

Data Envelopment Analysis (DEA)

August 2026



KONKURRENCE- OG FORBRUGERSTYRELSEN

Data Envelopment Analysis (DEA)**Konkurrence- og Forbrugerstyrelsen**

Carl Jacobsens Vej 35

2500 Valby

Tlf.: +45 41 71 50 00

E-mail: kfst@kfst.dk

Online ISBN 978-87-7029-886-5

Data Envelopment Analysis (DEA) er udarbejdet af
Konkurrence- og Forbrugerstyrelsen.

August 2026

Indhold

Kapitel 1	4
Data Envelopment Analysis (DEA)	4
1.1 Beskrivelse af DEA	4
1.2 Matematisk gennemgang af DEA.....	5
Kapitel 2	7
DEA-Order-M	7
2.1 Beskrivelse af DEA-Order-M.....	7
2.2 Matematisk gennemgang af Order-M.....	10
Kapitel 3	11
Vægtrestriktioner	11
3.1 Beskrivelse af vægtrestriktioner	11
3.2 Matematisk gennemgang af vægtrestriktioner.....	12
Referencer	14

Kapitel 1

Data Envelopment Analysis (DEA)

1.1 Beskrivelse af DEA

Data Envelopment Analysis (DEA) er en matematisk metode, der anvender lineær programmering til at sammenligne en række enheder – i denne sammenhæng vandselskaber – med henblik på at bestemme effektiviteten for hvert selskab.¹

Selskaberne sammenlignes ved at holde deres output op imod deres input. De selskaber, der producerer mest output i forhold til deres input vurderes at være effektive. DEA kan håndtere flere input og output samtidig. I dette notat anvendes til illustration en model med ét input (omkostninger) og to output (vand og kunder).

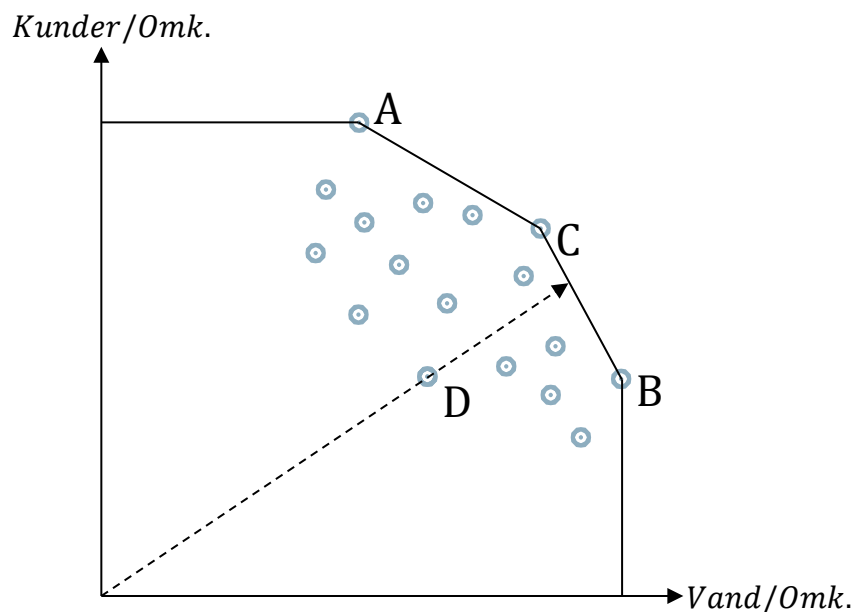
DEA fastsætter en front, som udgøres af de mest effektive selskaber, jf. Figur 1.1. I eksemplet er selskab A, B og C de mest effektive selskaber (frontselskaberne). Selskab A servicerer flest kunder i forhold til sine omkostninger og betegnes derfor som efficient (fuldt effektiv). Selskab B håndterer mest vand i forhold til sine omkostninger og betegnes derfor også som efficient. Selskab C er ligeledes efficient, fordi selskabet – givet en bestemt vægtning mellem vandmængde og kunder – producerer mest output i forhold til sine omkostninger.

