

Assignment 3, Økonometri A

20. oktober 2025

Introduktion til opgaven

En afgørende udfordring i analyser af bilmarkedet er at forstå, hvordan forbrugerne reagerer på ændringer i priser og produktkarakteristika, og hvordan virksomheder fastsætter deres priser under hensyntagen til disse reaktioner. Ved at kombinere data med økonomisk teori og statistiske metoder kan vi opnå en dybere indsigt i disse mekanismer og kvantificere effekterne af forskellige politikker på bilmarkedet.

Vi tager udgangspunkt i en strukturel model for udbud og efterspørgsel efter biler inspireret af Grieco, Murry and Yurukoglu (2024). Vi undersøger først, hvordan forbrugere træffer valg mellem forskellige bilmodeller baseret på bilernes priser og egenskaber. Under antagelse om optimal prissætning undersøger vi herefter, hvordan vi kan udlede virksomhedernes (uobserverede) marginalomkostninger. Denne strukturelle tilgang lader os estimere parametre, der kan bruges til at simulere effekterne af forskellige politikker på bilmarkedet, såsom skatter på hestekræfter eller ændrede tariffer.

Afleveringen rummer fire opgaver, som hver består af tre delopgaver. Opgave 1-3 fokuserer på empiriske analyser af bilmarkedet, mens opgave 4 omfatter en teoretisk og simulationsbaseret Monte Carlo-analyse. Opgave 4 kan løses uafhængigt af opgave 1-3. Nogle studerende vil have fordel af at starte med opgave 4, da arbejdet med teorien kan hjælpe med fortolkningen af de empiriske resultater.

Bemærk: Din besvarelse må maksimalt fylde 2600 ord (dvs. ca. 10 sider). Antallet af ord skal fremgå på din forside. Hvis du skriver din opgave i Overleaf, kan du tælle opgavens ord ved at klikke Menu → Word Count øverst til venstre. Figurer, tabeller og forside tæller ikke med i dit sideantal.

Det anbefales at du er kortfattet, præcis og formel i din argumentation, og at du konkret henviser til modellen og den økonometriske teori, når det er muligt.

Data

Til denne aflevering medfølger datasættet `cars.csv`, som er en modificeret udgave af datagrundlaget for Grieco, Murry and Yurukoglu (2024). Indlæses filen som DataFrame i Python med koden:

```
import pandas as pd
df = pd.read_csv("cars.csv") # Tilpas evt. stien til csv-filen
df.describe(percentiles=[]).T
```

Datasættet indeholder 9.694 observationer af markedsandele, priser og produktkarakteristika for forskellige bilmodeller på det amerikanske marked fordelt på årene 1980-2018. Hvert produkt j er identificeret ved en bilproducent (make) og en bilmodel (model) i år t (year). Observationsenheden er altså produkt-år kombinationer (j, t) , og et eksempel er: `Toyota Corolla, 2007`.

Priser (price) er målt i \$1.000 (2015-priser). year og number_households varierer kun over tid. Dummyvariablene car, suv, truck, og van er gensidigt udelukkende og summerer til 1. Dummyvariablene EV, HEV, og PHEV er også gensidigt udelukkende, men de summerer ikke til 1 (f.eks. mangler dummy for diesel- og benzinbiler). De resterende to dummier, sport og lux, er ikke gensidigt udelukkende og der er ingen dummy for referencekategorien.

Tabel 1 giver et overblik over alle de tilgængelige variabler.

