



Bruxelles, den **XXX**
SWD(2014) **XXX**

ARBEJDSDOKUMENT FRA KOMMISSIONENS TJENESTEGRENE

Fælles metodologi til evaluering af statsstøtte

Indholdsfortegnelse

1	Indledning.....	2
2	Formålene med evaluering af statsstøtten	3
3	Evalueringsplanen	5
3.1	Formålene med den støtteordning, som skal evalueres	5
3.2	Evalueringsspørgsmålene	5
3.3	Resultatindikatorer	6
3.4	Metoder: tilvejebringelse af et hensigtsmæssigt sammenligningsgrundlag.....	6
3.5	Dataindsamling: anvendelse af de bedste mulige kilder.....	9
3.6	Tidsplan for evalueringen	10
3.7	Evalueringsgruppens uafhængighed og ekspertise.....	11
3.8	Offentliggørelse med henblik på inddragelse af interessenterne	12
4	Kriterier for udvælgelse af støtteordninger, som skal evalueres.....	12
4.1	Store støtteordninger, herunder ordninger under den generelle gruppefritagelsesforordning	13
4.2	Nye typer støtteordninger.....	14
4.3	Støtteordninger, der påvirkes af større forudsete ændringer	14
4.4	Andre støtteordninger.....	14
Bilag I:	Teknisk appendiks om relevante metoder til måling af kausale effekter	16
Bilag II:	Liste over mulige resultatindikatorer	35
Bilag III:	Ordliste	39
Bilag IV:	Litteraturliste	40

1 Indledning

Medlemsstaterne yder statsstøtte for at bidrage til gennemførelsen af en lang række politiske mål, f.eks. for at mindske de regionale forskelle inden for et land, fremme forsknings-, udviklings- og innovationsaktiviteter eller fremme et højt miljøbeskyttelsesniveau.

EU's statsstøtteregler er baseret på et system med forudgående kontrol af, hvilke former for støtte der er forenelige med det indre marked: Godkendelsen af støtteordninger¹ sker på grundlag af foruddefinerede vurderingskriterier ud fra en antagelse om, at hvis de er i overensstemmelse med disse vurderingskriterier, vil deres positive virkninger opveje eventuelle negative virkninger. Denne vurdering af ordningerne udføres typisk, uden at der foretages en tilstrækkelig evaluering af deres faktiske indvirkning på markederne over tid.

Ved anvendelsen af EU's statsstøtteregler har man hidtil kun lagt forholdsvis begrænset vægt på efterfølgende at få dokumentation for, hvad der rent faktisk er opnået med de offentlige midler, eller hvilken indvirkning statsstøtten har haft på konkurrencen. Det er imidlertid vigtigt, at beslutningstagerne i både medlemsstaterne og på EU-niveau ser nærmere på de målbare resultater af tidligere ydet statsstøtte og erfaringerne med den. Det vil bidrage til at sikre, at ordninger, der finansieres via statsstøtte, bliver mere effektive og skaber mindre markedsfordrejning, og det vil også forbedre effektiviteten af fremtidige ordninger og eventuelt de fremtidige regler for tildeling af statsstøtte.

En række lande foretager allerede evalueringer af deres støtteforanstaltninger, selvom det ikke sker regelmæssigt.² På samme måde er EU's udgifter (herunder finansiering fra EU's struktur- og investeringsfonde som EFRU, ESF og ELFUL) genstand for forudgående, løbende og efterfølgende evaluering i henhold til de gældende forordninger og de vejledningsdokumenter, som Kommissionen udsender.³ For at undgå dobbeltarbejde i forhold til de evalueringer, der udføres af medlemsstaterne, præciseres det i vejledningsdokumentet "Concepts and Recommendations" om overvågning og evaluering, at evalueringskravene for de europæiske struktur- og investeringsfonde kan opfyldes ved at gennemføre de evalueringer, der kræves i henhold til statsstøttereglerne.

1 Støtteordninger tegner sig for størstedelen af al statsstøtte: Ifølge tal fra resultattavlen for 2013 udgør godkendte støtteordninger 23 % af alle støtteforanstaltninger og 55 % af støttebeløbene, og en yderligere gruppe af gruppefritagede støtteordninger udgør 63 % af alle støtteforanstaltninger og omkring 32 % af støttebeløbene. I Rådets forordning nr. 659/1999 defineres "støtteordning" som "enhver retsakt, på grundlag af hvilken der uden yderligere gennemførelsesforanstaltninger kan ydes individuel støtte til virksomheder, som inden for denne retsakt defineres generelt og abstrakt, og enhver retsakt på grundlag af hvilken støtte, der ikke er knyttet til et specifikt projekt, kan ydes til en eller flere virksomheder, uden at tidsrummet eller beløbet er nærmere fastsat".

2 I adskillige medlemsstater udarbejdes der f.eks. regelmæssigt statsstøtteevalueringsrapporter til statsrevisorerne eller parlamentet.

3 Kommissionens vejledningsdokumenter om evaluering for finansieringsperioden 2014-20 (findes her: http://ec.europa.eu/regional_policy/information/evaluations/guidance_en.cfm) indeholder en detaljeret beskrivelse af de relevante koncepter og anbefalinger.

Programmet for modernisering af statsstøttepolitikken⁴ tager sigte på at koncentrere Kommissionens håndhævelsesindsats om større støtteordninger, der efter al sandsynlighed vil have de største indvirkninger på det indre marked. Samtidig skal analysen af støtteforanstaltninger af mere lokal karakter og mindre eller mere begrænsede indvirkninger på samhandelen forenkles, bl.a. ved udvide anvendelsesområdet for den nye generelle gruppefritagelsesforordning⁵. For at sikre, at de positive virkninger af statsstøtten (i relation til dens oprindelige formål) samlet set fortsat vil opveje de potentielle negative virkninger for konkurrence og samhandel, og for at forhindre urimelig markedsfordrejning, bør den øgede forenkling kombineres med større gennemsigtighed, skærpet kontrol med overholdelsen af statsstøttereglerne på nationalt og europæisk niveau samt en effektiv evaluering⁶.

Dette arbejdsdokumentet indeholder en fælles metodologi for evaluering af statsstøtteordninger. Det har til formål at vejlede de offentlige myndigheder, der har ansvaret for at planlægge og gennemføre evalueringerne.

2 Formålene med evaluering af statsstøtten

Det overordnede formål med evaluering af statsstøtten er at vurdere de relative positive og negative virkninger af en ordning, dvs. støttens samfundsmæssige formål forhold til dens indvirkning på konkurrencen og samhandelen mellem medlemsstaterne. Denne evaluering kan give svar på spørgsmålene, om og i hvilket omfang målene for en støtteordning er nået (dvs. vurdering af de positive virkninger), og hvilke virkninger ordningen har på markeder og konkurrence (dvs. eventuelle negative virkninger). Evalueringen har derfor et andet formål end de to former for efterfølgende kontrol, som Kommissionen i øjeblikket foretager af statsstøtteordninger, nemlig overvågning⁷ og rapportering⁸.

Evalueringen af statsstøtte bør navnlig gøre det muligt at vurdere støttens direkte tilskyndelsesvirkning for støttemodtageren (dvs. om støtten har fået modtageren til at handle anderledes, og hvor stor indvirkning støtten har haft). Den bør også give en indikation af støtteordningens overordnede positive og negative virkninger for realiseringen af det tilsigtede politiske mål samt for konkurrence og samhandel og kan også afklare, hvorvidt det valgte støtteinstrument er proportionalt og hensigtsmæssigt.

4 Meddelelse fra Kommissionen til Europa-Parlamentet, Rådet, Det Europæiske og Sociale Udvalg og Regionsudvalget, "Modernisering af EU's statsstøttepolitik, 8.5.2012, COM(2012)0209 final.

5 Kommissionens forordning (EU) nr. .../2014 af XXX om visse former for støttes forenelighed med det indre marked i henhold til traktatens artikel 107 og 108.

6 Se også Rådets konklusioner om reform af kontrollen med statsstøtte af 13. november 2012.

7 Kommissionens overvågning består i en regelmæssig gennemgang af en stikprøve af statsstøtteforanstaltninger iværksat af medlemsstaterne. Formålet er at sikre, at medlemsstaterne gennemfører Kommissionens afgørelser korrekt og overholder de relevante retsregler (dvs. bestemmelserne i den generelle gruppefritagelsesforordning). Kommissionen vurderer også overholdelsen af ex ante-regler og -betingelser i en repræsentativ stikprøve af sager.

8 Det primære formål med medlemsstaternes årlige rapportering er at øge gennemsigtigheden omkring medlemsstaternes statsstøtte. Den er også en kilde til pålidelige statistikker til politisk beslutningstagning og tilsynsformål. Dataene i de årlige rapporter indeholder navnlig kvantitative oplysninger (f.eks. om de mål, statsstøtten er rettet imod, og om det budgetniveau der er tale om). Kommissionen bruger medlemsstaternes rapporter til at udarbejde resultattavlen for statsstøtte.

På grundlag af denne vurdering kan evalueringen bekræfte, hvorvidt de formodninger, der lå til grund for den forudgående godkendelse af støtteordningen, stadig gælder, og kan dermed bidrage til at forbedre udformningen af fremtidige støtteordninger og statsstøtteregler. Den kan danne grundlag for en tilpasning af fremtidige indgreb fra statens side med henblik på at forbedre støttens effektivitet og omkostningseffektivitet i den udstrækning, det er nødvendigt for at sikre, at de positive virkninger er tilstrækkelige til at opveje den markedsfordrejning, som dette indgreb forårsager. Sådanne forbedringer af fremtidige ordninger kan omfatte både tilpasninger af udformningen, herunder ændringer af udvælgelseskriterierne og en mere omfattende vurdering af tilskyndelsesvirkningen, og mere gennemgribende ændringer såsom overgang til en alternativ støtteform, omdefinering af mål eller støttemodtagere eller overvejelser om at anvende andre virkemidler end statsstøtte til at nå de samme politiske mål.

Det er vigtigt at fastsætte en hensigtsmæssig tidsplan for evalueringen, der giver tid nok til at indsamle tilstrækkelig dokumentation og samtidig gør det muligt at levere resultater til de politiske beslutningstagere så hurtigt som muligt, så de får mulighed for at indføre eventuelle forbedringer i tide.⁹ I lyset heraf bør evalueringer af statsstøtte normalt foregå løbende og foretages, mens støtteordningen stadig er i kraft, i stedet for kun at være efterfølgende evalueringer, der først foretages, efter at ordningen er ophørt. Der bør tages højde for særlige tilfælde, hvor de samlede virkninger af en foranstaltning måske først bliver synlige på længere sigt, og hvor evalueringen kun kan registrere og måle de første virkninger.

Evalueringen af statsstøtten skal i sidste ende fungere som en læreproces for såvel Kommissionen som medlemsstaterne. Det kan kun lade sig gøre, hvis evalueringen opfylder visse minimumskrav med hensyn til kvalitet. Derfor bør Kommissionen sikre, at der foretages en hensigtsmæssig kvalitetskontrol af evalueringerne. Kommissionen vil navnlig foretage en detaljeret analyse af evalueringens overordnede troværdighed og pege på eventuelle mangler i de to centrale faser, nemlig evalueringsplanen og den endelige rapport. Når det er relevant, kan Kommissionen søge bistand hos uafhængige eksterne eksperter, som kan bistå ved kvalitetskontrollen af evalueringen.

Kommissionen kan også afholde kurser og workshops for de nationale myndigheder om evalueringemetoder og -teknikker. Desuden kan vellykkede erfaringer og bedste praksis fra medlemsstaterne udveksles og anvendes til at udforme mere effektive støtteordninger for fremtiden.

Fordelene ved at gennemføre evalueringer vil blive indlysende inden for få år, når de første evalueringsrapporter er klar, og deres resultater og anbefalinger bliver tilgængelige. Herefter kan de bruges til at forbedre udformningen af efterfølgende støtteordninger og eventuelt statsstøttereglerne. På mellemlang til lang sigt kan evalueringerne gradvist føre til mere grundlæggende ændringer i den overordnede tilgang til statsstøtte.

⁹ I nogle statsstøttereigningsligner opereres der med en normal varighed af evaluerede støtteordninger på fire år.

3 Evalueringsplanen

Det er vigtigt, at en samlet plan for evaluering af en statsstøtteordning opstilles på et tidligt tidspunkt, sideløbende med udformningen af selve ordningen. For at sikre ligebehandling er det vigtigt, at evalueringsplanen bliver godkendt af Kommissionen. Planen skal derefter gennemføres stringent.

Det er almindeligt anerkendt, at evalueringer er mere effektive, når de planlægges ordentligt og forberedes på forhånd, navnlig fordi det gør det nemmere at indsamle de relevante data. Tidlig planlægning vil formentlig også betyde, at der er behov for væsentlig færre ressourcer til evalueringen, og vil i sidste ende forbedre dens kvalitet.

Evalueringsplanen, som medlemsstaten skal anmelde til Kommissionen i overensstemmelse med de relevante regler, skal som minimum indeholde følgende elementer.

3.1 Formålene med den støtteordning, som skal evalueres

Første fase i evalueringen af en ordning består af en klar beskrivelse af den underliggende "interventionslogik" for støtteordningen, hvor man beskriver de behov og problemer, som det er hensigten at afhjælpe med ordningen, støttemodtagerne og investeringerne, de overordnede og specifikke mål samt de forventede virkninger. De centrale antagelser vedrørende de eksterne faktorer, der kan påvirke ordningen, bør også nævnes.

3.2 Evalueringsspørgsmålene

Evalueringsplanen bør indeholde en definition af evalueringens omfang, og den bør derfor omfatte præcise spørgsmål, der kan besvares kvantitativt og med den nødvendige dokumentation. Disse evalueringsspørgsmål, som fokuserer på statsstøtteordningens virkninger, kan opdeles i følgende tre niveauer:

1. Støttens direkte virkning for støttemodtagerne, f.eks.:

- Har støtten haft betydelig indvirkning på støttemodtagernes handlinger (tilskyndelsesvirkning)?
- Har støtten haft indvirkning på støttemodtagernes situation (f.eks. ved at ændre deres konkurrenceposition eller kreditværdighed)?
- I hvilket omfang har støtten haft de ventede virkninger?
- Er støttemodtagerne blevet påvirket forskelligt af støtten (f.eks. alt efter deres størrelse, lokalisering eller sektor)?

2. Støttens indirekte virkning, f.eks.:

- Har ordningen haft afsmittende virkninger på aktiviteten i andre virksomheder eller på andre geografiske regioner? Har støtten fortrængt investeringer fra andre konkurrenter eller tiltrukket aktiviteter fra naboerområder?
- Har ordningen bidraget til det relevante politiske mål?
- Kan ordningens samlede virkninger for konkurrence og samhandel måles?

3. Støtteordningens proportionalitet og hensigtsmæssighed, f.eks.:

- Var støtteordningen proportional med det problem, der skal afhjælpes? Kunne de samme virkninger have været opnået med mindre støtte eller en anden form for støtte (f.eks. lån i stedet for tilskud)?
- Valgte man det mest effektive støtteinstrument? Ville andre støtteinstrumenter eller andre virkemidler have været mere hensigtsmæssige med henblik på at nå det pågældende mål?