DEA-fronten beregnes i eksemplet på baggrund af de tre efficiente selskaber og udgøres af de fire fuldt optrukne linjer i figuren. Selskab D ligger under fronten og er derfor inefficent. Afstanden til fronten, som er angivet med den stiplede linje, afgør selskab D's efficiensscore. En efficiensscore kan ligge mellem 0 og 100 pct. Hvis efficiensscoren fx er 80 pct., betyder det, at selskabet ifølge modellen kan reducere sine omkostninger til 80 pct. af de faktiske omkostninger og fortsat producere den samme mængde output.²

¹ Charnes, Cooper, & Rhodes (1978)

² Det antages her, at modellen er inputorienteret. Hvis modellen derimod er outputorienteret, defineres en efficiensscore til, hvor meget output kan øges for et fast niveau af input. Her vil efficiensscoren være større end 1.

Figur 1.1 Illustrativt eksempel af DEA



Anm.: Figuren viser en DEA-model med to outputs og ét input. De sorte optrukne linjer indikerer fronten som det ineffektive selskab (D) måles op imod.

Kilde: Egen tilvirkning

1.2 Matematisk gennemgang af DEA

Efficiensscorene i DEA findes ved at løse følgende lineære optimeringsproblem:

$$\min_{\lambda, E} E^0 \quad (1)$$

$$E^0 \mathbf{x}^0 \geq \sum_{k=1}^K \lambda^k \mathbf{x}^k \quad (2)$$

$$\mathbf{y}^0 \leq \sum_{k=1}^K \lambda^k \mathbf{y}^k \quad (3)$$

Optimeringsproblemet løses for hvert selskab, hvor E^0 angiver selskabets efficiensscore, $\mathbf{x}^0$ er selskabets input, K er antallet af selskaber i modellen, λ^k er en skaleringsvariabel, $\mathbf{x}^k$ er input for selskab k , $\mathbf{y}^0$ er output for det evaluerede selskab, og $\mathbf{y}^k$ er output for selskab k . Bemærk, at input og output kan være vektorer, hvis der er mere end henholdsvis ét input og ét output.

Ligning (1) angiver, at modellen finder den laveste efficiensscore, der opfylder de to bibetingelser. Ligning (2) angiver, at selskabets effektive input skal være mindst lige så stort som inputtet for det referencepunkt på fronten, som selskabet sammenlignes med. Det svarer til, at selskab D i Figur 1.1 følger den stiplede linje, indtil den rammer fronten. Ligning (3) angiver, at referencepunktets output skal være mindst lige så stort som selskabets eget output.

Ovenstående optimeringsproblem er baseret på en inputorienteret DEA-model med konstant skalaafkast, som er den modeltype, der anvendes i benchmarkingmodellen i den økonomiske regulering af vandselskaberne.³

³ For andre valg af orienteringer og antagelser om skalaafkast se fx (Bogetoft & Otto, 2010)

Kapitel 2

DEA-Order-M

2.1 Beskrivelse af DEA-Order-M

DEA-Order-M (herefter kaldet Order-M) er en udvidelse af standard-DEA, som har til formål at reducere afhængigheden af enkelte frontselskaber. Det gør modellen mindre følsom over for outliers, datausikkerhed og generelle udsving i datagrundlaget. Fronten i Order-M estimeres som en såkaldt partiel front på baggrund af gentagne udtræk af sammenlignelige selskaber.

Order-M blev første gang foreslået i 2002 som en robust tilgang til frontierestimering og outlieranalyse i DEA.⁴ Metoden er efterfølgende anvendt som benchmarkingmetode i analyser af flere sektorer.⁵

Order-M estimerer en robust partiel front ved at beregne DEA gentagne gange. I hver iteration udtrækkes M tilfældige selskaber, som udgør referencegrundlaget i den pågældende iteration. De mest effektive blandt de udtrukne selskaber bliver frontselskaber i iterationen. Det adskiller sig fra standard-DEA, hvor alle selskaber er med i referencegrundlaget. Hvis de mest effektive selskaber fra standard-DEA ikke udtrækkes i en Order-M-iteration, vil de næstmest effektive udtrukne selskaber danne fronten i den pågældende iteration. Den endelige Order-M-efficiensscore beregnes som gennemsnittet af efficiensscorerne over alle iterationer. I den input-orienterede fortolkning giver Order-M derfor efficiensscorer, der er mindst lige så høje som i standard-DEA, når scorerne sammenlignes på samme skala.⁶