Tabel 1: Variabelbeskrivelse

Variabel	Beskrivelse og Værdisæt
year	Kalenderår for observation. Interval, $t = 1980, 1981 \dots 2018$
product_id	Unik ID for make-model, $j = 0, \dots, J = 1066$
make	Bilproducent som Ford, Toyota, BMW, Volvo, etc.
model	Model af køretøj som BMW 3 Series, Chevrolet Suburban, Ford Mustang, etc.
number_households	Antal husholdninger i året, n_t
sales	Solgte enheder i et givent år
price	Anbefalet udsalgspris, p_{jt} , målt i \$1,000 (2015-priser)
height	Højde på køretøjet. Målt i tommer
footprint	Areal (længde \times bredde). Målt i kvadrattommer
weight	Vægt af køretøjet. Målt i pund
hp	Hestekræfter.
mpg	Brændstoffeffektivitet målt i miles per gallon
yearsSinceDesign	Antal år siden design blev introduceret. Interval: 0-25
number_trims	Antal trimniveauer for modellen (dvs. konfigurationsmuligheder)
releaseYear	Dummy for nyt modelår
lux	Dummy for luksusbiler
sport	Dummy for sportsbiler
HEV	Dummy for hybridbiler
PHEV	Dummy for plug-in hybridbiler
EV	Dummy for elbiler
car	Dummy for personbiler
suv	Dummy for SUV'er
truck	Dummy for lastbiler
van	Dummy for varevogne
market_share	Markedsandel, $s_{jt} \in (0, 1)$
no_car	Andel af husholdninger uden bil, $s_{0t} \in (0, 1)$
dlogS	Logaritme af markedsandele relativt til andelen uden bil, $\log\left(\frac{s_{jt}}{s_{0t}}\right)$
RXR	Real valutakurs
gh_log_g	GH-instrumentet for $\log g \in \{\text{height, footprint, hp, mpg, weight}\}$

Note: Nogle variabler er udeladt, omdøbt, skaleret anderledes eller tilføjet ift. det oprindelige datasæt.

Model (opsummering)

Afleveringen tager udgangspunkt i en strukturel model for bilmarkedet, som kan opsummeres med følgende tre lineære ligninger:

1. Markedsandele (efterspørgselskurve)

$$\log\left(\frac{s_{jt}}{s_{0t}}\right) = \beta \mathbf{x}_{jt} + \alpha p_{jt} + \xi_{jt}. \quad (1)$$

2. Prissætning (udbudskurve)

$$p_{jt} = mc_{jt} + \frac{1}{-\alpha(1 - s_{jt})}. \quad (2)$$

3. Marginalomkostninger

$$\log(mc_{jt}) = \gamma_x \mathbf{x}_{jt} + \gamma_w \mathbf{w}_{jt} + \nu_{jt}, \quad (3)$$

hvor

- s_{jt} er markedsandelen for produkt j i år t . s_{0t} er andelen af forbrugere uden bil i år t .
- \mathbf{x}_{jt} er de observerbare produktkarakteristika som også påvirker nytten (f.eks. størrelse).
- p_{jt} er den observerbare pris.
- ξ_{jt} er fejleddet i ligning 1, dvs. uobserverede produktkarakteristika som påvirker nytten.
- mc_{jt} er de uobserverede marginalomkostninger.
- \mathbf{w}_{jt} er udbudsspecifikke karakteristika (f.eks. råvarepriser, skatter, lønomkostninger, mv.).
- ν_{jt} er fejleddet i ligning 3, som også kan tolkes som et stokastisk udbudsstød.
- β , α , γ_x og γ_w er parametre, som vi ønsker at estimere.

Bemærk, at ovenstående kun er en kortfattet opsummering af modellen. En komplet beskrivelse af modellen kan findes i appendikset. Det anbefales, at man sætter sig ind i den fulde strukturelle model, så man kan relatere sin analyse til den økonomiske teori.

Ligevægt på bilmarkedet

Den centrale udfordring er, at de tre ligninger ovenfor introducerer en simultanitet mellem priser p_{jt} og markedsandele s_{jt} . Den simultanitet er vi nødt til at tage højde for, hvis vi ønsker at estimere modellens parametre med data på observerede priser p_{jt} , markedsandele s_{jt} , produktkarakteristika \mathbf{x}_{jt} og udbudsspecifikke variable \mathbf{w}_{jt} . Det virker umiddelbart som en vanskelig opgave, da både markedsandelene s_{jt} og priserne p_{jt} er endogene, fordi de er bestemt i en ligevægt. Heldigvis tager vi et skridt ad gangen.

Opgave 1: Indledende analyse (25%)

(1A) Udfør en deskriptiv analyse af datasættet `data.csv`.

- Illustrér udviklingen over tid i priser, markedsandele og andre centrale attributter som hestekræfter (`hp`), brændstoffeffektivitet (`mpg`), areal (`footprint`) og vægt (`weight`) for forskellige biltyper (f.eks. `car`, `suv`, `truck`, `van`).
- Undersøg sammenhængen mellem logaritmen af markedsandele relativt til andelen uden bil (`dlogS`) og bilpriser (`price`), samt mellem priser og fysiske attributter for bilerne. Kommentér kort på de observerede tendenser.

