Danmarks Statistik
MODELGRUPPEN

Arbejdspapir*

Grane Høegh

21. juli 2010

Økonometriske værktøjer

Resumé:

Dette arbejdspapir beskriver økonometriske værktøjer, som anvendes til formulering af ligninger i større makroøkonometriske modeller, såsom ADAM.

GRH21710

Nøgleord: Økonometriske værktøjer

1. Indledning

Dette papir er en opdatering af et undervisningspapir af lidt ældre dato med samme overskrift af Thomas Kyhl og Jacob Nielsen fra 2002. Meget af indholdet har ændret sig markant, men opbygningen minder meget op den oprindelige.

Noten er en praktisk gennemgang af de metoder, der allerede vil være gennemgået i økonometrikurserne på polit-studiet. Noten vil kun beskæftige sig med de områder, der er vigtige for økonometrisk arbejde med økonomiske modeller.

I afsnit 2 repeteres fejlkorrektionsmodellen, som er den mest benyttede model til økonometrisk modellering i makroøkonometrisk arbejde. Helt afgørende for estimationsresultaterne er, at kortsigtsleddene er stationære og at langsigtsvariablerne kointegrerer. Afsnit 3 beskæftiger sig med stationaritet. Det er vigtigt, at modellen ikke er misspecificeret, og i afsnit 4 repeteres de gængse misspecifikationstests, som man bør benytte. Parameterustabilitet er et problem, man ofte støder på, når man arbejder med tidsserier. Der kigges nærmere på, hvordan problemet identificeres i afsnit 5, hvor der også gives et par muligheder for at komme problemet til livs. Dummy'er er en måde at fjerne ekstreme observationer fra sit datasæt. Afsnit 6 beskriver, hvornår man bør og ikke bør benytte sig af dummy'er. ADAM er ikke estimeret på hele den tilgængelige dataperiode. Baggrunden for dette beskrives i afsnit 7. Kriterierne for at vælge konkurrerende modeller er beskrevet i afsnit 8. Til sidst kommer en lille krølle. Jeg viser i afsnit 9, at indførelse af et trendkorrektionsled hjælper med på fortolkningen af vores ligevægtsrelationer. Afsnit 10 konkluderer.

2. Fejlkorrektionsmodellen

En vilkårlig lineær økonometrisk model for tidsserier kan skrives på ADL-form[1]. Her vises et eksempel med 2 forklarende variabler og 1 lag:

$$\Delta y_t = \mu_0 + \mu_1 \Delta x_{1,t} + \mu_2 \Delta x_{2,t} + \mu_3 y_{t-1} + \mu_4 x_{1,t-1} + \mu_5 x_{2,t-1} \tag{2.1}$$

Det er denne ligning, som estimeres i et almindeligt simpelt lineært estimationsprogram som for eksempel Aremos har. I dette program vil man estimere parametrene μ_0, μ_1, μ_2, μ_3, μ_4 og μ_5.

Ved at rykke lidt rundt på ligningen fås fejlkorrektionsformen:

$$\Delta y_t = \mu_1 \Delta x_{1,t} + \mu_2 \Delta x_{2,t} - (-\mu_3)\left(y_{t-1} - \left(-\frac{\mu_0}{\mu_3} - \frac{\mu_4}{\mu_3} x_{1,t-1} - \frac{\mu_5}{\mu_3} x_{2,t-1} \right) \right) \tag{2.2}$$

som også kan opskrives ved at dele ligningen op i to dele:

$$\Delta y_t = \mu_1 \Delta x_{1,t} + \mu_2 \Delta x_{2,t} - \gamma\left(y_{t-1} - y_{t-1}^* \right) \tag{2.3}$$

$$y_t^* = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} \tag{2.4}$$

hvor $\gamma = -\mu_3$, $\beta_0 = -\frac{\mu_0}{\mu_3}$, $\beta_1 = -\frac{\mu_4}{\mu_3}$ og $\beta_2 = -\frac{\mu_5}{\mu_3}$.

Hvis alle serierne er stationære, vil ligning (2.4) angive ligevægtsrelationen for y_t. Hermed kan man tolke y_t^*, som ligevægtsniveauet for y_t. Ligning (2.3) siger således, at hvis $x_{1,t}$ ændrer sig med 1, vil y_t ændre sig med μ_1, og i ligevægt vil *y* have ændret sig med β_1. Typisk er $|\mu_1| < |\beta_1|$, hvilket betyder, at *y* ikke opnår hele ændringen i samme periode. Endvidere vil vi typisk have, at tilpasningshastigheden er mellem 0 og 1, $0 < \gamma < 1$, hvilket vil sige, at der er en gradvis tilpasning mod ligevægten.

Antag, at $0 < \gamma < 1$, $0 < \mu_1 < \beta_1 = 1$ og $y_{t-1} = y_{t-1}^*$. Vi starter altså i ligevægt, og på langt sigt skal *y* stige med det samme som x_1. Øges $x_{1,t}$ med 0.01, vil y_t øges med mindre end 0.01 i samme periode. I den følgende periode er vi ude af ligevægt, da er y_t er steget med mindre end $x_{1,t}$, så $y_t < y_t^*$. Dette medfører, at y_{t+1} også vil stige i forhold til y_t. Dog er det kun en fast andel af forskellen, $0 < \gamma < 1$, som indsnævres. Hermed går *y* asymptotisk mod y^* og ændringerne i *y* bliver mindre og mindre.

