

THE EFFECTIVENESS OF ACTIVE LABOR MARKET POLICIES: A META-ANALYSIS

Melvin Vooren 

*Department of Economics
University of Amsterdam*

Carla Haelermans and Wim Groot

*Top Institute for Evidence Based Education Research
Maastricht University*

Henriëtte Maassen van den Brink

*Top Institute for Evidence Based Education Research
University of Amsterdam*

Abstract. This paper provides a meta-analysis of microeconomic evaluation studies on the effectiveness of active labor market policies. The analysis is built upon a systematically assembled data set of causal impact estimates from 57 experimental and quasi-experimental studies, providing 654 estimates published between January 1990 and December 2017. We distinguish between the short and longer term impacts in our analysis; at 6, 12, 24, and 36 months after program start. After correcting for publication bias and country-specific macroeconomic characteristics, subsidized labor and public employment programs have negative short-term impacts, which gradually turn positive in the longer run. Schemes with enhanced services including job-search assistance and training programs do not have these negative short-term effects, and stay positive from 6 until 36 months after program start.

Keywords. Active labor market policy evaluation; Meta-analysis; Effect size; Publication bias; Meta-regression

1. Introduction

In Western countries, a considerable amount of public money is spent on active labor market policies (ALMPs) to enhance the labor market prospects and to decrease the welfare dependency of the unemployed. In 2013, average public expenditures on ALMPs amounted to 0.5% of Gross Domestic Product (GDP) in the OECD countries, which is equivalent to 1.0% of total government expenditures (OECD, 2017). Governments have various reasons to invest in ALMPs. Aside from the individual negative aspects of unemployment – for instance, the loss of income and depreciation of human capital – unemployment benefits also weigh heavily on the national budget.

Corresponding author contact email: M.Vooren@uva.nl.

The renewed interest in ALMPs in the 1990s – exemplified by the British New Deal, the Welfare-to-Work reforms signed under the Clinton administration, and similar efforts undertaken by other governments (see Robinson, 2000; Bonoli, 2010) – has led to a large number of microeconomic program evaluations that have been published in scientific journals. These evaluation studies can be informative for the design of future policy, as they can tell whether a previous intervention had been successful in improving the labor market outcomes of its participants. Proven effective programs can then be implemented elsewhere and scaled up. In this paper we present a meta-analysis based upon a new, systematically assembled data set of experimental and quasi-experimental impact evaluation studies.

Individual impact evaluations typically assess the effectiveness of a particular program, on a particular group, and in a particular period of time. Accordingly, individual studies do not provide a general answer to the question which program types are effective and under which circumstances, even though these studies have a solid experimental or quasi-experimental design. For example, Caliendo and Künn (2015) look at the effectiveness of a specific type of labor subsidy (a start-up subsidy) on the re-employment rates of unemployed females. Dorsett *et al.* (2013) on the other hand conduct a randomized experiment of a labor market program involving enhanced job-search services in the UK, while Alegre *et al.* (2015) perform an impact evaluation of training programs in the Spanish region of Catalonia. Some studies do evaluate multiple program types in different regions, such as Dyke *et al.* (2006) who assesses the effects of the Temporary Assistance for Needy Families (TANF) programs in Missouri and North Carolina, but it remains difficult to extrapolate these lessons to another region or time period. Meta-analysis provides a way to generalize the lessons from this heterogeneous collection of studies. A meta-analysis quantifies the size of the program effect, and controls for the implemented evaluation design, the background settings, and publication bias.

Card *et al.* (2010) conduct a meta-analysis based on the sign and significance of the effects of ALMPs, covering the period from 1995 up to 2007. In addition to this, Card *et al.* (2017) provide an update using a Google Scholar search to identify all studies citing Card *et al.* (2010), up until 2014. With respect to labor market outcomes, both studies point at the same pattern of results with respect to differences in the effectiveness between training, subsidized labor, and enhanced services programs. Also the short, medium, and long-term effectiveness differs: training programs appear to be ineffective in the short term, but they have a positive impact in the medium term. In the literature this phenomenon is known as the lock-in effect, which we further explain in Section 2. Kluve (2010) conducts a similar meta-analysis of the sign and significance of the effectiveness of European ALMPs. He also concludes that it is mostly the program type that explains the variation in program effectiveness. In a systematic review Filges *et al.* (2016) conclude that there is a general effect of participating in ALMPs, though the effect is small. Reviews by Crépon and van den Berg (2016) and McKenzie (2017) also highlight that ALMPs turn out to be far less effective than expected.

However, the meta-analysis of Card *et al.* (2010) is restricted to a partially complete sample of studies from the Institute for the Study of Labor (IZA) and the National Bureau of Economic Research (NBER). It is based on a survey of their fellows, of whom 55% responded. The data set used by Card *et al.* (2017) is similar to the one used in Card *et al.* (2010), but includes the studies citing Card *et al.* (2010), derived from Google Scholar. The analysis of Kluve (2010) not only includes experimental and quasi-experimental effect studies, but also non-experimental non-causal studies. Both of these previously published meta-analyses are based on samples assembled in the beginning of 2007, and do not measure the magnitude of the effect sizes, but only consider the sign and significance of the effects. Statistically significant yet relatively small impact estimators are treated in the same way as larger ones, which is a significant drawback. Since active labor market programs are costly, the magnitude of the effect is important for policymakers for cost-benefit assessments.

In this paper, we improve on the existing meta-analytic literature in multiple ways. We adopt a systematic search and reporting procedure in which we incorporate standardized effect size estimates, following

Stanley *et al.* (2013). These standardized effect sizes allow for a comparison between the different types of ALMPs in a meta-regression model, while controlling for study-specific characteristics (after Stanley and Doucouliagos, 2014). We also incorporate information on the macroeconomic conditions in the country the program was administered, as these might be related to the effectiveness as well. Furthermore, we base our analysis on a data set including only causal effect studies to rule out any selection effects at study level. We further extend the scope of the meta-analytic evidence by expanding the time period up to and until December 2017. Finally, we also include a theoretical framework, which we also use to relate our findings to.

To answer the question which programs are most effective and under which circumstances, we discriminate between the following program types: (i) training and retraining programs, which are aimed at the formation of human capital, (ii) subsidized labor schemes, including working tax credits and start-up subsidies, (iii) public sector employment schemes, in which the government attempts to directly hire the unemployed, and (iv) enhanced services schemes, including job-search assistance and regular encounters with caseworkers, sometimes accompanied by sanctions in case the participant does not fulfill certain participation criteria.

Many studies in our sample make a distinction between short and long-term program effects, allowing for the analysis of the effects over time as well. We extract the program impact estimates on the participants' labor market outcomes in terms of standardized effect sizes at 6, 12, 24, and 36 months after program start.

After controlling for country-specific macroeconomic background characteristics and publication bias, we find that ALMPs are generally successful in improving the labor market outcomes of their participants, yet the effects are small ($d < 0.1$). There is, however, a disparity in the effectiveness between the different program types. Public sector employment schemes, characterized by job creation in the public sector, as well as subsidized labor, have *negative* impacts in the short term. These negative “lock-in” effects turn into positive impacts over time. These lock-in effects of subsidized labor programs tend to last shorter than those of public employment schemes. The impact of subsidized labor turns positive after 12 months, whereas with public employment this is the case only after 36 months. Both enhanced services and training programs have positive impacts in both the short and longer run.

The remainder of this paper is structured as follows. In the next section we provide a brief theoretical framework. In Section 3 we describe our sample, including our search strategy and selection protocol. In Section 4 we briefly discuss our meta-analytic models, followed by our main results. We present our concluding remarks and policy implications in Section 5.

2. Theoretical Background

Over the past three decades, many OECD countries have transformed their policies with respect to the labor market from “passive”, i.e. unemployment insurance and welfare benefits, to “active”. The basic idea behind this “activation” concept is to reduce structural unemployment by steering the unemployed to find employment and/or to promote wage growth (Martin 2015). In order to understand and explain why particular activation programs work better than others, it is useful to consider a theoretical framework that identifies the core differences between these activation strategies.

Overviews by for example, Heckman *et al.* (1999) and Martin (2015) generally divide ALMPs in the following main categories: (a) classroom training; (b) subsidized employment, which both includes entirely subsidized temporary government jobs on the one hand, and subsidized entry-level jobs for young people on the other hand; (c) subsidies to private firms to promote the provision of on-the-job training; (d) training in how to obtain a job; (e) direct support to job search, including referrals to employers and free access to job listings. In this meta-analysis we discriminate between training, subsidized labor, public employment, and enhanced services/job-search assistance. Training corresponds to category (a),

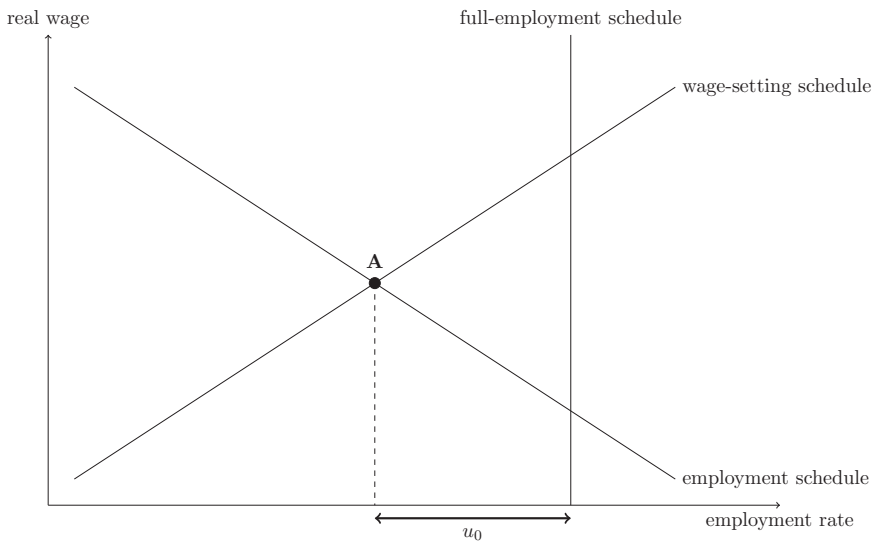


Figure 1. The Layard–Nickell (1986) Model.

