

# KLASSISKA EXPERIMENT INOM ARBETSMARKNADS- POLITIKEN

av  
Anders Björklund





## **Industriens Utredningsinstitut**

är en fristående vetenskaplig forskningsinstitution grundad 1939 av Svenska Arbetsgivareföreningen och Sveriges Industriförbund.

### **Syfte**

Att bedriva forskning rörande ekonomiska och sociala förhållanden av betydelse för den industriella utvecklingen.

### **Verksamhet**

Huvuddelen av arbetet inom institutet ägnas åt långsiktiga forskningsuppgifter. Man siktar härvid till ett studium av de grundläggande sammanhangen inom näringslivet och särskilt till att belysa de frågor som hör samman med strukturella och institutionella förändringar. Forskningsresultaten publiceras i institutets skriftserier.

### **Styrelse**

Tekn. dr Curt Nicolin, ordf.  
Direktör Gösta Bystedt  
Direktör Anders Carlberg  
Direktör Torbjörn Ek  
Direktör Per-Olof Eriksson  
Tekn. dr Lennart Johansson  
Direktör Ulf Laurin  
Direktör Olof Ljunggren  
Direktör Lars Nabseth  
Ekon. dr Hans Stahle  
Direktör Marcus Storch  
Ekon. dr Peter Wallenberg  
Direktör Sven Wallgren  
Docent Gunnar Eliasson, chef

### **Adress**

Industriens Utredningsinstitut  
Box 5501, 114 85 Stockholm  
Tel. 08-783 80 00

INDUSTRIENS UTREDNING SINSTITUT

KLASSISKA EXPERIMENT INOM  
ARBETSMARKNADSPOLITIKEN

av

Anders Björklund

Distribution:

Almqvist & Wiksell International, Stockholm

## FÖRORD

Den aktiva arbetsmarknadspolitiken har länge varit ett flaggskepp i svensk ekonomisk politik. Arbetsmarknadsinsatserna har varit ökande och omfattande och dragit årliga resurser som räknats i procent av BNP. Den svenska förebilden har manat till efterföljd i många industriländer. Man hade därför väntat sig att politiken följts av en betydande utvärderingsforskning. Så har dock inte alls varit fallet. Större delen av den empiriska forskningen på detta område har utförts utomlands, särskilt i USA. Docent Anders Björklund redovisar i denna skrift en översikt av kunskapsläget och resultaten från egen forskning, som till betydande del gäller just utvärdering av arbetsmarknadspolitikens effekter. Det är intressant att konstatera att vi för närvarande vet mycket litet om vilka effekter arbetsmarknadspolitiken har, att effekterna ibland kan vara rakt motsatta de förväntade och att klarhet förmodligen inte kan nås om man inte i betydande utsträckning övergår till okonventionella, ofta dyrbara experimentella utvärderingsmetoder. Skriften avslutas med konkreta förslag för den framtida arbetsmarknadsforskningen.

Stockholm i juli 1989

Gunnar Eliasson

## INNEHÅLL

	sida
<b>1</b>	<b>Inledning</b> 7
<b>2</b>	<b>Erfarenheter av icke-experimentella studier</b> 11
2.1	Amerikanska studier 11
2.2	Svenska studier 15
<b>3</b>	<b>Experiment inom utvärderingsforskningen</b> 19
3.1	Allmänna problem med experiment 19
3.2	National Supported Work Demonstration, USA 23
3.3	Löne-vouchers till socialbidragstagare i Dayton, Ohio, USA 27
3.4	Bonus till arbetslösa i Illinois, USA 30
3.5	Employment Opportunity Pilot Project (EOPP) 36
3.6	JTPA-experimentet i USA 39
3.7	Workfare-experimenten i USA 41
3.8	Intensifierad förmedlingsservice, Eskilstuna 44
<b>4</b>	<b>Utvärdering av utvärderingar med hjälp av experiment</b> 51
<b>5</b>	<b>Makroexperiment</b> 59
<b>6</b>	<b>Förslag för den framtida utvärderingsforskningen</b> 63
6.1	Experiment 64
6.2	Databaser för icke-experimentell utvärdering 72
<b>Appendix a</b>	<b>Metoder använda av Bassi (1984) och Dickinson, Johnson och West (1986) i de studier som presenterades i tabell 2.1.</b> 75
<b>Appendix b</b>	<b>Metoder använda av Björklund (1989) i den studie som presenterades i tabell 2.2.</b> 77
<b>Appendix c</b>	<b>Metoder använda av Heckman, Hotz och Dabos (1987) i den studie som presenterades i tabell 4.2.</b> 81

<b>Litteratur</b>	83
<b>Tabeller</b>	
2.1 Några resultat från två studier av CETA-programmet	14
2.2 Några resultat från en studie av den svenska arbetsmarknadsutbildningens effekter	17
3.1 Sysselsättningstal för experiment- och kontrollgrupper i Dayton	29
3.2 Några resultat från experimentet i Illinois	32
3.3 Några resultat från workfare-experimentet i San Diego	43
3.4 Några centrala resultat från Eskilstuna-studien	48
4.1 Estimerade effekter (i dollar) av National Supported Work Demonstration med hjälp av slumpmässigt utvald kontrollgrupp respektive jämförelsegrupp från Current Population Survey	53
4.2 Resultat från utvärdering av testmetodik i Heckman, Hotz och Dabos (1987). Estimerade effekter på inkomsterna 1978 resp 1979	56

## 1 Inledning

Många arbetsmarknadspolitiska åtgärder är avsedda att förbättra enskilda individers ställning på arbetsmarknaden. Det kanske bästa exemplet inom den svenska arbetsmarknadspolitiken är arbetsmarknadsutbildningen. Man brukar säga att det är en individinriktad åtgärd. Av politiska målformuleringar framgår inte särskilt väl hur "ställning på arbetsmarknaden" skall definieras operationellt. Ett ofta använt mått är dock arbetsinkomster under en viss period, exempelvis ett år. Detta mått innehåller en rad komponenter: antal dagar i sysselsättning under året, antal arbetade timmar och genomsnittlig timlön påverkar alla de sammanlagda arbetsinkomsterna under ett år. Sannolikt skulle många politiker fästa störst vikt vid att antalet dagar i sysselsättning ökar (och antalet dagar i arbetslöshet minskar) som en följd av politiska insatser, men även höjda timlöner skulle säkert av de flesta uppfattas som god måluppfyllelse. Även förbättrad arbetsmiljö kan ingå i målen för arbetsmarknadsutbildning.

En viktig uppgift för utvärderingsforskningen måste därför vara att kvantifiera hur mycket inkomsterna – eller olika komponenter därav – påverkas av arbetsmarknadsutbildning och liknande åtgärder. Den forskning som dominerat inom arbetsmarknadsekonomi (labor economics) under de senaste decennierna har, föga överraskande, byggt på någon form av jämförelse av inkomsterna efter utbildningsperioden mellan deltagare och icke-deltagare. En lämpligt gjord sådan jämförelse anses kunna bestämma hur stor effekten av åtgärden har varit. Det centrala metodproblemet inom denna forskning har därför varit att göra denna jämförelse på ett konstruktivt sätt.

Två olika ansatser har dominerat i litteraturen. Den ena har varit klassiska experiment, som innebär att prospektiva programdeltagare slumpmässigt indelas i en experimentgrupp som får delta i programmet och en kontrollgrupp som inte får delta i programmet. Eftersom alla egenskaper som påverkar utfallet (i termer av årsinkomster eller liknande), utom just programdeltagande, måste vara lika fördelade i de båda grupperna anses skillnaden kunna hänföras till just programmet. Ibland brukar experiment betecknas som den ideala metoden.

Den andra ansatsen innebär att man utnyttjar data för programdeltagare och icke-deltagare såsom de faktiskt valts ut i den selektionsprocess som avgör

deltagandet. I praktiken innebär denna selektionsprocess oftast beslut av både de myndigheter som handhar åtgärderna och de prospektiva deltagarna själva. Olika statistiska metoder har utvecklats för att kunna dra slutsatser om effekter även från data över icke slumpmässigt utvalda grupper. Hit hör "matchningstekniker" som innebär att deltagarna jämförs med en grupp icke-deltagare – en jämförelsegrupp<sup>1</sup> – som har liknande egenskaper som deltagarna. Den vanligaste metoden har dock varit att "kontrollera för" systematiska skillnader mellan deltagare och jämförelsegrupp med hjälp av någon form av regressionsanalys.

Under senare år har en betydande metodutveckling ägt rum inom den icke-experimentella forskningsansatsen. Översikter av denna litteratur återfinns (på svenska) i Björklund (1983) och (mer utförligt och mer tekniskt) i Heckman och Robb (1985a, 1985b). I USA har även betydande insatser gjorts för att få fram bättre databaser med vars hjälp de icke-experimentella metoderna kan tillämpas.

Trots dessa ambitiösa satsningar på både statistisk metod och på uppbyggnad av databaser råder idag bland många forskare en betydande misstro om möjligheterna att producera tillförlitliga skattningar av programmets effekter med hjälp av icke-experimentella tekniker. I en artikel som publicerades i *Evaluation Review* karakteriserar Heckman, Hotz och Dabos (1987) sinnesstämningarna på följande sätt:

"In the past three years, a pessimistic mood has become dominant in the evaluation literature. New studies have been put forth that claim the non-experimental methods of program evaluation produce 'unbelievable', 'sensitive', 'nonrobust' estimates of program impacts and that randomized experiments are necessary to produce reliable estimates."

Det "jämförelseproblem" som den ovan nämnda forskningen fokuserat på är emellertid inte det enda problem som bör beaktas när man försöker kvantifiera effekterna av program som arbetsmarknadsutbildning. En (ofta underförstådd)

---

<sup>1</sup>Genomgående kommer begreppet jämförelsegrupp att användas för en icke slumpmässigt utvald grupp av icke-deltagare, medan kontrollgrupp används för en slumpmässigt vald grupp.



förutsättning för såväl den experimentella som den icke-experimentella ansatsen är att effekterna begränsar sig till dem som faktiskt deltar i programmet. De som inte deltar i arbetsmarknadsutbildning antages alltså inte alls påverkas av att andra deltar. Detta antagande är långt ifrån självklart. Man kan mycket väl tänka sig såväl positiva som negativa effekter för andra grupper. Sådana indirekta effekter kan uppstå både under själva utbildningstiden och efteråt. Denna problematik betonades också i de första svenska utvärderingsstudierna och i EFAs betänkande "Att utvärdera arbetsmarknadspolitik" från 1974 (SOU 1974:29). Någon metodik för att kvantifiera sådana effekter kunde dock inte anvisas och i den internationella – huvudsakligen amerikanska – utvärderingsforskning, som återfinns i tidskriftslitteraturen, tas problemet normalt inte alls upp. Detta betyder således att de indirekta effekterna förutsätts vara små eller obetydliga.

Avsikten med denna skrift är att diskutera möjligheterna att komma vidare inom den utvärderingsforskning som syftar till att kvantifiera effekterna av åtgärder som arbetsmarknadsutbildning. Framställningen är präglad av den befintliga litteraturen som är inriktad på att lösa det så kallade jämförelseproblemet. Både den experimentella och den icke-experimentella ansatsen kommer att behandlas.

Som en bakgrund presenteras i avsnitt 2 en sammanfattning av de resultat som erhållits vid icke-experimentella studier av amerikanska och svenska utbildningsprogram. Det är osäkerheten och den låga precisionen i resultaten från dessa studier som är bakgrunden till önskemålen om att genomföra klassiska experiment.

I det långa avsnittet 3 presenteras sex olika experiment som genomförts, och ytterligare ett som planeras, på det arbetsmarknadspolitiska området. Bara ett av dessa avser visserligen arbetsmarknadsutbildning, men de metodologiska erfarenheterna är relevanta för de flesta individinriktade åtgärder.

Experiment kan även utnyttjas för att utvärdera icke-experimentella studier – en slags "utvärdering av utvärderingar". Denna ansats – och de studier som utnyttjat denna möjlighet – presenteras i avsnitt 4.

Därefter följer i avsnitt 5 en diskussion om så kallade "makro-experiment" som innebär att orter eller regioner slumpmässigt fördelas på experiment- och kontrollgrupper. De vanligaste experimenten har annars varit "mikro-experiment", dvs individer har slumpmässigt fördelats på experiment- och kontrollgrupper.

Skriften avslutas med en diskussion om hur experiment och icke-experimentella studier kan utvecklas var för sig och i kombination. Konkreta förslag att förbättra utvärderingarna av typiska arbetsmarknadspolitiska insatser presenteras.

## 2 Erfarenheter av icke-experimentella studier

### 2.1 Amerikanska studier

Kännetecknande för amerikansk arbetsmarknadspolitik sedan början av 1960-talet har varit att en mängd olika program har prövats, vanligen i liten skala och under relativt korta perioder. De olika programmen har blivit utvärderade på olika sätt; hundratals studier har genomförts. När det gäller metodologiska erfarenheter är emellertid studierna kring CETA-programmet i USA under 1970-talet av störst intresse, varför vi inskränker oss till att behandla dessa studier.

CETA (Comprehensive Employment and Training Act) startade 1973 och innehöll olika delprogram, såsom "public service employment", "work experience" och "direct referrals" (renodlad förmedlingsservice). Somliga deltog i flera av dessa aktiviteter. Department of Labor tog initiativ till att samla in data för kommande utvärderingar; databasen kom att kallas "Continuous Longitudinal Manpower Survey" eller CLMS. Först och främst insamlades information om dem som deltagit i CETA under 1975 (6 700 individer i urvalet) och under 1976 (13 300 individer i urvalet). Variabler som beskriver bland annat demografiska egenskaper, familjesituation, arbetsmarknadserfarenheter under fyra år med början ett år före programstarten ingick i CLMS. Till denna information om CETA-deltagarna har kopplats data från ett urval personer från den amerikanska arbetskraftsundersökningen (från mars 1976). Både för denna grupp och för CETA-deltagarna i CLMS insamlades registerdata om inkomster ("pensionsgrundande" inkomster från Social Security Administration) för perioden 1951–1978, dvs två år efter programdeltagandet och en mycket lång period före detsamma.

Även om den databas som konstruerades på detta sätt är det mest ambitiösa försöket att skapa en utvärderingsdatabas, så har den vissa brister. Några centrala variabler, som vissa arbetsmarknadserfarenheter, fackföreningstillhörighet och region saknas (dvs finns inte för samtliga individer) och det inkomstmått som utgör den centrala utfallsvariabeln har vissa brister. Den största bristen är emellertid att det inte varit möjligt att avgöra vilka av dem som ingår i jämförelsegruppen som också deltagit i CETA. Om detta varit möjligt – och med bättre registerinformation hade det varit möjligt – skulle dessa individer själv-

fallet ha tagits bort från jämförelsegruppen. Den analys som gjorts med den befintliga jämförelsegruppen är därför inte helt tillförlitlig; generellt sett är effekterna sannolikt något underskattade på grund av den så kallade "smittobias" (contamination bias) som uppstår (se Bassi (1983) för mer ingående diskussion).

En mängd studier har baserats på de data som samlats in på detta sätt. Att studierna blivit många har säkerligen underlättats av att materialet varit så kallade "public-use-data" och funnits tillgängliga för hela forskarsamhället till självkostnadspris.<sup>2</sup> Barnow (1987) innehåller en noggrann metodologisk granskning av dessa studier och en sammanfattning av resultaten.

En lättillgänglig översikt över resultaten från studierna av såväl CETA-programmet som andra amerikanska program återfinns i Bassi och Ashenfelter (1986). Kännetecknande för alla dessa studier är att de utnyttjar de möjligheter som longitudinella data, eller paneldata, erbjuder. Den utfallsvariabel som använts är årsinkomster (under 1977 eller 1978), och motsvarande variabel finns även tillgänglig för tidsperioder före programmet. Därmed finns möjlighet att "kontrollera för" såväl inkomsternas nivå som deras förlopp före programmet. Detta kan göras dels med hjälp av matchningstekniker, som innebär att deltagarnas utfall jämförs med en grupp som har liknande egenskaper bland annat avseende tidigare inkomster, dels med hjälp av regressionstekniker där tidigare inkomster utnyttjas som kontrollvariabler. I praktiken har en kombination av dessa metoder utnyttjats i de studier som Barnow sammanfattar. Först har ett antal individer från den totala jämförelsegruppen, som ju var ett slumpmässigt urval från hela befolkningen, valts ut till en mer avgränsad jämförelsegrupp. Därefter har regressionsanalys tillämpats.

Inom den tillämpade arbetsmarknadsforskningen har en mängd modeller med tillhörande estimationsmetoder utvecklats för att beräkna effekter av, exempelvis, arbetsmarknadsutbildning utifrån longitudinella data. Varje modell kan sägas ha två delar; en som beskriver inkomsternas dynamiska struktur (eller utveckling över tiden) och en som beskriver hur selektionen till programmet gått

---

<sup>2</sup>I ett sådant skick att enskilda individer inte kunnat identifieras.

till. Valet av modell styrs därför av hur dessa båda delar och sambandet mellan dem bör specificeras. Heckman och Robb (1985a och 1985b) innehåller en karta över modeller och tillhörande estimationstekniker som kan tillämpas.

Trots den relativt goda databasen och de nya sofistikerade metoder som utvecklats och kommit till användning är resultaten långt ifrån entydiga. Den osäkerhet som föreligger är av två slag, dels ger varje estimerad modell en skattning av programmets effekt som är förknippad med en viss osäkerhet, dels ger olika modeller olika resultat, och det är svårt att på a priori grunder avgöra vilken modell som är mest tillförlitlig. Båda källorna till osäkerhet illustreras av några av resultaten från två av de studier som Barnow granskar, se tabell 2.1. Det bör dock framhållas att varken Barnow eller andra uttryckligen betonar den första typen av osäkerhet. Detta är ju den klassiska osäkerhet (standardfelet) som varje skattning är förknippad med och som avgör bredden på konfidensintervall-  
et för skattningen. Mitt intryck från litteraturen är att standardfelen ofta är mycket höga. Likafullt diskuteras sällan hur denna osäkerhet kan reduceras. I Björklund (1989), kapitel 9, anger jag några olika möjligheter.