Er $|\mu_1| > |\beta_1|$, er der overshooting. Det betyder, at når $x_{1,t}$ stiger, så vil y_t umiddelbart stige(falde) kraftigt for herefter i en periode at falde(stige). Tilsvarende betyder $1 < \gamma < 2$, at fejlkorrektionen overkorrigerer. Hermed vil et stød til x_1 betyde, at *y* skiftevis vil stige og falde, indtil den når sin nye ligevægt. Estimerer man sådanne resultater, bør man overveje om overshooting er fornuftigt. Er det ikke, bør man teste om man kan restriktere parametrene så

[1] Jf. Verbeek (2004) 9.1 og Nielsen (2010b).

overshooting forsvinder eller prøve at opstille sin estimationsmodel med andre forklarende variabler. Er $\gamma < 0$ eller $\gamma > 2$, er modellen ustabil, og man bør helt sikkert overveje, hvad der er gået galt. Typisk vil $\gamma < 0$ betyde, at der ikke er fejlkorrektion, fordi serierne ikke umiddelbart har nogen niveausammenhæng.

I bilag A er givet et eksempel på estimation af en fejlkorrektionsligning i AREMOS.

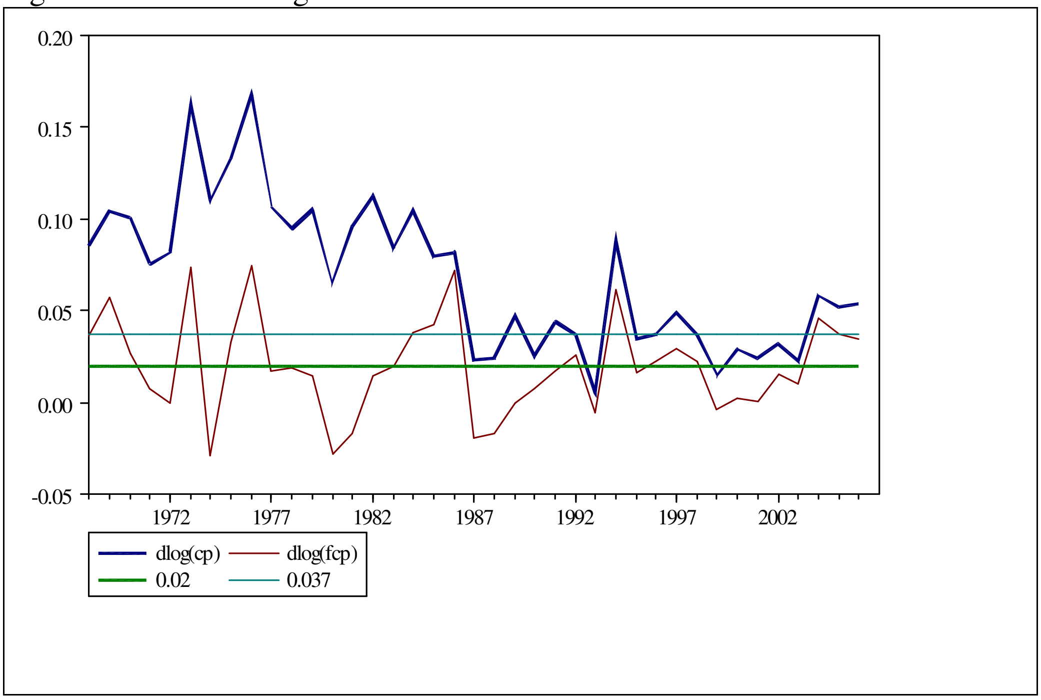
3. Stationaritet

En grundforudsætning for at estimere noget fornuftigt på baggrund af en fejlkorrektionsmodel er, at residualerne er stationære. I en fejlkorrektionsmodel er venstresidevariablen i differenser eller dlog'er. Hvis denne variabel er stationær, vil den kun kunne forklares af andre stationære sammenhænge. Det vil sige andre stationære serier, og kointegrerende relationer mellem ikke-stationære serier. For at sikre en fornuftig fortolkning af sine variabler, så skal man sørge for, at serierne i den kortsigtede del er stationære, og at niveau-variablerne kointegrerer.

Hvis Δy ikke er stationær, så kan man komme til at udsætte sig selv for en spurious regression – jf. Verbeek (2004) afsnit 9.2.1.

En serie er stationær, hvis dens middelværdi og varians er stabil over tid. Dette skal gælde uanset hvilken delperiode, der betragtes. Alle serierne i den kortsigtede del kan og bør tjekkes før estimationen foretages. Vi kan tage privatforbruget som eksempel. Figur 3.1 viser dlog til forbruget (ca. lig den procentvise stigning i forbruget) i henholdsvis faste og løbende priser. I faste priser ser serien stationær ud, mens den ikke ser stationær ud i løbende priser før 1986.

Figur 3.1. Privatforbruget i Danmark fra 1968-2006.



Kilde: ADAMs databank.

Dette er et typisk resultat. Rigtig mange økonomiske serier har konstante vækstrater i faste størrelser, men det holder ikke i løbende priser, så længe der forekommer systematiske ændringer, f.eks. regimeskift, i den underliggende prisstigning, hvilket har været tilfældet i den viste. Efter overgangen til fastkurspolitik i 80'erne fremstår prisstigningen som stationær om et forholdsvist lavt niveau. Dette betyder også, at dlog til de fleste prisserier ikke er stationære. Til gengæld er dlog til langt de fleste serier med relative priser stationære, da det inflationære led går ud.

