

pga. bedre erhvervsstatistik, hvorved δ er tættere ved nul. Udviklingen i parameteren δ kan altså forklares som en følge af datakonstruktionen.

Side 440 argumenterer Kærgård for, at autokorrelationen i investeringsfunktionen faktisk burde være negativ, da store investeringer ét år kan skyldes en ophobning, således at investeringerne året efter bliver små. Det viser sig, at det giver en svag forbedring af modeltilpasningen, hvis modellen udvides med et glidende gennemsnitsled, der tager højde for en anelse negativ autokorrelation i residualerne i (23). Det fører til en større estimeret værdi af $\delta = 1.07$. Tages der ikke hensyn til en eventuel negativ autokorrelation ved hjælp af glidende gennemsnitsled, vil den estimerede værdi af β_2 blive negativ, hvilket fører til endnu mindre værdier af δ . Kun i en rigtig tidsrække-model kan afvejningen af om autokorrelationen skal være positiv eller negativ foregå.

Som tidsrækkeanalytiker vil jeg alt i alt konkludere, at Kærgård med sin investeringsfunktion har fundet en god tidsrække-model – endda uden at bruge tidsrækkemetoder – med estimerede parameterværdier, der er rimelige fra et tidsrækkeanalytisk synspunkt. Investeringsfunktionen begrundes med økonomisk teoretiske argumenter, men da disse argumenter leder til antagelser om parametrene værdier, der ikke stemmer overens med de estimerede værdier, synes disse argumenter kun at være distraherende for den empiriske analyse.

Robuste metoder

Nils Kousgaard

Statistisk Institut, Københavns Universitet

Mange af disputatsens økonomiske relationer formuleres ved hjælp af lineære modeller, der for størstepartens vedkommende tilpasses til data, $(y_t, x_{t1}, \dots, x_{tp})$, $t = 1, \dots, T$, ved anvendelse af MK-metoden (mindste kvadraters metode). Ved tilpasning forstås her bestemmelsen af estimater for de $p + 1$ regressionsparametre $\beta_0, \beta_1, \dots, \beta_p$ i den lineære regressionsmodel

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_p x_{tp} + e_t, \quad i = 1, \dots, T.$$

I kapitel 9 forholder Kærgård sig dog kritisk til brugen af MK-metoden, og en række alternative, såkaldt robuste, estimationsmetoder trækkes frem. Imidlertid spiller disse metoder ikke nogen rolle i de senere kapitler, hvor de egentlige statistiske analyser

pga. bedre erhvervsstatistik, hvorved δ er tættere ved nul. Udviklingen i parameteren δ kan altså forklares som en følge af datakonstruktionen.

Side 440 argumenterer Kærgård for, at autokorrelationen i investeringsfunktionen faktisk burde være negativ, da store investeringer ét år kan skyldes en ophobning, således at investeringerne året efter bliver små. Det viser sig, at det giver en svag forbedring af modeltilpasningen, hvis modellen udvides med et glidende gennemsnitsled, der tager højde for en anelse negativ autokorrelation i residualerne i (23). Det fører til en større estimeret værdi af $\delta = 1.07$. Tages der ikke hensyn til en eventuel negativ autokorrelation ved hjælp af glidende gennemsnitsled, vil den estimerede værdi af β_2 blive negativ, hvilket fører til endnu mindre værdier af δ . Kun i en rigtig tidsrække-model kan afvejningen af om autokorrelationen skal være positiv eller negativ foregå.

Som tidsrækkeanalytiker vil jeg alt i alt konkludere, at Kærgård med sin investeringsfunktion har fundet en god tidsrække-model – endda uden at bruge tidsrækkemetoder – med estimerede parameterværdier, der er rimelige fra et tidsrækkeanalytisk synspunkt. Investeringsfunktionen begrundes med økonomisk teoretiske argumenter, men da disse argumenter leder til antagelser om parametrene værdier, der ikke stemmer overens med de estimerede værdier, synes disse argumenter kun at være distraherende for den empiriske analyse.

Robuste metoder

Nils Kousgaard

Statistisk Institut, Københavns Universitet

Mange af disputatsens økonomiske relationer formuleres ved hjælp af lineære modeller, der for størstepartens vedkommende tilpasses til data, $(y_t, x_{t1}, \dots, x_{tp})$, $t = 1, \dots, T$, ved anvendelse af MK-metoden (mindste kvadraters metode). Ved tilpasning forstås her bestemmelsen af estimater for de $p + 1$ regressionsparametre $\beta_0, \beta_1, \dots, \beta_p$ i den lineære regressionsmodel

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_p x_{tp} + e_t, \quad i = 1, \dots, T.$$

I kapitel 9 forholder Kærgård sig dog kritisk til brugen af MK-metoden, og en række alternative, såkaldt robuste, estimationsmetoder trækkes frem. Imidlertid spiller disse metoder ikke nogen rolle i de senere kapitler, hvor de egentlige statistiske analyser

udføres. Her benyttes MK-metoden, idet en række størrelser, der udtrykker indflydelsen fra de enkelte observationer på analysens resultater, de såkaldte "regression diagnostics", inddrages.