Ved denne evaluering bør man så vidt muligt vurdere støtteordningens indvirkning på alle tre niveauer og søge at afklare de relevante spørgsmål i forhold til målsætningerne for ordningen. Støttens direkte indvirkning på modtagerne er dog typisk den type indvirkning, der kan måles mest pålideligt. I praksis er størstedelen af evalueringsmetoderne udformet til at vurdere virkninger af denne art. Desuden er det af central betydning at evaluere støttens direkte virkninger, inklusive tilskyndelsesvirkningen, eftersom dette kan give et værdifuldt indblik i, hvilke typer af indirekte virkninger og fordrejninger der kan forventes. Især hvis støtten ikke skaber nogen tilskyndelsesvirkning, må den antages at have en fordrejende virkning i den forstand, at den giver de pågældende støttemodtagere ekstraordinære fordele.

3.3 Resultatindikatorer

Spørgsmålene i evalueringen bør føre til, at man vælger specifikke resultatindikatorer, der kan give kvantificerede oplysninger om resultaterne af statsstøtteordningen. Bilag II indeholder en vejledende og ikkeudtømmende liste over resultatindikatorer, der både dækker de direkte og indirekte virkninger af en ordning, inklusive de mulige indvirkninger på konkurrence og samhandel. Resultatindikatorerne vil afhænge af formålet med den støtte, der skal evalueres. I evalueringsplanen bør man forklare, hvorfor de valgte indikatorer er de mest relevante til måling af virkningerne af denne støtteordning.

3.4 Metoder: tilvejebringelse af et hensigtsmæssigt sammenligningsgrundlag

Evalueringer af statsstøtte bør kunne identificere selve ordningens kausale effekt uden påvirkning fra andre variabler, der kunne påvirke de konstaterede resultater, f.eks. de generelle makroøkonomiske betingelser eller heterogene virksomheder (f.eks. forskelle i virksomhedernes størrelse, lokalisering, økonomiske eller ledelsesmæssige ressourcer). Evalueringsplanen skal indeholde en beskrivelse af de vigtigste metoder, der vil blive benyttet til at måle støttens virkninger, og en diskussion af, hvorfor disse metoder formentlig vil være hensigtsmæssige for den pågældende ordning.

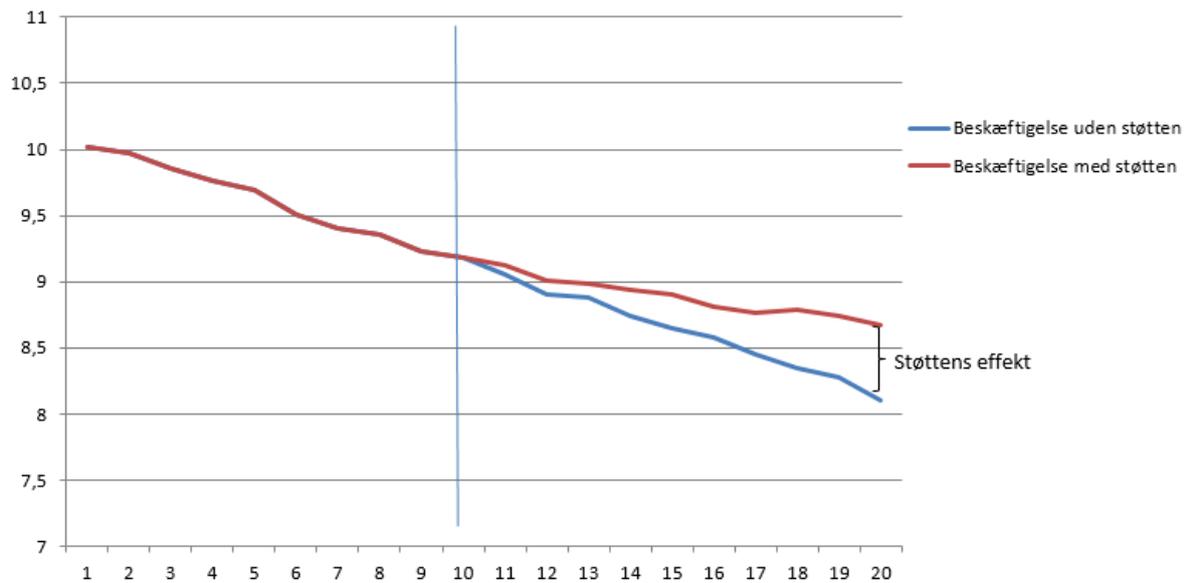
Denne kausale effekt er forskellen mellem resultaterne med støtten og resultaterne uden støtten. Mens resultaterne med støtten iagttages for virksomheder, der modtager støtten, måles resultaterne uden støtten kun for virksomheder, der ikke modtager støtte. Pr. definition iagttager vi ikke, hvad resultaterne ville have været uden støtten for de virksomheder, der modtog støtte. For at vurdere støttens virkninger for støttemodtagerne er det derfor nødvendigt at konstruere den kontrafaktiske situation baseret på de(n) mest sammenlignelige virksomhed(er) eller kontrolgruppe.

Kvaliteten af denne kontrolgruppe er afgørende for evalueringens gyldighed.

Virksomheder, der modtager støtte, kan meget vel befinde sig i en anden situation end virksomheder, der ikke modtager støtte. De kan f.eks. stå over for anderledes lokale forhold, hvad angår udbud og efterspørgsel, have vanskeligere ved at få adgang til kredit eller være mere eller mindre effektive. Disse faktorer kan alle påvirke virksomhedernes resultater eller aktivitetsniveau, både når de modtager støtte, og når de ikke gør. Sammenligningen af støttemodtagernes resultater med ikkestøttemodtageres vil sandsynligvis i højere grad afspejle denne virkelighed end virkningen af selve støtten. En evaluering af støtteordningen kan derfor ikke udelukkende baseres på en sammenligning mellem støttemodtagere og ikkestøttemodtagere, men skal også tage højde for de forskellige kendetegn ved de to grupper af virksomheder, både dem, der kan observeres, og dem, der ikke kan.

For regionalstøttes vedkommende er det f. eks. et typisk fænomen, at støttemodtagere i regioner med ugunstige markedsbetingelser (dvs. hvor de lokale produkt-, arbejds- eller kapitalmarkeder er svage) klarer sig dårligere end ikkestøttemodtagere i mere fremgangsrigge regioner. Dette er imidlertid på ingen måde en afspejling af selve støttens virkninger. Det relevante spørgsmål er, hvorvidt de klarer sig bedre, end de ville have gjort uden støtten, ikke om de klarede sig bedre end ikkestøttemodtagere i andre regioner.

Tilsvarende skal man også tage højde for generelle tendenser inden for branchen, når man ønsker at identificere støttens virkninger. Selv om modtagerne af regionalstøtte måske reducerer antallet af medarbejdere, kan støtten stadig have været effektiv. Når betingelserne inden for en bestemt branche som helhed f.eks. forværres, og alle virksomheder skærer arbejdspladser væk, kan det være, at støttemodtagerne reducerer medarbejderantallet mindre, end de ellers ville have gjort. Dette illustreres af nedenstående diagram, som viser en negativ tendens for antallet af arbejdspladser i virksomheder, der modtager støtte, både før og efter støtten blev tildelt. Men tendensen bliver mindre negativ, efter at virksomheden har modtaget støtten. Forskellen på den forlængede linje, der viser tendensen uden støtten, og den linje, der viser det faktiske antal arbejdspladser i virksomheden efter støtten, isolerer støttens positive indvirkning.



Figur 1 — Støttens positive indvirkning, når den aktuelle tendens er negativ

Der opstår et specifikt problem med at identificere en kontrolgruppe, når ikkestøttemodtagere selv har besluttet, om de vil ansøge eller ikke ansøge om statsstøtte. Hvis alle virksomheder er støtteberettigede (dvs. at alle virksomheder, der har et projekt og ansøger om støtte til det, modtager en vis støtte), vil de virksomheder, der ikke ansøger om støtte, sandsynligvis være dem uden projekter. Virksomhedernes resultater kan vise, at virksomheder, der ikke modtog støtte, klarede sig dårligere i absolutte og relative tal end virksomheder, der modtog støtte. Denne konstatering kan imidlertid forklares fuldt ud med, at den førstnævnte gruppe ikke havde noget projekt til at begynde med, mens sidstnævnte gruppe havde et projekt, dvs. at ledelsen i førstnævnte gruppe savner interesse eller kreativitet. Derfor er det vigtigt, at virksomhederne i kontrolgruppen (virksomheder, der ikke har modtaget støtte) indgår i denne gruppe af årsager, som ikke har nogen indvirkning på de målbare resultater. Når virksomhederne selv har truffet et bevidst valg om ikke at ansøge om støtte, er denne betingelse muligvis ikke opfyldt.

Der bør tages højde for enhver systematisk forskel mellem modtagere og ikkemodtagere af statsstøtte ved tilrettelæggelsen af evalueringen for at undgå en skævhed i resultaterne (skævhed i udvælgelsen). I de seneste årtier har man udviklet adskillige pålidelige metoder til at afhjælpe dette problem. Valget af metoden afhænger af udformningen af den enkelte statsstøtteordning og af de tilgængelige oplysninger. Metoderne har alle deres begrænsninger og er kun gyldige, når visse antagelser holder stik. Det er vigtigt for en undersøgelses troværdighed, at man anerkender og diskuterer disse begrænsninger og antagelser.

Randomisering af processen for udvælgelse af støttemodtagere er en af de måder, hvorpå man kan sikre, at evalueringen ikke bliver skæv. Hvis støttemodtagerne udelukkende udvælges tilfældigt, kan enhver systematisk forskel, som kan iagttages i virksomhedernes resultater, tilskrives støtten. Denne metode kan imidlertid være vanskelig at anvende i praksis, navnlig på store eksisterende ordninger. Andre metoder sigter mod at anvende eksisterende kilder til eksogen variation i virksomhedernes driftsmiljø (dvs. at variationen ikke bestemmes af

parametrene og variablerne i modellen) til at identificere kausalitet.¹⁰ Bilag I til denne vejledning indeholder nærmere oplysninger om de mest relevante metoder, hvor man fokuserer på de praktiske aspekter af deres anvendelse. Her diskuteres det, hvordan kausaliteten identificeres inden for de enkelte metoder, eftersom dette har særligt stor betydning i forbindelse med evaluering af statsstøtte, hvor den forudgående udformning af evalueringen har til formål at sikre, at der er mulighed for en korrekt evaluering af støttens virkninger.

Endelig skal man undersøge, om der forekommer virkninger som følge af flere forskellige slags støtte, enten fra forskellige ordninger eller ad hoc-støtte. Hvis ikkemodtagerne inden for det givne program modtager støtte under andre programmer, eller hvis modtagerne under et givet program modtager supplerende støtte fra andre programmer, vil evalueringen af virkningerne af den pågældende støtteordning sandsynligvis blive skævvredet.

3.5 Dataindsamling: anvendelse af de bedste mulige kilder

Der skal indsamles konsekvente og tilstrækkelige data om såvel støttemodtagerne som kontrolgruppen. Det er en del af planlægningen af evalueringen at identificere de krævede data og opnå adgang til datakilderne.

En effektiv overvågning af ordningen og en præcis dataindsamling og -behandling er af central betydning for evalueringens kvalitet. Så snart støtteordningen er godkendt, bør der derfor indføres en mekanisme for overvågning af dens gennemførelse og til indsamling og behandling af de pågældende data. Dette vil sandsynligvis nedbringe udgifterne til evalueringen betydeligt.

Det er et vigtigt skridt i evalueringsplanen at sikre, at de nødvendige data om støtteansøgere og -modtagere indsamles, hvis oplysninger herom kan gøres til en del af betingelserne for støtteberettigelse.

Med undtagelse af data om støtteansøgninger (herunder ansøgere, der har fået afslag, hvis dette er relevant) skal datakilderne for støttemodtagerne og kontrolgruppen være identiske, hvis dataene skal være sammenlignelige. Det er meget sandsynligt, at disse data vil skulle hentes fra flere forskellige kilder, f.eks. ved at kombinere data fra databaser med oplysninger om støttemodtagere med data fra virksomhedsregistre. Ved evalueringen kan det være nødvendigt at trække på eksisterende datakilder (f.eks. skattevæsenet, virksomhedsregistret, innovationsundersøgelser og patentkontoret). Derfor skal man i evalueringsplanen se på, hvilke datakilder der findes, tage stilling til, om de indeholder tilstrækkelige oplysninger til evalueringen, og sikre, at det er muligt at få adgang til dem inden for de relevante frister.

Der vil sandsynligvis kun være adgang til data fra kilder i den offentlige forvaltning, f.eks. nationale statistikkontorer, på særlige betingelser vedrørende beskyttelse af personoplysninger og virksomhedsoplysningers fortrolighed. Betingelserne for adgangen til disse data skal

¹⁰ De mest almindeligt benyttede metodologier er differences-in-differences, regression discontinuity design og instrumentvariabler.

beskrives i evalueringsplanen. Når det er nødvendigt, skal den myndighed, der giver adgang til dataene, sikre, at evalueringseksperterne har adgang til disse data.

Når man benytter data fra forskellige kilder, er det meget vigtigt, at de indsamles i et format, der giver mulighed for en konsekvent sammenligning af variabler. Det kan være nødvendigt at finde entydige identifikatorer for observationsenhederne i hvert af de benyttede datasæt. Identifikatorerne for virksomheder og anlæg skal f.eks. være de samme i alle datasæt, adresserne skal indsamles i et format, der giver mulighed for geolokalisering osv. Hvilke identifikatorer der benyttes, kan variere fra medlemsstat til medlemsstat. De kan f.eks. stamme fra skattemyndighederne (f.eks. et momsnummer) eller fra statistikkontoret (f.eks. SIREN og SIRET i Frankrig, som er de identifikationsnumre for henholdsvis virksomheder og virksomhedsenheder, der benyttes af det franske institut for statistik og økonomiske studier (INSEE)).

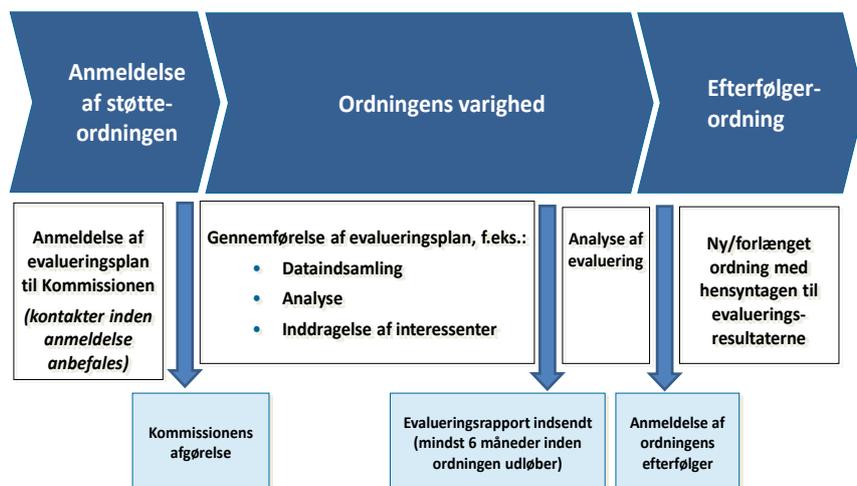
Evalueringen af statsstøtten kan suppleres med oplysninger fra undersøgelser af støttemodtagere og/eller interviews med de ansvarlige forvaltere af ordningen. Kvalitative oplysninger af denne art er naturligvis subjektive, og svarene kan afspejle støttemodtagernes strategiske interesser snarere end at være en egentlig vurdering af støttens virkninger. Denne risiko er særligt stor, hvis den interviewede antager, at et positivt vidnesbyrd vil forbedre mulighederne for at modtage støtte fremover. Men hvis oplysninger fra kvalitative kilder som interviews og casestudier behandles med den nødvendige forsigtighed, kan de bruges som et nyttigt supplement og bidrage til at fortolke resultaterne af evalueringen.

