

The Effectiveness of
Active Labor Market Programs:
What Have We Learned

David Card
UC Berkeley

(based on work with
Jochen Kluge and Andrea Weber)

Introduction

- ALMP's are widely used as a tool to combat both short-term and long-term joblessness

- programs have been controversial from the start (e.g., New Deal programs in the 1930s)

- evaluation of ALMP's has been goal of the applied econometrics literature since the 1980s

- now a large and "maturing" literature. Previous surveys: Freidlander, Greenberg, Robins (JEL, 1997); Heckman Lalonde Smith (HLE, 1999)....

-the call for credible evaluations of ALMP's (including the use of RCT's) has "spilled over" to other program interventions, particularly in developing country context

Today's goals

- summarize a recent meta-analysis
- draw some general lessons about evaluation methods used in the ALMP literature
- suggest some new directions

I. Institutional Context

-many ALMP's are designed for current participants in income support programs (unemployment and welfare benefit programs). E.g. "New Deal" program in U.K.; job search assistance programs for U.S. welfare recipients.

-extreme example: Swedish UI recipients ultimately forced to enter some type of program

-these ALMP's function both as a human capital development program and as a "cost" of program participation that limits moral hazard

-other ALMP's are designed as remedial education/training program for "disadvantaged" individuals. E.g. JPTA and Job Corps U.S.; *Juventud y Empleo* (JE) program in Dom Rep.

-a theme that will emerge is that the data sources and methods used in evaluation studies, and the impacts of these 2 broad classes are quite different

-a second theme is that more intensive programs take time, and in the short run may have negative effects (like any general human capital investment). So longer term evaluations are necessary

II. A New Sample of Studies

-we contacted all members of IZA network and the NBER Labor Studies program, asking for any studies (post 1995) of ALMP's, and for names of colleagues/co-authors/students who had done studies

-received approximately 200 responses (response rate = 53%), 156 separate studies

-in today's presentation: focus on 199 "program estimates" from 97 studies that met our inclusion criteria

-in scope:

- classroom or on-the-job training
- job search assistance/sanctions
- subsidized private or public employment

-out of scope:

- financial incentives that are universally available (e.g., EITC; enhanced disregards)
- routine services
- theoretical studies
- implementation studies
- studies with no comparison group
- not written in English

- from each study we extracted one or more “program estimates”
- preferred approach is to extract “effect sizes”
- because of the wide diversity of outcomes we could not devise a standardized “effect”
- instead: classify a program estimate as significantly positive, significantly negative, or insignificant
- extract separate estimates by gender and age (over/under age 25) if possible
- also extract short/medium/long-term (1,2,3+ years) effects: we will look at short and longer-term effects *within the same study*

-our studies are recent:

1996-99:	20 estimates
2000-03:	46 estimates
2004-05:	42 estimates
2006:	41 estimates
2007:	50 estimates

Leading “source countries”:

Germany	45	Austria	13
Denmark	25	Switz.	9
Sweden	19	Norway	7
France	14		
U.S.	10		
Slovakia	13		

-strong connection between:

- “meta-type” of program (integrated part of UI/welfare system vs. open enrollment),
- outcome measure
- evaluation method
- country

3 archetypes

“Swedish” model:

- program participants enter from UI rolls
- controls could enter the same program at later date, or some other program
- outcome = time to exit measure
- method = duration modeling

“German” model (largest group):

- program participants enter from UI rolls
- controls could enter the same program at later date, or some other program
- outcome = emp/earnings at later time
- method = longitudinal matching

“JPTA” model (minority):

- program participants = self-selected disadvantaged workers/non-workers
- control group = other “similar” workers
- outcome = employment/earnings some months in the future
- method = longitudinal emp/earnings model

