

Christophe Kolodziejczyk & Nicolai Kristensen

Husholdningsbudgetberegner

En model for husholdningers dagligvareforbrug udarbejdet for
Penge- og Pensionspanelet



Publikationen *Husholdningsbudgetberegner – En model for husholdningers dagligvareforbrug* kan downloades fra instituttets hjemmeside

Det Nationale Analyse- og Forskningsinstitut for Kommuner og Regioner
Købmagergade 22
1150 København K
Telefon: 444 555 00

AKF, Anvendt KommunalForskning er pr. 1. juli 2012 lagt sammen med DSI og KREVI. Det nye, fusionerede institut hedder Det Nationale Analyse- og Forskningsinstitut for Kommuner og Regioner.

© 2012

Mindre uddrag, herunder figurer, tabeller og citater, er tilladt med tydelig kildeangivelse. Skrifter, der omtaler, anmelder, citerer eller henviser til nærværende, bedes sendt til Det Nationale Analyse- og Forskningsinstitut for Kommuner og Regioner.

© Omslag: Monokrom, Lars Degnbøl

Forlag: Det Nationale Analyse- og Forskningsinstitut for Kommuner og Regioner
ISBN: 978-87-7509-472-1
i:\08 sekretariat\forlaget\nik\5228\5228_notat_husholdningsberegner.docx
August 2012

Christophe Kolodziejczyk & Nicolai Kristensen

Husholdningsbudgetberegner

En model for husholdningers dagligvareforbrug

Forord

Dette notat dokumenterer beregningerne bag en model for husholdningers dagligvareforbrug. Modellen er udarbejdet for Penge- og Pensionspanelet ud fra et ønske om, at den skal danne grundlag for en webbaseret applikation, der giver alle husholdninger et værktøj til at sammenligne deres faktiske forbrug med et gennemsnitligt eller "forventet" forbrug. Notatet er udarbejdet af forsker Christophe Kolodziejczyk og programleder Nicolai Kristensen. Vi takker professor Carsten Tangaard og professor Søren Leth-Petersen for faglige kommentarer.

Nicolai Kristensen

August 2012

Indhold

1	Indledning	7
2	Samlet præsentation af modellen	8
2.1	Præsentation af den endelige model	8
2.2	Spørgsmål til brugerne	9
3	Data og beskrivende statistik	11
4	Modeller.....	15
5	Prædiktionsmodel.....	17
5.1	Metode.....	17
5.2	Resultater	18
5.3	Prædiktion og usikkerhedsanalyse	21
5.4	Regneeksempel	24
6	Litteratur	26

1 Indledning

Dette notat dokumenterer beregningerne bag en model for husholdningers forbrug af dagligvarer. Notatet indeholder dels en beskrivelse af data, dels modelberegninger, der viser parameterestimater og prædiktionsusikkerhed for fire alternative specifikationer af forbrugsmodellen.

Formålet er at estimere en model for husholdningers dagligvareforbrug, der kan bruges til at lave prædiktioner for forbruget for typiske husholdninger med bestemte karakteristika, som er nærmere specificeret nedenfor. Specifikt skal der ud fra data fra Forbrugsundersøgelsen 2009 estimeres en forbrugsfunktion, som kan bruges til at lave prædiktioner af værdien af forbruget i kroner for en husholdning med givne karakteristika. Data for forbrugsundersøgelsen 2009 er indsamlet over årene 2008-2010.

De fire alternative modeller afspejler varianter af modeller for Engel-kurver, dvs. modeller, der beskriver, hvordan husholdningsudgifter på udvalgte varer varierer med husstandens indkomst. Der er grundlæggende to varianter. I den ene variant anvendes budgetandele, dvs. her beskrives, hvordan *andelen*, der anvendes på de udvalgte varer, afhænger af husstandsindkomsten. Alternativt kan man beskrive Engel-kurver ved at undersøge, hvordan forbrugsniveauet afhænger af indkomstniveauet. Baseret på denne analyse anbefales model 3.

Igennem hele notatet henvises der til forbruget eller det udvalgte forbrug af dagligvarer. Dette inkluderer *husstandens samlede årlige forbrug af fødevarer, drikkevarer og husholdningsartikler, målt i kr.*

2 Samlet præsentation af modellen

I dette kapitel præsenteres den endelige model til at forudsige husholdningers forbrug. Detaljerede forklaringer af data, modeller og metode fremgår af de efterfølgende kapitler.

2.1 Præsentation af den endelige model

Ud af de fire modeller, der er estimeret, anbefales model 3. Modellen er for forbruget i niveau og er specificeret således:

Model 3

$$\begin{aligned}\widehat{\ln C}_i = & \beta_0 + \beta_2 \times Enlig_i \\ & + \beta_3 \times \max(Alder1_i, Alder2_i) + \beta_4 (\max(Alder1_i, Alder2_i))^2 \\ & + \beta_{10} Nborn_5_i + \beta_{11} Nborn_611_i + \beta_{12} Nborn_1217_i + \beta_{13} Nborn_18_i \\ & + \beta_{14} \ln(Y_i) + \beta_{15} (\ln Y_i)^2 \\ & + \beta_{18} Dejer_i + \beta_{19} Dudderhv_i + \beta_{20} Duddkort_i + \\ & + \beta_{21} Duudml_i + \beta_{22} Dlang_i\end{aligned}$$

Den afhængige variabel er den naturlige logaritme til husstandens forbrug på udvalgte varegrupper, se kapitel 3 for detaljer. Indekset i henviser til den enkelte husstand. I modellen indgår:

- **Enlig:** En indikatorvariabel for, om man bor alene eller ej.
- **Max alder:** Den højeste alder i husstanden. Indgår desuden kvadreret.
- **Nborn:** Antallet af børn der bor hjemme. Opgøres på fire aldersgrupper.
 - 0-5 år, 6-11 år, 12-17 år, 18 år og derover.
- **Logaritmen til husstandens samlede disponible indkomst.** Indgår også kvadreret. Her inkluderes alle former for indkomst efter skat, såsom lønindkomst, overførselsindkomst, børnecheck mv.
- **Ejrbolig:** En indikatorvariabel for, om husstanden bor i ejerbolig eller ej.
- **Uddannelse:** Indikatorvariable for, hvad den højeste fuldførte uddannelse i husstanden er. Der indgår fem uddannelsesniveauer og fire indikatorvariable, idet referencegruppen er uddannelse på grundniveau (folkeskole samt gymnasialt niveau). De fire indikatorvariable inkluderer:
 - Erhvervsuddannelse; kort videregående; mellemlang videregående og lang videregående uddannelse.