Når behandling af personoplysninger indgår i evalueringerne, finder EU-lovgivningen om beskyttelse af personoplysninger anvendelse, navnlig direktiv 95/46/EF om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger i fællesskabsinstitutionerne og -organerne og om fri udveksling af sådanne oplysninger og den nationale gennemførelseslovgivning samt forordning (EF) nr. 45/2001 om beskyttelse af fysiske personer i forbindelse med behandling af personoplysninger i fællesskabsinstitutionerne og -organerne og om fri udveksling af sådanne oplysninger.

3.6 Tidsplan for evalueringen

En evalueringsplan skal indeholde oplysninger om den nøjagtige tidsplan for evalueringen, som fastlægges i henhold til ordningens godkendte varighed, og bør indeholde delfrister, bl.a. for indsamling af data, gennemførelsen af evalueringen og indsendelsen af den endelige rapport. Tidsplanen kan variere afhængigt af ordningen og bør derfor drøftes og aftales med Kommissionen i hvert enkelt tilfælde. Det tilrådes, at de ansvarlige for forvaltningen af ordningerne stiller sig åbne over for uformelle drøftelser af planens indhold, inden de indsender deres officielle anmeldelse til Kommissionen.

Hvis der er planer om en forlængelse af støtteordningen, bør den endelige evalueringsrapport indsendes til Kommissionen i god tid (f.eks. seks måneder inden ordningen normalt skulle ophøre). Hvis der ikke er planer om forlængelse, kan rapporten indsendes, når ordningen er afsluttet.



Figur 2 — Oversigt over evalueringsprocessen for en anmeldt ordning

3.7 Evalueringsgruppens uafhængighed og ekspertise

Evalueringen af statsstøtteordningers virkninger skal være objektiv, stringent, upartisk og gennemsigtig.¹¹ Hver enkelt evaluering skal gennemføres på grundlag af velfunderede metoder og af eksperter, der besidder tilstrækkelig og dokumenteret erfaring samt den metodologiske viden til at foretage evalueringen.

Evalueringerne skal udføres af en evalueringsgruppe, der som minimum er funktionelt uafhængigt af den myndighed, der tildeler støtten, og som besidder de nødvendige, dokumenterede færdigheder og kvalificeret personale til gennemførelse af sådanne evalueringer. Evalueringsgruppens funktionelle uafhængighed af den myndighed, der tildeler støtten, er af central betydning for at sikre evalueringens kvalitet og troværdighed. Dette betyder ikke nødvendigvis, at der skal oprettes et nyt organ, eller at evalueringsopgaverne skal udliciteres til kommercielle evalueringseksperter. Afhængigt af de specifikke organisationer, der er til stede i de enkelte medlemsstater, kan det f.eks. være muligt at udnytte uafhængigheden og ekspertisen hos organisationer som statistikkontorer, centralbanker, revisionsretter, offentlige eller private universiteter eller forskningscentre. Dette kan besluttes fra tilfælde til tilfælde for de enkelte ordninger.

¹¹ Se f.eks. Europa-Kommissionens evalueringsstandarder, OECD's Evaluation Norms and Standards, FN's Evaluation Standards og Verdensbankens Independent Evaluation: Principles, Guidelines and Good Practice.

For at opnå en vellykket evaluering er det vigtigt, at evalueringsgruppen inddrages tidligt, f.eks. i forbindelse med udformningen af ordningen. På den måde sikrer man, at statsstøtteordningen vil kunne evalueres på den planlagte måde, og at de nødvendige data vil blive indsamlet. Når det er muligt, bør evalueringsplanen derfor udformes af eller i det mindste i et meget tæt samarbejde med den udpegede evalueringsekspert. Den bør også omfatte oplysninger, eventuelt kun af vejledende karakter, om de nødvendige menneskelige og økonomiske ressourcer, der vil blive stillet til rådighed for gennemførelsen af evalueringen. Oplysninger om de enkelte nøgleeksperter, der skal deltage i evalueringen, og deres rolle samt forventede involvering har særligt stor relevans.

Evalueringsplanen bør indeholde en præcis beskrivelse af den evalueringsgruppe, der skal foretage evalueringen, og hvis den endnu ikke er udpeget, de detaljerede kriterier, der vil blive anvendt ved udvælgelsen, navnlig vedrørende uafhængighed, erfaring og ekspertise. Den bør omfatte eventuelle mulige alternativer. Hvis evaluatoren endnu ikke er udvalgt eller er blevet udvalgt, men ikke har deltaget aktivt i udformningen af evalueringsplanen, skal årsagerne hertil klart angives. Selv i denne situation skal evalueringsplanen være tilstrækkeligt detaljeret til at give mulighed for en korrekt vurdering af evalueringens validitet.

3.8 Offentliggørelse med henblik på inddragelse af interessenterne

Evalueringen bør offentliggøres. Dette indebærer, at både evalueringsplanen og den endelige evalueringsrapport, når denne er godkendt, skal offentliggøres på en hensigtsmæssig måde, ved at gøre den tilgængelig på de steder, der beskrives i evalueringsplanen, f.eks. på et websted. Kommissionen kan også offentliggøre disse dokumenter¹².

Hvis de data, der anvendes til evalueringen, er personlige og/eller fortrolige, skal fortroligheden sikres gennem hele evalueringsprocessen, jf. artikel 8, 16 og 17 i EU's charter om grundlæggende rettigheder. Fortroligheden gælder imidlertid ikke resultaterne af evalueringen. I særdeleshed kan der ikke indsættes nogen fortrolighedsklausul i kontrakten om evalueringen, bortset fra: 1. fortrolighedsforpligtelser vedrørende personoplysninger og/eller fortrolige oplysninger og 2. forpligtelser til at overholde de generelle bestemmelser i den nationale statistiklovgivning og den statistiske fortrolighed, f.eks. i forbindelse med fremlæggelse af resultaterne.

Data, der indsamles i forbindelse med evalueringen, bør stilles til rådighed med henblik på reproducering af resultaterne eller yderligere undersøgelser på betingelser, der ikke er mere restriktive end dem, evalueringsgruppen var underlagt ved den oprindelige evaluering.

Den støtteydende myndighed bør sikre en hensigtsmæssig inddragelse af relevante interessenter, som bør høres mindst en gang under gennemførelsen af evalueringsplanen.

12 Med undtagelse af forretningshemmeligheder og andre fortrolige oplysninger i behørigt begrundede tilfælde (Kommissionens meddelelse om tavshedspligt i statsstøttebeslutningers, C(2003) 4582, EUT C 297, 9.12.2003, s. 6). Enhver offentliggørelse af personoplysninger skal ske i overensstemmelse med EU-lovgivningen om beskyttelse af personoplysninger, navnlig direktiv 95/46/EF og den nationale gennemførelseslovgivning samt forordning (EF) nr. 45/2001.

Interessenterne kan f.eks. indbydes til at drøfte de første resultater af evalueringen på grundlag af en statusrapport. Sådanne tiltag bør medtages i evalueringsplanen.

4 Kriterier for udvælgelse af støtteordninger, som skal evalueres

I princippet kan enhver statsstøtteordning gøres til genstand for evaluering, men selvom evaluering anses for god praksis, er den ikke et krav i alle tilfælde i henhold til statsstøttereglerne. Evalueringen af statsstøtte bør fortsat tilgodesee proportionalitetsprincippet og bør generelt udføres for ordninger, der potentielt kan have betydelig indvirkning på det indre marked og udgøre en risiko for betydelige markedsfordrejninger, hvis deres gennemførelse ikke evalueres i tide. Fokus i de relevante retningslinjer for statsstøtte ligger derfor på støtteordninger, der er: (1) store, herunder ordninger under den generelle gruppefritagelsesforordning, (2) nye eller (3) rummer mulighed for betydelige (markeds-mæssige, teknologiske eller reguleringsmæssige) ændringer i den umiddelbare fremtid, der kan kræve en ny vurdering af ordningen. I retningslinjerne og rammebestemmelserne for de forskellige statsstøtteområder peges der også på andre typer af ordninger, der kan have gavn af evaluering.

4.1 Store støtteordninger, herunder ordninger under den generelle gruppefritagelsesforordning

Som anført i meddelelsen om modernisering af statsstøttepolitikken kan Kommissionen kræve, at de største støtteordninger evalueres, eftersom: (1) sådanne ordninger kan have de største negative konsekvenser for det indre marked, hvis de ikke er udformet korrekt, (2) de kan give de største effektivitetsgevinster på grund af deres store budgetter og (3) store ordninger med mange forskellige typer støttemodtagere kan levere tilstrækkelige data til evalueringen.

Visse støtteordninger vil muligvis stadig ikke skulle evalueres, hvis de trods deres størrelse ikke indebærer specifikke problematiske aspekter (f.eks. rutinesager, hvor et stort antal støttemodtagere hver modtager små støttebeløb, og ordninger, hvor der ikke er risiko for betydelige ændringer, eller hvor der ikke kan opstå alvorlige fordrejninger).

I den nye generelle gruppefritagelsesforordning defineres store støtteordninger ud fra deres budget (et gennemsnitligt årligt budget på over 150 mio. EUR), og for visse kategorier af støtte stilles der krav om evaluering¹³.

For ikke at forsinke disse store ordningers ikrafttræden, men også for at sikre, at de gøres til genstand for en effektiv evaluering, giver den generelle gruppefritagelsesordning mulighed for en undtagelse fra anmeldelseskravet i en periode på højst seks måneder, som

13 Regionalstøtte (undtagen regional driftsstøtte), støtte til SMV'er, støtte til adgang til finansiering for SMV'er, støtte til F&U&I, støtte til miljøbeskyttelse (undtagen støtte i form af nedsættelser af miljøafgifter i henhold til direktiv 2003/96/EF) og støtte til bredbåndsinfrastruktur.

Kommissionen kan forlænge efter godkendelse af evalueringsplanen¹⁴. Evalueringsplanen bør anmeldes hurtigst muligt og senest 20 arbejdsdage efter ordningens ikrafttræden.

I den nye generelle gruppefritagelsesordning tages der også højde for ændringer af eller efterfølgere til disse større ordninger, der skal evalueres, og som skal anmeldes, medmindre ændringerne udelukkende er af formel og administrativ karakter eller foretages inden for rammerne af foranstaltninger, der medfinansieres af EU.

4.2 Nye typer støtteordninger

Definitionen af "nye" kan variere mellem forskellige støtteinstrumenter og medlemsstater. Hvorvidt ordningen er ny, vil i princippet blive vurderet ud fra støtteordningens art eller de markeder, den er rettet imod, f.eks. nye vækstmarkeder, hvor markedet befinder sig på et meget tidligt udviklingsstade. Disse ordninger har potentiale til at præge de pågældende brancher varigt og grundlæggende. Derfor er mulighederne for såvel fordele som fordrejninger særligt store. Det nye aspekt kan f.eks. omfatte indførelse af en ny mekanisme for kapacitetstildeling i energisektoren, støtte til nye typer teknologier eller en ny type støtte til vedvarende energikilder i forbindelse med miljøstøtte. Evalueringen af sådanne nye ordninger hjælper også dem, der er i færd med at udforme nye ordninger, eftersom det giver dem mulighed for at tage højde for den seneste udvikling på markedet.

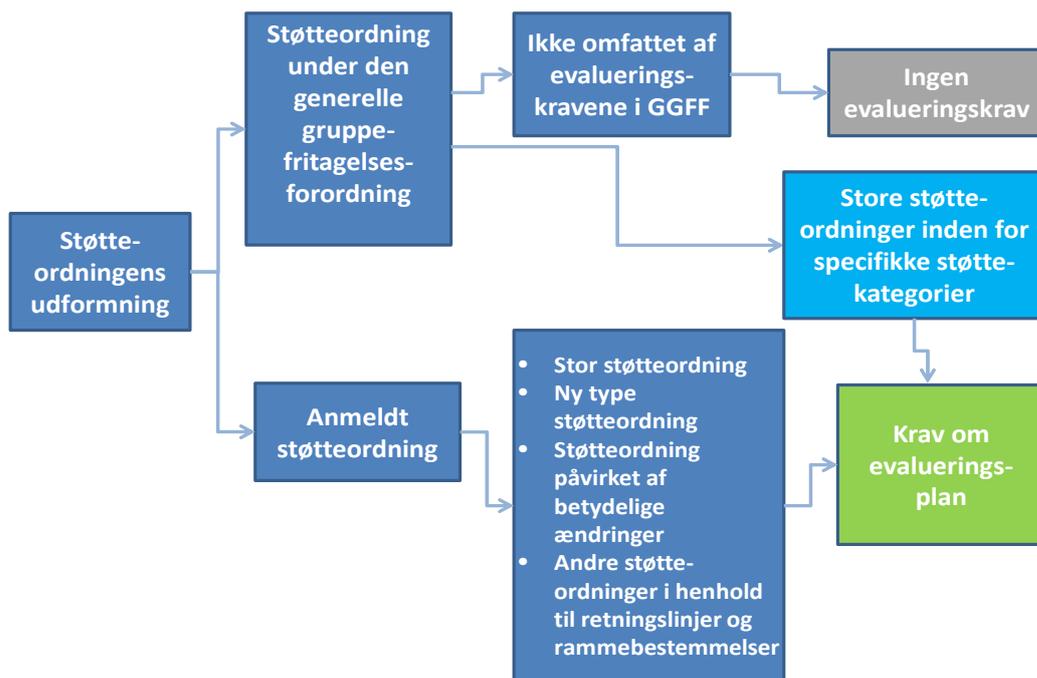
4.3 Støtteordninger, der påvirkes af forventede større ændringer

Muligheden for betydelige (markedsræssige, teknologiske eller reguleringsræssige) ændringer inden for den nærmeste fremtid skal vurderes fra sag til sag. Sådanne betydelige ændringer kan f.eks. omfatte en planlagt revision af gældende lovgivning eller støtte til brancher i hastig udvikling, hvor markeds miljøet og de tilgængelige teknologier udvikler sig hurtigt. Hvis ordningerne ikke er tilpasset virkningerne af disse betydelige ændringer, er der risiko for, at den offentlige finansiering ikke udnyttes effektivt (hvis der f.eks. ydes finansiering til en potentiel "markedsfiasko", som vil forsvinde), eller at der vil opstå betydelige fordrejninger, der påvirker nye markedsdeltagere anderledes end de eksisterende eller skaber ulige betingelser for nye teknologier og eksisterende teknologier. Som tydelige eksempler kan nævnes en revision af en eksisterende reguleringsramme (f.eks. inden for sektoren for elektronisk kommunikation), store udsving i input- eller outputpriser (f.eks. for solpaneler) eller lancering af en ny teknologi på markedet (f.eks. brugen af fjerdegenerations mobilnetværk til bredbåndstjenester), som alle er situationer, hvor der kan være grund til at foretage en evaluering, således at man i fremtidige ordninger kan tage hensyn til udviklingen på markedet.