Notes: The difference between the equilibrium employment rate A and the full–employment rate is the involuntary unemployment rate u_0 .

subsidized labor combines categories (b) and (c), public employment is in line with category (b), and enhanced services is a combination of categories (d) and (e).

Heckman *et al.* (1999) advocate the use of general equilibrium models to develop a theoretical rationale to the effects of ALMP, especially since most ALMPs target large populations, like the entire stock of unemployed in a particular country. To explain the effects of ALMP on equilibrium employment, Calmfors (1994) utilizes a model introduced by Richard Layard and Stephen Nickell (see Layard and Nickell, 1986; Johnson and Layard, 1986; Layard *et al.*, 2005). A graphical representation of this model is displayed in Figure 1. The employment schedule is downward-sloping and relates employment, which in this model equals labor demand, to the real wage. The upward-sloping curve represents the wage-setting schedule, showing that higher aggregate employment levels coincide with higher real wages.¹ The equilibrium values of the employment level and the real wage are at point A, where both curves intersect. The vertical line indicates full employment of the labor force, and involuntary open unemployment is then represented by u_0 , the difference between this line and point A.

ALMPs may improve the matching process on the labor market to resolve unemployment, through for instance the elimination of mismatch between the skills that are requested by employers and labor demand via training programs. Intensified and more active searching through enhanced services programs also improve the matching process. Subsidized labor and public employment schemes can have a screening function for uncertainty about the employability of applicants, where program participation provides a substitute for work experience.

Improved matching affects both the employment and wage-setting schedules in the Layard–Nickell model. When vacancies are filled more quickly, it becomes less costly for firms to open vacancies and recruit personnel. This incentivizes firms to open up more vacancies, which corresponds to an increase in labor demand and a shift of the employment schedule to the right (Calmfors and Lang, 1995). With increased matching efficiency and vacancies becoming filled in more quickly, firms do not have to

offer high wages to attract workers anymore. This mechanism causes the wage-setting schedule to shift downwards (Johnson and Layard, 1986). Both effects decrease involuntary employment, shifting point A more toward the full-employment schedule. There are, however, also effects working in the opposite direction. During the period of program participation, the participants are likely less actively engaged in the job-search process. These negative “lock-in” effects are an important aspect when considering the effectiveness of ALMPs. This meta-analysis reveals which program types are associated with the largest lock-in effects, and for which programs the positive unemployment-reducing effects dominate over time.

3. A Dataset of Effect Estimates

3.1 Search Strategy and Selection Protocol

To assemble a new data set of ALMP evaluations, we have searched the Web of Science core collection for studies containing either one of the terms “Active Labor Market”, “Active Labour Market”, “Welfare-to-Work”, “Activation Program”, or “Activation Programme” together with the term “effect”. This search was completed at the end of January 2018. Before extracting the causal treatment effect estimates from empirical studies, we first have to define our selection protocol. The inclusion criteria of this “protocol” allow us to make a clear demarcation of which studies to include or not to include. Our protocol includes only studies that satisfy all of the following criteria: (i) the study focuses specifically on the evaluation of ALMPs, (ii) the composition of the intervention is clear, (iii) the study is based on experimental or quasi-experimental data, (iv) it has been published in a peer-reviewed journal, (v) it is in the English language, and (vi) it has been published between January 1990 and the December 2017.

The reasoning behind imposing these restrictions is as follows. We only include studies that focus specifically on the evaluation of active labor market programs, excluding all papers focusing solely on methodological questions, while just providing an application of the discussed method to a program that has been evaluated in an earlier study. We also need to be sure of the composition of the intervention, in the sense that it has to fit in one of the categories stated in Section 2. This criterion boils down to the fact that we leave out any “mixed” program estimates, which fit in more than one of our categories. Because the compositions are not straightforward across the literature, our meta-analysis cannot synthesize this “mixed” category accurately.

To account for possible selection bias into an active labor market program, we only consider randomized control trials (RCT) and quasi-experimental studies, ensuring a proper identification of the program effect. Of the quasi-experimental identification approaches, we include (i) matching methods, (ii) difference-in-differences, (iii) regression discontinuity designs, and (iv) instrumental variables in our sample. Earlier meta-analyses, such as Kluge (2010) have not imposed this restriction on their data set. Yet, this is crucial because ALMP evaluations typically have to deal with selection biases. Without a causal identification strategy it is difficult to control for these selection biases. Frölich (2004) provides an overview of the main identification strategies applied in microeconomic policy evaluation.

We exclude working papers from our analysis, and only include peer-reviewed studies. We do so because the scientific quality of these studies has been assessed through the peer-review process. We only include studies published in English for the sake of accessibility and correct interpretation. We furthermore restrict our sample to studies published between January 1990 and December 2017. After filtering out the studies that do not satisfy our inclusion criteria, we have 57 studies left. Because we perform four separate analyses at 6, 12, 24, and 36 months after the start of the program, we only code the effect estimates at these points in time. We omit any available estimates at other moments after program start. In total, these 57 studies contain 799 effect estimates on different moments after the start of the program. We eventually end up with 654 unique program estimates. In the next subsection we elaborate on how we arrive at 654 observations for our final analyses.

3.2 Extraction of Standardized Effect Size Estimates

Before we can make an appropriate comparison between the different effect estimates, we computed standardized effect sizes using the standardized mean difference method. The effect size ES_{sm} , or Cohen's d , is then defined as the difference in the sample means of the control and treatment groups, divided by the pooled standard deviation. This pooled standard deviation is calculated as:

$$s_{pooled} = \sqrt{\frac{(n_{treatment} - 1)s_{treatment}^2 + (n_{control} - 1)s_{control}^2}{n_{treatment} + n_{control} - 2}}. \quad (1)$$

While most studies do not report the variances of the treatment and the control groups, they do report the results of a simple test comparing both means. The effect size and its standard error can then be calculated using the following formulas, using the t -statistic of this mean-difference test (Lipsey and Wilson, 2001, p. 198):

$$ES_{sm} = t \sqrt{\frac{n_{treatment} + n_{control}}{n_{treatment}n_{control}}}, \quad (2a)$$

$$SE(ES_{sm}) = \sqrt{\frac{n_1 + n_2}{n_1 n_2} + \frac{ES_{Psm}^2}{2(n_1 + n_2)}}. \quad (2b)$$

We have applied Equations 2a and 2b to all of the impact estimators from the studies in our data set to calculate the effect sizes, as well as their standard errors. When the t -statistics or p -values could not be extracted from the study, we have used the t -value corresponding to the reported significance level in the case of a statistically significant result, as this is the most accurate approximation we could make in that case. In the case of a statistically insignificant result we have coded a t -statistic corresponding to a p -value of 0.99. These approximations are both *underestimates* of the exact effect size. We have done this for approximately 35% of the observations.

Labor market outcomes can be captured in a variety of positively and negatively defined outcome variables. The probability of being employed is a positive outcome variable, whereas the probability of being unemployed is a negative outcome variable. Since these dependent variables vary significantly, we have defined the outcome variables in positive terms. Doing so, the marginal effects can generally be interpreted as an improvement in the overall labor market status of the participant. Also, some studies report multiple outcome variables for the same subgroup. Some studies for example report the effect on earnings next to the effect on the probability of being employed. In that case we have kept the outcome estimate we consider as the most important, in the order of the entries in Table 1, which shows the relative frequencies of the positive (+) and negative (−) outcomes after applying this procedure. For each study, we have used the outcome variable with the lowest “number”, as shown in the first column, resulting in 654 effect estimations.² We took this order because *not* being unemployed, regardless of earnings, prevents human capital depreciation through the loss of workers' job skills. The first column shows the distribution of outcome variables including duplicate effect sizes for the same treatment/subgroup. The second column shows the distribution of outcome variables after applying our procedure to remove these duplicates. In here we take the outcome variable with the highest “ranking” per treatment/subgroup. Thus, when one study reports both the effect on the probability of being employed as the effect on earnings, we take the effect on the probability of employment for our main analysis. We control for the category of outcome variable in the meta-analytic model later on to see whether this influences our results, by performing a separate analysis only for the studies that report the probability of being employed as a robustness check.

Table 1. Distribution of Outcome Variables.

Outcome variables by order of importance			Sign (+/–)	With duplicates		One outcome per treatment	
				Freq.	%	Freq.	%
1	Probability of being employed	+	544	66.09	544	83.18	
2	Probability of not being unemployed	+	30	3.75	0	0.00	
3	Probability of unemployment	–	68	8.51	25	3.82	
4	Duration of employment	+	36	4.51	36	5.50	
5	Duration of unemployment	–	19	2.38	17	2.60	
6	Earnings	+	102	12.77	32	4.89	
			799	100.00	654	100.00	

Notes: The first column shows the distribution of outcome variables including duplicate effect sizes for the same treatment/subgroup. The second column shows the distribution of outcome variables after applying a procedure to remove these duplicates, that is, by taking the outcome variable with the highest “ranking” per treatment/subgroup.

3.3 Background Characteristics

To determine which program types are the most effective in improving the labor market outcomes of the participants, we code a dummy for each program type as explanatory variables into our data set. Furthermore, we incorporate background characteristics, such as the duration of the intervention and the gender of the participants. We also encode dummy variables for the evaluation design, enabling us to distinguish between them in our meta-analytic model. The estimated program effects can be considered country-specific, due to differences in the macroeconomic conditions. A recession can be responsible for an increase in unemployment for which an ALMP has been designed. Due to the cyclical nature of unemployment, human capital-enhancing ALMPs aiming at structural unemployment can be ineffective. To take account for these conjectural variations, we include the unemployment rate in the year the program was introduced.³

4. Analysis and Results

4.1 Descriptive Statistics

Before proceeding to the weighted mean effects and the multivariate meta-analysis, we first present our sample descriptively. A substantial part of the sample consists of evaluation studies of German programs that have been implemented in the beginning of the 1990s. This may partly be attributable to the efforts undertaken in Germany following the reunification in 1990. There is also a noticeable increase in the amount of causal effect studies just before the beginning of the 21st century. This might be partly attributable to the ongoing methodological progress in the econometric evaluation literature, of which Imbens and Wooldridge (2009) present a historical review. The welfare reforms in the United States (Personal Responsibility and Work Opportunity Reconciliation Act, 1996) and the United Kingdom (New Deal for Young People, 1998; Working Families Tax Credit, 1999) also explain these developments. The German Harz reforms that have been implemented since 2003 have also been evaluated frequently.

