor mængde data, men ingen relevant a priori-viden, så vil resultatet fra den bayesianske metode ikke adskille sig meget fra maximum likelihood-estimation, da begge tilgange vil finde estimater ud fra datasættet alene.

I afsnit 6.2 vil trin et om modelspecifikation blive uddybet teoretisk, og i afsnit 6.4 vil teorien blive anvendt på SAR-modellen, hvorefter samme vil blive gennemgået for CAR-modellen. Trin to om modelsammenligning vil også blive uddybet teoretisk i afsnit 6.3 dog uden at gå i detaljer med modeltyperne SAR- og CAR-modellen.

6.2 Bayesiansk estimering af modelparametre

Dette afsnit er skrevet på baggrund af [Lee, 2004, s. 245-270].

I praksis tager posteriorfordelingen fra ligning (6.3) ikke altid form som en kendt fordeling, hvilket kan vanskeliggøre parameterestimeringen for en model. Derfor anvendes „Markov Chain Monte Carlo“-metoden (MCMC) til at finde estimaterne i den bayesianske metode, da MCMC danner en algoritme, som ud fra et vilkårligt begyndelsespunkt kan finde passende estimater for en model. Som det fremgår af navnet, består MCMC af en Markovkæde sat sammen med Monte Carlo-metoden, hvor den simpleste version af sidstnævnte gør det muligt at approksimere vanskelige integraler over eksempelvis en posteriorfordeling med en sum som

$$\int_a^b f(x)\, p(x)\, dx \cong \frac{1}{n} \sum_{i=1}^{n} f(x_i),$$

hvor $x_1, x_2, \ldots, x_n$ er valgte, uafhængige værdier, som har tætheden $p(x)$ på intervallet (a,b). I det mest simple tilfælde sættes tætheden $p(x) = \frac{1}{b-a}$ for et kontinuert x, hvilket svarer til en uniform fordeling. Anvendes Monte Carlo-metoden sammen med en Markovkæde, fås MCMC, der kan anvendes til at finde de ønskede estimater for en model. I MCMC ses der på sandsynligheden $p\left(y|x\right)$ for, at Markovkæden vil befinde sig i leddet y til tiden $t+1$ givet kæden er i x til tiden t. Det vil sige, at hvis tæthedsfunktionen for kæden er $p^{(0)}(x)$ til tiden 0 så er tætheden til tiden 1 givet ved

$$p^{(1)}\left(y\right) = \sum_x p^{(0)}(x)\, p\left(y|x\right). \tag{6.4}$$

Hvis Markovkædens mulige tilstande er kontinuerte og ikke diskrete, erstattes summen i (6.4) med et integral. Pointen er, at når tiden t vokser, vil $p^{(t)}\left(y\right)$ nærme sig en bestemt tæthedsfunktion $p\left(y\right)$ uanset begyndelsesværdien $p^{(0)}\left(y\right)$, hvorved der for et

tilstrækkeligt stort t vil opnås et passende estimat uanset begyndelsesværdien. MCMC-algoritmen for SAR- og CAR-modellerne kan eksempelvis bestå af en Gibbs-sampler, der genererer estimater for parametre med kendte fordelinger, samt Metropolis Hastings-metoden, der generere estimater for parametre med ukendte fordelinger. I det følgende vil Gibbs-sampling og Metropolis Hastings-metoden blive gennemgået, og derefter sat sammen i en MCMC-algoritme.

Gibbs-sampler

Gibbs-sampling er en algoritme til at generere estimater for elementerne i parametervektoren $\boldsymbol{\theta}$, der hører til en mængde observationer $\mathbf{y}$, hvor det antages, at de betingede fordelinger for parametrene $\theta_i \in \boldsymbol{\theta}$ er kendte for $i = 1,2,\ldots,k$. I tilfældet med SAR-modellen kan det for eksempel antages, at værdien ρ er kendt, hvorefter parametrene σ^2 og $\boldsymbol{\beta}$ kan estimeres med Gibbs-sampling. Ved Gibbs-sampling vælges først en mængde begyndelsesværdier $\theta_1^{(0)},\theta_2^{(0)},\ldots,\theta_k^{(0)}$ fra priorfordelingerne for parametrene $\theta_i \in \boldsymbol{\theta}$, $i = 1,2,\ldots,k$, hvorefter algoritmen er som følger:

1. Der genereres en værdi $\theta_1^{(t+1)}$ fra fordelingen $p\left(\theta_1|\theta_2^{(t)},\theta_3^{(t)},\ldots,\theta_k^{(t)},\mathbf{y}\right)$, hvor t er algoritmens nuværende opholdssted. Værdien $\theta_1^{(t+1)}$ erstatter $\theta_1^{(t)}$.
2. Værdien $\theta_2^{(t+1)}$ genereres fra fordelingen $p\left(\theta_2|\theta_1^{(t+1)},\theta_3^{(t)},\ldots,\theta_k^{(t)},\mathbf{y}\right)$ og erstatter værdien $\theta_2^{(t)}$.
3. Der fortsættes med at generere værdier for $\theta_i^{(t+1)}$ fra $p\left(\theta_i|\theta_1^{(t+1)},\ldots,\theta_{i-1}^{(t+1)},\theta_{i+1}^{(t)},\ldots,\theta_k^{(t)},\mathbf{y}\right)$ til der haves værdier for alle indgange i $\boldsymbol{\theta}$ til tiden $t+1$.

Metoden giver en kæde af værdier $\boldsymbol{\theta}^{(t)} = \begin{bmatrix}\theta_1^{(t)} & \theta_2^{(t)} & \ldots & \theta_k^{(t)}\end{bmatrix}$, der kun afhænger af værdierne til tiden $t-1$, hvilket er det samme som en Markovkæde. Algoritmen udgør Gibbs-sampleren og har den egenskab, at når antallet af gennemløb t vokser, vil $\boldsymbol{\theta}^{(t)}$ nærme sig en vektor af værdier $\boldsymbol{\theta}$, som har simultan fordeling $p\left(\boldsymbol{\theta}|\mathbf{y}\right)$.

Metropolis-Hastings

Det er også muligt at anvende Metropolis-Hastings til at finde estimater for parametre i en model, når den Bayesianske metode anvendes. Metropolis-Hastings anvendes, når der ikke haves kendte fordelinger for parametrene i en model, og den kan også sættes sammen med Gibbs-sampleren i det tilfælde, at en model både har parametre med kendte og nogle med ukendte fordelinger, hvor metoden kaldes Metropolis-i-Gibbs. Et eksempel på Metropolis-i-Gibbs er når et estimat for parameteren ρ i SAR-modellen findes ved Metropolis-Hastings, mens Gibbs-sampling anvendes til at finde estimater for parametrene σ^2 og $\boldsymbol{\beta}$.

Metropolis-Hastings tager udgangspunkt i posteriorfordelingen $p(\boldsymbol{\theta}|\mathcal{D})$ for modelparametrene $\boldsymbol{\theta}$ og en datamængde $\mathcal{D}$, og metoden har den fordel, at sandsynlighedstætheden kan approksimeres, hvis der haves nok simuleringer fra posteriorfordelingen $p(\boldsymbol{\theta}|\mathcal{D})$. Dermed forsvinder problemet med at finde den præcise analytiske udgave af sandsynlighedstætheden.

Metoden går ud på først at vælge en tæthed $f\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right)$, der opfylder $\int_{\boldsymbol{\theta}} f\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = 1$, og denne tæthed kaldes forslagsfordelingen for parametervektoren $\boldsymbol{\theta}^{(t)}$ til tiden t (mulige metoder til at vælge forslagsfordelingen kan ses i [Lee, 2004, s. 270]). Denne fordeling kan efterfølgende anvendes til at finde fordelingen for parametervektoren $\boldsymbol{\theta}^{(t+1)}$, som benyttes til at præcisere fordelingen for $\boldsymbol{\theta}^{(t+2)}$, og der haves dermed en Markovkæde, da fordelingen for fordelingen for $\boldsymbol{\theta}^{(t+1)}$ kun afhænger af fordelingen for $\boldsymbol{\theta}^{(t)}$.

Første skridt i Metropolis-Hastingsalgoritmen er at simulere en værdi $\boldsymbol{\theta}^*$, der kaldes en kandidat til punktet $\boldsymbol{\theta}^{(t+1)}$, fra tætheden $f\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right)$. Herefter kan værdien $\boldsymbol{\theta}^*$ enten accepteres som parameteren $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^*$ eller forkastes, hvorved Markovkæden beholder den nuværende værdi $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$. Tætheden $f\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)$ afgør, om kandidaten $\boldsymbol{\theta}^*$ accepteres eller forkastes. I det tilfælde tætheden er reversibel, hvilket svarer til

$$p\left(\boldsymbol{\theta}^{(t)}|\mathcal{D}\right) f\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right) = p\left(\boldsymbol{\theta}^*|\mathcal{D}\right) f\left(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*\right),$$

bliver $f\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)$ den endelige tæthed, hvormed sandsynligheden for kandidaten $\boldsymbol{\theta}^*$ forkastes er lig sandsynligheden for $\boldsymbol{\theta}^*$ accepteres. I det tilfælde tætheden $f\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)$ ikke er reversibel, så

$$p\left(\boldsymbol{\theta}^{(t)}|\mathcal{D}\right) f\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right) > p\left(\boldsymbol{\theta}^*|\mathcal{D}\right) f\left(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*\right), \tag{6.5}$$

er sandsynligheden for $\boldsymbol{\theta}^*$ accepteres frem for $\boldsymbol{\theta}^{(t)}$ mindre end den omvendte mulighed. Dette resulterer i et ujævnt traceplot for estimaterne for parameteren $\boldsymbol{\theta}$, hvorved det kan være svært at afgøre, om værdierne konvergere. Det er nemmere at vurdere graden af konvergens for estimaterne, hvis estimaterne for $\boldsymbol{\theta}$ fordeler sig jævnt på et traceplot, hvilket svarer til, at omkring halvdelen af kandidatværdierne accepteres. Det vil sige, det ønskes at få det irreversible tilfælde til at ligne det reversible mest muligt. Der defineres derfor en sandsynlighed $\psi_H\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)$, som udligner uligheden (6.5), så

$$\begin{aligned} p\left(\boldsymbol{\theta}^{(t)}|\mathcal{D}\right) f\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right) \psi_H\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right) &= p\left(\boldsymbol{\theta}^*|\mathcal{D}\right) f\left(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*\right) \Rightarrow \\ \psi_H\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right) &= \frac{p\left(\boldsymbol{\theta}^*|\mathcal{D}\right) f\left(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^{(t)}|\mathcal{D}\right) f\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)}. \end{aligned}$$

Sandsynligheden $\psi_H\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)$ har det formål at sikre, at sandsynligheden for $\boldsymbol{\theta}^*$ vælges frem for $\boldsymbol{\theta}^{(t)}$ er lige så stor som sandsynligheden for $\boldsymbol{\theta}^{(t)}$ vælges som $\boldsymbol{\theta}^{(t+1)}$ i stedet for $\boldsymbol{\theta}^*$. Det ønskes, at acceptsandsynligheden er så høj som mulig, hvilket svarer til der højst er 50% sandsynlighed for $\boldsymbol{\theta}^{(t+1)} := \boldsymbol{\theta}^{(t)}$, så sandsynligheden $\psi_H\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)$ sættes til

$$\psi_H\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right) = \min\left[\frac{p\left(\boldsymbol{\theta}^*|\mathcal{D}\right) f\left(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^{(t)}|\mathcal{D}\right) f\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)}, 1\right]. \tag{6.6}$$

Sandsynligheden, for kandidatværdien vælges som $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^*$, er dermed (6.6), og hvis den ikke vælges, beholdes den nuværende værdi så $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$.

Metropolis Hastings-algoritmen kan formelt formuleres som gentagne gennemløb af følgende trin, hvor begyndelsesværdien inden første gennemløb er en valgt værdi $\boldsymbol{\theta}^{(0)}$:

1. Først antages det, at kæden er i $\boldsymbol{\theta}^{(t)}$.

2. Derefter simuleres kandidatværdien $\boldsymbol{\theta}^*$ fra fordelingen $f\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)$.
3. Dernæst udregnes sandsynligheden $\psi_H\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)$ fra ligning (6.6).
4. Generer nu en værdi u mellem nul og et fra en uniform fordeling $U(0,1)$.
5. Hvis $u \leq \psi_H\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)$ accepteres kandidaten som den nye værdi $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^*$, og hvis $u \geq \psi_H\left(\boldsymbol{\theta}^*|\boldsymbol{\theta}^{(t)}\right)$ forkastes kandidatværdien, så $\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)}$.

Resultatet af algoritmen er en følge med vektorer $\{\boldsymbol{\theta}^{(0)},\boldsymbol{\theta}^{(1)},\ldots\}$, hvor antallet af vektorer svarer til antal gennemløb af algoritmen. Efter et passende antal gennemløb af algoritmen skulle følgen af simulerede værdier gerne konvergere mod en fast værdi, der vælges som det endelige estimat. Her er det en god idé at tage højde for et såkaldt burn-in, der er mængden af simulerede værdier, som fås inden algoritmen finder en ligevægtsfordeling, så resten af værdierne for parametrene approksimeret konvergerer mod posteriorfordelingen.

MCMC-algoritmen

Den samlede MCMC-algoritme afhænger af fordelingerne for de parametre, der ønskes estimater for, da Gibbs-sampling anvendes ved kendte fordelinger, mens Metropolis-Hastings anvendes ved ukendte fordelinger. Hvis det eksempelvis antages, at SAR-modellen har kendte fordelinger for parametrene σ^2 og $\boldsymbol{\beta}$, mens fordelingen for ρ er ukendt, kan MCMC-algoritmen opstilles i tre trin. Først vælges begyndelsesværdierne $\sigma^2_{(0)}$, $\boldsymbol{\beta}_{(0)}$ og $\rho_{(0)}$, og derefter gennemgås følgende:

1. Simuler vektoren $\boldsymbol{\beta}_{(t+1)}$ fra fordelingen $p\left(\boldsymbol{\beta}|\sigma^2_{(t)},\rho_{(t)}\right)$ og erstat $\boldsymbol{\beta}_{(t)}$ med $\boldsymbol{\beta}_{(t+1)}$.
2. Simuler parameteren $\sigma^2_{(t+1)}$ fra den inverse $p\left(\sigma^2|\boldsymbol{\beta}_{(t+1)},\rho\right)$ og lad den simulerede værdi erstatte $\sigma^2_{(t)}$.
3. Anvend Metropolis-Hastings til at simulere værdien $\rho_{(t+1)}$ fra $p\left(\rho|\boldsymbol{\beta}_{(t+1)},\sigma^2_{(t+1)}\right)$.

Algoritmen gennemløbes, indtil de simulerede værdier konvergerer, og de genererede værdier indtil da sorteres fra som burn-in. De efterfølgende værdier, der simuleres, anvendes til at finde posteriorestimater og lave inferens.

6.3 Bayesiansk modelsammenligning

Afsnittet er skrevet på baggrund af [LeSage and Pace, 2009, s. 168-175].

Modelsammenligning handler om at finde den model blandt flere mulige, der beskriver en givet mængde observationer mest nøjagtigt, så det er muligt at generere forudsigelser for fremtidige observationer. Modelsammenligning er i dette tilfælde fokuseret på at vælge den mest relevante vægtmatrix, hvis der haves flere forskellige, samt udvælge de baggrundsvariable til matricen X som giver den bedste model, når den overordnede model i form af SAR- eller CAR-modellen er valgt.

Bayesiansk modelsammenligning baseret på forskellige vægtmatricer

Hvis der haves flere forskellige, rumlige vægtstrukturer, kan modelsammenligning baseres på sammenligning af modeller tilhørende m forskellige vægtmatricer $W_1, W_2, \ldots, W_m$ for $m \in \mathbb{N}$, hvilket giver m mulige modeller. Ved modelsammenligning præciseres en priorfordeling til parametrene i hver model, samtidig med hver model tillægges en hyperpriorsandsynlighed, der ofte vælges til $\frac{1}{m}$, hvis hver model har lige stor sandsynlighed for at blive valgt. Posteriorfordelingerne udregnes da sammen med posteriorsandsynligheden for hver model, hvilket anvendes til at vælge den model, der beskriver observationerne bedst.

Det antages, at der haves $m \in \mathbb{N}$ forskellige vægtmatricer til et givet datasæt, hvilket giver m mulige modeller $\mathbf{M} = \begin{bmatrix} M_1 & M_2 & \ldots & M_m \end{bmatrix}^T$, hvor det ønskes at finde den model, der er mest sandsynlig givet data. Alle modellerne i $\mathbf{M}$ har de samme forklarende variable i matricen X, samme parametre og er samme type model, som for eksempel SAR- eller CAR-modellen, da der her er fokus på at finde den vægtmatrix, der beskriver dataene mest nøjagtigt. Hver model M_i, $i = 1, \ldots, m$, tillægges en hyperpriorsandsynlighed $\pi(M_i)$ samt priorfordelinger for alle parametrene $\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \ldots \end{bmatrix}^T$ givet ved $\pi(\boldsymbol{\theta})$, hvilket gør hver model har tilknyttet en udgave af ligning (6.3) i form af

$$p\left(\boldsymbol{\theta}^i \,|\mathcal{D}\,, M_i\right) = \frac{l\left(\boldsymbol{\theta}^i \,\middle|\, \mathcal{D}, M_i\right) \pi\left(\boldsymbol{\theta}^i \,|M_i\right)}{p\left(\mathcal{D}|\, M_i\right)},$$

hvor $l\left(\boldsymbol{\theta}^i \,\middle|\, \mathcal{D}, M_i\right)$ er likelihoodfunktionen og $\pi\left(\boldsymbol{\theta}^i \,|M_i\right)$ er priorfordelingen.

Hvis det ønskes, at de givne observationer skal afgøre posteriorsandsynlighederne for hver model, sættes priorsandsynligheden til $\frac{1}{m}$, hvilket svarer til, der ikke haves forhåndsviden om hvilken model, der er mest sandsynlig. Den samlede sandsynlighed for mængden af mulige modeller $\mathbf{M}$, parametrene $\boldsymbol{\theta}$ og en givet datamængde $\mathcal{D}$ er

$$p(\mathbf{M},\boldsymbol{\theta},\mathcal{D}) = p(\mathcal{D}|\mathbf{M},\boldsymbol{\theta})\,\pi(\mathbf{M},\boldsymbol{\theta}) = p(\mathcal{D}|\mathbf{M},\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\mathbf{M})\,\pi(\mathbf{M}).$$

Ved brug af ligning (6.1) fra Bayes sætning kan ovenstående skrives som

$$p(\mathbf{M},\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\mathbf{M},\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}|\mathbf{M})\,\pi(\mathbf{M})}{p(\mathcal{D})}. \tag{6.7}$$

Da det kun er vægtmatricerne W_i, hvor $i = 1,2,\ldots,m$, der adskiller modellerne $\mathbf{M}$, er parametrene $\boldsymbol{\theta}$ ens for alle modellerne og dermed uden indflydelse på posteriorsandsynligheden for hver enkelt model. Dermed kan (6.7) omskrives til posteriorsandsynligheden for hver model M_i ved

$$p(M_i|\mathcal{D}) = \frac{p(\mathcal{D}|M_i)\,\pi(M_i)}{p(\mathcal{D})} = \int p\left(M_i,\boldsymbol{\theta}^i|\mathcal{D}\right) d\boldsymbol{\theta}^i, \tag{6.8}$$

hvor der integreres over hver parameter i vektoren $\boldsymbol{\theta}^i$, der indeholder parametrene til model M_i. I ligningen (6.8) er $p(\mathcal{D})$ en konstant, der dog skifter alt afhængig af M_i, da $W_i \in \mathcal{D}$, mens $\pi(M_i)$ kaldes hyperpriorfordelingen og som tidligere nævnt kan sættes lig en uniform fordeling. I ligningen (6.8) indgår også $p(\mathcal{D}|M_i)$, der er den marginale

likelihoodfunktion, som jævnfør [Olofsson, 2005, s. 170] kan udregnes ved brug af almindelig sandsynlighedsregning ved

$$p(\mathcal{D}|M_i) = \int_{\Theta^i} p\left(\mathcal{D}|\boldsymbol{\theta}^i, M_i\right) p\left(\boldsymbol{\theta}^i|M_i\right) d\boldsymbol{\theta}^i.$$

Her er $\boldsymbol{\theta}^i \in \Theta^i$, mens Θ^i betegner mængden af mulige værdier i vektoren $\boldsymbol{\theta}^i$.