Modellens estimerede parametre er angivet i tabel 2.1.

Tabel 2.1 Parameterestimer for den valgte model, model (3)

In(C _i)	Koeff.	Std. Err.	t	
Enlig	-0,353	0,045	-7,83	***
Maxalder	0,038	0,005	7,40	***
Maxalder ²	-0,0003	0,00005	-6,95	***
Antal børn				
0-5 år	0,115	0,028	4,17	***
6-11 år	0,111	0,026	4,34	***
12-17 år	0,138	0,022	6,23	***
18+	0,155	0,029	5,43	***
ln(hushold. disponible indkomst)	-1,030	0,491	-2,10	**
(ln(hushold. disponible indkomst)) ²	0,049	0,019	2,55	**
Ejerbolig	0,046	0,029	1,58	
Uddannelse				
kort videregående	0,017	0,067	0,25	
erhvervsuddannelse	0,037	0,040	0,93	
mellemlang videregående	0,121	0,044	2,74	***
lang videregående	0,155	0,046	3,36	***
konstantled	14,668	3,145	4,66	***

Anm.: *** angiver signifikans på 1% (dvs. meget signifikant); ** på 5%. Insignifikante parameterestimer er angivet uden stjerne.

2.2 Spørgsmål til brugerne

For at beregne det forventede forbrug skal brugeren indtaste oplysninger om husstandens sammensætning. Givet at den endelige model er valgt til at være model 3, er der behov for svar på spørgsmål omkring de variable, der indgår i denne model.

Husholdningsbudgetberegneren skal udarbejdes således, at brugerne i første omgang bliver bedt om at indtaste følgende:

Indtast antal voksne i husstanden (hjemmeboende børn på 18 år og derover skal ikke medtages som voksne):

TAST TAL

[Hvis der tages >0 popper følgende bokse op (Hvis der tages 1, popper 1 boks op, hvis der tages 2, popper 2 bokse op etc.)

Skriv alderen på voksen 1:

Skriv alderen på voksen 2:

Er der hjemmeboende børn i husstanden? (Medregn også hjemmeboende "børn" på 18 år og derover)

Ja
nej

[Hvis der tages ja, popper disse bokse op]

Antal hjemmeboende børn i alderen 0-5 år: TAST TAL
Antal hjemmeboende børn i alderen 6-11 år: TAST TAL
Antal hjemmeboende børn i alderen 12-17 år: TAST TAL
Antal hjemmeboende børn i alderen 18+ år: TAST TAL

Skriv husstandens **årlige indkomst efter** skat. Du skal medtage alle former for indkomst, både lønindkomst, overførselsindkomst, børnecheck mv.:

Skriv tal _____ kr. pr. år

Bor husstanden i ejerbolig?

Ja
Nej

Hvad er den højeste gennemførte uddannelse blandt personer i husstanden?

- Folkeskole/gymnasium
- Kort videregående uddannelse
- Erhvervsuddannelse
- Mellemlang videregående uddannelse
- Lang videregående uddannelse

Når brugeren har indtastet oplysninger om vedkommendes husstand, så kan beregneren vise, hvor mange penge sammenlignelige husstande samlet set bruger på de udvalgte fødevarer, drikkevarer og husholdningsartikler. Med sammenlignelige husstande menes husstande, som både har samme indkomst og uddannelseskategori, samme antal voksne i samme alderskategorier og samme antal børn i samme alderskategorier.

I afsnit 5.4 vises et detaljeret regneeksempel på, hvordan det forventede forbrug beregnes på baggrund af de estimerede værdier.

3 Data og beskrivende statistik

Data stammer fra forbrugsundersøgelsen 2009, suppleret med registeroplysninger om indkomst, overførsler mv. Forbrugsundersøgelsen er gennemført som en stikprøveundersøgelse, hvor antallet af husstande i undersøgelsen udgør ca. 0,1% af Danmarks i alt ca. 2,5 mio husstande.

Forbrugsundersøgelsen bygger dels på spørgeskemadata dels på regnskabsføring af de enkelte husholdninger, der over en 14 dages periode holder et meget detaljeret regnskab med indkøb. Data for forbrugsundersøgelsen 2009 er indsamlet over årene 2008-2010. For årene 2008 og 2010 er data omregnet ved brug af en pris-mængde korrektion, der gør oplysningerne sammenlignelige med 2009 som basisår. Datasættet indeholder også et sæt vægte, der er baseret på alle tre år. Disse vægte sikrer, at undersøgelsen er repræsentativ for samtlige husstande i Danmark.

Data setup'et fra Danmarks Statistik er lidt særegent for disse data. De rå data ligger i filen `work.dta`.¹ Der er i alt 2.484 husstande i datasættet. Disse ligger i én lang "stacket" fil (`work.dta`), hvor hver husstand har ca. 900 rækker (husstandens "identifer" er variabelen `ip`). For hver række er angivet et beløb (variabelnavn `belob`) og en kode (variabelnavn `kodec`).² Kodec indeholder koder for både forbrug samt baggrundsvariable, der er koblet på forbrugsundersøgelsen.

Tabel 3.1 Struktur på rådata fra Danmarks Statistik

linje	ip	kodec	belob
1	1	se link til regneark med koder	
2	1		
...	...		
903	1		
904	2		
405	2		
906	2		
osv.	osv.		

I realiteten er `ip` en `string12`, så ingen husstand har koden 1, men strukturen på data er som angivet i tabel 3.1: hver husstand har ca. 900 linjer (i eksemplet her har husstand 1 903), og hver linje svarer til en bestemt variabel. En mere traditionel struktur på datasættet opnås ved at vælge de variable, vi skal bruge til at lave estimationerne og dernæst bruge kommandoen `reshape`, som gør, at hver husstand blot har én linje, og vi i stedet får flere variable. I det tilhørende stata-program konstruerer `_data.do` fremgår det præcist, hvordan dette er gjort. Fremtidige opdateringer af beregningerne kan udføres ved ganske simpelt at køre dette pro-

¹ Se filen `konstruerer_data.do`, som indeholder stata-koden til at sætte data op.

² Se alle koderne i `regneark`, der kan downloades fra <http://www.dst.dk/da/Statistik/emner/forbrug/forbrugsundersogelsen.aspx?tab=dok#coicop>.

gram, forudsat at datastrukturen og koderne i rådata fra Danmarks Statistik forbliver uændrede.³

Der er som nævnt knyttet en række baggrundsoplysninger til forbrugsdataene. Til budgetberegneren inddrages kun relativt få, idet brugerne af modellen efterfølgende skal indtaste oplysninger for alle variable, der indgår, og for at mindske omfanget af indtastningen inkluderes derfor kun de variable, der på forhånd skønnes mest væsentlige. Tabel 3.2 på næste side angiver de variable, der indgår i modellerne.