Den första typen av osäkerhet anges i tabellerna genom att varje punktskattning kompletteras med en uppgift som anger plus/minus 2 gånger standardfelet:  $740 \pm 260$  innebär således att punktskattningen är 740 dollar, men ett 95-procents konfidensintervall blir 480–1 000 dollar. Det smalaste konfidensintervallet i tabellen (som erhöles av Bassi för en "fixed effect model med seriekorrelerad slumpterm" för 1977) är 2 gånger 192 dollar, dvs nära 400 dollar. Med tanke på att inkomsterna detta år var mellan 3 500 och 4 700 för dessa personer motsvarar denna osäkerhet cirka tio procent av inkomsterna, vilket är mycket högt. För andra grupper och metoder kan konfidensintervallen få en bredd av 1 500 dollar, dvs omkring en tredjedel av inkomstnivån.

Nästa typ av osäkerhet följer av att olika metoder (och olika studier) ger olika resultat. De skattningar som presenteras från Bassis studie har alla utsatts för ett testförfarande, och ingen av modellerna har kunnat förkastas. Det betyder att det är omöjligt att hävda att en metod är mer tillförlitlig än den andra. För gruppen vita kvinnor varierar resultaten mellan 740 och 987 dollar (1977 års inkomster) respektive 1 108 och 1 621 (1978 års inkomster). För kvinnor från minoritetsgrupper är motsvarande variationer 426 och 626 dollar (1977 års

Tabell 2.1 Några resultat från två studier av CETA-programmet

A Effekt (i dollar) på årsinkomsterna 1977 och 1978 enligt Bassi (1984)			
Metod	Grupp av deltagare		
	Vita kvinnor	Kvinnor från minoritetsgrupp	Män från minoritetsgrupp
<u>Fixed effect model</u>			
Inkomster 1977	740 ± 260	426 ± 470	117 ± 632
Inkomster 1978	1 108 ± 336	531 ± 570	27 ± 772
<u>Fixed effect model med seriekorrelerad slumpterm</u>			
Inkomster 1977	987 ± 192	—	164 ± 426
Inkomster 1978	1 621 ± 220	—	132 ± 538
<u>Creaming model</u>			
Inkomster 1977	—	626 ± 326	271 ± 430
Inkomster 1978	—	947 ± 388	240 ± 540
B Effekt (i dollar) på årsinkomsterna 1978 enligt Dickinson, Johnson, West (1986)			
Metod	Grupp av deltagare		
	Vuxna män	Vuxna kvinnor	
State dependence model	-690 ± 278	13 ± 232	

Metoderna presenteras närmare i appendix a.

inkomster) respektive 531 och 947. För män är variationerna i punktskattningarna mindre, men å andra sidan är konfidensintervallen exceptionellt stora.

Osäkerheten till följd av metodskillnader förstärks om Bassis resultat jämförs med dem som Dickinson, Johnson och West (1986) erhöill. För vuxna män erhöill de relativt kraftiga negativa effekter, som dessutom var statistiskt signifikant skilda från noll, medan effekten för kvinnor var nära noll. Detta är den studie som ger sämst resultat av alla de studier som Barnow presenterar. Nu bygger de båda studierna på olika delgrupper ur CLMS-materialet, varför de inte är riktigt jämförbara. Därför kan möjligen resultaten vara förenliga med varandra. Barnow gör emellertid bedömningen att de effekter som Dickinson, Johnson och West erhöill är alltför låga p g a olika brister i materialet (se s 189 hos Barnow). Hans sammanfattande bedömning är därför att de flesta studier hamnar på effekter i storleksordningen 200 till 600 dollar (dvs ca 2,5 till 7,5 procent av årsinkomsten) med något högre tal för kvinnor och något lägre tal för män. Osäkerheten är emellertid betydande av de skäl som presenterats ovan. Barnow menar att experiment med randomiserat urvalsförfarande "appears to be the only method available at this time to overcome the limitations of nonexperimental evaluations".

Samma bedömningar om osäkerheten i resultaten och behovet av experiment gör Bassi och Ashenfelter (1986) i sin genomgång av forskningen. Det är dessa slutsatser som Heckman, Hotz och Dabos (1987) syftar på när de hävdar att en pessimistisk stämning råder inom utvärderingslitteraturen (jämför citatet ovan i avsnitt 1).

## **2.2 Svenska studier**

Tre studier av arbetsmarknadsutbildningen i Sverige genomfördes i början av 1970-talet, och alla pekade på att deltagarnas inkomster ökade tack vare utbildningen. Precisionen i statistiska termer av dessa skattningar var emellertid oklar. Främst är det emellertid två andra skäl som gör att studierna i dag har ett begränsat värde. För det första avsåg alla studierna den arbetsmarknadsutbildning som bedrevs på 60-talet, då nivån var betydligt lägre än i dag. Om det föreligger en avtagande avkastning, vilket är rimligt att tro, kan effekterna vara

mindre positiva i dag. För det andra var metodiken för att skatta effekterna, med dagens mått mätt, inte tillfredsställande. Studierna gjordes ju innan metodutvecklingen på detta område sköt fart på 1970-talet.

I en pågående studie (Björklund (1989)) har jag själv försökt tillämpa några av de metoder som använts i den amerikanska litteraturen. Studien har fördelen att kunna utnyttja longitudinella data, men materialet är inte i första hand uppbyggt för analys av arbetsmarknadspolitiska åtgärder. Urvalet är bland annat ganska litet.

Resultaten utvisar tyvärr samma typ och ungefär samma grad av osäkerhet som de amerikanska studierna. Tabell 2.2 rapporterar ett urval av resultaten. Varje enskild effektskattning uppvisar en osäkerhet som motsvarar konfidensintervall av en längd mellan cirka 10 och 38 procentenheter. (Notera att effekterna i tabellen är uttryckta i relativa termer, vilket beror på att inkomst- och lönestudierna formulerats med logaritmerade beroende variabler.)

Till denna osäkerhet kommer att det finns en besvärande skillnad i resultaten för de två olika metoderna. Detta gäller särskilt effekterna på timlönen och på årsinkomsterna.



Tabell 2.2 Några resultat från en studie av den svenska arbetsmarknadsutbildningens effekter

A Effekten av arbetsmarknadsutbildning 1976–80 på sysselsättning och timlöner 1981		
	State dependence model	Fixed effect model
Effekt på andelen med sysselsättning vid intervjutillfälle 1981	0,055 ± 0,060	0,080 ± 0,076
Relativ effekt på timlönen för dem med sysselsättning	-0,043 ± 0,054	0,065 ± 0,076
B Effekten av arbetsmarknadsutbildning 1976–82 (första halvåret) på inkomster under 1983		
	State dependence model	Fixed effect model
Effekt på andelen med inkomster någon gång under året	0,009 ± 0,048	0,013 ± 0,068
Relativ effekt på årsinkomsterna för dem med inkomster under året	0,044 ± 0,16	0,186 ± 0,190

Källa: Björklund (1989).

Metoderna beskrivs mer ingående i appendix b.

### 3 Experiment inom utvärderingsforskningen

Syftet med ett klassiskt experiment är att lösa det jämförbarhetsproblem som den icke-experimentella ansatsen lider av. Experiment kan emellertid vara förknippade med egna problem. I litteraturen om experiment finns utförliga diskussioner kring sådana problem (se t ex Burtless och Orr (1986)). Innebörden av dessa problem förstår man dock bäst om de diskuteras i anslutning till konkreta exempel, antingen experiment som redan genomförts eller konkreta förslag till nya experiment. Det främsta syftet med detta avsnitt är därför att presentera ett antal experiment som faktiskt har genomförts och att diskutera experimentens specifika problem i anslutning till dessa. Vi tar dock först upp några allmänna problem som experiment anses vara förknippade med.

#### 3.1 Allmänna problem med experiment

##### Experimentgruppens representativitet

För att ett experiment ska ge övertygande resultat krävs att experimentgruppen är representativ för den grupp som kommer att delta när programmet är i normal drift.

Man kan här skilja mellan två renodlade fall. Anta först att (praktiskt taget) alla i den population man studerar, som önskar delta i ett program också kan ges möjlighet att delta. Ett exempel är flyttningsstöden inom svensk arbetsmarknadspolitik som alla arbetslösa är berättigade till. Även arbetsmarknadsutbildningens kurser har i stor utsträckning varit tillgängliga för de flesta sökande under 1970- och 80-talen i Sverige. Anta vidare att man i utvärderingssyfte väljer att slumpmässigt fördela de sökande till dessa program på en experimentgrupp som tillåts delta (eller få del av servicen) och på en kontrollgrupp som ej tillåts delta. För att denna upplägning ska vara tillförlitlig krävs att själva förekomsten av experiment inte påverkar vilka som ansöker om att få delta. Om den osäkerhet, som introduceras av experimentet, påverkar vilka som ansöker så kan resultaten bli missvisande.

Ett annat fall gäller ett program där normalt bara en del av de sökande kan ges

möjlighet att delta. Om exempelvis tillgången på arbetsförmedlare är knapp så tvingas en arbetsförmedling normalt att prioritera bland de sökande så att somliga får relativt omfattande service medan andra kanske bara får begränsad service. Ett experiment med syfte att mäta effekterna av förmedlingsservice kan då leda till att en slumpmässigt utvald experimentgrupp, som erhåller service, får en annan sammansättning än den skulle ha haft när programmet är i normal drift. Detta kan också leda till att experimentet ger missvisande resultat.

Sammantaget belyser dessa exempel att en slumpmässigt utvald experimentgrupp inte behöver vara representativ för dem som deltar när programmet är i normal drift och besluten att delta fattas av programmyndigheter och prospektiva programdeltagare.

#### Hawthorne-effekter

Begreppet Hawthorne-effekt härstammar från en försöksverksamhet, vid Western Electric Company i Chicago på 1920-talet. Effekten av förändringar i arbetsmiljön på arbetsproduktiviteten undersöktes. De första studierna ledde till att även påtagliga försämringar i den fysiska arbetsmiljön gav höjd produktivitet.<sup>3</sup> Skälet skulle vara att själva försöksverksamheten uppfattades som en positiv omtanke om de anställda och att produktiviteten därför steg. Med Hawthorne-effekt har man därför kommit att mena en effekt som uppstår på grund av experimentet i sig snarare än på grund av den åtgärd som är föremål för en experimentell studie.

När det gäller arbetsmarknadspolitiska experiment måste man därför fråga sig om de speciella omständigheter som experimentet innebär leder till att de inblandade agenterna beter sig annorlunda. Till de inblandade agenterna hör både programdeltagarna själva och de som driver och administrerar programmen.

I första hand finns vissa risker för att de administrerande myndigheterna kan bete sig annorlunda. Ett exempel kan vara att det är frågan om ett nytt medel

---

<sup>3</sup>Senare har emellertid dessa paradoxala resultat ifrågasatts (se Franke och Kaul (1978)).

som är föremål för experimentet och att man inte hunnit lära sig hantera detta medel effektivt under experimentets gång. Ett annat exempel kan vara att goda resultat från experimentet kan leda till stora framtida resurstillskott till myndigheten och att de anställda därför anstränger sig extra för att uppnå positiva resultat. I båda dessa exempel kan man tro att riskerna för missvisande resultat är störst om experimentet är mycket kortvarigt.

Sammanfattningsvis är det viktigt att se till att samtliga inblandade agenter beter sig så "normalt" som möjligt under experimentets gång.

### Indirekta effekter

Vissa arbetsmarknadspolitiska åtgärder kan leda till att deltagarna erhåller förbättringar på arbetsmarknaden på bekostnad av andra. Såväl förmedlings-service, utbildning, som olika lönesubventioner kan ha sådana konsekvenser. En experimentell uppläggningsstudie kan därför innebära att man erhåller kraftiga förbättringar för experimentgruppen jämfört med kontrollgruppen, men att denna skillnad enbart återspeglar att en grupp fått de befintliga arbetena på bekostnad av andra.

En konkret form av sådana effekter kan vara så kallad "queueing-bias" (kö-bias). Om arbeten normalt fördelas enligt ett visst kösystem kan experimentet leda till att experimentgruppen hamnar först i kön medan kontrollgruppen hamnar sist. I så fall skulle resultaten framstå som mycket positiva utan att några effekter på den totala sysselsättningen uppnåtts.

Även positiva indirekta effekter är emellertid tänkbara. Utbildning och flyttningstöd kan leda till att flaskhalsar i produktionen reduceras och att andra grupper tack vare detta får sysselsättning. Utbildning kan också ha positiva externa effekter; den kunskap en viss person erhåller under en kurs kan förmedlas vidare till andra som ett led i det ordinarie arbetet.

En annan typ av indirekta effekter av experimentet kan verka via traditionella marknadsmekanismer. Om experimentgruppen erhåller en viss typ av yrkesutbildning kan det ökade utbudet av denna yrkeskategori leda till löneförändringar

både för denna yrkesgrupp och andra. Riktningen på sådana effekter – dvs om övriga individer vinner eller förlorar – är generellt sett svår att uttala sig om.

Indirekta effekter på andra än dem som tillhör experimentgruppen skapar således problem vid användningen av denna utvärderingsmetod. Det är emellertid viktigt att betona att dessa problem inte är specifika för experiment som sådana, utan är precis lika stora för den icke-experimentella ansatsen.

Slutsatsen är emellertid att varje studie som bygger på experimentell uppläggning måste kompletteras med en bedömning av huruvida indirekta effekter av olika slag kan ha förekommit, och i vilken riktning de i så fall kan ha verkat.

#### Etiska problem

En experimentell studie, med slumpmässig fördelning av individer på experiment- och kontrollgrupp, anses ibland omöjlig av etiska skäl. Många kanske därför slår bort tanken att genomföra experiment utan ingående analys av hur de etiska problemen egentligen ser ut och hur dessa förhåller sig till experimentets fördelar. Eftersom ett antal experiment inom arbetsmarknadspolitiken faktiskt har genomförts kan man inte utan vidare avfärda tanken med hänvisning till några diffusa etiska problem. Ett av dessa experiment genomfördes dessutom i Sverige i början av 70-talet (se nedan avsnitt 3.8).

Allmänt sett är huvudproblemet att det kan vara etiskt stötande att förvägra de personer som skall utgöra kontrollgrupp en viss form av service. Hur allvarligt detta är varierar dock från fall till fall. Om det handlar om en rättighet som funnits länge och är starkt förankrad i samhället kan det självfallet framstå som etiskt otillfredsställande med experiment. Att exempelvis förvägra ett slumpmässigt urval av arbetslösa arbetslöshetsersättning i syfte att studera ersättningsystemets effekter på arbetslöshetstiderna är nog etiskt oacceptabelt för de flesta. Däremot kan det vara möjligt att acceptera slumpmässig fördelning när det gäller mer avgränsade nya inslag i medelsarsenalen.

Det bör dock framhållas att det också kan vara oetiskt att avstå från experiment. Om en sådan studie är det enda realistiska sättet att skaffa kunskap om

effekterna av en dyrbar åtgärd, kan det vara oetiskt att förvägra arbetsmarknadsdepartement, riksdag och skattebetalarna kunskap om medlets effekter och lönsamhet.

### Kostnader

Ett praktiskt problem med experiment kan vara höga kostnader. Precis som fallet var med de etiska problemen är det svårt att uttala sig generellt om kostnadsproblemet. Experiment kan genomföras på olika sätt och de direkta kostnader som uppstår måste också vägas mot de potentiella vinster i form av bättre arbetsmarknadspolitik som kan uppnås.

### **3.2 *National Supported Work Demonstration, USA*<sup>4</sup>**

National Supported Work Demonstration, som genomfördes på 15 olika platser i USA under åren 1975 till 1978, är det mest kända och ambitiösa experiment som genomförts på det arbetsmarknadspolitiska området. Det finns dokumenterat i en bok (Hollister, Kemper och Maynard (1984)), det har givit upphov till åtskilliga artiklar (bl a ett temanummer i Journal of Human Resources, höstnumret (Fall) 1981) och det har utnyttjats till de "utvärderingar av utvärderingar" som genomförts under mitten av 1980-talet och som presenteras i avsnitt 4 nedan.

Det arbetsmarknadspolitiska instrument som testades i detta projekt riktades till grupper som var påtagligt svårplacerade på arbetsmarknaden. Fyra målgrupper identifierades:

1. Kvinnor med AFDC (Aid for Families with Dependent Children). Kravet var att kvinnorna uppbar AFDC och hade gjort detta under minst 30 av de senaste 36 månaderna. Vidare krävdes att det yngsta barnet var 6 år eller äldre.

---

<sup>4</sup>Denna presentation bygger främst på Hollister, Kemper och Maynard (1984).

2. Missbrukare. Dessa skulle vara lägst 18 år och delta i ett behandlingsprogram eller ha deltagit i ett sådant under de senaste 6 månaderna.
3. Ex-interner. Dessa skulle också vara lägst 18 år och ha varit internerade för brott inom de senaste 6 månaderna.
4. Ungdomar. Dessa skulle vara mellan 17 och 20 år, ha avbrutit gymnasieskolan (high school) och inte varit i skolan under de senaste 6 månaderna. För minst 50 procent av ungdomarna krävdes också att de gjort sig skyldiga till någon form av brottslig verksamhet.

För samtliga grupper krävdes dessutom att deltagarna var arbetslösa och hade arbetat mindre än 3 månader under de senaste 6 månaderna.

Det arbetsmarknadspolitiska instrument som testades på dessa grupper var en form av skyddad eller halvskyddad sysselsättning med fyra karakteristika. För det första var arbetet tidsbegränsat till 12 eller 18 månader. De anställda kunde självfallet frivilligt lämna sina arbeten tidigare och de kunde avskedas om de misskötte sig, men det var klart utsagt att arbetet var slut efter 12 (eller 18) månader. För det andra utfördes arbetet i grupper där deltagarna hade likartad bakgrund. Tanken med detta var att deltagarna skulle kunna stödja varandra både genom att ha förståelse för varandras problem och genom att vara förebilder för varandra. För det tredje stöddes arbetsgrupperna av personal som både behärskade arbetsuppgifterna och hade erfarenheter av programdeltagarnas speciella problem. För det fjärde skärptes successivt kraven på deltagarnas arbetsinsatser i syfte att så småningom nå upp till de krav som ställs på den öppna arbetsmarknaden. Ett instrument i detta syfte var ett speciellt löne- och bonussystem. Grundlönen tilläts icke understiga den lagstiftade minimilönen men låg något under marknadslönerna för icke-kvalificerad arbetskraft på orten. Till det kom en bonus som var kopplad till hur väl arbetskraven uppfylldes. Eftersom kraven skärptes successivt blev det också successivt svårare att erhålla bonus.