4.4 Andre støtteordninger

I retningslinjerne og rammebestemmelserne for de forskellige statsstøtteområder identificeres også bestemte støtteordninger, hvor det vil være særligt relevant med en evaluering.

¹⁴ Kommissionen kan også ekstraordinært beslutte, at der på grund af særlige forhold ikke er behov for en evaluering.



Figur 3 — Udvælgelse af støtteordninger til evaluering

Bilag I: Teknisk appendiks om relevante metoder til måling af kausale effekter

En statsstøtteordning kan have virkninger på mange forskellige niveauer. Den forventes normalt at have direkte indvirkning på støttemodtagerne. Det er vigtigt at forstå omfanget af denne virkning, hvis man ønsker at forstå graden af en offentlig foranstaltningens effektivitet og omkostningseffektivitet. Men eftersom støtten er rettet mod virksomheder, der indgår i et samspil på markeder eller i regioner, som konkurrerer om at tiltrække økonomisk aktivitet, har statsstøtte normalt også indirekte virkninger. Disse virkninger kunne f.eks. være følgevirkninger for andre virksomheder (f.eks. positive følgevirkninger af F&U eller fortrængning af investeringer fra konkurrerende virksomheder) eller flytningseffekter (f.eks. flytning af økonomisk aktivitet fra en region til en anden). Disse indirekte virkninger danner grundlag for såvel potentielle skadevirkninger som fordele ved den statslige indgriben i økonomien. Derfor kræver en evaluering af offentlige foranstaltninger også, at man evaluerer omfanget af disse indirekte virkninger.

Måling af de direkte og indirekte virkninger af en politik kræver normalt, at man benytter forskellige værktøjer. I de seneste årtier har man oplevet en betydelig udvikling i metoder og teknikker, der har til formål at vurdere politikkers direkte virkninger for målgrupperne. Disse teknikker beskrives nærmere senere i dette afsnit. Desværre giver disse teknikker kun sjældent mulighed for også at vurdere en støtteordningens indirekte virkninger på virksomheder og regioner. Evalueringen af de indirekte virkninger af statsstøtteordningen kræver normalt andre typer af dokumentation, end der bruges til at vurdere de direkte virkninger på modtagerne, og fortolkningen er normalt i højere grad baseret på økonomiske teorier og modeller. Det er vanskeligere at komme med præcise vejledninger om denne type evaluering, eftersom den skal være skræddersyet til de mulige og forventede positive og negative virkninger af politikken. Derfor skal denne evaluering udføres efter en omhyggelig og stringent analyse af politikens mest sandsynlige indirekte virkninger. På grundlag af denne analyse kan evalueringseksperterne foretage målinger baseret på mikrodata fra ikkestøttemodtagere, navnlig i den samme region, klynge eller branche samt i naboregionerne. Dette kan udgøre kernen i vurderingen af de indirekte virkninger af statsstøtteordningen. Om nødvendigt kan den suppleres med flere makroøkonomiske data og, vigtigst af alt, nøje udvalgte casestudier.

Evalueringen af de direkte effekter er et nødvendigt og vigtigt første skridt. En stringent udført vurdering af støttens indirekte virkninger kan dog fungere som vigtig dokumentation ved vurderingen af de bredere virkninger af ordningen. Hvis manglende supplerende investeringer fra støttemodtagerne generelt er tegn på, at politikken er slået fejl, er selv en positiv virkning ikke tilstrækkelig til, at man kan konkludere, at en politik har opfyldt sine mål. Navnlig gælder, at hvis det viser sig, at støttens direkte virkninger for støttemodtagerne er meget begrænsede eller endda ikkeeksisterende, er det meget sandsynligt, at ordningen må anses for ikke at opfylde sine mål, medmindre der kan fremføres meget overbevisende argumenter for store og gavnlige indirekte virkninger. Det modsatte gælder også: Selvom man ved evalueringen finder, at støtten har positive direkte virkninger, må man stadig stille

spørgsmålet, om der kan være negative indirekte virkninger, der udligner eller endda overstiger disse.

Desuden er det ikke altid nemt at foretage en klar adskillelse mellem direkte og indirekte virkninger. En virksomhed kan have investeret mere (påstået direkte effekt), fordi dens egne investeringer har fortrængt investeringer fra konkurrerende virksomheder (medvirkende indirekte virkning). En virksomhed investerer måske også mere, fordi den forventer afsmittende virkninger og investeringer fra andre virksomheder. Desuden kan det være, at støtten selv eller blot det forhold, at der er ydet støtte, kan have begge virkninger. Ved evalueringen af de direkte virkninger bør man se nærmere på den sandsynlige tilstedeværelse af indirekte virkninger og disses retning og forventede størrelse. Den økonomiske teori, der sammenkæder de indirekte virkninger med støtten, skal angives klart, og supplerende oplysninger, der kan fungere som dokumentation for denne teori, bør udgøre en integrerende del af evalueringen.¹⁵

Kausal inferens

Støttens kausale effekt er forskellen mellem resultaterne med støtten og resultaterne uden støtten. Resultatet med støtten iagttages for de virksomheder, der modtager støtte. Men resultatet uden støtte måles kun for de virksomheder, der ikke modtager støtte. Pr. definition kan vi ikke observere, hvad resultaterne ville være blevet uden støtten for de virksomheder, der modtog støtte. For at vurdere støttens virkning på støttemodtagerne er det derfor nødvendigt at konstruere denne *kontrafaktiske* situation, herunder opstille et rimeligt scenario for, hvad der formentlig ville være sket med støttemodtagerne, hvis de ikke havde modtaget støtten. Dette kræver, at man finder en *kontrolgruppe*, dvs. en gruppe af virksomheder, der i alle henseender minder så meget som muligt om den gruppe af virksomheder, der har modtaget støtte, når der ses bort fra selve støtten.

Kvaliteten af denne kontrolgruppe er afgørende for evalueringens gyldighed. Virksomheder, der modtager støtte, har typisk andre kendetegn end dem, der ikke modtager støtte. De kan f.eks. være aktive i et fattigere område med mindre markedspotentiale, være underlagt kraftigere kreditstramninger, være mere eller mindre effektive, have et projekt eller ej osv. Derfor vil en simplistisk sammenligning af støttemodtagere og ikkestøttemodtagere sandsynligvis i højere grad afspejle denne virkelighed end virkningerne af selve politikken.

Kerneproblemet i udførelsen af en gyldig evaluering er at sørge for, at denne systematiske forskel mellem modtagere og ikke-modtagere af statsstøtte (den såkaldte udvælgelseeffekt) ikke skævvrider resultaterne. I de seneste årtier har man udviklet adskillige pålidelige metoder til at afhjælpe dette problem. Valget af metoden afhænger af den politik, der skal evalueres, og af de tilgængelige oplysninger. Desuden har hver af metoderne deres begrænsninger og er kun gyldige, når visse antagelser holder stik. Troværdigheden af en undersøgelse kan styrkes

15 Selvom dette arbejdsdokument fokuserer på støttens direkte virkninger, udgør det forhold, at støtten kan have indirekte virkninger, visse analytiske udfordringer for vurderingen af de direkte effekter, og man skal være særligt opmærksom på effekten af interaktion på markedet.

ved eksplicit at identificere og diskutere disse begrænsninger. Dette tekniske bilag indeholder oplysninger om de mest relevante metoder, og det fokuserer på de praktiske aspekter og understreger betydningen af en god identificeringsstrategi.¹⁶

A. Randomiserede forsøg

Identificeringen af en korrekt kontrolgruppe er nøglen til at opnå gode (dvs. ikke skæve) skøn over politikkenes virkninger. Den gunstigste situation er, når der ikke er nogen udvælgelseeffekt, fordi støttemodtagerne er udvalgt tilfældigt.¹⁷ Så findes der ingen systemiske forskelle mellem støttemodtagere og ikke-støttemodtagere ud over støtten, og forskelle i resultaterne kan tilskrives politikken.

Men tilfældig udvælgelse af støttemodtagere kritiseres undertiden for at stride imod målsætningen for mange ordninger, som går ud på at udvælge de bedst mulige støttemodtagere på grundlag af objektive kriterier. Under visse omstændigheder kan det dog være muligt at indføre elementer af tilfældighed i udvælgeskriterierne eller i støttemodtagernes incitament til at deltage. Det kan eksempelvis gøres ved at afsætte et fast budget til den pågældende ordning. Hvis de indkomne støtteansøgninger overstiger budgettet, og ansøgernes karakteristika er nogenlunde de samme, kan det være nødvendigt at indføre tilfældighed i behandlingen. En anden mulighed kan bestå i en tilfældig formidling af forskellige mængder information om ordningen til de potentielle støttemodtagere.

Pilotprojekter giver yderligere muligheder for tilfældig tildeling af støtte. Når der er tale om innovative politikker, kan det være tilrådeligt at evaluere et mindre pilotprojekt først. Dette pilotprojekt kan have en mindre størrelse, og det kan være lettere at udvælge støttemodtagerne tilfældigt. En anden mulighed kunne være at indfase ordningen gradvis, således at f.eks. 25 % tilfældigt udvalgte virksomheder er støtteberettigede det første år og derefter henholdsvis 50, 75 og 100 % andet, tredje og fjerde år (eller alternativt at formidle information om ordningen til et større og større publikum). For en ny politik er en sådan indfasningsperiode i mange tilfælde en administrativ nødvendighed.

En sådan tilgang kan være bedre egnet til helt nye ordninger eller en bred vifte af eksisterende ordninger. Ved videreførelse af en eksisterende ordning er det formentlig forholdsvis vanskeligt (direkte eller indirekte) at randomisere udvælgelsen af støtteberettigede virksomheder. Men det betyder ikke, at randomisering ikke kan anvendes til dele af evalueringen. Det er navnlig fortsat muligt tilfældigt at udvælge støttemodtagere til potentielt mere effektive, mere målrettede og/eller mindre fordrejende varianter af ordningen. Er der tale om en tilskudsordning, kan randomiseringen f.eks. baseres på en nyligt indført låneordning i stedet.

16 Dette bilag indeholder en hurtig og ikketechnisk præsentation af de økonometriske metoder til evaluering af politikker. Denne præsentation rummer mange elementer fra Givord (2010), andre udmærkede præsentationer findes i Imbens and Wooldridge (2009) samt Angrist and Pischke (2008).

17 Randomiserede forsøg har f.eks. i årtier været den eneste acceptable metodologi til vurdering af virkningerne af lægemidler og medicinske behandlinger.

B. Kvasi-eksperimentelle metoder

Selvom randomiserede forsøg er den bedst mulige metode til evaluering af en politikks virkninger, er det ikke altid muligt at anvende dem. Der er udviklet andre metoder til at evaluere virkningerne af en politik set ud fra et ex post-perspektiv. Fælles for disse metoder er, at de benytter eksogene variationer af det miljø, virksomhederne opererer i, til at skabe situationer, der ligger meget tæt på eksperimenter (såkaldte naturlige eksperimenter eller kvasi-eksperimenter).

Generelt er det en udfordring for ex post-vurderinger at identificere naturlige eksperimenter eller kvasi-eksperimenter. Men en grundig analyse af politikken udformning kan føre til en analyse af, hvorvidt der forefindes tilstrækkelig eksogen variation. Om nødvendigt kan den oprindelige opsætning indstilles, så man indfører flere elementer, der kan give mulighed for at identificere politikken virkninger.

Kontrol af observerede forskelles indflydelse

Som forklaret ovenfor, er der normalt betydelige forskelle mellem støttemodtagere og ikke-støttemodtagere. Derfor er det nødvendigt at tage højde for disse forskelle, når man sammenligner resultaterne mellem de to grupper, dvs. mellem støttemodtagere og ikke-støttemodtagere.

Mange af forskellene mellem de to grupper kan normalt observeres. Den mest almindelige måde at tage højde for disse forskelle er at anvende lineær regression. Ved lineær regression søger man at tage højde for de observerede kendetegns indflydelse på resultaterne. Man antager, at der findes et lineært forhold mellem resultatet, f.eks. investeringen i F&U, og andre kendetegn ved virksomheden, f.eks. sektoren, alder, størrelse osv., herunder tildelingen af støtte. Lineær regression kan ses som en lineær approksimering af mere komplicerede sammenhænge.¹⁸ Lineære regressioner er almindeligt anvendte teknikker, der kan bruges i mange forskellige evalueringssammenhænge.

Et alternativ til lineær regression er brugen af matchingteknikker. Formålet med matchingteknikker er at matche den enkelte støttemodtager med en anden virksomhed, der "ligner meget", men ikke har modtaget støtte. De iagttagede kendetegn, der matches, kan være konkrete karakteristika ved virksomheden eller den skønnede sandsynlighed for at modtage støtte (*propensity score matching*). Matching kan være en nyttig måde, hvorpå der kan tages højde for observerbare kendetegn inden for rammerne af en valid empirisk strategi.

Både simpel lineær regression og matching har dog en række iboende begrænsninger. Begge er kun gyldige under forudsætning af den såkaldte betingede uafhængighed. Denne forudsætning indebærer, at når der er taget højde for effekten af de observerbare kendetegn, er resultatet uafhængigt af de observerbare kendetegn. I praksis kræver det normalt, at alle variabler, der påvirker såvel resultatet som udvælgelsen, kan observeres (og tages i

¹⁸ Desuden er det muligt at lade kendetegn interagere (f.eks. omsætning og sektor) og at indføre funktioner af disse kendetegn (f.eks. kvadrater af variabler).

betragtning i den korrekte funktionelle form). Hvis det ikke er tilfældet, vil alene den omstændighed, at en virksomhed deltager, afspejle visse (ikkeobserverede) kendetegn ved virksomheden, der også påvirker dens resultater. Hverken lineær regression eller matching vil give en gyldig evaluering. Hvis en virksomhed f.eks. har et "lovende projekt", påvirker det både sandsynligheden for, at den vil ansøge om (og få) støtte, og sandsynligheden for, at det vil lykkes virksomheden at skabe en god forretning. Hvis man ikke tager hensyn hertil, kan det skævvride resultaterne.

Især i forbindelse med matching vil sammenligningen mellem en støttemodtager og dens matchede "tvilling" uden støtte kun gøre det muligt at undgå udvælgelseeffekten, hvis støtten ydes uden nogen sammenhæng med ikkeobserverede variabler, som også påvirker resultatet. I virkeligheden vil denne antagelse sjældent være opfyldt. Det er sjældent muligt at måle alle de variabler, der har indvirkning på, om virksomheden vil ansøge om eller få tildelt støtte. Anvendelsen af matchingteknikker kræver desuden, at de støttemodtagende virksomheder er meget ens med hensyn til deres observerbare kendetegn i forhold til virksomheder, der ikke modtager støtte. Hvis de matchede virksomheder virkelig er ens med hensyn til alle observerbare aspekter, er årsagen til, at nogle virksomheder modtager støtte, mens andre ikke gør, pr. definition ikkeobserveret. Validiteten af en matchingbaseret evaluering eller en enklere klassisk lineær regression kan derfor ikke alene begrundes med, at der foreligger et meget fuldstændigt datasæt med mange observerede kendetegn.