³ Danmarks Statistik afleverer data i SAS-format. For at transformere til stas dta-format har vi anvendt StatTransfer. Bemærk, at det er nødvendigt at bruge optionen "optimize" for at sikre en korrekt transfer.

Tabel 3.2 Variabelbeskrivelse

Variabel-navn	Variabelbeskrivelse
C_i	Samlet årligt forbrug af fødevarer, drikkevarer og husholdningsartikler målt i kr. Det defineres ud fra følgende numre i forbrugsundersøgelsen: 11+12 +21+5611 +5612 +9913+9341 ⁴
Y_i	Den årlige indkomst efter skat. Her inkluderes alle former for indkomst, både løn-indkomst, overførselsindkomst, børnecheck mv.
Y1	Indikatorvariabel=1, hvis husstandens indkomst <C1, 0 ellers
Y2	Indikatorvariabel=1, hvis husstandens indkomst C1 <=y<C2, 0 ellers
Y3	Indikatorvariabel=1, hvis husstandens indkomst C2 <=y<C3, 0 ellers
Y4	Indikatorvariabel=1, hvis husstandens indkomst C3 <=y<C4, 0 ellers
Y5	Indikatorvariabel=1, hvis husstandens indkomst y>C4, 0 ellers
NumPers	Antal personer i husstanden
Alder1	Alder på voksen person 1
Numborn	Antal børn
DAlder1_29	Indikatorvariabel=1, hvis voksen person 1 er yngre end 30 år, 0 ellers
DAlder1_3044	Indikatorvariabel=1, hvis voksen person 1 er 30-44 år gammel, 0 ellers
Dalder1_4559	Indikatorvariabel=1, hvis voksen person 1 er 45-59 år gammel (referencegruppe), 0 ellers
Dalder1_60_	Indikatorvariabel=1, hvis voksen person 1 er ældre end 59 år, 0 ellers
DVoksen2	Indikatorvariabel=1, hvis der er to voksne i husstanden, 0 ellers
Alder2	Alder på voksen person 2
DAlder2_29	Indikatorvariabel=1, hvis voksen person 2 er yngre end 30 år, 0 ellers
DAlder2_3044	Indikatorvariabel=1, hvis voksen person 2 er 30-44 år gammel, 0 ellers
Dalder2_4559	Indikatorvariabel=1, hvis voksen person 2 er 45-59 år gammel (referencegruppe), 0 ellers
Dalder2_60_	Indikatorvariabel=1, hvis voksen person 2 er ældre end 59 år, 0 ellers
DBorn	Indikatorvariabel=1, hvis hjemmeboende børn i husstanden, 0 ellers
Nborn_5	Antal hjemmeboende børn i alderen 0-5 år
Nborn_611	Antal hjemmeboende børn i alderen 6-11 år
Nborn_1217	Antal hjemmeboende børn i alderen 12-17 år
Nborn_18_	Antal hjemmeboende børn i alderen 18 og ældre
Dejer	=1, hvis ejer af bolig, 0 ellers
Dudderhv	=1, hvis højeste gennemførte uddannelse er erhvervsfaglig, 0 ellers
Duddkort	=1, hvis højeste gennemførte uddannelse i husstanden er kort, 0 ellers
Duddml	=1, hvis højeste gennemførte uddannelse i husstanden er mellemlang, 0 ellers
Duddlang	=1, hvis højeste gennemførte uddannelse i husstanden er lang, 0 ellers

Y1-Y5 er funktioner af en række tærskelværdier, C1-C4. Disse er fastsat ud fra variationen i data. Vi finder følgende percentiler for den disponible husstandsindkomst, se tabel 3.3.

⁴ Forbruget er grupperet i henhold til den såkaldte COICOP-nomenklatur. Se nummereringsliste på <http://www.dst.dk/ext/171184426>

Tabel 3.3 Fordeling af disponible husstandsindkomst

Percentil	Indkomstgrænse
1%	65.587
5%	113.625
10%	141.140
25%	214.988
50%	344.536
75%	526.600
90%	679.372
95%	818.862
99%	1.258.600
Gennemsnit	404.078
Std. Afv.	321.586
Observationer	2.474

Ud fra ovenstående fordeling er C1-C4 grænseværdierne, der adskiller Indikatorvariabelvariablene, valgt som angivet i tabel 3.4.

Tabel 3.4 Valgte grænseværdier for indkomstdummies

	grænseværdi	Interval	% husstande
C1	150.000	<150.000	11
C2	250.000	150.000-250.000	21
C3	350.000	250.000-350.000	18
C4	550.000	350.000-550.000	27
		>550.000	22

De fem grupper, der kommer ud af de fire grænseværdier, er ikke lige store, og navnlig er der relativt få husstande i den laveste indkomstgruppe (11%), men den laveste grænseværdi på 150.000 kr. i disponibel husstandsindkomst synes relativt høj, og det vil måske sende et uhensigtsmæssigt signal om, at beregneren ikke er tiltænkt alle indkomstgrupper, hvis den sættes endnu højere blot for mere præcist at få 20% i hver af de fem grupper. Derudover er grænseværdierne valgt som "runde tal" for kognitivt at gøre indtastningen mere simpel.

4 Modeller

Der estimeres fire forskellige varianter af en forbrugsmodel. De fire modeller adskiller sig primært ved ændringer i den afhængige variabel, dvs. variabelen, der måler forbruget. Derudover varierer de fire modeller også lidt i forhold til specifikationen af de forklarende variable (men ikke i forhold til hvad de forklarende variable grundlæggende måler, eksempelvis måles indkomst i alle fire modeller, men indgår som Indikatorvariabelvariable i model 1 og 2 og logaritmisk i model 3 og 4).

De fire modelvarianter fremgår i formaliseret form nedenfor. Model 1 inkluderer niveauet af forbrug som den afhængige variabel. I model 2 og 3 indgår forbruget i logaritmisk form, mens model 4 anvender forbrugstilbøjeligheden som afhængig variabel, det vil sige forbrugets andel i forhold til den samlede disponible husstandsindkomst.