Denna typ av arbetsmarknadspolitiskt instrument genomfördes på 15 orter utspridda över hela USA och i 10 av dessa genomfördes randomiserade urvalsförfaranden, dvs klassiska experiment. På varje ort krävdes ett omfattande arbete

för att få projektet att fungera. Till att börja med var det nödvändigt med nära samarbete mellan forskarna och de myndigheter som kunde anvisa deltagare till programmet. Hit hörde (motsvarigheter till) socialbyråer, rehabiliteringscentra för missbrukare och kriminalvårdsmyndigheter. Vidare krävdes att potentiella programdeltagare fördelades slumpmässigt på experiment- och kontrollgrupper. Framför allt krävdes det arbetsplatser där deltagarna skulle kunna arbeta i grupper och ett antal handledare.

Slutligen krävde utvärderingen en betydande datainsamling. Till största delen utnyttjades intervjuer för att få information om inkomst- och sysselsättningsutvecklingen efter att deltagarna lämnat programmet. Dessa kompletterades med vissa registerdata, i första hand för att få bättre mått på företeelser som fortsatt brottslighet och missbruk. (Intervjumetoden ledde för övrigt till visst bortfall, men en närmare analys visade att detta sannolikt var mindre allvarligt.)

Verksamheten på de olika orterna måste dessutom samordnas så att en viss likformighet uppnåddes. Även finansieringen ställde krav på samordning. Olika departement samt Ford Foundation bidrog på central nivå med anslag, men en stor del av den lokala programverksamheten finansierades lokalt. Sammantaget var det således ett mycket stort projekt med betydande arbete i form av samordning. För att sköta denna samordning skapades MDRC (Manpower Demonstration Research Corporation). Den kompetensuppbyggnad som ägde rum under detta projekt bevarades genom att MDRC permanentades. MDRC har organiserat många utvärderingsprojekt sedan dess.

Den experimentella uppläggnings i mer statistisk mening innebar att 1 620 kvinnor med AFDC, 1 154 missbrukare, 1 458 ex-interner och 1 252 ungdomar valdes ut. Av dessa fördelades (med undantag för någon ort) hälften till experimentgrupp och hälften till kontrollgrupp. Det är viktigt att precisera i vilket stadium som det slumpmässiga urvalet skedde. I denna studie skedde detta efter en noggrann selektion av både de myndigheter som anvisade sökande och de som var ansvariga för programverksamheten; handledarna gavs möjlighet att intervjua alla deltagare och avvisa de som bedömdes olämpliga. Med andra ord skedde urvalet till experiment- och kontrollgrupp ur en grupp som både själva hade accepterat att delta och hade accepterats av programadministratörerna.



Det slumpmässiga urvalet innebar dock att bara cirka hälften av dem som bedömdes lämpliga av fältpersonalen fick delta. Resten måste ju utgöra kontrollgrupp. Detta skapade – se Hollister, Kemper och Maynard (1984, s 35) – ett visst missnöje, men det ansågs av författarna inte så allvarligt att det menligt påverkade resultaten.

### Resultat

Studierna av programmets effekter har fokuserats på om deltagarna fick högre sysselsättning och inkomster än vad de annars skulle ha fått samt om kriminellt beteende och missbruk reducerades. Något kortfattat är huvudresultaten att kvinnor med AFDC som deltagit i programmet fick en bättre sysselsättnings- och inkomstutveckling än kontrollgruppen, dvs en positiv effekt förelåg. Missbrukarna har noterats för lägre frekvens av arresteringar. Däremot kunde man inte finna några effekter på sysselsättning och inkomster för denna grupp. Inte heller påverkades användandet av droger för denna grupp trots att arresteringarna blev färre. För ex-internerna fann man heller inga effekter, men olika data-problem gjorde analysen av denna grupp svårare än för de övriga. Slutligen fann man heller inga effekter för gruppen ungdomar; deltagarna i kontrollgruppen uppvisade i stort sett samma utveckling som experimentgruppen.

Försök gjordes också att genomföra samhällsekonomiska kostnads-/intäktskalkyler av programmet. På plussidan medtogs då värdet av såväl den produktion som utfördes under själva projektet som värdet av den högre sysselsättningen och minskade brottsligheten efter projektet. På minussidan medtogs bland annat kostnaderna för handledare, lokaler m m. Kalkylerna, som förses med flera reservationer, visar på samhällsekonomiska vinster för åtgärderna riktade till grupperna kvinnor med AFDC och missbrukare, medan åtgärderna till gruppen ungdomar gav negativt utfall. För åtgärderna riktade till gruppen ex-interner var utfallet mycket oklart.

### Metodologiska begränsningar

National Supported Work Demonstration framstår i allt väsentligt som ett väl

genomfört experiment. Urvalet till experiment- och kontrollgrupperna gjordes från en population som genomgått de selektioner som både prospektiva deltagare gör (att anmäla sig till programmet) och som myndigheterna gör (att välja ut dem som är mest lämpade att delta). Därför är experimentgruppen säkert representativ för dem som skulle ha deltagit om programmet varit i normal drift.

En viss risk för Hawthorne-effekter kan ha förelegat. Det program som infördes och testades var nämligen nytt, och fältpersonalen kan därför ha varit ovan vid att hantera detta arbetsmarknads- och socialpolitiska instrument. Man kan också tänka sig att programmet hade kunnat förbättras om utvärderingsresultaten hade kunnat omsättas i praktiken.

Indirekta effekter på övriga arbetsmarknaden, inklusive kontrollgruppen, kan ha förekommit och kan också ha snedvridit resultaten. Det är emellertid omöjligt att säga något mer precist om detta.

### ***3.3 Löne-vouchers till socialbidragstagare i Dayton, Ohio, USA<sup>5</sup>***

Bakgrunden till detta projekt är att den amerikanska kongressen i samband med en omorganisation av CETA-programmet 1978 ålade Department of Labor att genomföra en studie av direkta löne-vouchers. En löne-voucher är ett certifikat som tilldelas den enskilde arbetssökanden och berättigar en arbetsgivare som anställer den sökande till en viss lönesubvention. Kongressen ålade Arbetsmarknadsdepartementet att testa instrumentet på grupper av arbetssökande som var berättigade till andra typer av service inom CETA-programmet.

Projektet genomfördes i samarbete med forskningsföretaget Mathematica Policy Research. Experimentets uppläggning var mycket enkel. Mellan december 1980 och maj 1981 utvaldes 916 deltagare som alla var bidragstagare; knappt hälften uppbar AFDC (Aid for Families with Dependent Children) och resten uppbar "general assistance payments". Dessa fördelades slumpmässigt på tre olika stora grupper. En grupp erhöll en "tax credit voucher", som innebar att den arbetsgiv-

---

<sup>5</sup>Denna presentation bygger på Burtless (1985).

are som anställde den arbetssökande var berättigad till en skattereduktion. En andra grupp erhöll en "direct cash rebate subsidy", som innebar att den arbetsgivare som anställde den arbetssökande var berättigad till en direkt subvention. För båda grupperna motsvarade subventionsbeloppen 50 procent av lönesumman under det första året (med en övre gräns på subventionen av \$3 000) och 25 procent av lönesumman det andra året (med en övre gräns på subventionsbeloppet av \$1 500). Det var administrativt ganska enkelt för företagen att erhålla dessa pengar. En tredje grupp erhöll ingen voucher alls och utgjorde därför kontrollgrupp.

Gemensamt för alla tre grupperna var att de deltog i en två veckor lång kurs som inriktades på undervisning i och träning på att söka arbete, en slags jobbsökar-klubbar. De båda experimentgrupperna fick under denna kurs också information om hur de kunde använda sina vouchers. Lärarna som förmedlade undervisningen cirkulerade mellan grupperna så att det inte skulle uppstå några systematiska "lärareffekter". På detta sätt framtogs också myndigheterna möjligheten att fördela de bästa lärarna på experimentgruppen så att resultaten skulle framstå som så bra som möjligt.

### Resultat

Resultaten från experimentet kan presenteras på ett mycket enkelt sätt utan några komplicerade statistiska eller ekonometriska formler. Faktum är att Tabell 3.1 är den enda dokumentation som publicerats om projektets resultat. Burtless har i ett annat sammanhang (Burtless 1987) hävdats att en av fördelarna med experiment är att resultaten kan förstås mycket enkelt även av dem som inte har statistiska metodkunskaper.

Som framgår av tabellen är resultaten minst sagt anmärkningsvärda. De som fått en förmån i form av en löne-voucher har ett klart sämre utfall än de som tillhör kontrollgruppen! Skillnaden är klart signifikant vilket betyder att det är högst osannolikt att slumpen kan förklara skillnaderna mellan grupperna. Där- emot är utfallen för de båda typerna av vouchers i stort sett desamma, vilket tyder på att det är likgiltigt i vilken form företagen erhåller stöden.

Tabell 3.1      Sysselsättningstal för experiment- och kontrollgrupper i Dayton

Grupp	Sampelstorlek	Antal placerade i arbete	Andel placerade
Tax credit voucher	247	32	13,0 %
Direct rebate subsidy	299	38	12,7 %
Kontrollgrupp	262	54	20,6 %
Totalt	808	124	15,3 %

Källa: Burtless (1985).

Noter: Sysselsättningstalet avser om arbete erhållits under en period av 8 veckor efter kursen i sökverksamhet.

De 808, av ursprungligen 916 deltagare, som ingår i resultatredovisningen är de som fullföljt kursen i sökverksamhet.

Frågan är hur man kan förklara det paradoxala resultatet att det gått sämre för dem som kunnat erbjuda arbetsgivarna ett ekonomiskt lockbete i form av en voucher. Någon mekanism måste inte bara ha uppvägt utan också varit starkare än den ekonomiska fördel som ligger i lönesubventionen. Burtless' förklaring bygger på att en löne-voucher också kan fungera som en signal för arbetsgivaren. När den arbetssökande presenterar sin voucher för en presumtiv arbetsgivare måste han även förklara varför just han kan erbjuda arbetsgivaren en sådan. Orsaken är ju att den sökande är socialbidragstagare. Detta betyder att arbetsgivaren får en typ av information om den sökande som han annars kanske inte hade fått. Om denna information uppfattas negativt på grund av att socialbidragstagare oftare har olika sociala problem än andra kan en voucher som utdelas efter sådana kriterier fungera som en negativ signal. Som Burtless skriver kan en löne-voucher i detta sammanhang fungera som en signal att den sökande utgör "damaged goods".

Av studien framgick även att bara en mindre del – ca en tredjedel – av de arbetsgivare som i princip skulle ha varit berättigade till lönesubventioner (70 stycken enligt tabellen) faktiskt krävde att få sina pengar. En förklaring kan vara att flera av de sökande med vouchers avstod från att utnyttja dessa. Enligt Burtless föreligger informella rapporter från den personal som genomfört experimentet, att vissa sökande avsiktligt avstod från att utnyttja sin voucher. Detta kan i sin tur bero på att de förutsåg att arbetsgivarna skulle uppfatta dessa som en negativ signal.

Resultat av detta slag ger i första hand inte upphov till metodologiska spekulationer utan snarare till arbetsmarknadspolitiska funderingar: Hur generellt kan resultatet vara? Kan liknande mekanismer ge negativa effekter även av andra typer av lönesubventioner? Kan resultaten vara giltiga även i andra länder? Detta kan åtminstone inte uteslutas. Eftersom lönesubventioner förekommer i betydande omfattning i många länder måste denna studie utgöra en varning till beslutsfattare i arbetsmarknadspolitiska frågor. Det bästa sättet att reagera på denna varning är självfallet att med hjälp av experiment utvärdera existerande lönesubventioner i andra länder.

En annan bedömning kan vara att lönesubventioner bör användas för att kompensera för synliga och uppenbara handikapp snarare än för osynliga och mindre uppenbara sådana. Om personer med påtagliga fysiska handikapp kan uppvisa en voucher behöver denna inte uppfattas som en signal om osynliga problem. Möjligen är det så att lönesubventioner kan fungera väl när de avser personer med synliga och påtagliga handikapp men att de kan ge negativa effekter om de avser personer med osynliga och mindre påtagliga handikapp.

### **3.4 Bonus till arbetslösa i Illinois, USA<sup>6</sup>**

Bakgrunden till detta experiment var en långvarig kritik av systemet för arbetslöshetsförsäkring och en otillfredsställelse med resultaten från icke-experimentella studier av hur arbetslöshetsunderstöd påverkar arbetslösas sökbeteende. Department of Employment Security i delstaten Illinois beslöt därför att med

---

<sup>6</sup>Denna presentation bygger på Woodbury och Spiegelman (1987).

hjälp av ett klassiskt experiment undersöka hur väl bonusbetalningar till arbetslösa kan fungera som instrument att reducera arbetslöshetens längd. Projektet genomfördes i samarbete med W.E. Upjohn Institute for Employment Research.

Mer specifikt testades två typer av premier till arbetslösa. Den ena innebar att en nybliven arbetslös med rätt till ersättning från arbetslöshetsförsäkringen erbjöds en extra bonus på \$500 om han eller hon inom 11 veckor skaffade ett arbete. Två krav på arbetet var dock att det skulle innehas under minst fyra månader och omfatta minst 30 arbetstimmar per vecka. Den andra premien innebar att den arbetslöse i stället erhöll en löne-voucher värd \$500 som han kunde ge till en arbetsgivare som anställde honom inom 11 veckor. Kraven på arbetet var desamma som i fallet med bonus till den arbetslöse själv. Storleken på bonusen respektive vouchern motsvarade ungefär fem procent av en årslön eller fyra veckors arbetslöshetsunderstöd. Beloppet bestämdes bland annat av den budget på \$750 000 som fanns att tillgå för premieutbetalningar till företag och arbetslösa.

För att vara berättigad till antingen ett erbjudande om bonus eller en voucher krävdes att individen (i) hade påbörjat en period med arbetslöshetsunderstöd mellan 29 juli och 17 november 1984, (ii) var berättigad till 26 veckor med ersättning, (iii) var registrerad vid någon av 22 olika Job Service Offices i norra eller centrala Illinois och (iv) var mellan 20 och 55 år. De som uppfyllde dessa villkor fördelades slumpmässigt på tre grupper; en experimentgrupp som erbjöds en bonus, en experimentgrupp som erbjöds en voucher, och en kontrollgrupp som inte fick något extra erbjudande utöver det reguljära arbetslöshetsunderstödet som självfallet även utgick till de båda experimentgrupperna.

Data om inkomster och bidrag både före och efter arbetslöshetsperioden insamlades från olika register. På detta sätt undveks det bortfall som intervju-tekniken normalt ger upphov till.

### Resultat

Studien ger en bred belysning av effekterna av extra incitament till intensifierat arbetssökande. Några av resultaten är sammanställda i tabell 3.2. Den första raden i tabellen talar om att cirka 4 000 individer återfanns i varje grupp.

Tabell 3.2 Några resultat från experimentet i Illinois

	Kontroll- grupp	Experiment- grupp som erbjudits bonus	Experiment- grupp som erbjudits voucher
Sampelstorlek (antal berättigade till stöden för experimentgrupperna)	3 952	4 186	3 963
Andel som accepterade erbjudandet att delta	–	0,84	0,65
Andel som erhöll bonus	–	0,14	0,03
Utbetalat arbets- löshetsunderstöd	2 558	2 329***	2 446**
Antal arbets- löshetsveckor	18,3	17,0***	17,7**
Andel vars understöd upphörde inom 11 veckor	0,353	0,408***	0,384**
Kvot mellan inbesparade understöd och utbetalad bonus		2,32 <sup>a</sup>	4,29 <sup>b</sup>

Källa: Woodbury och Spiegelman (1987).

a Signifikant skild från ett.

b Ej signifikant skild från ett.

\*\*\* Hypotesen att skillnaden mellan experiment- och kontrollgrupperna är noll kan förkastas på 1 procents signifikansnivå.

\*\* Motsvarande hypotes kan förkastas på 5 procents signifikansnivå.

Ett första resultat av intresse är att en betydligt högre andel accepterade erbjudandet om en bonuspremie än erbjudandet om en löne-voucher att ge till en arbetsgivare; 84 procent mot 65 procent. Skillnaden blev relativt sett ändå större mellan de båda experimentgrupperna när man kunde konstatera att 14 procent av dem som erbjöds möjligheten till en bonus också erhöll en sådan, medan motsvarande tal för gruppen med löne-vouchern var 3 procent. Två faktorer bidrog till att dessa tal blev så låga. För det första klarade många inte av att hitta ett arbete inom 11 veckor och blev därför inte berättigade. För det andra visade det sig att förvånansvärt många av dem som var berättigade att erhålla bidrag avstod från att kräva dem; bara 54 procent av de arbetslösa som kvalificerade sig för en bonusbetalning krävde att få sina pengar och bara 12 procent av de företag som var berättigade till pengar för den voucher de skulle ha erhållit från de nyanställda krävde att få dessa belopp. Det senare är ganska märkligt och kan bero på ovana med de administrativa rutinerna för utbetalningarna. Man kan befara att detta berodde på att experimentet var kortvarigt och att utfallet skulle bli annorlunda om det fått pågå en längre tid.

De tre följande raderna belyser på tre olika sätt hur arbetslöshetstiderna har påverkats av experimenten. Notera att siffrorna för experimentgrupperna är genomsnitt för dem som erbjudits möjligheten att utnyttja bonus eller voucher. Det visar sig att de båda experimentgrupperna fick lägre arbetslöshetsunderstöd, hade kortare genomsnittliga arbetslöshetstider och i större utsträckning lämnade arbetslösheten (upphörde att vara understödstagare) inom 11 veckor än kontrollgruppen. Effekterna var något kraftigare för gruppen som erbjöds bonus.

Senare, mer ingående, analys av materialet (Meyer (1988)) har övertygande visat att sannolikheten att lämna arbetslösheten ("hazarden") var högre för experimentgruppen<sup>7</sup> än för kontrollgruppen under de 11 första veckorna. Däremot kunde inga skillnader spåras efter den tidpunkt då bonusen upphörde att gälla.

Slutsatsen är således att extra incitament att skaffa arbete har betydelse för sökbeteendet. Utifrån traditionell sökteori kan man förvänta sig att dessa effekter uppkommit av två skäl; antingen genom att reservationslönerna sänkts i syfte

---

<sup>7</sup>Enbart gruppen som erhöll individuell bonus ingick i analysen.



att snabbt få ett jobb eller genom att sökintensiteten höjts eller en kombination av dessa mekanismer. För att undersöka detta närmare studerades också inkomsterna på de nya jobben för de olika grupperna. Det visade sig då att det inte förelåg några skillnader mellan grupperna, varför det är osannolikt att effekterna uppkommit genom att reservationslönerna sänkts. Den troligaste förklaringen är i stället att sökintensiteten har höjts.