(1B) Brug OLS til at estimere varianter af efterspørgselsligningen i ligning (1) som specificeret herunder. Rapporter for hver af specifikationerne dine parameterestimer, standardfejl, TSS, ESS, RSS og R^2 . I alle specifikationerne er den afhængige variabel logaritmen af markedsandelen relativt til andelen uden bil, dvs. $y_{jt} = \log\left(\frac{s_{jt}}{s_{0t}}\right)$ (variablen `dlogS`). Vi ser i første omgang bort fra endogenitetsproblemet, og antager derfor her, at prisen er eksogen.

- *Specifikation 1:* Estimer en variant uden produktkarakteristika (men med et konstantled β_0):

$$y_{jt} = \beta_0 + \alpha p_{jt} + \xi_{jt}$$

Fortolk fortegnet til $\hat{\alpha}$. Stemmer det overens med din forventning og almindelige økonomiske antagelser?

- *Specifikation 2:* Estimer en variant med produktkarakteristika:

$$y_{jt} = \beta \mathbf{x}_{jt} + \alpha p_{jt} + \xi_{jt}$$

hvor matricen \mathbf{x}_{jt} indeholder følgende karakteristika: et konstantled, variablene `hp`, `mpg`, `height`, `footprint`, `weight`, `number_trims`, `releaseYear`, og `yearsSinceDesign`, samt dummies for biltyperne `sport`, `EV`, `truck`, `suv` og `van`.

- *Specifikation 3:* Estimer en variant, hvor vi også kontrollerer for tidseffekten δ_t ved at inkludere et fuldt sæt af tids-dummies for kalenderår dannet ud fra variablen `year`:

$$y_{jt} = \beta \mathbf{x}_{jt} + \alpha p_{jt} + \delta_t + \xi_{jt}$$

Hint: Tidsdummies kan genereres i pandas ved at bruge `pd.get_dummies()`. Eksempelvis:

```
# Add dummies for 'year'
year_dummies = pd.get_dummies(df['year'], prefix='y', drop_first=True, dtype=int)
df = pd.concat([df, year_dummies], axis=1)

# Store year variable names for later use
years = list(year_dummies.columns) # ['y_1981', 'y_1982', ...]
```

- *Specifikation 4*: Estimer samme specifikation, men hvor du log-transformerer de kontinuerte forklarende variable `hp`, `mpg`, `height`, `footprint`, `weight` og `number_trims`.
- Sammenlign de fire specifikationer og fortolk dine resultater. Diskuter, hvordan kontrol for forskellige variable påvirker $\hat{\alpha}$. Vurder, hvilken model der bedst beskriver data.

(1C) Udfør et test af $H_0 : \alpha = 0$ mod det tosidede alternativ $H_A : \alpha \neq 0$.

- Opstil den relevante test statistik og redegør for de nødvendige antagelser bag testet.
- Konkluder på testresultatet og diskuter, hvordan testresultatet påvirkes af, om der inkluderes produktkarakteristika, biltyper og tidseffekter i modellen.
- Diskuter om resultatet er følsomt over for brugen af heteroskedasticitetsrobuste standardfejl.

Opgave 2: Fixed Effects og IV-estimation (25%)

(2A) Vi vil nu udvide vores model ved at inkludere bilmærke-specifikke fixed effects baseret på variabelen `make`. Vi dekomponerer altså efterspørgsels-støddet som $\xi_{jt} = \bar{\xi}_m + \tilde{\xi}_{jt}$, hvor $\bar{\xi}_m$ er bilmærke-specifikke effekter, som er konstante over tid for hvert bilmærke, og $\tilde{\xi}_{jt}$ er den resterende del af fejllædet, der varierer både med bilmodel og tid.

- Estimér først modellen ved at inkludere `make`-specifikke dummies som beskrevet ovenfor, dvs.

$$y_{jt} = \beta \mathbf{x}_{jt} + \alpha p_{jt} + \delta_t + \bar{\xi}_m + \tilde{\xi}_{jt}, \quad (4)$$

Inkluder fortsat alle kontrolvariablerne fra specifikation 4 i opgave 1B.