Table 3: Characteristics of Sample of Estimated Program Effects

	Overall Sample (1)	Austria Germany & Switzerland (2)	Scandinavia (3)	Anglo Countries (4)
1. Number of Estimates	199	67	53	20
2. Program Intake				
a. Drawn from Registered Unemployed	68.3	94.0	67.9	15.0
b. Long Term Unemployed (registered and other)	12.6	0.0	3.8	25.0
c. Other (Disadvantaged, etc.)	19.1	6.0	28.3	60.0
3. Type of Program				
a. Classroom/ Work Experience	41.7	62.7	26.5	35.0
b. Job Search Assistance	12.1	7.5	5.7	30.0
c. Subsidized Private Sector Emp.	14.6	3.0	20.8	10.0
d. Subsidized Public Sector Emp.	14.1	16.4	9.4	5.0
e. Threat of Assignment	2.5	0.0	7.5	0.0
f. Combination of Types	15.1	10.4	30.2	20.0

Table 3: Characteristics of Sample of Estimated Program Effects

	Overall Sample (1)	Austria Germany & Switzerland (2)	Scandinavia (3)	Anglo Countries (4)
4. Program Duration				
a. Unknown or Mixed (%)	26.1	11.9	32.1	45.0
b. 4 Months or Less (%)	20.6	26.9	20.8	25.0
c. 5-9 Months (%)	35.2	28.4	43.4	30.0
d. Over 9 Months (%)	18.1	32.8	3.8	0.0
5. Gender of Program Group				
a. Mixed (%)	59.3	55.2	73.6	40.0
b. Male Only (%)	20.6	22.1	13.2	25.0
c. Female Only (%)	16.6	21.0	13.2	35.0
6. Age of Program Group				
a. Mixed (%)	63.8	62.7	56.6	60.0
b. Age Under 25 Only (%)	14.1	0.0	18.9	25.0
c. Age 25 and Older Only (%)	21.6	35.8	24.5	15.0

Table 4: Evaluation Methods Used in Sample of Estimated Program Effects

	Overall Sample (1)	German Speaking (2)	Nordic (3)	Anglo (4)
1. Number of Estimates	199	67	53	20
2. Basic Methodology				
a. Cross Sectional with Comparison Group	3.0	0.0	5.7	0.0
a. Longitudinal with Comparison Group	51.3	80.6	30.2	75.0
c. Duration Model with Comparison Group	36.2	19.4	43.4	0.0
d. Experimental Design	9.1	0.0	18.9	25.0
3. Dependent Variable				
a. Probability of Emp at Future Date	45.7	71.6	17.0	40.0
b. Earnings at Future Date	11.6	4.5	20.8	25.0
c. Duration of Time in Registered Unemp. until Exit to Job	24.6	16.4	35.8	10.0
d. Duration of Time in Registered Unemp. (any type of exit)	6.0	1.5	22.6	0.0
e. Other Duration Measures	3.5	0.0	0.0	0.0
f. Probability of Registered Unemp. at Future Date	6.0	6.0	3.8	25.0

Table 5: Summary of Estimated Impacts of ALM Programs

	Percent of Estimates that are:		
	Positive	Insignificant	Negative
<u>I. Short Term Impact Estimates (~12 Months)</u>			
a. Overall Sample (N=183)	39.3	32.8	27.9
b. Austria, Germany & Switzerland (N=59)	28.8	33.9	37.3
c. Scandinavia (N=50)	46.0	30.0	24.0
d. Anglo Countries (N=17)	70.6	11.8	17.7
<u>II. Medium Term Impact Estimates (~24 Months)</u>			
a. Overall Sample (N=108)	50.0	39.8	10.2
b. Austria, Germany & Switzerland (N=45)	53.3	35.6	11.1
c. Scandinavia (24)	37.5	50.0	12.5
d. Anglo Countries (N=15)	73.3	26.7	0.0

Table 6a: Relation Between Short-Term and Medium-Term Impacts of ALM Programs

	<u>Pct of Medium-Term Estimates that are:</u>		
	Significantly Positive (1)	Insignificant (2)	Significantly Negative (3)
<u>Short Term Impact Estimate:</u>			
a. Significantly Positive (N=30)	90.0	10.0	0.0
b. Insignificant (N=28)	28.6	71.4	0.0
c. Significantly Negative (N=36)	30.6	41.7	27.8