Beräkningar har också gjorts av hur kostnaderna förhåller sig till intäkterna av de båda premieformerna. Till kostnadssidan hänförs då utbetalningar av premierna, medan inbesparade arbetslöshetsunderstöd utgör intäktssidan. Som framgår av sista raden i tabellen är kvoten mellan intäkter och kostnader klart större än 1,0 för båda experimenten. Trots att kvoten är störst för vouchergruppen är denna kvot ej signifikant skild från ett. Dessa beräkningar bygger dock på att enbart faktiska utbetalningar av premier ingår på kostnadssidan. Om man i stället utgår från att alla som vore berättigade till en premie också skulle ha utnyttjat denna möjlighet, vilket kanske skulle skett om de administrativa rutinerna förenklats, skulle enbart bonusexperimentet givit intäkts-/kostnadskvot högre än ett.

En slutsats av detta experiment är att sökbeteendet påverkas av ekonomiska incitament. Detta betyder dock inte att sökteorin i sig har testats. För att göra detta krävs att de explicita beteendesamband som följer av teorin blir föremål för test.

En annan slutsats är att effekten av att erbjuda en bonus, som går direkt till den arbetslöse, är större än effekten av en löne-voucher som utbetalas till en arbetsgivare. Det relativt låga intresset för löne-vouchern tyder på att många kanske inte tror att sådana kan bidra till att få ett arbete. Nettoeffekten för gruppen som erbjudits en voucher är dock positiv, låt vara ej kraftigt statistiskt signifikant skild från noll. Det betyder att denna studie givit ett annat resultat än experimentet i Dayton Ohio. En förklaring till denna skillnad kan vara gränsen på 11 veckor som skapar ytterligare incitament att snabbt skaffa ett arbete. En annan förklaring kan vara att de båda experimenten riktar sig till olika grupper: socialbidragstagare i Dayton och mottagare av arbetslöshetsunderstöd i Illinois. Det är betydligt mer stigmatiserande att erhålla socialbidrag än arbetslöshetsunderstöd i USA. En löne-voucher som avslöjar för en arbetsgivare att den

arbetssökande uppbär socialbidrag kan därför ha negativa effekter medan en voucher som avslöjar att den arbetssökande uppbär arbetslöshetsförsäkring kan ha positiva effekter.

Det bör också betonas att de effekter som presenterades i tabellen utgör effekter av att erbjuda en voucher. De effekter som då uppstår är rimligen koncentrerade till dem som faktiskt har accepterat erbjudandet att utnyttja denna möjlighet i sitt sökande. Som framgick av tabellen avstod 16 procent av dem som erbjöds att söka med möjlighet till en bonus och hela 35 procent av dem som erbjöds en voucher.

Generellt gäller att skillnaden i utfall mellan en experimentgrupp som erbjudits en förmån och en kontrollgrupp som icke erbjudits denna förmån är: (Andelen som accepterat förmånen) gånger (Effekten för dem som utnyttjat förmånen). Detta betyder att medan effekten på arbetslöshetstiderna av att erbjuda en bonus till en grupp arbetslösa är 1,3 veckor (18,3–17,0) är effekten för dem som faktiskt utnyttjat den 1,55 veckor (1,3 dividerat med 0,84). På motsvarande sätt är effekten av att erbjuda en voucher 0,6 veckor (18,3–17,7) medan de som faktiskt utnyttjat vouchern är 0,92 veckor (0,6 dividerat med 0,65).

#### Metodologiska begränsningar

Ett förvärande resultat var att så få arbetssökande och företag som var berättigade till ersättning faktiskt krävde att få denna. Detta kan bero på att de administrativa rutinerna för att hantera dessa vouchers inte hade hunnit utvecklas ordentligt under den relativt korta tid som experimentet pågick. Detta innebär i synnerhet att kostnads-/intäktskalkylerna av programmet måste tolkas med stor försiktighet; kostnaderna reducerades ju påtagligt av att många inte krävde den ersättning de var berättigade till.

En reservation måste också göras för eventuella indirekta effekter. Om sökintensiteten ökar påtagligt bland de arbetssökande kan i första hand lönestrukturen, men kanske också lönenivån, påverkas. Det är emellertid troligt att sådana effekter uppkommer efter en viss tid och inte visade sig under den tid experimentet pågick. Med nationalekonomisk terminologi är det troligt att experimen-

tet i första hand har belyst "partial-equilibrium-effects" snarare än "general-equilibrium-effects". Det har emellertid ett stort värde att få även sådana mer avgränsade effekter belysta. Eftersom dessa partiella effekter var positiva kan det finnas anledning att i ett nästa steg pröva bonusbetalningar i större skala för att söka belysa eventuella indirekta effekter.

Även effekter på kontrollgruppens möjligheter att få jobb kan tänkas, s k displacement-effekter. Ett experiment av detta slag kan inte fånga upp sådana effekter. I ett av tre nya liknande experiment i USA planeras emellertid en jämförelse med andra regioner för att kunna spåra eventuella effekter av detta slag. Med tanke på de problem som sådana regionala jämförelser är förknippade med (se avsnitt 5), kan det emellertid bli svårt att finna indirekta effekter med rimlig precision.

Slutligen finns skäl att betona att experimentet inte var tillräckligt omfattande för att belysa hur ett system med bonus för framgångsrikt arbetssökande kommer att fungera om det permanentas. Både arbetsgivare och arbetssökande får incitament att ändra sitt beteende. Arbetsgivare får större incitament att säga upp personal ofta för att kunna dra nytta av bonusbetalningar. Arbetssökanden som aldrig ansöker om arbetslöshetsunderstöd får incitament att göra detta och i vissa fall kan det t o m bli lönande att något dröja med att acceptera ett arbete för att först se till att man blir berättigad till bonusersättning.

Trots dessa metodologiska begränsningar har detta experiment givit värdefull kunskap om incitamentens betydelse för arbetslösas sökverksamhet.

### **3.5 *Employment Opportunity Pilot Project (EOPP)*<sup>8</sup>**

Detta projekt är inte ett renodlat experiment men är ändå av betydande intresse. I synnerhet de ursprungliga planerna för projektet är lärorika, även om förändrade politiska förutsättningar ledde till att planerna fick överges.

---

<sup>8</sup>Denna presentation bygger i första hand på Mathematica Policy Research (1983) och i andra hand på Burtless och Haveman (1984).

Syftet var från början att testa ett förslag från Carter-administrationen som gick ut på att erbjuda socialbidragstagare ett garanterat arbete. Rent politiskt var moralen bakom förslaget att de socialbidragstagare som kunde arbeta också borde göra det och att de som inte kunde finna ett arbete skulle erbjudas antingen ett subventionerat arbete – i första hand i den offentliga sektorn – eller en utbildning. Förslaget hade två delar. Först skulle deltagarna genomgå en period med intensivt sökande. Denna period innehöll också arbetsprövning och en ambitiös träning i att söka arbete. Träningen innehöll komponenter som övning i att kontakta arbetsgivare, att bli intervjuad, att skriva ansökningar och övningar som syftade till att öka motivationen och höja självförtroendet. Det skulle också finnas en möjlighet att genomföra denna del av programmet i så kallade jobb-sökar-klubbar.

Om denna sökverksamhet inte givit resultat i form av ett reguljärt arbete efter fem till åtta veckor skulle programmets andra del träda in. Då skulle nämligen ett arbete eller en utbildning kunna erbjudas. Efter ett år skulle emellertid detta upphöra och deltagarna skulle återigen genomgå en period med aktiv sökverksamhet.

Att utvidga ett system med socialbidrag med ett program av detta slag är självfallet en betydande uppgift med många potentiella problem. Syftet med EOPP var att pröva programmet på 14 orter och därvid studera hur det fungerar innan det skulle införas generellt över hela landet. De frågor som aktualiserades av förslaget var bland annat följande:

- o Hur väl fungerar den första delen av programmet med aktiverat sökande? Vilka former av sökande fungerar bäst?
- o Hur många kommer att erhålla jobb redan under den första fasen och hur stort kommer behovet av subventionerade arbets- och utbildningsplatser att bli?
- o Kommer det att bli möjligt att få fram de arbets- och utbildningsplatser som krävs?
- o Kommer arbetena att vara produktiva eller bara utgöra passiva förvaringsplatser?

- o Kommer programmet att förbättra deltagarnas arbetsmarknadssituation på lång sikt?
- o Vilka krav kommer att ställas på annan samhällsservice såsom barnomsorg, transporter och rådgivning?
- o Hur kommer arbetsmarknaden i övrigt att påverkas? Kommer möjligheterna för andra låginkomsttagare att få arbete att försvåras eller underlättas av de jobb som skapas för programdeltagarna?

Ur metodologisk synvinkel är det intressanta med detta projekt att det utvecklades en forskningsplan för att belysa dessa frågor. I denna plan ingick bland annat ett urval av tio jämförelseorter där programmet inte infördes. Genom att studera experiment- och jämförelsegrupperna skulle det bland annat bli möjligt att belysa indirekta effekter av olika slag. En jämförelse av arbetslösheten för låginkomsttagare mellan de båda grupperna skulle medge bedömningar av om indirekta effekter förekommit och vilken riktning de haft.

Tyvärr fullföljdes inte dessa planer på grund av att det politiska intresset för förslaget minskade. Redan under Carter-administrationens tid ändrades bland annat reglerna för tillträde till programmet flera gånger, vilket försvårade utvärderingen. Den stora förändringen kom emellertid när Reagan-administrationen tillträdde 1981. Intresset för offentlig sysselsättning minskade då och "public service employment" hade varit det centrala instrumentet för att upprätthålla jobb-garantin. Däremot kvarstod intresset för att utvärdera programmets första del, varför denna del av utvärderingsprojektet fullföljdes. Problemet var dock – enligt Burtless och Haveman (1984) – att forskningsplanen var mindre god när det gällde att utvärdera denna del, medan dess stora förtjänster låg i att utvärdera ett program med inslag av jobb-garanti. De resultat som föreligger är därför inte baserade på experimentell uppläggning, utan har producerats med icke-experimentell metodik baserad på i vissa avseenden bristfälliga data.

### Resultat

Programmet beräknas ha höjt sysselsättningen med 10–12 procentenheter för

ogifta kvinnor vilka utgjorde programmets största grupp. Däremot kunde någon minskning i bidragsberoendet inte samtidigt spåras. Positiva sysselsättnings-effekter erhöles även för andra grupper av låginkomsttagare än de som direkt deltagit.

Ett försök att göra en samhällsekonomisk kostnads-/intäktskalkyl gav till resultat att programmet sannolikt innebar en kortsiktig subvention från skattebetalare till programdeltagare. Om däremot de positiva sysselsättningseffekterna skulle kvarstå en längre tid kan resultatet ha blivit positivt även för skattebetalarna.

### *3.6 JTPA-experimentet i USA<sup>9</sup>*

1982 ersattes det amerikanska CETA-programmet, vars resultat presenterades ovan, med Job Training Partnership Act (JTPA). Sedan dess är JTPA den federala regeringens primära medel för att till låginkomsttagare förmedla arbetsrådgivning ("job search assistance"), arbetsmarknadsutbildning och arbetserfarenhet ("remedial education and classroom and on-the-job occupational skills"). Programmet stöds av det amerikanska arbetsmarknadsdepartementet och administreras av 600 lokala myndigheter ("Service Delivery Areas").

Kravet på ordentlig utvärdering finns inskrivet i lagstiftningen som säger att utvärderingen ska inriktas på de primära målen: att höja deltagarnas arbetsinkomster och minska deras bidragsberoende.

Mot bakgrund av erfarenheterna av utvärderingarna av CETA-programmet beslöts 1985 att genomföra en ambitiös satsning på klassiska experiment. 20 miljoner dollars avsattes för detta utvärderingsprojekt (Stromsdorfer (odaterad)). För närvarande föreligger enbart planer för projektet. Vi ger ändå en presentation av det eftersom det principiellt är mycket intressant; det är en utvärdering av ett pågående program som är inriktat på utbildning.

---

<sup>9</sup>Denna presentation bygger på Stromsdorfer (odaterad), Bangsor, Borus, Bloom och Orr (odaterad), Bloom, Borus, Orr (1987) och Barnow (1988).

Enligt planerna skall experimenten genomföras i 15–20 områden (Service Delivery Areas) och sammanlagt skall 30 000 individer under en period av 18 månader fördelas slumpmässigt på experiment- och kontrollgrupper. Mer specifikt kommer urvalet att gå till på följande sätt. Först ansöker enskilda individer om att få delta i JTPA. Därefter gör myndigheten en bedömning av om individen är berättigad att delta. Under de intervjuer som krävs för att göra denna bedömning informeras de sökande om utvärderingsprojektet och den slumpmässiga fördelningen. De får också underteckna ett medgivande ("informed consent statement") att de varit korrekt informerade om projektet och att information från offentliga register (främst inkomster och bidrag) får användas av forskarna.

De som därefter bedöms vara lämpade och berättigade att delta i JTPA fördelas slumpmässigt på en experimentgrupp som tillåts delta och en kontrollgrupp som ej tillåts delta. Två tredjedelar fördelas till experimentgruppen och en tredjedel till kontrollgruppen.

Eftersom JTPA egentligen innehåller olika typer av service – "classroom training", "on-the-job-training" och övrig service – sker fördelning på grupper bland dem som bedömts vara behöriga till en viss servicetyp. Den slumpmässiga fördelningen sker alltså för varje servicetyp, medan myndigheterna har beslutat om fördelningen på typ av service.

Efter den slumpmässiga fördelningen insamlas information under 30 månader för båda grupperna. Utöver arbetsinkomster, sysselsättning och bidrag planeras även att följa upp erfarenheterna vad gäller familjeförhållanden, fertilitet och fortsatt utbildningsverksamhet.

Till utvärderingen av JTPA hör även datainsamling till en icke-experimentell utvärderingsstudie samt en separat studie av hur selektionsprocessen går till. Vi återkommer i avsnitt 4 till hur experimentella och icke-experimentella studier kan kombineras på ett mycket konstruktivt sätt.

Frågan är nu om man kan finna några av experimentens svagheter i denna uppläggningsstudie. Ett problem har visat sig redan i en tidig fas; nära 70 procent av de utvalda områdena har vägrat att acceptera en slumpmässig fördelning (enligt Barnow (1988)). Det betyder att experimenten genomförs på ett skevt urval av

orter, nämligen de som avviker från övriga i och med att de accepterar experimentet. Man kan mycket väl tänka sig att denna skevhet innebär att enbart orter där de ansvariga vet med sig att programmen fungerar väl ställer upp. Bortfallet innebär i varje fall att resultaten inte kan generaliseras att gälla hela landet. Varje genomfört experiment kan däremot belysa effekten av JTPA på den egna orten.

Ytterligare ett problem har att göra med hur representativ experimentgruppen är för dem som deltar när JTPA är i normal drift. Om experimentgruppen är ett slumpmässigt urval av en grupp som skulle ha deltagit i normala fall är allt gott och väl. I så fall innebär denna design till experiment av ett pågående program att en tredjedel av dem som skulle ha deltagit i normala fall icke tillåts delta. Under experimentets gång måste därför programmet reduceras i omfattning för att "skapa plats för" kontrollgruppen.

Enligt Barnow (1988) är detta dock ej fallet. I stället har programadministratörerna ombetts att välja ut fler som behöriga för att experimentgruppen ska ha samma storlek som deltagargruppen när programmet är i normal drift. Effekten mäts därför för en större grupp än vad som normalt deltar i programmet. Därför mäter man effekten av ett större program än det som är i normal drift. Frågan är om detta har någon betydelse. Svaret är att det har betydelse om två villkor är uppfyllda samtidigt (se Björklund (1988)). För det första krävs att det finns en spridning i effekternas storlek mellan olika individer. För det andra krävs att programadministratörerna har en viss förmåga att rangordna individerna efter effekternas storlek och att de också utnyttjar denna möjlighet. I så fall kommer det tillskott av behöriga individer som väljs ut på grund av experimentet att ha lägre effekter än den grupp som skulle ha valts ut även om programmet varit i normal drift. Hur stort detta problem är i praktiken är dock svårt att säga.

### **3.7 Workfare-experimenten i USA**

Bakgrunden till dessa experiment är ett generellt problem för alla socialpolitiska system. Å ena sidan är det önskvärt att garantera alla människor en rimlig ekonomisk standard, men å andra sidan anses det också önskvärt att de som verkligen kan försörja sig själva ges incitament att göra detta i stället för att bli beroende av bidrag.



I USA har en av de stora socialpolitiska frågorna varit vilka krav som ska ställas på de ensamstående kvinnor som omfattas av AFDC-programmet (Aid for Families with Dependent Children). Detta infördes 1935 och då var syftet att ge ensamma mödrar den ekonomiska möjligheten att helt koncentrera sig på uppgiften som vårdnadshavare i hemmet (se Burtless (1988) och Gueron (1987)). Under 1960-talet förändrades inställningen till AFDC, bland annat till följd av en förändrad syn på kvinnans ställning i samhället men också som en följd av att detta bidragsprogram ansågs skapa ett onödigt beroende av bidrag. År 1967 kom därför WIN-programmet (WIN = Work Incentive) som dels innebar att reglerna för bidragets storlek gav ökade incitament till arbete, dels innehöll åtgärder för arbetsrådgivning och utbildning för bidragstagare med äldre barn.

"Arbetslinjen" förstärktes ytterligare 1981 då en ny lagstiftning gjorde det möjligt för delstaterna att experimentera med hårdare krav på arbete och utbildning för att erhålla bidrag än vad som tidigare varit fallet. Exempelvis kunde delstaterna kräva att bidragstagare vars yngsta barn var tre år gammalt skulle kunna acceptera erbjudande om arbete eller arbetsmarknadsutbildning. Åtskilliga stater har utnyttjat denna möjlighet och i synnerhet har de hårdare kraven på arbete införts i experimentell form på ett flertal platser i USA. Forskningsföretaget MDRC (Manpower Demonstration Research Company) har genomfört dessa experiment. Sammantaget genomfördes relativt storskaliga experiment i åtta stater under perioden 1982–87 och mindre experiment på ytterligare tre mindre orter. Experimenten går under beteckningen workfare-experimenten där workfare står för ett program som knyter krav på arbete till ekonomiska bidrag.