Hint: $\bar{\xi}_m$ kan estimeres ved at inkludere en dummy, som antager værdien 1 hvis bilmærket er m og 0 ellers. Variablen kan genereres med samme fremgangsmåde, som vi brugte til vores tidsdummies i opgave 1B.

- Estimer derefter modellen ved hjælp af within-transformationen:

$$y_{jt} - \bar{y}_m = \beta (\mathbf{x}_{jt} - \bar{\mathbf{x}}_m) + \alpha (p_{jt} - \bar{p}_m) + (\delta_t - \bar{\delta}_m) + (\tilde{\xi}_{jt} - \bar{\tilde{\xi}}_m), \quad (5)$$

hvor \bar{y}_m angiver gennemsnittet inden for hvert bilmærke m baseret på variabelen `make`. For eksempel beregnes $\bar{\mathbf{x}}_m = \frac{1}{T_m} \sum_{j \in m} \mathbf{x}_{jt}$, hvor T_m er antallet af observationer for bilmærke m . De øvrige gennemsnit, såsom \bar{p}_m , er defineret på tilsvarende måde. Within-transformationen kan implementeres i `pandas` ved at bruge `groupby()` og `transform()`. Eksempelvis:

```
df_within = df.copy()
vars_to_transform = [...] # List the names of the variables to transform

for var in vars_to_transform:
    make_mean = df.groupby('make')[var].transform('mean')
    df_within[var] = df[var] - make_mean
```

- Sammenlign parameterestimaterne for de to modeller. Er de ens? Forklar, hvorfor konstantleddet er nul i modellen med within-transformerede variable, og hvorfor $\bar{\xi}_m$ forsvinder efter within-transformationen (*Hint*: læg mærke til forskellen mellem $\bar{\xi}_m$ og $\tilde{\xi}_m$).
- Sammenlign også TSS, ESS, RSS og R^2 for de to modeller. Diskuter betydningen af eventuelle forskelle i goodness-of-fit mål. Forklar, hvorfor R^2 er højere i modellen med make-dummies sammenlignet med den within-transformerede model.

(2B) Forklar nødvendigheden af at inkludere de make-specifikke fixed effects $\bar{\xi}_m$. Tag udgangspunkt i efterspørgselsligningen, hvor vi har dekomponeret fejllædet som beskrevet i 2A:

$$\log\left(\frac{s_{jt}}{s_{0t}}\right) = \beta \mathbf{x}_{jt} + \alpha p_{jt} + \delta_t + \bar{\xi}_m + \tilde{\xi}_{jt}.$$

- Hvis de uobserverede bilmærke-specifikke effekter $\bar{\xi}_m$ (fx oplevet brand-image, kvalitetsfølelse) er positivt korrelerede med prisen p_{jt} , hvilken retning forventer du så, at bias i estimatet af α vil tage, hvis $\bar{\xi}_m$ udelades fra modellen?
- Er det konsistent med det, du finder? (*Hint*: Sammenlign dine resultater fra 1B med 2A)

(2C) Den resterende tids- og bilmodelvarierende komponent $\tilde{\xi}_{jt}$ kan stadig være korreleret med prisen p_{jt} . For at adressere dette anvender Grieco, Murry and Yurukoglu (2024) den reale valutakurs, RXR, som instrument. RXR påvirker især udenlandske bilproducenters marginalomkostninger og indgår derfor i de udbudsspecifikke faktorer \mathbf{w}_{jt} i ligning (3), som igen påvirker udbudsprisen p_{jt} gennem ligning (2).

Bemærk, at selv hvis den reducerede form for prisfunktionen p_{jt} ikke er lineær i de eksogene variable, så er en korrekt specificering af denne *ikke* et krav for at benytte RXR som instrument. Vi kan derfor anvende følgende lineære approksimation af den reducerede form for prisfunktionen:

$$p_{jt} = \gamma_0 + \gamma_1 \text{RXR}_{jt} + \gamma_2 \mathbf{x}_{jt} + \kappa_t + \phi_m + v_{jt}, \quad (6)$$

hvor \mathbf{x}_{jt} er de eksogene produktkarakteristika fra efterspørgselsligningen, κ_t og ϕ_m er henholdsvis tids- og mærkespecifikke effekter, og v_{jt} er et fejllædet.