Model 1

$$\begin{aligned} C_i = & \beta_0 + \beta_2 \text{Enlig}_i + \\ & + \beta_3 \text{DAlder1}_{-29}_i + \beta_4 \text{DAlder1}_{-3044}_i + \beta_5 \text{DAlder1}_{-60}_{-i} \\ & + \beta_6 \text{DAlder2}_{-29}_i + \beta_7 \text{DAlder2}_{-3044}_i + \beta_8 \text{DAlder2}_{-60}_{-i} \\ & + \beta_9 \text{Nborn}_{-5}_i + \beta_{10} \text{Nborn}_{-611}_i + \beta_{11} \text{Nborn}_{-1217}_i + \beta_{12} \text{Nborn}_{-18}_{-i} \\ & + \beta_{13} Y1_i + \beta_{14} Y2_i + \beta_{15} Y4_i + \beta_{16} Y5_i \\ & + \beta_{18} \text{Dejer}_i + \beta_{19} \text{Dudderhv}_i + \beta_{20} \text{Duddkort}_i + \beta_{21} \text{Duudml}_i + \beta_{22} \text{Dlang}_i \end{aligned}$$

Model 2

$$\begin{aligned} \ln C_i = & \beta_0 + \beta_2 \text{Enlig}_i + \\ & + \beta_3 \text{DAlder1}_{-29}_i + \beta_4 \text{DAlder1}_{-3044}_i + \beta_5 \text{DAlder1}_{-60}_{-i} \\ & + \beta_6 \text{DAlder2}_{-29}_i + \beta_7 \text{DAlder2}_{-3044}_i + \beta_8 \text{DAlder2}_{-60}_{-i} \\ & + \beta_9 \text{Nborn}_{-5}_i + \beta_{10} \text{Nborn}_{-611}_i + \beta_{11} \text{Nborn}_{-1217}_i + \beta_{12} \text{Nborn}_{-18}_{-i} \\ & + \beta_{13} Y1_i + \beta_{14} Y2_i + \beta_{15} Y4_i + \beta_{16} Y5_i \\ & + \beta_{18} \text{Dejer}_i + \beta_{19} \text{Dudderhv}_i + \beta_{20} \text{Duddkort}_i + \beta_{21} \text{Duudml}_i + \beta_{22} \text{Dlang}_i \end{aligned}$$

Model 3

$$\begin{aligned} \widehat{\ln C}_i = & \beta_0 + \beta_2 \times \text{Enlig}_i \\ & + \beta_3 \times \max(\text{Alder1}_i, \text{Alder2}_i) + \beta_4 (\max(\text{Alder1}_i, \text{Alder2}_i))^2 \\ & + \beta_{10} \text{Nborn}_{-5}_i + \beta_{11} \text{Nborn}_{-611}_i + \beta_{12} \text{Nborn}_{-1217}_i + \beta_{13} \text{Nborn}_{-18}_{-i} \\ & + \beta_{14} \ln(Y_i) + \beta_{15} (\ln Y_i)^2 \\ & + \beta_{18} \text{Dejer}_i + \beta_{19} \text{Dudderhv}_i + \beta_{20} \text{Duddkort}_i + \\ & + \beta_{21} \text{Duudml}_i + \beta_{22} \text{Dlang}_i \end{aligned}$$

Model 4

$$\begin{aligned} \frac{C_i}{Y_i} = & \beta_0 + \beta_2 \text{Enlig}_i + \\ & + \beta_3 \max(\text{Alder1}_i, \text{Alder2}_i) + \beta_4 (\max(\text{Alder1}_i, \text{Alder2}_i))^2 \\ & + \beta_{10} \text{Nborn}_5_i + \beta_{11} \text{Nborn}_611_i + \beta_{12} \text{Nborn}_1217_i + \beta_{13} \text{Nborn}_18_i \\ & + \beta_{14} \ln(Y_i) + \beta_{15} (\ln Y_i)^2 \\ & + \beta_{18} \text{Dejer}_i + \beta_{19} \text{Dudderhv}_i + \beta_{20} \text{Duddkort}_i + \beta_{21} \text{Duudml}_i + \beta_{22} \text{Dlang}_i \end{aligned}$$

Hvor $i = 1, \dots, N$ angiver observationer (husholdninger) i forbrugsundersøgelsen. Alle modeller estimeres med Ordinary Least Squares (OLS). Referencegrupperne er således valgt til at være DAlder1_4559, DAlder2_4559, Y3 samt almen uddannelse (dvs. folkeskole eller gymnasial uddannelse som højeste fuldførte uddannelse). De medtagne forklarende variable skal fortolkes i forhold til den valgte referencegruppe. Valg af referencegruppe er arbitrært. Ved fx at vælge Y3 som reference i model 1 og model 2 vil vi forvente, at parameterestimererne til Y1 og Y2 vil have et negativt fortegn, og at parameterestimererne til Y4 og Y5 vil have et positivt fortegn (men man kunne godt vælge en anden referencegruppe).

5 Prædiktionsmodel

5.1 Metode

Ud fra de estimerede forbrugsrelationer, modellerne 1-4 skal der konstrueres en prædiktionsmodel, som kan forudsige det forventede forbrug (målt i kroner) for husholdninger med givne karakteristika, der ligger inden for kombinationen af de variable, der indgår i de estimerede modeller.

Ud over at prædiktere det forventede forbrug beregnes der også konfidensintervaller for prædiktionerne generelt for hele stikprøven samt for udvalgte kombinationer af de forklarende variable. Dette gøres for at afklare, hvor sikre (eller usikre) de forventede forbrugsniveauer er, herunder for hvilke kombinationer af de forklarende variable, som indgår i modellerne, at de estimerede modeller giver meget usikre prædiktioner. Det er også et formål at afklare, hvilken af de ovennævnte specifikationer, 1-4, der giver de bedste/mest præcise prædiktioner.

Beregningerne af forventet forbrug og konfidensintervaller for prædiktionerne følger Wooldridge (2009: 206-212). Nogle få forhold ved disse beregninger skal nævnes her.

Usikkerheden ved beregninger af konfidensintervaller kan opdeles i to:

- (i) *Sampling error*, som kommer af, at parameterestimererne er estimerede. Denne fejlkilde afhænger af stikprøvestørrelsen og mindskes ved øget stikprøvestørrelse, så denne kilde til usikkerhed vil forventelig være beskedent givet stikprøvestørrelsen på 2.474 husstande.
- (ii) *Variansen på fejleddet*, som afspejler variansen i populationen. Denne fejlkilde er uafhængig af stikprøvestørrelsen og vil ofte dominere i forhold til sampling-error. Når vi medtager variansen i populationen, tager vi implicit højde for variansen i uobserverede faktorer, der påvirker forbruget. Da modellerne 1-4 inkluderer relativt få parametre (og kun bygger på en enkelt cross-section observation), må vi forvente, at denne usikkerhed vil være ganske stor. Dette analyseres i afsnit 5.3.

