

# Budgetanalyse FFL 2008

**Analyse af Danmarks Statistik:**

## **Effektivisering af stikprøveundersøgelser**

**Maj 2007**

## Budgetanalyse FFL 2008

Printet i  
Danmarks Statistik  
Juni 2007

Oplag 150 stk.

Danmarks Statistik  
Sejrøgade 11  
2100 København Ø

Tlf. 39 17 39 17  
[www.dst.dk](http://www.dst.dk)  
[dst@dst.dk](mailto:dst@dst.dk)

Økonomi- og Erhvervsministeriet  
Finansministeriet

10. maj 2007

**FFL2008:**

## **Budgetanalyse af Danmarks Statistik stikprøveundersøgelser**

Økonomiudvalget besluttede i februar 2007 at gennemføre en budgetanalyse af muligheden for effektiviseringsgevinster i Danmarks Statistik via justeringer/optimeringer af stikprøveundersøgelsesundersøgelsernes undersøgelsesdesign.

Der er sket en kortlægning af Danmarks Statistiks arbejde med optimering af stikprøver. Det fremgår, at langt de fleste stikprøver er optimeret inden for de seneste år. På baggrund af kortlægningen er der gennemført en nærmere analyse af de tre større obligatoriske tællinger på erhvervsområdet, der ikke er blevet optimeret de senere år; regnskabsstatistikken, den nationale godstransportstatistik samt den internationale godstransportstatistik.

Analysen af regnskabsstatistikken viser, at den økonomiske gevinst ved en optimering vil være begrænset (ca. 50.000 kr. årligt) og vil indebære, at flere små virksomheder skal indberette. Det er vurderingen, at det ikke vil være hensigtsmæssigt at realisere de beregnede optimeringstiltag på regnskabsstatistikken under hensyn til de begrænsede gevinster og de administrative byrder.

Analysen af transportstatistikkerne mhp. optimering har vist, at disse ikke på nuværende tidspunkt lever op til EU-kravene og der kan derfor ikke påvises effektiviseringsgevinster.

Det **indstilles** på den baggrund, at de begrænsede muligheder for effektivisering ikke realiseres.

**Indstillingen tiltrådt i maj 2007**



## Indholdsfortegnelse:

Sammenfatning og konklusioner .....	5
1. Baggrund .....	7
2. Danmarks Statistiks tællinger .....	7
3. Optimering af stikprøver .....	8
3.1 Kvalitet .....	9
3.2 Optimering af stikprøveudvælgelse .....	10
3.3 Population; udvælgelse og opregning .....	11
4. Løbende arbejde med optimering .....	12
5. Optimering af regnskabsstatistikken .....	14
6. Optimering af transportstatistikker .....	15
7. Muligheder for besparelser ved optimering .....	16
8. Konklusion .....	17
Bilag 1: Kommissorium for budgetanalyse .....	19
Bilag 2: Oversigt over de tællinger, der danner grundlag for budgetanalysen .....	21
Bilag 3: Analyse af Regnskabsstatistikken .....	29
Bilag 4: Analyse af transportstatistikken .....	37



## Sammenfatning og konklusioner

*Formål* Som led i forberedelsen af FFL 08 besluttede regeringens Økonomiudvalg i februar 2007 at gennemføre en budgetanalyse af muligheden for effektiviseringsgevinster i Danmarks Statistik via justeringer af stikprøveundersøgelsernes undersøgelsesdesign. En forudsætning for analysen skulle være, at stikprøvens samlede usikkerhedsniveau ikke blev forringet.

Baggrunden for budgetanalysen var, at Danmarks Statistik i forbindelse med drøftelser om en udvidelse af arbejdskraftundersøgelsen vurderede, at det var muligt at optimere stikprøven yderligere.

Budgetanalysen er gennemført i samarbejde mellem Finansministeriet (formand), Økonomi- og Erhvervsministeriets departement og Danmarks Statistik.

*Indhold* I analysen indgår pligtige tællinger på det erhvervsstatistiske område med en samlet administrativ byrde på mere end 300.000 kr. årligt. Hertil kommer alle større frivillige tællinger på såvel det erhvervs- som det personstatistiske område.

Der er indledningsvist redegjort for processer i forbindelse med indsamling og bearbejdning af statistiske data, herunder den omfattende brug af digitale løsninger og hensynet til begrænsning af de administrative byrder. Dernæst er redegjort for hovedprincipperne ved optimering.

Danmarks Statistiks arbejde med stikprøver i de seneste år er i analysen behandlet i et særligt kapitel. Det fremgår heraf, at der har været et omfattende arbejde med optimering. På baggrund af dette kapitel blev det besluttet at gennemføre en nærmere analyse af de få større obligatoriske tællinger på erhvervsområdet, der ikke var optimeret de senere år; nemlig regnskabsstatistikken, den nationale samt internationale godstransportstatistik.

*Optimering af stikprøver* For regnskabsstatistikken viser analysen, at der kan opnås en begrænset reduktion i stikprøven på 120-150 virksomheder (svarende til 2-3 pct.) ved en optimering. En gennemførelse vil indebære en forøgelse af antallet af små virksomheder (med mindre end 10 ansatte) på 140-170 i stikprøven med en samtidig reduktion i antallet mellemstore og større virksomheder på ca. 300.

Den økonomiske gevinst vil være begrænset (ca. 50.000 kr.)

For transportstatistikkerne viser analysen, at den nuværende usikkerhed i statistikkerne er væsentlig større end den, der er fastlagt i den gældende EU-forordning. Der vil kunne opnås en vis reduktion i stikprøveusikkerheden ved en optimering. Reduktionen vil dog langt fra være tilstrækkelig til at imødekomme EU-kravene.

*Konklusion* Samlet vurderes, at det ikke vil være hensigtsmæssigt at realisere de beregnede optimeringstiltag på regnskabsstatistikken under hensyn til de begrænsede gevinster sammenholdt med de øgede administrative byrder for de små virksomheder.

Tilsvarende vil der ikke kunne høstes effektiviseringsgevinster for transportstatistikkerne.





## 1. Baggrund

Der stilles fra både nationalt og internationalt hold store krav til det statistiske grundlag på en lang række områder. Med de store krav følger også i mange tilfælde store udgifter. Det er derfor af væsentlig betydning, at både nuværende og fremtidige statistikker udarbejdes så effektivt som muligt under hensynstagen til, at kvalitetskravene til statistikkerne fastholdes.

Danmarks Statistik arbejder løbende med at analysere og optimere stikprøver ved indhentning af oplysninger til tællinger. Arbejdet er et led i den løbende kvalitetssikring og effektivisering af statistikproduktionen. Det skal desuden sikres, at de administrative byrder for virksomhederne begrænses.

I forbindelse med drøftelser om en forbedring af arbejdskraftundersøgelsen (AKU) i foråret 2006 skrev Danmarks Statistik, at der med en 3-dobling - men i øvrigt uændret tilrettelæggelse - af undersøgelsen ville kunne opnås næsten en halvering af stikprøveusikkerheden. Det blev samtidig vurderet, at såfremt stikprøven blev optimeret ved øget anvendelse af registerinformation og en ændret panelstruktur, var der mulighed for at opnå samme effekt på stikprøveusikkerheden ved en forøgelse af stikprøven med en faktor 2,2.

Med disse erfaringer besluttede Økonomiudvalget at gennemføre en budgetanalyse for at vurdere om der kunne opnås besparelser ved optimering af andre af Danmarks Statistiks større stikprøver. Kommissoriet for budgetanalysen er vist i bilag 1.

## 2. Danmarks Statistiks tællinger

### *Tællinger i budgetanalysen*

Budgetanalysens udgangspunkt er overvejende fokuseret på erhvervsstatistiske tællinger med en årlig administrativ byrde (AMVAB) i 2005 på mere end 300.000 kr. (2004-priser). Dette indebærer en samlet dækningsgrad på 97 pct. af de obligatoriske indberetninger. Herudover inddrages frivillige tællinger til erhvervsstatistikken i analysen.

Derudover indgår de større tællinger på det personstatistiske område. Det bemærkes, at tællinger på dette område er frivillige.

### *Stikprøvestørrelse og svarprocenter*

Der er tale om meget varierende størrelse af de obligatoriske tællinger, hvor de største tællinger (intrastat, lønstatistik, struktur-tællingen i landbruget og detailomsætningsindekset) omfatter mere end 20.000 årlige indberetninger, mens den mindste tælling (international godstransport) omfatter ca. 2.400 årlige indberetninger.

Tilsvarende varierer tællingsomfanget i de frivillige tællinger fra 89.000 indberetninger (Arbejdskraftundersøgelsen) til 5.500 (konjunkturbarometret for industri)

For de obligatoriske tællinger på erhvervsområdet er svarprocenten typisk tæt på 100 pct., mens den for frivillige tællinger på såvel erhvervs- som personområdet er 60-70 pct.

<i>Indsamlingsmetode</i>	For alle erhvervsstatistikker (såvel obligatoriske som frivillige) indhentes oplysninger fra virksomhederne via udsendelse af spørgeskemaer. På personstatistikområdet indhentes oplysninger hovedsageligt ved interview.
<i>Indberetningsmuligheder</i>	Virksomhederne kan for de spørgeskemabaserede tællinger indsende oplysninger på de tilsendte blanketter. For de største pligtige tællinger er der samtidig mulighed for, at virksomhederne kan indberette via Virk.dk. For enkelte tællinger (Intrastat, Svinetællingen og Lønstatistikken) er der etableret særlige - og meget anvendte - digitale indberetningsløsninger.  Danmarks Statistik vil allerede for indberetninger til regnskabsstatistikken i 2007 åbne muligheder for indberetninger via en fælles XBRL-taxonomi, der er etableret i samarbejde med Erhvervs- og Selskabsstyrelsen og Skat.  For interviewundersøgelserne foregår langt hovedparten af databehandlingen digitalt via CAI (Computer Assisted Information) programmel.  Samlet er der mulighed for digital indberetning for i alt 70 pct. af de samlede indberetninger. Danmarks Statistik vil etablere nye indberetningsløsninger via virk.dk, når den nye platform er etableret.
<i>Bearbejdning af data</i>	Danmarks Statistik behandler de indkomne skemaer enten via scanning eller ved manuel indtastning. Scanning indebærer – på linje med digitale indberetninger – en umiddelbar mulighed for overførsel til Danmarks Statistiks databaser og dermed fejlsøgning og kvalitetssikring. Omkring 87 pct. af indberetningerne til frivillige og obligatoriske tællinger scannes i 2007.
<i>Outsourcing</i>	Danmarks Statistik har løbende fokus på mulige gevinster ved outsourcing. Således varetages langt hovedparten af interviewopgaverne af eksterne leverandører.  På erhvervsstatistikområdet indebærer en enstregen og integreret udsendelse af spørgeskemaer samt den omfattende digitalisering i behandling af data, at effektiviseringsgevinster ved outsourcing er tvivlsomme.
<i>Besparelsesmuligheder ved tilpasning af stikprøver</i>	De faste omkostninger ved en tælling (udtræk, opregning, kvalitetssikring, analyse og formidling) er ganske betydelige og er i stort omfang uafhængige af stikprøvens omfang. Samtidig indebærer den voksende brug af scanning af indberetninger, at de variable omkostninger/marginalomkostningerne udgør en begrænset og aftagende andel af de samlede omkostninger.
<i>Administrative byrder</i>	Danmarks Statistik har fokus på at sikre, at der af kvalitetsmæssige hensyn sker en optimering af stikprøver, så grundlaget for statistikudarbejdelsen tilvejebringes med så begrænset belastning af virksomhederne som muligt.

### 3. Optimering af stikprøver

*Formålet med optimering* Formålet med optimering er både at reducere stikprøveusikkerheden og den systematiske bias, så den samlede usikkerhed bliver begrænset og statistikken giver et retvisende billede af samfundsudviklingen.

I arbejdet med optimering af stikprøver inddrages relevant supplerende information fra registre om virksomhedernes branche, beskæftigelse, omsætning mv. På baggrund af disse kilder vurderes det, hvordan stikprøven kan

stratificeres for, at en række forskellige mål for kvalitetsniveau kan sikres på en omkostningseffektiv måde.

Omfanget af supplerende informationer ved udvælgelsen af stikprøven, påvirker kvaliteten i statistikken. Hvis stikprøven er meget skæv, vil anvendelse af gode registre først og fremmest kunne bidrage til en reduktion af bias og mindske stikprøveusikkerheden.

Fastlæggelse af den optimale stikprøvestørrelse er afhængig af de overordnede krav til usikkerhed samt fx krav udmøntet i bindende EU-retsakter.

Resultaterne fra stikprøveundersøgelserne skal efterfølgende opregnes. Også her er det afgørende, at grundlaget for opregningen tilrettelægges, så opregningen (estimationen) kan ske med begrænset usikkerhed.

Optimering sker individuelt for hver stikprøve. Det er ikke muligt ud fra en analyse af muligheder for optimering af en eller få stikprøver at generalisere til potentialet for alle stikprøver i Danmarks Statistik.