Den faktiska utformningen av de arbetsmarknadspolitiska insatserna varierade kraftigt mellan de olika staterna, varför experimenten inte är jämförbara (se Gueron (1987)). För att illustrera den metodologiska uppläggningsen väljer vi att kort presentera den studie som genomfördes i San Diego, Californien, (se Goldman m fl (1986)).

I San Diego prövades två typer av obligatoriska program. Det ena innebar enbart obligatoriskt arbetssökande ("mandatory job search") i strukturerad form under tre veckor. Den första veckan erhöll bidragstagaren allmän orientering om arbetsmarknaden och om hur man bör söka arbete. De två andra veckorna innebar aktivt sökande i "jobb-sökar-klubbar".

Tabell 3.3 Några resultat från workfare-experimentet i San Diego

	<u>Job search/Work exp</u>			<u>Job search</u>		
	Experiment- grupp	Kontroll- grupp	Skillnad	Experiment- grupp	Kontroll- grupp	Skillnad
Andel sysselsatta under						
1:a kvartalet	35,5	33,1	+2,4	32,9	33,1	-0,2
2:a "	35,6	28,7	+6,9***	37,2	28,7	+8,5***
3:e "	40,2	32,3	+7,8***	36,9	32,3	+4,6**
4:e "	42,4	36,9	+5,5***	38,4	36,9	+1,6
5:e "	42,9	37,5	+5,4***	37,9	37,5	+0,4
6:e "	41,9	38,1	+3,8*	37,4	38,1	-0,7
Inkomster under						
1:a kvartalet	359	337	+22	368	337	+31
2:a "	510	369	+141***	487	369	+118***
3:e "	701	538	+163***	656	538	+118**
4:e "	810	693	+117**	669	693	-24
5:e "	848	729	+119**	742	729	+13
6:e "	933	773	+160***	799	773	+26
Andel som erhöll AFDC under						
1:a kvartalet	78,3	80,3	-2,0	79,5	80,3	-0,7
2:a "	64,2	67,6	-3,4*	66,2	67,6	-1,4
3:e "	51,8	56,2	-4,5**	52,2	56,2	-4,0*
4:e "	45,8	47,9	-2,0	45,5	47,9	-2,4
5:e "	39,5	41,1	-1,7	42,3	41,1	+1,2
6:e "	35,0	36,2	-1,2	36,2	36,2	+0,0

Källa: Goldman m fl (1986).

- \* = signifikant skilt från noll på 10% nivå
- \*\* = signifikant skilt från noll på 5% nivå
- \*\*\* = signifikant skilt från noll på 1% nivå

Det andra programmet innebar ett krav att efter den första treveckorsperioden med arbetssökande delta i obetalt arbete upp till 13 veckor i en offentlig eller privat icke-vinstdrivande organisation. Arbetstiden bestämdes som det ekonomiska bidraget dividerat med den lagstiftade minimumlönen.

Den metodologiska uppläggnings kunde göras mycket enkel. De som ansökte om bidrag under perioden oktober 1982 till augusti 1983 indelades slumpmässigt i två experimentgrupper och en kontrollgrupp; en experimentgrupp ålades att delta i programmet med enbart arbetssökande, en experimentgrupp ålades att delta i det utvidgade programmet med arbete för att erhålla bidrag, och kontrollgruppen behandlades på samma sätt som tidigare.

Några av resultaten presenteras i tabell 3.3. Det framgår att den grupp som även ålades att arbeta i 13 veckor (job search/work exp.) genomgående har ett bättre utfall än kontrollgruppen. Såväl andelen sysselsatta, inkomsterna som andelen med bidrag uppvisar genomgående bättre resultat för denna experimentgrupp än för kontrollgruppen under hela uppföljningsperioden. Däremot finns inga påtagliga förbättringar för gruppen som enbart ålades obligatoriskt arbetssökande.

Ett genomgående resultat från de amerikanska Workfare-experimenten är att krav på arbete har ökat arbetsinkomsterna och något reducerat bidragsberoendet. Någon drastisk förbättring av bidragstagarnas ekonomiska situation har det dock inte varit fråga om. De sammanfattande bedömningar som Burtless (1988) och Gueron (1987) gjort är att dessa åtgärder varit "kostnadseffektiva" utan att för den skull utgöra någon lösning på de grundläggande socialpolitiska problemen. Helt klart är enligt deras uppfattning att experimenten givit värdefull kunskap.

### **3.8 Intensifierad förmedlingsservice, Eskilstuna<sup>10</sup>**

Huvudsyftet med detta projekt – den s k Eskilstunastudien – var att mäta, i

---

<sup>10</sup>Denna presentation bygger på Delander (1978).

både kvantitativa och kvalitativa termer, det bidrag en utökad förmedlings-service kan ge när det gäller att lösa arbetsmarknadsproblem för långtidsarbetslösa. Undersökningspopulationen, som uppgick till ca 400 personer, utgjordes av de arbetslösa sökande som varit registrerade vid arbetsförmedlingens distriktskontor i Eskilstuna tre månader eller längre. Kontoret i Eskilstuna erhöll en personalförstärkning under den tid försöksverksamheten pågick, 10 mars – 6 juni 1975. Experimentgruppen utnyttjade i genomsnitt förmedlingens tjänster i 7,5 timmar under försöksperioden jämfört med 1,5 timmar i genomsnitt för kontrollgruppen. Den senare gruppen erhöll en service som var av normal omfattning, varför studien är inriktad på att mäta effekterna av just utökad förmedlings-service.

Som grund för arbetet i experimentgruppen utarbetades ett handlingsprogram med stark tonvikt vid en utökad individuell information och rådgivning till såväl arbetssökande som företag. Programmet beskrivs i Delander (1978) s 192, på följande sätt:

1. Företagen bearbetas för att ändra sina krav på arbetskraften genom intensifierad information om de arbetssökandes kapacitet och förutsättningar.
  - a) Mottagning av personalrekryterare på arbetsförmedlingen (af) varvid presentation av arbetssökande sker med hjälp av arbetssökandebblankett och förmedlarens kunskaper.
  - b) Information om enskilda arbetssökande till af:s representanter i anpassningsgrupper på företag som söker personal.
2. Rörligt förmedlingsarbete i syfte att påverka de arbetssökandes och arbetsgivarnas preferenser, innebärande utökad information genom:
  - a) Intensiv direktkontakt mellan arbetssökande, arbetsförmedlare och företag, vilket bl a kan innebära att arbetsförmedlare introducerar arbetssökande på företag.
  - b) Gruppverksamhet för arbetslösa bl a i syfte att påverka sökandes yrkespreferenser. Studiebesök på arbetsplatser.

3. Efterfrågepåverkan med hjälp av lönekostnadsanknutna subventioner.
  - a) Ackvisition av halvskyddade platser, bidrag till arbetsbiträde, arbetstekniska hjälpmedel, särskilda anordningar på arbetsplatser och motorfordon.
  - b) Provanställning, enskild utbildning i företag för äldre och handikappade, köp av ledig utbildningskapacitet.
  - c) Bidrag till arbetsgivare som anställer och utbildar män respektive kvinnor i yrken som domineras av arbetstagare av motsatt kön.

De slumpmässigt utvalda experiment- och kontrollgrupperna kom att omfatta 216 respektive 194 personer. Information om dessa insamlades vid försöksperiodens början, vid dess slut och ungefär 9 månader senare med hjälp av en postenkät, varvid bortfallet bara blev 10 procent.

### Resultat

Överlag erhöles mycket påtagliga positiva effekter av denna försöksverksamhet. En översikt över de viktigaste resultaten presenteras i tabell 3.4. De fyra första raderna anger olika kvantitativa mått på försöksverksamhetens effekter, såsom sysselsättning och arbetslöshet. Alla måtten pekar på klara skillnader till experimentgruppens förmån, dvs positiva effekter. Det är intressant att notera från raderna 5–7 att de kvantitativa förbättringarna inte påverkat kvaliteten i arbetsplaceringarna i form av lön, anställningstrygghet och arbetsmiljö. Tvärtom har förbättringar uppkommit både när det gäller snabbheten och kvaliteten i placeringarna. Överlag rapporteras dessa skillnader mellan grupperna vara signifikant skilda från noll, dvs det är osannolikt att det är slumpen som orsakat skillnaderna mellan grupperna.

Delander genomförde också några mer övergripande kostnads-/intäktskalkyler av insatsernas lönsamhet. Även dessa gav ett kraftigt positivt utfall. Kostnaderna understeg inte bara det produktionsvärde som skapades utan de understeg

även den del av produktionsvärdet som tillföll "övriga samhället", dvs de som inte fick del av den intensifierade förmedlingsverksamheten.

#### Metodologiska begränsningar

En första invändning mot dessa mycket kraftiga positiva effekter är att experimentgruppen kan ha gynnats på kontrollgruppens bekostnad. Detta är ett typfall av experiment där så kallad kö-bias kan uppstå. Delander diskuterar denna möjlighet och påpekar att efterfrågetrycket var starkt vid tiden för experimentet, varför konkurrensen om de lediga platserna kanske var begränsad. Åtminstone kan man förmoda att kö-bias hade varit ett allvarigare problem i en djup lågkonjunktur.

Ett sätt att undersöka detta hade varit att studera tidsserier över "sannolikheten att lämna arbetslösheten givet att den pågått 3 månader" för Eskilstuna-distriktet. Om ett skift kan noteras för denna variabel under försöksperioden, som motsvarar den effekt som jämförelsen mellan grupperna angav, är det inte sannolikt att effekterna uppnåtts på kontrollgruppens bekostnad. Huruvida sökandestatistiken vid denna tidpunkt hade den kvalitet som krävs för en sådan analys är oklart. Allmänt sett kan man dock med hjälp av sökandestatistiken komplettera den analys som ett renodlat experiment erbjuder.

Sammantaget kan man emellertid inte utesluta att indirekta effekter förekommit i denna studie varför effekternas storlek måste tolkas med försiktighet. Detta utesluter dock inte att effekternas riktning och karaktär (i synnerhet kombinationen av kvantitativa och kvalitativa effekter) är av stort värde att få kunskap om. Exempelvis kan man helt utesluta den typ av negativa stigmatiseringseffekter som uppnåddes vid experimentet med löne-vouchers i Dayton, Ohio, USA.

En annan fråga gäller vilken betydelse det hade att en del lönesubventioner ingått i det handlingsprogram som varit utgångspunkten för arbetet med experimentgruppen. Ifall dessa varit av större omfattning än i kontrollgruppen är det effekten av både förmedlingsverksamheten och lönesubventionerna som analysen omfattar. Detta påverkar inte på något sätt värdet av experimentet när det

**Tabell 3.4 Några centrala resultat från Eskilstuna-studien**

	Experiment- grupp	Kontroll- grupp
1. Andel med arbete på öppna marknaden vid försöksperiodens slut	0,48	0,34
2. Andel med arbete på öppna marknaden efter 9 månader	0,50	0,38
3. Arbetslöshetsveckor från försökets start till uppföljningen 9 månader efter dess slut	11 veckor	18 veckor
4. Genomsnittligt antal sysselsättningsmånader för dem som var sysselsatta vid uppföljningen	9,0 månader	7,6 månader
5. Genomsnittlig månadslön vid uppföljningstillfället för dem som var sysselsatta då	3 588 kr	3 386 kr
6. Andel med fast anställning vid uppföljningstillfället bland dem som var sysselsatta då	0,92	0,68
7. Andel med negativa omdömen om arbetets kvalitet	0,12	0,27

Källa: Delander (1978).

gäller att belysa hur arbetsmarknadspolitik fungerar, utan bara hur det skall tolkas. Delander rapporterar emellertid att båda grupperna utnyttjat lönesubventioner i ungefär samma utsträckning.

Ytterligare en fråga är om experimentgruppen varit representativ för den grupp som skulle ha fått utökad förmedlingsservice om en permanent förstärkning skulle erhållits på kontoret. I ett avseende kan detta ifrågasättas. Om en sådan förstärkning erhållits för att stödja de långtidsarbetslösa skulle resurserna för-

modligen ha fördelats bland de drygt 400 som ingick i hela populationen efter de sökandes behov av extra service. Experimentet innebär emellertid att resurserna måste avgränsas till 200 personer. Därför kan det tänkas att en del med stort behov av service blev utan, medan en del "låg-behov-personer" i experimentgruppen fick service i stället. Generellt sett borde detta leda till en underskattning av effekterna, eftersom experimentet innebär att personalen fråntas möjligheten att allokera sina resurser till dem som har de största behoven. Ju större spridning i behoven (eller effekterna av förmedlingsservicen för att vara mer korrekt) och ju bättre personalen kan genomskåda denna spridning, desto större underskattning blir det fråga om.

En annan fråga man kan ställa sig är om det kan ha förekommit Hawthorne-effekter, vilket i detta fall i första hand kan ha gällt personalen och distriktskontorets ledning. Om dessa var exceptionellt entusiastiska över experimentet, eller såg en chans att få extra resurser i händelse av gott utfall, så kan resultatet vara missvisande. Ytterligare en förutsättning för att resultaten ska vara helt tillförlitliga är att personalen hade samma kvalitet i de båda grupperna. I experimentet i Dayton, Ohio eliminerades detta problem genom att rotera personalen mellan olika grupper. Rapporten diskuterar inte dessa problem, varför man kan förmoda att de inte bedömts som viktiga i praktiken.

De etiska invändningarna mot ett experiment av detta slag är självfallet lätt vägande. Visserligen förvägras kontrollgruppen service som experimentgruppen erhåller, men för det första sker detta med lottens hjälp (vilket ofta uppfattas som rättvist), för det andra sker det bara under en begränsad period och för det tredje har den kunskap som experimentet ger ett stort etiskt värde.



#### 4 Utvärdering av utvärderingar med hjälp av experiment

En oväntad och oplanerad användning av National Supported Work Demonstration är att dess data gjort det möjligt att jämföra resultaten mellan experimentella och icke-experimentella studier. Därmed har det varit möjligt att genomföra en slags "utvärdering av utvärderingar" tack vare detta experiment.

Ansatsen har varit följande: I stället för den slumpmässigt utvalda kontrollgruppen från experimentet har en jämförelsegrupp av den typ som man normalt använder i en icke-experimentell studie valts. Denna grupp av individer kopplas ihop med experimentgruppen från experimentet till ett analysmaterial. De regressionsmetoder och matchningstekniker som normalt utnyttjas vid icke-experimentella studier appliceras sedan på detta analysmaterial. En metod som ger samma resultat som det renodlade experimentet betecknas som tillförlitlig medan naturligtvis motsatsen gäller för metoder som kraftigt avviker från experimentets resultat. Detta bygger på antagandet att ett experiment, som är upplagt som National Supported Work Demonstration, ger en helt korrekt skattning av programmets effekt. Som framgick av presentationen i avsnitt 3.2 finns all anledning att ha stor tilltro till detta experiment.

Denna ansats till utvärderingar av icke-experimentella utvärderingsmetoder har tillämpats av Fraker och Maynard (1987) och Lalonde (1984 och 1986). Fraker och Maynard koncentrerade sig på två av delgrupperna från projektet, nämligen kvinnor med AFDC och ungdomar. Jämförelsegrupperna valdes från Current Population Survey (CPS), den amerikanska arbetskraftsundersökningen. Den grupp som användes i första hand valdes utifrån de grundläggande kriterierna för tillträde till programmet; ungdomarnas ålder skulle vara 17 till 20 år och kvinnorna med AFDC skulle inte ha barn yngre än 6 år.

De började med att studera resultaten från en regressionsmodell, där utfallet i inkomster (1977, 1978 respektive 1979) förklarades, förutom med programdeltagande, med demografiska karakteristika, tidigare arbetserfarenheter, arbetslösheterfarenheter, tidigare inkomster och förändring i inkomster (före programmet). I tabell 4.1 presenteras de effekter som estimerades med denna modell. Det mest anmärkningsvärda är den stora diskrepansen mellan de effekter som erhöles från kontrollgruppen jämfört med jämförelsegruppen för ungdomarna.

Faktum är att resultaten avviker så kraftigt att den icke-experimentella ansatsen måste betecknas som helt otillförlitlig.

Däremot överensstämmer resultaten från kontrollgruppen och jämförelsegruppen relativt bra för kvinnor med AFDC. Diskrepansen är störst för det tredje året, då effekten som erhålls då jämförelsegruppen utnyttjas är cirka 25 procent lägre (i dollar räknat) än den effekt som experimentet ger. Även om punktskattningarna är relativt samstämmiga kan man också se från standardfelen att precisionen är relativt låg. Denna källa till osäkerhet i de estimerade effekterna diskuteras dock inte av författarna.

Fraker och Maynard studerar också några alternativa regressionsmodeller<sup>11</sup> och jämförelsegrupper. De fann att resultaten var känsliga för valet av jämförelsegrupp, särskilt för ungdomar. En mindre känslighet för valet av regressionsmodell kunde också noteras.

Den huvudsakliga slutsatsen från denna studie är således att regressionsmodeller som utnyttjar longitudinella data kan replikera resultaten från klassiska experiment relativt väl för vuxna kvinnor, men dåligt för ungdomar. Detta är på sätt och vis knappast förvånande, eftersom inkomsterna före programmet kan förväntas vara mindre värdefulla som kontrollvariabler för ungdomar än för vuxna kvinnor. Syftet med att utnyttja tidigare inkomster som kontrollvariabler är att dessa utgör ett mått på individernas ställning på arbetsmarknaden och eftersom ungdomarna knappast hunnit etablera sig på arbetsmarknaden är denna information av ringa värde som kontrollvariabler för denna grupp.

Den andra studien – presenterad i Lalonde (1984 och 1986) – baserades på dels gruppen kvinnor med AFDC, dels män som ingick i någon av de andra grupperna. Lalonde konstruerade jämförelsegrupper genom slumpmässiga urval av kvinnor med AFDC respektive män ur 1976 års arbetskraftsundersökning (Current Population Survey). Kriterierna för att inkluderas i jämförelsegrupperna var att ha tillhört arbetskraften i mars 1976 och att ha haft en inkomst understigande 20 000 dollar under 1975.

---

<sup>11</sup>Så kallade "first-difference"-modeller (med och utan förändringar i förklaringsvariabler).

**Tabell 4.1** Estimerade effekter (i dollar) av National Supported Work Demonstration med hjälp av slumpmässigt utvald kontrollgrupp respektive jämförelsegrupp från Current Population Survey

År	Ungdomar		Kvinnor med AFDC	
	Kontrollgrupp	Jämförelsegrupp	Kontrollgrupp	Jämförelsegrupp
1977	313 (134)	-668 (310)	1 423 (162)	1 560 (400)
1978	-28 (135)	-1 191 (373)	505 (137)	537 (335)
1979	-18 (166)	-1 179 (375)	351 (174)	257 (465)

Not: Standardfel inom parentes.