- Estimer først first stage-regressionen (6), og test, om RXR er en signifikant prædiktør for prisen ved hjælp af et t-test af γ_1 .

- Estimer herefter den strukturelle efterspørgselsligning (4) ved hjælp af 2SLS, hvor RXR anvendes som instrument for den endogene pris p_{jt} . Sammenlign 2SLS-estimatet af α med OLS-estimatet og vurder, om 2SLS-estimatet stemmer overens med din forventning om endogenitetsbias i OLS-estimatet.
- Argumentér for, at RXR er et gyldigt instrument for prisen p_{jt} . Henvi til den teoretiske model.

Opgave 3: Gandhi Houde-instrumenterne (25%)

(3A) Gandhi and Houde (2019) foreslår en anden tilgang til endogenitetsproblemet, nemlig at anvende instrumenter baseret på ligheder i produktkarakteristika. Ideen bag Gandhi-Houde (GH) instrumenterne er, at mere intensiv konkurrence inden for specifikke karakteristika (f.eks. brændstofeffektivitet) reducerer producenternes markup og dermed prisen. Konkret måler et GH-instrument antallet af konkurrerende bilmodeller, der minder tilstrækkeligt meget om den undersøgte bilmodel med hensyn til en specifik karakteristika.

Datasættet indeholder fem GH-instrumenter, hhv. `gh_height`, `gh_footprint`, `gh_hp`, `gh_mpg` og `gh_weight`, som er konstrueret på baggrund af karakteristikaene `height`, `footprint`, `hp`, `mpg` og `weight`.

OBS: For at kunne løse denne opgave tilfredsstillende er det *ikke* nødvendigt at forstå de tekniske detaljer om, hvordan GH-instrumenterne er konstrueret. Men for en god ordens skyld følger den formelle definition af GH-instrumenterne alligevel her:

$$z_{jt}^{(g)} = \sum_{k \in J} 1 \left(|x_{jt}^{(g)} - x_{kt}^{(g)}| < \sigma_{gt} \right), \quad (7)$$

hvor $z_{jt}^{(g)}$ er GH-instrumentet for karakteristika g for bil j på tidspunktet t , $1(\cdot)$ er indikatorfunktionen, $x_{jt}^{(g)}$ er den normaliserede karakteristika g for bil j på tidspunkt t , og σ_{gt} er standardafvigelsen for karakteristika g beregnet for år t .

- Diskutér, hvorfor instrumenter som disse er relevante og forventes at være gyldige til at identificere parameteren α . Tag udgangspunkt i den økonometriske teori og den intuitive forklaring af GH-instrumenterne; du behøver ikke henvise til ligning (7).
- Estimér efterspørgselsligningen ved hjælp af 2SLS med hhv. GH-instrumenterne alene og GH-instrumenterne sammen med RXR-instrumentet. Udfør et overidentifikationstest for at vurdere instrumenternes eksogenitet.
- Sammenlign resultatet med tidligere estimater og diskuter eventuelle forskelle. Overvej fordele og ulemper ved at bruge GH-instrumenterne sammenlignet med RXR.

(3B) Vi skal nu bruge de estimerede efterspørgselsligninger til at beregne markedsandele, marginalomkostninger og priselasticiteter for forskellige bilmodeller over tid. Med et estimat for α , samt de observerede priser og markedsandele, kan vi beregne marginalomkostningerne som:

$$p_{jt} = mc_{jt} + \frac{1}{-\alpha(1 - s_{jt})} \quad (8)$$

$$\Leftrightarrow mc_{jt} = p_{jt} - \frac{1}{-\alpha(1 - s_{jt})}. \quad (9)$$

Egenpriselasticiteten kan beregnes som:

$$\varepsilon_{jt} = \frac{\partial s_{jt}}{\partial p_{jt}} \cdot \frac{p_{jt}}{s_{jt}} = \alpha \cdot p_{jt} \cdot (1 - s_{jt}).$$