Table 2 presents a breakdown of the effect estimates by sample characteristics. Training and retraining programs have been evaluated most frequently, followed by programs regarding labor subsidies. This group consists of subsidies for “regular” jobs or private work practice, in contrast to directly created public

Table 2. Sample Characteristics: Estimated Program Effects.

	Frequency (1)	Percentage (2)	Cumulative (3)
<i>(a) by program type</i>			
Enhanced services/sanctioning	82	12.54	12.54
Public employment	73	11.16	23.70
Subsidized labor	247	37.77	61.47
Training or retraining	252	38.53	100.00
<i>(b) by country</i>			
Austria	48	7.34	7.34
Belgium	1	0.15	7.49
Denmark	36	5.50	13.00
Germany	372	56.88	69.88
New Zealand	9	1.38	71.25
Norway	37	5.66	76.91
Poland	26	3.89	80.89
Portugal	4	0.61	81.50
Romania	4	0.61	82.11
Russia	2	0.31	82.42
Serbia	2	0.31	82.72
Slovak Republic	2	0.31	83.03
Spain	16	2.45	85.47
Sweden	35	5.35	90.83
Switzerland	19	2.91	93.73
United Kingdom	8	1.22	94.95
United States	33	5.05	100.00
<i>(c) by program duration</i>			
Up to 6 months	234	35.80	35.80
Between 7 and 12 months	143	21.87	57.67
Longer than 12 months	42	6.42	64.09
Not reported	268	40.98	100.00
<i>(d) by decennium of program introduction</i>			
1980s	25	3.82	3.82
1990s	224	34.25	38.07
2000s	399	61.01	99.08
2010s	6	0.92	100.00
<i>(e) by gender</i>			
Only men	190	29.05	29.05
Only women	179	27.37	56.42
Both men and women	285	43.58	100.00
<i>(f) by evaluation design</i>			
Difference-in-Differences	22	3.36	3.06
Instrumental Variables	8	1.22	4.59
Matching	576	88.07	92.66
Randomized Experiment	48	7.34	100.00

Notes: Based on 654 effect estimates from 57 studies.

employment schemes. Enhanced services, including job-search assistance, and sanctioning programs occur less frequently in our sample. This does not necessarily imply that these programs are less frequently implemented, nor that they are evaluated less often, but this pattern may also be attributable to our methodological inclusion criteria.

Roughly half of our sample consists of effect estimates of programs implemented in Germany. Many of these studies consider evaluations of programs implemented after the German reunification in 1990, followed by the East-German transition to a market economy and the analogous increase in unemployment. Furthermore, the German Hartz reforms came into effect in the first half of the 2000s. These reforms included many ALMP measures. Due to the controversies about the Hartz reforms, many evaluation studies have been performed.

Other Eastern European countries, that is, Poland, Romania, Russia, Serbia, and the Slovak Republic account for 36 program estimates, resembling roughly 5% of our sample. Anglo countries, being New Zealand, the United Kingdom, and the United States, contribute to 50 program estimates, which is roughly equivalent to another 7.5% of our sample. Nordic countries, consisting of Denmark, Norway, and Sweden together contribute to 108 program estimates, roughly 16.5% of our sample. The variety in countries and regions also indicates the importance of the different institutional environments. To take account of these institutional environments we include year and country indicators, or macroeconomic indicators, in our multivariate meta-regression further in this paper.

Over one-third of the effect estimates refer to programs that do not last longer than a year until completion. About 6% of the estimates concern programs that take longer than a year on average. For another one-third of the estimates the average program duration is not specified. Nearly all programs took place in the 1990s and the 2000s. With respect to the evaluation designs, more than 85% of the effect estimates are based on matching methods, followed by randomized experiments (7.5%) and difference-in-differences setups (3%).

4.2 Mean Effect Sizes

Table 3 displays some more basic sample characteristics of the effect sizes and the overall impacts of each program type at 12 and 24 months after the start of the program. The overall impacts at 6 and 36 months after program start are shown in Table A2. The unweighted mean effect sizes are shown in column 1. The weighted mean effect sizes under the fixed-effects model are shown in column 3. This fixed-effects model assumes homogeneity in the error term across studies. However, in our sample this assumption might not be realistic due to differences in settings and participants across studies.

We can check whether the distribution of the effect sizes is truly homogeneous by performing a statistical test for homogeneity developed by Hedges (1982, p. 493). This test is based on the Q -statistic, which has a chi-squared distribution and $k - 1$ degrees of freedom, where k equals the number of effect sizes.⁴ The null hypothesis of homogeneity is rejected in each subgroup, implying heterogeneous standard errors of the effect sizes in our sample. One approach to account for heterogeneity is by adding an additional error term that varies across studies. The random-effects model takes account for this, of which the results are shown in column 4. In addition to the random-effects weighted averages, we estimate an unrestricted fixed-effects weighted least squares (FE-WLS) model. This is obtained by regressing the effect size divided by its standard error on the inverse standard error and no constant, as proposed by Stanley and Doucouliagos (2015). Given the heterogeneity in our dataset, this is our preferred model, as from their simulations it seems that this approach produces the least biased results, compared to the random effects model. These WLS results are shown in column 5.

In the FE-WLS model, enhanced services are shown to have a statistically insignificant effect in the short term (up to one year after the start of the program) and a small negative effect in the longer term (from one year onwards). Public sector employment programs have a negative impact overall, although the

Table 3. Mean and Weighted Program Effects by Programme Type.

	Sample Mean (1)	SD (2)	Fixed-effects (3)	Random-effects (4)	FE-WLS (5)
<i>(i) 12 months after program start</i>					
Enhanced services (22)	0.023	0.063	0.004**	0.023**	0.004
Public employment (21)	−0.124	0.146	−0.048***	−0.056***	−0.048***
Subsidized labor (79)	−0.078	0.210	−0.054***	−0.076***	−0.054***
Training/retraining (83)	−0.020	0.123	−0.038***	−0.041***	−0.038***
Overall effect (205)	−0.048	0.123	−0.023***	−0.045***	−0.023***
<i>(ii) 24 months after program start</i>					
Enhanced services (23)	0.018	0.041	−0.009***	0.011***	−0.009**
Public employment (20)	−0.168	0.247	−0.032***	−0.040***	−0.032*
Subsidized labor (66)	0.000	0.138	−0.001	0.001***	−0.001
Training/retraining (64)	0.034	0.113	−0.001	0.011***	−0.001
Overall effect (173)	−0.004	0.150	−0.005***	0.004***	−0.005*

Notes: ***, **, * denote 1%, 5%, and 10% significance levels, respectively. Numbers of observations in parentheses. Mean and weighted program effects on 6 and 36 months after program start can be found in Table A2.

longer term impact is slightly less negative compared to the short-term impact. Private-sector subsidized labor seems to have a negative impact in the short run, which turns into a positive impact in the longer run. Training and retraining programs have negative effects in the short term, and insignificant effects in the longer term. The FE-WLS results for subsidized labor and training and retraining turn positive at 36 months after program start (see Table A2). However, in this mean-effect analysis we do not control for publication bias and background characteristics. In the next part of this paper, we address these issues and provide our main results.

4.3 Publication Bias

In our selection criteria, we stipulate that only published, peer-reviewed studies are to be included in the sample. This is to ensure the quality of the effect estimates we base this meta-analysis on. However, when academic journals not only select studies for publication based on the quality of the research design, this could constitute a selection bias. If studies that report relatively high effect sizes are more likely to be published, this would be also the case in our sample. These missing studies would threaten the internal validity of the meta-analysis, because then the sample would not reflect the true distribution of effect sizes. A way of visualizing this publication bias is to plot the effect sizes against their standard errors in a so-called funnel plot. In the absence of publication bias, the studies will be evenly distributed around the mean in this funnel plot (Stanley and Doucouliagos, 2010).

Next to this visual representation of the publication bias, we can formally test for it using a test that has been developed by Egger *et al.* (1997). The Egger test is based on a simple (meta-)regression of the standardized effect estimate against its standard error:

$$ES_i = \alpha_0 + \beta_0 SE_i + u_i. \quad (3)$$

Because this meta-regression model (MRA) contains heteroscedasticity, Equation 3 should be estimated by Weighted Least Squares (WLS) instead of Ordinary Least Squares (OLS), using $1/SE_i^2$ as weights (Stanley, 2005). Alternatively, WLS can be obtained by dividing Equation 3 by an estimate of the standard

deviation of this heteroscedasticity: SE_i . This brings us to the following model that can be estimated by a regular OLS procedure:

$$t_i = \beta_0 + \alpha_0 \frac{1}{SE_i} + v_i. \quad (4)$$

Here, $(1/SE_i)$ gives an indication of the precision of the estimate. Then, a rejection of the null hypothesis under which the intercept is equal to zero indicates publication bias. This test is known as the Funnel Plot Asymmetry Test (FAT) (Stanley, 2008). However, on the basis of Monte Carlo simulations, Stanley (2008) argues that this setup gives biased results. To improve on this, Stanley and Doucouliagos (2014) propose a quadratic version of the FAT–PET as a starting point of a meta-regression analysis (MRA) to get the Precision Effect Estimate with Standard Error (PEESE) estimate for $\hat{\beta}_0$ (Stanley and Doucouliagos, 2014), by applying WLS to the following model while using $1/SE_i^2$ as weights:

$$ES_i = \beta_0 + \alpha_0 SE_i^2 + u_i. \quad (5)$$

Or equivalently, by applying OLS after dividing Equation 5 by SE_i :

$$t_i = \alpha_0 SE_i + \beta_0 \frac{1}{SE_i} + v_i. \quad (6)$$

As a first approach to assess the possibility of a publication bias, we have applied FAT (Equation 4) to the entire data set to give a general indication. However, since many of our effect size estimates are derived from p -values based on confidence intervals, many estimates are on the 95% confidence band in the funnel plot, as can be seen in Figure 2, where the red x-marks represent the effect sizes that have been