Bayesiansk modelsammenligning baseret på MC^3

En anden mulighed er, at der haves k mulige baggrundsvariable, og der ønskes en model med de baggrundsvariable, som samlet beskriver den givne mængde observationer bedst, hvilket vil sige, at de 2^k mulige modeller sammenlignes. I modsætning til tilfældet med forskellige vægtmatricer er der her den ulempe, at antallet af mulige variable k ikke behøver være specielt stort, før mængden af mulige modeller bliver næsten uoverskuelig. Hvis der for eksempel haves 29 mulige baggrundsvariable, er der $2^{29} = 536.870.912$ mulige modeller, hvor det ville være meget tidskrævende at finde både prior- og posteriorfordelinger for hver enkelt, mulig model. Derfor anvendes i stedet en metode kaldet „Markov Chain Monte Carlo model composition“ $\left(MC^3\right)$, der konstruerer en algoritme, som udvælger de vigtigste forklarende variable. MC^3-metoden vil blive beskrevet overordnet i dette afsnit, mens yderligere information kan findes i [LeSage and Pace, 2009, s. 168].

Det antages, at der haves k mulige forklarende variable, som kan indgå i matricen X. MC^3 giver en strategisk, stokastisk Markovkædeproces, der kan gennemgå den muligvis store mængde af mulige modeller, hvor områder med stor posteriorsandsynlighed udvælges. Det vil sige, der konstrueres en algoritme til at finde den relevante andel af de mulige modeller 2^k. Algoritmen tager udgangspunkt i en Markovkæde, hvor hvert led er en mulig model, og det antages, at M er den nuværende model og dermed det led, som kæden befinder sig i. Det defineres, at nabolaget til M er

$$nbl(M) = \{M \wedge M(k \pm 1)\},$$

der indeholder M selv samt de modeller, der har enten en variabel mere end M eller en variabel mindre. Næste trin er at definere en overgangsmatrix K, hvor hver indgang er en sandsynlighed $P\left(M_j|M_i\right) = p_{ij}$, som er sandsynligheden for at springe til modellen M_j, som udgør j'te led i Markovkæden, når modellen M_i er den nuværende model. Matricens størrelse afgøres af Markovkædens længde, mens formen er

$$K = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \end{bmatrix}.$$

Her sættes $p_{ij} = 0$ for alle $M_j \notin nbl(M_i)$, og ellers er den konstant.

Der foreslås en ny model M' som erstatning til den nuværende M, og de to modeller sammenlignes ved

$$\min\left(1, \frac{p\left(M'|\mathbf{y}\right)}{p\left(M|\mathbf{y}\right)}\right) \tag{6.9}$$

Ved brug af metoder til univariat integration, der er beskrevet i [LeSage and Pace, 2009, s. 185-187], kan der konstrueres en Metropolis-Hastings sampler, der ved brug af MC^3 finder værdien

$$\frac{p\left(M'|\mathbf{y}\right)}{p\left(M|\mathbf{y}\right)}.$$

Brøken udgør acceptsandsynligheden, men i modsætning til almindelige regressionsmodeller er der her brug for, at der for hver enkelt model gemmes vektorer, som indeholder logmarginaltætheden fundet i MCMC-samplingen, da vektorerne bruges til at udregne posteriorsandsynlighederne for hver af de modeller, MC^3-algoritmen har undersøgt.

Algoritmen kan udvides ved at tilføje et såkaldt „bevægelsestrin“, så der ikke kun foreslås modeller M', der har præcist en variabel mere eller mindre end M. Jævnfør [LeSage and Pace, 2009, s. 174] vil et bevægelsestrin forbedre konvergensen i den samlede proces, da et bevægelsestrin erstatter en tilfældig indgang i overgangsmatricen K med en tilfældig variabel, som ikke er i modellen i forvejen. Et eksempel på et bevægelsestrin er, at hvis den foreslåede model M' har en variabel mindre end M, så lægger bevægelsestrinnet en variabel til, så det endelige forslag M' har samme antal variable som M. Anvendes bevægelsestrin i algoritmen, er metoden med Metropolis Hastings (6.9) kun gyldig så længe sandsynlighederne for, at M' har en variabel mindre end M, en variabel mere eller er underlagt et bevægelsestrin er lige store, hvilket vil sige, at hver har en sandsynlighed på $\frac{1}{3}$.

Sidste trin i algoritmen er at tage gennemsnittet af alle de modeller, MC^3 har gennemgået, hvor hver model vægtes med modellens posteriorsandsynlighed. Summen af alle posteriorsandsynlighederne er én, og den endelige model udgør dermed en linearkombination af alternative modeller med de relevante, forklarende variable. Denne metode sørger for, at der er taget højde for usikkerhed med hensyn til udvælgelsen af de forklarende variable i den endelige model, da denne består af et vægtet gennemsnit i stedet for kun at udvælge en enkelt model. Som i alle regressionsmodeller er der også usikkerhed forbundet med udvælgelsen af parameterværdierne, selvom usikkerheden muligvis formindskes en smule gennem metoden med at anvende gennemsnittet af de alternative modeller.

6.4 Bayesiansk parameterestimering for SAR-modellen

Afsnittet er skrevet på baggrund af [LeSage and Pace, 2009, s. 127-150].

Det vil nu blive vist, hvordan den bayesianske metode kan anvendes til at finde estimater for parametrene i SAR-modellen. Her kan det også bemærkes, at økonomiske datasæt tit har så meget information, at likelihoodfunktionen vil være den dominerende funktion, så a priori-viden har begrænset indflydelse på posteriorfordelingen.

SAR-modellens likelihoodfunktion blev fundet i sætning 5.4, men i bayesiansk statistik kan det i nogle tilfælde være nemmere at se sammenhængen med Bayes' formel, hvis

likelihoodfunktionen skrives som tæthedsfunktionen

$$p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} |I_n - \rho W| \cdot \exp\left\{\frac{-1}{2\sigma^2}\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)\right\}. \tag{6.10}$$

For at kunne finde posteriorfordelingen på formen (6.3) for SAR-modellen er det udover likelihoodfunktionen også nødvendigt at kende priorfordelingerne for parametrene $\pi(\boldsymbol{\theta})$, hvor $\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\beta}^T & \sigma^2 & \rho\end{bmatrix}^T$ for SAR-modellen. Da værdien for parameteren ρ antages at være uafhængig af værdierne for parametrene σ^2 og $\boldsymbol{\beta}$, kan priorfordelingen skrives

$$\pi\left(\boldsymbol{\beta},\sigma^2,\rho\right) = \pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right) = \pi\left(\boldsymbol{\beta}|\sigma^2\right)\pi\left(\sigma^2\right)\pi\left(\rho\right).$$

Parameteren ρ er et vigtigt element i modellen, da den beskriver graden af afhængighed mellem naboobservationer, og derfor kan det være en fordel at tildele denne en uniform fordeling, således at den estimerede værdi fortrinsvist vil baseres på likelihoodfunktionen og dermed data frem for a priori-viden. Den uniforme fordeling defineres på intervallet $\left(\lambda_{\min}^{-1},\lambda_{\max}^{-1}\right)$, hvor værdierne $\lambda_{\max}$ og $\lambda_{\min}$ er henholdsvis den største og den mindste egenværdi til vægtmatricen W. Priorfordelingen kan da skrives som

$$\begin{aligned}\rho &\sim U\left(\lambda_{\min}^{-1},\lambda_{\max}^{-1}\right) \Rightarrow \\ \pi\left(\rho\right) &= \frac{1}{\lambda_{\max}^{-1}-\lambda_{\min}^{-1}},\end{aligned} \tag{6.11}$$

hvor $\lambda_{\min}^{-1} < \rho < \lambda_{\max}^{-1}$, og U angiver en uniform fordeling (se appendiks B.2). Som tidligere nævnt sættes intervallet $\left(\lambda_{\min}^{-1},\lambda_{\max}^{-1}\right)$ ofte lig $[0,1)$ i praksis.

Den simultane priorfordelingen for parametrene σ^2 og $\boldsymbol{\beta}$ kan for eksempel vælges til at være en normal-invers gamma priorfordeling NIG. Det svarer til, at $\boldsymbol{\beta}$ har en normalfordeling (se appendiks B.1) som priorfordeling, der er betinget med σ^2, som har den inverse gammafordeling (se appendiks B.3) som priorfordeling. Dermed er den simultane priorfordeling er givet ved

$$\begin{aligned}\boldsymbol{\beta},\sigma^2 &\sim NIG\left(\mathbf{c},\Sigma,a,b\right) \Rightarrow \\ \pi\left(\boldsymbol{\beta},\sigma^2\right) &= \pi\left(\boldsymbol{\beta}|\sigma^2\right)\pi\left(\sigma^2\right) = N_k\left(\mathbf{c},\sigma^2\Sigma\right) IG\left(a,b\right).\end{aligned} \tag{6.12}$$

Her er $\sigma^2\Sigma$ og $\mathbf{c}$ henholdsvist kovariansmatricen og middelværdivektoren fra normalfordelingen for $\boldsymbol{\beta}$, mens hyperparametrene a og b er to konstanter fra den inverse gammafordeling for σ^2. Fordelingerne for $\boldsymbol{\beta}$ givet σ^2 og for σ^2 er dermed

$$\begin{aligned}\boldsymbol{\beta}|\sigma^2 &\sim N_k\left(\mathbf{c},\sigma^2\Sigma\right) \Rightarrow \\ \pi\left(\boldsymbol{\beta}|\sigma^2\right) &= \frac{1}{\left(2\pi\right)^{k/2}\left|\sigma^2\Sigma\right|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)\right\},\end{aligned} \tag{6.13}$$

og

$$\begin{aligned}\sigma^2 &\sim IG\left(a,b\right) \Rightarrow \\ \pi\left(\sigma^2\right) &= \frac{b^a}{\Gamma\left(a\right)}\left(\sigma^2\right)^{-(a+1)} \exp\left\{-\frac{b}{\sigma^2}\right\}.\end{aligned} \tag{6.14}$$

Her angiver $\Gamma(\cdot)$ gammafunktionen $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$. Det vil sige, at den simultane priorfordeling (6.12) kan præciseres ved at indsætte (6.13) og (6.14), hvilket giver

$$\begin{aligned}
&\boldsymbol{\beta},\sigma^2 \sim N_k\left(\mathbf{c},\sigma^2\Sigma\right) IG(a,b) \Rightarrow \\
&\pi\left(\boldsymbol{\beta},\sigma^2\right) = \frac{b^a}{(2\pi)^{k/2}|\Sigma|^{1/2}\Gamma(a)}\left(\sigma^2\right)^{-(a+(k/2)+1)} \exp\left\{-\frac{1}{2\sigma^2}\left(\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)+2b\right)\right\},
\end{aligned} \tag{6.15}$$

hvor $\sigma^2, a, b > 0$. Da matricen Σ i ovenstående er $k \times k$, giver en almindelig determinantregneregel det anvendte udtryk $|\sigma^2\Sigma|^{-\frac{1}{2}} = \sigma^{-k}|\Sigma|^{-\frac{1}{2}}$.

Sætning 6.2:
Når SAR-modellens parametre ρ, σ^2 og $\boldsymbol{\beta}$ har priorfordelingerne

$$\begin{aligned}
\rho &\sim U\left(\lambda_{\min}^{-1}, \lambda_{\max}^{-1}\right) \\
\boldsymbol{\beta},\sigma^2 &\sim NIG(\mathbf{c},\Sigma,a,b),
\end{aligned}$$

så har posteriorfordelingen formen

$$p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right) \propto \left(\sigma^2\right)^{-\left(a^*+(k/2)+1\right)}\left|I_n-\rho W\right|\exp\left\{-\frac{1}{2\sigma^2}\left(\left(\boldsymbol{\beta}-\mathbf{c}^*\right)^T\left(\Sigma^*\right)^{-1}\left(\boldsymbol{\beta}-\mathbf{c}^*\right)+2b^*\right)\right\},$$

hvor

$$\begin{aligned}
\Sigma^* &= \left(X^TX+\Sigma^{-1}\right)^{-1} \\
a^* &= a+\frac{n}{2} \\
b^* &= b+\frac{\mathbf{c}^T\Sigma^{-1}\mathbf{c}+\mathbf{y}^T\left(I_n-\rho W\right)^T\left(I_n-\rho W\right)\mathbf{y}-\left(\mathbf{c}^*\right)^T\left(\Sigma^*\right)^{-1}\mathbf{c}^*}{2} \\
\mathbf{c}^* &= \Sigma^*\left(X^T\left(I_n-\rho W\right)\mathbf{y}+\Sigma^{-1}\mathbf{c}\right).
\end{aligned}$$

Bevis:
Jævnfør Bayes' sætning 6.1 er SAR-modellens posteriorfordeling givet som

$$\begin{aligned}
p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right) &= \frac{p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right)\pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right)}{p\left(\mathcal{D}\right)} \\
&\propto p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right)\pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right),
\end{aligned} \tag{6.16}$$

hvor det er antaget, at parametrene a priori er uafhængige. Derudover kan leddet $p(\mathcal{D})$ udelades som en konstant, da det ikke afhænger af parametrene. Ligheden (6.17) gælder jævnfør [LeSage and Pace, 2009, s. 129] og viser sig at være nyttig, når posteriorfordelingen skal udregnes.

$$\begin{aligned}
&\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)+\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)+2b \\
&\equiv \left(\boldsymbol{\beta}-\mathbf{c}^*\right)^T\left(\Sigma^*\right)^{-1}\left(\boldsymbol{\beta}-\mathbf{c}^*\right)+2b^*,
\end{aligned} \tag{6.17}$$

hvor

$$\begin{aligned}
\Sigma^* &= \left(X^T X + \Sigma^{-1}\right)^{-1} \\
a^* &= a + \frac{n}{2} \\
b^* &= b + \frac{\mathbf{c}^T\Sigma^{-1}\mathbf{c} + \mathbf{y}^T\left(I_n - \rho W\right)^T\left(I_n - \rho W\right)\mathbf{y} - \left(\mathbf{c}^*\right)^T\left(\Sigma^*\right)^{-1}\mathbf{c}^*}{2} \\
\mathbf{c}^* &= \Sigma^*\left(X^T\left(I_n - \rho W\right)\mathbf{y} + \Sigma^{-1}\mathbf{c}\right).
\end{aligned} \tag{6.18}$$

Posteriorfordelingen kan nu findes ved at indsætte (6.10), (6.11) og (6.15) i udtrykket (6.16), hvilket giver

$$\begin{aligned}
p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right) &\propto p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right)\pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right) \\
&= \left(\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\left|I_n - \rho W\right|\exp\left\{-\frac{1}{2\sigma^2}\left(\left(I_n - \rho W\right)\mathbf{y} - X\boldsymbol{\beta}\right)^T\left(\left(I_n - \rho W\right)\mathbf{y} - X\boldsymbol{\beta}\right)\right\}\right)\cdot \\
&\quad \left(\frac{b^a}{(2\pi)^{k/2}|\Sigma|^{1/2}\Gamma(a)}\left(\sigma^2\right)^{-(a+(k/2)+1)}\exp\left\{-\frac{1}{2\sigma^2}\left(\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)+2b\right)\right\}\right)\cdot \\
&\quad \pi\left(\rho\right) \\
&\propto \left(\sigma^2\right)^{-\left(a^*+(k/2)+1\right)}\left|I_n - \rho W\right|\exp\left\{-\frac{1}{2\sigma^2}\left(\left(\boldsymbol{\beta}-\mathbf{c}^*\right)^T\left(\Sigma^*\right)^{-1}\left(\boldsymbol{\beta}-\mathbf{c}^*\right)+2b^*\right)\right\},
\end{aligned}$$

hvor a^*, c^*, b^* og Σ^* er defineret ved (6.18). Udtrykket (6.17) er anvendt til at opnå andet proportionalitetstegn, og derudover er alle konstanter udeladt, når der tages proportionalitet, hvilket også gælder $\pi\left(\rho\right)$, der er uniformt fordelt og dermed konstant. ■

I ovenstående sætning og bevis er det værd at bemærke, at hvor priorfordelingen NIG normalt fungere som en konjugeret prior i tilfældet med ikke-rumlige modeller, så gælder dette ikke i det rumlige tilfælde, som SAR-modellen hører under. Betegnelsen „konjugeret priorfordeling“ henviser normalt til situationer, hvor posteriorfordeling har samme form som priorfordelingen med den undtagelse, at modelparametrene er opdaterede. I SAR-modellens tilfælde ville det resultere i, at $\rho = 0$, hvormed $I_n - \rho W = I_n$.

A priori-viden medtages i modellen i form af hyperparametrene a, b, $\mathbf{c}$ og Σ fra NIG, mens priorfordelingen for ρ er valgt til at være uniform og dermed afspejler, at der på forhånd ikke haves nogen viden om, hvilken af værdierne fra intervallet $\left(\lambda_{\min}^{-1}, \lambda_{\max}^{-1}\right)$ som ρ vil blive tildelt. Hvis der derimod haves forhåndsviden om værdien for ρ, kan der for eksempel anvendes en betafordeling (se appendiks B.4), som er tilpasset intervallet $\left(\lambda_{\min}^{-1}, \lambda_{\max}^{-1}\right)$. Derudover blev det i beviset for sætning 6.2 antaget, at parameteren ρ a priori er uafhængig af parametrene σ^2 og $\boldsymbol{\beta}$, hvor det er vigtigt at bemærke, at antagelsen ikke nødvendigvis medfører uafhængighed i posteriorfordelingerne for parametrene.

Det er også muligt at vælge en priorfordeling for parameteren $\boldsymbol{\beta}$ i de tilfælde, hvor der ikke haves særlig forhåndsviden om $\boldsymbol{\beta}$. Dette gøres ved at sætte parameteren $\mathbf{c} = \mathbf{0}$, samtidig

med $\boldsymbol{\beta}$ tildeles stor varians og manglende kovarians mellem værdien i indgangene i $\boldsymbol{\beta}$, hvilket afspejles i en høje værdier i indgangene for matricen Σ, eksempelvis $\Sigma = I_k \cdot 10^{10}$. Priorfordelingen bliver dermed uegentlig i den forstand, at der ikke haves en veldefineret tæthed som har summen 1, hvormed det er nødvendigt at tilpasse beregningerne, så posteriorfordelingen summerer til 1 og dermed er veldefineret.

På samme måde kan der defineres en priorfordeling for parameteren σ^2 i det tilfælde, der ikke haves nogen særlig forhåndsviden for parameteren, hvilket gøres ved at lade $a,b \to 0$ i fordelingen (6.14) for σ^2.

I sætning 6.2 anvendes informative priorfordelinger for parametrene $\boldsymbol{\beta}$ og σ^2, hvilket kan medføre, at udledningen af posteriorfordelingen kompliceres på to måder. Første problem er at fastsætte værdier for parametrene i NIG-priorfordelingen ud fra, hvad der vides a priori. Det andet problem er analysen af posteriorfordelingen fra sætning 6.2, da posteriorfordelingen for $\boldsymbol{\beta}$ findes ved at integrere med hensyn til både parameteren σ^2 og parameteren ρ, hvilket kan være vanskeligt. På samme måde er det nødvendigt at integrere med hensyn til både ρ og $\boldsymbol{\beta}$ for at finde udtrykket for posteriorfordelingen til σ^2. Det er derfor værd at overveje, om den forhåndsviden, som udtrykkes i priorfordelingerne, har nogen væsentlig indflydelse på inferens i forhold til parameteren $\boldsymbol{\beta}$, for hvis den ikke har, er det nemmere at anvende priorfordelinger, som ikke er informative, i sætning 6.2.

Derudover kan det bemærkes, at hvis middelværdien fra henholdsvis posteriorfordelingen for $\mathbf{c}^*$ og den for ρ var kendt, ville leddet

$$\mathbf{y}^T \left(I_n - \rho W\right)^T \left(I_n - \rho W\right)\mathbf{y} - \left(\mathbf{c}^*\right)^T \left(\Sigma^*\right)^{-1} \mathbf{c}^*$$

fra b^* i (6.18) give den kvadrerede sum af residualerne for SAR-modellen. Det svarer til, at når $a,b,\Sigma^{-1} \to 0$, vil udtrykket $\frac{b^*}{a^*}$ approksimativt være lig den kvadrerede sum af SAR-modellens residualer. Hvis det er tilfældet for NIG-priorfordelingerne, vil parametrene σ^2 og $\boldsymbol{\beta}$ få priorfordelinger, som ikke er informative.