Udgangspunktet for diskussionen i kapitel 9 er et ønske om at "få en estimator, der ikke er så følsom overfor store afvigelser som mindste kvadraters metode" (side 222). Bag denne bemærkning skjuler sig det i dag velkendte forhold, at observationer, der er ekstreme i forhold til hovedparten af data, kan få en dominerende indflydelse på værdierne af MK-estimerne for regressionsparametrene. En situation, hvor værdierne af de estimerede parametre hovedsageligt bestemmes af en lille del af data, er selvfølgelig ikke ideel og kan så tvivl om analysens konklusioner. Dette har affødt en voksende interesse for anvendelsen af analysemetoder, der ikke i samme omfang som MK-metoden er følsomme overfor ekstreme observationer. Metoder, hvis egenskaber kun i begrænset omfang afhænger af fordelingen af regressionsmodellens restled, e_t , kaldes ofte robuste. Imidlertid er det ikke på nogen måde oplagt, at man blot ved at erstatte MK-metoden med en robust metode vil få en sundere analyse.

Mange statistiske analyser kan opdeles i to faser, nemlig en eksplorativ eller beskrivende fase og en konkluderende fase. I den eksplorative fase er man primært interesseret i at tilpasse regressionsfunktionen til data ved anvendelse af en metode, der ikke tager hensyn til atypiske observationer, d.v.s. observationer, der på en eller anden måde er ekstreme i forhold til hovedparten af data. Det undersøges om der er inhomogeniteter i data, f.eks. om der er år, hvor kombinationerne af modellens variable afviger markant fra de øvrige år. Man kan herved få et indtryk af restledsfordelingens form, om der er behov for ændringer i modelspecifikationen, eller om der bør lægges begrænsninger på modellens anvendelsesområde. I denne fase interesserer man sig kun i mindre grad for præcisionen af de estimerede regressionsparametre og lader de normale år bestemme tilpasningens form. I den konkluderende fase lægges hovedvægten derimod på fortolkningen af data, og der ønskes en præcis estimation af parametrene. Idet den optimale estimationsmetode vil være bestemt af restledsfordelingen, skal denne beskrives så detaljeret som muligt.

I analysens to faser stilles der således forskellige og måske uforenelige krav til estimations- eller tilpasningsmetoden. Mens man i den eksplorative fase har brug for en metode, der kan tage højde for den værst tænkelige restledsfordeling og som ikke påvirkes af forekomsten af ekstreme observationer af såvel responsvariablen som de forklarende variable, er man i den konkluderende fase interesseret i stærkt restriktive antagelser om restledsfordelingen med henblik på at sikre den optimale udnyttelse af informationen i data.

Ofte skal MK-metoden imidlertid optræde i to roller, nemlig som tilpasningsmetode i den eksplorative fase og som maksimum likelihood metode i den konkluderende fase under en forudsætning om at regressionsmodellens restled er normalfordelte. Desværre

kan MK-metoden fungere dårligt i den eksplorative fase på grund af dens følsomhed overfor ekstreme observationer. En dårlig tilpasning til observationerne i nogle få år tværes ud over de øvrige år, hvilket skyldes, at selve estimationskriteriet, nemlig minimeringen af de kvadrerede residualer, har en indbygget tendens til at undgå store afvigelser mellem tilpasning og data og i stedet producere en række halvstore afvigelser. Dette forhold gør det bl.a. vanskeligt at identificere år med afvigende værdier af en eller flere af modellens variable.

Er restleddene imidlertid normalfordelte og holder linearitetsantagelsen, er MK-estimatet identisk med maksimum likelihood estimatet, og har derfor optimale statistiske egenskaber, d.v.s. at informationen i data udnyttes optimalt. Gøres derimod restleddesfordelingen haler tungere end normalfordelingens haler, svarende til at der hyppigere end i normalfordelingen forekommer ekstreme værdier af responsvariablen, mister MK-metoden sine optimalitetsegenskaber. MK-metoden er således ikke robust, hvormed der menes, at dens statistiske egenskaber ikke er robuste overfor afvigelser fra normalitetsforudsætningen.