Nogle gange kan det ved et visuelt tjek være svært at se, om en serie er stationær, og man kunne være fristet til at tro, at stationaritetstests kan give et entydigt svar i disse tilfælde. I Nielsen (2010c) gennemgås tests for stationaritet. Desværre er styrken på disse tests meget lav, hvilket betyder, at man ofte accepterer hypotesen om enhedsrod, hvor den ikke er der. Herved kan man komme til at forkaste mange gode modeller. Mit råd vil være at lægge mere vægt på visuel inspektion end på traditionelle enhedsrodstests. Nielsen (2010c) henviser til mere avancerede tests, disse ligger dog uden for vores pensum.

Hvis niveauvariablerne ikke kointegrerer, kan vi ikke tage estimationsresultaterne for gode varer. Dette kan testes ved et PcGive test omtalt i Nielsen (2010d). Dette kan have svag styrke, så jeg ville alternativt plotte residualerne, og se om de er stationære. Ser de stationære ud, så er de det nok. Vi har trods alt opstillet en model, hvis teoretiske grundlag vi tror på.

Ser en serie ikke stationær ud skyldes det typisk: Trender, niveauskift, variansændringer eller enhedsrod. Som nævnt tidligere er det stationaritet for dlog (og nogle gange diff) af serierne, som i første omgang er essentiel. En løsning til alle fire problemer kan være at tage serien relativt til en anden serie, der indeholder samme underliggende trend. For eksempel kan privatforbruget i løbende priser divideres med forbrugerprisindekset. De indeholder begge den prismæssige trend.

Alternativt kan årsagen modelleres. Man kan i sin model inkludere en lineær trend, indsætte en dummy for en eller flere variabler ved niveauskift. Alternativt kan man estimere de to perioder hver for sig. Denne mulighed kan også benyttes for variansændringer. Typisk vil variansændringer ikke påvirke parameterestimaterne, men estimater på variansen og de kritiske værdier er ikke valide. Det kan muligvis hjælpe at ændre funktionel form, f.eks. ved at prøve om variablerne skal være i logaritmer.

Typisk vil afvist stationaritet betyde, at man hverken kan benytte sine parameterestimater, standardafvigelser eller tests til noget som helst. Hovedkonklusionen må være, at man skal tage afvist stationaritet seriøst.

4. Misspecificerede modeller

Efter estimationen bør der foretages misspecifikationstest. Nielsen (2010a) nævner 4 tests for autokorrelation, heteroskedasticitet, funktionel form og normalitet. Det mest essentielle er at teste for autokorrelation. I fejlkorrektionsmodeller betyder autokorrelation, at estimatoren er inkonsistent. Dette er ret skidt, men problemet kan ofte løses ved flere forklarende variabler eller flere lags. Sidstnævnte kan dog give problemer med modellens egenskaber, men mere om det senere.

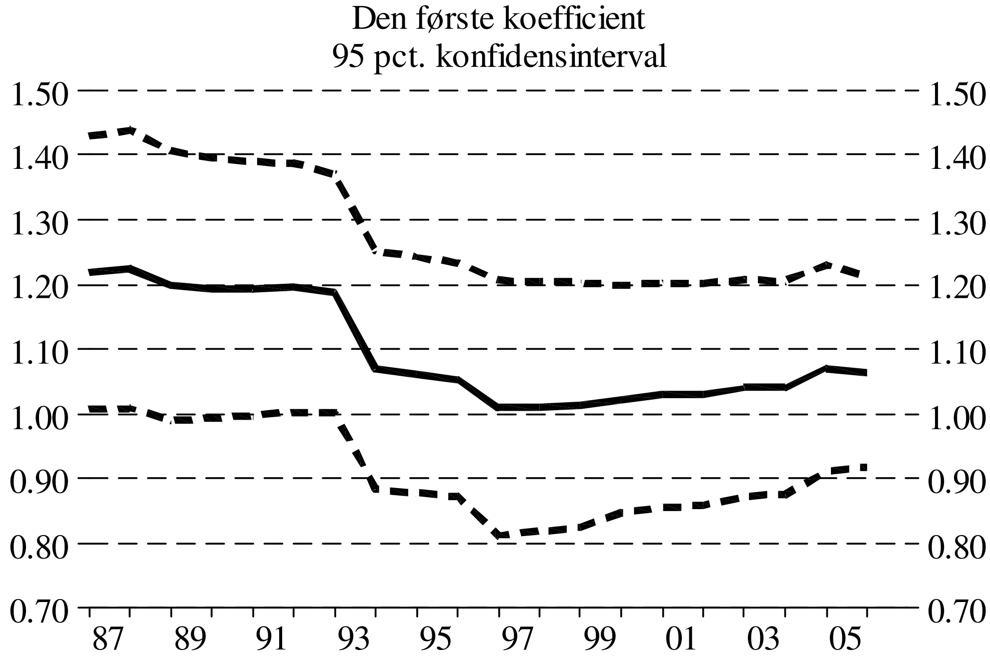
Et af de mest almindelige test er et Durbin-Watson test. Det kommer også automatisk ud med estimationsresultaterne, når man estimerer en ligning i Aremos. Man skal dog være opmærksom på, at Durbin-Watson teststørrelsen er invalid, når man inkluderer en lagget endogen variabel. Det er standard at benytte den laggede endogene i fejlkorrektionsmodeller, hvilket gør, at man aldrig i praksis bør benytte denne teststørrelse. Bilag A (boks 2) viser, hvordan man bør teste for autokorrelation i AREMOS.