*Særlige regler for små virksomheder*

Af hensyn til at begrænse de administrative byrder har Danmarks Statistik i dataleverandørpolitikken fastlagt særlige regler for udvælgelse af mindre virksomheder. Det indebærer, at disse underrepræsenteres i de udvalgte stikprøver i forhold til, hvad der vil være optimalt. For at sikre den nødvendige kvalitet må der derfor inddrages et større antal mellemstore og store virksomheder i stikprøven.

### 3.1 Kvalitet

*Eksterne krav*

Udgangspunktet for Danmarks Statistiks stikprøveundersøgelser er enten fastlagt eksternt eller ved en række generelle kvalitetskrav til usikkerheden.

*Statistikken skal kunne give et retvisende billede*

På numeriske variabler (fx beløb, arbejdstid eller indeks) er kravet en *relativ* standardfejl (varianskoefficient) på højst 1 pct. af niveauet for hele populationen samt evt. hovedbrancher og op til 5 pct. på undergrupper for de vigtigste variabler.

For kategoriske variabler (fx ja/nej spørgsmål) er kravet en *absolut* stikprøvefejl på højst 2 pct. med et 95 pct. statistisk sikkerhedsinterval for hele populationen samt evt. hovedbrancher og højst 10 pct. for en undergruppe.

Hvis statistikkens hovedformål er at beskrive udviklingen, skal kravet opfyldes både for den aktuelle statistik og for stigningen i forhold til den sidste periode.

*Ikke større stikprøve end der er nødvendig*

Danmarks Statistik søger at opfylde kvalitetskravene med så små stikprøver som muligt ved at optimere stikprøverne vha. registre, jfr. nedenfor.

Hvis kvalitetskravene ikke kan opfyldes med den aktuelle stikprøve, skal mulighederne for at optimere denne være udtømte, før det eventuelt besluttes at forøge stikprøven. En endelig beslutning om en forøgelse af stikprøven vil samtidig afhænge af økonomiske hensyn og omfanget af indberetningsbyrden.

Hvis en analyse viser, at det er muligt at optimere stikprøven og opfylde kvalitetskravene med en mindre stikprøve, bliver stikprøven reduceret.

Når der er tale om variationer i populationen (eksempelvis hugsttællinger) baseres stikprøver typisk på flere års erfaringer. Såfremt der kommer en periode med en mere stabil udvikling vil dette medføre, at spredningen i population reduceres. Når der er rimelig sikkerhed for, at ændringen er varig ned-sættes stikprøven.

### 3.2 Optimering af stikprøveudvælgelse

Ved optimering af stikprøver skal der tages hensyn til den samlede stikprøve-usikkerhed, der afhænger af stikprøvedesignet samt –størrelsen, og den systematiske fejl (bias) fra bl.a. bortfald, ufuldstændig population og underrapportering. Tilsammen giver de to fejltypen den samlede statistiske usikkerhed, der afgør, hvor retvisende statistikken er.

*Hvilke variabler skal der optimeres efter*

Udgangspunktet for arbejdet med optimering er sikkerheden af den eller de variabler, der indsamles eller ønskes optimeret.

I næsten alle stikprøvetællinger er der flere variabler, fx omsætning, ordrer, investeringer eller henlæggelser i erhvervsstikprøver. Hvis man optimerer efter den ene variabel, bliver sikkerheden for de andre ikke optimale (mindst mulige).

*Niveau eller forskel?*

Der skal herefter tages stilling til måling af variablerne fx. niveau eller forskel i forhold til den forrige tælling. De to forskellige mål har stor betydning for, hvordan en given stikprøvestørrelse optimeres (fordeles) på fx brancher, da der alt andet lige skal fordeles flere enheder til brancher med størst usikkerhed. Hvis ændringer skal belyses, vil det normalt forøge kravet til stikprøvestørrelsen.

*Optimering af stratificering*

En optimal allokering (Neumann allokering) af en stikprøve betyder, at der allokeres en større stikprøve til de undergrupper, der har størst spredning, og hvor populationen er størst. Men det vil samtidig medføre, at eksempelvis små brancher, der publiceres efter, får en lille stikprøve og en for stor usikkerhed. Af hensyn til en samlet retvisende statistik indebærer det derfor, at noget af stikprøven bruges til at sikre, at alle brancher har en rimelig minimumssikkerhed.

Stratificeringen kan bruges til enten at optimere den totale stikprøveusikkerhed eller til at optimere usikkerheden i de enkelte undergrupper og sammenligningen mellem dem. De to formål er i direkte modstrid med hinanden, så der må prioriteres.

I virksomhedsstikprøver er brancher – og ikke mindst størrelse – normalt positivt korreleret med alle variabler, der skal måles. Stratificering af virksomhedsstikprøven vil derfor have en positiv virkning for alle variabler.

I personstikprøver er der ikke en tilsvarende sammenhæng med alle variablerne i de enkelte skemaer og registervariabler. En optimering efter én variabel kan derfor være skadelig for andre variabler.

I personstikprøver vil der derfor først og fremmest ske en efterstratifikation, hvor der ved opregning af de enkelte variable mod registeroplysninger, kan reduceres i den tilfældige og den systematiske fejl. Dette skyldes, at personregistrene er meget omfattende og altid opdateret med civilstand, alder og køn, og derudover findes historiske oplysninger om fx uddannelse, indkomst og beskæftigelse.

*Paneler* Paneler, hvor de samme personer eller virksomheder indgår flere gange, kan reducere den statistiske usikkerhed af ændringer fra den ene tællingsperiode til den anden. Paneler vil dog også have en negativ effekt på biasen eller stikprøveusikkerhed, idet færre svarer anden eller tredje gang ligesom det giver øget bias, at ændringer i populationen såsom indvandring, nye firmaer og firmaer i fremgang ikke er med i stikprøven.

*Cut-off* Hvis grænsen for, hvorfra der udvælges virksomheder, er høj, har det kritisk betydning for, hvor retvisende statistikken er. I mange virksomhedsundersøgelser er det besluttet, at små virksomheder under en given størrelse ikke kan udvælges. Hvis de små virksomheder udgør en synlig del af den samlede omsætning, har cut-off-grænsen en kritisk negativ betydning for biasen i undersøgelsen. Strukturændringer, hvor fx flere virksomheder kommer over cut-off-grænsen pga. opkøb, vil således resultere i en tilsyneladende stigning, hvis der ikke forsøges korrigeret herfor. Opregning for cut-off-grænser giver øget behov for flere udvalgte virksomheder lige over cut-off-grænsen.

### 3.3 Population; udvælgelse og opregning

*Opdateret, korrekt og fuldstændig population*

Registerdata bliver brugt både til før- og efterstratifikation, herunder korrektion for bortfald. Det sidste er specielt vigtigt i de frivillige undersøgelser. Når kvinder er mere tilbøjelige til at svare end mænd, kan hjælpeinformation fra registre bruges til at justere vægte, så de mænd, som har svaret, får højere vægte. Er stikprøven meget skæv i forhold til populationen, kommer vægtene til at være meget forskellige. I virksomhedsundersøgelser kan der vægtes for nye virksomheder. Stor spredning i vægtene som følge af skævt bortfald eller nye virksomheder giver en øget stikprøveusikkerhed.

*Registerdata på personstatistikområdet*

På personområdet er der gode registre. Personregistrene anvendes til at øge kvaliteten på målvariablene (fx ledige, indkomst, forbrug). Kvalitetsøgningen er både i form af mindre stikprøveusikkerhed og mindre bias. Hvis der er bias i stikprøven (fx som følge af bortfald, ufuldkomne registre eller forskerbeskyttelse) vil registrene kunne rette herpå. En korrektion for stort bortfald vil dog markant reducere mulighederne for samtidig at reducere stikprøveusikkerheden.

*Registerdata på erhvervsstatistikområdet*

På virksomhedsområdet er der langt færre oplysninger til rådighed. Det Erhvervs Statistiske Register (ESR) har oplysninger om branche, omsætning, beskæftigelse og ejerform. Traditionelt bliver erhvervsstikprøver stratificeret efter branche og størrelse.

Som oftest er der meget lidt at vinde ved at anvende samme variable til opregning, som der er stratificeret efter. Informationen fra registre kan i alt væsentlighed kun bruges én gang til at reducere usikkerheden eller mindske stikprøvestørrelsen. Den primære effekt ved at udnytte registerdata yderligere en gang til opregning er, at det reducerer skævheden pga. tilgangen af nye virksomheder.

*Registerdata på  
Landbrugsområdet*

På landbrugs- og transportområdet er der gode registre til rådighed. For landbrugsområdet skyldes det i mange tilfælde EU-støtteordninger, hvor areal, antal dyr og støtte registreres.

Landbrugsregisterene giver gode muligheder for at stratificere og beregne avancerede vægte, så stikprøveusikkerhed og bias mindskes.

#### **4. Løbende arbejde med optimering**

*Oversigt over obligatoriske og frivillige tællinger*

I bilag 2 er der i en række tabeller sammenfattet de væsentligste resultater af arbejdet med optimering for de tællinger, der er indgået i budgetanalysen.

*Tællinger, hvor stikprøven er optimeret inden for de seneste år*

Der er for en række tællinger med repræsentative stikprøver i de seneste år foretaget et omfattende analysearbejde med henblik på optimering og gennemgang af opregningsprocedurer, jfr. beskrivelsen i bilag 2, tabel 1. Formålet har været at sikre, den nødvendige kvalitet i statistikkerne opnås på en omkostningseffektiv måde. Arbejdet har omfattet hovedparten af Danmarks Statistiks stikprøver.

Som en opfølgning af optimeringsarbejdet sker der en løbende reoptimering, hvilket erfaringsmæssigt ikke ændrer stikprøvernes størrelse nævneværdigt.

Resultaterne har været, at for enkelte tællinger er stikprøven nedsat, mens det for andre vurderes, at stikprøven burde øges for at leve op til kravene om retvisende statistik. Indtil nu er der dog ikke sket en forøgelse af stikprøver, på grund af konsekvenser for ressourcer eller virksomhedernes indberetningsbyrde.

For hovedparten af de berørte tællinger er stikprøven fastholdt, men som et resultat af optimeringen er det lykkedes at opnå en reduktion i usikkerheden på de detaljerede variable i statistikkerne.

Særligt bemærkes, at der i 2005 blev fremsat et politisk ønske om en forbedret statistisk belysning af arbejdsmarkedet. Herunder blev mulighederne for en væsentlig udvidelse af arbejdskraftundersøgelsen drøftet, idet denne undersøgelse i en årrække ikke har levet op til EU-kravene om stikprøven. Usikkerhedsniveauet var tidligere omtrent det dobbelte af de tilsvarende undersøgelser i de øvrige nordiske lande.

Det betydelige arbejde med optimering af øvrige stikprøver i Danmarks Statistik indebar sammen med usikkerheden om det fremtidige omfang af arbejdskraftundersøgelsen, at arbejdet med en større optimering af stikprøven blev udsat. I 2006 forelå en regeringsbeslutning om en forbedret beskæftigelsesstatistik. I forbindelse med drøftelserne om udvidelsen af stikprøven vurderede Danmarks Statistik, at en forøgelse af stikprøven med en faktor 2,2 og en samtidig optimering kunne indebære næsten en halvering af stikprøveusikkerheden. Arbejdskraftundersøgelsen bliver derfor optimeret i 2007.

*Tællinger, der ikke er stikprøver, men er undergivet særlige krav eller har en særlig tilrettelæggelse*

For en række tællinger er der krav fra EU til dækningsomfang (fx udmøntet i krav om dækning af samlet omfang af udenrigshandel eller omsætning), jf. beskrivelsen i Bilag 2, tabel 2.

For at sikre den nødvendige dækning på disse områder udvælges virksomheder efter størrelse, så alle virksomheder, der er større end den angivne "cut-off" grænse, spørges. På denne måde sikres det, at små og en række mellemstore virksomheder fritages for indberetning.

Også på andre områder har Danmarks Statistik valgt at pålægge de større virksomheder indberetning, mens de mindste fritages. Udover hensynet til de administrative byrder ved indberetning er et centralt hensyn, at der skal sikres et grundlag for konsistente tidsserier til brug for bl.a. nationalregnskabet. En oversigt er angivet i Bilag 2, tabel 3.

Tællinger underkastet særlige EU-krav eller hvor der sker en særlig tilrettelæggelse, er ikke stikprøver. Udvalgelse af undersøgelsesgrundlag og statistisk tilrettelæggelse af opregningen indebærer derfor ikke en optimering.

*Øvrige tællinger* I Bilag 2, tabel 4 og 5 er vist en oversigt over de øvrige tællinger med angivelse af særlige forhold.

For at sikre tilstrækkeligt pålidelige tidsserier (herunder på individniveau) i regnskabsstatistikken arbejdes med faste panelstrukturer, hvor de større virksomheder over en længere årrække skal indberette. Dette kan indebære begrænsninger i kriterierne for at optimere stikprøven.

For udarbejdelsen af datagrundlaget til stikprøver for prisindeks bemærkes, at dataleverandørerne typisk skal levere en række prisoplysninger, så der sikres en bred dækning af prisudviklingen. Stikprøven justeres løbende for sikre, at alle varegrupper er repræsenteret.