MCMC anvendt på bayesiansk SAR-model

I det følgende vil teorien om Markov Chain Monte Carlo-estimering fra afsnit 6.2 blive anvendt på SAR-modellen.

I langt de fleste tilfælde er det nemmere at anvende MCMC til at finde de ønskede estimater i stedet for at estimere parametrene analytisk. Jævnfør [LeSage and Pace, 2009, s. 123] gælder det generelt for MCMC, at hvis metoden anvendes sammen med Gibbs-sampling og metropolis-Hastings på en mængde betingede fordelinger for alle parametrene i en model, så vil de fundne estimater konvergerer mod den simultane posteriorfordeling for parametrene. Det vil sige, MCMC først adskiller den simultane posteriorfordeling for modellens parametre i betingede fordelinger og derefter simulerer estimater fra disse, hvor estimaterne vil konvergere mod den „ægte“ fordeling.

Som nævnt i afsnit 6.2 kan MCMC anvende Gibbs-sampling til at simulere parametre fra kendte fordelinger, hvorefter Metropolis-Hastings kan anvendes til at simulere parameterestimater fra ukendte fordelinger. Da fordelingerne for parametrene $\boldsymbol{\beta}$ og σ^2

ofte har kendte priorfordelinger, vil der først blive set på Gibbs-sampling, hvorefter der ses på Metropolis-Hastings med henblik på at finde et estimat for parameteren ρ, da det sjældent vides på forhånd, hvilken type fordeling, denne parameter har. Til sidst vil de to metoder blive samlet i en MCMC-algoritme.

Gibbs-sampling for parametrene $\boldsymbol{\beta}$ og σ^2 i SAR-modellen

Det antages, at parametrene i SAR-modellen er fordelt som i sætning 6.2, og dermed har den simultane fordeling $p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right)$. Gibbs-sampling starter med at finde fordelingen for hver parameter givet de resterende parametre, der antages at være kendte og dermed konstante. De betingede fordelinger for SAR-modellens σ^2 og $\boldsymbol{\beta}$ findes ud fra på NIG-priorfordelingen for samme. Det antages for nu, at parameteren ρ er kendt og dermed konstant, selvom det i de fleste situationer ikke er tilfældet, men da ρ alligevel estimeres efter σ^2 og $\boldsymbol{\beta}$ i disse situationer, fungerer antagelsen også her.

Den betingede multivariate normalfordeling for parameteren $\boldsymbol{\beta}$ er givet ved

$$p\left(\boldsymbol{\beta}|\rho,\sigma^2_{(0)}\right) \sim N_k\left(\mathbf{c}^*,\sigma^2_{(0)}\Sigma^*\right), \tag{6.19}$$

hvor

$$\begin{aligned}\Sigma^* &= \left(X^TX+\Sigma^{-1}\right)^{-1}\\ \mathbf{c}^* &= \Sigma^*\left(X^T\left(I_n-\rho W\right)\mathbf{y}+\Sigma^{-1}\mathbf{c}\right).\end{aligned}$$

Her er $\sigma^2_{(0)}$ en valgt begyndelsesværdi for parameteren σ^2, der dermed kan betragtes som en konstant i ovenstående ligning. Det ses også af ovenstående, at begyndelsesværdien $\sigma^2_{(0)}$ sammen med en konstant værdi for ρ kan anvendes til at finde middelværdien $\mathbf{c}^*$ og matricen Σ^* til fordelingen for $\boldsymbol{\beta}$. Herefter anvendes en algoritme, som konstruerer en vektor bestående af stokastiske, multivariate normale afledede med middelværdi og kovariansmatrix som i (6.19), til at opnå den simulerede værdi $\boldsymbol{\beta}_{(1)}$.

På samme måde er det muligt at simulere en værdi for σ^2 givet en konstant værdi for ρ og værdien $\boldsymbol{\beta}_{(1)}$. Den betingede, inverse gammafordeling for σ^2 er givet ved

$$p\left(\sigma^2|\rho,\boldsymbol{\beta}_{(1)}\right) \sim IG\left(a^*,b^{**}\right),$$

hvor

$$\begin{aligned}a^* &= a+\frac{n}{2}\\ b^{**} &= b+\frac{\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}_{(1)}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}_{(1)}\right)}{2},\end{aligned}$$

som er fundet ud fra værdierne ρ og $\boldsymbol{\beta}_{(1)}$. Den føromtalte algoritme anvendes igen til at simulere en stokastisk værdi fra fordelingen $IG(a^*,b^{**})$, hvor den simulerede værdi $\sigma^2_{(1)}$ erstatter $\sigma^2_{(0)}$.

Gibbs-sampleren gentager nu første trin i algoritmen, og opdaterer værdien $\boldsymbol{\beta}_{(1)}$ til $\boldsymbol{\beta}_{(2)}$, der efterfølgende anvendes til at opdatere $\sigma^2_{(1)}$ til $\sigma^2_{(2)}$. Algoritmen fortsættes, så der til sidst haves en stor mængde simulerede værdier for parametrene σ^2 og $\boldsymbol{\beta}$.

Det er her værd at bemærke, at estimater for $\boldsymbol{\beta}$ og σ^2 fundet ved brug af Gibbs-sampling, som har gennemløbet algoritmen tilstrækkeligt mange gange, giver samme estimater som hvis de var fundet ved brug af eksempelvis Metropolis-Hastings på hele posteriorfordelingen.

Metropolis-Hastings for parameteren ρ fra SAR-modellen

I praksis er det som regel ikke tilfældet, at parameteren ρ er kendt, som det var antaget ved Gibbs-sampling, og derfor ses der nu på, hvordan ρ fra SAR-modellen kan estimeres ved brug af Metropolis-Hastings. Først er det nødvendigt at se på den betingede fordeling for ρ, som er givet ved

$$\begin{aligned} p\left(\rho|\boldsymbol{\beta},\sigma^2,\mathcal{D}\right) &\propto \frac{p\left(\rho,\boldsymbol{\beta},\sigma^2|\mathcal{D}\right)}{\pi\left(\boldsymbol{\beta},\sigma^2|\mathcal{D}\right)} \\ &\propto \left|I_n-\rho W\right| \exp\left\{-\frac{1}{2\sigma^2}\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)\right\}, \end{aligned} \tag{6.20}$$

hvor tælleren er posteriorfordelingen fra sætning 6.2, nævneren er fordelingen (6.15), og $\mathcal{D}$ som tidligere angiver mængden af data. Fordelingen (6.20) ligner ikke en kendt fordeling, så der må anvendes Metropolis-Hastings i stedet for Gibbs-sampling. Metropolis-Hastings starter med at vælge en kandidatfordeling, der jævnfør [Lee, 2004, s. 270] eksempelvis kan være en standard normalfordeling $N(0,1)$, og denne sættes sammen med en justeret random walk, hvorfra der genereres værdien ρ^* ved

$$\rho^* = \rho^t + c \cdot N(0,1)\,. \tag{6.21}$$

Her er ρ^t værdien for parameteren ρ til tiden t, mens c er en konstant kaldet justeringsparameteren, da denne tilpasses, så Metropolis-Hastings kommer til at dække hele den betingede fordeling. Justeringsparameteren anvendes i praksis ved at ændre på c til den værdi, der virker bedst, findes. Dette vil blive beskrevet nærmere i eksempel 6.3. Herefter findes acceptsandsynligheden ved først at indsætte ρ^* og derefter ρ^t i (6.20), hvorefter de to udregnede værdier indsættes i

$$\psi_H\left(\rho^t,\rho^*\right) = \min\left[1, \frac{p\left(\rho^*|\boldsymbol{\beta},\sigma\right)}{p\left(\rho^t|\boldsymbol{\beta},\sigma\right)}\right], \tag{6.22}$$

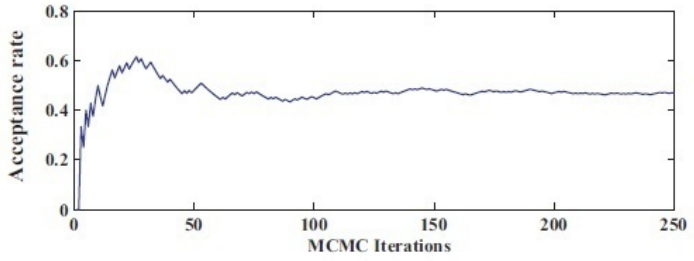
som er acceptsandsynligheden. Værdien (6.22) er sandsynligheden for, at det nye værdi ρ^* erstatter den nuværende ρ^t. Ved at justere konstanten c i (6.21) er det muligt at sikre, at estimatet for ρ simuleres fra den tætte del af fordelingen, og simuleringerne ikke „sætter sig fast“ i en af fordelingens ender, da sandsynlighedsmassen er lav i de områder.

Eksempel 6.3:

Som regel ønskes det, at acceptområdet $\psi_H\left(\rho^t,\rho^*\right)$ er tæt på 50%, idet det vil medføre, at traceplottet for de simulerede værdier hverken springer lidt eller meget, og det dermed er nemmere at vurdere grafisk, om de simulerede værdier for ρ konvergerer. Acceptområdet $\psi_H\left(\rho^t,\rho^*\right)$ vælges derfor til at ligge et sted mellem 40% og 60%, hvilket overholdes ved at ændre c til $c_1 = \frac{c}{1{,}1}$, hvis sandsynligheden i $\psi_H\left(\rho^t,\rho^*\right)$ er mindre end 40% for adskillige simulerede værdier i træk. Når $c_1 = \frac{c}{1{,}1}$, vil $\rho_1^* = \rho_1^t + c_1 \cdot N(0,1)$ være nærmere

ρ^t, så acceptområdet $\psi_H(\rho^t, \rho^*)$ øges. Ligeledes sættes c til $c_1 = 1{,}1 \cdot c$, i det tilfælde $\psi_H(\rho^t, \rho^*)$ er større end 60% for flere simuleringer af ρ, hvorved acceptsandsynligheden mindskes. Hvad enten der haves en stor eller lille mængde data, er det ofte nødvendigt at ændre på c, da en stor datamængde kan give lille spredning i den betingede fordeling for ρ, samtidig med det modsatte ofte er tilfældet for en lille mængde data. Det er ikke nødvendigt at justere c lige meget for alle datasæt, selvom den benyttede metode er ens for datasættene, så i praksis prøver man sig som regel frem.

Figur 6.1 viser, at hvis de første ca. 50 gennemløb af MCMC med Metropolis-Hastings tages fra som burn-in, hvor der muligvis er justeret på parameteren c, så vil resten af simuleringerne være i en ligevægt, hvor $\psi_H(\rho^t, \rho^*)$ ligger mellem 40% og 60%.



Figur 6.1: Acceptsandsynlighed ved gentagne gennemløb af MCMC, [LeSage and Pace, 2009, s. 138].

Det er også værd at bemærke, at når $\psi_H(\rho^t, \rho^*)$ vælges til at være mellem 40% og 60%, vil de simulerede værdier for ρ blive forkastet omkring halvdelen af tiden, hvormed det er nødvendigt at gennemløbe MCMC-algoritmen endnu flere gange for at have nok simulerede værdier til at finde en posteriorfordeling. □

MCMC-algoritmen for SAR-modellen

Gibbs-sampling kan nu sættes sammen med Metropolis-Hastings i en MCMC-algoritme, der finder estimater for alle parametrene σ^2, $\boldsymbol{\beta}$ og ρ i SAR-modellen. Det skal her bemærkes, at det kan være en hjælp at udregne en $(q \times 1)$-vektor ved brug af q tilfældigt udvalgte værdier for $\rho \in [0,1)$, som kan anvendes til at udregne logdeterminantleddet $\ln|I_n - \rho W|$. Vektoren

$$\left[\ln|I_n - \rho_1 W| \quad \ln|I_n - \rho_2 W| \quad \ldots \quad \ln|I_n - \rho_q W|\right]^T$$

indgår i posteriorfordelingen $p(\theta|\mathcal{D})$ i Metropolis-Hastings-metoden, hvor det også er en fordel at anvende regneregler til forenkling af udregning af determinanten. Som tidligere nævnt kan det være vanskeligt at fortolke værdier af ρ uden for intervallet $[0,1)$, så det kan være en hjælp at afvise de værdier af ρ, som ikke er i intervallet, når de genereres ved brug af Metropolis-Hastings-sampling. Herefter kan posteriorsandsynligheden for, at $\rho \in [0,1)$, findes ved at notere, hvor stor en andel af de simulerede værdier, der er blevet afvist.

MCMC-algoritmen for SAR-modellen er inddelt i tre trin, hvor der i hvert trin simuleres fra en betinget fordeling, og inden algoritmen gennemløbes første gang, fastsættes

begyndelsesværdier for parametrene $\boldsymbol{\beta}$, σ^2 og ρ som $\boldsymbol{\beta}_{(0)}$, $\sigma^2_{(0)}$ og $\rho_{(0)}$. Det antages som før, at den simultane priorfordeling for $\boldsymbol{\beta}$ og σ^2 er $NIG(\mathbf{c},\Sigma,a,b)$. Algoritmens tre trin er som følger:

Trin 1
Der simuleres en vektor for $\boldsymbol{\beta}$ fra den betingede posteriorfordeling $p\left(\boldsymbol{\beta}|\sigma^2_{(t)},\rho_{(t)}\right)$, der er givet ved normalfordelingen

$$N\left(\mathbf{c}^*,\sigma_{(t)}\Sigma^*\right),$$

hvor

$$\Sigma^* = \left(X^TX+\Sigma^{-1}\right)^{-1}$$
$$\mathbf{c}^* = \Sigma^*\left(X^T\left(I_n-\rho_{(t)}W\right)\mathbf{y}+\Sigma^{-1}\mathbf{c}\right).$$

Begyndelsesværdien $\boldsymbol{\beta}_{(t)}$ erstattes af den simulerede vektor $\boldsymbol{\beta}_{(t+1)}$.

Trin 2
Der simuleres en værdi for parameteren σ^2 fra den inverse gammafordeling

$$p\left(\sigma^2|\boldsymbol{\beta}_{(t+1)},\rho_{(t)}\right) \sim IG\left(a^*,b^{**}\right),$$

hvor

$$a^* = a+\frac{n}{2}$$
$$b^{**} = b+\frac{\left(\left(I_n-\rho_{(t)}W\right)\mathbf{y}-X\boldsymbol{\beta}_{(t+1)}\right)^T\left(I_n-\rho_{(t)}W\right)\mathbf{y}-X\boldsymbol{\beta}_{(t+1)}}{2}.$$

Begyndelsesværdien $\sigma^2_{(t)}$ erstattes nu af den simulerede værdi $\sigma^2_{(t+1)}$.

Trin 3
Metropolis Hastings-sampling anvendes til at simulere værdien $\rho_{(t+1)}$ fra den betingede fordeling $p\left(\rho|\boldsymbol{\beta}_{(t+1)},\sigma^2_{(t+1)}\right)$.

Algoritmens tre trin gennemløbes adskillige gange for $t = 0,1,\ldots$, hvor de første simuleringer sorteres fra som burn-in, og resten af de simulerede parametre bruges til at konstruere posteriorestimater og lave inferens, da estimaterne efter burn-in approksimeret konvergerer mod posteriorfordelingen. I praksis er det muligt at lave en test for at se, om algoritmen har nået sin ligevægtstilstand, hvilket gøres ved at gennemløbe algoritmen to gange med forskellige startværdier. Eksempelvis kan det første gennemløb være relativt kort på 3500 gange, hvoraf de første 800 sorteres fra som burn-in, mens det andet kan være relativt langt gennemløb på 9000, hvor de første 3500 estimater frasorteres som burn-in. Ligner middelværdien og variansen for hvert af de to gennemløb hinanden, kan det konkluderes, at MCMC-algoritmen konvergerer, og de værdier, som simuleres efter burn-in, vil udgøre estimaterne fra posteriorfordelingen. Dernæst mangler der kun at blive lavet modeltest for de simulerede parametre eksempelvis gennem residualer eller hypotesetest.

Generelt er der ikke de store problemer med konvergens for simple, rumlige modeller som SAR-modellen.

6.5 Bayesiansk parameterestimering for CAR-modellen

Parameterestimeringen for CAR-modellen i den bayesianske tilgang forløber på tilsvarende måde som ved SAR-modellen. Først bemærkes det, at likelihoodfunktionen også kan noteres $p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right)$, hvilket kan være en fordel, hvis Bayes sætning benyttes, og for CAR-modellen er likelihoodfunktionen

$$p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} |\rho W|^{\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}\sigma^{-2}\left(\mathbf{y}-\left(I-\rho W\right)^{-1}X\boldsymbol{\beta}\right)^T \rho W\left(\mathbf{y}-\left(I-\rho W\right)^{-1}X\boldsymbol{\beta}\right)\right), \quad (6.23)$$

som blev bevist i sætning 5.10. Her er W nabomatricen, ρ og $\boldsymbol{\beta}$ er parametre og X er en matrix med baggrundsvariable, mens σ^2 er variansen.

Som i tilfældet med SAR-modellen antages det a priori, at parameteren ρ er uafhængig af $\boldsymbol{\beta}$ og σ^2, så

$$\pi\left(\boldsymbol{\beta},\sigma^2,\rho\right) = \pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right).$$

Da parametrene $\boldsymbol{\beta}$, σ^2 og ρ repræsenterer det samme som i SAR-modellen, antages det, at de har samme priorfordelinger, der blev defineret i (6.11) og (6.15). Det er nu muligt at finde posteriorfordelingen for CAR-modellen.

Sætning 6.4:
Lad parametrene $\boldsymbol{\beta}$, σ^2 og ρ være fordelt a priori som

$$\rho \sim U\left(\lambda_{\min}^{-1},\lambda_{\max}^{-1}\right)$$
$$\boldsymbol{\beta},\sigma^2 \sim N_k\left(\mathbf{c},\sigma^2\Sigma\right) IG\left(a,b\right),$$

hvor $\lambda_{\max}$ og $\lambda_{\min}$ betegner den største og den mindste egenværdi for matricen W, mens $\sigma^2, a, b > 0$ og Σ er en $(k \times k)$-matrix. Da er posteriorfordelingen for CAR-modellen givet ved

$$p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right) \propto \left(\sigma^2\right)^{-\left(a+(k/2)+1+\frac{n}{2}\right)} |\rho W|^{\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}\sigma^{-2}\left(\left(\mathbf{y}-\left(I-\rho W\right)^{-1}X\boldsymbol{\beta}\right)^T \rho W\left(\mathbf{y}-\left(I-\rho W\right)^{-1}X\boldsymbol{\beta}\right)+\left(\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)+2b\right)\right)\right) \quad (6.24)$$

Bevis:
Anvendes Bayes sætning 6.1, fås resultatet (6.3), der også kan skrives som

$$\begin{aligned} p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right) &\propto l\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right)\pi\left(\boldsymbol{\beta},\sigma^2,\rho\right) \\ &= p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right)\pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right). \end{aligned}$$

Indsættes likelihoodfunktionen (6.23) og priorfordelingerne fra (6.11) og (6.15) givet ved

$$\pi\left(\rho\right) = \frac{1}{\lambda_{\max}^{-1}-\lambda_{\min}^{-1}}$$
$$\pi\left(\boldsymbol{\beta},\sigma^2\right) = \frac{b^a}{(2\pi)^{k/2}|\Sigma|^{1/2}\Gamma\left(a\right)}\left(\sigma^2\right)^{-(a+(k/2)+1)} \exp\left\{-\frac{1}{2\sigma^2}\left(\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)+2b\right)\right\},$$

fås posteriorfordelingen til

$$p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right) \propto \left(2\pi\sigma^2\right)^{-\frac{n}{2}} |\rho W|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sigma^{-2}\left(\mathbf{y}-\left(I-\rho W\right)^{-1}X\boldsymbol{\beta}\right)^T \rho W\left(\mathbf{y}-\left(I-\rho W\right)^{-1}X\boldsymbol{\beta}\right)\right)$$

$$\cdot \frac{b^a}{(2\pi)^{k/2}|\Sigma|^{1/2}\Gamma(a)}\left(\sigma^2\right)^{-(a+(k/2)+1)} \exp\left\{-\frac{1}{2\sigma^2}\left(\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)+2b\right)\right\}$$

$$\cdot \frac{1}{\lambda_{\max}^{-1}-\lambda_{\min}^{-1}}$$

$$\propto \left(\sigma^2\right)^{-\left(a+(k/2)+1+\frac{n}{2}\right)} |\rho W|^{\frac{1}{2}} \exp\left(-\frac{1}{2}\sigma^{-2}\left(\left(\mathbf{y}-\left(I-\rho W\right)^{-1}X\boldsymbol{\beta}\right)^T \rho W\right.\right.$$

$$\left.\left.\cdot\left(\mathbf{y}-\left(I-\rho W\right)^{-1}X\boldsymbol{\beta}\right)+\left(\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)+2b\right)\right)\right)$$