Udover autokorrelation er det også en god idé at teste for korrekt funktionel form. Heteroskedasticitet er kun et problem, hvis man ønsker at vide noget om parameterusikkerheden for eksempel ved test af restriktioner på modellen. Forekommer heteroskedasticitet skal man rense ud for det ved eksplicit at modellere det og/eller benytte robuste fejlled alt efter årsagen. Test for normalitet er vigtigt, hvis man estimerer en fejlkorrektionsmodel med ML under antagelse af normalfordelte fejlled.

5. Parameterstabilitet

Et typisk problem er, at parametrene i ens opstillede model ikke er stabile over tid. Nogle gange afsløres det ved autokorrelation, men ikke altid. Man kan teste for parameterstabilitet ved at foretage forlæns og baglæns rekursiv estimation. Man reestimerer modellen med 1,2,… færre observationer henholdsvis først og sidst i perioden. Herefter plotter man alle estimaterne for hver parameter, hvis de estimerede parametre ligger indenfor det konfidensinterval, som er estimeret i den oprindelige model, så er parametrene stabile. I bilag A (boks 3) vises, hvordan man laver figur 9.1 og 9.2.

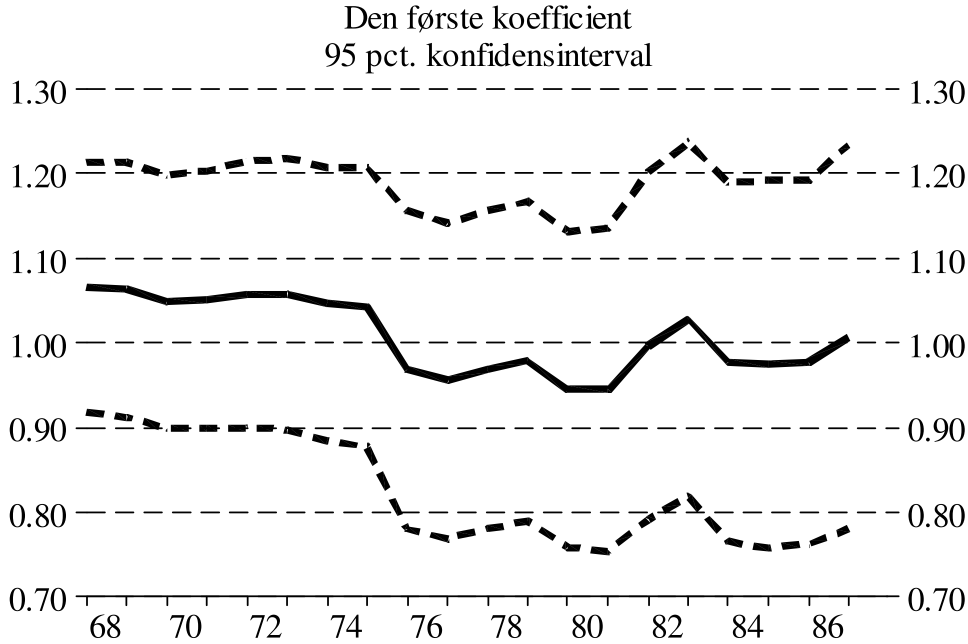
Figur 9.1. Rekursiv estimation.



Den første parameter tages som eksempel. Figur 9.1 viser rekursiv estimation af denne parameter med konfidensbånd, mens figur 9.2 viser den invers rekursive. X-aksen på figur 9.1 viser den sidste observation (der startes i 1968), mens X-aksen på figur 9.2 viser den første observation (der sluttes i 2006). Ifølge figur 9.1 ser der ud til at være et skift i parameteren omkring 1994, og parameterestimatet for 1968-87 er lige akkurat på grænsen af konfidensintervallet for hele samplet. Skiftet ligger dog inden for konfidensbåndet.

I nyeste version af ADAM har jeg bundet koefficienten til at være 1. Dette understøttes i nogen grad af figuren – måske ikke så meget når der estimeres på en helt kort periode, men her er der også kun lidt over 20 observationer. Figur 9.2 ser pænere ud. Her ser parameteren ud til at være helt stabil, uanset hvornår man starter estimationsperioden, så det største problem er den omtalte tendens til brud i 1994.

Figur 9.2. Invers rekursiv estimation



Man kan mere formelt teste for strukturelle brud med et Chow test, jf. Verbeek (2004) 3.3.3. Et Chow-test er et helt almindeligt F-test, hvor man tester om parametrene for de forklarende variabler er konstante før og efter et angivet brud. På baggrund af figur 9.1 ville man teste om, der er et strukturelt brud mellem 1993 og 1994.

Manglende parameterstabilitet kan skyldes en misspecificeret model eller et strukturelt brud, f.eks. pga. en grundlæggende ændring i den økonomiske politik, den tyske genforening eller lignende. Et strukturelt brud betyder, at parameteren har en anden værdi i en del af perioden. Sker der et skift i parameteren, som intutivt kan forklares og som kan virke fornuftigt, kan man indføre en dummy, som sikrer, at parameteren kan skifte værdi. Alternativt kan man vælge at estimere alene på perioden efter det strukturelle brud.

6. Dummy'er

Der er typisk tre situationer, hvor man vil overveje at bruge dummy'er:

1. Data er helt øjensynligt fejlbehæftet
2. Man ved apriori, at der er sket noget ekstræordinært
3. Der er en stor residual

Hvis data helt oplagt ser fejlbehæftet ud, så er data sandsynligvis fejlbehæftet. Fremgangsmåden for at afgøre dette er, at ringe/skrive til den ansvarlige person. Typisk er det en medarbejder fra Danmarks Statistik, som efter den første umiddelbare irritation over at have lavet en fejl, faktisk vil være glad for at få fejlen rettet. Godt råd: Vær høflig. Regn ikke med at fejlen bliver rettet umiddelbart, men hvis vedkommende medgiver, at det ser "mystisk" ud, så skal man i hvert fald indsætte en dummy.