Table 6b: Relation Between Short-Term and Long-Term Impacts of ALM Programs

	Pct of Long-Term Estimates that are:		
	Significantly Positive (1)	Insignificant (2)	Significantly Negative (3)
<u>Short Term Impact Estimate:</u>			
a. Significantly Positive (N=19)	73.7	21.1	5.3
b. Insignificant (N=13)	30.8	69.2	0.0
c. Significantly Negative (N=16)	43.8	43.8	12.5

III. Meta Analysis of Studies

Estimating Model

- assume i^{th} program estimate (b_i) generated by

$$b_i = \beta_i + k_i \frac{\sigma_i}{\sqrt{N_i}} \epsilon_i ,$$

β_i = true treatment effect for program i

σ_i = std. dev. of outcome var used in evaluation

N_i = sample size

k_i = design effect (e.g., comparison of means between 2 groups of equal size, $k_i = \sqrt{2}$)

$\epsilon_i = N(0,1)$ reflecting random factors

In this case t-statistic for i^{th} estimate is:

$$t_i = \sqrt{N_i / k_i} \times (\beta_i / \sigma_i) + \epsilon_i$$

β_i / σ_i = “effect size” of the i^{th} program

$\sqrt{N_i / k_i}$ = “effective sample size”

If the effect size in study i depends on X 's:

$$\beta_i / \sigma_i = X_i \alpha' + \lambda_i ,$$

and the effective sample size is (roughly) constant across studies then

$$t_i = X_i \alpha + \eta_i \quad (\eta_i = \epsilon_i + \lambda_i)$$

(otherwise, t_i will depend on $X_i \times$ eff. sample size)

We implement this as an “ordered probit” for t_i ($t_i < -2$; $-2 \leq t_i \leq 2$; $t_i > 2$).

We can test the ordered probit setup by fitting separate probits for $t_i < -2$ and $t_i > 2$. Our simple model asserts the probits will have equal and opposite coefficients on the X 's.

Using the probits we can also *test* if effective sample size is constant. If not, studies with large N_i 's will tend to lead to more significant positive (and significant negative) impact estimates.

Table 7: Ordered Probit Models for Sign/Significance of Estimated Short-term Program Impacts

	Dep var = indicator for sign/significance of estimated impact					
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Dummies for Dependent Variable (omitted=Post-program employment)</u>						
1. Time in Reg. Unemp. Until Exit to Job	0.59 (0.21)	--	--	--	0.45 (0.23)	0.29 (0.26)
2. Time in Registered Unemp.	1.05 (0.33)	--	--	--	1.00 (0.38)	0.99 (0.44)
3. Other Duration Measure	0.38 (0.42)	--	--	--	0.34 (0.44)	0.03 (0.49)
4. Prob. Of Registered Unemp.	1.43 (0.49)	--	--	--	1.37 (0.50)	1.11 (0.53)
5. Post-program Earnings	0.29 (0.30)	--	--	--	0.21 (0.32)	0.03 (0.37)
<u>Dummies for Type of Program (omitted=Mixed and Other)</u>						
6. Classroom or On-the-Job Training	--	-0.40 (0.26)	--	--	-0.04 (0.31)	0.03 (0.36)
7. Job Search Assistance	--	0.38 (0.33)	--	--	0.54 (0.37)	0.65 (0.44)
8. Subsidized Private Sector Job	--	-0.43 (0.31)	--	--	-0.11 (0.34)	-0.12 (0.38)
9. Subsidized Public Sector Job	--	-0.71 (0.32)	--	--	-0.50 (0.37)	-0.46 (0.42)

Table 8: Ordered Probit Models for Sign/Significance of Estimated Medium-term Program Impacts

	Dep var = indicator for sign/significance of estimated impact					
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Dummies for Dependent Variable (omitted=Post-program employment)</u>						
1. Time in Reg. Unemp. Until Exit to Job	0.55 (0.26)	--	--	--	1.21 (0.69)	0.90 (0.73)
2. Other Duration Measure	0.28 (0.84)	--	--	--	0.38 (0.99)	0.45 (0.99)
3. Prob. Of Registered Unemp.	0.63 (0.74)	--	--	--	0.33 (0.77)	0.38 (0.79)
4. Post-program Earnings	0.22 (0.31)	--	--	--	0.04 (0.38)	0.09 (0.38)
<u>Dummies for Type of Program (omitted=Mixed and Other)</u>						
6. Classroom or On-the-Job Training	--	0.56 (0.40)	--	--	0.86 (0.51)	0.95 (0.51)
7. Job Search Assistance	--	0.66 (0.58)	--	--	0.48 (0.69)	0.53 (0.78)
8. Subsidized Private Sector Job	--	0.24 (0.53)	--	--	0.25 (0.61)	0.32 (0.62)
9. Subsidized Public Sector Job	--	-0.58 (0.47)	--	--	-0.82 (0.60)	-0.80 (0.60)