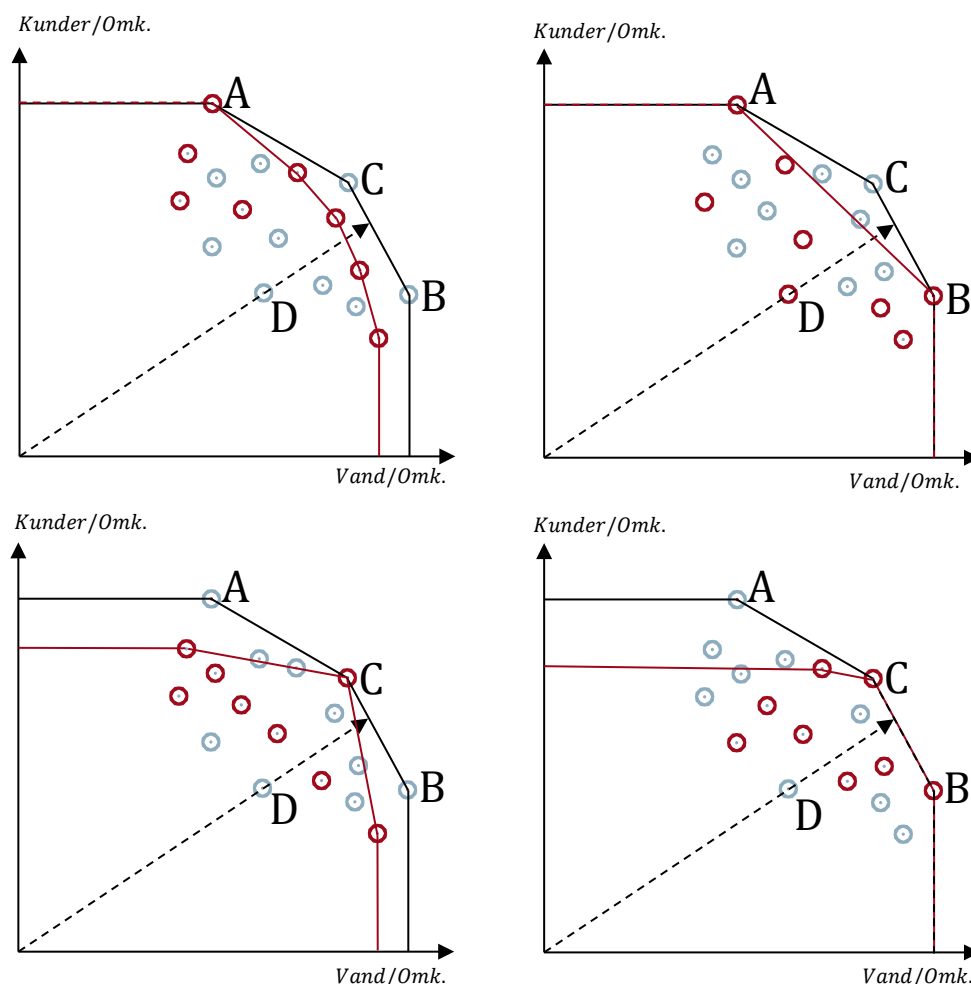
Figur 2.1 illustrerer, hvordan fronten i fire tilfældige Order-M-iterationer påvirkes af, hvilke selskaber der indgår i udtrækket. Figuren viser samtidig samme DEA-front som Figur 1.1. De røde punkter indikerer de selskaber, der er tilfældigt udtrukket, og de blå punkter indikerer de selskaber, som ikke indgår i det konkrete udtræk. I figuren øverst til venstre er selskab C og B fx ikke udtrukket. Det betyder, at selskab C og B ikke kan danne front i den pågældende iteration. Fronten, som er angivet ved de røde linjestykker, dannes derfor på baggrund af selskab A og de mest effektive af de øvrige udtrukne selskaber. De inefficiante selskaber sammenlignes dermed med den røde front i stedet for den sorte, hvilket giver en højere efficiensscore. Fronten kan variere betydeligt mellem iterationer. I nogle iterationer vil fronten være identisk med DEA-fronten, fx hvis selskab A, B og C alle udtrækkes. Fronten kan dog ikke ligge nordøst for DEA-fronten. Derfor er den gennemsnitlige efficiens i Order-M mindst lige så høj som i standard-DEA.