Hvis der er sket en ekstraordinær begivenhed, kan den enten være indeholdt i ligningens forklarende variabler eller fremstå som en eksogen begivenhed. Er begivenheden eksogen, kan man benytte en dummy. Er residualen lille det pågældende år, og får man stort set samme parametre med og uden observationen, kan man overveje at undlade dummy'en.

Med årsdata har man ikke mange observationer, så det er kun, når det virker helt oplagt, at man skal benytte dummy'er til at fjerne store residualer.

7. Estimationsperiode

Typisk vil man ønske en så lang estimationsperiode som muligt. Dette skyldes både, at ens estimationsresultater vil blive mere præcise, samt at der vil være en vis "small sample bias" i modeller med laggede endogene variabler – jf. Nielsen (2010a). Man starter med det først mulige år og slutter med det sidst mulige år. Normalt mister man et par år i starten på grund af lags, og herudover kan det være nødvendigt at smide første del af en serie væk, hvis den skal være stationær i ligningen og ikke har været stationær i den første del af perioden. Denne estimationsteknik benyttes de fleste steder. Hermed får man også mest mulig ny information med i estimationen.

I nationalregnskabet offentliggøres de første tal primært på baggrund af indikatorer. Herefter oparbejdes og forbedres de, når der kommer flere og bedre kilder. I processen afstemmes de også, og der fejlsøges. Cirka 2 måneder efter udløbet af et kvartal offentliggøres første nationalregnskabstal, men først 3 år efter årets udgang er tallene "endelige". De endelige tal kan revideres ved større nationalregnskabsrevisioner, men de har en endelig kvalitet, og alle informationerne vedrørende de endelige tal er indhentet og bearbejdet.

ADAM er kun estimeret på endelige år. Dette betyder, at der er 2-3 observationer, som ikke benyttes. Der er tre grunde til dette. For det første kan man altid genskabe ADAMs ligninger fra de gældende data, så længe der ikke er større revisioner af Nationalregnskabet (og når der er større revisioner, så reestimeres ADAM). For det andet er datakvaliteten mindre i de foreløbige år. For det tredje kan man benytte de foreløbige år, som en out-of-sample periode til at se hvordan ens model klarer sig.

8. Modelvalg

Indtil nu har jeg gennemgået, hvordan man opstiller og estimerer fejlkorrektionsmodeller. Typisk kan man opstille en del modeller, som har stort set de samme forklarende variabler og samme karakteristika. Én model bruger forbrugerprisen, mens en anden model for det samme bruger forbrugerprisen uden boliger. Én model bruger usercost, mens en anden model bruger logaritmen til usercost. I én model er en koefficient restrikteret til at være en, mens den tilsvarende koefficient i en anden model er estimeret til 1,06. Hvordan vælger man mellem disse modeller?

Før man kan svare på spørgsmålet, skal man afgøre, om man kun ønsker en ligning der kan bruges til forecast, eller om ligningen skal indgå i en større sammenhæng. For eksempel om det skal være en ligning i ADAM. Skal den indgå i en større sammenhæng, skal man vælge ud fra hensyn til:

1. Teori: Den langsigtede ligevægtsrelation skal kunne udledes på baggrund af økonomisk teori.
2. Fortolkning: Parametrene skal være fortolkelige. Derfor er det ofte pænest at have dem i logaritmer. Endvidere er det ofte ønskværdigt at lægge (acceptable) restriktioner på parametrene.
3. Simplicitet: Inkluder kun variabler, som bidrager til at forklare de variabler, man interesserer sig for. Ikke alle signifikante variabler bidrager i størrelsesorden, så de af denne baggrund er værd at inkludere. (De kan dog være nødvendige som kontrolvariabler, hvis de ændrer parametrene på de interessante variabler).
4. Modelegenskaber: Mange lags i modellen kan give store og cykliske sving i modellen. Disse sving gør det sværere at arbejde med og fortolke modellen. Undgå derfor ekstra lags, hvis de ikke er strengt nødvendige på grund af for eksempel autokorrelation.

Man skal aldrig vælge en model alene efter R^2, og man skal heller ikke vælge en model alene efter andre statistiske udvælgningskriterier som for eksempel Akaike's. Selvfølgelig kan man skele til disse, men de bør ikke være den vigtigste faktor i ens afgørelse.

For en indikatormodel - som ikke skal indgå i en formel model, men derimod kun skal benyttes til at forecaste nogle få kvartaler frem - skal man i høj grad gå efter en høj forklaringsgrad. Det er dog ikke godt kun at se på R^2, da man så vil ende op med for mange irrelevante forklarende variabler. Man kan med fordel også benytte et udvælgelseskriterium – som for eksempel Aikake's, men bør kun inkludere variabler, som rent teoretisk bør være der. Med teoretisk forstås ikke nødvendigvis overensstemmende med optimerende agenter, men man skal kunne forklare hvorfor de inkluderede variabler kunne tænke at påvirke venstreside-variablen. Man kan sagtens inkludere adskillige lags, men man må ikke miste overblikket, hvilket betyder, at mange indikatormodeller i praksis er simple.

Det bør nævnes, at indikatormodeller oftest er på kvartaler eller måneder. Dette er også tilfældet, hvis hovedmodellen er en årsmodel. Baggrunden er, at de fleste indikatorer er på måneder og kvartaler. Det er nemt at sammenkæde de forudsagte måneder og kvartaler til at give et årstal.