Her er alle konstanter udeladt ved det andet proportionalitetstegn, inklusive priorfordelingen $\pi\left(\rho\right)$. ■

MCMC for CAR-modellen

Der kan nu anvendes MCMC til at estimere parametrene i CAR-modellen, hvor der simuleres fra de betingede fordelinger for hver parameter givet ved $p\left(\rho|\boldsymbol{\beta}_{(t)},\sigma^2_{(t)}\right)$, $p\left(\boldsymbol{\beta}|\rho_{(t)},\sigma^2_{(t)}\right)$ og $p\left(\sigma^2|\rho_{(t)},\boldsymbol{\beta}_{(t)}\right)$ til tiden t, hvor $\boldsymbol{\beta}_{(0)}$, $\sigma^2_{(0)}$ og $\rho_{(0)}$ er valgte begyndelsesværdier. Den betingede posteriorfordeling for parameteren ρ har formen

$$p\left(\rho|\boldsymbol{\beta},\sigma^2,\mathcal{D}\right) = \frac{p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right)}{\pi\left(\sigma^2,\boldsymbol{\beta}|\mathcal{D}\right)}$$

hvor priorfordelingen for $\boldsymbol{\beta}$ og σ^2 kendes fra (6.15), da det er antaget, parametrene i CAR-modellen har samme priorfordelinger som parametrene i SAR-modellen. Anvendes Metroolis-Hastings til at estimere parameteren ρ, er det dog ikke nødvendigt at kende den eksakte form af den betingede sandsynlighed $p\left(\rho|\boldsymbol{\beta},\sigma^2,\mathcal{D}\right)$, da Metropolis-Hastings simulere estimater fra brøken

$$\frac{p\left(\rho^*|\boldsymbol{\beta},\sigma^2,\mathcal{D}\right)}{p\left(\rho^t|\boldsymbol{\beta},\sigma^2,\mathcal{D}\right)} = \left(\frac{p\left(\boldsymbol{\beta},\sigma^2,\rho^*|\mathcal{D}\right)}{p\left(\sigma^2,\boldsymbol{\beta}|\mathcal{D}\right)}\right)\cdot\left(\frac{p\left(\boldsymbol{\beta},\sigma^2,\rho^t|\mathcal{D}\right)}{p\left(\sigma^2,\boldsymbol{\beta}|\mathcal{D}\right)}\right)^{-1} = \frac{p\left(\boldsymbol{\beta},\sigma^2,\rho^*|\mathcal{D}\right)}{p\left(\boldsymbol{\beta},\sigma^2,\rho^t|\mathcal{D}\right)}$$

fra (6.22). Parametrene $\boldsymbol{\beta}$ og σ^2 har ikke nødvendigvis nogen kendt fordeling, selvom det var tilfældet i SAR-modellen, og det kan derfor være en fordel også at anvende Metropolis-Hastings til at finde disse estimater, da det kun kræver kendskab til den samlede posteriorfordeling (6.24). Dermed kan MCMC-algoritmen for CAR-modellen opskrives i tre trin lignende dem for SAR-modellen, hvor der først vælges tre begyndelsesværdier $\boldsymbol{\beta}_{(0)}$, $\sigma^2_{(0)}$ og $\rho_{(0)}$.

Trin 1
Først simuleres vektoren for $\boldsymbol{\beta}$ ved brug af Metropolis-Hastings og den samlede posteriorfordeling $p\left(\boldsymbol{\beta},\sigma^2_{(t)},\rho_{(t)}|\mathcal{D}\right)$ fra (6.24), og den simulerede værdi $\boldsymbol{\beta}_{(t+1)}$ erstatter værdien $\boldsymbol{\beta}_{(t)}$.

Trin 2
Dernæst simuleres en ny værdi for σ^2 også ved brug af Metropolis-Hastings og den samlede posteriorfordeling $p\left(\boldsymbol{\beta}_{(t+1)},\sigma^2,\rho_{(t)}|\mathcal{D}\right)$. Værdien $\sigma^2_{(t)}$ opdateres med den simulerede værdi $\sigma^2_{(t+1)}$.

Trin 3
Tredje trin anvender Metropolis-Hastings til at simulere værdien $\rho_{(t+1)}$ fra den betingede fordeling $p\left(\rho|\boldsymbol{\beta}_{(t+1)},\sigma^2_{(t+1)}\right)$.

De tre trin i algoritmen gentages tor $t = 0,1,\ldots$, indtil estimaterne konvergerer, og de første simuleringer tages fra som burn-in.

Databehandling 7

I dette kapitel vil der blive set på, hvordan der kan estimeres henholdsvis en SAR- og CAR-model for et konkret datasæt for at forsøge at vurdere hvilken model, der passer bedst i det givne tilfælde. Til parameterestimeringen vil der både blive anvendt maximum likelihood-estimation og den bayesianske metode, for at se om de to metoder giver lignende parametre i hver af modellerne. Det anvendte datasæt er fra ejendomsmægleren HOME, der har solgt boliger i en stor del af Danmark, så datasættet kan opfattes som en stikprøve for priser i landets boligmarked. Datasættet indeholder oplysninger om 96.010 boliger, der er blevet solgt i perioden januar 2002 til og med januar 2009, hvor der til hvert salg er knyttet flere baggrundsvariable, heriblandt boligens tilstand, boligareal og om der er tale om en villa eller en lejlighed. En del observationer har dog ikke registrerede værdier for samtlige baggrundsvariable, så inden parameterestimeringen er det nødvendigt at sortere i datasættet.

7.1 Datasortering

Det oprindelige datasæt fra HOME er et Excel-regneark, der indeholder flere tastefejl og generelt inkonsistent notation, da nogle afstande er registreret i meter og andre i kilometer, og derudover er det ret forskelligt, hvilke baggrundsvariable, der er udfyldt ved hver adresse. Det oprindelige datasæt er derfor først blevet renset for tastefejl, og derefter er alle afstande konverteret til meter, mens arealer er i kvm. Datasættet indeholder i alt 44 baggrundsvariable, hvor der her vælges at fokusere på boligens beliggenhed i forhold til eksempelvis skoler, offentlig transport og indkøbsmuligheder, samt de variable, der fortæller noget om boligens type og stand, heriblandt om det er en villa eller lejlighed, boligareal og boligens tilstand. At en baggrundsvariabel er fravalgt skyldes for det meste, at der kun er noteret værdier ved relativt få af de solgte huse, eller at baggrundsvariablen ikke giver særlig meget information om boligens beliggenhed eller stand. Af fravalgte variable kan nævnes vurderingsår, salgsdato, region, udbudspris, antal plan, nyt køkken, nyt tag, garage-carport og lignende. De baggrundsvariable, der er udvalgt til den videre databehandling, er

- Postnummer
- Grundareal
- Boligareal
- Antal værelser
- Afstand til City
- Afstand til dagligvarer
- Afstand til skole
- Afstand til offentlig transport
- Afstand til strand
- Boligtilstand
- Kontantpris
- Ejendomstype

Efter at baggrundsvariablene er blevet valgt, er alle de observationer fra datasættet, hvor der mangler værdier for en eller flere af baggrundsvariablene, blevet slettet, så datasættet

består nu af 6.496 observationer. Alle salgene, der er indeholdt i datasættet fra HOME, har naturligvis tilknyttet en salgspris, og vektoren **y** i både SAR- og CAR-modellen vil derfor bestå af boligernes salgspriser. Resten af de valgte baggrundsvariable fra datasættet, udgør søjlerne i matricen X, på nær variablen ejendomstype. Da lejligheder som oftest ligger tættere på centrum af en by end villaer, og det som regel kun er villaer der har tilhørende have, som afspejles i grundarealet, kan det muligvis give en misvisende model at samle villaer og lejligheder under samme kategori. Derfor bliver datasættet nu delt i to, nemlig et med alle priser og tilhørende baggrundsvariable for lejligheder og et med samme værdier for villaer, hvor variablen „grundareal“ kun er med i kategorien villaer. Da databehandlingen består af de samme trin for begge datasæt, vil den her kun blive gennemgået for datasættet med villaer.

Baggrundsvariablen med postnumre ændres nu manuelt til at bestå af to søjler indeholdende længde- og breddegraderne for hvert postnummer, der er fundet ved brug af hjemmesiden [Wick]. Dermed er hver observeret huspris knyttet til et punkt i et koordinatsystem, så det er muligt at definere naborelationerne mellem observationerne, hvilket gøres i afsnit 7.2.

Datasættet med priser for villaer er derefter blevet konverteret fra Excel-regneark til en csv-fil og er efterfølgende blevet indlæst i programmet R, hvor koden er vist i kode 7.1.

```
1  ## Først defineres datamængden til at bestå af alle værdierne fra csv−filen med data, hvor rækken med
       variabelnavne er fjernet
2  data = read.table("villa.csv",sep=";")[-1,]
3
4  ## Længde− og breddegrader befinder sig i søjle 13 og 14, og kommaer konverteres til punktummer,
       hvorefter koordinaterne konverteres til tal og samles i en matrix med 2 søjler
5  coord = cbind(as.numeric(sub(",",".",data[,13])),
       as.numeric(sub(",",".",data[,14])))
6
7  ## Koordninaterne laves til en vektor med komplekse tal, så de kun udgør en enkelt søjle, hvilket er
       lettere at anvende i tapply−funktionen, som bruges senere
8  coordc = complex(real=coord[,1],imaginary=coord[,2])
9
10 ## Der laves en matrix laves med to søjler med koordinaterne, der har samme rækkefølge som de
       tilhørende værdier i y−vektoren og X−matricen, så de rigtige variable tilknyttes det korrekte punkt
11 coordkort = matrix(c(tapply(coord[,1],coordc,mean),
       tapply(coord[,2],coordc,mean)), ncol=2)
12
13 ## X−matricen defineres som Xkort med 10 baggrundsvariable
14 Xkort = matrix(0,dim(coordkort)[1],10)
15 for (i in 3:10){
16   Xkort[,i−2] = as.vector(tapply(as.numeric(as.matrix(data[,i])),
         coordc,mean)) # søjle 3−10 i data bliver indsat som søjle 1−8 i X
17 }
18
19 ## Boligtilstanden deles i to søjler, der udgør de sidste to søjler i X−matricen
20 tilstand = as.numeric(as.matrix(data[,11]))
21 tilstandkort = ceiling(tapply(tilstand,coordc,median))
22 tilstandkort1 = as.numeric(tilstandkort==1) # giver 1 for tilstand 1, 0 ellers
23 tilstandkort3 = as.numeric(tilstandkort==3) # giver 1 for tilstand 3, 0 ellers
24 Xkort[,9] = tilstandkort1
```

```
25 Xkort[,10] = tilstandkort3
26
27 ## y-vektoren defineres til at bestå af huspriser, hvor der er udregnet gennemsnittet for alle priser
       tilknyttet samme koordinat
28 y = as.vector(tapply(as.numeric(as.matrix(data[,12])),coordc,mean))
```

R-kode 7.1: Definition af X, **y** og matrix over koordinater for villaer.

Boligens stand oprindeligt er angivet som tre diskrete variable, hvor vurderingen god har værdien 1, middel har værdien 2 og dårlig har værdien 3. For at lave en mere præcis indikation på boligens stand laves i stedet to søjler sidst i matricen X, hvor den første søjle har et 1-tal når boligtilstanden er ét, anden søjle har et 1-tal når boligtilstanden er tre og værdierne (0,0) angiver boligtilstanden to. Det vil sige at der haves en indikatorfunktion på formen

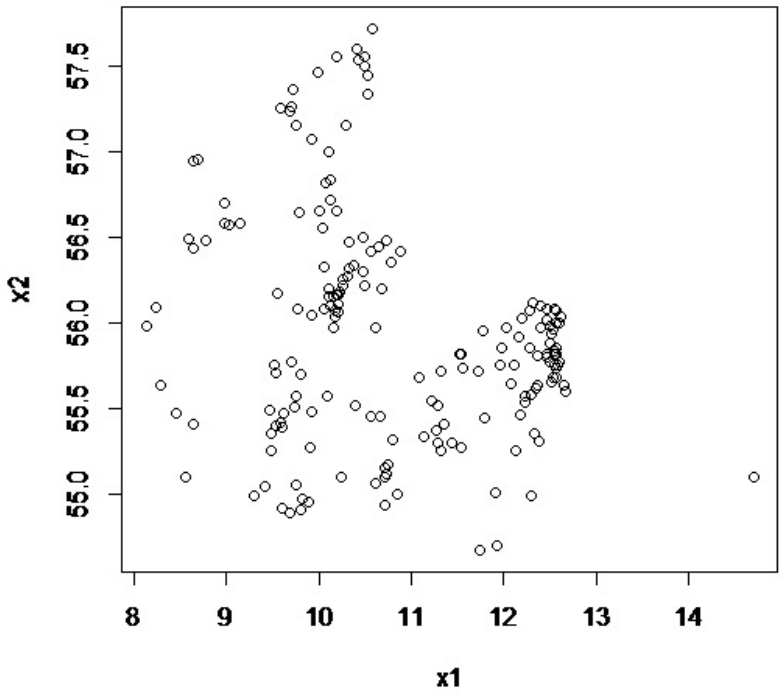
$$\gamma_1 \cdot f_1(x_i) + \gamma_2 \cdot f_2(x_i),$$

hvor $f_1(x_i) = 1$ hvis boligtilstanden er 1 og $f_2(x_i) = 1$ hvis boligtilstanden er 3. Parameteren γ_1 er dermed forskellen i pris mellem tilstand 1 og 2, mens γ_2 er forskellen mellem tilstand 2 og 3, hvor forskellen i pris mellem tilstand 1 og 2 ikke nødvendigvis er den samme som forskellen i pris mellem tilstand 2 og 3.

Derudover er der i koden 7.1 taget gennemsnittet over alle de observationer, der er tilknyttet samme koordinatsæt, så datasættet med villaer indeholder nu 180 observationer. Koden 7.1 giver dermed en vektor y med huspriser, der er vektoren **y** i både SAR- og CAR-modellen, en matrix $Xkort$ der er X-matricen og en matrix $coordkort$, der indeholder to søjler med henholdsvis længde- og breddegrader for observationerne i samme rækkefølge, som de står i vektoren **y**.

7.2 Definition af nabomatricen W

Naborelationerne mellem observationerne **y** bliver både i SAR- og CAR-modellen angivet i nabomatricen W, der kan konstrueres ved brug af et Voronoi-diagram, hvilket blev forklaret i afsnit 4.1. Det kan give et overblik over observationernes beliggenhed at afbilde dem i et koordinatsystem, inden Voronoi-diagrammet konstrueres. I koden 7.1 blev matricen $coordkort$ defineret til at indeholde observationernes koordinater, og denne matrix kan nu anvendes til at afbilde villaernes beliggenhed som vist på figur 7.1.



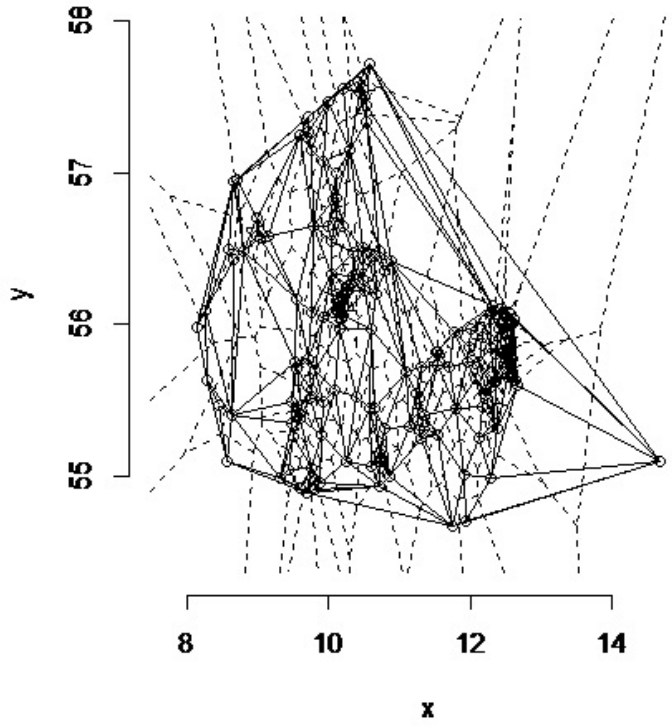
Figur 7.1: Koordinaterne til observationer for 180 solgte villaer i Danmark.

Koden 7.2 anvender længde- og breddegraderne fra matricen $coordkort$ til at lave et Voronoi-diagram over de 180 punkter, som efterfølgende anvendes til at definere (180×180)-matricen W.

```
x1 <- coord[,2] # x1 defineres som en vektor med længdegrader
x2 <- coord[,1] # x2 defineres som en vektor med breddegrader

plot(x1,x2) # Koordinaterne til huspriserne plottes

install.packages("deldir") # Der installeres pakke til at lave Voronoi-diagrammer
library(deldir)

V = deldir(x1,x2) # Der laves et Voronoi-diagram over koordinaterne
plot(V) # V plottes, hvor de stiplede linjer angiver Delaunay-trekanter og fuldtoptrukne linjer viser nabopar

nb = V$delsgs[,5:6] # Udtager informationen af V om naborelationer via indeksnumrene på nabopar, der er i 5. og 6. kolonne i V$delsgs

## Nabomatricen W defineres for de 180 observationer
W = matrix(0,180,180)
for (i in 1:dim(nb)[1]) W[nb[i,1],nb[i,2]] = 1 # 1-taller på nabopars indgange i en nedre triangulær matrix
W = W + t(W) # Inkludering af den øvre triangulære matrix
W = W/apply(W,1,sum) # W gøres rækkestokastisk
apply(W,1,sum) # Det undersøges om rækkesummerne er 1
```

R-kode 7.2: Definition af W for villaer ved brug af Voronoi-diagram.

Voronoi-diagrammet over punkterne fra figur 7.1 ses på figur 7.2.



Figur 7.2: Voronoi-diagram for 180 koordinater for solgte villaer i Danmark. De stiplede linjer er Delaunay-trekanter, mens de fuldt optrukne er naborelationer.

Det ses, at koden også har defineret naboforbindelser over Kattegat, og blandt andet har en lidt urealistisk forbindelse mellem Nordjylland og en observation på Bornholm. Det er naturligvis muligt at fjerne ulogiske forbindelser fra W manuelt, men det vil samtidig kræve en ny definition af naboer, som må tage stilling til spørgsmålet om naboforbindelser langs kyststrækninger eller langs grænsen er mere logiske end forbindelser over vandet, og derefter om hver enkelt naboforbindelse dannet af Voronoi-diagrammet beskriver det givne datasæt præcist nok. Da det er muligt at anvende metoden Voronoi-diagrammet på endnu større datasæt end de 180 observationer her og derfor må betragtes som en meget praktisk metode at kende, og det desuden vil være ret tidskrævende at definere nabopar manuelt, vil metoden blive anvendt i uredigeret form på datasættet fra HOME.

Nabomatricen W konstrueres ud fra et Voronoi-diagram, som programmet R i koden 7.2 danner ud fra matricen med koordinater $coordkort$ fra koden 7.1. Her anvendes pakken „deldir“ i R til at udtrække informationer om punkternes naborelationer, hvorefter der indsættes et 1-tal på en givet plads i matricen W, hvis observationen ud for den tilhørende søjle er nabo til observationen ud for den tilhørende række. Sidst men ikke mindst sørger koden 7.2 for, at W gøres rækkestokastisk og det tjekkes, at rækkesummerne er 1.

7.3 Maximum likelihood-estimation

Når der er defineret en rækkestokastisk nabomatrix W over naborelationerne mellem n observationer i vektoren $\mathbf{y}$, er det muligt at lave en SAR- og en CAR-model over datasættet, hvor modelparametrene kan estimeres med maximum likelihood-estimation. Først vil der blive anvendt maximum likelihood-estimation til at finde parametrene i en SAR-model for de 180 observationer med villapriser, hvorefter, der vil blive estimeret en CAR-model for samme. Der vil også blive givet et eksempel på påvirkningerne mellem observationerne, som er beskrevet i teoriafsnittet 4.2.