Et andet forhold, der kort skal nævnes, er beregningen af det forventede forbrug, når $\ln C_i$ er den afhængige variabel (dvs model 2 og 3). I dette tilfælde kan vi ikke nøjes med blot at beregne $\widehat{C}_i = \exp(\widehat{\ln C}_i)$, idet dette systematisk vil underestimere det forventede forbrug. I stedet omregnes til det forventede forbrug ved at anvende formlen

$$\widehat{C}_i = \exp(\widehat{\ln C}_i) \times \exp(\hat{\sigma}^2 / 2),$$

hvor $\hat{\sigma}^2$ er et estimat af variansen for fejleddet fra populationen, kaldet u . Denne korrektion er imidlertid følsom over for afvigelser fra en underliggende antagelse om normalitet i

fejleddet. Derfor er der korrigeret for dette ved at estimere en såkaldt auxiliary regression. Udgangspunktet for denne ekstra regression er

$$\widehat{C}_i = \exp(\ln \widehat{C}_i) \times \alpha_0,$$

hvor α_0 er den forventede værdi af $\exp(u)$. For at anvende denne tilgang skal α_0 estimeres. Dette gøres her ved at anvende Wooldridge (2009: 212)'s formel (6.44), dvs. estimere følgende OLS uden et konstantled

$$\widehat{\alpha}_i = \left(\sum_{i=1}^n \widehat{m}_i^2 \right)^{-1} \left(\sum_{i=1}^n \widehat{m}_i C_i \right),$$

hvor \widehat{m}_i beregnes ved at indsætte de estimerede $\widehat{\beta}$ -værdier fra regressionen $\ln C_i = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$, så vi opnår $\widehat{m}_i = \exp(\ln \widehat{C}_i)$.

Efterfølgende sammenligninger med de faktiske forbrugsniveauer viste, at dette førte til en (overraskende) stor ændring, hvilket indikerer, at normalfordelingsantagelsen ikke holder i disse data.

En omregning er også relevant for model 4. Her beregnes først den forventede forbrugstilbøjelighed som efterfølgende multipliceres med husholdningens indkomst.

5.2 Resultater

Parameterestimerne for de fire modeller er angivet i tabel 5.1.

Tabel 5.1 Parameterestimer fra OLS-beregninger, model (1) – (4)

	(1) C_i	(2) lnC_i	(3) lnC_i	(4) C_Y
Denlig	-13288.0*** (-4.27)	-0.424*** (-6.60)	-0.353*** (-7.83)	-0.0622*** (-6.46)
DAlder1_29	-6587.5*** (-4.27)	-0.294*** (-5.51)		
DAlder1_3044	-254.5 (-0.15)	-0.0424 (-1.05)		
DAlder1_60_	-1771.5 (-1.18)	-0.0839 (-1.83)		
DAlder2_29	-9256.3*** (-3.75)	-0.0791 (-1.45)		
DAlder2_3044	-3620.0 (-1.46)	-0.0260 (-0.60)		
DAlder2_60_	-3377.1 (-1.37)	-0.00520 (-0.10)		

	(1) C_i	(2) lnC_i	(3) lnC_i	(4) C_Y
Maxalder			0.0385*** (7.40)	0.00653*** (6.88)
Maxalder^2			-0.00034*** (-6.95)	-0.000058*** (-6.67)
Nborn_5	3999.9* (2.30)	0.0762** (2.64)	0.115*** (4.17)	0.0160*** (3.93)
Nborn_611	7159.4*** (3.89)	0.106*** (4.11)	0.111*** (4.34)	0.0120** (2.83)
Nborn_1217	9499.8*** (6.49)	0.136*** (6.14)	0.138*** (6.23)	0.0147*** (3.98)
Nborn_18_	7116.3*** (3.82)	0.126*** (4.22)	0.155*** (5.43)	0.0110* (2.07)
Y1	-2866.6 (-1.85)	-0.113 (-1.82)		
Y2	-1453.9 (-1.04)	-0.0526 (-1.16)		
Y4	6236.1*** (3.82)	0.143*** (3.63)		
Y5	16391.1*** (7.43)	0.288*** (6.95)		
lnY_i			-1.030* (-2.10)	-1.261*** (-4.64)
lnY_i2			0.0487* (2.55)	0.0447*** (4.28)
Dejer	2713.5* (2.34)	0.0391 (1.28)	0.0463 (1.58)	0.0187** (3.10)
Duddkort	2198.4 (0.88)	0.0483 (0.71)	0.0168 (0.25)	0.0115 (1.13)
Dudderhv	364.6 (0.29)	0.0663 (1.61)	0.0371 (0.93)	0.00978 (1.71)
Duddml	4821.0** (2.99)	0.147** (3.22)	0.121** (2.74)	0.0245** (3.28)
Duddlang	6621.5** (2.85)	0.168*** (3.57)	0.155*** (3.36)	0.0256*** (3.43)
Constant	39829.4*** (12.25)	10.48*** (157.96)	14.67*** (4.66)	8.737*** (5.00)
N	2474	2474	2474	2474

t statistik i parentes. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Referencegruppen for uddannelse er "almen uddannelse", som inkluderer ikke-kompetencegivende uddannelse inklusive folkeskolen samt gymnasiale uddannelser. Referencegruppen for DAlder er aldersgruppen 45-59, mens referencegruppen for Y1-Y5 er Y3.

Resultaterne viser generelt de forventede fortegn og forventede signifikansniveauer. Indikatorvariabelvariablen for enlig (Denlig) er negativ og signifikant på 1%-niveau (dvs. meget signifikant) i alle fire modeller. Model 1 viser, at enlige, alt andet lige, bruger ca. 13.000 kr. mindre om året på de udvalgte varer sammenlignet med ikke-enlige.

Alder indgår på forskellig vis i de fire modeller. I model 1 og 2 medtages tre Indikatorvariabelvariable for person 1, og i de tilfælde hvor der er mere end én voksen i husstanden, medtages alderen for person 2 også. Referencegruppen er 45-59 årige, og det er derfor ikke overraskende, at forbruget for den yngste aldersgruppe, personer yngre end 30 år, er signifikant lavere end for referencegruppen. Dette gælder både for person 1- og person 2-estimatene. De øvrige aldersgrupper er ikke signifikant forskellige fra referencegruppen.

I model 3 og 4 inkluderes i stedet alderen for den person i husstanden, der har den højeste alder og denne alder kvadreret (kaldet maxalder og maxalder²). Det leder til meget signifikante parameterestimer, der viser, at forbruget stiger med alderen, men med en aftagende grad. Toppunktet beregnes til 111 henholdsvis 112 år, hvilket betyder, at forbruget er stigende med alderen inden for observationsområdet.