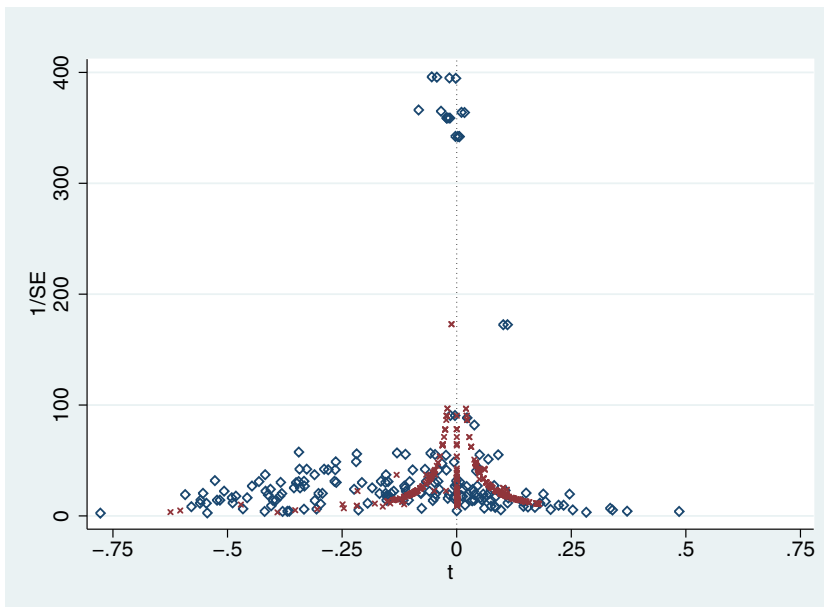


Figure 2. Funnel Plot, Entire Sample. [Colour figure can be viewed at wileyonlinelibrary.com]

Notes: On the vertical axis $1/SE$ depicts the inverse standard error of the effect size, and on the horizontal axis t represents the effect size divided by its standard error. The red x-marks depict effect size approximations based on 95% confidence intervals, and the blue diamonds depict exact effects sizes.

Table 4. FAT–PET and PEESE Results.

Panel A: FAT–PET				
	FAT–PET ($n = 654$)		FAT–PET ($n = 220$)	
	Coef.	p -value	Coef.	p -value
precision (β_0)	−0.0140***	(0.000)	−0.0101***	(0.015)
bias (α_0)	−0.2686	(0.118)	−2.2160***	(0.000)
Panel B: PEESE				
	PEESE ($n = 654$)		PEESE ($n = 220$)	
	Coef.	p -value	Coef.	p -value
precision (β_0)	−0.0155***	(0.000)	−0.0194***	(0.000)
standard error (α_0)	−2.8916	(0.174)	−10.4586**	(0.012)

Notes: The left column shows the results for the complete sample ($n = 654$). The right column shows the results for a restricted sample, omitting the effect sizes approximated from a 5% confidence interval ($n = 231$).

derived from 95% confidence levels. This may bias the results of the FAT, as the MRA procedure will then center the constant around zero, where both frontiers cancel out. This may lead to a non-rejection of $\alpha_0 = 0$, while there still could be a publication bias.

To circumvent this problem we also perform the FAT on a subsample of effect sizes that are derived from exact standard errors and t -values. The results of the FAT–PET and PEESE are shown in Table 4. The left column shows the results for the entire sample, and the right column shows the results for the subsample of effect sizes based on exact standard errors or t -values. In the right column, the null hypothesis of no publication bias is rejected at $p < 0.000$, giving us a clear indication for the presence of a publication bias. The Precision-Effect Test (PET) rejects $\beta_0 = 0$, implying that a genuine effect beyond the selection bias exists. The results of the PEESE approximations are in line with the FAT–PET results.

4.4 Multivariate Analysis

A multivariate MRA approach allows for an accurate description and interpretation of subgroup differences in the presence of across-study heterogeneity, by incorporating dummy variables for different subgroups. It furthermore allows us to control for macroeconomic background characteristics. In the next part of this paper we built upon the PEESE–MRA model to control for and explain the “genuine” variation the effect sizes beyond publication bias in a multivariate meta-regression model. In line with other meta-analyses, we include SE^2 instead of a linear term in SE .⁵

We extend model 5 with explanatory variables X_j , to accommodate for “genuine” variation among the effect sizes, more specifically the variation among the subgroups we are interested in (see Table A1 for descriptive statistics for our MRA variables):

$$ES_i = \beta_0 + \alpha_0 SE_i^2 + \sum_{j=1}^K \gamma_j X_{ji} + u_i. \quad (7)$$

We have estimated this meta-regression model by WLS using $1/SE_i^2$ as weights, while suppressing the constant term. This allows us to include dummy variables for all program types. The estimates then show

Table 5. Results Multivariate MRA.

	6 months		12 months		24 months		36 months	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
z_outcome_se_sq	0.187 (0.194)	0.000408 (0.176)	0.0931 (0.0730)	0.0136 (0.0535)	0.0343 (0.0350)	−0.00327 (0.0205)	0.00112 (0.0564)	0.0356 (0.0543)
subsidy	0.0993 (0.129)	−0.0769 (0.0859)	0.0335 (0.0422)	0.00491 (0.0457)	0.0359 (0.0301)	0.0196 (0.0227)	−0.0194 (0.0484)	0.0522** (0.0227)
services	0.223** (0.104)	0.0784 (0.0606)	0.105*** (0.0265)	0.0682** (0.0264)	0.0150 (0.0237)	0.00527 (0.0154)	−0.0531 (0.0514)	0.0232 (0.0215)
public	0.0995 (0.0989)	−0.0416 (0.0459)	0.0102 (0.0249)	−0.0181 (0.0196)	−0.0195 (0.0325)	−0.0253 (0.0219)	−0.0622 (0.0479)	0.00832 (0.0183)
training	0.162 (0.103)	0.0159 (0.0605)	0.0670*** (0.0245)	0.0264 (0.0217)	0.0249 (0.0236)	0.0145 (0.0149)	−0.0415 (0.0514)	0.0348 (0.0214)
male	−0.119 (0.0721)	−0.0837 (0.0539)	−0.0757** (0.0330)	−0.103*** (0.0341)	−0.0347 (0.0231)	−0.0417* (0.0234)	−0.00564 (0.0200)	−0.0358* (0.0204)
female	−0.134* (0.0743)	−0.0999** (0.0445)	−0.0796** (0.0382)	−0.101*** (0.0276)	−0.0139 (0.0272)	−0.0206 (0.0182)	0.0185 (0.0201)	−0.0150 (0.0144)
z_max_age	0.0258 (0.0342)	0.0101 (0.0168)	−0.00523 (0.0167)	0.00588 (0.00904)	−0.0159 (0.0155)	−0.00681 (0.00626)	−0.0432 (0.0320)	−0.00250 (0.00391)
z_year	0.0576 (0.0618)		0.0528* (0.0313)		0.0287 (0.0194)		0.0106 (0.0124)	
did	−0.00413 (0.0265)	0.0182 (0.0235)	−0.224 (0.202)	−0.0523 (0.0512)	−0.0393 (0.0495)	−0.0313 (0.0284)	0.0382 (0.0463)	−0.0491 (0.0364)
iv	0 (.)	0 (.)	−0.0678* (0.0398)	−0.0825** (0.0332)	0.00526 (0.0315)	0.00790 (0.0210)	0 (.)	0 (.)
rct	−0.238 (0.145)	−0.0736 (0.0569)	−0.0273 (0.0324)	−0.0257 (0.0347)	0.0477* (0.0243)	0.0426** (0.0168)	0.0932*** (0.0325)	0.0349 (0.0208)
country FE	Included		Included		Included		Included	
z_unemployment		Included		Included		Included		Included
N	154	154	205	205	173	173	122	122
Clusters	28	28	43	43	38	38	23	23
R ²	0.629	0.574	0.567	0.453	0.244	0.162	0.366	0.278

Notes: (1) Country fixed effects with normalized year of implementation; (2) No country fixed effects and no normalized year of implementation, but with the normalized employment rate on country level during the year of implementation.

Cluster—robust standard errors on study level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

their absolute levels of effectiveness, instead of relative differences, which simplifies the interpretation of the coefficients. We estimate a fixed-effects model, and cluster the standard errors on study level (Stanley and Doucouliagos, 2017).

Table 5 shows the results at 6, 12, 24, and 36 months after program start. Columns 1, 3, 5, and 7 show the results of model 7 without controlling for the country-specific unemployment rate in the year of program implementation, but including country dummies and a variable for the year of program implementation. Columns 2, 4, 6, and 8 show the results after controlling for the unemployment rate in the year of program implementation. Here, we exclude the country dummies and the year variable, since the country- and year-specific unemployment rates capture both the variation across countries as in the years

of implementation. This is our preferred specification, since this captures more than just the variation in countries and years.

Because we normalized all continuous MRA variables, we can interpret the coefficients of the different program types as the mean program impact on a mixed male and female target group of the average age. As we omitted the constant from model 7, we can interpret the program type coefficients as the general impacts.

At 6 months after program start, none of the program type coefficients are statistically significant. However, with 154 observations this may be attributable to low power. Subsidized labor and public employment have negative coefficients, whereas enhanced services and training programs both have positive impacts. The effects of enhanced services and training programs remain positive at 12, 24, and 36 after program start. The impact of subsidized labor turns positive at 12 months, and remains positive at 24 and 36 months as well. The impact of public employment stays negative at 12 and 24 months, but becomes positive at 36 months.