Markup kan beregnes som:

$$\text{Markup}_{jt} = \frac{p_{jt} - mc_{jt}}{p_{jt}}$$

- Beregn marginalomkostningerne, markups og priselasticiteter for alle observationer i datasættet baseret på både OLS og IV-estimerne opnået i 2A og 2C, $\hat{\alpha}^{OLS}$ og $\hat{\alpha}^{IV}$. Har du ikke estimeret α , så brug værdierne $\hat{\alpha}^{OLS} = -0.03$ og $\hat{\alpha}^{IV} = -0.17$.
- Udfør en deskriptiv analyse for at undersøge, hvordan fordelingen af marginalomkostninger, priselasticiteter og markups varierer mellem dine OLS- og IV-estimer. Illustrer også udviklingen i gennemsnitlige marginalomkostninger, markups og priselasticiteter for biltyperne *car*, *suv*, *truck* og *van* over tid på baggrund af dine IV-estimer. Brug gerne visualiseringer.
- Reflektér over, hvor afhængige vi er af modeller og antagelser, når vi estimerer marginalomkostninger. Diskutér, hvilke konsekvenser det kan have, hvis modellen er fejlspecificeret, eller hvis instrumenterne ikke er valide eller stærke nok til at korrigere for prisendogenitet.

(3C) Vi kan bruge prædikterede værdier af marginalomkostningerne fra 3B baseret på IV-estimerne for α til at estimere marginalomkostningsfunktionen (3) med bilmærke- og tidsspecifikke effekter:

$$\log(mc_{jt}) = \gamma_x \mathbf{x}_{jt} + \gamma_{RXR} RXR_{jt} + \phi_m + \kappa_t + \nu_{jt} \quad (10)$$

hvor \mathbf{x}_{jt} indeholder et konstantled og de samme produktkarakteristika, som vi anvendte i specifikation 4 i opgave 1B.

- Estimer marginalomkostningsfunktionen og fortolk parameterestimerne γ_x og γ_{RXR} . Hvordan varierer marginalomkostningerne med produktkarakteristika og RXR?
- Er marginalomkostningerne steget eller faldet over tid, når vi kontrollerer for produktkarakteristika, RXR, og mærkespecifikke effekter?

Opgave 4: Teoretiske opgaver (25%)

(4A) Betragt følgende simple strukturelle model for udbud og efterspørgsel på bilmarkedet:

$$y_i^D = \beta_0 + \beta_1 p_i + \xi_i, \quad i = 1, \dots, n \quad (11)$$

$$y_i^S = \gamma_0 + \gamma_1 p_i + \gamma_2 z_i + \nu_i, \quad (12)$$

hvor y_i^D og y_i^S repræsenterer hhv. den totale efterspørgsel og det totale udbud for bil i . p_i er prisen, z_i er en forklarende variabel, og ξ_i, ν_i er to fejllid. Vi antager at $\beta_1 < 0$ og $\gamma_1 > 0$, således at efterspørgslen falder og udbuddet stiger med prisen.

- Vis, at når der er ligevægt på markedet ($y_i = y_i^D = y_i^S$), så er den reducerede form for p_i :

$$p_i = \eta_0 + \eta_1 z_i + u_i, \quad (13)$$

hvor vi har defineret

$$\eta_0 \equiv \frac{\gamma_0 - \beta_0}{\beta_1 - \gamma_1}, \quad \eta_1 \equiv \frac{\gamma_2}{\beta_1 - \gamma_1}, \quad u_i \equiv \frac{\nu_i - \xi_i}{\beta_1 - \gamma_1}.$$

Er prisen eksogen i efterspørgselsligningen, eller er den korreleret med fejllidet ξ_i ?

- Indsæt ligning (13) i ligning (11) og udled den reducerede form for y_i som en funktion af z_i .
- Diskutér med afsæt i den økonometriske teori, hvorvidt det er muligt at identificere β_1 udelukkende baseret på den reducerede form for y_i . Hvis ikke, hvad er årsagen til dette, og hvordan kan vi løse identifikationsproblemet?

(4B) Antag nu, at z_i er et eksogent instrument, dvs. $\text{Cov}(z, \xi) = \text{Cov}(z, \nu) = 0$. Antag desuden, at fejllidene er ukorrelerede, dvs. $\text{Cov}(\xi, \nu) = 0$.