Antallet af hjemmeboende børn vil forventeligt have en meget stor betydning for en husstands forbrug af dagligvarer, og denne forventning bekræftes da også af parameterestimerne for alle fire modeller. Forbruget er 4.000-9.500 kr. højere pr. år (model 1) for hvert barn afhængigt af barnets alder. For modellerne 1, 2 og 4 findes, at et barn i alderen 12-17 har den største betydning for forbruget, mens model 3 finder børn over 17 øger forbruget mest, alt andet lige. Samtlige parameterestimer for hjemmeboende børn er signifikante, og de fleste er signifikante på 1%-niveau.

Husstandsindkomsten er ligeledes stærkt korreleret med forbruget. De fire modeller varierer, med hensyn til hvordan husstandsindkomsten indgår. I model 1 og 2 indgår fire dummys som evalueres i forhold til referenceindkomsten Y_3 , dvs. husstande der har 250.000-350.000 i disponibel indkomst. Parameterestimerne for både model 1 og 2 viser, at indkomstniveauer der ligger lavere end referencegruppens ikke har et signifikant lavere forbrug. Dette er konsistent med, at mange dagligdagsvarer kan opfattes som nødvendige forbrugsgoder, der nok kan substitueres med billigere produkter, men i mange tilfælde ikke fravælges helt. Bemærk også, at parameterestimerne for Y_1 og Y_2 er negative, og Y_1 er signifikant lavere hvis man anvender et 10%-signifikansniveau. Husstande med disponible indkomster, der er højere end referencegruppen ($Y_4 = 350.000-550.000$ og $Y_5 > 550.000$), har signifikant højere forbrug, selv om stigningen ikke er lige så markant som forskellen i disponibel indkomst. Dette kan formentlig igen til dels tilskrives, at vi her ser på dagligvareforbruget, som i høj grad adskiller sig ved substitution mellem produkter i forskellige prisklasser.

I modellerne 3 og 4 bekræftes ovenstående fortolkning. I disse modeller indgår forbrugstilbøjeligheden som den afhængige variabel, og vi finder, at både $\ln Y_i$ og $(\ln Y_i)^2$ er yderst signifikante. Dette bekræfter resultater fra litteraturen på området, fx Banks et al. (1997), om at Engel-kurver bør inkludere et kvadratisk led, således at man tillader "Engel curvature".

Vi genfinder her det klassiske Engel-resultat, at forbrugstilbøjeligheden er faldende med indkomsten. Dette er konsistent med Engel's lov om, at lavindkomstfamilier anvender en større budgetandel på fødevarer.

Modellerne inkluderer også en indikatorvariabel for, om personen bor i ejerbolig eller ikke. I modellerne 1 og 4 kommer parameterestimatet for denne variabel ud signifikant, mens den er insignifikant for model 2 og 3. Insignifikansen skyldes nok til dels, at ejerbolig også er

korreleret med indkomst og uddannelse. Variablen vil desuden opsamle noget af den geografiske variation, som ikke ellers opfanges i disse modeller.

Slutteligt er uddannelse inkluderet. Alle fire modeller inkluderer fire Indikatorvariabelvariable, der opfanger forskellige uddannelsesniveauer. For hver husholdning medtages kun den højeste fuldførte uddannelse inden for hver husholdning. Da samboende ofte vil have nogenlunde samme uddannelsesniveau kan parameterestimerne således ikke direkte fortolkes som "øget forbrug som følge af uddannelse". Disse parameterestimer skal i øvrigt fortolkes i forhold til referencegruppen, som er det laveste uddannelsesniveau, dvs. uddannelse på grundniveau (folkeskole) inklusive gymnasialt niveau. Ikke mindre end 30% af stikprøvens husholdninger angiver "erhvervsrettet" uddannelse som husholdningens højeste uddannelsesniveau, og derfor medtages en særskilt Indikatorvariabelvariabel for denne gruppe. Resultaterne er enslydende på tværs af de fire modeller. Husholdninger med (minimum) én person, der har taget en mellemlang eller lang videregående uddannelse, har et signifikant højere forbrug af de udvalgte dagligvarer – også når der er taget højde for indkomst samt de øvrige variable inkluderet i modellerne.

5.3 Prædiktion og usikkerhedsanalyse

I dette afsnit beskrives modellernes prædiktion af forbrugsniveauet samt de konfidensintervaller, der knytter sig til det forventede forbrug.

Tabel 5.2 Forventet forbrug af udvalgte varer (målt i kr/år) og 95% prædiktionskonfidensinterval, model 1-4

	Model 1	Model 2	Model 3	Model 4
Faktisk gennemsnitligt forbrug	49.385	49.385	49.385	49.385
Modelprædikterede	49.200	48.793	48.687	51.426
Forskel (faktisk-prædikterede)	185	592	698	-2.041
Nedre grænse for prædiktionskonfidensinterval	4.262	1.576	1.571	-33.008
Øvre grænse for prædiktionskonfidensinterval	94.139	96.011	95.802	135.860
R2	0,4599	0,5304	0,5245	0,3854

Resultaterne i tabel 5.2 er baseret på et gennemsnit over individuelle prædiktioner. Det fremgår, at alle fire modeller rammer gennemsnitlige forbrug ganske pænt, hvilket ikke er overraskende for en OLS-regression baseret på cross-section data. Forklaringsgraden er størst for model 2 og 3, hvilket også er forventeligt, da variationen i data mindskes, når vi anvender en log-lineær model.

Det fremgår også, at 95% konfidensintervallerne for alle fire modeller endog er meget brede. Det skyldes altovervejende variansen på fejlleddet, dvs. variansen i populationen, som implicit tager højde for variansen i uobserverede faktorer, der påvirker forbruget. Da modellerne 1-4 inkluderer relativt få parametre (og kun bygger på en enkelt cross-section observation) får vi som forventet, at denne usikkerhed er ganske stor. Hvis vi kun ser på prædiktionskonfidensintervallerne for den givne sample, får vi konfidensintervaller inden for nogle få

tusinde fra den forventede værdi. Men disse in-sample konfidensintervaller er ikke de relevante, når vi som her ønsker at udbrede brugen af estimerne til befolkningen generelt.

Variansen i populationen indgår også i beregningerne af forventet værdi og konfidensinterval for forskellige typer af personkarakteristika.