Der er på nuværende tidspunkt et arbejde i gang mhp. at analysere Forbrugerforventningsundersøgelsen for at begrænse skævheder pga. den udbredte anvendelse af forskerbeskyttelsesordningen.

De frivillige tællinger på erhvervsområdet er overvejende tællinger med få spørgsmål og hvor spørgsmålene er af kvalitativ karakter. Der er tale om skemaer med en lav kompleksitetsgrad samtidig med at de udsendes i et betydeligt omfang til – især – større virksomheder.

*Analyse og optimering* På baggrund af den samlede gennemgang af tællinger blev det besluttet at afgrænse arbejdet med optimering af stikprøver til de stikprøver, der ikke var optimeret de seneste år, dvs. regnskabsstatistikken og den nationale og internationale transportstatistik.

Tællinger, der er optimeret inden for de seneste år, holdes uden for budgetanalysen. Baggrunden for beslutningen var at omfanget af supplerende information til arbejdet med optimering vurderedes at være stort set udtømt, samt at de hidtidige opfølgende optimeringer kun har indebåret en ændring i stikprøvestørrelsen i meget begrænset omfang.

Samtidig blev det besluttet at holde cut-off tællinger uden for budgetanalysen, idet der dels var tale om eksterne (især EU) krav dels krav fastlagt under hensyn til såvel nationalregnskabet og en reduktion af de administrative byrder for små virksomheder.

Endelig indgår konjunkturbarometre og forbrugerforventningsundersøgelsen ikke. Der er her tale om tællinger vundet i EU-udbud, og hvor det ikke vides,

om Danmarks Statistik fortsat skal varetage disse opgaver i de kommende år, samt en tælling, der er ved at blive optimeret.

## 5. Optimering af regnskabsstatistikken

### *Data om regnskabsstatistikken*

Regnskabsstatistikken baseres på ca. 8.750 spørgeskemaindberetninger, svarende til 4,6 pct. af den samlede population. Alle virksomheder med mere end 50 ansatte skal indberette, mens de resterende ca. 5.400 virksomheder udvælges ved en stikprøve.

For de øvrige mellemstore EU-lande er den procentvise andel af indberettere til regnskabsstatistikken 8,4 pct. Den lille danske andel skyldes en omfattende brug af registerinformation fra de administrative registre i Skat.

Det vurderes, at der allerede på nuværende tidspunkt sker en fuld anvendelse af tilgængelig registerinformation og at der ikke foreligger yderligere information på dette område, der kan anvendes i optimeringen.

### *Etablering af indikatorer for effektiviteten i stikprøven*

Anvendelsen af registerinformation ved etablering af stikprøven indebærer, at det ikke er muligt med standardmetoder at beregne stikprøveusikkerheden. I stedet må man basere sig på en simulering af usikkerheden og på denne baggrund udarbejde en række indikatorer for effekten af en omlægning af stikprøven.

Stratificeringen sker for tre beskæftigelsesgrupper<sup>1</sup> og fordelt på brancher, så stikprøven inden for den enkelte branche stort set er proportional med antal ansatte og omsætning.

Som konsekvens af dataleverandørpolitikken er der dog en vis begrænsning i stikprøveudvælgelsen, idet virksomheder med under 50 ansatte kun skal deltage i et begrænset antal indberetninger i løbet af en 10 årsperiode.

### *Optimering af stikprøven*

Arbejdet med optimering af stikprøven skete med udgangspunkt i regnskabsstatistikens oplysninger om investeringer. Der blev udarbejdet simple estimater af usikkerheden på baggrund af de aktuelle stikprøvestørrelser og den anvendte registerinformation. Inden for de enkelte størrelsesgrupper og hovedbrancher blev der beregnet den teoretisk optimale fordeling af stikprøven (en Neumann allokering). Denne stikprøve blev sammenlignet med den aktuelle stikprøve.

Det vurderes, at der kan være begrænsede muligheder for optimering og dermed en reduktion af stikprøven for regnskabsstatistikken. Resultaterne er dog behæftet med usikkerhed, og det kan ikke udelukkes, at der er tale om tilfældige udsving. Samtidig er analyseresultaterne ikke entydige mht. muligheder for optimering inden for de enkelte størrelsesgrupper.

### *Faglig vurdering af optimeringstiltag*

Ud fra en samlet faglig vurdering bør der ikke ske en reduktion i antallet af mindre virksomheder i nogen hovedbranche (pga. den lille andel i stikprøven og en stor variation i omsætningen i denne gruppe). Usikkerheden vedrørende især de større virksomheders andel af stikprøven stiller yderligere spørgsmål ved, om der skal ske en ændring af den eksisterende udvælgelse her.

---

<sup>1</sup> Beskæftigelsesgrupper: 5-9, 10-19 og 20-49 ansatte. Herudover gælder, at alle virksomheder med mere end 50 ansatte altid vil skulle indberette til regnskabsstatistikken.



*Muligheder for optimering* Under disse forudsætninger vil stikprøven kunne nedsættes med 120 - 150 virksomheder, hvilket svarer til en reduktion af stikprøven på 2-3 pct. af den nuværende stikprøvestørrelse.

Optimeringen kan ske ved at øge antallet af mindre virksomheder (med mindre end 10 ansatte) i stikprøven med 140–170 (svarende til 12-15 pct. i denne gruppe), samtidig med en reduktion i antallet af mellemstore og større virksomheder på omkring 300 virksomheder (svarende til ca. 7 pct. i denne gruppe).

## 6. Optimering af transportstatistikker

*Data om transportstatistikkerne* Transportstatistikkerne belyser den godstransport danske lastbiler gennemfører hhv. nationalt (i Danmark) og internationalt. Statistikkerne er organiseret som to selvstændige undersøgelser, hvor der foretages separate indsamlinger af data.

Stikprøveudvælgelsen til national godstransport udgør under 2 pct. af den samlede population. Stikprøven anvendes samtidig som udgangspunkt for tællingen vedrørende international godstransport.

*EU-krav til sikkerhed* For transportstatistikker er der EU-krav om maksimale variationskoefficienter på 2,5 pct. for oplysninger om hhv. antal kørte kilometer med læs, pålæst godsmængde og transportarbejde i alt.

Med den nuværende stikprøve er de aktuelle variationskoefficienter beregnet. Det bemærkes dog, at den indirekte udvælgelse af stikprøven for international godstransport indebærer, at der er betydelig bias i denne tælling og det dermed kun er muligt at skønne over variationskoefficienterne for denne tælling.

*Delvis manglende dansk opfyldelse af krav til variationskoefficienter* Sammenfattende er EU-kravene til transportstatistikkerne opfyldt vedrørende antal kørte kilometer med læs, mens det modsatte gør sig gældende for pålæst godsmængde og transportarbejdet i alt, der er de to vigtigste variabler.

Med de nuværende stikprøver er der derfor behov for en markant forøgelse af stikprøven for at kunne leve op til EU-kravene.

*Optimering af stikprøver* Ved optimering af stikprøverne til transportstatistikkerne er søgt inddraget hjælpeinformation, der kan bidrage til en reduktion af variansen og/eller reduktion i bias. Der er her søgt inddraget variable, der ikke tidligere har været benyttet i optimeringen (alder, lastevne og aksler).

Det konstateres, at der ved inddragelse af supplerende information kun kan opnås en begrænset (5-10 pct.) reduktion i variansen. Dermed er mulighederne for reduktion af stikprøven ved optimering meget begrænsede.

Ved at sammenholde de beregnede variationskoefficienter med EU-kravene kan det konstateres, at selv efter en gennemført optimering vil der skulle ske op til en fordobling af stikprøven ved national godstransport for at overholde EU-kravene, mens der for international godstransport er behov for en udvidelse med ca. 50 pct.

*Sammenfatning* Analysen viser, at der er behov for at ændre tællingen vedrørende international godstransport, at mulighederne for optimering i begge tællinger er begrænsede og at der selv efter optimering er behov for en væsentlig udvidelse af stikprøverne til begge transportstatistikker for at leve op til EU-kravene.

## 7. Muligheder for besparelser ved optimering

Sammenfattende vurderes i afsnit 5 og 6, at der er begrænsede effekter for størrelsen af stikprøven ved optimering. Hertil kommer, at der forud for beslutning om eventuel ændring af stikprøvestørrelsen skal inddrages en række supplerende forhold.

*Internationale forpligtelser* Der er forpligtende EU-krav til variationskoefficienter i transportstatistikkerne. Den gennemførte analyse viser, at på en række områder er koefficienterne langt højere end de krævede.

*Besparelsesmuligheder* Danmarks Statistik har vurderet marginalomkostningerne pr. skema til regnskabsstatistikken ud fra omkostninger til print, forsendelse af skemaer og rykning med inddragelse af den i institutionen faktiske anvendte arbejdstid.

Det skønnes, at omkostningerne ved en optimering af regnskabsstatistikken kan reduceres med godt 50.000 kr. i dataindsamlingen. Dette skal sammenholdes med betydelige afledte meromkostninger i Danmarks Statistik vedrørende administration af tællingen.

De potentielle besparelsesmuligheder skal sammenholdes med usikkerheden i beregningerne. Danmarks Statistik vil normalt først tilpasse stikprøven, såfremt gevinsterne ved en optimering er signifikante (dvs. i praksis en ændring i stikprøven på over 5 pct.) eller viser sig sikre.

*Administrative byrder* Regeringens arbejde med reduktion af virksomhedernes administrative byrder er i høj grad rettet mod lettelser for mindre virksomheder, der typisk har begrænsede administrative ressourcer.

Dette er afspejlet generelt i dataleverandørpolitikken, hvortil kommer, at den nuværende praksis for regnskabsstatistikken indebærer, at mindre virksomheder maksimalt kan pålægges at indberette til statistikken hvert 10. år, mens mellemstore og store virksomheder kan pålægges at indberette til statistikken i hhv. 2 og 5 ud af 10 år.

En forøgelse af antal indberetninger for mindre virksomheder vil indebære et forøget arbejdspress hos de små virksomheder, et arbejde der typisk skal udføres manuelt.

*Arbejde i Danmarks Statistik* For arbejdet i Danmarks Statistik vil en ændring af optimering af stikprøven indebære et øget antal indsendte papirskemaer og dermed større manuel behandling af skemaer, idet det overvejende er mindre virksomheder der indsender deres interne regnskab, mens større virksomheder udfylder blanketten til spørgeskemaet.

## 8. Konklusion

<i>Løbende optimering af stikprøver</i>	<p>Danmarks Statistik arbejder løbende med at optimere stikprøver, så indsamling af oplysninger til en korrekt og retvisende statistik sker med en begrænset påvirkning af dataleverandører.</p> <p>I budgetanalysen er redegjort for Danmarks Statistiks arbejde med optimering samtidig med, at de løbende arbejdsprocesser med indsamling og bearbejdning af data er beskrevet.</p>
<i>Udvælgelse af statistikker til analyse</i>	<p>På baggrund af gennemgangen er der peget på tre tællinger (regnskabsstatistikken, transportstatistikkerne vedrørende hhv. international og national transport) med henblik på at undersøge, om der kan opnås effektiviseringsgevinster ved en optimering.</p>
<i>Ikke hensigtsmæssigt at optimere regnskabsstatistikken</i>	<p>For regnskabsstatistikken kan der opnås en begrænset effektivisering ved optimering af stikprøven. Den økonomiske gevinst er imidlertid begrænset (ca. 50.000 kr.) samtidig med at en optimering vil indebære, at et væsentligt større antal små virksomheder (med under 10 beskæftigede) skal pålægges indberetning.</p> <p>Samlet vurderes det ikke hensigtsmæssigt at realisere optimeringstiltag på regnskabsstatistikken.</p>
<i>Behov for markant øget stikprøve i transportstatistikkerne</i>	<p>For transportstatistikkerne viser analysen, at den nuværende usikkerhed i statistikkerne er væsentlig større end den, der er fastlagt i en EU-forordning. Der vil kunne opnås en vis reduktion i stikprøveusikkerheden ved en optimering. Reduktionen vil dog langt fra være tilstrækkelig til at imødekomme EU-kravene.</p>



## Bilag 1: Kommissorium for budgetanalyse

### FFL 2008: Analyse af Danmarks Statistik – effektivisering af stikprøveundersøgelser

#### Analysens indhold

Budgetanalysen skal undersøge og medvirke til realisering af effektiviseringspotentialer i et antal af de nuværende større stikprøveundersøgelser, der foretages af Danmarks Statistik, herunder for eksempel den løbende undersøgelse af forbrugerforventninger samt konjunkturbarometrene inden for bygge og anlæg, industri og serviceerhverv. Analysen skal afdække muligheden for effektiviseringsgevinster via justeringer af stikprøveundersøgelsernes undersøgelsesdesign. Der skal bl.a. ses nærmere på effektiviseringspotentialet ved øget brug af registerdata, ændringer i panelstrukturen og anvendelse af forskellige interviewformer.

Analysen vil kortlægge nettogevinsten ved justeringer af stikprøver. Det vil være en forudsætning for analysen, at stikprøveundersøgelsens samlede usikkerhedsniveau ikke forringes, således at analysen ikke medfører ændringer i det statistiske output, ligesom der skal tages hensyn til en eventuel påvirkning af virksomhedernes administrative byrder.