SAR-model

Som vist i definition 2.4 har SAR-modellen formen

$$\mathbf{y} = \rho W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\boldsymbol{\varepsilon} \sim N_n\left(\mathbf{0}, \sigma^2 I_n\right),$$

hvor $\mathbf{y}$ indeholder kontantpriserne tilknyttet 180 placeringer, for salg af villaer i Danmark, mens W er nabomatricen defineret i koden 7.2. Matricen X med baggrundsvariable, der er defineret som $Xkort$ i koden 7.1, indeholder søjler med værdier for grundareal, boligareal, antal værelser, afstand til City, afstand til dagligvarer, afstand til skole, afstand til offentlig transport, afstand til strand og to søjler med indikatorer for boligtilstanden i den nævnte rækkefølge.

Maximum likelihood-estimation for parametrene i en SAR-model for de 180 observationer for villapriser foretages med koden 7.3.

```
1  ## Først installeres pakken spdep, der indeholder SAR- og CAR-modellen
2  install.packages("spdep")
3  library(spdep)
4
5  Wl = mat2listw(W, style="W") # Nabomatricen W konverteres til brugbar form i SAR- og
       CAR-modellen
6
7  ## Estimater for SAR-modellen, hvor MC betyder der anvendes Monte Carlo-approksimation til
       estimering af logdeterminanten
8  sarmodel<-spautolm(y~Xkort[,1] + Xkort[,2] + Xkort[,3] + Xkort[,4] +
       Xkort[,5] + Xkort[,6] + Xkort[,7] + Xkort[,8] + Xkort[,9] +
       Xkort[,10], listw=Wl, family = "SAR", method="MC")
9  summary(sarmodel)
10
11 ## Estimater for CAR-modellen
12 carmodel<-spautolm(y~Xkort[,1] + Xkort[,2] + Xkort[,3] + Xkort[,4] +
       Xkort[,5] + Xkort[,6] + Xkort[,7] + Xkort[,8] + Xkort[,9] +
       Xkort[,10], listw=Wl, family = "CAR", method="MC")
13 summary(carmodel)
```

R-kode 7.3: Maximum likelihood-estimation for parametrene i SAR- og CAR-modellen.

Her anvendes pakken „spdep“ i R, hvor funktionen „spautolm“ kan anvendes til at finde både SAR- og CAR-modellen. I koden 7.3 er der anvendt Monte Carlo-approksimation

af logdeterminanten, selvom der som oftest ikke er de store problemer med at udregne determinanten, når antallet af observationer $n = 180$ er relativt lille. Outputtet fra koden er vist på figur 7.3.

```
Call: spautolm(formula = y ~ Xkort[, 1] + Xkort[, 2] + Xkort[, 3] +
    Xkort[, 4] + Xkort[, 5] + Xkort[, 6] + Xkort[, 7] + Xkort[,
    8] + Xkort[, 9] + Xkort[, 10], listw = Wl, family = "SAR",
    method = "MC")

Residuals:
     Min       1Q   Median       3Q      Max
-1873970  -653998  -124351   475726  3875265

Coefficients:
              Estimate  Std. Error z value  Pr(>|z|)
(Intercept)  9.9202e+05  3.9555e+05  2.5079 0.0121439
Xkort[, 1]   1.5780e+01  4.1928e+00  3.7635 0.0001676
Xkort[, 2]   6.5115e+03  2.7592e+03  2.3599 0.0182779
Xkort[, 3]   1.0472e+05  9.3695e+04  1.1177 0.2637127
Xkort[, 4]   2.2275e+01  1.2016e+01  1.8538 0.0637651
Xkort[, 5]  -1.5179e+02  1.0665e+02 -1.4233 0.1546519
Xkort[, 6]  -7.5609e+01  6.5489e+01 -1.1545 0.2482893
Xkort[, 7]  -2.3575e+02  2.1615e+02 -1.0906 0.2754303
Xkort[, 8]  -7.8550e+01  2.3585e+01 -3.3305 0.0008671
Xkort[, 9]   2.2270e+05  1.4794e+05  1.5053 0.1322445
Xkort[, 10] -1.0276e+06  3.8517e+05 -2.6678 0.0076346

Lambda: 0.54199 LR test value: 32.779 p-value: 1.0327e-08
Numerical Hessian standard error of lambda: 0.080643

Log likelihood: -2741.42
ML residual variance (sigma squared): 9.3311e+11, (sigma: 965980)
Number of observations: 180
Number of parameters estimated: 13
AIC: 5508.8
```

Figur 7.3: Output maximum likelihood-estimation for SAR-model.

Estimaterne for værdierne i parametervektoren $\boldsymbol{\beta}$ til hver søjle i X-matricen er angivet ud for $Xkort[,1]$, $Xkort[,2]$ og så videre, mens parameterestimatet for ρ her hedder „lambda“.

Hvis det vælges, at p-værdien for et estimat skal være under 0,05 for at parameteren er signifikant, kan de signifikante værdier findes ved at afprøve forskellige kombinationer af baggrundsvariable i matricen X. Dette kan gøres ved først at tage den baggrundsvariabel ud, der har størst p-værdi, hvorefter koden i 7.3 køres igen, og derefter tages den baggrundsvariabel ud, der nu har størst p-værdi, hvilket fortsættes, indtil alle de tilbageværende baggrundsvariable har p-værdi mindre end 0,05. I SAR-modellens tilfælde fås dette ved at anvende baggrundsvariablene grundareal, boligareal, afstand til dagligvarer, afstand til strand og boligtilstand, hvor den første boligtilstandsvariabel stadig er med, selvom den ikke er signifikant, da den sammen med den anden boligtilstandsvariabel viser boligens tilstand. Det nye output ses på figur 7.4.

```
Residuals:
     Min       1Q   Median       3Q      Max
-2042827  -591018  -150266   416984  3992927

Coefficients:
              Estimate  Std. Error z value  Pr(>|z|)
(Intercept)  1.1661e+06  3.5091e+05  3.3230 0.0008906
Xkort[, 1]   1.4838e+01  4.1843e+00  3.5461 0.0003910
Xkort[, 2]   8.6527e+03  1.9615e+03  4.4113 1.028e-05
Xkort[, 5]  -2.2567e+02  9.3483e+01 -2.4140 0.0157799
Xkort[, 8]  -7.9358e+01  2.3691e+01 -3.3496 0.0008092
Xkort[, 9]   2.0559e+05  1.4862e+05  1.3833 0.1665831
Xkort[, 10] -1.0226e+06  3.8488e+05 -2.6569 0.0078860

Lambda: 0.57706 LR test value: 43.537 p-value: 4.1612e-11
Numerical Hessian standard error of lambda: 0.075149

Log likelihood: -2744.441
ML residual variance (sigma squared): 9.5834e+11, (sigma: 978950)
```

Figur 7.4: Endeligt udvalgte baggrundsvariable

Estimaterne for koefficienterne indeholdt i $\boldsymbol{\beta}$-vektoren for SAR-modellen fås dermed til

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 1.166.068 \\ 14{,}84 \\ 8652{,}70 \\ -225{,}67 \\ -79{,}36 \\ 205.585 \\ -1.022.591 \end{bmatrix},$$

mens variansen σ^2 og parameteren ρ estimeres med

$$\hat{\sigma^2} = 978.950^2 = 958.343.102.500$$
$$\hat{\rho} = 0{,}5771.$$

Parameteren $\hat{\beta}_1$ er skæringen, der ikke har nogen videre praktisk fortolkning, da den er kontantprisen for en villa, når alle baggrundsvariable har værdien nul, og det giver ikke meget mening at have en pris for en villa med boligareal nul og grundareal nul.

Parameteren $\hat{\beta}_2$ tilhører første søjle fra $Xkort$-matricen, der indeholder grundarealet for hver observation. Da parameteren har en positiv værdi, vil kontantpriserne i **y** være større, des større grundarealet er.

Parameteren $\hat{\beta}_3$ hører sammen med anden søjle fra matricen $Xkort$, der indeholder boligarealet for hver observation, og da parameteren er positiv, vil prisen stige, når boligarealet stiger, ligesom ved grundarealet.

Parameteren $\hat{\beta}_4$ er tilknyttet femte søjle fra matricen $Xkort$ og angiver sammenhængen mellem afstand til dagligvarer og kontantprisen for en villa. Den estimerede værdi er negativ, hvilket svarer til, at når afstanden til dagligvarer er stor, er prisen for en villa mindre end hvis afstanden var mindre, hvilket også stemmer overens med den intuitive opfattelse af sammenhængen mellem boligpris og indkøbsmuligheder.

Parameteren $\hat{\beta}_5$ angiver sammenhængen mellem kontantpris og afstand til strand, der er ottende søjle i matricen $Xkort$. Da værdien er negativ, vil kontantprisen falde, når afstanden til strand stiger, hvilket også stemmer overens med den generelle opfattelse af, at en bolig med havudsigt er dyrere end en bolig langt fra vandet.

De sidste to parametre $\hat{\beta}_6$ og $\hat{\beta}_7$ angiver forskellen i pris forårsaget af boligens tilstand. Parameteren $\hat{\beta}_6$ er positiv, og udgør forskellen mellem en bolig med tilstand 1 og tilstand 2, hvor tilstand 1 resultere i en kontantpris, der er 205.585kr. højere end ved en bolig med tilstand 2. Tilsvarende vil prisen for en bolig med tilstand 3 være 1.022.591 kr. lavere end for en tilsvarende bolig i tilstand 2. Hvis boligens tilstand er i kategorien 2 udgår begge parametre fra modellen.

Parameteren $\hat{\rho}$ er positiv og mindre end én, så der er positiv korrelation mellem priser på villaer, der tilhører naboliggende postnumre.

Variansen for modellen virker umiddelbart som et overvældende stort tal, men ses der på spredningen $\hat{\sigma}$ = 978.950, har den en passende størrelse taget i betragting, at priserne i vektoren **y** varierer fra 325.000kr. til 7.295.000kr.

CAR-model

Den datagenererende funktion for CAR-modellen er vist i definition 3.18 og har formen

$$\mathbf{y} = \left(I - \rho W\right)^{-1} X\boldsymbol{\beta} + \left(\left(I_n - \rho W\right)^{-1}\right)^{\frac{1}{2}} \boldsymbol{\varepsilon}.$$

Som ved estimering af parametrene i SAR-modellen, indeholder **y** i CAR-modellen kontantpriserne tilknyttet 180 placeringer for salg af villaer i Danmark, mens nabomatricen W er defineret i koden 7.2. Matricen X indeholder de samme baggrundsvariable som ved SAR-modellen og er i den følgende kode givet ved matricen $Xkort$. I koden 7.3 er også vist, hvordan der kan foretages maximum likelihood-estimation for parametrene i en CAR-model for samme data som ved SAR-modellen. Der sorteres i de estimerede værdier på samme måde som før ved at udelade baggrundsvariable med p-værdi over 0,05 én for én, da signifikansen for andre variable kan ændre sig, når en variabel udelades. De resulterende baggrundsvariable for CAR-modellen er vist på figur 7.5.

```
Residuals:
     Min       1Q   Median       3Q      Max
-2181079  -511903   -17696   437350  3986210

Coefficients:
               Estimate  Std. Error z value  Pr(>|z|)
(Intercept)  9.3137e+05  3.8177e+05  2.4396 0.0147032
Xkort[, 1]   1.4058e+01  4.1116e+00  3.4191 0.0006284
Xkort[, 2]   9.9644e+03  1.9418e+03  5.1316 2.874e-07
Xkort[, 5]  -2.5185e+02  9.3109e+01 -2.7049 0.0068319
Xkort[, 7]  -4.3734e+02  2.1245e+02 -2.0585 0.0395423
Xkort[, 8]  -7.1538e+01  2.3468e+01 -3.0483 0.0023014
Xkort[, 9]   1.7152e+05  1.4584e+05  1.1761 0.2395646
Xkort[, 10] -1.3181e+06  3.7988e+05 -3.4698 0.0005208

Lambda: 0.90899 LR test value: 44.124 p-value: 3.0814e-11
Numerical Hessian standard error of lambda: 0.12164

Log likelihood: -2741.8
ML residual variance (sigma squared): 8.7805e+11, (sigma: 937050)
```

Figur 7.5: Output for maximum likelihood-estimation for CAR-modellen.

Som ved SAR-modellen er begge baggrundsvariable for boligtilstanden med, selvom den ene har en p-værdi over 0,05, da de to variable tilsammen fortæller hvilken stand, boligen er i. Det ses, at CAR-modellen har en signifikant baggrundsvariabel mere end SAR-modellen i form af syvende søjle fra $Xkort$, der indeholder afstande til offentlig transport. For CAR-modellen er estimaterne givet ved

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 931.366{,}10 \\ 14{,}06 \\ 9.964{,}40 \\ -251{,}85 \\ -437{,}34 \\ -71{,}54 \\ 171.517{,}10 \\ -1.318.130 \end{bmatrix},$$

mens variansen σ^2 og parameteren ρ har estimaterne

$$\hat{\sigma^2} = 937.050^2 = 878.062.702.500$$
$$\hat{\rho} = 0{,}9090.$$

Som ved SAR-modellen er $\hat{\beta}_1$ skæringen for modellen, mens $\hat{\beta}_2$ er parameteren til grundareal, og $\hat{\beta}_3$ er parameteren til boligarealet. Både parameteren for grundareal og for boligareal er positive, hvilket som før svarer til, at prisen stiger, hvis et af arealerne vokser. Parameteren $\hat{\beta}_4$ for afstand til dagligvarer har negativt fortegn som i SAR-modellen, så des større afstanden til dagligvarer er, des lavere er villaens pris.

Parameteren $\hat{\beta}_5$ er negativ, så des større afstanden til offentlig transport er, des lavere er prisen på villaen, hvilket er logisk, da de fleste ønsker at have offentlig transport i nærheden af boligen. Parameteren $\hat{\beta}_6$ er også negativ, så des længere der er til

stranden, des lavere er prisen for villaen. De to sidste parametre $\hat{\beta}_7$ og $\hat{\beta}_8$ repræsenterer boligtilstanden, og da de har samme fortegn som de to sidste værdier i $\hat{\boldsymbol{\beta}}$-vektoren for SAR-modellen, fortolkes de på samme vis.

Parameterestimatet $\hat{\rho}$ er positivt og mindre end én, så der er positiv korrelation mellem priserne for villaer med naboliggende postnumre. Variansen $\hat{\sigma}^2$ er desuden en smule mindre end ved SAR-modellen, hvilket kunne antyde en smule mere præcise estimater for CAR-modellen.

Modeltest

Der vil nu blive set på modeltest for både SAR- og CAR-modellen, for at se hvilken model, der ud til at beskrive det givne datasæt bedst. I outputtet fra koden 7.3, som for SAR-modellen ses på figur 7.4, og for CAR-modellen på figur 7.5, er der også værdier til brug ved modeltest. En metode er likelihood ratio-test, hvor en nulhypotese som eksempelvis kan være $H_0 : \theta_i = 0$ testes mod en alternativ hypotese $H_1 : \theta_i \neq 0$.

Det gøres ved at undersøge, hvor stor likelihoodfunktionen er i et punkt $\hat{\theta}_0$, som er den værdi, der maksimerer likelihoodfunktionen $l\left(\theta_i|y\right)$ under H_0. Forholdet mellem $l\left(\hat{\theta}_0|y\right)$ og maksimum likelihood-estimatet for samtlige, mulig værdier for θ_i er givet ved

$$LR = \frac{l\left(\hat{\theta}_0|y\right)}{l\left(\hat{\theta}_i|y\right)},$$

hvor $\hat{\theta}_i$ er den værdi for θ_i, som maksimerer likelihoodfunktionen. Det gælder her, at $0 < LR \leq 1$, som er en teststørrelse. Hvis værdien LR er tæt på 1, betyder det, at parameterestimatet $\hat{\theta}_0$ er tæt på at gøre likelihoodfunktionen maksimal, hvormed nulhypotesen kan accepteres. Hvis LR derimod er tæt på 0, er der grund til at tvivle på nulhypotesen.

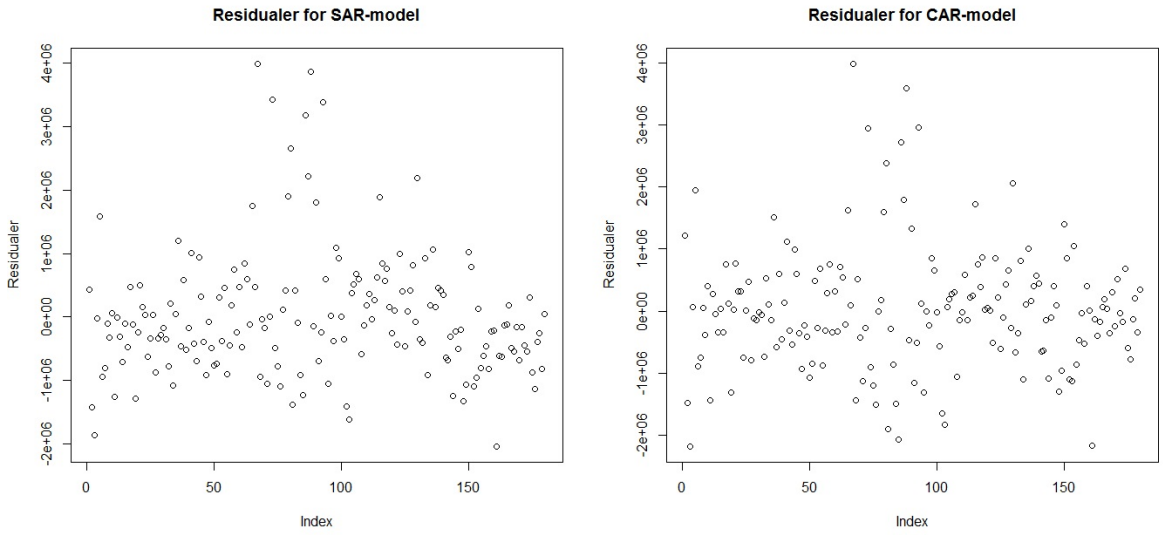
Det ses på figur 7.4 at estimaterne for SAR-modellen $LR = 43{,}54$ med p-værdien $4{,}1612 \cdot 10^{-11}$ som er tæt på nul, hvormed en nulhypotese på $\rho = 0$ kan forkastes. Det samme er tilfældet for CAR-modellen, der har $LR = 44{,}12$ med p-værdien $3{,}0814 \cdot 10^{-11}$, som er lidt tættere på nul end p-værdien for SAR-modellen.

En anden metode er at teste en model ved brug af residualplot, hvor residualerne for hver model findes med koden 7.4.

```
1 residuals(sarmodel) # Residualer findes
2 residuals(carmodel)
3
4 par(mfrow=c(1,2)) # Residualer plottes
5 plot(residuals(sarmodel),ylab="Residualer",main="Residualer for
      SAR-model")
6 plot(residuals(carmodel),ylab="Residualer",main="Residualer for
      CAR-model")
```

R-kode 7.4: Residualer for SAR- og CAR-modellerne.

Residualerne er afbildet på figur 7.6.

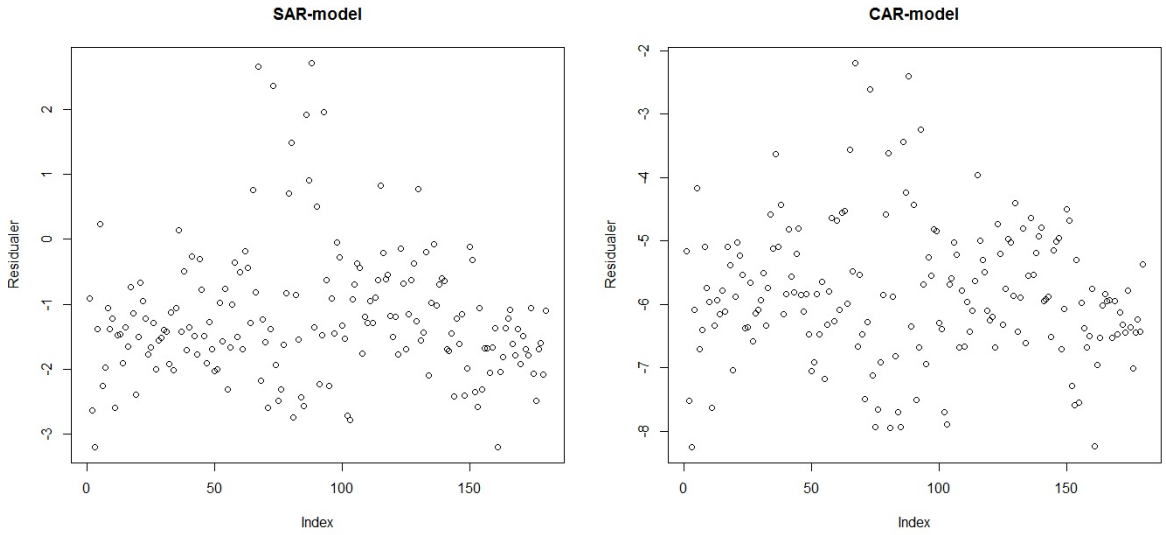


Figur 7.6: Residualer for de estimerede modeller.