Tværtimod beror argumentet for at bruge matching eller den simple lineære regression til evaluering på, at disse ikkeobserverede årsager til støtteberettigelse eller støttetildeling ikke har nogen direkte eller indirekte indflydelse på resultatet (når der er taget højde for observerbare elementer). For at en evaluering udelukkende baseret på matching eller lineær regression skal være gyldig, må der være sikkerhed for, at den gruppe af virksomheder, der ikke har modtaget støtte, er blevet fastlagt eksogent. Det betyder, at når der er taget højde for de observerbare elementer, må der ikke være nogen ikkeobserverede årsager til støtteberettigelse eller støttetildeling, der også vil påvirke resultaterne direkte eller indirekte. Generelt vil dette sidste kriterium ikke være opfyldt, hvis man matcher virksomheder med samme grad af støtteberettigelse. Hvis alle virksomheder f.eks. er støtteberettigede, vil virksomheder, der modtager investeringsstøtte, med større sandsynlighed have et projekt end virksomheder, der ikke modtager støtte (da disse ellers ville have søgt om og fået tildelt støtte). Normalt er der større sandsynlighed for, at virksomheder med et projekt vil øge deres omsætning eller medarbejderantal, men det hænger ikke sammen med støtten, og de to aspekter kan ikke adskilles gennem en matching ud fra de observerbare elementer (medmindre vi måler eksistensen af et sammenligneligt investeringsprojekt).

I mange situationer vil formodningen om betinget uafhængighed ikke holde stik. Derfor kan det være nødvendigt at anvende andre teknikker end simpel lineær regression eller matching for at tage højde for en ikkeobserveret udvælgelse.

I resten af dette afsnit præsenteres de mest almindelige metodologier, der anvendes til at foretage en nøjere vurdering af en politiks effekt i denne sammenhæng, nemlig Differences-

in-Differences, Regression Discontinuity Design (RDD) og instrumentvariabelmetoden (instrumental variables, IV). Validiteten af disse metodologier er baseret på forskellige antagelser, og det bedste valg vil normalt afhænge af baggrunden for den politik, der skal evalueres, og adgangen til data. I denne præsentation beskriver vi fordele og svagheder ved de enkelte teknikker. Med de ovenfor beskrevne randomiserede kontrollerede forsøg som en klar undtagelse er ingen af disse teknikker bedre end de andre på alle punkter. Valget af en bestemt teknik bør baseres på en grundig analyse af baggrunden for foranstaltningen og de tilgængelige data.

Det er i denne forbindelse værd at understrege, at det ikke er valget af en specifik økonometrisk teknik, der er afgørende for klarlægningen af virkningerne af en bestemt politik – derimod har kontrolgruppens eksogenitet og dermed kvaliteten af den kontrafaktiske situation helt afgørende betydning. Evalueringens kvalitet vil derfor i høj grad afhænge af, hvor stor sikkerhed der er for kontrolgruppens eksogenitet. I de tilfælde, hvor der fortsat kan være residuelle skævheder, er det vigtigt at diskutere disse skævheder grundigt, herunder deres kilder samt retning og den indvirkning, de kan antages at få på resultaterne.

a. Difference-in-difference

Rationale og identifikation

Som beskrevet tidligere vil en simpel sammenligning mellem støttemodtagere på den ene side og en nøje udvalgt gruppe ikkemodtagere på den anden side næppe føre til en gyldig evaluering. Dette skyldes, at det ikke er muligt at udelukke forekomsten af uobserverede forskelle mellem de to grupper, som også uden statsstøtten fører til vedvarende forskelle i deres resultater. På samme måde vil en simpel sammenligning af støttemodtagernes resultater før og efter støtten formentlig også føre til en upræcis evaluering. En sådan sammenligning gør det vanskeligt at skelne mellem støttens virkninger og virkningerne af andre faktorer, som også påvirker de to gruppers resultater, f.eks. den generelle økonomiske udvikling, ændringer i lovgivningen eller øgede lønomkostninger.

Ved at kombinere de to tilgange bliver det imidlertid muligt at måle støttens kausale effekt, og det er, hvad den såkaldte difference-in-difference-metode drejer sig om. Denne metode går ud på at betragte forskellen i virksomhedernes resultater over tid. Allerede eksisterende forskelle kan dermed tilskrives andre faktorer end statsstøtte. Kun en ændring i disse forskelle (difference-in-difference) kan henføres til statsstøtten. Metoden sammenligner med andre ord forskellen i støttemodtagernes og kontrolgruppens resultater *før* og *efter* statsstøtten, hvorved ændringer heri kan tilskrives støtten. Metoden virker, hvis såvel støttemodtagerne som kontrolgruppen over tid påvirkes af andre faktorer, som påvirker deres resultater på samme måde. Det giver mulighed for at konkludere, at støtten er den eneste relevante faktor, der kan forklare den observerede ændring i støttemodtagernes resultater i forhold til kontrolgruppen.

Det er en afgørende forudsætning, at forskellene mellem støttemodtagerne og kontrolgruppen er stabile over tid, og at begge grupper i den pågældende periode påvirkes ens af de samme chokpåvirkninger. Denne forudsætning holder imidlertid ikke nødvendigvis stik i praksis.

Hvis støttemodtagergruppen f.eks. består af mere sårbare virksomheder, vil de formentlig blive mere påvirket af økonomisk nedgang og det generelle erhvervsklima. I så fald skal kontrolgruppen også udgøres af sårbare virksomheder. Overordnet set er valget af kontrolgruppe afgørende for metodens gyldighed. Målingen af en politiks effekt kan ikke baseres på difference-in-difference-metoden i sig selv, for den er blot en teknik – det afgørende er derimod valget af kontrolgruppe.

Hvis de virksomheder, der ikke modtager støtten, selv har valgt ikke at ansøge om den, kræver kontrolgruppens sammensætning særlig opmærksomhed. Der er formentlig en sammenhæng mellem beslutningen om at ansøge om støtte eller ej og de fordele, støtten ventes at kunne give. Der er derfor grund til at tro, at de virksomheder, der ikke søger om støtte, forventer resultater (i relation til antal arbejdspladser, produktivitet, salg osv.), som adskiller sig fra støttemodtagernes forventede resultater. Hvis f.eks. alle virksomheder, der ansøger om støtte, modtager en form for støtte, vil de eneste støtteberettigede virksomheder, der ikke ansøger om støtte, være virksomheder, som ikke har et konkret projekt (forudsat at ansøgningsomkostningerne er lave). Sidstnævnte virksomheder vil ikke kun klare sig dårligere i absolutte tal, men vil også være relativt dårligere stillet over tid, eftersom de mere succesfulde virksomheder vil gennemføre projekter og opleve vækst. Udviklingen i antallet af medarbejdere, produktivitet eller salg kan ikke forventes at forblive parallel, og en dobbelt differentiering vil normalt ikke løse problemet.

De virksomheder i kontrolgruppen, der ikke har modtaget støtte, skal derfor udvælges på grundlag af faktorer, som ikke har indvirkning på de målbare resultater. De må ikke frivilligt have valgt ikke at deltage. Det mest overbevisende scenario er, når deres manglende deltagelse skyldes manglende støtteberettigelse, som konsekvens af et naturligt eksperiment. I sådanne tilfælde skyldes den manglende støtteberettigelse næppe uobserverede faktorer, som også vil påvirke resultaterne. Kontrolgrupperne kan f.eks. udgøres af virksomheder, der er etableret i regioner, som ikke længere er berettiget til støtte (i så fald er støtteberettigelsen ikke relateret til virksomhedernes egne resultater, men derimod til en eksogen årsag).

Gennemførelse

Ud fra et teknisk synspunkt kan difference-in-difference-metoderne enten gennemføres inden for rammerne af en lineær regressionsmodel eller på grundlag af matching. I det første tilfælde udvælges kontrolgruppen uafhængigt af de observerede karakteristika og er dermed overordnet set sammenlignelig med den samlede støttemodtagergruppe. Derefter tages der højde for de observerede forskelle i en klassisk lineær regressionsmodel. I det andet tilfælde udgøres kontrolgruppen af virksomheder, der er individuelt sammenlignelige med hver af virksomhederne i støttemodtagergruppen ud fra de observerbare faktorer. Den enkelte virksomheds resultat sammenlignes med resultatet for de(n) mest sammenlignelige virksomhed(er), og resultaterne lægges sammen. De to metoder repræsenterer to forskellige tilgange til observerbare forskelle, men der er ingen grundlæggende forskel mellem dem i relation til måling af politikens kausale effekt.

Afhængigt af omstændighederne kan det være hensigtsmæssigt at sammenligne forskellene i støttemodtagernes og kontrolgruppens resultater før støtten. Hvis resultaterne allerede er systematisk divergerende før støtten, skyldes det formentlig forskelle mellem kontrolgruppen og støttemodtagergruppen, som ikke har relation til støtten, og derfor vil metoden ikke give en gyldig evaluering af støttens kausale effekt. Men selv om det ikke åbner mulighed for en stringent test af antagelsens gyldighed (en sådan findes ikke), kan det give et nyttigt fingerpeg.

Der er også yderligere metoder og pålidelighedsundersøgelser, der kan benyttes, når der er flere potentielle kontrolgrupper, som umiddelbart forekommer brugbare. Den første og mest naturlige pålidelighedsundersøgelse går ud på at integrere en række difference-in-difference-estimerer og efterfølgende sammenligne resultaterne heraf. Det er ligeledes muligt at benytte de forskellige kontrolgrupper til at nå frem til et mere pålideligt estimat. Et eksempel herpå kunne være en ordning målrettet SMV'er i en bestemt region. De to potentielle kontrolgrupper udgøres af ikke-SMV'er i regionen og SMV'er i en naboregion. Ingen af disse virksomheder har frivilligt valgt ikke at ansøge om støtte, de var ganske enkelt ikke berettiget hertil. Ingen af disse kontrolgrupper fungerer dog perfekt: de større virksomheder i samme region påvirkes formentlig anderledes af den generelle økonomiske udvikling, mens SMV'erne i en naboregion kan opleve forskellige regionale udsving. Frem for at vælge mellem de to mulige difference-in-difference-estimerer er det muligt at kombinere dem og indføre et tredje estimat (difference-in-difference-in-difference, DDD). Denne metode tager udgangspunkt i den klassiske difference-in-difference-model for SMV'er og ikke-SMV'er i den pågældende region fratrukket difference-in-difference i naboregionen med det formål at mindske resultatforskellene mellem SMV'er og ikke-SMV'er i støtteregeion¹⁹. Alternativt kan man søge at konstruere en syntetisk kontrolgruppe bestående af SMV'er fra flere naboregioner og ikke-SMV'er fra samme region for derigennem bedre at reproducere støttemodtagernes resultater før støtten (se Abadie, Diamond & Hainmuller, 2010).

Inferens

Ud over nøje at overveje valg af kontrolgruppe og dens sammensætning bør man også være særligt opmærksom på spørgsmålet om inferens. I denne sammenhæng drejer begrebet inferens sig om, hvorvidt den estimerede effekt virkelig er signifikant. Statistisk signifikans er ikke det samme som økonomisk signifikans. Økonomisk signifikans relaterer til omfanget af den estimerede effekt i forhold til andre parametre, der er relevante ud fra en økonomiteoretisk synsvinkel. Som beskrevet tidligere spiller den økonomiske signifikans en afgørende rolle. Denne diskussion er dog i princippet kun relevant, hvis effekter rent statistisk kan estimeres ret nøjagtigt, dvs. når det kan udelukkes, at der slet ikke er nogen effekt.

19 Et eksempel kunne være en regional støtteordning for SMV'er med henblik på jobskabelse. Ved ordningens udløb viser det sig, at regionens SMV'er har skabt 20 % flere arbejdspladser end de store virksomheder i regionen. Hvis SMV'erne i en sammenlignelig naboregion (hvor der ikke blev ydet støtte) også havde skabt flere arbejdspladser end de store virksomheder (f.eks. 15 % flere), kan støttens virkning sættes til omkring 5 %.

Der er grund til at antage, at ved standardhypoteser (f.eks. om homoskedasticitet eller ingen autokorrelation) vil simpel inferens sandsynligvis føre til en overvurdering af den statistiske signifikans.²⁰

Det første problem vedrører dataclustering. Hvis både kontrolgruppen og gruppen af støttemodtagere er meget homogene (selvom de adskiller sig fra hinanden), vil virksomhederne i begge grupper formentlig blive påvirket af de samme afvigelser fra gennemsnittet (udsving). Det betyder statistisk set, at fejlleddet har en fælles komponent. Hvis den fælles komponent udviser stor varians i forhold til variationen i de resultater, der observeres for de enkelte virksomheder, vil der være en skævhed i inferensen. Hvis der opereres med to perioder og to grupper, kan det være et alvorligt problem, grænsende til et identificeringsproblem, fordi det bliver umuligt at isolere politikken effekt fra virkningerne af de fælles chokpåvirkninger i hver gruppe. Dette problem behøver ikke at være så alvorligt, hvis grupperne ikke er så homogene. Det er dog altid nødvendigt at tage højde for fælles chokpåvirkninger i homogene undergrupper. Hvis efterspørgslen f.eks. er lokal, vil det normalt være nødvendigt at korrigere for fejlleddets clusterstruktur på det lokale plan. Det samme kan gøre sig gældende for brancher og sektorer.

Det andet problem opstår, når der anvendes paneldata. Fejlleddet i de fleste virksomhedsdata om f.eks. beskæftigelse, produktivitet og investering er normalt autokorreleret. Det betyder, at afvigelser fra gennemsnittet i en periode formentlig også vil gøre sig gældende i den næste periode. Hvis man ignorerer dette problem, kan det føre til, at nøjagtigheden af den estimerede effekt overvurderes, og at hypotesen, om at politikken ikke har haft nogen effekt, afvises hyppigere, end det er nødvendigt. Det kan udvikle sig til et alvorligt problem, jf. Bertrand, Duflo & Mullainathan (2004).

b. Instrumentvariabler

Rationale og identifikation

Instrumentvariabler (*instrumental variables*, IV) er en klassisk metode til at tage højde for, at forklarende variabler er endogene. I en lineær regressionsmodel kan støtten betragtes som en endogen forklarende variabel for en virksomheds resultater, og det er derfor naturligt at anvende instrumentvariabler til evaluering af støttens virkning.

En variabel er endogen, når den er korreleret med et uobserveret element, der ligeledes er bestemmende for resultatet. Man kunne f.eks. forsøge at vurdere statsstøttens virkning på beskæftigelsen i virksomhederne ved at regressere jobskabelse på programdeltagelse og øvrige observerbare faktorer. Lad os forestille os, at støtteprogrammet er målrettet virksomheder, der underpræsterer, og som formentlig står over for vanskelige lokale markedsforhold. Markedsforholdene kan ikke observeres af evaluatoren og kan dermed ikke kontrolleres direkte. Men hvis denne variabel ignoreres, vil evaluatoren formentlig

²⁰ Det er et problem, der især er blevet fremdraget i forbindelse med difference-in-difference-metoden, men samme problemer kan opstå ved de andre metoder, der er beskrevet i dette arbejdsdokument.

undervurdere støttens virkning på grund af programdeltagelsens endogene karakter. Uanset om en virksomhed står over for favorable eller vanskelige markedsforhold, vil det have indflydelse på såvel programdeltagelsen som jobskabelsen, eftersom programdeltagelsen korreleres med det fejlede, der forklarer beskæftigelsen. Markedsforholdenes indvirkning på programdeltagelsen gør det umuligt at henføre hele korrelationen mellem programdeltagelse og beskæftigelse til støttens kausale effekt.