Tabel 5.3 Eksempler på konfidensberegninger for konkrete personkarakteristika

Type 1	Type 2
Ikke enlig	Ikke enlig
Ældre end 29	Ældre end 29
Ét barn (i aldergruppen 18+)	Ét barn (i aldergruppen 18+)
Indkomstgruppe 3 (Y3=1)	Indkomstgruppe 3 (varierer over alle fem grupper)
Ejerbolig	Ejerbolig
Uddannelse (varierer over alle fem grupper)	Erhvervsuddannelse (Dudderhv=1)

Tabel 5.4 Forventet forbrug af udvalgte varer (målt i kr./år) og 95% konfidensintervaller for Type 1

Model 1	Uddannelse				
	Almen	Kort videregående	Erhvervsrettet	Mellem videregående	Lang videregående
Faktisk gennemsnitligt forbrug	38.609	47.714	47.498	62.414	69.832
Modelprædikterede	45.509	47.835	45.894	49.311	52.417
Forskel (faktisk-prædikterede)	-6.900	-121	1.604	13.103	17.415
Modelafvigelse i procent	18%	0%	-3%	-21%	-25%
Nedre grænse for prædiktionskonfidensinterval	663	2.776	1.049	4.420	7.396
Øvre grænse for prædiktionskonfidensinterval	90.355	92.893	90.740	94.201	97.437
Model 2					
Faktisk gennemsnitligt forbrug	38.609	47.714	47.498	62.414	69.832
Modelprædikterede	42.705	44.435	45.944	48.980	51.147
Forskel (faktisk-prædikterede)	-4.096	3.279	1.554	13.434	18.685
Modelafvigelse i procent	11%	-7%	-3%	-22%	-27%
Nedre grænse for prædiktionskonfidensinterval	-4.509	-2.780	-1.271	1.764	3.930
Øvre grænse for prædiktionskonfidensinterval	89.920	91.649	93.159	96.196	98.363
Model 3					
Faktisk gennemsnitligt forbrug	38.609	47.714	47.498	62.414	69.832
Modelprædikterede	43.257	46.138	47.550	49.554	53.027
Forskel (faktisk-prædikterede)	-4.648	1.576	-52	12.860	16.805
Modelafvigelse i procent	12%	-3%	0%	-21%	-24%
Nedre grænse for prædiktionskonfidensinterval	-3.855	-975	437	2.440	5.912
Øvre grænse for prædiktionskonfidensinterval	90.369	93.251	94.664	96.668	100.142
Model 4					
Faktisk gennemsnitligt forbrug	38.609	47.714	47.498	62.414	69.832
Modelprædikterede	43.686	49.713	49.576	51.736	54.058
Forskel (faktisk-prædikterede)	-5.077	-1.999	-2.078	10.678	15.774
Modelafvigelse i procent	13%	4%	4%	-17%	-23%
Nedre grænse for prædiktionskonfidensinterval	-39.274	-33.355	-33.343	-31.250	-28.926
Øvre grænse for prædiktionskonfidensinterval	126.648	132.782	132.495	134.722	137.042

Tabel 5.5 Forventet forbrug og 95% konfidensintervaller for Type 2

Model 1	Disponibel husstandsindkomst (1000 kr.)				
	< 150	150-250	250-350	350-550	> 550
Faktisk gennemsnitligt forbrug	47.558	41.940	47.498	51.863	67.298
Modelprædikterede	41.765	43.555	45.895	55.317	66.380
Forskel (faktisk-prædikterede)	5.793	-1.615	1.603	-3.454	918
Modelafvigelse i procent	-12%	4%	-3%	7%	-1%
Nedre grænse for prædiktionskonfidensinterval	-3.176	-1.268	1.049	10.432	21.442
Øvre grænse for prædiktionskonfidensinterval	86.706	88.378	90.740	100.203	111.318
Model 2					
Faktisk gennemsnitligt forbrug	47.558	41.940	47.498	51.863	67.298
Modelprædikterede	40.529	42.993	45.944	55.600	65.239
Forskel (faktisk-prædikterede)	7.029	-1.053	1.554	-3.737	2.059
Modelafvigelse i procent	-15%	3%	-3%	7%	-3%
Nedre grænse for prædiktionskonfidensinterval	-6.684	-4.221	-1.271	8.381	18.016
Øvre grænse for prædiktionskonfidensinterval	87.743	90.207	93.159	102.818	112.461
Model 3					
Faktisk gennemsnitligt forbrug	47.558	41.940	47.498	51.863	67.298
Modelprædikterede	43.263	43.432	47.550	54.379	61.696
Forskel (faktisk-prædikterede)	4.295	-1.492	-52	-2.516	5.602
Modelafvigelse i procent	-9%	4%	0%	5%	-8%
Nedre grænse for prædiktionskonfidensinterval	-3.849	-3.680	437	7.263	14.577
Øvre grænse for prædiktionskonfidensinterval	87.743	90.207	93.159	102.818	112.461
Model 4					
Faktisk gennemsnitligt forbrug	47.558	41.940	47.498	51.863	67.298
Modelprædikterede	39.227	44.767	49.576	56.421	65.927
Forskel (faktisk-prædikterede)	8.331	-2.827	-2.078	-4.558	1.371
Modelafvigelse i procent	-18%	7%	4%	9%	-2%
Nedre grænse for prædiktionskonfidensinterval	-43.778	-38.161	-33.343	-26.538	-17.368
Øvre grænse for prædiktionskonfidensinterval	122.232	127.696	132.495	139.380	149.223

Tabel 5.4 og tabel 5.5 giver et indtryk af, hvordan modellernes evne til at prædiktere forbruget samt de tilhørende 95% konfidensintervaller er. Afvigelserne er angivet både i den faktiske forskel målt i kroner samt den relative afvigelse fra det observerede forbrugsniveau. For hver af de to typer, Type 1 og Type 2 (se tabel 5.3), lader vi uddannelse henholdsvis husstandsindkomst variere og ser, hvordan hver af de fire modeller prædikterer. Bemærk her, at der er meget få observationer for hver af disse udvalgte typer (fx er der 29 personer, som har karakteristika som Type 1 specificeret med uddannelsesniveau "almen" og blot 6 personer Type 1 karakteristika og med uddannelsesniveau "kort videregående").

Tabellerne viser, at de relative modelafvigelser generelt er størst for meget høje eller slet ingen (almen) uddannelse og mindst for erhvervsrettet. Dette afspejler, at der er mange personer med fx erhvervsrettet uddannelse (30% i stikprøven), og OLS-estimerne bliver derfor mere sikre for denne gruppe. Desuden ligger denne gruppe "midt i mellem" de øvrige grupper, og derfor vil OLS også af denne grund tendere til at gøre modelafvigelsen mindre for denne gruppe. Et tilsvarende mønster gælder for Type 2-beregningerne (tabel 5.5), hvor afvi-

gelskerne er størst for de meget lave indkomster, som kun udgør 11% af stikprøven, jf. tabel 3.4.