En metode, der hverken er følsom overfor forekomsten af ekstreme observationer af responsvariablen eller af de forklarende variable, kaldes også resistent. I erkendelse af at MK-metoden hverken er robust eller resistent, har man udviklet en række numeriske størrelser, de såkaldte diagnostics, der sigter mod at identificere observationer som ved anvendelsen af MK-metoden har en særlig stærk indflydelse på de estimerede regressionsparametre. Anvendelsen af disse diagnostics, herunder bl.a. Cook's afstand, der måler ændringen i de estimerede parametre ved udeladelse af en enkelt observation, er i de senere år blevet almindelig i forbindelse med MK-analyser. Mange diagnostics kan imidlertid kun benyttes ved vurderingen af indflydelsen fra et enkelt observationssæt, f.eks. et år.

Udviklingen af robuste observationsmetoder har især fundet sted indenfor de seneste 20 år. Det første skridt bestod i udviklingen af M-estimatorerne, der er robuste overfor tunghalede afvigelser fra antagelsen om, at restleddene er normalfordelte. I sammenligning med MK-metoden lægger disse metoder mindre vægt på forekomsten af store residualer. Dette gælder bl.a. den metode, der af Kærgård kaldes A_1 -metoden (s. 226) eller MAD-metoden. Den minimerer summen af de numeriske residualer, d.v.s.

$$\sum |y_t - \beta_0 - \beta_1 x_{t1} - \dots - \beta_p x_{tp}|$$

i stedet for som MK-metoden at minimere summen af de kvadrerede residualer

$$\sum (y_t - \beta_0 - \beta_1 x_{t1} - \dots - \beta_p x_{tp})^2$$

Sammenlignet med MK-metoden er M-metoderne mere beregningskrævende, uden at dette dog på nogen måde udgør en hindring for deres anvendelse. Estimationen udføres ved gentagne minimeringer af en vejet kvadratafvigelsessum med vægte, der successivt justeres under hensyntagen til størrelserne af de beregnede residualer. De robuste metoder er i varierende omfang mindre følsomme overfor ekstreme værdier af responsvariablen end MK-metoden, og en række af disse, bl.a. minimeringen af summen af de numeriske residualer, har i dag en vis udbredelse. M-estimatorer er imidlertid ikke resistente overfor forekomsten af ekstreme værdier af de forklarende variable, omend de i konkrete situationer ikke er helt så følsomme som MK-metoden.

Ønsker man at anvende metoder, der er resistente overfor enhver form for ekstreme observationer, kan man f.eks. benytte de metoder, som Kærgård beskriver i kapitel 9, afsnit 2. De har imidlertid den ulempe, at de vanskeligt lader sig generalisere til modeller med mere end en forklarende variabel. Til gengæld er de relativt simple at benytte. Derimod vil den metode, der minimerer kvadratet på medianen i residualfordelingen være yderst velegnet, idet den i en vis forstand har den højest opnåelige grad af resistens. Metoden er imidlertid meget beregningstung, specielt i situationer med store datamaterialer og mange forklarende variable, og endvidere giver den estimerer med en ringe præcision. (Rousseeuw and Leroy (1987)). Herved illustreres det dilemma, man står overfor ved valget af tilpasnings- eller estimationsmetode. Jo mere resistent en metode er, jo mere velegnet vil den være til at identificere inhomogeniteter i data, d.v.s. jo mere velegnet vil den være i den eksplorative fase af den statistiske analyse. Til gengæld er resistente metoder uegnede som grundlag for den fortolkningsmæssige del af analysen. Dette skyldes dels, at de giver usikre bestemmelser af regressionsparametrene (stor varians), dels at deres fordelingsmæssige egenskaber er vanskelige at håndtere.

I kapitel 9 foretages i afsnittene 5 og 6 en række empiriske sammenligninger af MK-metoden med MAD-metoden. Først tilpasses en model til data fra én periode ved hver af de to metoder, og dernæst beregnes udtryk for den præcision, hvormed responsvariablen kan forudsiges i en senere periode. Jo bedre forudsigelser en metode giver, jo mere velegnet anses den så for at være til estimation af modellens parametre.

Sammenligninger af denne art har imidlertid kun begrænset værdi og kan ikke fortælle meget om, hvilke estimationsmetoder, der bør foretrækkes. I den eksplorative fase bør man såvidt muligt benytte en resistent metode. Fremgår det heraf, at restleddene med rimelighed kan anses for at være normalfordelte, skal man anvende MK-metoden som grundlag for den konkluderende del af analysen. Har derimod restledsfordelingen haler, der er tungere end normalfordelingens haler, kan man overveje anvendelsen af f.eks. A_1 -metoden, der er identisk med maksimum likelihood metoden, såfremt restleddene følger en Laplace fordeling. I den konkluderende fase er det, som tidligere nævnt, af stor betydning, at informationen i data udnyttes optimalt, og dette sikres netop igennem anvendelse af den korrekte restledsfordeling.