We do not find hard evidence for heterogeneous impacts for male and female participants. The coefficients for male and female participant groups with respect to mixed participant groups do not differ substantially. However, these coefficients should be interpreted with care and not as a causal relation. The gender coefficients may also be driven by selection effects, since some studies report impact estimates of a mixed program separate for men and women, whereas other studies report the effects for male or female only programs. We only include them for a more precise estimate of the program type coefficients. The same applies for the coefficient on the unemployment rate variable and the coefficients for the evaluation design dummies.

In Tables A3, A4, A5, and A6, we provide several robustness checks for the multivariate MRA. As mentioned above, the heterogeneity in the institutional environments is an important factor. In the main specification, we include country dummies and a year indicator to account for this. However, because of the over-representation of German studies in our data set (roughly 55%, see Table 2), we can also estimate the effects separately for only the German studies (Table A3), to look further into this. Next, we analyze the effects without the German studies (solely the non-German studies, Table A4) to see to what extent the overrepresentation of German studies influences our general results. It seems that when we only consider German studies, the negative lock-in effects of subsidized labor and public employment do not turn into positive long-term effects, while in the non-German subset they do. Training programs are also less effective in this German subsample. We do not have sufficient German-only estimates for enhanced services to say something about the differences for this program category.

Since this meta-analysis considers multiple labor market outcomes at once, we re-estimated the model for the 83% of studies that report the probability of employment as outcome variable (see Table 1 for the distribution of outcome variables). These results are shown in Table A5. This does not change the results, except that the lock-in effects of subsidized labor last slightly longer. The impact at 12 months is very slightly negative, which turns into slightly positive after 24 months. Another interesting question is whether the duration of the program is correlated with the effectiveness. Not all studies report this; roughly 40% of the observations do not contain information about the average duration. Adding the average program duration to the specification and narrowing down the sample to the 60% of observations that do report average program duration does not alter the results. The coefficient of program duration is close to zero, and only significantly negative at 36 months. The results are available in Table A6.

5. Discussion and Conclusions

This meta-analysis points to the following main results: (i) enhanced services programs are the most effective in the short run, at 12 months after program start, (ii) subsidized labor and public job creation

schemes have negative impacts in the short run, which eventually become positive in the longer run, where subsidized labor is the only program type that has a statistically significant positive impact at 36 months after program start, (iii) the impact of subsidized labor turns positive after 12 months, while for public job creation this is only the case after 36 months, (iv) the average duration is not correlated with the effectiveness of a program, but only slightly negatively correlated in the longer run. Similar to our findings, earlier reviews by for example Card *et al.* (2010, 2017) and Kluve (2010) also point out that enhanced services and sanctioning schemes are the most effective in improving the labor market outcomes of the participants. We also find positive coefficients for training programs, but these are not statistically significant. Both subsidized labor and public employment have insignificant results in both the short run, but only subsidized labor turns significantly positive in the longer run. In general, we find that the average effects of ALMPs are relatively small ($d < 0.10$).

As for the differences in short and long term effectiveness, the literature provides some explanation for this. Many training and employment schemes show negative effects in the short run, which can be attributed to the fact that during this period the participants are still in the program and not on the labor market, which might explain why they perform worse than the non-participating control group. So-called lock-in effects (Calmfors, 1994) can explain the differences in the short term impacts, because enhanced services and sanctioning schemes demand less from the participants than more intensive employment programs in terms of time that could also be spent on looking for a job (see Section 2). The results also show that subsidized labor is associated with shorter lock-in effects than public employment/job creation. This could be caused by the fact that subsidized labor usually includes an on-the-job training component, which is tailored to labor demand (Heckman *et al.*, 1999).

This also points at an important critique to classroom training programs. Since these programs are relatively standardized, they are for that reason less tailored to the requirements of firms (Heckman *et al.*, 1999). On-the-job training, as comprised by wage and employment subsidies which we cover in the subsidized labor category, is much more adapted to the needs of firms and therefore expected to be more effective than classroom training. However, our results do not suggest negative lock-in effects for training programs, but positive, non-significant effects from 6 until 36 months after program start. This absence of lock-in effects can most likely be explained by the fact that most studies covering on-the-job training cover both types of training, but do not always explicitly mention this, such that we include these studies in the general training category.

Our results are somewhat different from earlier overview studies on the effectiveness of ALMPs. Card *et al.* (2010, 2017), for example, find negative short term effects of training programs, at 12 months after program start. However, given the large set of differences between previous meta-analyses (for example, Card *et al.* (2010, 2017) and Kluve (2010)) and our meta-analysis, it is difficult to determine exactly where this difference comes from.

The improvements of our meta-analysis over these earlier meta-analyses are both methodological and theoretical. Our methodological improvements consist first of all a systematically assembled data set of studies that satisfy the inclusion criteria stated in Section 3. In order to ensure the quality of the effect estimates and to rule out any selection effects, we only include published experimental and quasi-experimental studies. Furthermore, contrary to Card *et al.* (2010), we consider standardized effect sizes, which allows us to give an indication of the size of the effects of ALMP. Our theoretical contribution consists of the inclusion of a theoretical framework, to which we explicitly relate our results.

Our findings provide a general guideline for the design of future policy. In the design of future ALMP, one should keep in mind that public employment and subsidized labor programs are characterized by negative short-term effects. Subsidized labor, however, should be preferred over public employment, since the effects of subsidized labor turn positive much earlier than public employment schemes. Also, the effect of subsidized labor is significantly positive at 36 months after program start. Enhanced services, including job-search assistance, have positive effects in the short run. These effects remain positive but

become statistically insignificant in the longer run. Therefore, job-search assistance is more likely to line up with policy goals than subsidized labor and public employment in the short run. Subsidized labor turns out to have the greatest longer run effect. Finally, the fact that we find such small average effects highlight the need of thorough cost-benefit analyses, and the continuation of reliable randomized evaluations.

Acknowledgements

We would like to thank conference participants at the 2017 Research Workshop on Economics, Statistics and Econometrics of Education in Lisbon, the 2017 LEER Workshop on Education Economics in Leuven, the 22nd Annual Meetings of the Society of Labor Economists in Raleigh, and the 2017 MAER-Net Colloquium at Zeppelin University in Friedrichshafen, for their relevant comments and suggestions. We also wish to thank the editor Les Oxley, the associate editor Chris Doucouliagos, and two anonymous referees for their helpful comments and suggestions. Any remaining errors are our own.

Notes

1. Arguments for this relation include that unions are less motivated to advocate for wage increases when employment and labor demand are low, and that employers have to pay higher wages to attract sufficient labor when employment and labor demand are high.
2. The data set of 654 effect estimates can be found through ResearchGate with <https://doi.org/10.13140/RG.2.2.13919.97449>.
3. Unemployment data derived from the IMF's World Economic Outlook (IMF, 2017).
4. The Q-statistic is computed using the following formula, where ES_i stands for effect size i and SE_i for its standard error:

$$Q = \sum_{i=1}^k \frac{(ES_i - \overline{ES})^2}{SE^2}.$$

5. Many studies apply FAT–PET and PEESE in a multivariate framework (for example, Rose and Stanley, 2005; Ugur, 2014; Awaworyi Churchill and Yew, 2017). Doucouliagos *et al.* (2014) also applies SE^2 instead of a linear term in SE , like we do in our model. We have also run the analysis with a linear term in SE , and this does not make a difference.

References

- Alegre, M.A., Casado, D., Sanz, J. and Todeschini, F.A. (2015) The impact of training-intensive labour market policies on labour and educational prospects of NEETs: evidence from Catalonia (Spain). *Educational Research* 57: 151–167.
- Awaworyi Churchill, S. and Yew, S.L. (2017) Are government transfers harmful to economic growth? A meta-analysis. *Economic Modelling* 64: 270–287.
- Bonoli, G. (2010) The political economy of active labor-market policy. *Politics & Society* 38: 435–457.
- Caliendo, M. and Künn, S. (2015) Getting back into the labor market: the effects of start-up subsidies for unemployed females. *Journal of Population Economics* 28: 1005–1043.
- Calmfors, L. (1994) Active labour market policy and unemployment - a framework for the analysis of crucial design features. *OECD Economic Studies* 22: 7–47.
- Calmfors, L. and Lang, H. (1995) Macroeconomic effects of active labour market programmes in a union wage-setting model. *The Economic Journal* 105: 601–619.