Alle modellerne giver meget brede konfidensintervaller. Navnlig resulterer model 4 i ganske ekstreme konfidensintervaller. De smalleste intervaller ses generelt for model 2 og 3, som ligeledes har de laveste afvigelser mellem faktisk og forventet forbrug. Fortegnet og størrelsen på den relative modelafvigelse er stort set identisk for modellerne 2 og 3. Afvigelserne er betydelige for variation i uddannelse (i flere tilfælde 20-30%, se tabel 5.4), mens de er beskedne for variation i indkomsten (i flere tilfælde 3-8%, se tabel 5.5).

Sammenholdes disse resultater med resultaterne i tabel 5.2, indikerer det samlet set, at man bør vælge model 2 eller model 3, og at de hver især vil være stort set lige gode. Da modellen skal anvendes til at angive husholdningernes forventede forbrug foretrækkes (marginalt) model 3, som har lidt lavere procentuel afvigelse i prædikterede værdier sammenholdt med model 2.

5.4 Regneeksempel

Her gives et eksempel på, hvordan man kan beregne det forventede forbrug ud fra nogle indtastede data.

Da modellen er estimeret med den naturlige logaritme til forbruget som afhængig variabel, er det nødvendigt at anvende den skalering, der blev omtalt i afsnit 5.1, dvs. det er ikke tilstrækkeligt blot at beregne det forventede forbrug som $\exp(\widehat{\ln C_i})$. I stedet beregnes det forventede forbrug, $\widehat{C_i}$, som $\widehat{C_i} = \exp(\widehat{\ln C_i}) \times \hat{\alpha}_0$.

Lad os først se på beregningen af $\widehat{\ln C_i}$.

Ved at indsætte parameterestimerterne for model 3-parametrene, se kolonne (3) i tabel 5.1, får vi følgende

$$\begin{aligned} \widehat{\ln C_i} = & 14,67 \\ & - 0,353 \times \text{Enlig}_i \\ & + 0,0385 \times \max(\text{Alder1}_i, \text{Alder2}_i) - 0,000344 \times (\max(\text{Alder1}_i, \text{Alder2}_i))^2 \\ & + 0,115 \times \text{Nborn}_{-5}_i + 0,111 \times \text{Nborn}_{-611}_i + 0,138 \times \text{Nborn}_{-1217}_i + 0,155 \times \text{Nborn}_{-18}_i \\ & - 1,030 \times \ln(Y_i) + 0,0463 \times (\ln Y_i)^2 \\ & + 0,0463 \times \text{Dejer}_i + 0,0371 \times \text{Dudderhv}_i + 0,0168 \times \text{Duddkort}_i + \\ & + 0,121 \times \text{Duudml}_i + 0,155 \times \text{Dlang}_i \end{aligned}$$

I dette eksempel forestiller vi os, at en husstand har følgende karakteristika

Spørgsmål	Svar
Bor du alene?	Nej
Hvor gammel er den ældste person i husstanden?	52
Hvor mange hjemmeboende børn er der i alderen 0-5?	0
Hvor mange hjemmeboende børn er der i alderen 6-11?	1
Hvor mange hjemmeboende børn er der i alderen 12-17?	2
Hvor mange hjemmeboende børn er der i alderen 18+?	0
Hvad er husstandens samlede årlige indkomst <u>efter</u> skat? (her inkluderes alle former for indkomst såsom lønindkomst, overførselsindkomst, børnecheck osv.)	680.000 kr.
Hvad er den højeste gennemførte uddannelse i husstanden?	Erhvervsfaglig
Bor husstanden i ejerbolig?	Ja

Anm.: Bemærk, at spørgsmålene her ikke er skrevet som forslagene til spørgsmål i afsnit 2.2. I eksemplet her er det udelukkende for at vise, hvordan beregningen laves med en given husstand.

Givet en husstand med disse karakteristika kan vi nu indsætte værdier for variablene og multiplicere dem med de angivne parameterestimer. Det giver følgende (bemærk, at $\ln(680.000)=13,42985$):

$$\begin{aligned} \widehat{\ln C}_i &= 14,67 \\ &- 0,353 \times 0 \\ &+ 0,0385 \times 52 - 0,000344 \times (52)^2 \\ &+ 0,115 \times 0 + 0,111 \times 1 + 0,138 \times 2 + 0,155 \times 0 \\ &- 1,030 \times \ln(680.000) + 0,0487 \times (13,42985)^2 \\ &+ 0,0463 \times 1 + 0,0371 \times 1 + 0,0168 \times 0 + \\ &+ 0,121 \times 0 + 0,155 \times 0 \end{aligned}$$

Ud fra dette udregnes $\widehat{\ln C}_i = 11,16158$.

Det næste skridt er at inkludere korrektionsfaktoren, $\hat{\alpha}_0$, som for model 3 er beregnet til at være $\hat{\alpha}_0 = 1,0927$.

Vi indsætter nu i formlen $\hat{C}_i = \exp(\widehat{\ln C}_i) \times \hat{\alpha}_0$. Det giver et forventet forbrug

$$\begin{aligned} \hat{C}_i &= \exp(11,16158) \times 1,0927 \\ &= 76.898 \end{aligned}$$

For husstanden i dette eksempel finder vi altså, at det forventede forbrug af de udvalgte dagligvarer kan beregnes til 76.898. kr. Det vil nok være hensigtsmæssigt at anvise et afrundet beløb til nærmeste hele 1.000 kr. pr. år, således at man i dette eksempel vil anslå det forventede forbrug af de udvalgte dagligvarer til at være 77.000 kr. pr. år.

6 Litteratur

Banks, J., R. Blundell & A. Lewbel (1997): Quadratic Engel Curves and Consumer Demand. *The Review of Economics and Statistics*, LXXIX(4):527-539.

Wooldridge, J. M. (2009): *Introductory Econometrics – A Modern Approach. Fourth edition.* South-Western Cengage Learning.

Husholdningsbudgetberegner

I dette notat dokumenteres resultaterne af estimationer af en model for husholdningers dagligvareforbrug, det vil sige husstandenes samlede årlige forbrug (målt i kr.) af fødevarer, drikkevarer og husholdningsartikler. Fire forskellige modeller testes. De fire alternative modeller afspejler varianter af modeller for Engel-kurver, dvs. modeller der beskriver, hvordan husholdningsudgifter på udvalgte varer varierer med husstandens indkomst. Den udvalgte model anvendes til at lave prædiktioner for forbruget for typiske husholdninger med bestemte karakteristika. Disse prædiktioner skal efterfølgende danne baggrund for en webapplikation af husholdningers forventede forbrug.