Källa: Fraker och Maynard (1987).

När Lalonde estimerar effekten av programdeltagande med hjälp av regressionsmodeller och den ovan nämnda jämförelsegruppen, finner han återigen att det är lättast att replikera resultaten från det renodlade experimentet för gruppen kvinnor med AFDC. Eftersom gruppen män har en ganska låg genomsnittsålder – 24 mot 34 för kvinnor med AFDC – är orsaken troligen densamma som ovan. Tidigare inkomster som utnyttjas som kontrollvariabler är troligen mer informativa för vuxna än för unga.

Lalonde jämför också hur väl olika regressionsmodeller förmår replikera experimentens resultat. Den första utnyttjade enbart tvärsnittsdata; inkomsterna efter programmet förklarades med programdeltagande samt ålder, ålder i kvadrat, ras och utbildning. Den andra utnyttjade longitudinella data genom att förklara förändringen i inkomster (från en tidpunkt före programmet till en tidpunkt efter detsamma) med ålder och programdeltagande.<sup>12</sup> Den tredje utnyttjade longitudinella data genom att förklara inkomsterna efter programmet med program-

<sup>12</sup>En så kallad "first difference" eller "fixed effect model". Se appendix.

deltagande, kontrollvariablerna ovan samt inkomsternas nivå före programmet.<sup>13</sup> Den fjärde utvidgade den tredje modellen genom en två-steps-metod för att beakta selektions-bias.<sup>14</sup> De viktigaste slutsatserna var: (i) longitudinella data ger mer tillförlitliga resultat än tvärsnittsdata, och (ii) två-steps-metoden för att beakta selektions-bias kan förbättra resultaten något.

Resultaten från dessa studier är delvis konstruktiva för fortsatt utvärdering med icke-experimentella metoder men delvis också nedslående. Det är konstruktivt att få veta att metoder som utnyttjar longitudinella data ger relativt tillförlitliga resultat för kvinnor. Problemet är emellertid att det är svårt att bedöma hur generellt detta resultat är, om det är konstant över tiden och om det gäller för alla kvinnor eller för alla vuxna. Därför kvarstår att det är ett starkt önskemål att kunna välja mellan de icke-experimentella metoder som används inom forskningen i dag, att finna teoretiska eller empiriska kriterier för val av metod. Utan kriterier av detta slag är det onekligen nedslående att olika metoder ger resultat som kraftigt avviker från varandra.

Nyligen har emellertid Heckman, Hotz och Dabos (1987) utvecklat ett testförfarande för att välja mellan alternativa metoder. De tillämpar detta förfarande på exakt samma data som Fraker och Maynard (1987). Den spännande frågan blir nu om den metod som enligt detta testförfarande är "bäst" ger samma resultat (dvs samma effekt av programmet) som experimentet. I detta fall utnyttjas således experimentet för att utvärdera det testförfarande som Heckman, Hotz och Dabos föreslår.

Tre metoder granskas i denna studie, nämligen en "linear control function model", en "fixed effect model" och en "random growth model" (se appendix c för en mer fullständig teknisk beskrivning). Två grupper från National Supported Work Demonstration studeras, ungdomar och kvinnor med AFDC. Huvuddragen i resultaten återfinns i tabell 4.2.

---

<sup>13</sup>En så kallad "state dependence model". Se appendix.

<sup>14</sup>En korrektionsterm ("Heckman-lambda") konstruerades med hjälp av probit-analys. Se Lalonde (1986) för närmare förklaring och referenser.

De tester som föreslås bygger på att varje modell har falsifierbara implikationer, dvs om modellen utgör en korrekt specifikation av inkomsternas dynamiska struktur och selektionsprocessen får vissa koefficienter icke anta vissa värden. Därför testas dessa koefficienter. För alla tre modellerna gäller att om de estimeras på data som helt och hållet ligger före eller efter programmet så skall koefficienten för programdeltagande vara noll eftersom det då inte föreligger någon effekt. Om koefficienten inte är noll är modellen felaktigt specificerad. För två av modellerna – "fixed effect model" och "random growth model" – kan ytterligare ett test tillämpas.<sup>15</sup>

Experimentet gav för ungdomar resultatet att programmets effekt på inkomsterna låg nära noll. De tre icke-experimentella metoderna gav dock negativa effekter, allra störst negativa effekter gav "fixed effect model", följd av "linear control function model" och "random growth model". Modelltesten gav till resultat att de två första modellerna förkastas, medan "random growth model" icke kan förkastas. Testförfarandet kan sägas ha varit framgångsrikt eftersom den metod vars resultat ligger närmast experimentets valts ut som den bästa. Med tanke på de stora standardfelen kan man säga att resultaten från experimentet och från "random growth model" är helt förenliga med varandra. Det bör dock betonas att standardfelen är mycket höga, i synnerhet om man betänker att de totala inkomsterna 1978 och 1979 endast uppgick till drygt \$2 000 för programdeltagarna. Precisionen i de skattningar som experimentet och den bästa metoden gav är därför mycket låg.

För gruppen kvinnor med AFDC är mönstret annorlunda. Experimentet ger signifikant positiva effekter på \$439,6 (1978) och \$267,2 (1979). De tre icke-experimentella metoderna ger klart större effekter, låt vara att standardfelen är så pass höga att experimentets resultat täcks av de konfidensintervall som följer av de icke-experimentella resultaten.

Om en forskare varken haft tillgång till experimentet eller till modelltesterna skulle resultaten förmodligen betecknas som "robusta" eftersom de tre icke-experimentella metoderna ger i stort sett samma resultat – positiva effekter mellan 600 och 900 dollar. Modelltesterna gav dock i utslag att alla modellerna

---

<sup>15</sup>Testet utnyttjar det faktum att en koefficient för inkomsterna från en annan period än den som ingår i modellen skall vara noll om modellen är korrekt.

**Tabell 4.2 Resultat från utvärdering av testmetodik i Heckman, Hotz och Dabos (1987). Estimerade effekter på inkomsterna 1978 resp 1979**

(Standardfel inom parentes.)

Metod	Ungdomar	Kvinnor med AFDC
<u>Experiment<sup>a</sup></u>		
Inkomster 1978	–47,6 (144,1)	439,6 (141,6)
Inkomster 1979	9,5 (172,9)	267,2 (161,9)
<u>Linear control function model<sup>b</sup></u>	Denna modell förkastas av testet	Denna modell förkastas med tvekan av testet
Inkomster 1978	–889,1 (380,0)	906,6 (335,3)
Inkomster 1979	–1 330,9 (388,5)	586,0 (385,6)
<u>Fixed effect model<sup>b</sup></u>	Denna modell förkastas av båda testerna	Denna modell förkastas av båda testerna
Inkomster 1978	–963,7 (276,21)	800,5 (321,5)
Inkomster 1979	–1 383,0 (312,0)	708,5 (328,1)
<u>Random growth model<sup>b</sup></u>	Denna modell kan inte förkastas av någon av de två testerna	Denna modell förkastas av båda testerna
Inkomster 1978	–776,9 (536,8)	880,5 (869,5)
Inkomster 1979	–849,6 (629,8)	808,5 (918,3)

a) Dessa effekter baseras på en enkel skattning av skillnaden mellan experiment- och kontrollgrupperna.

b) Här presenteras de skattningar som bygger på maximalt antal förklaringsvariabler samt interaktionstermer.

Källa: Heckman, Hotz och Dabos (1987).

måste förkastas – dock med viss tvekan för "linear control function model". Eftersom punktskattningarna från de icke-experimentella metoderna ligger för högt kan man i och för sig säga att testerna varit framgångsrika eftersom de leder till att man tvingas ta avstånd från skattningar som annars gjort ett mycket robust intryck. Däremot utgör testerna inte något belägg för att den icke-experimentella forskningsansatsen utgör en framkomlig väg. För detta krävs att man kan finna någon annan modell som icke låter sig förkastas av tester av detta slag.

Sammanfattningsvis är det för tidigt att uttala sig om hur framgångsrikt tester av ovanstående slag kan komma att välja ut den bästa icke-experimentella metoden. Ett problem med den "utvärdering av modelltester" som Heckman, Hotz och Dabos presenterat är att experimentet i sig ger skattningar av låg precision. Ytterligare försöksverksamhet med de tester som dessa forskare föreslagit framstår dock som mycket angelägna.

## 5 Makroexperiment

Med ett undantag – Employment Opportunity Pilot Project – har experimenten som presenterades i avsnitt 3 varit så kallade mikroexperiment. Detta betyder att individer slumpmässigt har fördelats på experiment- och kontrollgrupper. Ett alternativ är makroexperiment, där olika grupper, orter eller delmarknader utgör de enheter vilka slumpmässigt utsätts för olika former av politik. Jag skall här belysa för- och nackdelar med dessa båda typer av experiment samt ta ställning till i vilka sammanhang den ena eller andra typen är att föredra.<sup>16</sup>

En potentiell fördel med makroexperiment är att problem med indirekta effekter kan undvikas. Ta Eskilstuna-studien av utökad förmedlingsservice (avsnitt 3.8) och studien av extra bonus till arbetslösa i Illinois (avsnitt 3.4) som exempel. I det första exemplet kan kö-bias misstänkas på grund av att de som erhåller utökad service kommer först i kön till de bästa jobben och därmed försvårar för andra. I det andra exemplet kan man misstänka att de som erhåller bonus tack vare ökad sökintensitet kommer först till jobben och därför försvårar för andra.

Ett alternativ i båda fallen hade varit att testa de nya programmen – utökad förmedlingsservice respektive bonus – på ett slumpmässigt urval av delmarknader, exempelvis orter. Dessa orter hade då fått utgöra experimentorter och för att kunna göra en jämförelse hade ett urval av kontrollorter, som inte fick utnyttja dessa möjligheter, fått väljas. Eftersom de indirekta effekter, som kan snedvrida resultaten från mikroexperimenten i första hand kan förväntas verka inom en ort kan detta problem elimineras. I värsta fall kan emellertid andra indirekta effekter uppstå vid makroexperiment. Om exempelvis förmedlingsservice och bonusmöjligheter på en ort blir så attraktiva att de drar till sig sökande från andra orter uppstår genast problem. En förutsättning för ett meningsfullt makroexperiment är med andra ord att orterna är avgränsade från varandra.

En annan fördel med makroexperiment är om man – som i Employment Opportunity Pilot Project – vill undersöka vilket deltagande som ett nytt program ger

---

<sup>16</sup>Problemen diskuteras ingående av Harris (1985) och Orr (1985). Harris argumenterar starkt för makroexperiment. Hans exempel är främst hämtade från sjukvårdsekonomi. Orr är mer skeptisk i sin kommentar till Harris och han bygger delvis på erfarenheter från arbetsmarknadsekonomiska tillämpningar.



upphov till. Detta program innehöll ju en jobb-garanti och för att undersöka hur många som kunde förväntas utnyttja denna möjlighet måste programmet få verka i sin helhet under en rimlig tid. Därför är det naturligt att välja ut ett antal orter där programmet testas i sin helhet.

En tredje fördel med makroexperiment uppstår ifall den åtgärd, som är föremål för studie, rent administrativt kan hanteras på ett mer naturligt sätt om det får tillämpas i sin helhet på en ort eller i ett distrikt inom den arbetsmarknads-politiska organisationen. Det kan tänkas att uppdelningen på en experiment- och en kontrollgrupp inom ett distrikt, som ju mikroexperimenten kräver, är administrativt besvärlig. Det är svårt att säga något generellt om betydelsen av detta problem utan det måste bedömas från fall till fall.

Det centrala problemet med makroexperiment är att det i praktiken är svårt att erhålla rimliga stickprovsstorlekar för experiment- och kontrollgrupperna. Skälet till detta är att en ort rent statistiskt måste betecknas som en observation. Även om en ort, på säg 25 000 inneväånare, utgör experimentort för en viss åtgärd och om denna ort är slumpmässigt utvald kan den bara betraktas som en oberoende observation. Om däremot 25 000 individer slumpmässigt väljs ut för en viss åtgärd, utgör dessa rent statistiskt 25 000 oberoende observationer. När man skattar skillnaden i utfallet mellan experiment- och kontrollgrupperna för en viss variabel – säg sysselsättningsgraden – måste skattningen i praktiken baseras på ett fåtal observationer. Då blir precisionen låg och det kan bli svårt att göra meningsfulla uttalanden om effekter. I praktiken uppmäts självfallet alltid någon skillnad mellan grupperna. Om man emellertid vill bedöma sannolikhet av att den uppmätta skillnaden är ett utfall av slumpen eller en renodlad effekt kommer små stickprov att skapa problem. Problemet förvärras ifall den "naturliga" spridningen i utfallsvariabeln är stor, vilket ofta är fallet när det gäller orter eller regioner.

Ett annat problem med makroexperiment är att det kan vara svårt att verkligen genomföra ett slumpmässigt urval av orter. En förutsättning är att en population av orter kan definieras på ett meningsfullt sätt. Om detta är möjligt blir resultatet sannolikt ett antal orter med mycket varierande egenskaper och av olika storlek. Två slumpmässigt valda grupper från en sådan population kommer sannolikt att uppvisa en betydande "naturlig" variation i utfallsvariabler som

sysstämning, arbetslöshet och arbetslöshetstider. En frestande lösning kan då vara att välja orter med likartade egenskaper till de båda grupperna, ett slags matchningsförfarande.<sup>17</sup> Problemet med detta förfarande är att man måste frågå principen om obundet slumpmässigt urval, vilket försvårar, och i värsta fall omöjliggör, en statistisk analys. Det bör noteras att de orter som i EOPP-projektet valts att ingå i experiment- och kontrollgrupperna, inte valts slumpmässigt.

Sammantaget kan makroexperiment framstå som ett elegant sätt att undgå problemen med indirekta effekter, men i praktiken kan det bli svårt att statistiskt testa skillnaderna i utfall mellan experiment- och kontrollgrupperna med god precision. Det betyder med statistisk terminologi att det blir svårt att "förkasta noll-hypotesen att inga skillnader (effekter) föreligger". Bara om de verkliga effekterna är mycket stora kan det bli möjligt att avslöja dessa med hjälp av ett makroexperiment.

För vissa typer av arbetsmarknadspolitiska åtgärder kan dock makroexperiment rent logiskt framstå som den enda lösningen. Det gäller exempelvis om man ska studera effekterna av nya arbetsformer på förmedlingarna. Då är just orter den naturliga analysenheten och makroexperiment den naturliga analysformen.

---

<sup>17</sup>Ett svenskt exempel på detta är det så kallade LYZS-projektet som utfördes inom AMS. Här valdes – sannolikt icke slumpmässigt – fyra län där nya regler för arbetslösas återbesök infördes samt fyra "likartade" län som arbetade enligt tidigare regler. Att tillämpa statistisk metod för att studera skillnaderna i utfallsvariablerna mellan grupper utvalda på detta sätt är inte lätt.

## 6 Förslag för den framtida utvärderingsforskningen

Den utvärderingsforskning kring arbetsmarknadsutbildning och andra individinriktade åtgärder som dominerat inom den nationalekonomiska tidskriftslitteraturen har varit relativt ambitiös. Den statistiska metodiken har utvecklats ordentligt både när det gäller estimationstekniken och när det gäller koppling till ekonomisk teori. Dessutom har – åtminstone för det amerikanska CETA-programmet – initiativ tagits till omfattande datainsamlingar vilka givit forskarna tillgång till longitudinella databaser med uppgifter om sysselsättning och inkomster både före och efter programmet för deltagare och icke-deltagare.

Trots detta har det, som framgått ovan, varit svårt att få fram riktigt tillförlitliga skattningar av programmens effekter. De amerikanska studierna av CETA-programmet visade ju någorlunda entydiga resultat enbart för kvinnor.

Frågan är därför om det är möjligt att gå vidare och i så fall hur. Inom forskningen kan man här urskilja två läger; ett med Orley Ashenfelter i spetsen, som förespråkar klassiska experiment; och ett med Jim Heckman i spetsen, som förespråkar fortsatt arbete med icke-experimentella metoder. Detta kan förefalla välbekant från 60-talet och tidigt 70-tal då motsvarande läger kunde urskiljas. Har då inga framsteg ägt rum? Utan tvekan har betydande framsteg gjorts. Insikterna om metodproblemens karaktär har ökat betydligt. Vad gäller experiment finns exempelvis, som framgick ovan, ett drygt halvdussin genomförda projekt. Dessa är väl dokumenterade och har givit värdefull praktisk kunskap om hur experiment bör genomföras. Dessutom finns en mer teoretiskt inriktad litteratur kring experiment som givit värdefulla insikter (se bland annat volymen som redigerats av Hausman och Wise). Därför är förutsättningarna för att genomföra experiment, rent metodologiskt, bättre i dag än för 15–20 år sedan.

Vad gäller icke-experimentella metoder har insikterna också ökat. Man kan mer precist säga vilken kunskap om inkomsternas dynamiska struktur och om selektionsprocessen som krävs för att välja mellan de metoder som står till buds.

Dessa ansatser är självfallet inte varandra uteslutande utan kompletterar varandra, vilket framgick av avsnitt 4. De gånger det är praktiskt och politiskt möjligt att genomföra experiment är det därför en god idé att samla in data som

också gör det möjligt att utnyttja icke-experimentella metoder. Experimentet ger då en möjlighet att utvärdera de icke-experimentella metoderna på det sätt som presenterades i avsnitt 4. Detta ingår i planerna för JTPA-experimentet i USA.

Vi avslutar nu med att mer i detalj precisera hur experiment kan utformas och hur icke-experimentella metoder kan utvecklas för att ge bättre kunskap om program som den svenska arbetsmarknadsutbildningen.

### **6.1 Experiment**

Hur kan man lägga upp experiment med vars hjälp man kan kvantifiera effekterna av individinriktade åtgärder som arbetsmarknadsutbildning? Problemen är av både etisk, organisatorisk och metodologisk natur. De etiska problemen gäller om experimentet berövar vissa medborgare förmåner av olika slag. Organisatoriska problem kan uppstå inom den arbetsmarknadsmyndighet som administrerar programmet och som normalt måste ändra sitt arbetssätt. De metodologiska problemen slutligen, är de problem som forskarna konfronteras med.