- Udled biaset for OLS, når y_i 'naivt' regresseres mod prisen og en konstant. Vis at vi kan nå følgende udtryk:

$$\text{plim } \hat{\beta}_1^{OLS} - \beta_1 = \frac{-\text{Var}(\xi)}{(\beta_1 - \gamma_1) \text{Var}(p)}.$$

I hvilken retning er OLS biased? (*Hint*: Husk at $\beta_1 < 0$ og $\gamma_1 > 0$ per standardantagelse).

- Vis, at IV-estimatoren er konsistent for priskoefficienten β_1 .
- Under antagelse af homoskedasticitet kan variansen af IV-estimatoren skrives som

$$\text{Var}(\hat{\beta}_1^{IV}) = \frac{\text{Var}(\xi)}{n \text{Cov}(z, p)^2},$$

Vis, at vi i vores model kan udlede følgende udtryk for IV-estimatorens varians:

$$\text{Var}(\hat{\beta}_1^{IV}) = \frac{\text{Var}(\xi)}{n \eta_1^2 \text{Var}(z)^2} = \frac{(\beta_1 - \gamma_1)^2 \text{Var}(\xi)}{n \gamma_2^2 \text{Var}(z)^2}.$$

Hvordan påvirker styrken af instrumentet z IV-estimatorens konsistens og varians? Hvordan bliver variansen påvirket af hældningen på udbuds- og efterspørgselskurverne β_1 og γ_1 ? Fortolk.

(4C) Lad os undersøge de teoretiske forudsigelser fra opgave 4B gennem et Monte Carlo eksperiment.

- Simuler data med $n = 500$ observationer og $r = 500$ replikationer baseret på ligevægtsmodellen givet ved (11) og (13), dvs. følgende ligninger:

$$\begin{aligned}y_i &= \beta_0 + \beta_1 p_i + \xi_i \\p_i &= \eta_0 + \eta_1 z_i + u_i,\end{aligned}$$

hvor vi har defineret

$$\eta_0 \equiv \frac{\gamma_0 - \beta_0}{\beta_1 - \gamma_1}, \quad \eta_1 \equiv \frac{\gamma_2}{\beta_1 - \gamma_1}, \quad u_i \equiv \frac{\nu_i - \xi_i}{\beta_1 - \gamma_1}.$$

Brug følgende baseline parameterværdier: $\beta_0 = 0.5$, $\beta_1 = -0.1$, $\gamma_0 = 0.25$, $\gamma_1 = 1$, $\gamma_2 = 1$ og antag, at $z_i \sim N(0, 1)$, $\xi_i \sim N(0, 1)$, og $\nu_i \sim N(0, 1)$.

Estimer for hver replikation β_1 ved OLS og ved IV, hvor z_i bruges som instrument for p_i , og beregn bias og varians af OLS- og IV-estimatorerne på tværs af replikationerne.

- Lad os kalde ovenstående scenarie 1. Gentag nu eksperimentet i følgende supplerende scenarier:
 - Scenarie 2: Brug baseline parametrene, men sæt $\gamma_2 = 0.25$
 - Scenarie 3: Brug baseline parametrene, men sæt $\gamma_1 = 0.25$

Hvad er den økonomiske fortolkning af forskellene mellem de tre scenariers parameterværdier?

- Diskutér simulationsresultaterne med fokus på, hvordan parameterværdierne påvirker estimatorernes varians og bias på tværs af de tre scenarier.

Appendix

Model for udbud, efterspørgsel og ligevægt på bilmarkedet (til opgave 1-3)

Forbrugernes efterspørgsel efter biler

Vi starter med at modellere forbrugernes valg af biler. Forbrugerne maksimerer deres nytte ved valg mellem forskellige bilmodeller, $j = 0, 1, \dots, J$. Vi antager, at forbrugerne vælger det produkt, j^* , der maksimerer deres nytte, givet prisen og produktkarakteristika for produkt j på tidspunkt t :

$$j_{it}^* = \arg \max_{j=0,1,\dots,J} U_{ijt},$$

hvor $j = 0$ repræsenterer et outside good (dvs. valget at ikke købe en bil).