Analysen skal ses i sammenhæng med, at Danmarks Statistik hvert år for centrale statistikker gennemfører en optimering af stikprøver og de tilknyttede analyser som led i det løbende arbejde med effektivisering af statistikproduktionen.

#### Fordele

Der stilles i dag fra både nationalt og internationalt hold store krav til det statistiske grundlag på en lang række områder. Med de store krav følger også i mange tilfælde store udgifter. Det er derfor af væsentlig betydning, at både nuværende og fremtidige statistikker udarbejdes så effektivt som muligt under hensynstagen til, at kvalitetskravene til statistikkerne fastholdes.

#### Provenu

Budgetanalysens provenu er uvist.

Tabel 1.

#### Provenuvirkning

Mio.kr. 2008 pl.	2008	2009	2010	2011
Budgetanalyse af Danmarks Statistik - effektivisering af stikprøveundersøgelser	?	?	?	?
I alt				

En gennemførelse af budgetanalysens forslag forventes ikke at medføre behov for lovændring.

#### Eventuelle risici

Budgetanalysens forslag forventes ikke at medføre politiske risici, da der ikke forventes ændringer i det statistiske output.



Tabel 1. Stikprøver der er analyserede mhp. optimering og opregning i de seneste år

	Beskrivelse af optimeringsprincipperne	Effekt på stikprøve, formål med optimeringen og effekt på den samlede usikkerhed	Overvejelser om ændret stikprøvestørrelse i forhold til usikkerhedsniveau
Udenrigshandel med tjenester	Stikprøven er optimeret i 2004 på baggrund af virksomhedernes indberetninger til bankerne om handel med tjenester med udlandet. Stikprøven i 2006 og 2007 har større usikkerhed end den fra det første produktionsår 2005, som følge af vanskeligheder med at opdatere populationen.	Stikprøvestørrelsen er bibeholdt af ressource- og byrdemæssige årsager, selv om usikkerheden har været stigende - både stikprøveusikkerhed og bias pga. ufuldstændig population.	Burde øges.
Svinebestanden	Optimering/opregningsprincipper er fastlagt i 2003. Bedre udnyttelse af registre. Ingen panelstruktur, da det er niveau, som er målvariablene. Der er anvendt en Neumann allokering til fordeling af bedrifter i de 24 strata.	Stikprøven er reduceret fra ca. 18.000 til 12.000 om året. Strukturel udvikling i stikprøvegrundlaget, har medført stigende samlet usikkerhed.	Burde øges, men fastholdes af byrdemæssige årsager.
Landbrugs- og gartneritællingen	Stikprøvedesignet følger EU's principper, der selv opregner, når de modtager data. DST opregner efter samme principper så tallene er sammenlignelige. Landbrugs- og gartneritællingen bruges som udtræksgrundlag for svinetællingen og høsttællingen. Stratificeringen efter region, størrelsesgrupper og landbrugsbranche opfylder EU's krav og optimeringen af stikprøven er en Neuman-allokering efter standarddækningsbidraget.	Stikprøvestørrelsen er fastlagt af aftalen med EU og stratificeringen, optimeringen og opregningen opfylder EU's krav. EU stiller krav om udvælgelse af 25.000 bedrifter og betaler omkostningerne ved tællingen.	Ingen ændring.
Høst af korn	Optimering/opregning er gennemført i 2004/2005. Der er ændret i stratainddelingen. Ingen paneler, da det er niveauet, der er målvariablen. Der er anvendt en Neuman allokering til fordeling af stikprøven i 120 strata.	Stikprøven er reduceret fra ca. 3 900 til ca. 3 300 om året. Stikprøven er yderligere reduceret til 3.000 i 2007.	B bliver nedsat i 2007.
Lønstatistik, kvartalsvis	Danmarks Statistik indsamler en stikprøve blandt de virksomheder, der ikke er medlem af DA og FA. Stikprøven er stratificeret efter 50 branchekoder og 4 størrelsesgrupper og optimeringen af stikprøven er en Neuman-allokering efter lønudgiften. Der opregnes med ratioestimer, der bruger virksomhedernes ATP indbetalinger som hjælpeinformation.	Effekten af optimeringen er at udnytte stikprøven på 2.000 bedst muligt så der kan produceres sikre lønindeks.	Ingen ændring.
Fraværstatistik	Der er tale om en ny statistik, hvis første indsamlingsår er 2006. Stikprøven er optimeret allokeret proportionalt efter antal ATP-beskæftigede. Der opregnes med den samme hjælpeinformation som for lønstatistikken.	Stikprøvestørrelsen er fastsat efter forhandlinger med Beskæftigelsesministeriet og Dansk Arbejdsgiverforening.	Ingen ændring.

	Beskrivelse af optimeringsprincipperne	Effekt på stikprøve, formål med optimerin- gen og effekt på den samlede usikkerhed	Overvejelser om ændret stikprøvestør- relse i forhold til usikkerheds- niveau
Arbejdskraftundersøgelsen (AKU)	AKU blev optimeret efter CRAM-ledighed i 2003 med en Neuman-allokeringen efter andelen af ledige. Der opregnes efter køn, alder, beskæftigelsesstatus, indkomst og uddannelse. Denne opregning vil i 2007 blive optimeret yderligere, da det vurderes, at det er muligt at udnytte registerinformationerne bedre. Der bliver tale om en meget omfattende analyse for at hente al den relevante information fra registrene mht. hovedvariablerne i AKU og begrænse en evt. negativ indflydelse på andre variabler.	Den tidligere optimering i 2003 resulterede i en optimering på ca. 7 %. AKU analyseres for tiden og stikprøvestørrelsen er aftalt med Finansministeriet og Økonomi- og Erhvervsministeriet.	Ingen ændring (udover det aftalte).
Industriens produktion og ordreindgang	Optimering gennemført og estimationsprogrammet er udarbejdet i 2005/06. Strata ændret fra ca. 250 til 77 (brancher). Firmaerne er udvalgt proportionalt efter størrelse.	Betydelig ændring i panelstrukturen (mindre brug af paneler). Stikprøvestørrelsen er ikke ændret, da undersøgelsen offentliggøres på detaljeret niveau.	Ingen ændring.
Beskæftigelse inden for bygge- og anlægsvirksomhed	Stikprøven blev i 2004 optimeret med en Neuman-allokeringen efter antal beskæftigede. Stikprøven er stratificeret efter 8 størrelsesgrupper og 9 brancher. Virksomheder med under 5 beskæftigede kan ikke udvælges. Optimeringen bliver opdateret ved nye tællinger for at bibeholde den opnåede gevinst.	Optimeringen af udvælgelsen i 2004 reducerede stikprøven fra 3.000 til 2.500. Det har siden været muligt at fastholde den reducerede stikprøve.	Ingen ændring.
Indkomststatistik (EU-SILC)	Opregningen gennemført i 2005. Der er individuelle vægte til hver husholdning (adresse), så populationstotaler for mange forskellige baggrundsvARIABLE (køn, alder, indkomst, socio, ...) genskabes vha. stikprøven.	Opregningen har betydet en væsentlig reduktion i stikprøveusikkerheden, fordi målvariablene er meget korreleret med registervariablene.	Ingen ændring.
It-udgifter	Optimering/opregningsprincipper er fastlagt i 2005. Der er stratificeret efter 5 beskæftigelsesgrupper og 18 brancher. Stikprøven er fordelt efter en Neuman allokering. IT udgifter varierer mere fra virksomhed til virksomhed end mange andre økonomiske variable. Virksomheder med under 10 ansatte udvælges ikke. Der opregnes ikke for denne skævhed.	Kvalitetsgevinsten fra ændret fordeling af stikprøven er anvendt i nødlidende brancher. Opregningen giver meget lille reduktion i stikprøveusikkerheden.	Ingen ændring
Danske virksomheders brug af it	Stikprøven er stratificeret efter branche og størrelse. Stikprøven er allokeret i forhold til tællingens forskellige formål, hvor der afhængig af formålet opdeles på forskellige brancher. Der opregnes med et regressionsestimat efter omsætning og antal beskæftigede. Stikprøven kan ikke opfylde EU-krav. Virksomheder med under 10 ansatte kan ikke udvælges, og der opregnes ikke herfor.	Stikprøven er stratificeret mht. tællingens forskellige formål og hjælpeinformation bruges til at reducere stikprøveusikkerheden.	Burde øges.



	Beskrivelse af optimeringsprincipperne	Effekt på stikprøve, formål med optimerin- gen og effekt på den samlede usikkerhed	Overvejelser om ændret stikprøvestør- relse i forhold til usikkerheds- niveau
Ferie- og forretningsrejser	Simpel tilfældig stikprøve. Opregningsprincipper er fastlagt i 2005. Estimatene har betydelig samlet usikkerhed og Nationalregnskabet ønsker mere sikre estimater end der kan opnås med den nuværende stikprøve. Undersøgelsen bidrager til opgørelsen af udenrigshandelen med tjenester.	Effekten af ændringen af opregning har ikke givet mindre stikprøveusikkerhed, men der har været en væsentlig opretning af biasproblemer som følge af forskerbeskyttelse og skævt bortfald.	Ingen ændring.
Befolkningens brug af internet	Der er fokus på optimeringen af opregningen, hvor der efterstratificeres efter køn, alder, civilstand og indkomst. Herved korrigeres der også for skævt bortfald.	Antal opnåede interview på 3.000 opfylder lige EU-krav.	Ingen ændring.
Forbrugs- undersøgelsen	Opregningsprincipper indført i 2005/2006. Der beregnes individuelle vægte til hver adresse, så populationstotaler genskabes for køn, alders, indkomst og sociogrupper vha. stikprøven. Der er ikke paneler i denne stikprøve.	Effekten af opregningen er mindre bias som følge af forskerbeskyttelse og andet bortfald. Bortfaldet er ca. 50 %. Tællingen har en betydelig samlet usikkerhed, samt bias.	Burde øges.
Detail- omsætningsindeks	Optimering/oprekningsprincipper er fastlagt i 2004 optimeringen revurderet i 2006. Stikprøven er stratificeret efter 63 forskellige brancher og 4 størrelsesgrupper (efter momsomsætningen) i alt 252 strata plus 2 strata til særlige firmaer. Stikprøven er optimeret efter Neuman allokering. Der er 3 årlige paneler.	Effekten af optimeringen er brugt i nødlidende brancher. Meget detaljeret statistik. Stikprøven er ikke øget af ressourcemæssige årsager.	Burde øges, men fastholdes af byrdemæssige årsager.

Tabel 2.

## Tællinger med EU-krav til dækningsomfang

	Baggrund/lovgrundlag for udvælgelse af virksomheder, krav til statistik	Overvejelser om ændret antal indberettende virksomheder
Intrastat, import	EU-krav om dækning af 97 pct. af udenrigshandel mellem EU-lande.	Justeres hvert år.
Intrastat, eksport	EU-krav om dækning af 97 pct. af udenrigshandel mellem EU-lande.	Justeres hvert år.
Industriens salg af varer	EU-krav om dækning af 90 pct. af den samlede omsætning på 4-cifret NACE-brancheniveau og skal omfatte alle virksomheder med mindst 20 ansatte.	Den danske opgørelse lever samlet op til kravene til dækningsgrad, men opfylder ikke kravene på 4-cifret brancheniveau. Hensyn til administrative byrder indebærer, at der ikke ønskes inddraget virksomheder med under 10 ansatte.
Fragt- og krydstogtskibes anløb på danske havne	EU-krav om, at alle danske trafikhavne med en årlig godsomsætning på mindst 1 mio. ton indgår.	Ingen ændring.

Tabel 3.

**Tællinger, hvor Danmarks Statistik har valgt at indsamle oplysninger udelukkende fra de store og mellemstore virksomheder**

	<b>Baggrund/lovgrundlag for udvælgelse af virksomheder, krav til statistik</b>	<b>Overvejelser om ændret antal indberettende virksomheder</b>
Lønstatistik, årlig	Alle virksomheder med mindst 10 fuldtidsansatte. Statistikken er koordineret med DA's lønstatistik.	Ingen ændring.
Industriens køb af varer og tjenester	Detaljeringsgraden til de indsamlede råvare- og tjenestekøb er fastsat af hensyn til Nationalregnskabet. Stikprøven består af virksomheder med mindst 50 beskæftigede.	Ingen ændring.
Overnatning på campingpladser	Alle virksomheder med mindst 75 teltpladser deltager. Statistikken bidrager til opgørelsen af betalingsbalancens rejsepost.	Ingen ændring.
Overnatninger på hoteller	Alle hoteller med over 40 sengepladser deltager. Statistikken bidrager til opgørelsen af betalingsbalancens rejsepost.	Ingen ændring.
Produktstatistik for it-brancher	EU stiller i en kommende forordning krav om deltagelse for virksomheder med mindst 20 ansatte. Indsamlingen bruges som grundlag for Nationalregnskabet.	Ingen ændring.

Tabel 4.