Table 8: Ordered Probit Models for Sign/Significance of Estimated Medium-term Program Impacts

	Dep var = indicator for sign/significance of estimated impact					
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Dummies for Age and Gender of Participants (omitted=Pooled Age, Pooled Gender)</u>						
10. Age Under 25 Only	--	--	-0.83 (0.36)	--	-0.89 (0.41)	-0.87 (0.41)
11. Age 25 and Older Only	--	--	-0.39 (0.30)	--	-1.12 (0.41)	-1.21 (0.42)
12. Men Only	--	--	-0.40 (0.34)	--	-0.04 (0.40)	-0.17 (0.42)
13. Women Only	--	--	0.28 (0.32)	--	0.51 (0.37)	0.41 (0.39)
<u>Dummies for Program Duration (omitted=5-9 month duration)</u>						
14. Unknown or Mixed	--	--	--	-0.72 (0.33)	-1.10 (0.41)	-1.05 (0.42)
15. Short (≤ 4 Months)	--	--	--	0.26 (0.34)	-0.43 (0.41)	-0.53 (0.43)
16. Long (> 9 Months)	--	--	--	-0.06 (0.33)	-0.32 (0.39)	-0.28 (0.39)
17. Dummies for Intake Group and Timing of Program, and Country Grp.	No	No	No	No	No	Yes
18. Dummy for Experimental Design	--	--	--	--	--	0.15 (0.83)
19. Square Root of Sample Size (Coefficient $\times 1000$)	--	--	--	--	--	1.13 (0.87)

Table 9: Ordered Probit Models for Change in Program Impacts from Short-term to Medium-term

	Dep var= change in sign/significance of estimated impact					
	(1)	(2)	(3)	(4)	(5)	(6)
<u>Dummies for Dependent Variable (omitted=Post-program employment)</u>						
1. Time in Reg. Unemp. Until Exit to Job	-1.71 (0.66)	--	--	--	-1.36 (0.76)	-1.48 (0.81)
2. Other Duration Measure	-0.03 (0.81)	--	--	--	0.46 (1.00)	0.48 (1.00)
3. Prob. Of Registered Unemp.	-1.12 (0.82)	--	--	--	-0.76 (0.90)	-0.64 (0.92)
4. Post-program Earnings	-0.17 (0.32)	--	--	--	-0.18 (0.38)	-0.16 (0.39)
<u>Dummies for Type of Program (omitted=Mixed and Other)</u>						
6. Classroom or On-the-Job Training	--	0.88 (0.44)	--	--	0.92 (0.57)	0.96 (0.58)
7. Job Search Assistance	--	0.10 (0.62)	--	--	0.35 (0.75)	0.54 (0.82)
8. Subsidized Private Sector Job	--	0.27 (0.57)	--	--	0.24 (0.68)	0.27 (0.68)
9. Subsidized Public Sector Job	--	0.67 (0.52)	--	--	0.55 (0.65)	0.59 (0.66)

IV Conclusions

1. outcome measures matter. Register-based measures seem to overstate short term positive effects
2. human capital enhancing programs are more effective in the medium/long run than they appear to be after 1 year. Evaluations need to have longer time frames to capture the benefits of the programs
3. recent non-experimental estimators do not appear to be biased

4. Evaluations often lack the required information to conduct even a “back of the envelop” cost-benefit assessment. Authors should do a better job of translating program effects into “welfare-relevant” outcomes (employment or earnings), and should try to at least crudely measure costs

5. More work on the channels by which programs affect outcomes is useful.