- Card, D., Kluve, J. and Weber, A. (2010) Active labour market policy evaluations: a meta-analysis*. *The Economic Journal* 120: F452–F477.
- Card, D., Kluve, J. and Weber, A. (2017) What works? A meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association* forthcoming.
- Crépon, B. and van den Berg, G.J. (2016) Active labor market policies. *Annual Review of Economics* 8: 521–546.
- Dorsett, R., Smeaton, D. and Speckesser, S. (2013) The effect of making a voluntary labour market programme compulsory: evidence from a UK experiment. *Fiscal Studies* 34: 467–489.
- Doucouliagos, H., Stanley, T.D. and Viscusi, W.K. (2014) Publication selection and the income elasticity of the value of a statistical life. *Journal of Health Economics* 33: 67–75.
- Dyke, A., Heinrich, C., Mueser, P., Troske, K. and Jeon, K. (2006) The effects of welfare to work program activities on labor market outcomes. *Journal of Labor Economics* 24: 567–607.
- Egger, M., Davey Smith, G., Schneider, M. and Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 315: 629–634.
- Filges, T., Smedslund, G. and Jørgensen, A.M.K. (2016) Active labour market programme participation for unemployment insurance recipients: a systematic review. *Research on Social Work Practice* forthcoming.
- Frölich, M. (2004) Programme evaluation with multiple treatments. *Journal of Economic Surveys* 18: 181–224.
- Heckman, J.J., LaLonde, R.J. and Smith, J.A. (1999) The economics and econometrics of active labor market programs. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. III, (pp. 1865–2097) chapter 31. Amsterdam: North-Holland.
- Hedges, L.V. (1982) Estimation of effect size from a series of independent experiments. *Psychological Bulletin* 92: 490–499.
- Imbens, G.W. and Wooldridge, J.M. (2009) Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47: 5–86.
- IMF (2017) *World Economic Outlook April 2017* is available at: <http://www.imf.org/en/Publications/WEO/Issues/2017/04/04/world-economic-outlook-april-2017> (Last accessed 22 August 2017).
- Johnson, G. and Layard, R. (1986) The natural rate of unemployment: explanation and policy. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Volume II, (pp. 921–999) chapter 16. Amsterdam: North-Holland.
- Kluve, J. (2010) The effectiveness of European active labor market programs. *Labour Economics* 17: 904–918.
- Layard, R. and Nickell, S. (1986) Unemployment in Britain. *Economica* 53: S121–S169.
- Layard, R., Nickell, S. and Jackman, R. (2005) *Unemployment: Macroeconomic Performance and the Labour Market*. Oxford: Oxford University Press.
- Lipsey, M.W. and Wilson, D.B. (2001) Practical meta-analysis. In C. Laughton (ed.), *Applied Social Research Methods Series*, Vol. 49. Thousand Oaks: SAGE.
- Martin, J.P. (2015) Activation and active labour market policies in OECD countries: stylised facts and evidence on their effectiveness. *IZA Journal of Labor Policy* 4: 4.
- McKenzie, D. (2017) How effective are active labor market policies in developing countries? A critical review of recent evidence. *The World Bank Research Observer* 32: 127–154.
- OECD (2017) Social Expenditure Database is available at: <http://doi.org/10.1787/els-socx-data-en> (Last accessed 1 March 2017).
- Robinson, P. (2000) Active labour market policies: a case of evidence-based policy-making? *Oxford Review of Economic Policy* 16: 13–26.
- Rose, A.K. and Stanley, T.D. (2005) A meta-analysis of the effect of common currencies on international trade. *Journal of Economic Surveys* 19: 347–365.
- Stanley, T.D. (2005) Beyond publication bias. *Journal of Economic Surveys* 19: 309–345.
- Stanley, T.D. (2008) Meta-regression methods for detecting and estimating empirical effects in the presence of publication selection. *Oxford Bulletin of Economics and Statistics* 70: 103–127.
- Stanley, T.D. and Doucouliagos, H. (2010) Picture this: a simple graph that reveals much ado about research. *Journal of Economic Surveys* 24: 170–191.

- Stanley, T.D. and Doucouliagos, H. (2014) Meta-regression approximations to reduce publication selection bias. *Research Synthesis Methods* 5: 60–78.
- Stanley, T.D. and Doucouliagos, H. (2015) Neither fixed nor random: weighted least squares meta-analysis. *Statistics in Medicine* 34: 2116–2127.
- Stanley, T.D. and Doucouliagos, H. (2017) Neither fixed nor random: weighted least squares meta-regression. *Research Synthesis Methods* 8: 19–42.
- Stanley, T.D., Doucouliagos, H., Giles, M., Heckemeyer, J., Johnston, R., Laroche, P., Nelson, J., Paldam, M., Poot, J., Pugh, G., Rosenberger, R. and Rost, K. (2013) Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys* 27: 390–394.
- Ugur, M. (2014) Corruption's direct effects on per-capita income growth: a meta-analysis. *Journal of Economic Surveys* 28: 472–490.

List of studies in sample

- Aber, J.L., Brooksgunn, J. and Maynard, R.A. (1995) Effects of welfare-reform on teenage parents and their children. *Future of Children* 5: 53–71.
- Achdut, N. (2017) Financial incentives to work: a quasi-experimental analysis of in-work cash benefits for single mothers. *International Journal of Social Welfare* 26: 21–36.
- Alegre, M.A., Casado, D., Sanz, J. and Todeschini, F.A. (2015) The impact of training-intensive labour market policies on labour and educational prospects of NEETs: evidence from Catalonia (Spain). *Educational Research* 57: 151–167.
- Alfonso Arellano, F. (2010) Do training programmes get the unemployed back to work? A look at the Spanish experience. *Revista de Economía Aplicada* 18: 39–65.
- Autor, D.H. and Houseman, S.N. (2010) Do temporary-help jobs improve labor market outcomes for low-skilled workers? Evidence from “Work First”. *American Economic Journal: Applied Economics* 2: 96–128.
- Autor, D.H., Houseman, S.N. and Pekkala Kerr, S. (2017) The effect of work first job placements on the distribution of earnings: an instrumental variable quantile regression approach. *Journal of Labor Economics* 35: 149–190.
- Baumgartner, H.J. and Caliendo, M. (2008) Turning unemployment into self-employment: effectiveness of two start-up programmes. *Oxford Bulletin of Economics and Statistics* 70: 347–373.
- Bergemann, A., Fitzenberger, B. and Speckesser, S. (2009) Evaluating the dynamic employment effects of training programs in East Germany using conditional difference-in-differences. *Journal of Applied Econometrics* 24: 797–823.
- Blien, U. and Caliendo, M. (2009) Startup subsidies in East Germany: finally, a policy that works? *International Journal of Manpower* 30: 625–647.
- Bonin, H. and Rinne, U. (2014) Beautiful Serbia – objective and subjective outcomes of active labour market policy in a transition economy. *Economics of Transition* 22: 43–67.
- Brock, T. and Harknett, K. (1998) A comparison of two welfare-to-work case management models. *Social Service Review* 72: 493–520.
- Caliendo, M., Hujer, R. and Thomsen, S.L. (2006) Sectoral heterogeneity in the employment effects of job creation schemes in Germany. *Jahrbücher für Nationalökonomie und Statistik* 226: 139–179.
- Caliendo, M. and Künn, S. (2011) Start-up subsidies for the unemployed: long-term evidence and effect heterogeneity. *Journal of Public Economics* 95: 311–331.
- Caliendo, M. and Künn, S. (2015) Getting back into the labor market: the effects of start-up subsidies for unemployed females. *Journal of Population Economics* 28: 1005–1043.
- Caliendo, M., Künn, S. and Mahlstedt, R. (2017) The return to labor market mobility: an evaluation of relocation assistance for the unemployed. *Journal of Public Economics* 148: 136–151.
- Centeno, L., Centeno, M. and Novo, A.A. (2009) Evaluating job-search programs for old and young individuals: heterogeneous impact on unemployment duration. *Labour Economics* 16: 12–25.
- Dahl, E. and Lorentzen, T. (2005) What works for whom? An analysis of active labour market programmes in Norway. *International Journal of Social Welfare* 14: 86–98.

- Dengler, K. (2015) Effectiveness of sequences of one-euro-jobs for welfare recipients in Germany. *Applied Economics* 47: 6170–6190.
- Doerr, A., Fitzenberger, B., Kruppe, T., Paul, M. and Strittmatter, A. (2016) Employment and earnings effects of awarding training vouchers in Germany. *ILR Review* 70: 767–812.
- Dorsett, R. (2006) The New Deal for Young People: effect on the labour market status of young men. *Labour Economics* 13: 405–422.
- Dorsett, R., Smeaton, D. and Speckesser, S. (2013) The effect of making a voluntary labour market programme compulsory: evidence from a UK experiment. *Fiscal Studies* 34: 467–489.
- Dyke, A., Heinrich, C., Mueser, P., Troske, K. and Jeon, K. (2006) The effects of welfare-to-work program activities on labor market outcomes. *Journal of Labor Economics* 24: 567–607.
- Eichler, M. and Lechner, M. (2002) An evaluation of public employment programmes in the East German state of Sachsen-Anhalt. *Labour Economics* 9: 143–186.
- Fitzenberger, B., Orlanski, O., Osikominu, A. and Paul, M. (2012) Déjà vu? Short-term training in Germany 1980–1992 and 2000–2003. *Empirical Economics* 44: 289–328.
- Fitzenberger, B. and Prey, H. (2000) Evaluating public sector sponsored training in East Germany. *Oxford Economic Papers* 52: 497–520.
- Frölich, M. and Lechner, M. (2010) Exploiting regional treatment intensity for the evaluation of labor market policies. *Journal of the American Statistical Association* 105: 1014–1029.
- Gerfin, M. and Lechner, M. (2002) A microeconomic evaluation of the active labour market policy in Switzerland. *The Economic Journal* 112: 854–893.
- Gerfin, M., Lechner, M. and Steiger, H. (2005) Does subsidised temporary employment get the unemployed back to work? An econometric analysis of two different schemes. *Labour Economics* 12: 807–835.
- Giorgi, G. (2005) The new deal for young people five years on. *Fiscal Studies* 26: 371–383.
- Graversen, B.K. and van Ours, J.C. (2008) How to help unemployed find jobs quickly: experimental evidence from a mandatory activation program. *Journal of Public Economics* 92: 2020–2035.
- Hagglund, P. (2014) Experimental evidence from active placement efforts among unemployed in Sweden. *Evaluation Review* 38: 191–216.
- Hamersma, S. (2008) The effects of an employer subsidy on employment outcomes: a study of the work opportunity and welfare-to-work tax credits. *Journal of Policy Analysis and Management* 27: 498–520.
- Hamersma, S. and Heinrich, C. (2008) Temporary help service firms' use of employer tax credits: implications for disadvantaged workers' labor market outcomes. *Southern Economic Journal* 74: 1123–1148.
- Huber, M., Lechner, M., Wunsch, C. and Walter, T. (2011) Do German welfare-to-work programmes reduce welfare dependency and increase employment? *German Economic Review* 12: 182–204.
- Hujer, R. and Thomsen, S.L. (2010) How do the employment effects of job creation schemes differ with respect to the foregoing unemployment duration? *Labour Economics* 17: 38–51.
- Jaenichen, U. and Stephan, G. (2011) The effectiveness of targeted wage subsidies for hard-to-place workers. *Applied Economics* 43: 1209–1225.
- Jagannathan, R. and Camasso, M. (2006) Beyond intention to treat analysis in welfare-to-work studies. *Journal of Social Service Research* 31: 43–60.
- Jespersen, S.T., Munch, J.R. and Skipper, L. (2008) Costs and benefits of Danish active labour market programmes. *Labour Economics* 15: 859–884.
- Johansson, P. (2008) The importance of employer contacts: evidence based on selection on observables and internal replication. *Labour Economics* 15: 350–369.
- Kluve, J., Lehmann, H. and Schmidt, C.M. (1999) Active labor market policies in Poland: Human capital enhancement, stigmatization, or benefit churning? *Journal of Comparative Economics* 27: 61–89.
- Kluve, J., Lehmann, H. and Schmidt, C.M. (2008) Disentangling treatment effects of active labor market policies: the role of labor force status sequences. *Labour Economics* 15: 1270–1295.
- Landeghem, B.V., Corvers, F. and Grip, A.D. (2017) Is there a rationale to contact the unemployed right from the start? Evidence from a natural field experiment. *Labour Economics* 45: 158–168.