## Øvrige obligatoriske tællinger

	Beskrivelse af optimeringsprincipperne	Effekt på stikprøve, formål med optimeringen og effekt på den samlede usikkerhed	Overvejelser om stikprøvestørrelse i forhold til usikkerhedsniveau
National transport med danske lastbiler	Stikprøven er optimalt allokeret i 2004 med en Neuman-allokering efter transportarbejde.	Der er et EU-krav om en variationskoefficient for den nationale transport med lastbiler på 2,5 % for ton, ton km og km. Stikprøven for national transport opfylder ikke dette krav.	Kan optimeres yderligere.
International transport med danske lastbiler	Stikprøven allokeres så byrden spredes på små og store firmaer på en hensigtsmæssig måde.	Der er et EU-krav om en variationskoefficient for den samlede transport (nationalt og internationalt) på 2,5 % mht. ton, ton km og km.	Kan optimeres yderligere.
Prisindeks i producent- og importleddet	Indeksundersøgelse (hvor der ikke kan beregnes stikprøveusikkerhed) der bygger på lange tidsserier. Stikprøven sammensættes, så indeksene offentliggøres med den sikkerhed som EU-forordningen stiller; men stor samlet usikkerhed på det detaljerede niveau, som der offentliggøres nationalt.		Ikke repræsentativ stikprøve. Burde øges, men fastholdes af byrdemæssige årsager.
Forbrugerprisindeks	Indeksundersøgelse (hvor der ikke kan beregnes stikprøveusikkerhed) som øvrige prisindeks, der bygger på lange tidsserier, hvor datagrundlaget løbende suppleres, hvis usikkerheden stiger eller designet ikke opfylder kravene til dækningsgrad. En række EU-forordninger angiver en lang række varegrupper, som skal dækkes med prisindsamlingen.		Ikke repræsentativ stikprøve. Ingen ændring.
Regnskabsstatistik	Stikprøven udgør 4,6 % af populationen, hvor gennemsnittet for andre mellemstore EU-lande er 8,4 %. Der er en panelstruktur. Opregning foretages med gennemsnits masse-imputering, hvilket gør at stikprøveusikkerheden kun kan simuleres med Jackknife-metoden. Som udgangspunkt allokeres stikprøven på baggrund af beskæftigelsen, hvilket på en robust måde medfører at tællingen kan opfylde de forskellige formål.		Kan optimeres.

Tabel 5.

## Øvrige frivillige tællinger

	Beskrivelse af optimeringsprincipperne	Kontrakt med EU	Overvejelser om stikprøvestørrelse i forhold til usikkerhedsniveau
Konjunkturbarometer for bygge og anlæg	Konjunkturbarometeret måler forventninger med en kategoriell variabel med 4 eller 5 udfald.	Nyt udbud i sept. 2007 for perioden 2008-12.	
Konjunkturbarometer for serviceerhverv	Konjunkturbarometeret måler forventninger med en kategoriell variabel med 4 eller 5 udfald.	Ingen kontrakt med EU.	En optimering af stikprøven er undervejs
Konjunkturbarometer for industri	EU-finansiering. Baseres på de 500 største virksomheder, der ønsker at deltage.	Nyt udbud i sept. 2007 for perioden 2008-12.	
Forbrugerforventninger	Forbrugerforventninger måles med en kategoriell variabel med 4 eller 5 udfald.	Nyt udbud i sept. 2007 for perioden 2008-12.	



### Bilag 3: Analyse af Regnskabsstatistikken

<i>Formål med statistikken</i>	Regnskabsstatistikken er – udover EU kravene – grundlaget for flere andre stikprøveundersøgelser og indgår i Nationalregnskabet samt Virksomhedsdata-basen. For en række variabler er regnskabsstatistikken den eneste kilde, fx for investeringer, hvor den samlede standardfejl (CV) er 5 pct. Statistikken underopdeles på sekscifret DB03.
<i>Formålet med analysen</i>	Formålet med denne analyse er at vurdere den samlede effekt af en optimal allokering efter størrelse inden for hovedbrancherne. I alt otte hovedbrancher analyseres, da råvareudvikling er udeladt af analysen pga. den meget lille population. Udgangspunktet for analysen er at bibeholde den samme sikkerhed og opfylde formålene med statistikken.
<i>Analysens forudsætninger</i>	Regnskabsstatistikken er baseret på ca. 8.750 spørgeskemaindberetninger fra virksomhederne og en omfattende brug af registeroplysninger. Begge de to kilder bruges til imputering af værdierne for de ca. 180.000 virksomheder, der ikke indgår i stikprøven. Stikprøven udgør 4,6 pct. af populationen, hvor gennemsnittet for andre mellemstore EU-lande er 8,4 pct. Dette har været muligt, fordi statistikkontoret vha. imputeringsalgoritmerne udnytter al den registerinformation, der findes.
<i>Jack-knife simuleringer</i>	Pga. imputeringerne er det ikke muligt med standardmetoder at beregne den statistiske stikprøveusikkerhed. Der er derfor udviklet en simulering af usikkerheden baseret på den såkaldt Jack-knife teknik. Da beregningerne er meget komplekse, tager det op mod fem timer at afvikle programmet – når det er blevet gennemtestet og færdigudviklet til produktion.
<i>Indikatorer på effekten af en stikprøveændring</i>	Imputeringerne bygger på en fagstatistisk gennemgang af datagrundlaget, hvor små brancher og undergrupper slås sammen, så imputeringerne bliver robuste. Det betyder også, at det ikke er muligt sikkert at beregne virkningen af store ændringer af stikprøven, da imputeringsprogrammet udvikles særskilt til hver årstællings datagrundlag. Det blev derfor aftalt på mødet i Finansministeriet, at Danmarks Statistik skulle udarbejde <i>indikatorer</i> for effekten af en omlægning af stikprøven, der kunne danne baggrund for den videre diskussion.
<i>Stikprøveallokeringen i dag</i>	Stikprøven stratificeres efter fire beskæftigelsesgrupper, nemlig 5-10, 10-20, 20-50 og mindst 50 ansatte. Inden for hver branche fordeles stikprøven, så den i al væsentlighed er proportional efter antal ansatte og omsætning. Dette er også erfaringsmæssigt en meget robust måde at allokere en stikprøve på, men derfor behøver den ikke at være optimal. Specielt er der grund til at analysere, om der med denne metode udtrækkes nok små virksomheder, da de små virksomheder typisk varierer mere omkring deres niveau end større virksomheder. Eller sagt fagstatistisk: Har en større varianskoefficient (CV). Varianskoefficientens betydning fremgår, når man omskriver Neumannallokeringen, hvor $n$ er den samlede stikprøvestørrelse, $N_h$ er populationen og $S_h$ spredningen i det $h$ 'de stratum:

$$\begin{aligned}
n_h &= n \cdot \frac{N_h \cdot S_h}{\sum_{h=1}^H N_h \cdot S_h} \\
&= n \cdot \frac{N_h \cdot \bar{x}_h \cdot S_h / \bar{x}_h}{\sum_{h=1}^H N_h \cdot S_h} \\
&= n \cdot \frac{\text{"Omsætning"} \cdot CV_h}{\sum_{h=1}^H N_h \cdot S_h}
\end{aligned}$$

**Tabel 1. Oversigt over Regnskabsstatistikens stikprøve**

	5-10 ansatte			10-20 ansatte			20-50 ansatte			Samlet		
	Population	Omsætning (mio. kr.)	Stikprøve 2004	Population	Omsætning (mio. kr.)	Stikprøve 2004	Population	Omsætning (mio. kr.)	Stikprøve 2004	Population	Omsætning (mio. kr.)	Stikprøve 2004
Industri	2 233	15 941	184	1 738	26 723	330	1 432	54 564	711	5 403	97 228	1 225
Bygge og anlæg	3 087	19 205	195	1 797	22 062	264	924	27 505	446	5 808	68 771	905
Autohandel, service og tankstationer	875	14 610	74	463	17 399	98	278	37 652	136	1 616	69 661	308
Engroshandel	1 940	59 210	176	1 313	59 292	250	890	104 728	429	4 143	223 230	855
Detailhandel	2 165	22 831	177	816	17 315	154	431	26 505	168	3 412	66 651	499
Hoteller og restauranter	837	4 191	70	423	3 930	77	203	4 293	96	1 463	12 414	243
Transport, post og tele	1 115	12 402	77	706	15 905	110	440	26 649	216	2 261	54 957	403
Fast ejendom og forretningservice	2 577	20 880	225	1 427	21 275	299	882	29 296	409	4 886	71 452	933
I alt	14 829	169 270	1 178	8 683	183 901	1 582	5 480	311 192	2 611	28 992	664 363	5 371

I tabel 1 er stikprøven, dvs. virksomheder med under 50 ansatte, fordelt efter beskæftigelsen i hver hovedbranche. Det fremgår af tabel 2 nedenfor, at stikprøven ikke ligger langt fra denne proportionale fordeling efter omsætning. Den manglende overensstemmelse skyldes bl.a. udviklingen, siden stikprøven blev udvalgt, og hensyn til panelforpligtigelser. Virksomhederne får således lovning på, at de kun vil blive udvalgt hver tiende år, hvis de har 5-10 ansatte, to år i træk og derefter fri i otte år, hvis de har 10-20 ansatte, og inde i tre år og ude i tre år, hvis de har 20-50 ansatte. Hvis virksomheden bliver større (eller mindre), har den stadig ret til at blive friholdt i den periode, der er tilbage, med mindre den får over 50 ansatte, hvor man deltager hvert år. Det medfører, at stikprøven kun delvist ændres, når fordelingstallene opdateres, eller stikprøvedesignet ændres.



Tabel 2.

## Proportional fordeling af stikprøven efter omsætning

	5-10 ansatte			10-20 ansatte			20-50 ansatte			Samlet		
	Omsætning (mio. kr.)	Stikprøve 2004	Proportional stikprøve	Omsætning (mio. kr.)	Stikprøve 2004	Proportional stikprøve	Omsætning (mio. kr.)	Stikprøve 2004	Proportional stikprøve	Omsætning (mio. kr.)	Stikprøve 2004	Proportional stikprøve
Industri	15 941	184	201	26 723	330	337	54 564	711	687	97 228	1 225	1 225
Bygge og anlæg	19 205	195	253	22 062	264	290	27 505	446	362	68 771	905	905
Autohandel, service og tankstationer	14 610	74	65	17 399	98	77	37 652	136	166	69 661	308	308
Engroshandel	59 210	176	227	59 292	250	227	104 728	429	401	223 230	855	855
Detailhandel	22 831	177	171	17 315	154	130	26 505	168	198	66 651	499	499
Hoteller og restauranter	4 191	70	82	3 930	77	77	4 293	96	84	12 414	243	243
Transport, post og tele	12 402	77	91	15 905	110	117	26 649	216	195	54 957	403	403
Fast ejendom og forretningsservice	20 880	225	273	21 275	299	278	29 296	409	383	71 452	933	934
I alt	169 270	1 178	1 363	183 901	1 582	1 533	311 192	2 611	2 476	664 363	5 371	5 372

*Hensyn til effektiv stikprøve, administration og byrder på samme tid*

Undersøgelingsdesignet tager på samme tid hensyn til virksomhedernes byrder, søger at begrænse ressourcerne til administrationen af dataindsamlingen og opnå en effektiv stikprøve. Når man vælger hver tiende af virksomheder med 5-10 ansatte og dobbelt så ofte virksomhederne med 10-20 ansatte er det samtidig proportionalt med antal ansatte, da midtpunktet i det første interval er 7,5 og 15 i det andet interval. Det tilsvarende gælder næsten også mht. virksomheder med 10-20 ansatte og 20-50 ansatte idet midtpunktet her er hhv. 15 og 35 og man 2½ gang så ofte vælger virksomheder med 20-50 ansatte. Samtidig loves virksomheder, at de fx kun er med i to år og ud i otte år, hvis de har 10-20 ansatte. Systemet er samtidigt nogenlunde enkelt at administrere, fordi princippet er ens på tværs af brancher og styres af virksomhedens størrelse. Det interne ressourceforbrug til administration af dataindsamling vil påvirkes, hvis allokering gøres kompleks og afhænger af både branche og størrelse og ændres fra år til år.

*Analysen*

Jack-knife programmet er udviklet til at beskrive usikkerheden for 17 variabler, men kun en af dem findes ikke i registeret fra Skat, nemlig investeringer. Af de ca. 180.000 virksomheder, der ikke udvælges, er der oplysninger fra Skat for ca. 85.000. Udover investeringer er der ca. 75 andre variabler i spørgeskemaet, der ikke findes i registeret fra Skat. Da det er disse variabler, som det er regnskabsstatistikens formål at belyse, tager analysen udgangspunkt i spørgsmålet om *investeringer*.

*Analysestrategien*

Først var det planen at undersøge forskellige alternative stikprøvedesign ved i stikprøven på ca. 8.750 at trække fx 50 ekstra i hver branche blandt virksomheder med 5-10 ansatte som dubletter (ekstra stikprøve) og reducere med fx 100 i hver branche for virksomheder med 20-50 ansatte. For at få stabile resultater måtte hvert stikprøvedesign trækkes et antal gange, da der er tale om simuleringens simulering, fx 10 gange. Udover, at det ville tage ca. 50 timer at teste hvert stikprøvedesign, viste det sig, at det ikke ville give en rimelig sikker indikator. Derfor måtte denne tilgang opgives. Metodens styrke var, at den indarbejdede betydningen af imputeringerne, men datagrundlagets opdeling i undergrupper kunne ikke bære denne analytiske tilgang.