Diskussionen underlättas av några grundläggande begrepp och distinktioner. Först kan man skilja mellan pågående (och etablerade) program å den ena sidan, och nya program å den andra. Till nya program kan också hänföras nya inslag i etablerade program. Beroende på hur selektionen till programmet går till kan man dessutom skilja mellan efterfrågestyrda och utbudsstyrda program. I efterfrågestyrda program är det de potentiella deltagarnas intresse för programmet som avgör deltagande. Detta är fallet om exempelvis tillgången på kurser är mycket god i förhållande till efterfrågan så att alla som vill får delta. Den svenska arbetsmarknadsutbildningen har under det senaste decenniet – åtminstone tidvis – i stort sett kännetecknats av en sådan efterfrågestyrning. I utbudsstyrda program finns däremot inte plats för alla som har anmält sig och vill delta varför myndigheten som administrerar programmet gör ett urval bland dem som vill delta. Detta kan gälla om kurskapaciteten är otillräcklig inom ett utbildningssystem eller om tillgången på arbetsförmedlare är knapp inom ett förmedlingssystem. I ett utbudsstyrt program sker selektionen således genom beslut av både

de prospektiva deltagare som anmält sitt intresse och av myndigheterna som gör ett val bland dem som anmält sig.<sup>18</sup> I ett efterfrågestyrt program är det däremot enbart de prospektiva deltagarnas intresse som styr selektionen.

#### Efterfrågestyrda program

Betrakta först ett experiment som skall belysa effekten av ett efterfrågestyrt program. Den naturliga population och urvalsram från vilken experiment- och kontrollgrupperna slumpmässigt bör väljas är en lista över dem som anmält sig till ett program.<sup>19</sup> Det är då viktigt att själva anmälningsförfarandet inte påverkas av att ett experiment skall genomföras utan att såväl information om programmet som anmälningar till detsamma sker på ett helt realistiskt sätt. Detta bör leda till att slumpmässigt utvalda experiment- och kontrollgrupper i genomsnitt har samma egenskaper; de har ju genomgått samma selektionsprocess, dvs fattat beslutet att anmäla sig till programmet. Experimentgruppen har dessutom samma egenskaper som deltagarna skulle ha haft om programmet är i normal drift; den är ju slumpmässigt vald ur den grupp som anmält sig och skulle normalt ha deltagit i programmet. Skillnaden i utfall – sysselsättning och inkomster m m – mellan grupperna utgör därför en tillförlitlig skattning av programmets effekt. Mera precist belyser denna skillnad hur programmet i genomsnitt påverkar utfallet för programdeltagarna. (Eftersom det är ett mikroexperiment måste dock en reservation göras för att indirekta effekter kan snedvrida resultaten.)

Ett experiment av detta slag stöter dock på både etiska och organisatoriska problem. Det etiska problemet är att de som anmält sig till programmet men hamnat i kontrollgruppen måste förvägras att delta. Om experiment- och kontrollgrupperna skall vara lika stora måste således hälften förvägras delta, men ju

---

<sup>18</sup>Detta val från myndigheterna kan antingen avse det dikotoma valet delta/icke delta (gäller arbetsmarknadsutbildning) eller valet hur mycket service en viss kund skall ha (gäller förmedlingsservice).

<sup>19</sup>I praktiken måste naturligtvis en granskning göras av om dessa uppfyller grundläggande kriterier för att få delta. Arbetslöshet är ett vanligt sådant kriterium. Listan måste därför rensas från icke-berättigade personer som anmält sig.

mindre kontrollgruppen kan göras desto färre behöver avvisas.<sup>20</sup> Hur oetiskt det är att förvägra somliga att delta är självfallet svårt att uttala sig om generellt. Sannolikt är det mer allvarligt i fallet med ett pågående och etablerat program, som kan uppfattas som en väl etablerad rättighet. Detta gäller dock inte för nya program, som därför kan vara mer lämpliga att utvärdera med hjälp av experiment.

Det organisatoriska problemet är att programmet under experimentets gång måste bedrivas i reducerad omfattning, vilket kan leda till att kurslokaler inte utnyttjas ordentligt och i värsta fall till att vissa lärare och annan personal blir sysslösa. Man kan då invända att man kan göra en större grupp berättigade till programmet och göra ett urval ur denna större grupp vilket är tillräckligt stort för att utnyttja kurskapaciteten. Problemet är dock att detta kan förändra hela programmets karaktär. Sammantaget kan de etiska och organisatoriska problemen att genomföra ett experiment under de förhållanden som kännetecknar ett efterfrågestyrt program vara alltför stora.

De politiska frågorna kring ett efterfrågestyrt program är inte alltid om programmet skall behållas eller läggas ned. I stället är frågan ofta om det bör göras mer eller mindre attraktivt för de prospektiva deltagarna, dvs om det på marginalen bör ökas eller minskas. Ett typexempel är just den svenska arbetsmarknadsutbildningen. AMS har länge velat höja utbildningsbidragen för att locka fler att delta. Effekten på marginalen för nya deltagare behöver inte alls vara densamma som den genomsnittliga effekt som experimentet ovan kan hjälpa till att skatta.

Kan man då lägga upp experiment som belyser dessa marginella effekter av ett efterfrågestyrt program? I princip är detta möjligt, men problemen är stora både etiskt och organisatoriskt. Vad som krävs är att olika experimentgrupper, som är slumpmässigt utvalda, erbjuds att delta på olika villkor. Åtminstone krävs tre slumpmässigt utvalda grupper. En som erbjuds delta på de villkor som gäller

---

<sup>20</sup>Detta reser den statistiska frågan vad som är den optimala fördelningen mellan experiment- och kontrollgrupp och vilka förluster i termer av statistisk precision man drabbas av om man avviker från den optimala fördelningen och gör kontrollgruppen mindre.

när programmet är i normal drift. En annan grupp som erbjuds delta på de alternativa villkor (exempelvis högre utbildningsbidrag), som är föremål för politisk diskussion. Och slutligen en kontrollgrupp som icke tillåts delta. Genom att jämföra utfallet mellan kontrollgruppen och en av experimentgrupperna erhålls en skattning av den genomsnittliga effekten av att erbjuda en grupp ett program. En jämförelse av utfallet mellan två experimentgrupper som erbjudits olika utbildningsbidrag ger en skattning av den marginella effekten av att höja (eller sänka) bidragen. (Se Björklund (1988) för en mer formell analys.)

Ett sådant experiment förutsätter således särbehandling av olika grupper, låt vara att det är slumpen som avgör vilka som får extra förmåner och att slumpen ofta uppfattas som en rättvis fördelningsmekanism. Även organisatoriskt ställer det till vissa problem att erbjuda och informera olika grupper om olika förmåner.

#### Utbudsstyrda program

Betrakta nu ett experiment som skall belysa effekten av ett utbudsstyrt program. Såväl experiment- som kontrollgrupp skall i genomsnitt ha samma egenskaper som deltagarna har när programmet är i normal drift, varför urvalet måste ske från en population som genomgått exakt de selektionsprocesser som deltagarna normalt gör. Detta betyder att de prospektiva deltagarna först måste anmäla sig till programmet på vanligt sätt. Därefter skall den programansvariga myndigheten välja ut dem som skulle ha fått delta om allt varit normalt. En lista<sup>21</sup> över dem som skulle ha fått delta måste överlämnas till forskarna som från denna lista gör ett slumpmässigt urval till experiment- och kontrollgrupperna. Det är viktigt att myndighetens urval går till på samma sätt som när programmet är i normal drift.<sup>22</sup>

---

<sup>21</sup>Rent fysiskt behöver det inte vara en lista, det viktiga är att ett slumpmässigt urval görs från dem som programmyndigheten väljer ut.

<sup>22</sup>Det kan vid utbudsstyrda program vara lockande att göra urvalet från dem som anmält sig till programmet och hoppa över den selektion myndigheterna står för. Detta skapar dock en källa till fel (se Björklund(1988) som behandlar detta fall).

Vid ett experiment som läggs upp på detta sätt uppstår samma problem som i det efterfrågestyrda fallet, nämligen att kontrollgruppen förvägras delta (vilket de skulle ha fått göra om experimentet inte genomförts) och programmet måste bedrivas i en lägre omfattning än normalt med åtföljande organisatoriska problem. Men till skillnad från det efterfrågestyrda fallet finns en möjlighet att undvika dessa problem. Man kan nämligen uppdra åt programmyndigheten att välja ut fler deltagare bland dem som anmält sig än vad som normalt görs. Ur denna större grupp kan sedan ett urval till experiment- och kontrollgrupperna göras. (Notera att detta förfarande ingick i planerna för JTPA-experimentet, se avsnitt 3.6.) Genom en lämplig avvägning kan experimentgruppen bli lika stor som den programkapacitet som normalt utnyttjas. Det organisatoriska problemet kan därmed undvikas. Även det etiska problemet blir mindre; visserligen förvägras kontrollgruppen att delta, men denna grupp blir precis lika stor som den grupp som skulle ha förvägrats delta när programmet är i normal drift.<sup>23</sup>

Detta sätt att undvika det organisatoriska problem som experimentet ger upphov till skapar både problem och möjligheter. När en större grupp än vanligt väljs ut av programmyndigheten erhålls en annorlunda grupp än den vanliga. Om det finns en spridning i de effekter som programmet ger olika individer och om myndigheterna har en förmåga att välja deltagare efter effekternas storleksordning, kommer de ytterligare deltagare som väljs till följd av experimentet att ha ett lägre utbyte av programmet än de som normalt skulle fått delta. Man kan tala om en avtagande avkastning av programmet och ett större program ger därför lägre genomsnittliga effekter. Ett experiment som bygger på ett större urval mäter därför programmets effekter givet att programmet har en större omfattning än normalt.

Under vissa omständigheter kan emellertid detta vara en fördel. Om det finns ett politiskt intresse av att utöka kapaciteten för programmet kan ett experiment som belyser dess effekter vid en högre volym vara av stort intresse. Med tanke på att det i dessa fall, när programmet är i normal drift, finns köer av

---

<sup>23</sup>En viss skillnad mellan de grupper som förvägras delta kan ändå uppstå. När programmet är i normal drift blir det de som enligt myndigheterna är i minst "behov" av programmet som förvägras delta. Under experimentet blir det en slumpmässigt utvald grupp som inte tillåts delta.



personer som vill, men inte får, delta är ett sådant politiskt intresse ganska naturligt.

Ytterligare en värdefull analytisk möjlighet uppstår om programmyndigheten kan gradera de sökande efter hur angeläget det är att de deltar. Säg att 100 personer normalt får delta, men att ett experiment ska genomföras så att 200 personer ska väljas ut varav 100 ska fördelas till vardera experiment- och kontrollgrupperna. Antag vidare att programmyndigheten kan gradera de 200 utvalda efter angelägenhetsgrad; en gradering i två grupper är fullt tillräckligt, nämligen de 100 personer som skulle få delta under normala förhållanden och de ytterligare 100 personer som skulle få delta om programmet fördubblas i omfattning. Forskarna kan då välja experiment- och kontrollgrupper ur båda grupperna. Detta ger möjlighet att skatta effekten av programmet både för den grupp som deltar när programmet drivs i normal omfattning och för den större grupp som skulle delta om det skulle drivas i fördubblad omfattning. Därmed kan experimentet ge information både om programmets genomsnittliga effekt när det drivs i normal omfattning och utbytet av att utöka programmets omfattning.

Om en sådan gradering inte är möjlig är det mer komplicerat att mäta de marginella effekterna vid ett utbudsstyrt program. Problemet är att mäta effekterna för dem som skulle få delta om kapaciteten ökar. Den enda möjligheten synes vara med hjälp av ett makroexperiment. Genom att slumpmässigt fördela olika programkapacitet till olika orter kan effekterna vid olika nivåer beräknas, och därmed kan man också beräkna utbytet av att förändra kapaciteten från en nivå till en annan. Att genomföra ett sådant experiment kan dock både etiskt och organisatoriskt vara mycket problematiskt. Det skulle dessutom vara förknippat med de speciella problem som makroexperiment lider av – se avsnitt 5.

Distinktionen mellan efterfråge- och utbudsstyrda program behöver inte alltid vara helt klar i praktiken. Inom ramen för ett arbetsmarknadsutbildningsprogram kan det finnas vissa kurser där alla sökande får plats och andra kurser där bara en del av de sökande kan beredas plats. Detta förändrar inte de slutsatser som drogs ovan. Ett experiment som skall belysa programmets effekter under normal drift bör bygga på urval från de populationer som genomgått de selektionsprocesser som normalt förekommer. Detta betyder att urvalet till experiment- och kontrollgrupper för de kurser som har fritt tillträde bör baseras på

den population som anmält sig till kurserna. Urvalet till kurserna med ransonering bör baseras på den population som både anmält sitt intresse och valts ut att få delta av den administrerande myndigheten.

### Sammanfattning

Diskussionen ovan har illustrerat att det är långt ifrån problemfritt att konstruera experiment i praktiken. De största problemen sammanhänger med att en grupp – kontrollgruppen – inte kan tillåtas delta i programmet. Detta kan uppfattas som oetiskt och det kan skapa organisatoriska problem om programverksamheten måste reduceras kraftigt under den tid experimentet pågår. Men problemen behöver inte alltid vara lika stora och kan under vissa omständigheter vara negligerbara.

För det första är problemen mindre när det gäller nya program, eller nya inslag i gamla program. Detta illustreras också av att alla experiment – utom JTPA-experimentet som presenterades i avsnitt 3 – kan sägas utgöra studier av nya program.<sup>24</sup> Inga etiska eller organisatoriska invändningar kunde således förhindra dessa projekt. Tvärtom kanske man kan hävda att det hade varit oetiskt att avstå från den värdefulla information och kunskap som experimenten givit.

För det andra är problemen mindre för utbudsstyrda än för efterfrågestyrda program. Detta beror på att det är möjligt att erhålla en experimentgrupp som är tillräckligt stor för att utnyttja den befintliga kapaciteten inom programmet, exempelvis de lärare och de lokaler som används inom arbetsmarknadsutbildningen. En så stor experimentgrupp erhålls om programmyndigheten från populationen av sökande till programmet väljer ut fler än vad som normalt deltar. Ur denna grupp kan forskarna sedan slumpmässigt fördela individerna på experiment- och kontrollgrupper. Detta betyder i och för sig att experimentet mäter effekten av ett större program än det som är i normal drift, men sådana resultat kan ha ett stort värde för beslutsfattare som överväger att utöka programmet.

---

<sup>24</sup>Eskilstuna-studien var visserligen en studie av effekterna av arbetsförmedling, men experimentet knöts till en utvidgning av sådan service.

Sammantaget har vi därför kunnat identifiera två situationer då de etiska och organisatoriska problemen att genomföra klassiska experiment är små eller obefintliga, nämligen (i) när nya program övervägs och (ii) när ett befintligt program kännetecknas av betydligt fler sökande än vad som kan tillåtas delta och det finns ett politiskt intresse av att utöka programmet. Även i dessa gynnsamma situationer är det emellertid i första hand programmets genomsnittliga effekter som kan beräknas med hjälp av experimenten. Frågor av typen "Är det lönsamt att höja (eller sänka) utbildningsbidragen?" och "Är det lönsamt att öka (eller minska) kurskapaciteten?" är betydligt svårare att besvara. De experimentella uppläggningar som krävs för att besvara sådana frågeställningar är betydligt mer komplicerade.

Ytterligare en fråga som gäller de praktiska och organisatoriska möjligheterna att genomföra experiment är i fall ett (mikro)experiment behöver genomföras över hela landet eller om det räcker med en eller ett fåtal orter (eller distrikt). Med tanke på det extra arbete som ett experiment skapar för den programadministrerande myndigheten och med tanke på det arbete som forskargruppen får för att genomföra programmet underlättar det onekligen om experimentet kan genomföras bara på ett mindre antal platser.

Rent formellt kan resultaten från ett experiment inom ett distrikt inte generaliseras till att gälla hela landet. I praktiken kan det dock vara en rimlig strategi att börja med att genomföra experimenten på ett mindre antal orter. Om resultaten skulle avvika betydligt finns det anledning att fundera över vilka ortsspecifika faktorer som kan tänkas påverka utfallet. I annat fall kan det vara rimligt att generalisera resultaten.

Slutligen finns det skäl att betona att ett väl genomfört experiment av ett program som arbetsmarknadsutbildning skulle belysa flera viktiga frågor. Först och främst skulle effekterna på aggregerad nivå kunna belysas: "Utbildningen påverkar sysselsättningsgraden med X procent efter ett år, med Y procent efter två år osv." Sådan information kan utnyttjas för en allmän dimensionering av åtgärderna. Lika viktigt kan det emellertid vara att jämföra olika kurser med varandra, eftersom ett program som arbetsmarknadsutbildning innehåller en mängd olika kurser. Om experimentet är någorlunda stort så att stickprovsstorlekarna (för experiment- och kontrollgrupperna) för varje kurs inom det totala

programmet också blir rimligt stora kan sådana resultat erhållas. För dem som ansvarar för att välja och utforma kurserna är sådan information självfallet av mycket stort värde.

## **6.2 Databaser för icke-experimentell utvärdering**

Slutligen, hur kan den icke-experimentella utvärderingsmetodiken utvecklas? Denna fråga är viktig inte minst mot bakgrund av att det kan vara praktiskt svårt att genomföra experiment för att utvärdera pågående etablerade program.

Metodologiskt sett rymmer denna metodik mycket svårare tekniska problem. Varje "modell" bör vara rimligt formulerad ur ekonomisk synvinkel, dvs modellen måste specificera dels utfallsvariabelns (vanligen inkomstens) utveckling över tiden, dels selektionsprocessen att delta i programmet. Dessutom måste varje modell kunna estimeras på ett konsistent sätt, något som vanligen rymmer de allra svåraste tekniska problemen. Experimentens metodologiska problem ligger, som framgick ovan, snarare på de etiska och organisatoriska planen.

Vi nöjer oss här med att begränsa diskussionen till frågan om databaser för icke-experimentella studier. Med andra ord: givet den kunskap som finns om ekonomiska modeller och estimationsmetoder, vilka data bör insamlas för att metoderna ska kunna användas?

För det första är det önskvärt med longitudinella data, dvs data behövs för utfallsvariabeln (inkomster, sysselsättning, löner m m) inte bara efter programmet utan också före. De mest generella modeller som utnyttjats i tidigare studier<sup>25</sup> kräver data för åtminstone 2 perioder före programmet. Ytterligare data ökar precisionen i skattningarna av dessa modeller. Till detta kommer att det är önskvärt att kunna genomföra de modell-test som utvecklats under senare år. Detta ställer krav på data från fler perioder än som krävs för att estimeras modellerna som sådana.