Der findes imidlertid også andre faktorer, der kan forklare programdeltagelse, men ikke jobskabelse. Se f.eks. Criscuolo et al (2012), hvoraf det fremgår, at den geografiske lokalisering kan have betydning for det samlede beløb, der er til rådighed for programmet i regionen. Desuden ændrer listen over regioner, der er omfattet af programmet, sig over tid. Hvis programmets budget for en given region ændrer sig over tid af eksterne årsager (f.eks. hvis EU's gennemsnitlige BNP pr. indbygger falder), har det betydning for programdeltagelsen, men ikke for virksomhedens lokale markedsforhold. Ændringer i beskæftigelsen, der er relateret til den eksogene ændring i programdækningen, beror ikke på de lokale markedsforhold. Ved at fokusere på denne "del" af programdeltagelsesvariablen er det muligt at isolere deltagelsens egentlige virkning på beskæftigelsen i en virksomhed, uden at det berører de lokale markedsforhold. Dette er logikken bag instrumentvariabler.

Ved evaluering af statsstøtte er en instrumentvariabel en variabel, der kan forklare modtagelsen af støtte, men som ikke har direkte indvirkning på de andre uobserverede faktorer, der er bestemmende for det resultat, det drejer sig om at måle. Instrumentvariabler gør det således muligt at fokusere på deltagelsen i en ordning uden interferens fra udvælgelsesvirkningerne. Logikken bag en instrumentvariabel kan illustreres på følgende måde.²¹ På første trin regresseres programdeltagelsen på alle eksogene variabler, herunder instrumentvariabler. På næste trin erstattes deltagelsesvariablen (den variabel, der angiver, om der er ydet støtte) med deltagelsen som forudsagt på trin 1. Den forventede deltagelse er ikke korreleret med de uobserverede elementer, der også er bestemmende for udfaldet.

Problemer med svage instrumenter

En instrumentvariabel er en variabel, der kan forklare modtagelsen af støtte, men som ikke har direkte indvirkning på de andre uobserverede elementer, der er bestemmende for det resultat, der skal måles. Denne enkle og klassiske definition skjuler imidlertid en række praktiske problemer. Der findes tests, der kan anvendes til at kontrollere sammenhængen mellem instrumenter, når der anvendes flere instrumenter, end hvad der er strengt nødvendigt for at identificere en model. Der findes dog ingen tests af instrumenternes gyldighed. Hovedformålet med en undersøgelse, der anvender instrumentvariabler, er generelt at forklare, hvorfor det enkelte instrument kan formodes at være ukorreleret med de uobserverede bestemmende elementer i relation til virksomhedens resultater, hvad enten der er tale om beskæftigelse, produktivitet, salg, investering osv. Sådanne forklaringer, som både er baseret på økonomiske argumenter og faktuelle elementer, er nødvendige for at evaluere

21 På grund af inferens gennemføres 2SLS-modellen i praksis normalt i ét trin.

vurderingens gyldighed. De er imidlertid ikke tilstrækkelige, især ikke hvis der anvendes flere instrumenter.

Diskussionen vedrørende kvaliteten af instrumentvariabler bør omfatte problemet med svage instrumenter, dvs. instrumenter, der er svagt korreleret med resultatvariablen. Når instrumentvariablerne er svagt korreleret med den endogene variabel, vil det sandsynligvis resultere i en upræcis evaluering. I denne situation kunne man fristes til at tilføje yderligere instrumentvariabler. Det er velkendt, at anvendelsen af et tilstrækkeligt stort antal instrumentvariabler kan bidrage til at genetablere den oprindelige variabel og dermed opnå statistisk signifikante resultater. Samtidig vil 2SLS-modellen (*two stage least squares*) have stadig større tendens til at give de samme skævheder som OLS-estimatet (*ordinary least squares*).²² Det er noget, der må tages højde for i enhver evaluering, der anvender IV-metoden, specielt med hensyn til troværdigheden af ikke alene den individuelle eksogenitet, men også den fælles eksogenitet.

Et særlig situation opstår, når den endogene variabel antages at være autokorreleret. Hvis kilden til denne endogenitet kan antages at være helt samtidig, kan tidligere værdier anvendes som instrumentvariabler. Men det vil i så fald være nødvendigt at se nærmere på denne metodes gyldighed. Hvis de forklarende variabler f.eks. er autokorreleret, vil det også være tilfældet med de målte resultater. De tidsforskudte variabler vil så også være endogene. Mere generelt gælder, at eksogenitetsantagelserne risikerer at slå fejl, hvis de forklarende variabler i stor udstrækning er autokorrelerede. Hvis autokorrelationen er lav, kan en løsning være at anvende flere tidsforskydninger (og evt. fremtidige værdier). Som nævnt tidligere kan anvendelsen af mange svage instrumenter imidlertid føre til problemer. Alt i alt kan instrumenter baseret på tidligere værdier udgøre en gyldig strategi, men bør anvendes med forsigtighed.

For at undgå de problemer, der er beskrevet tidligere, er det generelt yderst tilrådeligt kun at anvende et begrænset antal overbevisende instrumenter. Men det skal i så fald også påvises, at disse instrumenter er egnede til at forudsige endogene forklarende variabler.²³

Varianter af totrinsmodeller: Heckmans udvælgelsesmodel (1979)

Når den endogene variabel er en uafhængig variabel (dummy for deltagelse), kan den første regression i 2SLS-modellen ses som en lineær sandsynlighedsmodel for sandsynligheden for deltagelse. Denne lineære sandsynlighedsmodel er en lineær approksimering. I nogle tilfælde

22 En meget interessant praktisk diskussion vedrørende skævheder som følge af svage instrumenter findes i Bound, Jaeger & Bakers (1995) diskussion af statistiske skævheder i Angrist & Krueger (1991). Desuden kan estimater af instrumentvariabler være misvisende ved finit afstand. Selv hvor der er tilstrækkeligt store mængder datasæt til opnåelse af statistisk signifikans kan der derfor opstå væsentlige ikkeasymptotiske skævheder.

23 Dette kan illustreres med Fishers signifikanstest på regressionsmodellens første trin. Jo højere denne Fisher er, desto mindre sandsynligt er det, at instrumenterne er svage. Stock, Wright & Yogo (2002) foreslår en formel test. Et eksempel herpå kan være, at et instrument ifølge Fisher-testen på regressionsmodellens første trin skal være højere end 10.

kan denne sandsynlighed dog være lav, også selv om man kun ser på støtteberettigede virksomheder. I så fald er den lineære approksimering formentlig for bred til effektivt at beskrive denne sandsynlighed og fokusere på fordelingen, som netop er det interessante her. Dette problem kan løses på flere måder. De drejer sig alle om at erstatte den lineære sandsynlighedsmodel med en ikke-lineær funktion.²⁴

En klassisk metode er at bruge en udvælgelsesmodel til at løse dette evalueringsproblem (Heckman, 1979). Ved denne metode behandles udvælgelseseffekten som et problem med en udeladt variabel i den lineære regression af resultatet på de observerbare elementer og deltagelsen.²⁵ Der findes flere varianter af denne metodologi, f.eks. en estimering af hele modellen på basis af den maksimale sandsynlighed eller en instrumentering af støtten med den forventede værdi i udvælgelsesligningen.²⁶

Det er imidlertid meget vigtigt nøje at overveje identificeringen og ikke mindst valget af variabler. Det giver ikke noget tilfredsstillende resultat, hvis man anvender de samme variabler i begge trin af estimatet, selv om resultaterne er tilstrækkeligt nøjagtige.¹

Den eneste pålidelige udvælgelsesmodel opnås ved at bruge en såkaldt udelukkelsesvariabel. En udelukkelsesvariabel er en variabel, der forklarer udvælgelsen af støttemodtageren, men ikke resultatet. Det er ikke tilstrækkeligt at fjerne en variabel fra hovedligningen for derefter at lade den indgå blandt de forklarende variabler i udvælgelsen. Udelukkelsesvariablen skal derimod forklare udvælgelsen, men ikke indvirke på det resultat, der skal forklares. Denne variabel minder på mange punkter meget om en gyldig instrumentvariabel. Valget af den må ikke ske ud fra bekvemmelighedshensyn, men bør være baseret på økonomisk teori, en institutionel struktur og/eller erfaring.

c. RDD-metoden (Regression Discontinuity Design)

24 Dette afsnit giver en meget kort beskrivelse af udvælgelsesmodellerne. For en mere detaljeret beskrivelse se Heckmans (1979) skelsættende artikel og navnlig Wooldridge (2002), kapitel 17.

25 Denne udeladte variabel repræsenterer forskellen i den betingede resultatforventning for den udvalgte gruppe (her støttemodtagerne). Ved visse processer for udvælgelse af støtteberettigede (f. eks. en probit- eller logitmodel) kan denne forskel udledes formelt (*inverse Mills ratio*) som funktion af udvælgelsesparametrene. Politikens effekt kan måles ved at tilføje den udeladte variabel i regressionen. Udvalgsparametrene kendes ikke, men det er muligt at rekonstruere sammenhængende parametre i en førstetrinsestimering af udvælgelsesproceduren. Det fører til den estimeringsmetode, der ofte beteges Heckit-metoden. Den indebærer først og fremmest en rekonstruktion af de parametre, der er interessante for udvælgelsen af støttemodtagere med en probit- eller logitmodel. Herefter kan der udledes et estimat af politikens effekt ved at indregne den estimerede *inverse Mills ratio* i den lineære regression. Statistikprogrampakker har normalt en funktion til denne Heckman-estimering.

26 En mere detaljeret beskrivelse af metoderne kan findes i Wooldridge (2002), kapitel 17.

27 Når udvælgelsesligningen er ikke-lineær, er *inverse Mills ratio* ikke kollinear med de øvrige forklarende variabler, heller ikke selv om den første ligning kun omfatter en undergruppe af disse forklarende variabler. I teorien er modellen dermed allerede identificeret. I så fald har *inverse Mills ratio* ikke tilstrækkelig variation, hvilket fører til meget upræcise estimeringer. Især med store datamængder kan estimeringen dog give signifikante resultater. Når alle variabler i udvælgelsesmodellen også indgår i hovedligningen, kan modellen kun identificeres ved hjælp af de ikke-lineære elementer af en specifik parameterform.

RDD (Regression Discontinuity Design) er det senest udviklede evalueringsredskab.²⁸ Dette redskab har igennem det sidste årti vundet stor udbredelse i akademiske kredse, navnlig på grund af dets enkelhed. Metoden udnytter forekomsten af en variabel, som har en diskontinuerlig indvirkning på sandsynligheden for, at en politik påvirker en virksomhed. For statsstøtteordningers vedkommende kan der være tale om flere forskellige typer diskontinuitet. Den første type er geografiske grænser - ordningernes regler for støtteberettigelse er knyttet til præcise administrative områder såsom kommuner, NUTS-regioner osv. Den anden drejer sig om de krav, der stilles til de støttemodtagende virksomheder, navnlig med hensyn til alder og størrelse.

Dette kan beskrives med følgende eksempel: Man kan forestille sig en støtteordning, hvor de projekter, der indsendes af virksomhederne, tildeles point (ud af 100), og hvor de virksomheder, der opnår mindst 70 point, vil modtage støtte, mens de øvrige virksomheder ikke får nogen støtte. En virksomhed, der får 71 point, har et marginalt bedre projekt end en virksomhed, der får 69 point. Konsekvensen af denne marginale forskel er imidlertid væsentlig: den ene virksomhed modtager støtte, mens den anden virksomhed ikke får nogen støtte overhovedet. En sammenligning af udviklingen i de to virksomheders resultater giver dermed en meget tydelig indikation af støttens kausale effekt.

RDD kræver formelt, at sandsynligheden for at modtage støtte er diskontinuerlig, mens alle de øvrige variable er kontinuerlige.²⁹ Den tekniske gennemførelse minder meget om metoden med instrumentvariablerne, med en dummyvariabel for, hvornår der nås op over den fastsatte tærskel, som instrument. Der er imidlertid to væsentlige forskelle. For det første bygger RDD på svagere antagelser. Instrumentet behøver dermed ikke være uafhængigt på forhånd. Hvis der f.eks. gives point, vil virksomheder med gode projekter nok være mere tilbøjelige til at ansøge end virksomheder med dårlige projekter. Det eneste krav er, at sandsynligheden for, at en virksomhed, der ligger omkring tærsklen, ansøger, ikke må være diskontinuerlig. For det andet er estimaterne udelukkende baseret på virksomheder, som ligger tæt på begge sider af den fastsatte tærskel. Dermed har de svagere antagelser en pris: Generelt er RDD-metoden mere lokal end instrumentvariablerne. Hvis virkningerne af støtten er anderledes for virksomheder, der befinder sig længere væk fra tærsklen, er RDD ikke den rette metode til evaluering af virkningen for alle støttemodtagere.

Dette fokus på området lige omkring tærsklen kan give problemer, hvis der kan forventes væsentlige forskelle længere væk fra tærsklen. Desuden kan virksomheder på den anden side af grænsen blive væsentligt berørt af politikken. Det er f.eks. tilfældet, når der er tale om

28 En mere formel og fuldstændig beskrivelse af RDD kan findes i Imbens et Lemieux (2008).

29 Der findes formelt to forskellige former for RDD'er: henholdsvis et skarpt RD-design og et vagt (fuzzy) RD-design. I det skarpe RD-design, som er det, der implicit beskrives her, undersøger man alle virksomheder over en bestemt tærskel. I det vage (fuzzy) RD-design er diskontinuiteten mindre drastisk: der er en diskontinuitet i sandsynligheden for deltagelse, men den ændrer sig ikke fra 0 til 1. Når det drejer sig om statsstøtteordninger baseret på støtteberettigelseskriterier, kan RD-designet kun betegnes som skarpt, hvis det er støtteberettigelsen, der undersøges. Hvis det derimod er støttemodtagelse, er der tale om et vagt (fuzzy) design. Hvis støttetildeling foregår på basis af pointmekanisme, undersøges kun de virksomheder, der ansøger om støtte, og her er RD-designet skarpt.

væsentlige flytningseffekter. Brugen af RDD ved geografiske grænser er derfor ikke nogen god empirisk strategi. RDD's styrke ligger i fokusering på en smal båndbredde lige omkring diskontinuitet. Hvis båndbredden er stor, kan virkningen af de øvrige karakteristika ikke forventes at være konstante. Dette problem kan normalt ikke løses ved at kontrollere for de observerbare elementer, hvilket forudsætter en bestemt funktionel form.

En grafisk inspektion af dataene kan give nyttige garantier for pålideligheden af de underliggende antagelser ved RDD-metoden. Der er især tre faktorer, det er vigtigt at kontrollere. Den første er, om der reelt er diskontinuitet i ydelsen af støtte på tærsklen. Den anden er, om de resultater, der skal måles, er diskontinuerlige på samme tidspunkt, og om der ikke er lignende diskontinuitet andre steder. Den tredje er, om der ikke er diskontinuitet i andre parametre, som er korreleret med resultatet, herunder tilbøjeligheden til at søge om støtte.

Endelig kan der bevidst skabes diskontinuitet for at muliggøre en evaluering af ordningen. Især kan en gradvis indfasning af en politik anvendes til at skabe diskontinuitet og dermed bidrage til at identificere politikens effekt.