- Lange, T., Perry, G. and Maloney, T. (2007) Evaluating active labour market programmes in New Zealand. *International Journal of Manpower* 28: 7–29.
- Larsson, L. (2003) Evaluation of Swedish youth labor market programs. *The Journal of Human Resources* 38: 891.
- Lechner, M., Miquel, R. and Wunsch, C. (2007) The curse and blessing of training the unemployed in a changing economy: the case of East Germany after unification. *German Economic Review* 8: 468–509.
- Lechner, M. and Wiehler, S. (2009) Kids or courses? Gender differences in the effects of active labor market policies. *Journal of Population Economics* 24: 783–812.
- Lechner, M. and Wunsch, C. (2009) Active labour market policy in East Germany. *Economics of Transition* 17: 661–702.
- Lindley, J., McIntosh, S., Roberts, J., Czoski Murray, C. and Edlin, R. (2015) Policy evaluation via a statistical control: A non-parametric evaluation of the ‘Want2Work’ active labour market policy. *Economic Modelling* 51: 635–645.
- López Moureló, E. and Escudero, V. (2017) Effectiveness of active labor market tools in conditional cash transfers programs: evidence for Argentina. *World Development* 94: 422–447.
- Lorentzen, T. and Dahl, E. (2005) Active labour market programmes in Norway: are they helpful for social assistance recipients? *Journal of European Social Policy* 15: 27–45.
- Maibom, J., Rosholm, M. and Svarer, M. (2017) Experimental evidence on the effects of early meetings and activation. *The Scandinavian Journal of Economics* 119: 541–570.
- Malmberg-Heimonen, I. and Tge, A.G. (2016) Effects of individualised follow-up on activation programme participants’ self-sufficiency: a cluster-randomised study. *International Journal of Social Welfare* 25: 27–35.
- Markussen, S. and Red, K. (2016) Leaving poverty behind? The effects of generous income support paired with activation. *American Economic Journal: Economic Policy* 8: 180–211.
- Neubäumer, R. (2012) Bringing the unemployed back to work in Germany: training programs or wage subsidies? *International Journal of Manpower* 33: 159–177.
- Nivorozhkin, A. (2005) An evaluation of government-sponsored vocational training programmes for the unemployed in urban Russia. *Cambridge Journal of Economics* 29: 1053–1072.
- Peck, L.R. (2007) What are the effects of welfare sanction policies? *American Journal of Evaluation* 28: 256–274.
- Raam, O. and Torp, H. (2002) Labour market training in Norway – effect on earnings. *Labour Economics* 9: 207–247.
- Rinne, U., Uhlendorff, A. and Zhao, Z. (2012) Vouchers and caseworkers in training programs for the unemployed. *Empirical Economics* 45: 1089–1127.
- Rodríguez-Planas, N. and Jacob, B. (2009) Evaluating active labor market programs in Romania. *Empirical Economics* 38: 65–84.
- Sianesi, B. (2008) Differential effects of active labour market programs for the unemployed. *Labour Economics* 15: 370–399.
- Stephan, G. (2008) The effects of active labor market programs in Germany: an investigation using different definitions. *Jahrbücher für Nationalökonomie und Statistik* 228: 586–611.
- Stephan, G. and Pahnke, A. (2011) The relative effectiveness of selected active labor market programs: an empirical investigation for Germany. *The Manchester School* 79: 1262–1293.
- Sorensen, K.L. (2016) Heterogeneous impacts on earnings from an early effort in labor market programs. *Labour Economics* 41: 266–279.
- Štefánik, M. (2014) Estimating treatment effects of a training programme in Slovakia using propensity score matching. *Ekonomický Časopis – Journal of Economics* 62: 631–645.
- Winterhager, H., Heinze, A. and Spermann, A. (2006) Deregulating job placement in Europe: a microeconomic evaluation of an innovative voucher scheme in Germany. *Labour Economics* 13: 505–517.

Appendix

Table A1. Descriptive Statistics of Variables in Multivariate MRA (Complete Sample).

Variable	Definition	N	Mean	SD
outcome	Cohen's <i>d</i>	654	-0.0296	0.1558
z_outcome_se_sq	SE^2 (standardized)	654	0.0000	1.0000
subsidy	=1 if category is subsidized labor	654	0.3777	0.4852
services	=1 if category is enhanced services	654	0.1254	0.3314
public	=1 if category is public employment	654	0.1116	0.3151
training	=1 if category is training or retraining	654	0.3853	0.4870
male	=1 if gender="male"	654	0.2905	0.4543
female	=1 if gender="female"	654	0.2737	0.4462
z_max_age	maximum age of participants (standardized)	654	0.0000	1.0000
z_year	year of implementation (standardized)	654	0.0000	1.0000
did	=1 in case of a DiD design	654	0.0336	0.1804
iv	=1 in case of an IV design	654	0.0122	0.1100
rct	=1 in case of an RCT	654	0.0734	0.2610
z_unemployment	unemployment rate (standardized)	654	0.0000	1.0000
z_duration	average program duration (standardized)	419	0.0000	1.0000

Notes: All continuous variables have been standardized to mean zero and standard deviation one.

Table A2. Mean and Weighted Program Effects by Programme Type.

	Sample Mean (1)	SD (2)	Fixed-effects (3)	Random-effects (4)	FE-WLS (5)
<i>(i) 6 months after program start</i>					
Enhanced services (22)	-0.029	0.080	0.002	-0.007	0.002
Public employment (15)	-0.061	0.062	-0.059***	-0.060***	-0.059***
Subsidized labor (58)	-0.145	0.243	-0.137***	-0.146***	-0.137***
Training/retraining (59)	-0.019	0.082	-0.057***	-0.027***	-0.057***
Overall effect (154)	-0.072	0.017	-0.038***	-0.073***	-0.038***
<i>(ii) 36 months after program start</i>					
Enhanced services (15)	0.020	0.047	-0.004**	0.006	-0.004
Public employment (17)	-0.063	0.177	-0.014	-0.013	-0.014
Subsidized labor (44)	0.032	0.081	0.026***	0.027***	0.026***
Training/retraining (46)	0.039	0.072	0.008***	0.017***	0.008***
Overall effect (122)	0.020	0.099	0.004***	0.016***	0.004*

Notes: ***, **, * denote 1%, 5%, and 10% significance levels, respectively. Numbers of observations in parentheses.

Table A3. Results Multivariate MRA – German Studies Only.

	6 months		12 months		24 months		36 months	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
z_outcome_se_sq	0.273 (0.244)	0.298 (0.228)	0.159 (0.127)	0.163 (0.134)	0.121 (0.0840)	0.200** (0.0740)	−0.0270 (0.0514)	−0.00570 (0.0683)
subsidy	0.0291 (0.0396)	−0.128 (0.130)	−0.130** (0.0559)	−0.0997 (0.107)	0.0612** (0.0276)	−0.0269 (0.0608)	0.0445 (0.0286)	−0.0277 (0.0354)
services	0.188*** (0.0500)	0.0764 (0.121)	0 (.)	0.0651 (0.0924)	−0.0224 (0.0504)	−0.0772 (0.0541)	0 (.)	0 (.)
public	0 (.)	−0.145 (0.117)	−0.189*** (0.0473)	−0.150 (0.0961)	0 (.)	−0.0837 (0.0639)	0 (.)	−0.0732* (0.0358)
training	0.0773* (0.0418)	−0.0497 (0.127)	−0.109*** (0.0144)	−0.0566 (0.0967)	0.0556** (0.0215)	−0.0286 (0.0686)	0.0491*** (0.00703)	−0.0226 (0.0333)
male	−0.108 (0.0698)	−0.0353 (0.0436)	−0.0586 (0.0353)	−0.00287 (0.0315)	−0.0433 (0.0303)	0.0394 (0.0340)	0.00849 (0.0318)	0.0600** (0.0238)
female	−0.153* (0.0744)	−0.0644 (0.0668)	−0.0941** (0.0438)	−0.0253 (0.0302)	−0.0526 (0.0362)	0.0382 (0.0320)	−0.000617 (0.0243)	0.0558 (0.0308)
z_max_age	0.0205 (0.0621)	−0.0434 (0.0542)	−0.0249 (0.0552)	−0.0778 (0.0550)	0.000878 (0.0484)	−0.0546 (0.0375)	−0.0212 (0.0324)	−0.0537* (0.0252)
z_year	0.0606 (0.0615)		0.0604 (0.0515)		0.0193 (0.0228)		0.00945 (0.0162)	
did	0 (.)	0 (.)	−0.307 (0.297)	−0.225 (0.330)	−0.717** (0.264)	−0.880*** (0.270)	0.0581 (0.0416)	0.0595 (0.0462)
iv	0 (.)	0 (.)	0 (.)	0 (.)	0.115*** (0.0273)	0.000816 (0.0436)	0 (.)	0 (.)
rct	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
Country FE	Included		Included		Included		Included	
z_unemployment		Included		Included		Included		Included
N	104	104	112	112	93	93	63	63
Clusters	16	16	18	18	14	14	9	9
R ²	0.528	0.668	0.424	0.595	0.189	0.459	0.253	0.374

Notes: (1) Country fixed effects with normalized year of implementation; (2) No country fixed effects and no normalized year of implementation, but with the normalized employment rate on country level during the year of implementation.

Cluster-robust standard errors on study level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A4. Results Multivariate MRA – Non-German Studies Only.