---

<sup>25</sup>Så kallade "random growth model" och "fixed effect model" med seriekorrelerade slumpfel.

För det andra är det viktigt med en noggrann kartläggning även av andra arbetsmarknadserfarenheter före programdeltagandet. Ofta har de som deltar i arbetsmarknadspolitiska program drabbats av någon form av chock på arbetsmarknaden, exempelvis blivit uppsagda i samband med en företagsnedläggning. Konsekvenserna av en uppsägning är bland annat beroende av hur länge man arbetat på företaget. Lång anställningstid i samma företag och bransch kan betyda att tidigare yrkeskunskaper blivit helt obsoleta. Ett annat exempel kan vara kvinnor som varit borta från arbetsmarknaden en längre tid. Förvärvs-avbrott leder normalt till en inledande period av låga inkomster. För att kunna kontrollera för sådana faktorer vid regressionsanalys eller liknande teknik krävs variabler som beskriver sådana arbetsmarknadserfarenheter för både experiment- och kontrollgrupperna.

För det tredje kan ifrågasättas om inte den periodlängd – ett år – som utnyttjats i framför allt de amerikanska studierna är alltför lång. Skälet till att årsinkomster utnyttjats så flitigt som utfallsvariabel är självfallet att registerdata finns tillgängliga på årsbasis. Som en följd av att data har denna egenskap betraktas normalt året före programdeltagandet som selektionsåret. Eftersom många program är korta – den typiska längden för en kurs inom svensk arbetsmarknadsutbildning är fem månader – blir analysen alltför grov. De chocker som givit upphov till programdeltagande under slutet av ett kalenderår kan ha inträffat under samma år, medan de som startar programmet i början av året drabbades av en chock under det föregående året. En övergång till att utnyttja kvartals- eller månadsdata skulle kunna förbättra kvaliteten i analysen påtagligt.

Slutligen är frågan om kvaliteten i data utomordentligt viktig. Som framgått ovan är precisionen i de skattningar av programmens effekter som erhållits från såväl experimenten som från de icke-experimentella studierna minst sagt besvärande låg. Sannolikt beror detta på mätfel i de variabler – både beroende och oberoende variabler – som använts i analysen. Vid datainsamlingar måste därför de rena kvalitetsaspekterna ägnas betydligt större uppmärksamhet än hittills.

## Appendix a

Metoder använda av Bassi (1984) och Dickinson, Johnson och West (1986) i de studier som presenterades i tabell 2.1.

### Bassi (1984)

#### *Fixed effect model*

$$(Y_{it}-Y_{is}) = (e_t-e_s) + (X_{it}-X_{is}) \gamma + P_i\beta_t + (\epsilon_{it}-\epsilon_{is})$$

där

$Y_{it}$  = inkomster för individ i under år t (efter programmet).

$Y_{is}$  = inkomster för individ i under år s (före programmet). (Bassi använde inkomsterna 2 år före programmet.)

$X_{it}(X_{is})$  = observerbara karakteristika som påverkar inkomsterna för individ i under år t (och s).

$P_i$  = dummyvariabel för programdeltagande.

$e_t(e_s)$  = en felterm som är specifik för period t(s) och lika för alla individer.

$\epsilon_{it}(\epsilon_{is})$  = en felterm som är specifik för individ i under period t(s).

Enbart ålder i kvadrat ingick bland X-variablerna.

#### *Fixed effect model med seriekorrelerad slumpterm*

Samma modell som ovan men slumptermen  $\epsilon_{it}(\epsilon_{is})$  tillåts ha följande dynamiska struktur:

$$\begin{aligned}\epsilon_{it} &= \zeta\epsilon_{it-1} + V_{it} \\ \epsilon_{is} &= \zeta\epsilon_{is-1} + V_{is}\end{aligned}$$

### *Creaming model*

Denna modell är ytterligare en generalisering som tillåter att urvalet av deltagare baseras på  $V_{is}$ , dvs de transitoriska slumptermerna perioden före programmet.

$E(V_{is}|\text{deltagande})$  estimeras från data före programmet och utnyttjas sedan i ett andra estimationssteg för att estimeras programmets effekt givet denna typ av selektion.

### Dickinson, Johnson och West (1986)

#### *State dependence model*

$$Y_{i78} = \sum_{t=70}^{75} Y_{it}\alpha_t + X_i\gamma + P_i\beta + \epsilon_i$$

där

$Y_{i78}(Y_{it}) =$  inkomster för individ  $i$  under år 1978 ( $t$ ).

$X_i =$  ett mycket stort antal bakgrundsvariabler för demografiska karakteristika, arbetsmarknadserfarenheter m m.

$P_i =$  dummyvariabel för programdeltagande.

$\epsilon_i =$  en slumpterm för individ  $i$ .

## Appendix b

Metoder använda av Björklund (1989) i den studie som presenterades i tabell 2.2.

### Panel A

#### *State dependence model*

Effekt på andelen sysselsatta:

$$Y_{i81} = Y_{i74}\alpha + X_i\gamma + P_i\beta + \epsilon_i$$

där

$Y_{i81}(Y_{i74}) =$  dummyvariabel för sysselsättning 1981 respektive 1974.

$X_i =$  ett antal bakgrundskaraktistika; ålder, kön, utbildningsår 1974, civilstånd 1981, antal barn 1981, invandrarstatus, arbetslöshetsunderstöd 1974–80.

$P_i =$  deltagande i arbetsmarknadsutbildning.

$\epsilon_i =$  slumpterm.

Effekten på timlön:

$$\text{Ln}W_{i81} = \text{Ln}W_{i74}\alpha + X_i\gamma + P_i\beta + \epsilon_i$$

där

$\text{Ln}W_{i81}(\text{Ln}W_{i74}) =$  log av timlön 1981 respektive 1974.

$X_i =$  ett antal bakgrundsvariabler; ålder, kön, utbildningsår 1974, arbetserfarenhet 1974, invandrarstatus, arbetslöshet och arbetslöshetsunderstöd 1974–80, arbetsförhållanden (fem olika indikatorer) 1981.

$P_i =$  deltagande i arbetsmarknadsutbildning.



*Fixed effect model*

Effekt på andelen sysselsatta:

$$Y_{i81} - Y_{i74} = (X_{i81} - X_{i74})\gamma + Z_i\delta + P_i\beta + \epsilon_i$$

där

$(X_{i81} - X_{i74}) =$  förändring i bakgrundskaraktistika; antal hemmabarn, civilstånd.

$Z_i =$  ett antal bakgrundskaraktistika; ålder, kön, utbildningsår 1974, invandrarstatus, arbetslöshetsunderstöd 1974–80.

$P_i =$  deltagande i arbetsmarknadsutbildning.

$\epsilon_i =$  slumpterm.

Effekt på timlön:

$$\ln W_{i81} - \ln W_{i74} = (X_{i81} - X_{i74})\gamma + Z_i\delta + P_i\beta + \epsilon_i$$

där

$(X_{i81} - X_{i74}) =$  förändring i arbetsförhållanden (fem olika indikatorer).

$Z_i =$  ett antal bakgrundskaraktistika; ålder, kön, utbildningsår 1974, arbetserfarenhet 1974, invandrarstatus, arbetslöshet och arbetslöshetsunderstöd 1974–80.

Panel B

*State dependence model*

Effekt på andelen med inkomst under år 1983:

$$Y_{i83} = Y_{i74}\alpha + X_i\gamma + P_i\beta + \epsilon_i$$

där

$Y_{i83} (Y_{i74}) =$  andelen med inkomster 1983 (1974).  
 $X_i =$  ett antal bakgrundskaraktistika; ålder, kön, utbildningsår 1974, civilstånd 1981, antal hemmabarn 1981, invandrarstatus, arbetslöshetsunderstöd 1974–82.  
 $P_i =$  deltagande i arbetsmarknadsutbildning.  
 $\epsilon_i =$  slumpterm.

Effekt på årsinkomsterna för dem med inkomster under året:

$$\text{Ln}E_{i83} = \text{Ln}E_{i74}\alpha + X_i\gamma + P_i\beta + \epsilon_i$$

där

$\text{Ln}E_{i83} (\text{Ln}E_{i74}) =$  log av årsinkomster av tjänst 1983 (1974).  
 $X_i =$  ett antal bakgrundskaraktistika (se närmast föregående ekvation).  
 $P_i =$  deltagande i arbetsmarknadsutbildning.  
 $\epsilon_i =$  slumpterm.

#### *Fixed effect model*

Effekt på andelen med inkomst under år 1983:

$$(Y_{i83} - Y_{i74}) = (X_{i81} - X_{i74})\gamma + Z_i\delta + P_i\beta + \epsilon_i$$

där

$Y_{i83} (Y_{i74}) =$  andelen med inkomster 1983 (1974).  
 $(X_{i81} - X_{i74}) =$  förändring i bakgrundskaraktistika; antal hemmabarn, civilstånd.  
 $Z_i =$  ett antal bakgrundskaraktistika; ålder, kön, utbildningsår 1974, invandrarstatus, arbetslöshetsunderstöd 1974–82.  
 $P_i =$  deltagande i arbetsmarknadsutbildning.  
 $\epsilon_i =$  slumpterm.

Effekt på årsinkomster för dem med inkomster under året:

$$\text{Ln}E_{i83} - \text{Ln}E_{i74} = (X_{i81} - X_{i74})\gamma + Z_i\delta + P_i\beta + \epsilon_i$$

där

$(X_{i81} - X_{i74}) =$  förändring i bakgrundskaraktistika; antal hemmabarn, civilstånd.

$Z_i =$  ett antal bakgrundskaraktistika; ålder, kön, utbildningsår 1974, invandrarstatus, arbetslöshetsunderstöd 1974–82.

$P_i =$  deltagande i arbetsmarknadsutbildning.

$\epsilon_i =$  slumpterm.

## Appendix c

Metoder använda av Heckman, Hotz och Dabos (1987) i den studie som presenterades i tabell 4.2.

### *Linear control function model*

I denna modell utnyttjas tidigare års inkomster (1972, 1973, 1974 respektive 1975 års inkomster) och arbetsmarknadssituation som kontrollvariabler i ekvationer med inkomstnivån 1978 respektive 1979 som beroende variabler. Dessutom ingick en uppsättning personliga karakteristika i ekvationerna.

### *Fixed effect model*

Denna modell har samma grundstruktur som den som användes i Bassis studie. Det framgår inte klart vilka X-variabler som ingår som förändringar i ekvationerna.

Den beroende variabeln är förändringen i inkomster från 1974 till 1978, respektive 1974 till 1979.

### *Random growth model*

Denna modell är en generalisering av fixed effect model. Slumftermerna  $\epsilon_{it}$  respektive  $\epsilon_{is}$  från Bassis specifikation i appendix a ovan antas ha dessa egenskaper:

$$\epsilon_{it} = \phi_{1i} + t\phi_{2i} + V_{it}$$

$$\epsilon_{is} = \phi_{1i} + s\phi_{2i} + V_{is}$$

Varje individ antas således ha en egen nivå –  $\phi_{1i}$  – och en egen tillväxt i inkomsterna  $\phi_{2i}$ . Genom att substituera in dessa slumftermer i den ursprungliga ekvationen erhåller man:

$$\begin{aligned} [(Y_{it}-Y_{is}) - (t-s)(Y_{is}-Y_{is-1})] &= P_i\beta_t + [(X_{it}-X_{is}) - (t-s)(X_{is}-X_{is-1})]\gamma + \\ &+ (V_{it}-V_{is}) - (t-s)(V_{is}-V_{is-1}). \end{aligned}$$

## LITTERATUR

- Bangsor, M, Borus, M, Bloom, H och Orr, L, odatrad, "The National JTPA Evaluation", Evaluation Forum.
- Barnow, B, 1987, "The Impact of CETA Programs on Earnings – A Review of the Literature", Journal of Human Resources, Vol. 22, No. 2, Spring 1987, s 157–193.
- Barnow, B, 1988, "The Uses and Limits of Social Experiments", uppsats presenterad vid AEAs möte i Chicago, december 1987.
- Bassi, L, 1983, "The Effect of CETA on the Postprogram Earnings of Participants", Journal of Human Resources, Fall 1983, s 539–556.
- Bassi, L, 1984, "Estimating the Effect of Training Programs with Non-Random Selection", The Review of Economics and Statistics, February 1984, s 36–43.
- Bassi, L, och Ashenfelter, O, 1986, "The effect of Direct Job Creation and Training Programs on Low-Skilled Workers"; i Fighting Poverty, Danziger, S, och Weinburg, D (redaktörer), Harvard University Press, London.
- Björklund, A, 1983, "Nya metoder inom utvärderingsforskningen", EFA, Arbetsmarknadsdepartementet.
- Björklund, A, 1988, "What Experiments are Needed for Manpower Policy", Journal of Human Resources, Vol. 23. No. 2, Spring 1988.
- Björklund, A, 1989, "Evaluation of Labor Market Training Programs – The Swedish Experience", stencil, The Industrial Institute for Economic and Social Research (Industriens Utredningsinstitut), Stockholm.
- Bloom, H, Borus, M, och Orr, L, 1987, "Using Random Assignment to Evaluate an Ongoing Program: The National JTPA Evaluation", uppsats presenterad vid ASAs möte i San Francisco, augusti 1987.
- Burtless, G och Haveman, R, 1984, "Policy Lessons From Three Labor Market Experiments" i Employment and Training R&D: Lessons Learned and Future Directions, W E Upjohn Institute for Employment Research.
- Burtless, G, 1985, "Are Targeted Wage Subsidies Harmful? Evidence from a Wage Voucher Experiment", Industrial and Labor Relations Review, Vol. 39, No. 1, October 1985, s 105–114.
- Burtless, G och Orr L, 1986, "Are Classical Experiments Needed for Manpower Policy?", Journal of Human Resources, Vol. 21, No. 4, Fall 1986, s 606–639.
- Burtless, G, 1987, "The Social and Scientific Value of Controlled Experimentation", uppsats presenterad vid AEAs möte i Chicago, december 1987.
- Burtless, G, 1988, "The Impact of Welfare Reform on Recipient Employment, Earnings and Income", stencil, Brookings Institution, Washington.
- Delander, L, 1978, "Studier kring den arbetsförmedlande verksamheten", SOU 1978:60.
- Dickinson, K P, Johnson, T R, och West, R W, 1986, "An Analysis of the Impact of CETA Programs on Participants Earnings", Journal of Human Resources, XXI, Winter 1986, s 63–91.
- Fraker, T och Maynard R, 1987, "The Adequacy of Comparison Group Designs for Evaluations of Employment – Related Programs", Journal of Human Resources, Vol. 22, No. 2, Spring 1987, s 194–227.
- Franke, R H och Kaul J D, 1978, "The Hawthorne Experiments: First Statistical Interpretations" American Sociological Review, 43, s 623–643.
- Goldman, B, Friedlander D och Long D, 1986, "Final Report on the San Diego Job Search and Work Experience Demonstration", Manpower Demonstration Research Corporation, New York.
- Gueron, J M, 1987, Reforming Welfare with Work, Ford Foundation, New York.

- Harris, J, 1985, "Macroexperiments versus Microexperiments for Health Policy", i Social Experimentation, J Hausman och D Wise (redaktörer), The University of Chicago Press, Chicago och London.
- Heckman, J J, och Robb, Jr R, 1985a, "Alternative Methods for Evaluating the Impact of Interventions" i Longitudinal Analysis of Labor Market Data, Heckman, J J och Singer B (redaktörer), Cambridge University Press.
- Heckman, J J, och Robb R, Jr, 1985b, "Alternative Methods for Evaluating the Impact of Interventions – an Overview", Journal of Econometrics, 30 (1985), s 239–267.
- Heckman, J J, Hotz V J och Dabos M, 1987, "Do We Need Experimental Data to Evaluate the Impact of Manpower Training on Earnings?", Evaluation Review, Vol. 11, No. 4, August 1987, s 395–427.
- Hollister, R, Kemper P och Maynard R (redaktörer), 1984, The National Supported Work Demonstration, University of Wisconsin Press, Madison, Wisconsin.
- Lalonde, R J, 1984, "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", Industrial Relations Section, Working Paper, No. 183, Princeton University.
- Lalonde, R J, 1986, "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", American Economic Review, September 1986, Vol. 76, No. 4, s 604–620.
- Mathematica Policy Research, 1983, "Final Report: Employment Opportunity Pilot Project: Analysis of Program Impacts", Princeton.
- Meyer, B D , 1988, "Implications of the Illinois Reemployment Bonus Experiments for Theories of Unemployment and Policy Design", Working Paper, No. 242, Industrial Relations Section, Princeton University.
- Orr, L, 1985, "Choosing between Macroexperiments and Microexperiments – a comment on Harris", Social Experimentation, Hausman, J, och Wise, D (redaktörer), The University of Chicago Press, Chicago och London, 1985.
- SOU 1974:29, "Att utvärdera arbetsmarknadspolitik".
- Stromsdorfer, E, odaterad, "The Origin of the National JTPA Evaluation", Evaluation Forum.
- Woodbury, S och Spiegelman R, 1987, "Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois", American Economic Review, Vol. XX, No. 4, September 1987, s 513–530.

## **Publikationer på engelska**

**1989**

**MOSES Code.**

James W. Albrecht et al. Research Report No. 36. 354 pp.

**MOSES Handbook.**

Fredrik Bergholm. Research Report No. 35. 213 pp.

**1988**

**Factor Demand in Swedish  
Manufacturing: Econometric Analyses.**

Joyce Dargay. Research Report No. 34.  
138 pp.

**The Knowledge Base of an Industrial Economy.**

Gunnar Eliasson. Research Report No. 33. 100 pp.

**The Political Economy of the Welfare State.**

James Buchanan. 31 pp.

## **Publikationer på svenska**

**1989**

**Klassiska experiment inom arbetsmarknadspolitiken.**

Anders Björklund. Forskningsrapport nr 37. 84 s.

**1988**

**Den svenska industrins utlandsinvesteringar 1960–1986.**

Birgitta Swedenborg, Göran Johansson-Grahn, Mats Kinnwall.  
208 s.

**Finansiell integration – en studie av svenska marknaders  
internationella beroende.** Lars Oxelheim. 387 s.

**Expansion, avveckling och företagsvärdering i svensk industri.  
En studie av ägarformens och finansmarknadernas betydelse  
för strukturomvandlingen.** Johan Örtengren m fl. 455 s.