9. Fejlkorrektionsmodellen med trendkorrektion

Tag udgangspunkt i ADAMs forbrugsrelation:

$$D\log\left(C_t\right)=\phi_1 D\log\left(Y_t\right)-\gamma\left(\log\left(C_{t-1}\right)-\log\left(C_{t-1}^*\right)\right)+gc+\varepsilon_t \tag{9.1}$$

hvor *C* er forbruget, *Y* er indkomsten, *gc* er trendkorrektionen, ε er fejlleddet, og C^* er ligevægtsniveauet for *C* givet ved:

$$\log\left(C_t^*\right)=\left(1-\beta_1\right)\log\left(Y_t\right)+\beta_1\log\left(W_t\right)+\beta_0 \tag{9.2}$$

hvor *W* er formuen. Overalt er græske bogstaver uden fodtegn parametre.

Antag, at indkomsten, *Y*, og formuen, *W*, stiger med 2 procent om året. I ligevægt skal forbruget også stige med 2 procent om året – jf. ligevægtsrelationen.

Antag, $\phi_1=0.4$, og at vi starter med *C*=C^*. Er der ingen trendkorrektion *gc*=0, så vil forbruget, *C*, umiddelbart stige med ϕ_1*0.02=0.4*2 procent=0.8 procent. Hermed vil *C* stige med mindre en C^*. Perioden efter vil der komme et positivt bidrag fra fejlkorrektionsleddet, og forbruget vil stige med mere end 0.8 procent, men mindre end 2 procent. Igen vil forskellen mellem *C* og C^* øges, hvilket igen betyder øget fejlkorrektion og øget stigning i *C*. Når forskellen er blevet så stor mellem *C* og C^*, at *C* stiger med 2 procent om året, så bibeholdes forholdet mellem *C* og C^*. Dette betyder, at uden trendkorrektion kan C^* ikke tolkes som ligevægtsniveauet for *C*.

I ovenstående eksempel kunne problemet være løst ved at sætte *gc* lig vækstraten i C i ligevægt fratrukket bidragene fra kortsigtsdynamikken: *gc*=0.02-0.4*0.02=0.012. I fremskrivninger sørger vi for at sætte *gc*, så den passer med ligevægtsvækstraten. For den historiske periode sætter vi den lig gennemsnittet over den estimerede periode *gc*=(gns. vækstrate i *C*)-0.4*(gns. vækstrate i *Y*).

Man kender typisk ikke parameteren til kortsigtsdynamikken, før man har estimeret ligningen. Derfor kan man ikke beregne *gc* før efter estimationen. I praksis vil man derfor estimere ligningen:

$$D\log\left(C_t\right)=\phi_1 D\log\left(Y_t\right)+\phi_2\log\left(\frac{C_{t-1}}{Y_{t-1}}\right)+\phi_3\log\left(\frac{W_{t-1}}{Y_{t-1}}\right)+\phi_0+\varepsilon_t \tag{9.3}$$

Parametrene til ADAMs ligninger findes ud fra simpel algebra som $\gamma=-\phi_2$, $\beta_1=\phi_3/\gamma$, $gc=\left(\sum D\log C_t \big/ T\right)-\phi_1\left(\sum D\log\left(Y_t\right)\big/ T\right)$ og $\beta_0=\left(\phi_0-gc\right)/\gamma$.

Alternativt til at beregne gc efter estimationen kan man detrende alle variabler i estimationen. Dette vil give samme resultat, men er lidt mere besværligt.

10. Konklusion

Der er mange overvejelser i forbindelse med estimation af en ligning, og denne note har kun berørt en del af dem. Forhåbentligt kan den hjælpe til estimation af simple fejlkorrektionsmodeller. Især håber jeg, at koden vil hjælpe til at få lidt bedre indsigt i Aremos og lette arbejdet med estimationen lidt. Er man vant til et andet estimationsredskab, så skal man dog ikke være bange for at benytte dette i stedet.

Literatur

- Kyhl, T. og J. Nielsen (2002), “Økonometriske værktøjer”, undervisningsnote.
- Nielsen, H. B. (2010a) ”Linear Regression with Time Series Data”, Quantitative Methods 3, Lecture Note 2
- Nielsen, H. B. (2010b) ”Dynamic Models for Stationary Time Series”, Quantitative Methods 3, Lecture Note 4
- Nielsen, H. B. (2010c) ”Non-Stationary Time Series and Unit Root Testing”, Quantitative Methods 3, Lecture Note 5
- Nielsen, H. B. (2010d) ”Cointegration and Common Trends”, Quantitative Methods 3, Lecture Note 6
- Verbeek, M. (2004), ”A Guide to Modern Econometrics”, John Wiley & Sons, Ltd

Bilag A – Eksempel på estimation i AREMOS

Fokus for de fleste makroøkonomiske størrelser er den årlige vækst i procent - vækstraterne. Dette gælder for eksempel BNP, forbrug, investeringer, import og eksport.

En simpel fejlkorrektionsmodel, som symboliserer dette, er materialeforbruget i diverse fremstillingserhverv. Den er (lidt omskrevet i forhold til ADAMs formelfil) givet ved:

$$\begin{aligned} D\log(fVmnz) &= D\log(fXnz) + gfvmnz \\ &\quad -0.63825\left(\log(fVmnz_{-1}) - \log(fVmnzw_{-1})\right) \end{aligned} \tag{9.4}$$

$$\log(fVmnzw) = -0.54197 + \log(fXnz) - \log(dtmnz) \tag{9.5}$$

hvor *fVmnz* er materialeforbruget i diverse fremstillingserhverv, *fXnz* er deres produktion, *gfvmnz* er trendkorrektionen, *fVmnzw* er ligevægtsrelationen, og *dtmnz* er en effektivitetstrend for materialer.