For SAR-modellen er det maksimale residual givet ved 3.992.927, mens det mindste er −2.042.827, hvor medianen −150.266 er relativt tæt på nul, når spredningen $\hat{\sigma} = 978.950$ tages i betragtning. Det maksimale residual ligger dog mere end tre standardafvigelser $3 \cdot \hat{\sigma} = 2.936.850$ væk, hvilket tyder på outliers. For CAR-modellen er det mindste residual −2.181.079, mens det maksimale er 3.986.210, hvor medianen −17.696 er tættere på nul end medianen for SAR-modellen. Der er dog stadig outliers i CAR-modellen, da det maksimale residual ligger mere end tre standardafvigelser $3 \cdot \hat{\sigma} = 2.811.150$ væk.

Residualplottene spreder sig over samme område for SAR- og CAR-modellen og ligner desuden hinanden ret meget, så modellerne kan her siges at beskrive data med cirka samme grad af præcision. Begge modeller har en mindre samling punkter over resten, der muligvis kan stamme fra, at datasættet med villaer indeholder nogle enkelte nedlagte landbrug, der har et meget stort grundareal, og muligvis ikke prissættes helt som villaer generelt. Det kan overvejes at fjerne de nedlagte landbrug fra kategorien villaer for at undgå outliers, men på den anden side kan der også argumenteres for at lade dem blive, da et nedlagt landbrug i princippet er en villa beliggende på landet med et meget stort, tilhørende grundareal.

For at se lidt tydeligere, hvordan residualerne fordeler sig omkring nul, kan residualerne standardiseres ved at dividere med spredningen, hvilket giver residualplottet 7.7. De standardiserede residualer er fundet ved at indsætte de estimerede parametre for SAR-modellen i koden 7.7, der ændres en smule når residualerne for CAR-modellen findes, hvilket forklares senere.



Figur 7.7: Standardiserede residualer for de estimerede modeller.

Her ses det tydeligt, at residualerne for SAR-modellen er tættere på nul end dem for CAR-modellen, hvilket kan tyde på, at den estimerede SAR-model passer bedre på datasættet med villapriser end den estimerede CAR-model.

Påvirkninger

Det er også muligt at udregne de gennemsnitlige påvirkninger mellem observationerne, som er beskrevet i afsnit 4.2. Da det i sætning 4.1 blev vist, at SAR- og CAR-modellen har de samme partielt afledede, og dermed at de gennemsnitlige påvirkninger udregnes på samme måde for de to modeller, udregnes der her kun påvirkningerne for SAR-modellen ved brug af kommandoen „impacts(sarmodel, listw = Wl)“ i programmet R. Outputtet er vist på figur 7.8.

```
Impact measures (lag, exact):
                    Direct      Indirect         Total
Xkort[, 1]   1.637753e+01  1.651398e+01  3.289151e+01
Xkort[, 2]   9.433842e+03  9.512443e+03  1.894628e+04
Xkort[, 5]  -2.517633e+02 -2.538609e+02 -5.056242e+02
Xkort[, 8]  -8.514701e+01 -8.585643e+01 -1.710034e+02
Xkort[, 9]   1.827912e+05  1.843142e+05  3.671054e+05
Xkort[, 10] -1.204228e+06 -1.214261e+06 -2.418489e+06
```

Figur 7.8: Gennemsnitlige påvirkninger for den estimerede SAR-model.

Den gennemsnitlige, direkte påvirkning, som står ud for $Xkort[,2]$, betyder, at hvis boligarealet stiger med δ kvm for observation y_i, så vil prisen for villa y_i stige med $\delta \cdot 9.434$kr. Den gennemsnitlige, indirekte påvirkning fra samme ændring vil medfører at naboobservationen y_j til y_i får en prisstigning på $\delta \cdot 9.512$kr., så den gennemsnitlige, samlede påvirkning fra en ændring i boligarealet er en stigning huspriser på $\delta \cdot 18.946$kr.

7.4 Bayesiansk parameterestimering

I praksis er det ikke kun interessant at se på hvilken model der beskriver et datasæt mest nøjagtigt, men også hvilken metode til parameterestimering, der giver de bedste resultater. Derfor vil der nu blive set på bayesiansk estimering af parametrene i en SAR-model og en CAR-model for villapriser i Danmark. Da den eneste viden, der haves om priserne, er givet i datasættet, vælges priorfordelingerne til at være ikke-informative ved at vælge hyperparametre tæt på nul og stor varians. Dermed skulle de endelige estimater gerne ligne estimaterne fra maximum likelihood-estimation, så det er nemmere at sammenligne resultaterne fra de to estimeringsmetoder.

SAR-model

Først ses der på SAR-modellen, der findes ved brug af koden 7.5.

```
1  install.packages("gibbs.met") # Først installeres en pakke til algoritme med
       Metropolis−Hastings i Gibbs−sampling
2  library(gibbs.met)
3
4  ## Der defineres nu en funktion, som udregner loglikelihoodfunktionen for den simultane
       posteriorfordeling for SAR−modellen
5  sarbayespostford <- function(x){
6  l = length(x) # Antallet af parametre i modellen
7  beta = x[-c(l-1,l)] # Hver af modellens parametre tildeles en plads i x, der er modellens input
8  sigma = x[l-1]
9  rho = x[l]
10 if (sigma<0) return(-Inf) # Det sikres, at variansen sigma er positiv
11 if (rho<1/min(eigen(W)$values)) return(-Inf) # Her sikres at rho er mindre end den
       inverse af den mindste egenværdi for W
12 if (rho>1) return(-Inf) # Det sikres, at parameteren rho ikke er større end 1
13 ## Herefter defineres hyperparametrene i posteriorfordelingen
14 a1 = a + n/2
15 Sigma1=solve(t(X)%*%X + solve(Sigma))
16 c1=Sigma1%*%(t(X)%*%(diag(180)-rho* W)%*%y + solve(Sigma)%*%c)
17 b1= b + (t(c)%*%solve(Sigma)%*%c+ t(y)%*%t(diag(180)-rho*
       W)%*%(diag(180)-rho* W)%*%y-t(c1)%*%solve(Sigma1)%*%c1)/2
18 ## Posteriorfordelingen for SAR−modellen indsættes
19 return(-(a1 + k/2 + 1)*log(sigma^(2)) + log(det(diag(180)-rho* W)) -
       (1/(2*sigma^2)*(t(beta - c1)%*%solve(Sigma1)%*%(beta-c1)+2*b1)))
20 }
21
22 ## Nu vælges begyndelsesværdier for hyperparametrene
23 a=0.01
24 b=0.01
25 c=c(0,0,0,0,0,0,0,0,0,0,0)
26 Sigma= 10^10*diag(11)
27
28 ## X−matricen, vektoren y og dimensionerne heraf indsættes
29 n=180 # Antal postnumre
30 k=11 # antallet af parametre i betavektoren
31 y=y # prisen for boliger
32 X<-matrix(0,180,11)# Der defineres en ny X−matrix, så skæringen også kan estimeres
33 X[,1:10]<-Xkort
```

```
34 X[,11]<-1
35
36 ## Der simuleres fra posteriorfordelingen
37 sim = gibbs_met(sarbayespostford, no_var=13, ini_value=c( 92, 2, 105,
       16, -33, -35, 154, -18,-100, 14,1,460556,0.5),iters=500,iters_met=2,
       stepsizes_met=c(100,1,10,1,10,5,8,5,10,10,10,2500,0.05))
38 # hvor no_var er antallet af komponenter i Gibbs-sampleren, ini_value er vektor med
       begyndelsesværdier, iters er antallet af iterationer og stepsizes_met er standardafvigelserne i
       forslagsfordelingerne.
```

R-kode 7.5: Bayesiansk estimering af parametrene i SAR-modellen.

Der simuleres fra posteriorfordelingen for SAR-modellen, som er givet i sætning 6.2. Via kommandoen „gibbs.met“ i R anvendes både Metropolis-Hastings og Gibbs-sampling til at finde modelestimaterne, da $\boldsymbol{\beta}$ og σ^2 har kendte fordelinger, mens fordelingen for ρ ikke har nogen velkendt form, hvilket blev vist i afsnit 6.4. Først foretages 500 iterationer ved brug af koden 7.5 for at finde ud af hvilke baggrundsvariable i matricen X, der er signifikante. Signifikansen undersøges ved at lave 95%-CPI for hver parameter, hvilket er vist i koden 7.6.

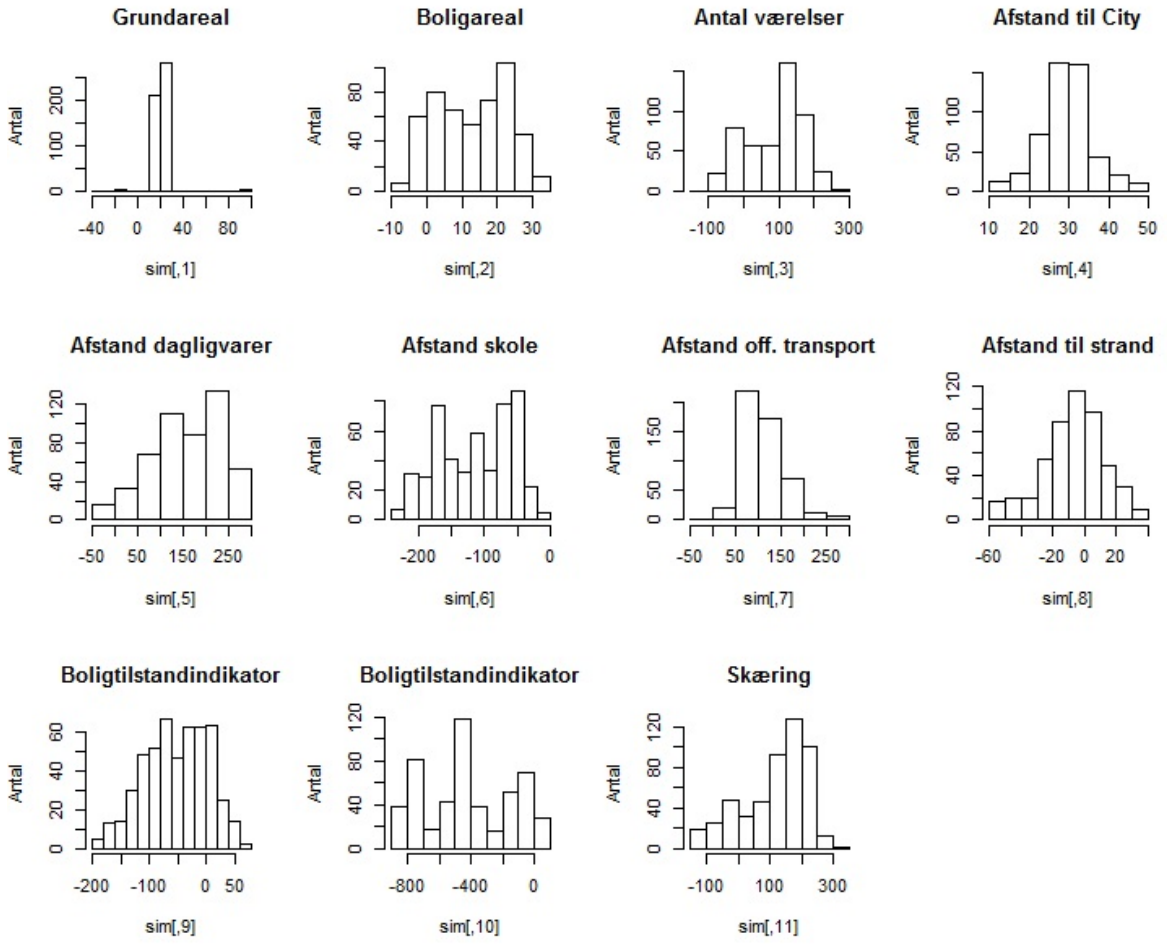
```
1 ## 95% CPI til at finde signifikante parameterværdier
2 simm1<-sort(sim[,1]) # Først sorteres de simulerede værdier for den første indgang i
       beta-vektoren
3 simm1[floor(500*0.025)] # 2.5%-fraktilen findes for de 500 simuleringer
4 simm1[floor(500*0.975)] # 97.5%-fraktilen for 500 simuleringer. Samme kode gennemløbes for
       hver af de 11 parametre
5
6 ## Histogrammet for simuleringerne for hver parameter
7 hist(sim[,1] , xlab="sim[,1]",ylab="Antal",main="Grundareal")
8 hist(sim[,2] , xlab="sim[,2]",ylab="Antal",main="Boligareal") # Samme kode
       gentages for resten af parametrene ved at ændre sim[,2] til sim[,3] osv.
```

R-kode 7.6: CPI og histogrammer for SAR-modellen.

Der fås følgende intervaller for 95%-CPI, det vil sige, at der er 95% sandsynlighed for den ægte værdi for hver parameter befinder sig i det givne interval:

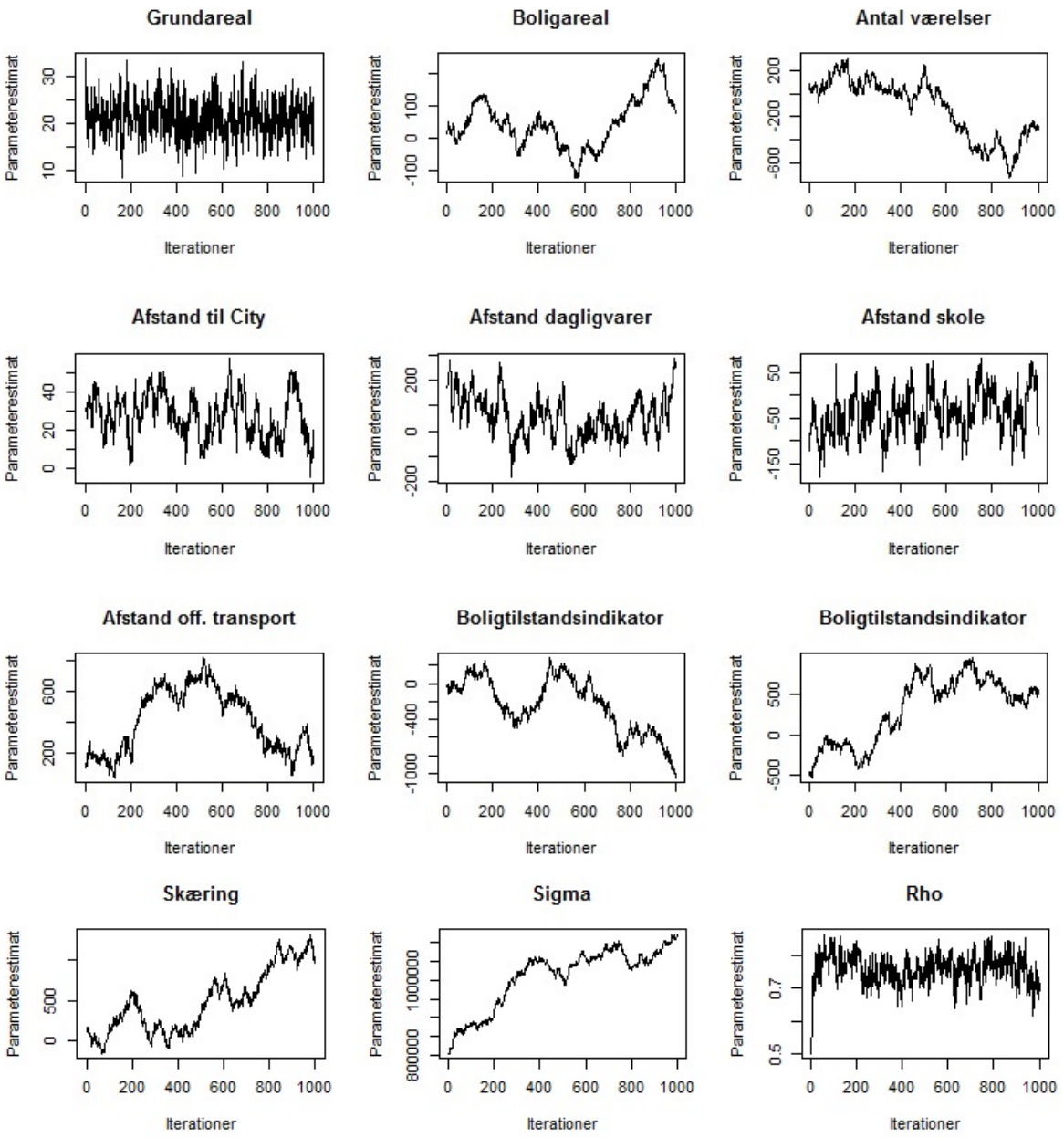
$$\beta_1 \in [15{,}97; 26{,}81] \text{ med 95\% sikkerhed}$$
$$\beta_2 \in [-4{,}31; 29{,}65] \text{ med 95\% sikkerhed}$$
$$\beta_3 \in [-64{,}64; 211{,}48] \text{ med 95\% sikkerhed}$$
$$\beta_4 \in [14{,}57; 43{,}92] \text{ med 95\% sikkerhed}$$
$$\beta_5 \in [-5{,}46; 264{,}70] \text{ med 95\% sikkerhed}$$
$$\beta_6 \in [-213{,}80; -32{,}48] \text{ med 95\% sikkerhed}$$
$$\beta_7 \in [37{,}18; 214{,}29] \text{ med 95\% sikkerhed}$$
$$\beta_8 \in [-51{,}83; 25{,}76] \text{ med 95\% sikkerhed}$$
$$\beta_9 \in [-165{,}08; 41{,}15] \text{ med 95\% sikkerhed}$$
$$\beta_{10} \in [-824{,}14; 45{,}08] \text{ med 95\% sikkerhed}$$
$$\beta_{11} \in [-111{,}03; 248{,}84] \text{ med 95\% sikkerhed}$$

Desuden fås histogrammerne som vist på figur 7.9.



Figur 7.9: Histogrammer over 500 simuleringer for parametrene i SAR-modellen.

Ses der på intervallerne for CPI, indeholder intervallerne for β_2, β_3, β_5, β_8, β_{10} og β_{11} værdien nul, men ses der på histogrammerne, er det kun parameteren β_8 for afstand til strand, der med sikkerhed er centreret omkring nul. Derfor sorteres denne parameter fra som værende ikke signifikant, hvorefter koden 7.5 køres igen for de signifikante baggrundsvariable, denne gang med 1000 iterationer. Traceplottene er vist på figur 7.10.



Figur 7.10: Traceplot for de simulerede parametre til SAR-modellen.

De bayesianske parameterestimater for SAR-modellen udregnes ved brug af koden 7.7.

```
1  ## Estimatet for hver simuleret parameter udregnes som gennemsnittet af de simulerede værdier minus
       burn-in
2  sarbeta1<-mean(sim[100:1000,1])
3  sarbeta2<-mean(sim[100:1000,2]) #og på samme måde estimeres resten af de 11 værdier i
       beta-vektoren
4
5  sarsigma2<-(mean(sim[300:1000,11]))^2 # Variansen estimeres
6  sarrho<-mean(sim[100:1000,12]) # Parameteren rho estimeres
7
8  estimaterbeta <-c(sarbeta1, sarbeta2, sarbeta3, sarbeta4, sarbeta5,
       sarbeta6, sarbeta7, sarbeta8, sarbeta9, sarbeta10) # Parametervektoren beta
```

```
9
10 ## Residualerne plottes
11 plot(((diag(180)-sarrho*W)%*%(y -
       (solve(diag(180)-sarrho*W)%*%X%*%estimaterbeta)))/sqrt(sarsigma2),
       xlab="Index",ylab="Residualer")
```

R-kode 7.7: Bayesianske estimater og tilhørende residualer for SAR-modellen.