C. Strukturel estimering

I nogle tilfælde er det muligt at gå et skridt videre og sammenholde dataene med en teoretisk model, f.eks. en model for virksomhedernes investeringer, for at kunne udlede de vigtigste nøgleparametre. Denne tilgang adskiller sig kvalitativt fra de metoder, der er nævnt tidligere. Strukturel estimering bygger på en fuldt specificeret teoretisk model for virksomhedernes adfærd. Estimeringen gør det således muligt at rekonstruere de parametre, der er bestemmende for virksomhedernes adfærd. Derefter kan der foretages en præcis vurdering af de faktorer, der er afgørende for de enkelte virksomheders adfærd, samt simuleringer af effekten af andre virkemidler. Strukturel estimering stiller imidlertid generelt større krav til ressourcer, data og antagelser.

Det er umuligt at give nøjagtige retningslinjer for strukturel estimering, eftersom identifikation, estimering og inferens må udledes fra sag til sag. De ovenstående generelle retningslinjer finder imidlertid også anvendelse her. For det første er det nødvendigt at sikre, at den teoretiske model stemmer overens med de væsentligste stiliserede markedsforhold. For det andet bør spørgsmål omkring uobserverede karakteristika og udvælgelse afklares fuldt ud.

D. Yderligere metodologiske bemærkninger

Heterogene effekter

De foregående afsnit har fokuseret på en estimering af en støtteordnings gennemsnitlige effekter på den målgruppe, den er rettet mod. Som det fremgår af selve ordet "gennemsnitlige", er det klart, at støttens virkninger varierer fra den ene støttemodtager til den anden. Denne heterogenitet har mange årsager og konsekvenser. Den første konsekvens er, at hvis støtten er meget effektiv for nogle virksomheder, men mindre for andre, kan den gennemsnitlige virkning være statistisk insignifikant. Men manglen på statistisk signifikante

virksomheder. Ud fra et politisk synspunkt er en ordnings gennemsnitlige resultater en meget interessant foreløbig indikator. En bedre forståelse af årsagerne til de heterogene aspekter er imidlertid vigtig for at kunne udforme bedre støtteordninger, så der kan fokuseres direkte på de virksomheder, hvor støtten er mest effektiv og mindst fordrejende.

Støttens virkninger bør så vidt muligt vurderes for de forskellige virksomhedstyper, herunder små virksomheder i forhold til store virksomheder, nystartede virksomheder i forhold til etablerede virksomheder, innovative virksomheder, virksomheder med ringe adgang til kredit osv.³⁰

Fordrejninger i forhold til virksomheder, der ikke får støtte

En vurdering af en støtteordnings direkte eller indirekte virkning på dem, der ikke modtager støtte, kan give nyttige oplysninger til evalueringen af ordningen. Statsstøtte kan føre til fordrejninger på markedet som følge af indvirkningen på ikkestøttemodtagere, f. eks. i form af videnafsmitning fra støttemodtagerne eller mindskelse af ikkestøttemodtagernes relative konkurrencedygtighed i forhold til støttemodtagerne.

Virkningerne på ikkestøttemodtagende virksomheder eller lokaliteter kan også påvirke evalueringens gyldighed. En del af en regionalstøttes virkninger kan f.eks. tage form af bevægelser ved den pågældende regions grænser - virksomheder, der historisk set har ligget på den "forkerte" side af grænsen, kan vælge at flytte deres virksomhed til den anden side. Hvis der foretages en RDD ved grænsen, vil den ofte hovedsagelig indfange denne flytningseffekt og dermed risikere at overvurdere politikkenes egentlige samlede virkning. I en sådan situation bør der anvendes en anden empirisk strategi (det kan f.eks. være nyttigt at undersøge evalueringens pålidelighed på bredere regioner).

E. Data

For at kunne foretage en effektevaluering er det afgørende vigtigt at have adgang til de rette mikroøkonomiske data. Der skal være kohærens i dataene for henholdsvis støttemodtagere og ikkestøttemodtagere. De må derfor stamme fra den samme kilde, dog med undtagelse af oplysninger om selve støtten. Disse data bør være tilgængelige på det mest detaljerede niveau, selv om der måske på et senere tidspunkt kan blive behov for en aggregering.

Det er nødvendigt at have data, der åbner mulighed for at udlede resultatindikatorer for såvel støttemodtagergruppen som kontrolgruppen, også på det tidspunkt, hvor resultatet måles. Der er ligeledes behov for så mange data som muligt om de faktorer, der potentielt kan påvirke

30 En anden tilgang kan være systematisk at estimere de forskellige virkninger for virksomheder med forskellige placeringer i den betingede fordeling. En stigende mængde litteratur beskæftiger sig med sådanne fraktile virkninger, herunder Abadie, Angrist & Imbens (2002). Det kan være et meget nyttigt redskab til at forstå denne heterogene karakter. Ud fra et strengt politisk synspunkt er redskabet dog mindre nyttigt, medmindre det er muligt at målrette det direkte mod forskellige virksomheder afhængig af deres placering i den betingede fordeling.

resultaterne og virksomhedernes beslutning om at deltage i statsstøtteprogrammer. Disse data anvendes til at tage højde for forskellene mellem støttemodtagergruppen og kontrolgrupperne. På virksomhedsplan kan der være tale om data om lokalisering, størrelse og demografiske forhold samt de inputs, der anvendes i produktionen.

De mest naturlige datakilder er kilder inden for den offentlige forvaltning, f.eks. skatteregnskaber og nationale undersøgelser. De kan levere oplysninger om virksomhedernes lokalisering og aktiviteter og undertiden også om enkelte virksomhedsenheder. De åbner ofte mulighed for at spore investeringer eller salg samt beregne finansielle nøgletal. Store nationale eller europæiske undersøgelser som f.eks. EU's innovationsundersøgelser er ligeledes interessante. De omfatter et stort og repræsentativt udsnit og giver supplerende oplysninger om specifikke emner. Endelig kan også en sammenholdelse af datasæt om arbejdsgivere/arbejdstagere være en relevant kilde. Disse datasæt kan åbne mulighed for at sammenkæde arbejdskraftoplysninger med bestemte virksomhedsenheder. Sådanne oplysninger er vigtige, hvis man interesserer sig for beskæftigelsens geografiske konstellation.

Ud over resultatindikatorer og oplysninger om støttemodtagerne er der behov for data om selve støtten og støttetildelingsprocessen. Disse oplysninger kommer normalt fra den støttetildelende myndighed. Der er bl.a. brug for oplysninger om støttebeløb og støttetidspunkt. Generelle oplysninger om støttetildelingsprocessen er også særdeles nyttige. Oplysninger om afviste ansøgere er vigtige, især hvis tildelingen af støtte er baseret på en pointmekanisme.³¹

Adgangen til disse fortrolige oplysninger er normalt reguleret. Det er derfor vigtigt at sikre rettidig adgang til disse data for alle parter i den videnskabelige gruppe, der foretager evalueringerne. Desuden tager det normalt også tid at få adgang til kilderne i det offentlige. Det er derfor vigtigt at tage hensyn til datatilgængeligheden ved udformning af evalueringsplanen.

31 Oplysninger om afviste ansøgere er især værdifulde for undersøgelser baseret på RDD-metoden.

F. Eksempler

Eksempel 1 (regionalstøtte): Criscuolo m. fl. (2012)³² har undersøgt den britiske regionalstøtteordning *Regional Selective Assistance* i perioden mellem 1986 og 2004. I denne periode blev der under denne ordning ydet skønsmæssig støtte til virksomheder i ugunstigt stillede områder. Det var den vigtigste erhvervsstøtteordning i Det Forenede Kongerige. Ydelsen af støtte inden for rammerne af ordningen skete i henhold til EU's regionalstøtteretningslinjer og især regionalstøttekortene. Disse kort har udviklet sig over tid. Kriterierne for en regions støtteberettigelse afhænger i det store og hele af regionens relative situation med hensyn til BNP pr. indbygger eller arbejdsløshed. Regionens status kan dermed ændre sig, enten fordi regionen har udviklet sig over tid eller på grund af ændringer i EU's gennemsnitlige BNP pr. indbygger (hvilket f.eks. var tilfældet, da nye medlemsstater tiltrådte EU i 1995). De anvendte indikatorer for støtteberettigelse har ligeledes ændret sig over tid. En del af de ændringer, der sker med hensyn til en virksomheds støtteberettigelse, afhænger derfor ikke af den pågældende virksomheds situation, men derimod af begivenheder uden for Det Forenede Kongerige eller af ændrede administrative bestemmelser. Ved at fokusere på ændringerne med hensyn til støtteberettigelse og vurdere, hvordan de påvirker faktorer som investeringsaktivitet, beskæftigelse og produktivitet, kan Criscuolo m. fl. (2012) på overbevisende måde identificere støttens effekt.³³

Eksempel 2 (erhvervsstøtte): Martini & Bondonio (2012)³⁴ har undersøgt to former for erhvervsstøtte: en investeringsstøtteordning gældende for hele Italien (lov 488), og forskellige SMV-ordninger i regionen Piemonte. Især den første undersøgelse er interessant. Undersøgelsen sammenligner virksomheder, hvis ansøgning om støtte er blevet godkendt (dvs. støttemodtagere), og virksomheder, hvis ansøgning er blevet afvist, fordi midlerne på det afsatte budget var opbrugt. Det er særdeles nyttigt at medtage afviste ansøgere i undersøgelsen for at undgå skævheder i udvælgelsen, som typisk opstår, hvis man udelukkende sammenligner ansøgere med ikkeansøgere. Denne gruppe af virksomheder havde bestået den første kvalitetskontrol, hvilket betyder, at de havde fremlagt et troværdigt investeringsprojekt. De havde ligesom de øvrige støttemodtagere en ambition om at investere i et troværdigt projekt. De fik imidlertid ikke tildelt støtte på grund af budgetmæssige begrænsninger (rationering). Det var muligt at foretage en pålidelig estimering af støttens effekt ved at vurdere resultatforskellene mellem henholdsvis godkendte og afviste ansøgere.

32 Criscuolo, C., R. Martin, H. Overman & J. Van Reenen, 2012. 'The causal effects of an industrial policy,' CEPR Discussion Papers 8818, C.E.P.R. Discussion Papers.

33 Criscuolo et al. (2012) bruger teknisk set en tilgang baseret på instrumentvariabler som omhandlet tidligere i det tekniske tillæg.

34 Rapport til GD REGIO. A. Martini, D. Bondonio: "Counterfactual impact evaluation of cohesion policy: impact and cost effectiveness of investment subsidies in Italy" (2012).

Eksempel 3 (lånegarantier): Lelarge, Sraer & Thesmar (2010) har undersøgt virkningen af en fransk ordning med lånegarantier, SOFARIS-programmet. Denne ordning sikrer långiverne mod risikoen for låntagernes misligholdelse gennem garantier. Låntagerne betaler en forsikringspræmie, men denne præmie er subsidieret. Lelarge, Sraer & Thesmar (2010) giver en detaljeret beskrivelse af udvælgelseeffekten. Det har for det første vist sig, at virksomheder med mere rentable projekter er mere tilbøjelige til at betale det gebyr, der kræves for den pågældende garanti. For det andet er programmets forvaltere mere tilbøjelige til at udvælge samfundsmæssigt ønskværdige projekter, som måske ellers ikke ville have adgang til privat finansiering. Generelt vælger virksomhederne selv, om de vil deltage i programmet, og i tildelingsfasen foretages der en udvælgelse. Disse omstændigheder må antages at ville påvirke resultaterne af simplistiske evalueringer, f.eks. evalueringer baseret på en klassisk lineær regressionsmodel eller på sammenligninger med de mest sammenlignelige virksomheder.³⁵ Programmets faktuelle og institutionelle kontekst kan imidlertid også hjælpe med til at identificere politikens effekt. Programmet blev indført i slutningen af 1980'erne og var oprindeligt begrænset til virksomheder inden for fremstillingsindustri eller erhvervstjenester. I 1995 blev de offentlige bevillinger til programmet øget, og nye sektorer (byggeri, detailhandel, engroshandel, transport, hotel, restaurant og personlige tjenester) blev dermed berettiget til støtte. Lelarge, Sraer & Thesmar (2010) sammenligner de nye støtteberettigede virksomheder med de tidligere for at vurdere programmets indvirkning på forskellige indikatorer såsom gæld, beskæftigelse, kapitalvækst, finansielle udgifter og konkursrisiko. Virksomhederne i de to grupper udviser visse forskelle, men berøres imidlertid af de samme makroøkonomiske udsving, og forskellene mellem dem skulle derfor ikke ændre sig over tid, undtagen i kraft af politikens forventede effekt.³⁶

Eksempel 4 (Creative Credit): Bakhshi m. fl.³⁷ anvender randomiserede kontrollforsøg for at vurdere virkningen af en innovativ erhvervsstøtteordning. Pilotprojektet, som blev indledt i Manchester i 2009, var opbygget således, at tilfældigt udvalgte små og mellemstore virksomheder, der ønskede at investere i kreative projekter såsom udvikling af websider, videoproduktion og kreative markedsføringskampagner, fik udbetalt et tilskud - de såkaldte *creative credits* - for at se, hvilke virkninger det havde på deres innovation. Det skabte helt nye forbindelser mellem SMV'er og kreative virksomheder, og sandsynligheden for, at en virksomhed ville indlede et innovativt projekt med en kreativ virksomhed, som den ikke

35 Dette er et eksempel, hvor matchingteknikker, dvs. nabomatching i forholdet en til en, ikke udgør en bedre metode til løsning af udvælgelsesproblemer end OLS (ordinary least squares). Som beskrevet tidligere i det tekniske tillæg udgør matchingteknikker generelt ikke en metode til løsning af udvælgelsesproblemer, hvor der ikke foreligger naturlige forsøg.

36 Forfatteren anvender i praksis Heckmans udvælgelsesmodel med en udelukkelsesvariabel på virksomhedsniveau samt en klassisk strategi baseret på instrumentvariabler på sektorspecifikt plan. En mere detaljeret beskrivelse af metodologierne kan findes i det tekniske tillæg.

37 Rapporten til Nesta, "Creative Credits, a randomised controlled industrial policy experiment", Bakhshi, H., J. Edwards, S. Roper, J. Scully, D. Shaw, L. Morley & N. Rathbone, juni 2013, er tilgængelig på http://www.nesta.org.uk/sites/default/files/creative_credits.pdf.

tidligere havde samarbejdet med, blev øget med mindst 84 % som følge af disse *creative credits*. Undersøgelsen påviste desuden, at de virksomheder, der fik tilskuddene, på kort sigt oplevede en stigning i deres innovation og salg i de efterfølgende seks måneder. De positive virkninger varede imidlertid ikke ved. Efter 12 måneder var der ikke længere nogen signifikant statistisk forskel mellem de grupper, der havde modtaget tilskuddene, og dem, der ikke havde. Rapporten hævder, at disse resultater ville være forblevet skjult, hvis myndighederne havde anvendt den almindelige vurderingsmetode, og opfordrer derfor til øget anvendelse af randomiserede kontrolforsøg til evaluering af erhvervsfremmepolitikker.