⁴ Cazals, Florens, & Simar (2002)

⁵ Silva, Martins-Filho, & Ribeiro (2016) og Daraio & Simar (2007)

⁶ Selskaberne trækkes med tilbagelægning, så ét selskab kan trækkes flere gange.

Figur 2.1 Sammenligning af en DEA-front og fronten fra enkelte Order-M-iterationer

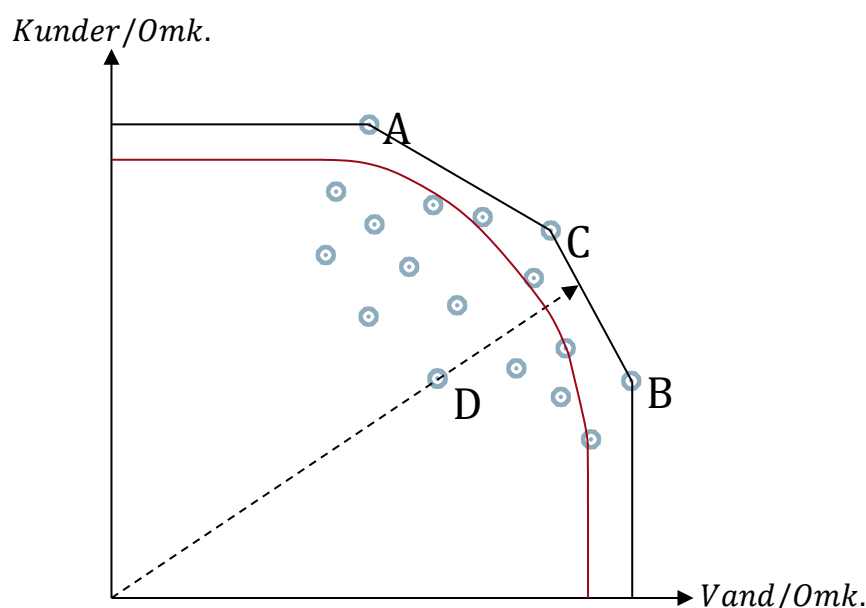


Anm.: Figuren sammenligner fronten i forskellige Order-M-iterationer med en standard DEA. Hvert punkt indikerer et selskab, hvor de røde punkter er de selskaber, der er tilfældigt udtrukket til at må indgå på fronten i den givne iteration, og de blå punkter er de selskaber, der ikke indgår i udtrækket. De sorte optrukne linjer angiver DEA-fronten, og de røde linjer angiver fronten i den givne Order-M-iteration. Figuren er lavet på simulerede data.

Kilde: Egen tilvirkning

Gennemsnittet af efficiensscorene fra alle iterationer kan illustreres som en gennemsnitlig partiel front, jf. Figur 2.2. Formelt beregnes den endelige Order-M-score dog som et gennemsnit af scorerne fra de enkelte iterationer. I Figur 2.2 ligger Order-M-fronten sydvest for DEA-fronten, hvilket gør den mere lempelig over for selskaberne end standard-DEA-fronten, hvilket gør den mere lempelig over for selskaberne end standard-DEA-fronten. Selskaberne får derfor højere efficiensscore og dermed lavere individuelle effektiviseringskrav end ved brug af standard-DEA. Figuren viser samtidig, at enkelte selskaber kan ligge nordøst for Order-M-fronten og derfor få en efficiensscore over 1. Det skyldes, at Order-M tager højde for støj, som lader selskaberne fremstå mere effektive, end hvad der ifølge modellen som udgangspunkt er teknisk muligt.

Figur 2.2 Sammenligning af fronten i DEA og Order-M



Anm.: Figuren sammenligner fronten i DEA og Order-M. Hvert punkt indikerer et selskab. De sorte optrukne linjer angiver DEA-fronten, og de røde linjer angiver Order-M-fronten. Figuren er lavet på simulerede data.

Kilde: Egen tilvirkning

Order-M bruger samme datagrundlag og de samme grundlæggende modelantagelser som DEA. Ved anvendelse af Order-M som benchmarkingmetode skal der derfor kun tages stilling til to ekstra parametre.