I stedet måtte analysen baseres på simple estimater af spredningen i stikprøven, der byggede på Jack-knife imputeringer med den anvendte stikprøvestørrelse og derfor ikke nødvendigvis kan generaliseres til andre stikprøvestørrelser.

**Tabel 3. Teoretisk optimal allokering efter investeringer**

	5-10 ansatte			10-20 ansatte			20-50 ansatte			Samlet		
	Stikprøve 2004	Optimal	Ændring	Stikprøve 2004	Optimal	Ændring	Stikprøve 2004	Optimal	Ændring	Stikprøve 2004	Optimal	Ændring
Industri	184	212	28	330	199	-131	711	755	44	1225	1166	-59
Bygge og anlæg	195	291	96	264	222	-42	446	181	-265	905	694	-211
Autohandel, service og tankstationer	74	71	-3	98	95	-3	136	142	6	308	308	0
Engroshandel	176	224	48	250	279	29	429	291	-138	855	794	-61
Detailhandel	177	188	11	154	147	-7	168	162	-6	499	497	-2
Hoteller og restauranter	70	25	-45	77	33	-44	96	123	27	243	181	-62
Transport, post og tele	77	84	7	110	89	-21	216	225	9	403	398	-5
Fast ejendom og forretningsservice	225	210	-15	299	176	-123	409	482	73	933	868	-65
I alt	1178	1305	127	1582	1240	-342	2611	2361	-250	5371	4906	-465

*Usikre resultater* Når man kigger på tabel 3, ses det hurtigt, at der for nogle brancher er tale om statistisk uforklarlige resultater. For tre brancher foreslås der for virksomheder med 20-50 ansatte en markant reduktion af stikprøven. For de fem andre brancher er stikprøven derimod uforandret eller svagt større. Det sidste giver generelt mest mening, fordi den proportionale allokering i tabel 2 generelt ligger over den faktiske, hvilket burde indikere, at den anvendte stikprøvestørrelse i dette størrelsesstrata ligger tæt på det optimale. Omvendt ses det for virksomheder med 5-10 ansatte, at der er hovedbrancher, hvor der foreslås en mindre stikprøve, hvilket er i uoverensstemmelse med erfaringen med tilsvarende optimeringer i andre erhvervsstikprøver. Samlet giver analysen et uklart og næppe fuldt retvisende billede over de mulige optimeringsgevinster.

*Anvendelsen af Neumann allokeringen* Neumann-allokeringen er et teoretisk forslag til en optimering i den forstand, at analysen bygger på, at man kendte den efterfølgende varians, før man trak stikprøven i 2004. Derfor vil en Neumann-allokering altid være mindst lige så god som det, man gjorde. Men der er ingen sikkerhed for, at Neumann-allokeringen baseret på 2004 viser sig at være optimal i 2005 – rent faktisk er der garanti for, at den ikke er det. Så hvis man år efter år antog, at man ville opnå den gevinst Neumann allokeringen antyder, ville man reelt hvert år få en dårligere og dårligere statistik. I den forstand er Neumann allokering for optimistisk.

Hvis man har foretaget en Neumann allokering kan man som udgangspunkt kun forvente, at en ny Neumann allokering sikrer, at den allerede opnåede gevinst kan bibeholdes. Kun hvis variationen i datamaterialet reduceres eller der er nye registerinformationer at optimere efter kan man forvente yderligere gevinster.

For at man kan forvente at opnå en del af gevinsten ved Neumann allokeringen skal man være sikker på, at variansestimaterne er sikret estimeret og stabile over tid. Det sidste er altid svært at vurdere. Man skal derudover altid regne med, at 2-5 procent af stikprøvestørrelsen skal afsættes til at tage højde for, at Neumann allokeringen er for optimistisk. I den konkrete situation må man derfor foretage en statistisk vurdering af, hvor sikre og stabile de variansestimater er, som Neumann allokeringen bygger på.

*Mulige faglige konklusioner* Hvis man tager udgangspunkt i analysen i tabel 3, er der en række umiddelbare konklusioner.

*Ikke færre små* For det første vil det være skadeligt at reducere stikprøven i størrelsesgruppen 5-10 ansatte. Primært, fordi denne størrelsesgruppe også bruges til at imputere værdierne for alle de virksomheder, der ligger under cut-off grænsen på fem ansatte. Dette tages der ikke højde for, når stikprøven allokeres, hvilket ville betyde en dramatisk forøgelse af stikprøvestørrelsen. Virksomheder med under fem ansatte repræsenterer 10-20 pct. af den samlede omsætning. Derudover vil det være meget overraskende, hvis CV-erne næste gang skulle vise samme billede for de brancher, hvor stikprøven foreslås reduceret, da man erfaringsmæssigt for små virksomheder skal vælge mindst den andel, som deres omsætning repræsenterer.

*Usikkert resultat for de største virksomheder* For det andet er resultaterne i størrelsesgruppen 20-50 ansatte meget ustabile, hvilket indikerer, at variansskønnet af imputeringerne ikke er fuldt retvisende. Man kan så vælge at se bort fra denne størrelsesgruppe og fastholde det nuværende stikprøvedesign for disse virksomheder. For denne faglige vurdering taler også, at forskere og brugere af virksomhedsdatabasen ofte er mere interesseret i de lidt større virksomheder med mindst 20 ansatte, når de analyser innovation, patenter, organisationsændringer eller lignende emner.

*Korrektion for, at Neumann allokeringen er for optimistisk* For det tredje skal man tage højde for, at Neumann-allokeringen er for optimistisk. Enten ved at afsætte et par procent af stikprøvestørrelsen hertil, eller fx halvere de foreslåede ændringer.

*Fagstatistisk konklusion* Om man foretrækker at se helt bort fra størrelsesgruppen 20-50 ansatte, eller man fx vil halvere de foreslåede ændringer pga. de usikre variansestimater, er en faglig vurdering. Den sikre løsning er i første omgang at se bort fra størrelsesgruppen 20-50 ansatte. Det giver (se tabel 4) en samlet reduktion på 120 virksomheder, fordelt med 173 flere af de små virksomheder med 5-10 ansatte og 293 færre af de mellemstore virksomheder med 10-20 ansatte.

Alternativt kan man vælge at halvere ændringerne og samtidig undlade at reducere stikprøven for virksomheder med 5-10 ansatte, jf. cut-off problemet ovenfor. Det giver en samlet gevinst på ca. 200 virksomheder, hvis man sammenholder med tabel 3. Man kan også vælge kun at lade store ændringer slå igennem med halv kraft, bibeholde de mindre ændringer og ligesom ovenfor ikke sætte stikprøven ned for virksomheder med 5-10 ansatte pga. cut-off opregningen. Af tabel 5 fremgår, at det giver 163 færre virksomheder, fordelt med 142 flere små virksomheder med 5-10 ansatte, 213 færre virksomheder med 10-20 ansatte og 92 færre virksomheder med 20-50 ansatte.

Tabel 4.

**Teoretisk optimal allokering efter investeringer for virksomheder med 5-20 ansatte**

	5-10 ansatte			10-20 ansatte			Samlet		
	Stikprøve 2004	Optimal	Ændring	Stikprøve 2004	Optimal	Ændring	Stikprøve 2004	Optimal	Ændring
Industri	184	239	55	330	225	-105	514	464	-50
Bygge og anlæg	195	240	45	264	184	-80	459	424	-35
Autohandel, service og tankstationer	74	74	0	98	98	0	172	172	0
Engroshandel	176	189	13	250	235	-15	426	424	-2
Detailhandel	177	185	8	154	145	-9	331	330	-1
Hoteller og restauranter	70	64	-6	77	82	5	147	146	-1
Transport, post og tele	77	89	12	110	94	-16	187	183	-4
Fast ejendom og forretningsservice	225	271	46	299	226	-73	524	497	-27
I alt	1178	1351	173	1582	1289	-293	2760	2640	-120

Tabel 5.

**Allokering på baggrund af den teoretiske analyse, halvering af de største ændringer og ingen reduktion af virksomheder med 5-10 ansatte**

	5-10 ansatte			10-20 ansatte			20-50 ansatte			Samlet		
	Stikprøve 2004	Ny stikprøve	Ændring	Stikprøve 2004	Ny stikprøve	Ændring	Stikprøve 2004	Ny stikprøve	Ændring	Stikprøve 2004	Optimal	Ændring
Industri	184	212	28	330	266	-64	711	711	0	1 225	1 189	-36
Bygge og anlæg	195	243	48	264	222	-42	446	334	-132	905	799	-126
Autohandel, service og tankstationer	74	74	0	98	95	-3	136	136	6	308	305	3
Engroshandel	176	224	48	250	279	29	429	360	-69	855	863	8
Detailhandel	177	188	11	154	147	-7	168	162	-6	499	497	-2
Hoteller og restauranter	70	70	0	77	33	-44	96	96	27	243	199	-17
Transport, post og tele	77	84	7	110	89	-21	216	216	9	403	389	-5
Fast ejendom og forretningsservice	225	225	0	299	238	-61	409	409	73	933	872	12
I alt	1 178	1 320	142	1 582	1 369	-213	2 611	2 424	-92	5 371	5 113	-163

*Sammenfatning* Analysen af indikatorer for effekten af en omlægning af stikprøvedesignet giver en gevinst på 2 pct. til 3 pct. af den samlede stikprøvestørrelse. Den største gevinst kan kun opnås, hvis man antager, at de usikre variansestimater for virksomheder med 20-50 ansatte delvist holder. Forudsætningen er, at stikprøven blandt virksomheder på 5-10 ansatte øges med knap 150. Denne ændring vil så kunne implementeres over de næste år, efterhånden som de lovede tilsagn om friholdelse udløber. En mere kompleks stikprøveplan med varierende udvalgssandsynligheder på tværs af brancher og størrelsesgrupper, der kan forandres hvert år, vil udover nyudvikling af administrationssystemet, også betyde øgede ressourcer til administration af den årlige dataindsamling og stikprøveudvælgelse.



## Bilag 4: Analyse af transportstatistikken

*Formålet med National og Internationalt vejgodstransport* Transportstatistiktællingen gennemføres hvert kvartal. Formålet med stikprøveundersøgelsen er at bestemme godstransporten med danske lastbiler i Danmark og i udlandet.

*Formålet med dette notat* Dette notat giver nogle estimerede usikkerheder og undersøger, hvilke optimeringsmuligheder der er i statistikken *National* og *International Godstransport*. Optimering af en stikprøve kan overvurdere eller undervurdere effekten af en omlægning.

I dette notat er estimeret varianser og variationskoefficienter (CV) på det foreliggende datamateriale. Disse bliver brugt som indikatorer for den fremtidige varians, men der kan først beregnes et estimat for variansen, når data er indsamlet efter det nye stikprøvedesign. Variansen vil ændre sig hvert kvartal – hver gang der indsamles nyt datamateriale, på samme måde som transportarbejdet ændrer sig hvert kvartal.

*EU-krav* Der er EU-forordning forbundet med *National* og *Samlet Godstransport*. Statistikken har flere forskellige målvariable, men der er kun tre, der er bundet af krav om maksimal stikprøveusikkerhed på  $\pm 5$  pct. på et 95 pct. konfidensinterval, svarende til en variationskoefficient på 2,5 pct. Denne analyse vil primært omhandle disse variable, men en ændring af stikprøvedesignet vil også have betydning for de andre variable. Det antages at være afspejlet i nedenstående målvariable.

*Tabel 1.* **EU-krav til statistikken på årsbasis**

	National transport	Samlet transport (Nationalt og internationalt)
	Variationskoefficient (CV)	Variationskoefficient (CV)
Kørte km med læs i alt	2,5 pct.	2,5 pct.
Pålæsset godsmængde i alt	2,5 pct.	2,5 pct.
Transportarbejde i alt (tonkm) <sup>1</sup>	2,5 pct.	2,5 pct.

<sup>1</sup> 1 tonkm svarer til 1 km kørt med 1 ton læs.

Når der er et krav om CV på 2,5 pct. til samlet transport, et krav om CV på 2,5 pct. til *National Godstransport*, og *National godstransport* udgør 2/3, og *International Godstransport* udgør 1/3 af den samlede transport mål, CV til *International Godstransport* være 5,0 pct. (se appendiks).

### Nuværende metode

*Nuværende metode for stikprøve* Transportstatistikken er organiseret i to dele, så der bliver indsamlet data for transport i Danmark med danske lastbiler i et *set-up* (Nationalt), og et helt andet *set-up* for transport med danske lastbiler i udlandet (Internationalt).

*Register (hjælpeinformation)* På transportområdet er der et lastbilregister til rådighed. Der bruges følgende variable:

- Registreringsnummer (identifikation af lastbilen)
- Registreringsdato (giver alder på lastbilen)

- Lastbil/sættevognstrækker
- Egenvægt
- Totalvægt (lasteevne = totalvægt – egenvægt)
- Aksler
- Trækkrog/ikke trækkrog
- Kørselsart (vognmandskørsel/firmakørsel)

I undersøgelsen for *National Godstransport* er lastbilen styrende enhed. Det er således lastbiler (registreringsnumre), der udvælges. Stikprøven er stratificeret efter størrelse, lastbiltype og firmatype (om der udføres firmakørsel eller vognmandskørsel).