Kærgård skal have ros for at han er opmærksom på at tilpasningen af den lineære regressionsmodel til data i princippet kan bestemmes ved andre metoder end MK-metoden. Desværre har denne opmærksomhed ikke sat sig synlige spor, når det gælder de i disputatsen udførte statistiske analyser, ligesom der heller ikke er taget nogen principiel stilling til, hvilken rolle sådanne metoder bør spille ved analysens udførelse.

Litteratur

Rousseeuw, P. J. and Leroy, A. M. 1987. *Robust Regression and Outlier Diagnostics*. New York.

Økonomisk historie

Carl-Axel Nilsson

Institut for Historie, Københavns Universitet

For en økonomisk historiker er Niels Kærgårds bog i høj grad interessant. Der pågår fortløbende siden mange år i disciplinen økonomisk historie en debat om forholdet mellem økonomisk analyse med brug af økonometriske teknikker og historisk rekonstruktion. Helt klart aktualiseres denne debat i Danmark i anledning af Kærgårds disputats. Der er også tendenser til en sådan debat hos Kærgård (NK).

Han påpeger i introduktionskapitlet (s. 19), at hans "analyse vil være af interesse for både økonomer, statistikere og historikere". Begrundelsen herfor er: "Selve den økonometriske teknik er i sig selv et grænseområde mellem økonomi og statistik, og når den skal bruges på data så langt tilbage som 1870, må faget økonomisk historie også være relevant". To sider senere præciserer han sin ambition om at "skematisere og kvantificere nogle få centrale årsagssammenhænge, som har været fælles for hele den undersøgte periode, det vil her sige fra 1870 til 1970. Om noget sådant kan gøres, og eventuelt oven i købet være nyttigt også for historiske analyser, er imidlertid selvfølgelig diskutabelt."

Men hvad forstår NK mere præcist med udtrykket historisk analyse? Der er et afsnit om Cleo og den økonomiske historie i afhandlingen, et afsnit, som desværre er meget kortfattet. Vi får dog følgende udtalelse (2. 22): "Når undersøgelsesperioden, som her går mere end 100 år tilbage, er det imidlertid oplagt, at nogle af de fundne sammenhænge *kun* (min kurs.) har historisk interesse, og man bliver tvunget til at overveje, hvad historikere mon kan mene om den økonometriske metode."

Uden helt at have forstået indholdet i denne sætning, forekommer det mig dog rimeligt at koble den sammen med hvad NK udtaler til allersidst i afhandlingens konklude-

Kærgård skal have ros for at han er opmærksom på at tilpasningen af den lineære regressionsmodel til data i princippet kan bestemmes ved andre metoder end MK-metoden. Desværre har denne opmærksomhed ikke sat sig synlige spor, når det gælder de i disputatsen udførte statistiske analyser, ligesom der heller ikke er taget nogen principiel stilling til, hvilken rolle sådanne metoder bør spille ved analysens udførelse.

Litteratur

Rousseeuw, P. J. and Leroy, A. M. 1987. *Robust Regression and Outlier Diagnostics*. New York.

Økonomisk historie

Carl-Axel Nilsson

Institut for Historie, Københavns Universitet

For en økonomisk historiker er Niels Kærgårds bog i høj grad interessant. Der pågår fortløbende siden mange år i disciplinen økonomisk historie en debat om forholdet mellem økonomisk analyse med brug af økonometriske teknikker og historisk rekonstruktion. Helt klart aktualiseres denne debat i Danmark i anledning af Kærgårds disputats. Der er også tendenser til en sådan debat hos Kærgård (NK).

Han påpeger i introduktionskapitlet (s. 19), at hans "analyse vil være af interesse for både økonomer, statistikere og historikere". Begrundelsen herfor er: "Selve den økonometriske teknik er i sig selv et grænseområde mellem økonomi og statistik, og når den skal bruges på data så langt tilbage som 1870, må faget økonomisk historie også være relevant". To sider senere præciserer han sin ambition om at "skematisere og kvantificere nogle få centrale årsagssammenhænge, som har været fælles for hele den undersøgte periode, det vil her sige fra 1870 til 1970. Om noget sådant kan gøres, og eventuelt oven i købet være nyttigt også for historiske analyser, er imidlertid selvfølgelig diskutabelt."

Men hvad forstår NK mere præcist med udtrykket historisk analyse? Der er et afsnit om Cleo og den økonomiske historie i afhandlingen, et afsnit, som desværre er meget kortfattet. Vi får dog følgende udtalelse (2. 22): "Når undersøgelsesperioden, som her går mere end 100 år tilbage, er det imidlertid oplagt, at nogle af de fundne sammenhænge *kun* (min kurs.) har historisk interesse, og man bliver tvunget til at overveje, hvad historikere mon kan mene om den økonometriske metode."

Uden helt at have forstået indholdet i denne sætning, forekommer det mig dog rimeligt at koble den sammen med hvad NK udtaler til allersidst i afhandlingens konklude-