Når relationen er skrevet op i logaritmer, så kan man fortolke den som relative ændringer. Når produktionen øges med 1 procent, så øges materialeforbruget også umiddelbart med 1 procent. Dette er også tilfældet i ligevægt. Når materialer bliver 1 procent mere effektive, så vil man i ligevægt skulle bruge 1 procent færre materialer. Dette er dog en effekt, som kun slår igennem på langt sigt. Der er ingen virkning 1. år, men 2. år betyder det et fald på 0,63825 procent, 3. år er der et yderligere fald på (0,63825*(1-0,63825)) = 0,23 procent – og så videre. Fortolkningen er altså ret lige til, når modellen er skrevet op i logartimer.

Boks 1 - estimation i AREMOS:

En model som den ovenstående ville estimeres i Aremos ved:

$$\begin{aligned} D\log(fVmnz) &= \mu_0 + \mu_1 D\log(fXnz) + \mu_2 D\log(dtmnz) \\ &\quad + \mu_3 \log(fVmnz_{-1}) + \mu_4 \log(fXnz_{-1}) + \mu_5 \log(dtmnz_{-1}) \end{aligned} \tag{9.6}$$

Dette gøres ved at skrive:

```
close *;
clear work;
open adambk;
set per 1968 2006;
equation dlfvmnz dlog(fvmnz) = dlog(fxnz), dlog(dtmnz), log(fvmnz.1), log(fxnz.1),
log(dtmnz.1);
fit;
```

Outputtet bliver:

```
dlog(fvmnz)

  =   1.06508 * dlog(fxnz) + 0.29461 * dlog(dtmnz)
     (14.1695)            (0.37564)

    - 0.57854 * log(fvmnz.1) + 0.61028 * log(fxnz.1)
     (3.27424)              (3.52301)
```

- 0.46374 * log(dtmnz.1) - 0.71644
(2.02910) (1.81703)

Sum Sq 0.0073 Std Err 0.0148 LHS Mean 0.0280
R Sq 0.9124 R Bar Sq 0.8992 F 5, 33 68.7773
D.W.(1) 1.9839 D.W.(2) 2.0361

Det vigtigste at lægge mærke til er koeffcienterne og t-værdierne (der står i parentes under koeffcienterne). Endvidere kan man bemærke summen af de kvadrerede afvigelser (Sum Sq), standardafvigelsen (Std Err) og R^2 (R Sq).

Da t-værdien på andet led er betydeligt under 1.96, kan vi godt sætte denne størrelse lig 0. Ulempen ved, at Aremos angiver t-værdier, er, at man selv skal omregne til standardfejl på estimaterne. Den er 1,06508/14,1695 altså lidt over 0,07. Altså ligger koefficienten rigeligt under 1.96 standard-fejl af 1. Man kan ikke umiddelbart se om langsgtsrestriktionerne må påføres. Dog kunne det godt ud fra fortegnene og størrelsesforholdene se fornuftigt ud. Det bemærkes også, at alle koefficienter har de forventede fortegn.

R^2-værdien gemmes som en ny serie;
series rsq0 = dlfvmnz.stats[5];

For at påføre restriktioner i ligning (9.4) og (9.5) skal man teste:

$$H_0 : \mu_1 = 1 \wedge \mu_2 = 0 \wedge \mu_4 = -\mu_3 \wedge \mu_5 = \mu_3 \tag{9.7}$$

Restriktionerne påføres i Aremos ved:
impose 1,m,m,m,m,m,1;
impose m,1,m,m,m,m,0;
impose m,m,1,1,m,m,0;
impose m,m,1,m,-1,m,0;
fit;

Output bliver i dette tilfælde:
DLFVMNZ
Restricted Ordinary Least Squares
ANNUAL data for 39 periods from 1968 to 2006
Date: 22 JUL 2010

dlog(fvmnz)

= 1.00000 * dlog(fxnz) + 0.00000 * dlog(dtmnz)
(NC) (NC)

- 0.63828 * log(fvmnz.1) + 0.63828 * log(fxnz.1)
(4.61669) (4.61669)

- 0.63828 * log(dtmnz.1) - 0.34283
(4.61669) (4.57315)

Sum Sq 0.0075 Std Err 0.0143 LHS Mean 0.0280
R Sq 0.9092 R Bar Sq 0.9067 F 1, 37 370.315
D.W.(1) 1.9226 D.W.(2) 2.0714

Vi ser umiddelbart, at den estimerede tilpasningshastighed 0,63828 parameter stemmer rimelig godt overens med de 0,63825 fra modellen. Den lille forskel skyldes, at i estimationen til modellen er trenden endogen.

Man kan beregne F-teststørrelsen ved:

```
series rsq1 = dlfvmnz.stats5];
series Ftest = ((rsq1-rsq0)/4)/((1-rsq1)*(dlfvmnz.stats[15]-dlfvmnz.stats[14]);
! 4 ovenfor da 4 restriktioner
print Ftest;
```

Dette skal være lig med eller mindre end den kritiske værdi, så accepteres hypotesen.