## National Godstransport

*Stikprøve af biler –  
National godstransport*

Der udvælges 884 biler hvert kvartal ud af ca. 45 000 lastbiler i registeret (populationen). Undersøgelsen gennemføres for en stikprøve af lastbiler med en maksimal tilladt totalvægt over 6 ton i stikprøven, valgt ud fra et register (som bygger på dansk motorregister (CRM)). Der er således et *cut-off* i stikprøven, som der ikke estimeres for i statistikken. Det er en postalundersøgelse, hvor der telefonrykkes.

*Indberetningen – National  
godstransport*

Respondenterne skal indberette kørsel, mængde og typen af last, der er kørt med i en bestemt uge i kvartalet. Ugen, der skal indberettes for, er bestemt af DST, og er fordelt jævnt over kvartalet for alle de biler, som skal indberettes. Der antages at være præcis 13 uger i et kvartal, og et år består af 4 gange 13 uger (364 dage).

*Bilen deltager kun et  
kvartal på et år – National  
godstransport*

Hvis bilen er udvalgt et kvartal, kan den ikke deltage i andre kvartaler i løbet af året. Estimerne for de fire kvartaler er uafhængige. Det gør det nemt at beregne variansen for årsestimerne, da estimerne og varianserne blot lægges sammen for de fire kvartaler. Derimod giver det biasproblemer<sup>2</sup> at friholde nogle biler, når de er udvalgt i et kvartal. I lyset af den forholdsvis lille stikprøve sammenlignet med hele populationen må problemerne antages at være meget begrænsede, da det må være sjældent, at en bil ville være udvalgt flere gange på et år.

Stikprøven udgør under to procent af populationen, hvilket faktisk er en ret lille stikprøve sammenlignet med andre erhvervsundersøgelser.

---

<sup>2</sup> Biasproblemer er, at estimatet ikke er middeltet estimeret.



Tabel 2. **Sammensætningen af stikprøven for *National Godstransport* (gennemsnit af 1.-4. kvartal)**

<i>National Godstransport</i>		Vognmandskørsel		Firmakørsel	I alt
		Totalvægt	Svar (Population)	Svar (Population)	Svar (Population)
Lastbil			103 (7 117)	107 (11 427)	
Lastbil med trækkrog		0 – 18 000	18 (1 688)	12 (1 003)	
		18 001 – 24 000	91 (2 663)	25 (1 025)	
		24 001 +	96 (4 401)	22 (1 146)	
Sættevognstrækker		0 – 18 000	147 (4 687)	17 (851)	
		18 001 – 24 000	46 (3 035)	6 (362)	
		24 001 +	89 (4 752)	4 (309)	
I alt					783 (44 466)

*Ikke korrekt håndtering af populationen*

Der er et metodemæssigt problem i håndteringen af populationen. Problemet er versioner af lastbilregistret. Stikprøven af lastbiler bliver valgt i tidligere version af registret end det, der estimeres ud fra. Det er muligt at have en udvalgspopulation og en estimeringspopulation, men der skal tages højde for dette i estimeringen, og det bliver der ikke. Det betyder, at lastbilerne indgår med en anden vægt i estimeringen, end de skal. Problemet er af mindre betydning, da lastbiler næppe skifter størrelse eller type. Herudover er det normalt at ændre registreringsnummer, hvis en lastbil skifter ejer. På den måde kan overgang fra vognmandskørsel til firmakørsel opfanges. Derfor er det sjældent, at en lastbil skifter stratum fra en version af lastregistret til næste version af registret. Derfor er problemet mindre i denne statistik. I andre undersøgelser med andre registre vil denne praksis føre til alvorligere undervurderinger af fremgang og overvurdering af tilbagegang.

## International Godstransport

*International Godstransport*

I undersøgelsen for *International Godstransport* er firmaet med internationale aktiviteter styrende enhed. Firmaer med internationale aktiviteter er overvejende fundet ved, at firmaerne i *National Godstransport* på et tidspunkt har angivet, at de har kørt internationalt. Desuden udnyttes informationer fra ITD's medlemsregistre<sup>3</sup>, fagtidsskrifter mv. Denne population udgør så grundlaget for stikprøven. Stikprøven er stratificeret efter firmaets størrelse, målt på antallet af biler. Firmaer med 10+ biler udvælges totalt, firmaer med 3-9 biler gennemsnitligt hvert andet kvartal og små virksomheder hvert fjerde kvartal. Firmaer, hvor der ikke har kunnet findes biler i bilregistret, udvælges totalt.

<sup>3</sup> ITD står for International Transport Danmark, det er en brancheorganisation.

Tabel 3. **Sammensætningen af stikprøven for *International Godstransport* (gennemsnit af 1.-4. kvartal)**

<i>International Godstransport</i>	Stratum	Antal svar	Antal firmaer i alt
	"0" biler	68	61
1- 2 biler	540	131	
3- 9 biler	242	118	
10 + biler	259	252	
I alt	1 108	562	

De 1 108 firmaer i populationen har ca. 7 500 biler og de 562 firmaer som har, indberettet har rapporteret for ca. 500 biler.

*Indberetningen – International godstransport*

Firmaerne indberetter kørsel, mængde og typen af last, der er kørt med i en bestemt uge eller halvuge i kvartalet, for hver af deres biler. Perioden, der skal indberettes for, er bestemt af DST og er fordelt jævnt over kvartalet for alle de firmaer, som skal indberette. Små firmaer får tildelt en fuld uge. De øvrige en halv uge – mandag til torsdag eller fredag til søndag. Firmaerne skal alene indberette kørsler for de biler, der krydser landegrænsen i den tildelte periode. Indberetningen skal dække kørslen tilbage til Danmark, selv om bilen først returnerer efter periodens udløb.

### National og International Godstransport

Transportarbejde, kørte km med læs og pålæsset mængde beregnes nu, vha. det direkte estimat i en stratificeret stikprøve (se appendiks for yderligere forklaring).

Ved den nuværende metode er der estimeret følgende punktestimater og CV. Der er angivet hovedresultater for udvalgte målvariable.

Tabel 4. **Usikkerheden på estimerterne 2006, med nuværende metode (1.-4. kvartal sammenlagt)<sup>1</sup>**

	National transport		International transport		Samlet transport	
	Punkt-estimat	Estimeret CV	Punkt-estimat	Estimeret CV	Punkt-estimat	Estimeret CV
Kørte km med læs i alt (mio. km)	1 208	2,5 pct.	618	ca. 4,9 pct.	1 826	ca. 2,4 pct.
Pålæsset godsmængde i alt (1000 ton)	177 632	3,7 pct.	15 368	ca. 5,4 pct.	193 000	ca. 3,5 pct.
Transportarbejde i alt (mio. tonkm)	11 494	3,3 pct.	9 759	ca. 6,1 pct.	21 253	ca. 3,3 pct.
Kørte km i alt (mio. km)	1 498	2,4 pct.	713	4,6 pct.	2 211	2,2 pct.

<sup>1</sup> Tal med *kursiver* pålagt EU-krav – CV max 2,5 pct.

I *National Godstransport* skal der også angives, hvis bilen har været i udlandet. Det er fra *National Godstransport*, at 713 mio. km stammer fra. 'Kørte km i alt' international, bliver også indberettet i *International Godstransport*. Dette tal er bindeleddet mellem *International* og *National Godstransport*.

*Opjusteringsfaktor* Tallet fra den *Nationale Godstransport* på 713 mio. km er 5-10 gange større end det indberettede tal fra den *Internationale Godstransport*. Da man tager udgangspunkt i *National Godstransport* sker der en opjustering af tallene fra *International Godstransport*. Det er denne opjustering, som ikke kan indsættes i gænge metoder, men forskellen på de to tal afspejler en bias. Denne ekstremt store bias må skyldes underrapportering, forkert register til *International Godstransport* eller fejl ved *National Godstransport*. Når denne metode er valgt til estimering af *international Godstransport*, bliver tallet for de internationalt 'kørte km i alt' kun bestemt af *National godstransport* med en CV på 4,6 pct..

Det er dette tal (713 mio. km i 2006) med tilhørende usikkerhed, der er styrende for hele *International Godstransport*, idet faktor på 5-10 bruges til at justere tallet for 'km med læs', 'pålæsset godsmængde', 'transportarbejde' og alle de andre målvariable om aktiviteter i udlandet.

*Biaskorrektion?* I den nuværende metode bliver det altså antaget, at indberetningen fra *National Godstransport* er den rigtige. Men sandheden kunne meget vel være et sted mellem tallet fra *International* og *Nationale Godstransport*. Det er meget nemt i *National Godstransport* at skrive, at bilen kun har kørt udlandsture, ligesom det er nemt at skrive i *International Godstransport*, at bilen kun har kørt indenlands, for så slipper man for at udfylde mere præcist, hvor langt, med hvad og hvorhen bilen har kørt.

*Usikkerheden er tilnærmet for International Godstransport* Usikkerheden på en stikprøve, som er styret af en stikprøve fra et helt andet register, kan ikke beregnes med gængse metoder. Det er kun en indikator for CV (tabel 4), kvaliteten af denne kendes ikke og er derfor angivet med cirka-tal. CV er beregnet på følgende måde:

$$CV_{\text{variabel, internat., hel population}} = \frac{CV_{\text{kørte km i alt, nat., hel population}}}{CV_{\text{kørte km i alt, internat., afgrænset populaion}}} \cdot CV_{\text{variabel, internat., afgrænset population}}$$

*EU-krav langt fra opfyldt med nuværende metode* Af tabel 4 fremgår det, at EU-kravene langt fra er opfyldt på 'Pålæsset godsmængde' og 'Transportarbejde', men EU-kravene er opfyldt for 'Kørte km med læs'.

## Bedre udnyttelse af hjælpeinformation i transportstatistikken

*Afdækning af problemer* I den indledende fase af en analyse af en statistik prøver DST's metodeenhed at afdække, hvilke muligheder der er for variansreduktion, som kan berettige en reduktion af stikprøven. Findes der registerinformationer, som ikke bruges i dag, kan der anvendes paneler eller er der andre forbedringsmuligheder? Der forsøges også afdækket andre problemer i statistikken, som kan bidrage til biasproblemer. Eksempler på dette er angivet i afsnittet Nuværende Metode, ovenfor.

*Register og hjælpevariable* Hjælpevariable fra registre er mere eller mindre korreleret med, hvor meget lastbilen kører, og hvor meget last den kan have med. Disse sammenhænge kan bruges til at reducere variansen ved fx et regressionsestimat i forhold til det direkte estimat (se appendiks).

*Alder på bilen* Af tabel 2 fremgår det, at stikprøven er stratificeret efter mange af de variable, der findes i registret. Der er nogle enkle hjælpevariable, der kan tilføjes i forbindelse med et regressionsestimater – alder, lasteevne og aksler. Lasteevne og antallet af aksler er stærkt korrelerede med totalvægt, som er en af de variable, der er stratificeret efter. Når hjælpevariable er stærkt korreleret indbyrdes betyder det, at det ikke gør stor forskel på variansen fra regressionsestimater, om alle hjælpevariablene er med eller kun en enkelt af dem. Alder er en hjælpevariabel, som man kan gøre sig håb om vil betyde noget for regressionsestimater, da den ikke indgår i nuværende stratificering.

Der er forsøgt opstillet forskellige regressionsestimater, og denne har vist sig mest effektiv:

$$y_i = (a1_{stratum} + a2_{aksler}) + b1_{stratum} \cdot lasteevne_i + b2_{stratum} \cdot alder_i + b3 \cdot totalvægt_i + e_i$$

$stratum = 1, 2, \dots, 14, aksler = 2, 3, 4$  og  $e_i$  er residualledet.

I denne model er konstanten  $a1_{stratum}$  og hældningskoefficienterne  $b1_{stratum}$  og  $b2_{stratum}$  bestemt inden for hvert af de 14 strata, mens konstanten  $a2_{aksler}$  er bestemt inden for hver af de tre grupper, som aksler udgør.

Det er ikke muligt at beregne punkt estimater og CV for International Gods-transport, men det er muligt at give en omtrentlig vurdering af usikkerheden. Dette er helt tilsvarende argument som i afsnittet Nuværende Metode, ovenfor.

Tabel 5. **Regressionsestimater<sup>1</sup>**

	National transport		International transport		Samlet transport	
	Punkt-estimat	Estimeret CV	Punkt-estimat	Estimeret CV	Punkt-estimat	Estimeret CV
Kørte km med læs i alt (mio. km)	1 218	2,4 pct.		ca. 4,8 pct.		ca. 2,3 pct.
Pålæsset godsmængde i alt (1000 ton)	179 572	3,6 pct.		ca. 5,2 pct.		ca. 3,3 pct.
Transportarbejde i alt (mio. tonkm)	11 628	3,3 pct.		ca. 5,9 pct.		ca. 3,2 pct.
Kørte km i alt (mio. km)	1 509	2,3 pct.	713	4,4 pct.	2 222	2,1 pct.