	6 months		12 months		24 months		36 months	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
z_outcome_se_sq	−0.0443 (0.111)	−0.104 (0.105)	0.0708* (0.0381)	0.0371* (0.0215)	0.0171 (0.0227)	0.00875 (0.0167)	0.0609** (0.0207)	0.138*** (0.0427)
subsidy	0.247* (0.127)	−0.0702* (0.0385)	0.0271 (0.0249)	−0.0279* (0.0157)	0.0222 (0.0293)	0.0306** (0.0122)	−0.202*** (0.0636)	0.0909*** (0.0159)
services	0.350** (0.125)	0.00276 (0.0468)	0.0567** (0.0271)	0.00552 (0.0150)	−0.0105 (0.0241)	−0.0147 (0.0167)	−0.256*** (0.0659)	0.0393** (0.0174)
public	0.284* (0.119)	−0.0369 (0.0313)	0.0239 (0.0270)	−0.00525 (0.0190)	−0.0194 (0.0344)	−0.00664 (0.0119)	−0.254*** (0.0669)	0.0395** (0.0132)
training	0.295** (0.123)	−0.0520 (0.0456)	0.0281 (0.0267)	−0.0238 (0.0148)	−0.00143 (0.0241)	−0.00548 (0.0167)	−0.245*** (0.0658)	0.0505** (0.0174)
male	−0.397*** (0.0939)	−0.0939*** (0.0292)	−0.0555** (0.0209)	−0.0804*** (0.0142)	−0.0375 (0.0235)	−0.0532*** (0.0160)	0.000101 (0.0103)	−0.0433*** (0.0109)
female	−0.388*** (0.0931)	−0.0950*** (0.0272)	−0.0400** (0.0169)	−0.0593*** (0.0123)	0.00301 (0.0240)	−0.0130 (0.0135)	0.0384*** (0.00950)	−0.00574 (0.0110)
z_max_age	0.00842*** (0.00267)	0.000695 (0.00478)	−0.00853 (0.00701)	0.00201 (0.00383)	−0.0244* (0.0121)	−0.0118*** (0.00410)	−0.239*** (0.0507)	−0.00406** (0.00158)
z_year	0.219*** (0.0512)		0.0453*** (0.0145)		0.0416* (0.0210)		0.110*** (0.0260)	
did	0.0749*** (0.0105)	0.0218** (0.00879)	−0.0846*** (0.0234)	−0.0113 (0.0197)	−0.00448 (0.0280)	−0.0162 (0.0138)	0.0204 (0.0267)	−0.0564*** (0.0133)
iv	0 (.)	0 (.)	−0.0257* (0.0135)	−0.0445*** (0.0129)	0.0145 (0.0302)	0.00682 (0.0165)	0 (.)	0 (.)
rct	−0.500*** (0.139)	−0.0427 (0.0440)	0.00415 (0.0236)	0.0200 (0.0255)	0.0516** (0.0212)	0.0491*** (0.0163)	0.194*** (0.0307)	0.0227 (0.0239)
Country FE	Included		Included		Included		Included	
z_unemployment		Included		Included		Included		Included
N	50	50	93	93	80	80	59	59
Clusters	12	12	25	25	24	24	14	14
R ²	0.909	0.893	0.839	0.805	0.458	0.374	0.580	0.502

Notes: (1) Country fixed effects with normalized year of implementation; (2) No country fixed effects and no normalized year of implementation, but with the normalized employment rate on country level during the year of implementation.

Cluster-robust standard errors on study level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A5. Results Multivariate MRA – Only Probability of Employment as Outcome Variable.

	6 months		12 months		24 months		36 months	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
z_outcome_se_sq	0.208 (0.203)	0.0311 (0.177)	0.139* (0.0709)	0.0231 (0.0527)	0.0888* (0.0450)	0.0135 (0.0277)	0.119*** (0.0349)	0.0866* (0.0404)
subsidy	0.111 (0.161)	−0.0699 (0.0810)	0.0399 (0.0863)	−0.0149 (0.0392)	0.0723* (0.0387)	0.00164 (0.0292)	0.0287 (0.0455)	0.0463 (0.0272)
services	0.241* (0.125)	0.102 (0.0682)	0.139*** (0.0486)	0.0916*** (0.0321)	0.0410 (0.0313)	0.00884 (0.0196)	0.0180 (0.0475)	0.0556** (0.0249)
public	0.105 (0.135)	−0.0348 (0.0447)	0.0186 (0.0527)	−0.0242 (0.0205)	0.0137 (0.0376)	−0.0269 (0.0253)	−0.00719 (0.0443)	0.0184 (0.0166)
training	0.171 (0.129)	0.0212 (0.0603)	0.0693 (0.0462)	0.0184 (0.0215)	0.0585* (0.0301)	0.0118 (0.0170)	0.0227 (0.0478)	0.0518** (0.0203)
male	−0.121 (0.0712)	−0.0834 (0.0609)	−0.0756* (0.0406)	−0.0704* (0.0381)	−0.0449 (0.0346)	−0.0211 (0.0319)	−0.0130 (0.0301)	−0.0192 (0.0322)
female	−0.136* (0.0737)	−0.0939 (0.0624)	−0.0799* (0.0464)	−0.0728 (0.0490)	−0.0259 (0.0399)	−0.000704 (0.0360)	0.0113 (0.0308)	0.00628 (0.0314)
z_max_age	0.0278 (0.0604)	0.0113 (0.0273)	−0.00642 (0.0362)	0.00428 (0.0199)	−0.0171 (0.0161)	−0.0107 (0.0125)	−0.0248 (0.0344)	−0.000112 (0.0111)
z_year	0.0563 (0.0629)		0.0564 (0.0467)		0.0258 (0.0192)		0.0203 (0.0175)	
did	0 (.)	0 (.)	0 (.)	0 (.)	0.0280 (0.0653)	0.0416 (0.0409)	0.0587* (0.0328)	0.0492** (0.0193)
iv	0 (.)	0 (.)	−0.0778 (0.0547)	−0.0711** (0.0274)	0.0481 (0.0579)	0.0227 (0.0280)	0 (.)	0 (.)
rct	−0.167* (0.0920)	−0.0335 (0.0406)	−0.0294 (0.0558)	−0.0301 (0.0366)	0.0854* (0.0446)	0.0377* (0.0189)	−0.00635 (0.0797)	0.00246 (0.0298)
Country FE	Included		Included		Included		Included	
z_unemployment		Included		Included		Included		Included
N	138	138	171	171	142	142	93	93
Clusters	24	24	35	35	29	29	15	15
R ²	0.507	0.435	0.385	0.303	0.206	0.099	0.329	0.257

Notes: (1) Country fixed effects with normalized year of implementation; (2) No country fixed effects and no normalized year of implementation, but with the normalized employment rate on country level during the year of implementation.

Cluster-robust standard errors on study level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A6. Results Multivariate MRA – Only Observations that Include the Average Program Duration.

	6 months		12 months		24 months		36 months	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
z_outcome_se_sq	0.183 (0.226)	0.164 (0.178)	0.0957 (0.0774)	0.0233 (0.0455)	0.0326 (0.0384)	−0.0196 (0.0267)	0.0825*** (0.0222)	0.0875* (0.0482)
subsidy	0.100 (0.202)	−0.268* (0.129)	0.0666 (0.209)	−0.0574 (0.0665)	0.142 (0.0964)	−0.0435 (0.0453)	0.110 (0.160)	0.0463 (0.0292)
services	0.274 (0.212)	−0.000642 (0.0731)	0.0932 (0.153)	−0.0775 (0.0455)	0.0570 (0.0591)	−0.104*** (0.0272)	0.174 (0.162)	0.170*** (0.0142)
public	0.235 (0.184)	−0.122 (0.0812)	0.111 (0.177)	−0.0631* (0.0354)	0.0942 (0.0694)	−0.0950*** (0.0302)	0.0515 (0.158)	−0.00823 (0.0201)
training	0.288 (0.189)	−0.0520 (0.0834)	0.173 (0.170)	0.0121 (0.0384)	0.154** (0.0668)	−0.0190 (0.0271)	0.0916 (0.163)	0.0363 (0.0218)
male	−0.0264 (0.0817)	0.0480 (0.0803)	−0.00208 (0.0498)	−0.0577 (0.0532)	−0.0270 (0.0435)	−0.00189 (0.0397)	−0.0347 (0.0405)	−0.00174 (0.0342)
female	−0.0615 (0.0829)	0.0184 (0.0866)	−0.0304 (0.0523)	−0.0853 (0.0646)	−0.0281 (0.0505)	−0.00193 (0.0440)	−0.0396 (0.0304)	−0.0103 (0.0345)
z_max_age	0.0216 (0.0334)	0.0168 (0.0173)	−0.0198 (0.0468)	−0.000533 (0.0257)	−0.000192 (0.0259)	0.00119 (0.0119)	0.00363 (0.0343)	−0.00460 (0.00815)
z_year	0.0156 (0.0408)		0.0332 (0.0354)		0.0183 (0.0219)		0.00680 (0.0170)	
did	−0.146 (0.192)	0.178* (0.0996)	−0.305** (0.145)	−0.155 (0.122)	−0.457*** (0.134)	−0.258** (0.115)	−1.022*** (0.162)	−1.064*** (0.333)
iv	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)	0 (.)
rct	−0.171 (0.196)	0.101 (0.158)	0 (.)	0.178 (0.107)	0 (.)	0.221*** (0.0507)	0 (.)	0 (.)
z_duration	0.00563 (0.0320)	0.00791 (0.0270)	−0.00693 (0.0291)	−0.0172 (0.0188)	−0.0166 (0.0181)	−0.0150 (0.00879)	−0.0490* (0.0231)	−0.0400* (0.0195)
country FE	Included		Included		Included		Included	
z_unemployment		Included		Included		Included		Included
N	106	106	127	127	115	115	71	71
Clusters	18	18	24	24	22	22	12	12
R ²	0.618	0.659	0.427	0.374	0.194	0.245	0.429	0.400

Notes: (1) Country fixed effects with normalized year of implementation; (2) No country fixed effects and no normalized year of implementation, but with the normalized employment rate on country level during the year of implementation.

Cluster-robust standard errors on study level in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.