Ved hvert parameterestimat er der forsøgt at tage højde for burn-in ved at udelade de første 100 til 300 iterationer, alt efter hvornår traceplottet for hvert enkelt estimat ser ud til at nærme sig en ligevægt. Dermed fås parameterestimaterne

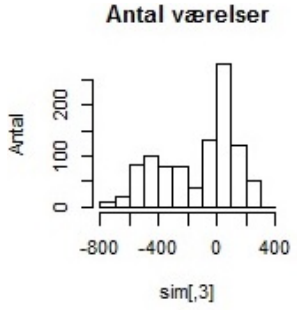
$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 21{,}15 \\ 49{,}08 \\ -147{,}86 \\ 27{,}06 \\ 46{,}15 \\ -34{,}08 \\ 468{,}36 \\ -310{,}89 \\ 538{,}91 \\ 534{,}99 \end{bmatrix} \tag{7.1}$$

$$\hat{\sigma^2} = 1.055.512^2 = 1.114.105.582.144$$

$$\hat{\rho} = 0{,}7576.$$

Første indgang i vektoren $\hat{\boldsymbol{\beta}}$ er parameteren tilhørende baggrundsvariablene for grundarealer, mens anden indgang i $\hat{\boldsymbol{\beta}}$ er boligarealet. Begge parameterværdier er positive og fortolkes derfor ligesom ved maximum likelihood-estimation.

Parameteren $\hat{\beta}_3$ tilhører baggrundsvariablen med antal værelser, der her er negativ, hvilket svarer til, at des flere værelser en bolig har, des billigere er den. Dette virker ikke helt logisk og ses der på histogrammet for de 1000 iterationer for denne parameter, der er vist på figur 7.11, ses det også, at mange værdier er centreret omkring nul. Det kan derfor overvejes at udelade variablen.



Figur 7.11: Histogram for 1000 simuleringer af parameteren for antal værelser.

Parameteren $\hat{\beta}_4$ hører sammen med baggrundsvariablen med afstand til City, der her er positiv, så des længere der er til centrum, des dyrere er boligen. Det kan muligvis skyldes,

at hvis villaerne ligger lidt uden for byen, kan de have større grundareal end villaer tættere på bymidten, og dermed også en højere pris.

Parameteren $\hat{\beta}_5$ tilhører variablen med afstand dagligvarer, der også er positiv, så des længere der er til indkøbsmuligheder, des dyrere er boligen. Intuitivt er det det modsatte af hvad man kunne forvente, og ses der på traceplottet, er de simulerede værdier ret centreret om nul, så det kan også overvejes at tage denne parameter ud af modellen.

Parameteren $\hat{\beta}_6$ hører til baggrundsvariablen afstand skole, og er negativ, så des længere der er til en skole, des lavere er villaens pris.

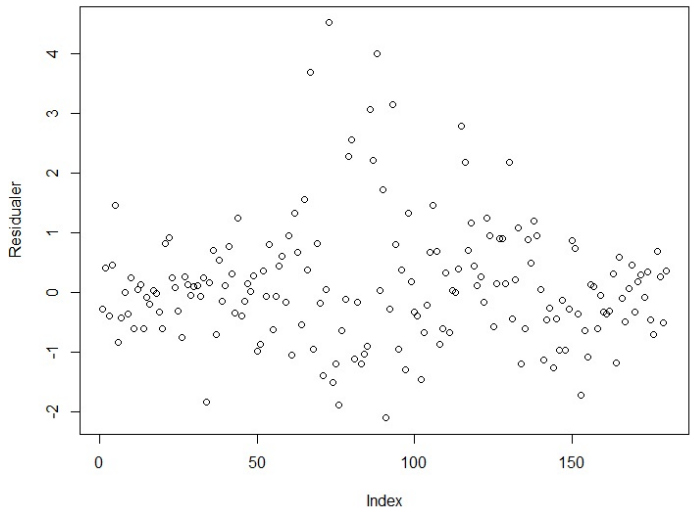
Parameteren $\hat{\beta}_7$ tilhører afstanden til offentlig transport, der her er positiv, så des længere der er til transport, des dyrere er villaen, hvilket ligesom afstanden til centrum kan virke en smule ulogisk. Det kan muligvis forklares med, at de lidt større og dermed dyrere villaer kan ligge lidt uden for byen og dermed også længere fra transport, dagligvarer og centrum.

De to parametre $\hat{\beta}_8$ og $\hat{\beta}_9$ hører til boligtilstanden, der ville give mere mening, hvis fortegnene for boligtilstandene have været omvendt, da den negative parameter $\hat{\beta}_8$ og den positive parameter $\hat{\beta}_9$ antyder, at prisen falder, når boligtilstanden er god, og at prisen stiger, når boligstanden er dårlig. Det kan skyldes, at den ene af boligtilstandsindikatorerne jævnfør histogrammerne 7.9 muligvis ikke er signifikant, hvilket kunne medfører, at ingen af boligtilstandsindikatorerne i sidste ende er signifikante, da de kan anses for at høre sammen. Derfor kunne det overvejes at tage dem ud af modellen. Sidste parameter $\hat{\beta}_10$ er skæringen.

Modellen kan testes ved at se på residualerne, der her er de standardiserede residualer, der er forskellen mellem de observerede værdier $\mathbf{y}$ og den estimerede model, hvor forskellen divideres med spredningen σ. For SAR-modellen fra sætning 2.6 fås de standardiserede residualer $\mathbf{r}$ ved

$$\begin{aligned}
\mathbf{y} &= \left(I_n - \rho W\right)^{-1}\left(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}\right) \Rightarrow \\
\boldsymbol{\varepsilon} &= \left(I_n - \rho W\right)\left(\mathbf{y} - \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}\right) \Rightarrow \\
\mathbf{r} &= \frac{\left(I_n - \hat{\rho} W\right)\left(\mathbf{y} - \left(I_n - \hat{\rho} W\right)^{-1} X\hat{\boldsymbol{\beta}}\right)}{\hat{\sigma}},
\end{aligned}$$

hvor $\boldsymbol{\varepsilon} \sim N_n\left(\mathbf{0}, \sigma^2 I_n\right)$, $\hat{\rho}$ er den estimerede værdi for ρ og $\hat{\boldsymbol{\beta}}$ er de estimerede værdier for parametervektoren $\boldsymbol{\beta}$. Residualerne $\mathbf{r}$ afbildes ved brug af koden 7.7 og er vist på figur 7.12.



Figur 7.12: Residualer for bayesianske estimater for SAR-modellen.

Det maksimale residual 4,51 og det mindste residual −2,10, hvilket ikke er særlig store værdier, mens medianen 0,0310 for residualerne er tæt på nul, så modellen kan siges at passe ret godt på data. Det ses desuden, at residualerne ligger tættere på nul end for de to modeller estimeret med maximum likelihood-estimation, der er vist på figur 7.7.

CAR-model

Fremgangsmåden til at finde modelparametrene i CAR-modellen ved brug af bayesiansk parameterestimering er den samme som for SAR-modellen. Posteriorfordelingen for CAR-modellen er vist i sætning 6.4, og koden til at simulere fra posteriorfordelingen er givet i koden 7.8.

```
carbayespostford <- function(x){
l = length(x) # Antal af parametre
beta = x[-c(l-1,l)] # Hver parameter tildeles en plads i x
sigma = x[l-1]
rho = x[l]
if (sigma<0) return(-Inf) # Det sikres, at variansen sigma er positiv
if (rho<1/min(eigen(W)$values)) return(-Inf) # Her sikres at rho er mindre end den
    inverse af den mindste egenværdi for W
if (rho>1) return(-Inf) # Her sikres at rho ikke er større end 1
a1 = a + n/2 # Hyperparametre i posteriorfordelingen defineres
## Logaritmen for posteriorfordelingen defineres
return(-(a1 + k/2 + 1)*log(sigma^(2)) + 0.5*log(det(rho* W)) -
    (1/(2*sigma^2) * (t(y-solve(diag(180)-rho* W)%*%X%*%beta)
    %*%(rho*W)%*%(y-solve(diag(180)-rho* W)%*%X%*%beta) +
    (t(beta-c)%*%solve(Sigma)%*%(beta-c)+2*b))))
}
a=0.01 # selvvalgte værdier for hyperparametre
b=0.01
c=c(0,0,0,0,0,0,0,0,0,0,0)
n=180 # antal postnumre
Sigma= 10^10*diag(11)
```

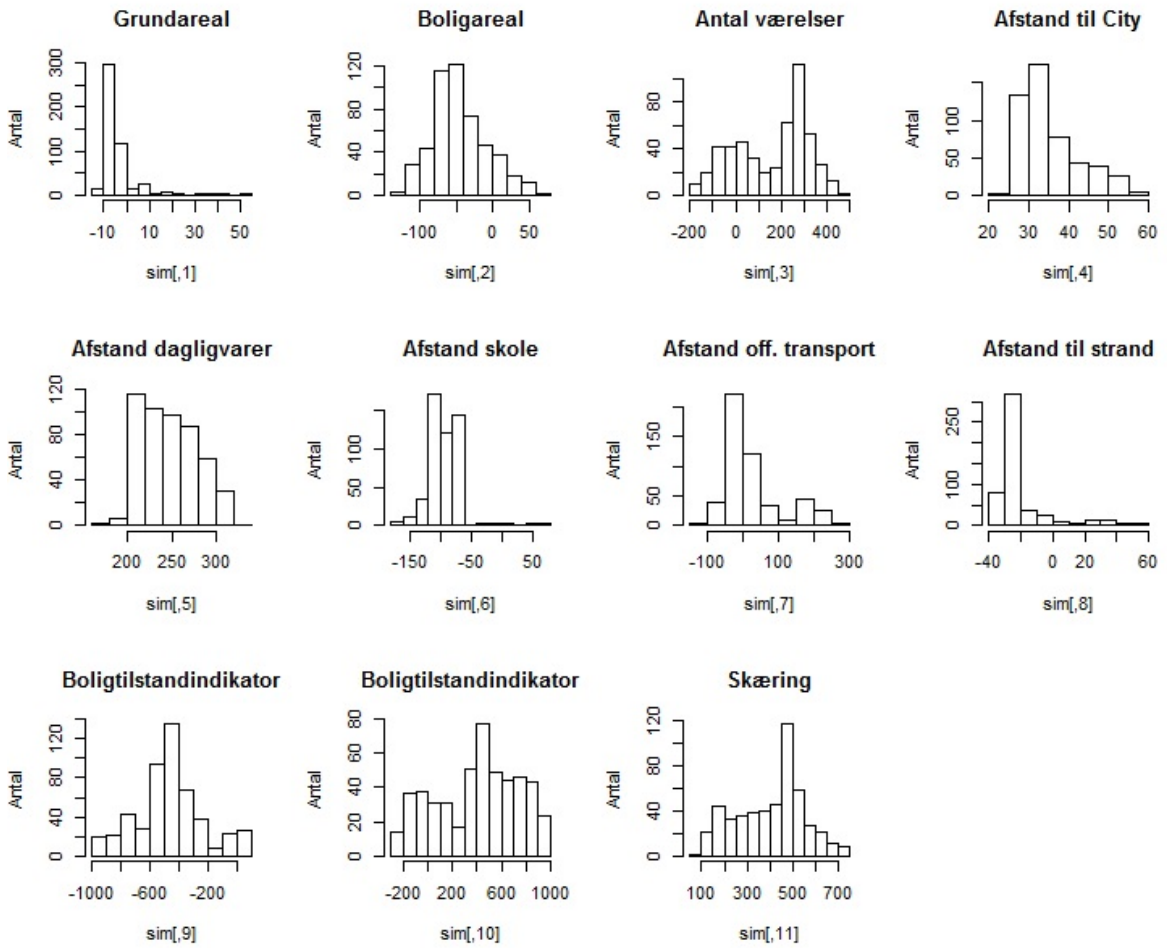
```
k=11 # antallet af parametre i betavektoren
y=y # prisen for boliger
## Laver ny x-matrix med skæring som sidste indgang
X<-matrix(0,180,11)
X[,1:10]<-Xkort
X[,11]<-1
X[1:5,]

## Der simuleres fra posteriorfordelingen
sim = gibbs_met(carbayespostford, no_var=13, ini_value=c( 92, 2, 105,
   16, 1, 1, 154, 1,1, 14,460556,1,0.5),iters=500,iters_met=2,
   stepsizes_met=c(100,1,10,1,10,5,8,5,10,10,10,2500,0.05))
```

R-kode 7.8: Simulering af estimater for CAR-modellen fra den bayesianske posteriorfordeling.

Her indsættes også en ekstra søjle i X-matricen af ettaller, så modellen får et skæringspunkt. Først simuleres 500 parameterværdier, hvor 95%-CPI for parametrene i CAR-modellen er givet ved

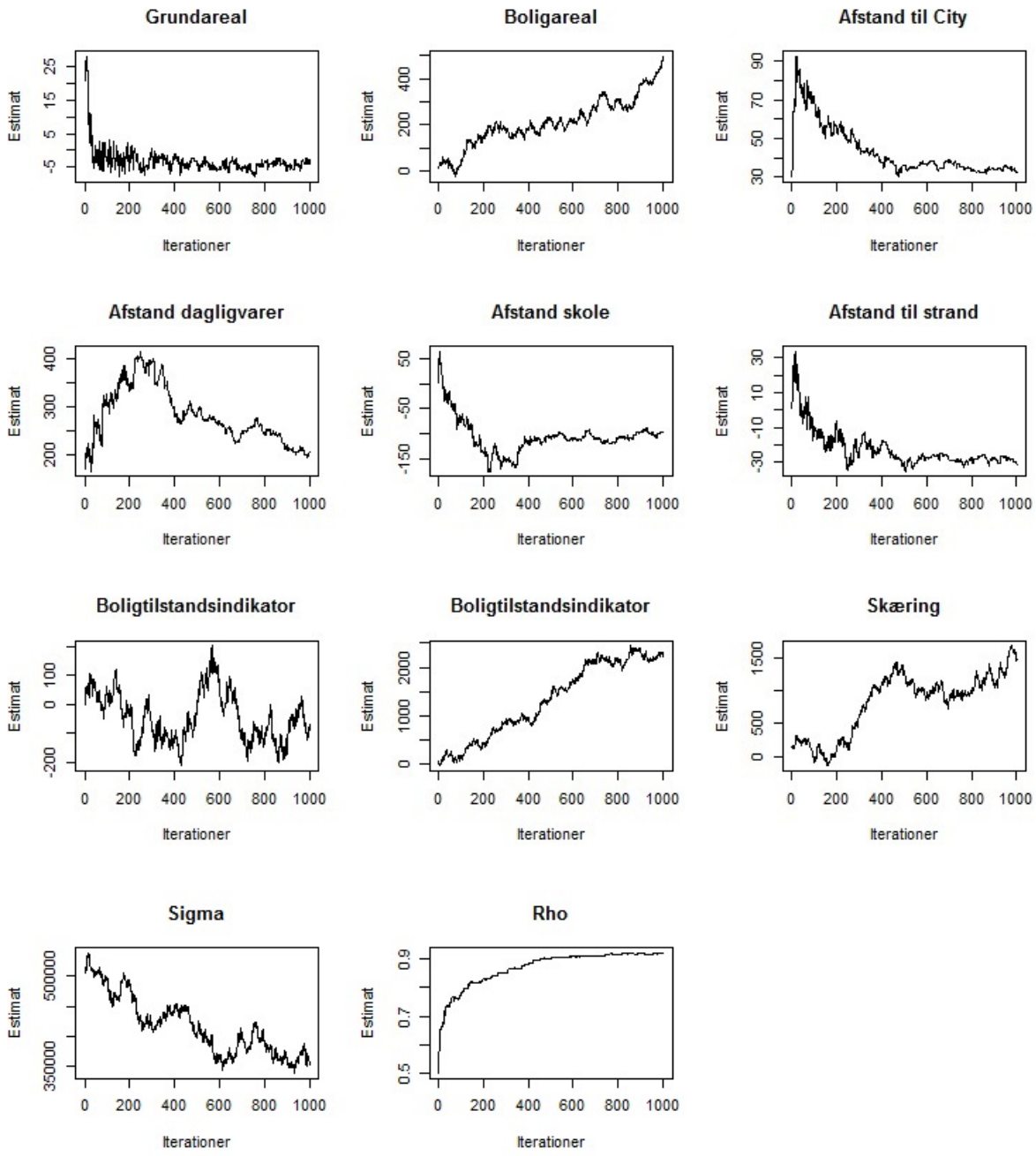
$$\beta_1 \in [-10{,}21355; 22{,}94238] \text{ med 95\% sikkerhed}$$
$$\beta_2 \in [-114{,}55462; 39{,}92595] \text{ med 95\% sikkerhed}$$
$$\beta_3 \in [-142{,}63065; 395{,}06887] \text{ med 95\% sikkerhed}$$
$$\beta_4 \in [27{,}83265; 51{,}82665] \text{ med 95\% sikkerhed}$$
$$\beta_5 \in [206{,}11799; 307{,}44329] \text{ med 95\% sikkerhed}$$
$$\beta_6 \in [-152{,}14204; -41{,}73992] \text{ med 95\% sikkerhed}$$
$$\beta_7 \in [-67{,}04206; 223{,}22883] \text{ med 95\% sikkerhed}$$
$$\beta_8 \in [-33{,}10241; 36{,}66470] \text{ med 95\% sikkerhed}$$
$$\beta_9 \in [-938{,}49244; 41{,}81791] \text{ med 95\% sikkerhed}$$
$$\beta_{10} \in [-203{,}43978; 929{,}79038] \text{ med 95\% sikkerhed}$$
$$\beta_{11} \in [132{,}13896; 680{,}03412] \text{ med 95\% sikkerhed.}$$

Intervallerne er fundet ud fra samme type kode som i koden 7.6, og histogrammerne for de simulerede parametre er vist på figur 7.13.



Figur 7.13: Histogrammer for de simulerede parametre i CAR-modellen.

Ses der på CPI, indeholder intervallerne for parametrene β_1, β_2, β_3, β_7, β_8, β_9 og β_{10} værdien nul, mens det kun er histogrammet for β_7, der er ret centreret om værdien nul, så denne parameter tages ud af modellen, inden der simuleres nye estimater for parametrene. Derudover indeholder histogrammet for β_3 en del estimater omkring værdien nul, så denne parameter tages også ud af modellen, mens resten af parametrene beholdes. Det vil sige, der defineres en ny X-matrix, hvorefter der foretages 1000 iterationer for parametrene i CAR-modellen. Traceplottene er vist på figur 7.14.



Figur 7.14: Traceplot for parameterestimater for CAR-modellen for villaer.

At flere traceplot ser ud til ikke at konvergere inden for de første 1000 iterationer, kan skyldes at MCMC-estimering her kræver positive begyndelsesværdier og det derfor tager negative parametre flere iterationer at nå ligevægten. Hvis der køres flere iterationer end 1000 kan ligevægten sandsynligvis ses tydeligere end her.

Parameterestimaterne udregnes ved brug af kode lig den for SAR-modellen i kode 7.7, hvor de første 100 til 300 simuleringer tages fra som burn in, vurderet ud fra traceplottene.

Parameterestimaterne for CAR-modellen bliver

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} -3{,}83 \\ 248{,}24 \\ 36{,}21 \\ 286{,}92 \\ -114{,}40 \\ -25{,}18 \\ -51{,}91 \\ 1.733{,}62 \\ 1.096{,}16 \end{bmatrix} \begin{matrix} \text{Grundareal} \\ \text{Boligareal} \\ \text{Afstand til City} \\ \text{Afstand til dagligvarer} \\ \text{Afstand til skole} \\ \text{Afstand til strand} \\ \text{Boligtilstandsindikator} \\ \text{Boligtilstandsindikator} \\ \text{Skæring} \end{matrix}$$

$$\hat{\sigma}^2 = 392.926^2 = 154.390.528.905$$

$$\hat{\rho} = 0{,}9053.$$

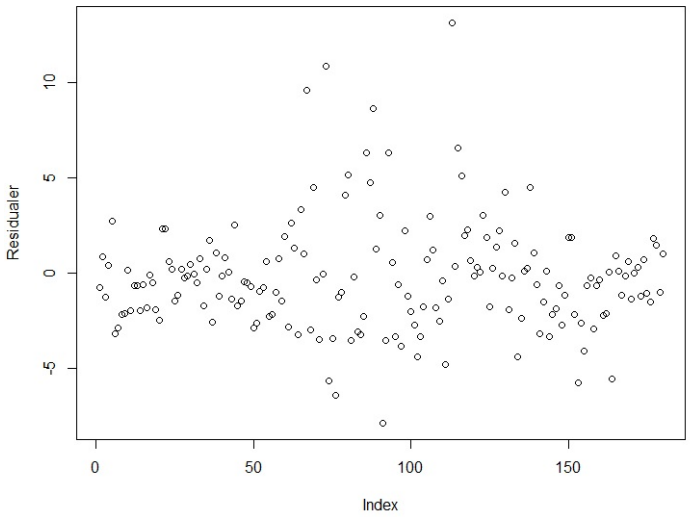
Her ses det at parameteren for grundarealet er negativ og tæt på nul, hvilket ikke stemmer overens med de tidligere fortolkninger af grundarealets sammenhæng med boligens pris. Det kunne muligvis skyldes, at der kræves endnu flere iterationer før ligevægten for grundareal nås, eller at der er medtaget baggrundsvariable, der ikke er signifikante, og dermed påvirker andre parameterværdier uheldigt. Parametrene for afstand til City, dagligvarer og skole, samt parametrene for boligstanden har samme fortegn som estimaterne for parametrene i SAR-modellen fra (7.1) og kan derfor fortolkes på samme vis. Parameteren for baggrundsvariablene med afstand til strand er her negativ, hvilket ikke er hvad man ville forvente intuitivt. Som nævnt tidligere kan det skyldes, at parameterestimaterne kræver flere end 1000 simuleringer for at nå en ligevægt, da startværdierne i flere tilfælde ligger relativt langt fra de parameterværdier, der er opnået ved eksempelvis maximum likelihood-estimation.