- 1) Antallet af iterationer: Antallet af iterationer skal sættes så højt, at resultaterne er stabile, hvis modellen genberegnes. Order-M kan dog kræve betydelig regnekraft. Valget af antal iterationer er derfor en afvejning mellem konvergens og beregnings-tid. I reguleringen af den danske vandsektor er der forholdsvis få selskaber, hvilket betyder, at der ikke er et problem med regnekraft. Antallet af iterationer kan derfor sættes meget højt (fx 10.000), hvilket sikrer konvergens uden at give problemer med regnekraften.
- 2) Antal udtrukne selskaber (M): M angiver, hvor mange selskaber der trækkes i hver iteration. Jo lavere M er, desto mere lempelig bliver den partielle front typisk. Det skyldes, at et lavt M øger sandsynligheden for, at de mest effektive selskaber ikke ud-trækkes i en given iteration. Der er ikke én generelt accepteret værdi for M ; paramete-
ren skal vælges ud fra datagrundlag, modelvalg og formålet med benchmarkingen.⁷

⁷ Bemærk, at Vandsektortilsynet trækker selskaber med tilbagelægning. Det vil sige, at det samme selskab kan trækkes flere gange, hvilket medfører, at der i alt trækkes færre selskaber end M tilsiger.

2.2 Matematisk gennemgang af Order-M

Efficiensscoren i Order-M findes ved at tage gennemsnittet af N DEA-beregninger, hvor hver beregning gennemføres på et tilfældigt udtrukket referencegrundlag.

I hver iteration løses følgende lineære optimeringsproblem:

$$\min_{\lambda, E} E_i^0 \quad (4)$$

$$E_i^0 x^0 \geq \sum_{k \in \bar{M}_i} \lambda^k x^k \quad (5)$$

$$y^0 \leq \sum_{k=1} \lambda^k y^k \quad (6)$$

Optimeringsproblemet løses for hvert selskab, hvor parametrene er de samme som i DEA, jf. ligning (1) - (3). Den centrale forskel er, at antallet af selskaber (K) i standard-DEA erstattes af den udtrukne referencegruppe $\bar{M}_i$. Gruppen $\bar{M}_i$ findes ved at trække M tilfældige selskaber med tilbagelægning fra hele populationen af selskaber. Bemærk, at efficiensscoren nu er indekseret med i , som refererer til den givne iteration.

Ovenstående optimeringsproblem løses for hvert selskab N gange, hvor referencegruppen genberegnes i hver iteration. Den endelige Order-M-efficiensscore for selskab 0 er givet ved gennemsnittet af efficiensscorerne for alle iterationer, jf. ligning (7).

$$E^0 = \frac{\sum_{i=1}^N E_i^0}{N} \quad (7)$$

Anm.: Figuren viser en DEA-model med to outputs og ét input. De sorte optrukne linjer indikerer fronten fra en standard DEA som det ineffektive selskab (D) måles op imod. De røde og grønne linjestykker er eksempler på vægtrestriktioner, som udvider fronten fra de sorte linjestykker.

Kilde: Egen tilvirkning

Nedenfor gennemgår vi to typer vægtrestriktioner: relative vægtrestriktioner og virtuelle vægtrestriktioner.

Relative vægtrestriktioner

Relative vægtrestriktioner angiver, hvordan forholdet mellem to eller flere vægte må fastsættes. Fx kan en relativ vægtrestriktion angive, at antal kunder som minimum skal vægte med halvdelen af vægten på vandmængden. Det svarer til, at det nederste røde linjestykke i Figur 3.1 får en hældning på -2. Dermed antager DEA, at et selskab, der reducerer antallet af kunder med 2, skal øge vandmængden med mindst 1 for at bevare den samme efficiensscore ved et uændret omkostningsniveau.⁸

Virtuelle vægtrestriktioner

Virtuelle vægtrestriktioner påvirker ligesom relative vægtrestriktioner fronten, men de formuleres og fortolkes anderledes.

En virtuel vægtrestriktion angiver, hvor stor en del af den samlede vægtede outputværdi der må komme fra et eller flere bestemte output. Fx kan en virtuel vægtrestriktion angive, at antal kunder højst må udgøre 20 pct. af den samlede vægtede outputværdi. Det vil sige, at modellen ikke må estimere en vægt (også kaldet skyggepris) for det givne output, som gør, at outputværdien bliver for stor relativt til de øvrige output. Grafisk gælder samme princip som for de relative vægtrestriktioner i Figur 3.1.