Alternativt kan man have estimeret modellen i niveau:

```
equation lfvmnz log(fvmnz) = dlog(fxnz), dlog(dtmnz), log(fvmnz.1), log(fxnz.1), log(dtmnz.1);
fit;
```

Output giver:

```
LFVMNZ
 Ordinary Least Squares
 ANNUAL data for   39 periods from 1968 to 2006
 Date: 22 JUL 2010

 log(fvmnz)

  =   0.42146 * log(fvmnz)[-1] + 1.06508 * dlog(fxnz)
     (2.38522)                  (14.1695)

    + 0.29461 * dlog(dtmnz) + 0.61028 * log(fxnz.1)
     (0.37564)               (3.52301)

    - 0.46374 * log(dtmnz.1) - 0.71644
     (2.02910)                (1.81703)

 Sum Sq     0.0073   Std Err    0.0148   LHS Mean  11.8933
 R Sq       0.9978   R Bar Sq   0.9975   F  5, 33  3012.43
 D.W.( 1)   1.9839   D.W.( 2)   2.0361
```

Eneste forskel bør være, at koefficienten til log(fVmnz.1) er steget med 1. Kigger vi på koefficienterne er dette også den eneste forskel. Dog er R^2 steget fra 91,24 procent til 99,78 procent. Dette skyldes, at den laggede værdi i sig selv forklarer en stor del af niveauet i fVmnz, og her beskriver R^2, hvor godt man kan forudsige niveauet og ikke ændringen. Moralen er, at man ikke skal lade sig nare af en høj R^2 for niveaurelationer eller en lille R^2 for relationer, der beskriver ændringer i forhold/andele.

Boks 2 - autrokorrelationstest i AREMOS:

Opfølges eksemplet fra tidligere med estimation af materialeforbrug, så kan man i Aremos teste for autokorrelation med følgende tillægskode:

```
series res = dlfvmnz.residual;
equation LM res = dlog(fxnz), dlog(dtmnz), log(fvmnz.1), log(fxnz.1), log(dtmnz.1), res.1;
fit;
print LM.stats[5]*(2006-1968+1);
```

Outputtet bliver:

```
   ANNUAL Data for   39 periods from 1968  to 2006

LM
 Ordinary Least Squares
 ANNUAL data for   38 periods from 1969 to 2006
 Date: 23 JUL 2010

 res

  =    0.04965 * res[-1] + 0.06073 * dlog(fxnz)
      (0.09020)          (0.73956)

     + 0.33414 * dlog(dtmnz) + 0.01655 * log(fvmnz.1)
      (0.32908)              (0.03078)

     + 0.00941 * log(fxnz.1) + 0.11206 * log(dtmnz.1) - 0.32298
      (0.01804)              (0.18672)              (0.65931)

 Sum Sq     0.0073   Std Err    0.0153   LHS Mean   0.0002
 R Sq       0.0278   R Bar Sq  -0.1603   F  6, 31    0.1480
 D.W.( 1)   1.9408   D.W.( 2)   2.0312

                                   LM.stats[5
                                   ]*(2006-19
                                     68+1)


                            1968         1.09
```

I dette tilfælde er der ikke problemer med autokorrelation, da teststørrelsen er noget mindre end 3,84.

Boks 3 - test for parameterstabilitet i AREMOS:

Dette kan gøres ved i Aremos ved følgende kode:

```
equation xx dlog(fvmnz) = dlog(fxnz), dlog(dtmnz), log(fvmnz.1), log(fxnz.1), log(dtmnz.1);
!Rekursiv estimation
for i = 1987 to 2006;
 set per 1968 #i;
 fit <coeffs=y covar=yes>;
 ser <#i #i> koef1_h=xx.coeff[1,1];
 ser <#i #i> koef2_h=xx.coeff[2,1];
 ser <#i #i> koef3_h=xx.coeff[3,1];
 ser <#i #i> koef4_h=xx.coeff[4,1];
 ser <#i #i> koef5_h=xx.coeff[5,1];
 ser <#i #i> std1_h=xx.covar[1,1]**0.5;
 ser <#i #i> std2_h=xx.covar[2,2]**0.5;
 ser <#i #i> std3_h=xx.covar[3,3]**0.5;
 ser <#i #i> std4_h=xx.covar[4,4]**0.5;
 ser <#i #i> std5_h=xx.covar[5,5]**0.5;
end;

set per 1987 2006;
gra using #gdrev\rekur1;
gra using #gdrev\rekur2;
gra using #gdrev\rekur3;
gra using #gdrev\rekur4;
gra using #gdrev\rekur5;

!Rekursiv estimation - invers
```

```
for i = 1968 to 1987;
 set per #i 2006;
 fit <coeffs=y covar=yes>;
 ser <#i #i> koef1_h=xx.coeff[1,1];
 ser <#i #i> koef2_h=xx.coeff[2,1];
 ser <#i #i> koef3_h=xx.coeff[3,1];
 ser <#i #i> koef4_h=xx.coeff[4,1];
 ser <#i #i> koef5_h=xx.coeff[5,1];
 ser <#i #i> std1_h=xx.covar[1,1]**0.5;
 ser <#i #i> std2_h=xx.covar[2,2]**0.5;
 ser <#i #i> std3_h=xx.covar[3,3]**0.5;
 ser <#i #i> std4_h=xx.covar[4,4]**0.5;
 ser <#i #i> std5_h=xx.covar[5,5]**0.5;
end;

set per 1968 1987;
gra using #gdrev\rekur1;
gra using #gdrev\rekur2;
gra using #gdrev\rekur3;
gra using #gdrev\rekur4;
gra using #gdrev\rekur5;
```