Eksempel 5 (F&U&I-støtte): Einiö (2013) har undersøgt virkningerne for F&U-investeringer, beskæftigelse og produktivitet i perioden 2000-2006 af den F&U-støtte, som blev ydet af den finske innovationsfond Tekes, der står for størstedelen af Finlands F&U-støtte. Undersøgelsen tog udgangspunkt i regionale forskelle i de midler, Tekes potentielt kan bruge til F&U-støtte, og som hænger sammen med større EFRU-finansiering i dele af det nordlige og østlige Finland (mål 1-områder). Disse områder blev oprindeligt fastlagt i forbindelse med forhandlingerne om Finlands tiltrædelse i 1995 og var baseret på princippet om en befolkningstæthed på højst otte personer pr. kvadratkilometer. Sandsynligheden for at modtage støtte var dermed større i mål 1-områder end i andre dele af landet som følge af det relativt forhøjede F&U-støttebudget. Dette førte til regional forskelsbehandling, eftersom en væsentlig større andel af virksomhederne i mål 1-området var omfattet af støtteordningen. Eftersom den regionale tildeling var baseret på et forudbestemt princip om befolkningstæthed fra 1993 (og ikke på de forventede fremtidige F&U-investeringer eller økonomiske resultater), blev der i undersøgelsen korrigeret for befolkningstætheden i 1993, hvilket afklarede spørgsmålene omkring regional udvælgelse. I praksis blev støtteeffekten estimeret på grundlag af instrumentvariabler, hvor indikatoren for mål 1-området anvendtes som instrument for programdeltagelse. Denne tilgang identificerede støttens effekt blandt de virksomheder, der deltog i støtteordningen på grund af den øgede finansiering til mål 1-områder. Gyldigheden heraf blev fastslået ved at påvise, at der forud for programmet ikke var forskel mellem de virksomheder, der deltog i programmet, og kontrolgruppen. Einiö (2013) har konstateret positive virkninger for så vidt angår F&U-investeringer, beskæftigelse og salg for de virksomheder, der har fået tildelt F&U-støtte som følge af den yderligere F&U-støtte i regionen. Det har imidlertid ikke været muligt at påvise øjeblikkelige virkninger på produktiviteten, selv om undersøgelsen peger på langsigtede produktivitetsfordele.

Bilag II: Liste over mulige resultatindikatorer

Det bemærkes, at den vejledende liste nedenfor udelukkende er et eksempel. De resultatindikatorer, der anvendes, bør fastlægges ud fra formålet med støtteordningen og evalueringen.

Støttens direkte effekt på støttemodtagerne

	RESULTATDIMENSION	RESULTATINDIKATORER
Regionalstøtte	Positive virkninger	Private investeringer matcher offentlig støtte Stigning i beskæftigelsen i de støttemodtagende virksomheder
Støtte til forskning, udvikling og innovation	Øget F&U&I-aktivitet	Private investeringer matcher offentlig støtte Øgede F&U&I-udgifter i de støttemodtagende virksomheder Antal nye forskere ansat i de støttemodtagende virksomheder Antal nye registrerede patenter Antal støttemodtagende virksomheder, der markedsfører nye produkter
Miljøstøtte	Positive miljøvirkninger	Reducerede CO ₂ -emissioner i de støttemodtagende virksomheder Øget kapacitet til vedvarende energiproduktion Reduktion af den mængde affald, der deponeres eller forbrændes Antal oprensede forurenende grunde
	Tidlig indførelse af miljøstandarder	Procentdelen af virksomheder, der har opfyldt de nye miljøstandarder mindst X måneder/år, før de træder i kraft [krav om, at de opfyldes mindst 1 år før, og mulighed for større støtte, hvis de er opfyldt tre år inden ikrafttrædelsen]
Energistøtte (infrastruktur)	Reduceret energiforbrug	Antal husstande med lavere energiforbrug Reduktion af det årlige primærenergiforbrug i offentlige bygninger Antal nye energibrugere tilsluttet intelligente net

	Støtte til vedvarende energi	Vedvarende energikilders andel af energiproduktionen
Risikofinansiering	Positive virkninger	Opnået afkast i fonden Mobilisering af private investeringer Antal virksomheder, der har modtaget risikokapital
	Investering i tabere	Ring gennemsnitligt resultat i de virksomheder, der investeres i, på grund af mangelfuld ledelse/utilstrækkelig privat deltagelse
	Utilstrækkelig diversificering	Fondene er for små eller for regionalt orienterede, giver ringe afkastudsigter og kan derfor ikke tiltrække private investorer
Bredbåndsstøtte	Øget bredbåndsdækning	Øget antal husstande med en bredbåndsforbindelse på mindst 30 Mbps Øget antal husstande eller brugere med en bredbåndsforbindelse på mindst 100 Mbps
	Effektivitet	Investeringsomkostninger/støtte pr. tilsluttet husstand (husstande med adgang til bredbånd) Antal husstande, der vælger de nye tjenester
Genopretning og omstrukturering	Positive virkninger	Bevarelse af arbejdspladser i virksomheden og i regionen Ændringer i de støttemodtagende virksomheders markedsandel og produktivitet
Luftfart	Positive virkninger	Antal luftfartsselskaber, der bruger lufthavnen Private investeringer matcher offentlig støtte Stigning i den regionale produktivitet og/eller bruttoværditilvæksten
	Negative virkninger	Duplikering af underskudsgivende infrastruktur eller flyruter Tilbagegang i trafikken i den eksisterende infrastruktur (f.eks. andre lufthavne i området eller andre transportmidler)

Støttens indirekte effekt

	RESULTATDIMENSION	RESULTATINDIKATORER
Mulige positive virkninger	Makroøkonomiske fordele	Stigning i beskæftigelsen Stigning i produktiviteten og/eller bruttoværditilvæksten
	Diversificering af den regionale økonomi	Antal erhvervssektorer (forskellige NACE-koder)
	Øget samarbejde mellem private og offentlige partnere	Antal virksomheder, der samarbejder med forskningsinstitutter
	Positive eksterne/afsmittende virkninger	Antal indirekte støttemodtagere (f.eks. antal tredjeparter med adgang til faciliteten) Ændringer i andre virksomheders eller regioners beskæftigelse eller aktivitet (Luftfart) Antal indbyggere med forbedrede transportforbindelser i området
Mulige negative virkninger for konkurrence og samhandel	Sektorspecifikke skævheder	Ordningen er tværsektoriel, men støtten ydes hovedsagelig til en enkelt industri
	Skævheder i forhold til tabsgivende virksomheder eller virksomheder med lav produktivitet (forebyggelse af lukning)	Andel af virksomheder med henholdsvis høj og lav produktivitet
	Skævheder i forhold til etablerede virksomheder	Andel af henholdsvis etablerede og nystartede virksomheder
	Forstærket markedsposition	Ændring af dominerende virksomheds markedsposition
	Lokaliseringsevner	Flytning fra en fattig region til en mere udviklet region
	Forsyningssikkerhed	Fastlåsning af kulstofintensive energikilder Vurdering af, om risiko for strømsvigt er reel og vil fortsætte Afskærmning af de nationale elmarkeder
	Energiinfrastruktur	Afskærmning af de nationale elmarkeder Styrkelse af en etableret virksomheds markedsposition

	Redning og omstrukturering	<p>Ændringer i andre virksomheders eller regioners beskæftigelse eller aktivitet</p> <p>Ændringer i de støttemodtagende virksomheders markedsandel og produktivitet</p>
	Luftfart	<p>Duplikering af underskudsgivende infrastruktur eller flyruter</p> <p>Tilbagegang i trafikvolumen i den eksisterende infrastruktur (f.eks. andre lufthavne i området eller andre transportmidler)</p>

Bilag III: Ordliste

Udgangsværdi	Indikatorens værdi inden det politiske tiltag
Kontrolgruppe	En kontrafaktisk analyse forudsætter, at man finder frem til den mest sammenlignelige virksomhed eller kontrolgruppe, dvs. en gruppe virksomheder, der minder så meget som muligt om den gruppe virksomheder, der har modtaget støtte – altså bortset fra, at de ikke har modtaget støtte.
Kontrafaktisk scenarie	For at vurdere støttens virkning på støttemodtagerne er det nødvendigt at konstruere et "kontrafaktisk" scenarie, dvs. opstille et rimeligt scenario for, hvad der formentlig ville være sket med støttemodtagerne, hvis de ikke havde modtaget støtten.
Evaluering	En systematisk indsamling og analyse af oplysninger om programmer og projekter, deres formål og gennemførelse. Evalueringen tilvejebringer viden om deres virkning, som kan danne grundlag for en bedømmelse. Evalueringerne har til formål at fremme effektivitet og velfunderede beslutninger om aktuelle og fremtidige programmer.
Effekt	Den ændring, der på troværdig vis kan tilskrives et politisk tiltag. Synonymer kan være "politikens virkning" eller "bidrag til ændring".
Indikator	En variabel, der giver kvantitative eller kvalitative oplysninger om et fænomen. Variablen omfatter normalt en værdi og en måleenhed.
Metode	Metoder er grupper af evalueringsteknikker og -redskaber, som opfylder forskellige formål. De består normalt af procedurer og protokoller, der sikrer systematisering og sammenhæng i måden, hvorpå evalueringerne foretages. Metoderne kan fokusere på indsamling eller analyse af oplysninger og data og kan være kvantitative eller kvalitative. De kan desuden tage sigte på at beskrive, forklare, forudsige eller underbygge et givet tiltag. Valget af metode sker ud fra tiltagets art, de stillede evalueringsspørgsmål og undersøgelsens art – kausal, undersøgende, normativ osv.
Resultat	Den specifikke aspekt af befolkningens velfærd, som det pågældende politiske tiltag tager sigte på at påvirke. Det kan f. eks. dreje sig om mobiliteten i et område eller om kvalifikationer og kompetencer i en given sektor.
Resultatindikator	En indikator, der beskriver et specifikt aspekt af et resultat eller en målbar funktion. Som eksempel på et mobilitetsaspekt kan nævnes den tid, det tager at rejse fra W til Y med en gennemsnitlig hastighed, mens et kompetenceaspekt kan være resultatet af prøver i et givet fag, og en given undersøgelse af et givent emne. Et aspekt, der f.eks. beskriver bankernes kreditrationering, kan være antallet af virksomheder, der ikke kan opnå lån til nogen rente.

Bilag IV: Litteraturliste

Abadie, A., J. Angrist & G. W. Imbens (2002), "Instrumental Variables Estimates of the Effect of Subsidised Training on the Quantiles of Trainee Earnings," *Econometrica*, 70(1), 91–117.

Abadie, A., A. Diamond & J. Hainmueller (2007), "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Programme," *Journal of the American Statistical Association*, June 2010, Vol. 105, No 490.

Angrist, J. & A. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings," *Quarterly Journal of Economics*, 106.

[Angrist, J. & J. Pischke \(2008\), "Mostly Harmless Econometrics: An Empiricist's Companion", Princeton University Press.](#)

Angrist, J. D., & J. Pischke (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics." *Journal of Economic Perspectives*, 24(2): 3-30.

Bakhshi, H., J. Edwards, S. Roper, J. Scully, D. Shaw, L. Morley & N. Rathbone (2013), "Creative credits, a randomized controlled industrial policy experiment", rapporten til Nesta er tilgængelig på http://www.nesta.org.uk/sites/default/files/creative_credits.pdf.

Bertrand, M., E. Duflo & S. Mullainathan (2004), "How much should we trust differences-in-differences estimates ?," *The Quarterly Journal of Economics*, 119, 249–275.

Bound, J., D. Jaeger & R. Baker (1995), "Problems with Instrumental Variable Estimation When the Correlation Between the Instruments and the Endogenous is weak," *Journal of the American Statistical Association*, 90(430), 443–450.

Criscuolo, C, R. Martin, H. Overman & J. Van Reenen (2012), "The causal effects of an industrial policy," CEPR Discussion Papers 8818, C.E.P.R. Discussion Papers.

Duflo, E., R. Glennerster & M. Kremer (2007), "Using Randomisation in Development Economics Research: A Toolkit,2" CEPR Discussion Papers 6059, C.E.P.R. Discussion Papers.

Duflo, E., & M. Kremer (2005), "Use of Randomisation in the Evaluation of Development Effectiveness," in *Evaluating Development Effectiveness*, ed. by O. Feinstein, G. K. Ingram, & G. K. Pitman. New Brunswick, New Jersey & London, U.K.: Transaction Publishers, vol. 7, 205{232}.

Einiö, Elias (2013), "R&D Subsidies and Company Performance: Evidence from Geographic Variation in Government Funding Based on the ERDF Population-Density Rule", *The Review of Economics and Statistics* (forthcoming).

Kommissionens evalueringsstandarder er tilgængelige på http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/standards_c_2002_5267_final_en.pdf.

Garicano, L. C. Lelarge & J. Van Reenen, (2012), "Firm Size Distortions and the Productivity Distribution: Evidence from France," CEP Discussion Papers dp1128, Centre for Economic Performance, LSE.

Givord, P. (2010), "Méthodes économétriques pour l'évaluation de politiques publiques", WPD3E n° G2010-08.

Givord, P., Rathelot, R. & P. Sillard (2013), "Place-based tax exemptions and displacement effects: An evaluation of the Zones Franches Urbaines programme", *Regional Science and Urban Economics*, Volume 43, Issue 1, January 2013, 151-163.

Heckman, J. J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica* 47, 153–161.

Imbens, G. & J. Wooldridge (2009), *Recent Developments in the Econometrics of Programme Evaluation*, *Journal of Economic Literature*, 47:1, 5-86.

Imbens, G. W., et T. Lemieux (2008), "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, 142(2), 615–635.

C. Lelarge, D. Sraer & D. Thesmar (2010), "Entrepreneurship and Credit Constraints: Evidence from a French Loan Guarantee Programme," NBER Chapters, in: *International Differences in Entrepreneurship*, s. 243-273, National Bureau of Economic Research, Inc.

Keane, M. P. (2010), "A Structural Perspective on the Experimentalist School." *Journal of Economic Perspectives*, 24(2): 47-58.

Martini, A. & D. Bondonio (2012), "Counterfactual impact evaluation of cohesion policy: impact and cost effectiveness of investment subsidies in Italy", rapport til Kommissionen, DG REGIO.

Nederlandse Rijksoverheid (2012), "Durf te meten", Eindrapport Expertwerkgroep Effectmeting, tilgængelig på <http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2012/11/23/durf-te-meten-eindrapport-expertwerkgroep-effectmeting.html>.

Nevo, A. & M. D. Whinston, (2010), "Taking the Dogma out of Econometrics: Structural Modeling and Credible Inference." *Journal of Economic Perspectives*, 24(2): 69-82.

OECD's evalueringsnormer og standarder er tilgængelige på <http://www.oecd.org/dac/evaluation/dcdndep/41612905.pdf>.

Sims, C. A. (2010), "But Economics Is Not an Experimental Science." *Journal of Economic Perspectives*, 24(2): 59-68.

Stock, J., J. Wrightand & M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalised Method of Moments," Journal of Business and Economic Statistics, 20(4), 518–29.

United Nations Evaluation Group (2005), ' "Standards for Evaluation in the UN System" tilgængelig på: http://www.uneval.org/papersandpubs/documentdetail.jsp?doc_id=22.

Wooldridge, J. (2002), "Econometric Analysis of Cross Section and Panel Data", Cambridge: MIT Press.

World Bank (2003), "Independent Evaluation: Principles, Guidelines and Good Practice". Tilgængelig på <http://siteresources.worldbank.org/INTDGF/Resources/Evaluation&LearningNote.pdf>.