3.2 Matematisk gennemgang af vægtrestriktioner

Vægtrestriktioner defineres typisk i multiplikatorformuleringen, som ofte omtales som den duale DEA-formulering. I kapitel 1.2 og 2.2 tog vi udgangspunkt i indhyldningsformuleringen, som ofte betegnes som den primale formulering i DEA-litteraturen. De to formuleringer giver samme efficiensscore, men de kan fortolkes forskelligt og giver forskellig information. Det primale problem bruges især til at analysere, hvilke selskaber der sammenlignes med hinanden, og til at illustrere DEA grafisk. Det duale problem anvendes derimod til at analysere input- og outputvægtene på fronten.

For at introducere vægtrestriktioner til DEA, er det ofte nemmest at anvende den duale formulering, hvorfor vi nu skifter til den i resten af notatet. Den duale formulering af DEA er givet ved ligning (8)-(10).⁹

$$\max_{u,v} v \cdot y^0 \quad (8)$$

$$u \cdot x^0 \leq 1 \quad (9)$$

$$-u \cdot x^k + v \cdot y^k \leq 0, \quad k = 1, \dots, K \quad (10)$$

⁸ Mere teoretisk siges det, at transformationskurven mellem de to variable ændres med relative vægtrestriktioner.

⁹ Her anvendes konstant skalaafkast og en inputorienteret DEA.

Optimeringsproblemet løses for hvert selskab, hvor $\mathbf{v} \cdot \mathbf{y}^0$ angiver selskabets efficiensscore, $\mathbf{x}^0$ er selskabets input, $\mathbf{x}^k$ er input for selskab k , K er antallet af selskaber i modellen, $\mathbf{y}_0$ er output for det evaluerede selskab, $\mathbf{y}^k$ er output for selskab k , $\mathbf{v}$ angiver vægten på de tilhørende output, og $\mathbf{u}$ angiver vægten på de tilhørende input. Bemærk, at input og output kan være vektorer, hvis der er mere end henholdsvis ét input og ét output.

Det duale problem finder de vægte, der stiller det enkelte selskab bedst muligt, ved at maksimere selskabets vægtede output relativt til dets vægtede input under betingelse af, at ingen andre selskaber med de samme vægte kan fremstå som mere end fuldt effektive. Dermed vælges de mest fordelagtige vægte for selskabet inden for modellens fælles rammer.

Vægtrestriktioner kan nu anvendes til at begrænse, hvilke værdier vægtene fra den duale formulering må antage. Nedenfor viser vi eksempler på de to tidligere nævnte typer vægtrestriktioner, som kan tilføjes som bibetingelser til DEA i ligning (8)-(10).

$$\text{Relativ vægtrestriktion:} \quad a \leq \frac{v_1}{v_2} \leq b \quad (11)$$

$$\text{Virtuel vægtrestriktion:} \quad y_1^0 \cdot v_1 \leq c \cdot y^0 \cdot \mathbf{v}, \quad c \in [0:1] \quad (12)$$

Den relative vægtrestriktion i ligning (11) angiver, at vægten til output 1 i forhold til vægten for output 2 skal ligge mellem a og b , hvor a og b er konstanter, som fastsættes på forhånd. Den virtuelle vægtrestriktion i ligning (12) angiver, at det vægtede bidrag fra output 1 maksimalt må udgøre en bestemt andel (c) af den samlede vægtede outputværdi for det evaluerede selskab.

Referencer

- Bogetoft, P., & Otto, L. (2010). *Benchmarking with DEA, SFA, and R*. Springer Science & Business Media.
- Cazals, C., Florens, J.-P., & Simar, L. (2002). Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*, 106(1), s. 1-25.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), s. 429-444.
- Daraio, C., & Simar, L. (2007). Conditional nonparametric frontier models for convex and nonconvex technologies: a unifying approach. *Journal of productivity analysis*, 28(1), s. 13-32.
- Silva, T. D., Martins-Filho, C., & Ribeiro, E. (2016). A comparison of nonparametric efficiency estimators: DEA, FDC, DEAC, FDCH, order-m and quantile. *Economic Bulletin*, 36(1).
-