<sup>1</sup> Tal med kursiver pålagt EU-krav – CV max 2.5 pct.

*Bedre udnyttelse af hjælpeinformation giver skuffende lidt*

Af tabel 5 fremgår det, at regressionsestimater giver et meget skuffende resultat med hensyn til usikkerheden. Punktestimaterne i tabel 5 og tabel 4 er meget ens, så der er ikke biaskorrektion ved denne model. Det er kun alder, som yderligere er tilføjet i modellen i forhold til nuværende metode. Hvis der var store forskelle på punktestimaterne fra de to tabeller, ville det betyde at gamle eller nye biler var underrepræsenterede i stikprøve, og det havde en betydning for estimaterne.

*Tilnærmet usikkerhed for International Gods-transport*

For at give et bud på variationskoefficienter for International Gods-transport og Samlet Gods-transport er samme metode som tidligere anvendt. Det er ikke muligt at beregne regressionsestimater for International Gods-transport, hvor der samtidig er anvendt opjusteringsfaktoren fra National Gods-transport.

Til sammenligning af regressionsestimatet og EU-krav anvendes designeffekten (se appendiks). Designeffekten angiver, hvor meget det er nødvendigt at øge eller reducere en stikprøve, for at opnå det man vil (her EU-krav). Med en designeffekt på 48 pct. betyder det, at stikprøven skal fordoble, for at opfylde EU-krav.

Hvis designeffekten bruges til at vurdere, hvor lang DST er fra at opfyldes EU-krav, fås følgende:

Tabel 6. **Designeffekt (EU-krav mod regressionsestimatet)**<sup>1</sup>

	National transport	International transport	Samlet transport
Kørte km med læs i alt (mio. km)	<i>110 pct.</i>	110 pct.	<i>123 pct.</i>
Pålæsset godsmængde i alt (1000 ton)	<i>48 pct.</i>	91 pct.	<i>56 pct.</i>
Transportarbejde i alt (mio. tonkm)	<i>58 pct.</i>	72 pct.	<i>60 pct.</i>
Kørte km i alt (mio. km)	121 pct.	126 pct.	142 pct.

<sup>1</sup> Tal med *kursiver* pålagt EU-krav – CV max 2.5 pct.

*Nuværende stikprøve skal fordobles*

Når stikprøvedesignet beholdes, som det er, og der indføres regressionsestimat, vil stikprøven stadig være halv størrelse af, hvad den skal være for at opfylde EU-krav. 'Kørte km med læs', opfylder bekvemt EU-krav, men det har ingen betydning når 'Pålæsset godsmængde' er halvt så god som DST er forpligtet til (designeffekt 48 pct. og 56 pct.).

## Konklusion

Transportstatistikken er som udgangspunkt organiseret i to forskellige statistikker *National Godstransport* og *International Godstransport*.

*EU-krav lang fra opfyldt*

Der er EU-krav til statistikken, som langt fra er opfyldt med den nuværende metode. Heller ikke med optimering af opregningen kommer statistikken i nærheden af at opfylde EU-kravene. I det værste tilfælde er det nødvendigt at fordoble stikprøvestørrelsen for at opfylde EU-kravene.

*Vanskeligt design af International Godstransport*

På nuværende tidspunkt ser det ud til, at de to største problemer er *International Godstransports* meget specielle design og EU-krav, der ikke er opfyldt. Ud fra denne første analyse ser det desværre ikke ud til, at EU-krav kan opfyldes inden for de økonomiske rammer, der er stillet til rådighed i øjeblikket.

*Rigtig håndtering af populationen*

Der er behov for en bedre styring af versionerne af lastbilregistret, så udvalgs- og opregningspopulationer er adskilt. Det har nok mindre betydning, at der på nuværende tidspunkt ikke er rigtig håndtering af dette, men ved en omlægning statistikken vil dette også blive rettet.

*Overvejelse: Nedlæg International og udvid National Godstransport*

Det næste vil så være overvejelser om en omlægning af hele statistikken. Vi vil overveje, om *International* og *National Godstransport* kan foregå i samme undersøgelse. Risikoen ved denne tilgang er, at man finder meget få biler, der har kørt international – med meget stor usikkerhed til følge. Denne tilgang vil også indebære et krav om en øget stikprøve, men *International Godstransport* bliver lukket, og oplysningerne indsamles samlet. Dette giver givetvis en øget

respondentbyrde, ud over den direkte konsekvens af en øget stikprøve. Herudover skal der måske ændres helt på stikprøvedesign, så det bliver bedre til at fange biler, som kører internationalt.

## Appendiks

### Det direkte estimat

Det forudsættes, at der findes et register med alle lastbiler i Danmark, og at der trækkes en stikprøve fra dette register.

Antallet af lastbiler i populationen betegnes med  $N$ , antallet af lastbiler i stikprøven  $S$  betegnes med  $n$ , og antallet af responderende lastbiler betegnes med  $m$  ( $m \leq n$ ). I det direkte estimat bruges ikke nogen hjælpeinformation i opregningen. Det direkte estimat på variabel  $Y$ , er

$$\hat{t}_y = \frac{N}{m} \sum_{i \in S} y_i = \sum_{i \in S} d_i y_i,$$

*Designvægt* hvor  $y_i$  er bil  $i$ 's score på variabel  $Y$ , og  $d_i = N/m$  dennes (nonresponskorrigerede) designvægt.

### Variationskoefficient (CV)

Variationskoefficienten er en måde at gøre usikkerheden uafhængig af den enhed, målvariablen bliver beregnet i. Spredningen (kvadratet af variansen) bliver sat i forhold til estimatet af målvariablen.

$$CV(\hat{t}_y) = \frac{\sqrt{\text{Var}(\hat{t}_y)}}{\hat{t}_y},$$

hvor  $\hat{t}_y$  er estimatet på målvariablen, og  $\sqrt{\text{Var}(\hat{t}_y)}$  er spredningen på estimatet.

### Variationskoefficient for summen af to variationskoefficienter

Vi ønsker at kende variationskoefficienten for summen af to variationskoefficienter. I udregningen nedenfor er antaget at de to målvariable er uafhængige. Da fås følgende:

$$\begin{aligned}
CV(\hat{t}_{y_1} + \hat{t}_{y_2}) & \stackrel{uafh.}{=} \frac{\sqrt{Var(\hat{t}_{y_1}) + Var(\hat{t}_{y_2})}}{\hat{t}_{y_1} + \hat{t}_{y_2}} \\
& = \sqrt{\frac{Var(\hat{t}_{y_1}) + Var(\hat{t}_{y_2})}{(\hat{t}_{y_1} + \hat{t}_{y_2})^2}} \\
& = \sqrt{\frac{Var(\hat{t}_{y_1})}{(\hat{t}_{y_1} + \hat{t}_{y_2})^2} + \frac{Var(\hat{t}_{y_2})}{(\hat{t}_{y_1} + \hat{t}_{y_2})^2}} \\
& = \sqrt{\frac{\hat{t}_{y_1}^2 \cdot Var(\hat{t}_{y_1})}{\hat{t}_{y_1}^2 \cdot (\hat{t}_{y_1} + \hat{t}_{y_2})^2} + \frac{\hat{t}_{y_2}^2 \cdot Var(\hat{t}_{y_2})}{\hat{t}_{y_2}^2 \cdot (\hat{t}_{y_1} + \hat{t}_{y_2})^2}} \\
& = \sqrt{CV^2(\hat{t}_{y_1}) \frac{\hat{t}_{y_1}^2}{(\hat{t}_{y_1} + \hat{t}_{y_2})^2} + CV^2(\hat{t}_{y_2}) \frac{\hat{t}_{y_2}^2}{(\hat{t}_{y_1} + \hat{t}_{y_2})^2}}
\end{aligned}$$

Variationskoefficienten for summen af de to målvariable bliver altså mindre end summen af de to variationskoefficienter. Der ganges nogle korrektionsfaktorer på, som er afhængige af størrelsen på de to målvariable.

### Stratificeret direkte estimat

Det forudsættes, at registeret med alle lastbiler i Danmark indeholder forskellige variable, som man med fordel kan inddele sin population i. Det er en fordel at gruppere lastbiler, der ligner hinanden, fordi de må forventes at køre mere ens end resten af populationen. Det er muligt at inddele populationen i  $h=1, \dots, H$  disjunkte strata. Antallet af biler i populationen i stratum  $h$  angives ved  $N_h$  og antallet af biler i stikprøven i stratum  $h$  angives med  $n_h$ , og antallet af responderende biler i stratum  $h$  er  $m_h$ . Det direkte estimat på variabel  $Y$ , er

$$\hat{t}_y = \sum_{i=1}^H \frac{N_h}{m_h} \sum_{i \in S_h} y_i,$$

hvor  $y_i$  er bil  $i$ 's score på målvariabel  $Y$  og  $S_h$  er stikprøven i stratum  $h$ .

### Regressionsestimater

*Hjælpeinformation*

I regressionsestimater kan man bruge hjælpevariable. Man kan bruge interaktionerne af kategoriske hjælpevariable i opregningen (multiplikativ model), eller man kan nøjes med de marginale effekter (lineær model), 2-dimensionale effekter eller andre. Desuden kan man bruge numeriske variable som hjælpeinformation. En udvidet beskrivelse af regressionsestimater kan findes i fx Särndal et al. (1992). Uden at gå alt for meget i detaljer beskrives vægtene som beregnes vha. regressionsestimater her.

*Vægte*

Regressionsestimater går ud fra, at man kan forklare en målvariabel  $Y$  ud fra  $J$  udsøgte hjælpevariable  $X_1, \dots, X_J$ . Hvis man indsætter bilens værdier på de  $J$  hjælpevariable i en vektor

$$\mathbf{x}_i = (x_{i1}, \dots, x_{iJ})'$$



(hvor en del af de  $J$  variable kunne være dummyvariable for at indikere om bil  $i$  tilhører bestemte kategorier af en kategorisk variabel, og resten af de  $J$  variable kunne være numeriske variable), så kan man skrive regressionsmodellen som:

$$y_i = \mathbf{B}'\mathbf{x}_i + e_i,$$

hvor  $\mathbf{B}$  er regressionskoefficienten, og  $e_i$  er residualen. Regressionsvægten til bil  $i$  er givet ved:

$$w_i = \frac{N}{m} \left( 1 + (\mathbf{t}_x - \hat{\mathbf{t}}_x)' \left( \sum_{i \in S} \frac{N}{m} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \mathbf{x}_i \right) \equiv d_i g_i,$$

hvor  $d_i = N/m$  er bilens (nonresponskorrigerede) designvægt, og faktoren  $g_i$  betegnes som 'korrektionsvægten'. Dette er korrektionen i vægten på grund af regressionsestimatet. I denne formel indeholder vektoren  $\mathbf{t}_x$  populationstotalerne på de  $J$  hjælpevariable (der er kendt fra registret), og

$$\hat{\mathbf{t}}_x = \frac{N}{m} \sum_{i \in S} \mathbf{x}_i$$

er en vektor med estimerede populationstotaler på de  $J$  hjælpevariable (de bliver estimeret ud fra stikprøven med det direkte estimat).

*Kalibrering på registertotaler*

En vigtig egenskab af regressionsvægtene er, at hvis man estimerer totalerne på hjælpevariablene med regressionsvægte, så reproduceres de kendte populationstotaler  $\mathbf{t}_x$  præcist.

$$\sum_{i \in S} w_i \mathbf{x}_i' = \frac{N}{m} \sum_{i \in S} \mathbf{x}_i' + (\mathbf{t}_x - \hat{\mathbf{t}}_x)' = \mathbf{t}_x'.$$

Det betyder, at ved at bruge regressionsestimatet bliver stikprøven kalibreret på de kendte populationstotaler fra registret, og man reproducerer altså populationstotaler  $\mathbf{t}_x$ .

Estimater fås på samme måde som i poststratifikationen, blot bruges regressionsvægtene. Estimatet på variabel  $Y$  (hvor  $Y$  fx er kørte km i alt) bliver således

$$\hat{t}_y = \sum_{j \in S} w_j y_j.$$

*Nogle bemærkninger*

I regressionsestimatet kan man bruge både numeriske og kategoriske variable som hjælpeinformation. Et eksempel på en numerisk variabel er alder på lastbilen. Et eksempel på en kategorisk variabel er antallet af hjul, lastbilen har.

## Designeffekt

*Faktor til vurdering af forskellige metoder*

Designeffekten er en faktor, der kan bruges til at vurdere forskellige metoder. Det er en sammenligning af variansen mellem to forskellige tilgange. Designeffekten (Deff) er defineret som:

$$Deff = \frac{Var_1(\hat{t}_y)}{Var_2(\hat{t}_y)} = \frac{CV_1^2(\hat{t}_y)}{CV_2^2(\hat{t}_y)},$$

hvor  $Var_1(\hat{t}_y)$  er variansen ved første metode, som sammenlignes med  $Var_2(\hat{t}_y)$ , som er variansen af  $t_y$  ved anden metode. Dette er det samme som forholdet mellem kvadratet af variationskoefficienterne (CV) fra de to metoder.

## Referencer

Särndal, C.-E., Swensson, B. og Wretman, J. (1992). Model Assisted Survey Sampling, Springer.