De estimerede værdier for CAR-modellen kan testes ved at se på de standardiserede residualer **r**, der findes ved at forskellen mellem dataene **y** og den estimerede model divideres med spredningen $\hat{\sigma}$. Den datagenererende funktion for CAR-modellen har jævnfør sætning 3.18 formen

$$\mathbf{y} = \left(I - \rho W\right)^{-1} X\boldsymbol{\beta} + \left(\left(I_n - \rho W\right)^{-1}\right)^{\frac{1}{2}} \boldsymbol{\varepsilon} \Rightarrow$$

$$\mathbf{r} = \frac{\left(I_n - \hat{\rho} W\right)^{\frac{1}{2}} \left(\mathbf{y} - \left(I - \hat{\rho} W\right)^{-1} \hat{X}\hat{\boldsymbol{\beta}}\right)}{\hat{\sigma}}.$$

Plottes de standardiserede residualer, fås outputtet på figur 7.15.



Figur 7.15: Residualer for bayesianske estimater til CAR-modellen.

Det maksimale residual for CAR-modellen er 13,14, mens det mindste er −7,85 og medianen −0,4908 er tæt på værdien nul. Residualerne er jævnt fordelt omkring værdien nul og tyder på en mere præcis model for de givne data end de standardiserede residualer for parametrene fra maksimum likelihood-estimation, der er vist på figur 7.7, mens den bayesianske SAR-model har residualer en smule tættere på nul end den bayesianske CAR-model.

Det givne datasæt med salgspriser og tilhørende baggrundsvariable for villaer i Danmark kan med udgangspunkt i residualplottene derfor beskrives mest præcist med en SAR-model med estimater fra bayesiansk MCMC-estimering, hvorimod fortegnene for parametrene intuitivt giver mest mening for både SAR-og CAR-modellen med maximum likelihood-estimaterne. Der kan givetvis opnås andre resultater i begge estimeringsmetoder, hvis der medtages andre kombinationer af baggrundsvariable, hvilket er oplagt at forsøge sig med ved de bayesianske estimater, hvor simuleringerne dog tog for lang tid til, at det var muligt at afprøve flere kombinationer i forbindelse med denne rapport. Udregningerne kunne muligvis gøres hurtigere ved at approksimere nogle af udtrykkene med store matricer i posteriorfordelingerne for både SAR- og CAR-modellen, så det var muligt at foretage endnu flere iterationer end 1000, hvilket sandsynligvis ville resultere i mere præcise estimater.

Konklusion 8

Rumligt data med indbyrdes afhængige punkter som eksempelvis huspriser kan beskrives ved brug af rumlige modeller, heriblandt også rumlige, økonometriske modeller. To af denne slags modeller er SAR- og CAR-modellen.

SAR-modellen er en rumlig, autoregressiv model på formen

$$\begin{aligned}\mathbf{y} &= \rho W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon} &\sim N_n\left(\mathbf{0},\sigma^2 I_n\right),\end{aligned}$$

der er defineret i definition 2.4. Her er ρ er en parameter for graden af rumlig afhængighed mellem de n observationer i vektoren $\mathbf{y}$, mens matricen W angiver naborelationerne mellem observationerne. Matricen X indeholder observationernes baggrundsvariable, og $\boldsymbol{\beta}$ er den tilhørende parametervektor, mens $\boldsymbol{\varepsilon}$ er fejlvektoren. SAR-modellen er en model for rumligt, afhængige data, der følger en normalfordeling, hvor det generelt antages, at når der haves mange observationer n i et datasæt, vil fordelingen for observationerne tilnærmelsesvis være en normalfordeling.

En anden model, der kan anvendes på rumlige, indbyrdes afhængige observationer, er CAR-modellen, der anvendes i forbindelse med Markovfelter. Et Markovfelt er, når n observationer har en betinget fordeling, som definerer nabostrukturen mellem punkterne. CAR-modellen er derfor en betinget normalfordeling for en mængde observationer $\mathbf{y}$, hvor hver observation er betinget med resten af observationerne. Den datagenererende funktion for CAR-modellen er givet i sætning 3.18 som

$$\mathbf{y} = \left(I - \rho W\right)^{-1} X\boldsymbol{\beta} + \left(\left(I_n - \rho W\right)^{-1}\right)^{\frac{1}{2}} \boldsymbol{\varepsilon},$$

hvor X, $\mathbf{y}$, W og parametrene er defineret som ved SAR-modellen.

Estimater for parametrene i hver model kan findes ved brug af enten maximum likelihood-estimation eller bayesiansk estimering af parametrene. Maximum likelihood-estimation tager udgangspunkt i likelihoodfunktionen $l(\boldsymbol{\theta}|\mathbf{y})$ for hver model, hvor $\boldsymbol{\theta}$ er en vektor indeholdende modellens parametre. Her findes det estimat $\hat{\theta}_i$, der maksimerer likelihoodfunktionen, ved at differentiere loglikelihoodfunktionen $l(\theta_i|y_i)$ med hensyn til hver parameter og sætte differentialet lig nul. Når der haves flere parametre at estimere som i SAR-og CAR-modellen, er det muligt at anvende profilloglikelihoodfunktionen til at forsimple udregningerne, hvilket er beskrevet i kapitel 5.

Bayesiansk estimering af modelparametrene medtager i modsætning til maximum likelihood-estimation eventuel forhåndsviden, som medtages i priorfordelingerne for hver parameter. Herefter findes posteriorfordelingen for alle parametrene i modellen ud fra priorfordelingerne og likelihoodfunktionen, og der kan konstrueres en MCMC-algoritme til at simulere estimater for parametrene fra posteriorfordelingen. I MCMC-

algoritmen kan parametre med kendte fordelinger simuleres ved brug af en Gibbs-sampler, mens parametre med ukendte fordelinger kan simuleres ved brug af Metropolis-Hastings. For SAR- og CAR-modellerne er MCMC-algoritmen konstrueret på to forskellige måder, da SAR-modellens parametre $\boldsymbol{\beta}$ og σ^2 har kendte fordelinger og kun parameteren ρ har ukendt fordeling, mens alle parametrene i CAR-modellen tilsyneladende har ukendte fordelinger, så alle parametre simuleres ved brug af Metropolis-Hastings.

I kapitel 7 blev der givet et praktisk eksempel på, hvordan der kan estimeres en SAR- og en CAR-model for den samme datamængde, der bestod af priser på villaer i Danmark solgt af ejendomsmægleren HOME. Først blev parametrene i hver model estimeret ved brug af maximum likelihood-estimation, hvor CAR-modellen havde én signifikant baggrundsvariabel mere end SAR-modellen, mens de standardiserede residualer var tættest på nul for SAR-modellen, der dermed kan siges at være en smule mere præcis end den fundne CAR-model. Dernæst blev modelparametrene estimeret ved brug af MCMC, der simulerede estimater fra posteriorfordelingen til hver model. Ved de bayesianske estimater blev priorfordelingerne valgt til ikke at være informative, så posteriorfordelingen fik størstedelen af sin information fra likelihoodfunktionen. Dermed skulle de bayesianske estimater gerne ligne estimaterne fra maximum likelihood, men de fundne estimater varierede en del både i størrelse og fortegn, selvom estimaterne fra de to maximum likelihood-estiomationer lignede hinanden ret godt. Parameteren ρ, som beskriver graden af afhængighed mellem naboerne, var positiv og mindre end én for alle de estimerede modeller, hvor den for begge CAR-modeller lå omkring 0,9 og dermed var højere end for de to estimerede SAR-modeller. De standardiserede residualer for hver estimeret model viste at den model, der passede bedst til data, var en SAR-model med estimater fra bayesiansk MCMC-estimering, selvom fortegnene for parametrene ikke intuitivt gav lige så god mening som fortegnene for parametrene i begge modeller med maximum likeihood-estimation. De bayesianske modeller kan sandsynligvis forbedres ved at foretage flere iterationer end 1000, så det er tydeligere hvilke værdier simuleringerne konvergerer mod, og det er også en mulighed at afprøve andre kombinationer af baggrundsvariable.

Det kan ikke sige at enten SAR- eller CAR-modellen er bedre end den anden, da det er to mulige måder at modellere rumligt afhængigt data på, og derfor vil passe bedre på nogle data end andre. Det er dog praktisk at have flere, mulige modeller at vælge imellem, når der ønskes en estimeret model for et givet datasæt, ligesom det er praktisk at have flere estimeringsmuligheder til at finde modelparametrene. Det kan være en fordel at anvende maximum likelihood-estimation, når der ikke haves nogen forhåndsviden om parametrene i en model, mens bayesiansk parameterestimering kan anvendes, hvis der haves forhåndsviden.

Litteratur

Luc Anselin. *Thirty years of spatial econometrics.* Blackwell Publishing, 2010.

Portal Danmark A/S. *rejs.dk.* `http://www.rejs.dk/regioner.asp`. Dato: 16-12-2012.

Sheldon Axler. *Linear Algebra Done Right.* Springer, 1997. ISBN 978-0-387-98258-8.

Adelchi Azzalini. *Statistical Inference - Based on the likelihood.* Chapman & Hall/CRC, 1996. ISBN 978 0412606502.

Julian Besag. *Spatial Interaction and the Statistical Analysis of Lattice Systems.* Journal of the Royal Statistical Society, 1974.

Julian Besag and Charles Kooperberg. *On conditional and intrinsic autoregressions.* Biometrika, 1995.

Renjie Chen, Craig Gotsman, and et al. *2012 Ninth International Symposium on Voronoi Diagrams in Science and Engineering.* The Institute of Electrical and Electronics Engineers, Conference Publishing Services, 2012. ISBN 978-0-7695-4724-4.

Noel A. C. Cressie. *Statistics for Spatial Data.* A Wiley-Interscience publication, 1993. ISBN 0 471 00255 0.

Wikipedia The Free Encyclopedia. *Wikipedia The Free Encyclopedia.* `http://en.wikipedia.org/wiki/Main_Page`. Anvendt foråret 2013.

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis.* Chapman & Hall, 1995. ISBN 0 412 03991 5.

Christiaan Heij, Paul de Boer, Philip Hans Franses, Teun Kloek, and Herman K. van Dijk. *Econometric Methods with Applications in Business and Economics.* Oxford University Press, 2004. ISBN 978-0-19-926801-6.

Zsusanna Horvath and Ryan Johnston. *AR(1) Time Series Process.* `http://www.math.utah.edu/~zhorvath/ar1.pdf`, Dato 19/11-2011 .

Krak. Krak. Dato: 05-12-2012.

Steffen L. Lauritzen. *Graphical Models.* Oxford University Press, 1996. ISBN 0 19 852219 3.

Peter M. Lee. *Bayesian Statistics.* Wiley, 2004. ISBN 978 0 470 68920 2.

James P. LeSage and Robert Kelley Pace. *Introduction to Spatial Econometrics.* CRC Press, Chapman & Hall, 2009. ISBN 978 1 4200 6424 7.

Peter Olofsson. *Probability, Statistics, and Stochastic Processes.* Wiley, 2005. ISBN 978 0471679691.

Lawrence Perko. *Differential Equations and Dynamical Systems.* Springer, 2001.

Kenneth H. Rosen. *Discrete Mathematics and its Applications.* McGraw-Hill, 2007.

Danmarks Statistik. Statistikbanken. Dato: 05-12-2012.

Melanie M. Wall. *A close look at the spatial structure implied by the CAR and SAR models.* Elsevier B.V., 2003.

Marc Wick. *GeoNames.* `http://www.geonames.org/postal-codes/postal-codes-denmark.html`. Dato: 29-11-2012.

Appendiks A

A.1 Tidskompleksitet

Afsnittet tager udgangspunkt i [Rosen, 2007, kapitel 3].

Når matematiske udtryk med kompliceret matrixregning med store matricer behandles i praksis, anvendes ofte algoritmer, og det er derfor praktisk at kunne sammenligne hastigheden af forskellige algoritmer med forskellige beregningsmetoder. Antallet af operationer, som en algoritme bruger til at løse et givet problem, angives ved algoritmens *tidskompleksitet*, der er et udtryk for, hvor meget arbejdskraft computeren har brug for, og den angiver derfor ikke det reelle tidsforbrug. Her ses der også bort fra, hvilken type software og hardware, computeren har, og det antages at alle operationer tager lige lang tid at udføre, hvilket selvfølgelig er en forenkling af virkeligheden. Tidskompleksitet kan opfattes som et estimat for sværhedsgraden af en algoritme, der beskrives ved Store-O-notationen.

Definition A.1 (Store-O):
Lad f og g være reelle funktioner defineret for en mængde bestående af heltal eller reelle tal, så $x \in \mathbb{R}$ eller $x \in \mathbb{Z}$. Funktionen $f(x)$ er $O(g(x))$, hvis der findes konstanter C og k, så

$$\left|f(x)\right| \leq C \cdot \left|g(x)\right| \quad \text{når } x > k.$$

Tidskompleksistet angiver hvordan antallet af operationer, algoritmen udfører, vokser når inputtet n bliver større. Almindelige tidskompleksiteter er vist i tabellen herunder.

Kompleksitet	Terminologi
$O(1)$	konstant kompleksitet
$O(\ln(n))$	logaritmisk kompleksitet
$O(n)$	lineær kompleksitet
$O(n\ln(n))$	$n\ln(n)$ kompleksitet
$O(n^k)$	polynomisk kompleksitet
$O(k^n)$	eksponentiel kompleksitet
$O(n!)$	faktoriel kompleksitet

Appendiks B

B.1 Normalfordelingen

Definition B.1 (Normalfordeling):
En stokastisk vektor $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n$ siges at være n-dimensionalt normalfordelt med parametrene $\boldsymbol{\mu}$ og Σ, hvis den har den simultane tæthedsfunktionen

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T\Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\},$$

hvor $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T \in \mathbb{R}^n$ er middelværdivektor, og hvor Σ er en positivt semidefinit $(n \times n)$-matrix. Dette noteres

$$\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \Sigma).$$

Jævnfør [Lee, 2004, s. 304] er det muligt at vise, at $\boldsymbol{\mu}$ og Σ er henholdsvist middelværdivektoren og kovariansmatricen udtrykt ved

$$\boldsymbol{\mu} = \begin{bmatrix} \mathrm{E}[Y_1] \\ \vdots \\ \mathrm{E}[Y_n] \end{bmatrix} \quad \text{og} \quad \Sigma = \begin{bmatrix} \mathrm{Cov}[Y_1, Y_1] & \cdots & \mathrm{Cov}[Y_1, Y_n] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[Y_n, Y_1] & \cdots & \mathrm{Cov}[Y_n, Y_n] \end{bmatrix}.$$

B.2 Den uniforme fordeling

Definition B.2 (Uniform fordeling):
En stokastisk variabel $Y \in \mathbb{R}$ siges at være uniformt fordelt på intervallet $[a,b]$, hvis den har tæthedsfunktionen

$$p(y) = \frac{1}{b-a} \quad a \leq y \leq b,$$

hvilket noteres som

$$Y \sim U(a,b).$$

Jævnfør [Lee, 2004, s. 296-297] er middelværdien og variansen givet ved

$$\mathrm{E}[Y] = \frac{a+b}{2}$$
$$\mathrm{Var}[Y] = \frac{(b-a)^2}{12}.$$

B.3 Invers gammafordeling

Definition B.3:
En stokastisk variabel $Y \in \mathbb{R}$ siges at være invers gammafordelt med parametrene a og b, hvis den har tæthedsfunktionen

$$p(y) = \frac{b^a}{\Gamma(a)} y^{-(a+1)} \exp\left\{-\frac{b}{y}\right\} \qquad 0 < y < \infty,$$

hvor Γ er gammafunktionen givet ved $\Gamma(a) = (a-1)!$ for positive heltal, og $a, b \in \mathbb{R}_+$. Dette noteres

$$Y \sim IG(a,b).$$

Gammafunktionen er defineret for positive heltal og alle komplekse tal, men ikke for negative heltal og nul. For komplekse tal med positiv realdel er den defineret ved det uegentlige integrale

$$\Gamma(n) = \int_0^\infty e^{-t} t^{n-1} dt.$$

Jævnfør [Gelman et al., 1995, s. 474] er middelværdi og varians for den inverse gammafordeling givet ved

$$\mathrm{E}[Y] = \frac{b}{a-1} \qquad \text{givet } a > 1$$
$$\mathrm{Var}[Y] = \frac{b^2}{(a-1)^2(a-2)} \qquad \text{givet } a > 2.$$

B.4 Betafordeling

Definition B.4:
En stokastisk variabel $Y \in \mathbb{R}$ siges at være betafordelt med parametrene a og b, hvis den har tæthedsfunktionen

$$p(y) = \frac{1}{B(a,b)} Y^{a-1} (1-Y)^{b-1} \qquad 0 < y < 1,$$

hvor $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}(1-t)^{b-1} dt$ og $\Gamma(n) = (n-1)!$ for positive heltal. Dette noteres

$$Y \sim B(a,b).$$

Jævnfør [Lee, 2004, s. 292-293] er middelværdi og varians for en betafordeling givet ved

$$\mathrm{E}[Y] = \frac{a}{a+b}$$
$$\mathrm{Var}[Y] = \frac{ab}{(a+b)^2(a+b+1)}.$$

Desuden skal det bemærkes, at for $a = b = 1$ haves en standard uniform fordeling $U(0,1)$. Betafordelingen ligger som udgangspunkt i intervallet [0,1], men hvis den skaleres, kan den komme til at ligge i andre intervaller, hvis det ønskes. Når $a = b = 0$ haves en fordeling, der er proportional med $p^{-1}(p-1)^{-1}$, og dermed repræsenteres en fuldstændig usikkerhed.

Appendiks C

C.1 Matrixregneregler

Regnereglerne er fundet i [Encyclopedia, Opslag: Invertible matrix], [Encyclopedia, Opslag: Matrix calculus] og [Encyclopedia, Opslag: Determinant].

Regneregel C.1:
Det gælder for en matrix $A(a)$, hvor a er et reelt tal, at

$$\begin{aligned}\frac{\partial |A|}{\partial a} &= |A| \operatorname{tr}\left(A^{-1}\frac{\partial A}{\partial a}\right) \Rightarrow \\ \frac{\partial \left|I_n - \rho W\right|}{\partial \rho} &= \left|I_n - \rho W\right| \operatorname{tr}\left(\left(I_n - \rho W\right)^{-1}\frac{\partial \left(I_n - \rho W\right)}{\partial \rho}\right) \\ &= \left|I_n - \rho W\right| \operatorname{tr}\left(-\left(I_n - \rho W\right)^{-1} W\right).\end{aligned}$$

Regneregel C.2:
Der gælder for en matrix A at $I_n = AA^{-1}$, hvormed det haves, at

$$\begin{aligned}0 = \frac{\partial}{\partial a} I_n = \frac{\partial}{\partial a}\left(AA^{-1}\right) = \frac{\partial A}{\partial a}A^{-1} + A\frac{\partial A^{-1}}{\partial a} \Rightarrow \\ \frac{\partial A^{-1}}{\partial a} = -A^{-1}\frac{\partial A}{\partial a}A^{-1}.\end{aligned}$$

Regneregel C.3:
For matricer A, B, C og D gælder det, at

$$\operatorname{tr}(ABCD) = \operatorname{tr}(DABC) = \operatorname{tr}(CDAB) = \operatorname{tr}(BCDA).$$