Vi antager, at forbrugernes præferencer kan beskrives ved en lineær nyttefunktion

$$U_{ijt} = \beta \mathbf{x}_{jt} + \alpha p_{jt} + \xi_{jt} + \epsilon_{ijt},$$

hvor \mathbf{x}_{jt} er observerbare karakteristika for produkt j på tidspunkt t (f.eks. hestekræfter, brændstoffektivitet), p_{jt} er prisen på produkt j på tidspunkt t , ξ_{jt} er uobserverbare karakteristika, som kan være brand-image eller andre kvaliteter, der ikke er direkte målt, og ϵ_{ijt} er den idiosynkratiske præference for forbruger i , som antages at følge en ekstremværdifordeling.

Antagelsen om, at ϵ_{ijt} følger en ekstremværdifordeling, leder til aggregerede markedsandele på logit-formen

$$s_{jt} = \frac{\exp(\beta \mathbf{x}_{jt} + \alpha p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^J \exp(\beta \mathbf{x}_{kt} + \alpha p_{kt} + \xi_{kt})},$$

hvor vi har normaliseret nytten af vores outside good til 0, og hvor s_{jt} er markedsandelen for produkt j på tidspunkt t .

Vi kan skrive log-markedsandelen relativt til andelen af husholdninger uden bil:

$$\log \left(\frac{s_{jt}}{s_{0t}} \right) = \beta \mathbf{x}_{jt} + \alpha p_{jt} + \xi_{jt}. \quad (\text{A1})$$

hvilket resulterer i en lineær regressionsmodel, hvor $\log \left(\frac{s_{jt}}{s_{0t}} \right)$ er den afhængige variabel, \mathbf{x}_{jt} og p_{jt} er observerbare forklarende variable, ξ_{jt} er fejleddet, og β og α er parametre vi ønsker at estimere.

Virksomhedernes prissætning og marginalomkostninger

Vi antager, at virksomhederne sætter prisen p_{jt} , der maksimerer profit givet efterspørgselsfunktionen for produkt j på tidspunktet t

$$p_{jt} = \arg \max_p [(p - mc_{jt})s_{jt}],$$

hvor mc_{jt} er marginalomkostningerne for produkt j på tidspunktet t . Førsteordensbetingelsen for optimal prissætning er givet ved

$$s_{jt} + (p_{jt} - mc_{jt}) \frac{\partial s_{jt}}{\partial p_{jt}} = 0.$$

Denne betingelse giver den optimale prissætning som en funktion af marginalomkostningerne og efterspørgslen:

$$p_{jt} = mc_{jt} - s_{jt} \left(\frac{\partial s_{jt}}{\partial p_{jt}} \right)^{-1} = mc_{jt} + \frac{1}{-\alpha(1 - s_{jt})}. \quad (\text{A2})$$

hvor det sidste lighedstegn følger ved at differentiere efterspørgselsfunktionen i (A1) mht. til prisen $\frac{\partial s_{jt}}{\partial p_{jt}} = \alpha s_{jt}(1 - s_{jt})$.

Vi antager, at logaritmen til marginalomkostningerne for virksomhed j i marked t kan beskrives som en lineær funktion af produktkarakteristika, \mathbf{x}_{jt} og udbudsspecifikke karakteristika, \mathbf{w}_{jt} :

$$\log(mc_{jt}) = \gamma_x \mathbf{x}_{jt} + \gamma_w \mathbf{w}_{jt} + \nu_{jt}, \quad (\text{A3})$$

hvor \mathbf{x}_{jt} er de observerbare produktkarakteristika som også påvirker nytten (f.eks. størrelse, kvalitet), \mathbf{w}_{jt} er udbudsspecifikke karakteristika (f.eks. råvarepriser, skatter, toldmure, lønomkostninger, mv.), ν_{jt} er et stokastisk udbudsstød, og γ_x og γ_w er parametre, vi ønsker at estimere.

Således kan modellen opsummeres ved ligningerne (A1), (A2) og (A3).

Litteratur

- GANDHI, A. AND J.-F. HOUDE (2019): “Measuring substitution patterns in differentiated-products industries,” *NBER Working paper*, (w26375).
- GRIECO, P. L., C. MURRY AND A. YURUKOGLU (2024): “The evolution of market power in the us automobile industry,” *The Quarterly Journal of Economics*, 139(2), 1201–1253.