Assessing Active Labor Market Policies in Transition Economies[*]

Hartmut Lehmann
DARRT, Department of Economics, University of Bologna and IZA, Bonn

and

Jochen Kluve
RWI Essen and IZA, Bonn

October 2008

Abstract

With the beginning of economic reform in the formerly centrally planned economies of Central and Eastern Europe (CEE), open unemployment rapidly reached comparable levels to those in Western economies. Governments in the region reacted to this rise by adopting active labor market policies (ALMP) as an important tool in the fight against unemployment. Before reviewing the evidence on the efficacy of such policies we look at the scope and the rationale of ALMP measures in a transitional context. Since government budgets are very tight in these countries it is important to evaluate ALMP in a rigorous fashion. The paper analyzes macroeconometric and microeconometric methods of program evaluation, as they were applied in transition economies. Both these approaches have a raison d'être and should be understood as complementing. Providing a selective review of the literature, some of the strengths and the pitfalls of the two approaches are highlighted. We also point to the lessons one can draw from the surveyed studies for a better understanding of how active measures affect labor market outcomes in this set of countries.

[*] This paper is an extended version of the invited lecture that Hartmut Lehmann gave at the XXII Conference of the Italian Association of Labor Economists (AIEL) in Naples in September 2007. He is grateful to the organizers for their hospitality and to Floro Ernesto Caroleo, Sergio Destefanis and Francesco Pastore for insightful discussions. The authors also would like to acknowledge an intellectual debt to Christoph M. Schmidt many of whose ideas can be found in this paper but who, of course, is not responsible for any errors committed by us. Lehmann also thanks the Volkswagen Foundation for financial support within the project "The political economy of labor market reform in transition: A comparative perspective."

1. Introduction

The evaluation of active labor market policies (ALMP) has been an important research area in North-America and Western Europe for more than two decades. While most of the early work on developing rigorous methods of evaluation was produced in the United States and Canada, the last fifteen years has seen a rapidly increasing share of important work by European based researchers who have contributed to the refinement of evaluation methods as well as to applying rigorous evaluation methods to a different context than the one we find in the flexible labor markets of North-America.[1] In Central and Eastern Europe, where the transition from a centrally planned economy to an economy dominated by market forces started in the early 1990s, some important studies on the efficacy of ALMP have also been undertaken. These studies have increased our understanding of the workings of labor markets and are thus genuine contributions to the general literature on labor markets and the evaluation of ALMP.

The adoption of ALMP, which had been developed and applied in Western OECD countries, was seen by policy makers as well as pundits in Central and Eastern Europe (CEE) as an important weapon in the fight against open unemployment. It is, therefore, worthwhile to ask whether this perception was justified and to what extent labor market policies developed in mature market economies could be legitimately implemented in the context of transitional labor markets, questions which we address in the first part of the paper.

We then proceed to a brief history of the evaluation of ALMP in transition countries. There are essentially two types of approaches to evaluate policy interventions, a macroeconometric approach that uses aggregated administrative data, and a microeconometric approach that is based on individual level data from either unemployment registers or, more frequently, from labor force surveys. As individual level data became available only several years after the regime switch, the first approach dominated the literature in the early years of transition. With the arrival of large micro data sets in the middle of the 1990s, microeconometric evaluation studies started to be a lot more frequent than studies based on the macroeconometric approach. However, as we will argue, one should comprehend the two approaches as complementing tools in policy evaluation. The "working horse" underlying the macroeconometric approach has been the augmented matching function. We will derive a simple version of this function and then discuss several seminal applications and how they dealt with estimation problems arising from the use of regional administrative panel data. The microeconometric evaluation studies have relied either on hazard rate analyses, often trying to model unobserved heterogeneity between program participants and non-participants econometrically, or on different variants of selection on observables, most

[1] See Kluve and Schmidt (2002) and Kluve (2006) for an overview of European evaluation studies.

prominently matching estimators. A first generation of such studies has, for example, been collected in the symposium edited by Boeri and Lehmann (1999), and one of these studies will be discussed in some detail.

The third section of the paper will look in a selective fashion at some recent microeconometric studies in transition economies, with an eye on the value added that these studies produce. In other words, we discuss papers that we consider important contributions to the literature as they help us better understand how labor markets in transition countries function. The subsequent section then focuses on one important aspect of the recent ALMP literature by discussing one of our own papers (Kluve, Lehmann and Schmidt, 2008) in detail: Labor force status might be a particularly good predictor of program participation, and labor force status sequences pre- and post-treatment seem highly correlated (Heckman and Smith, 2004). By conditioning on pre-treatment history when matching participants and controls one might substantially reduce selection bias, especially against the background of a rapidly changing macroeconomic environment that is typical for economies in transition. We try to demonstrate the validity of this proposition for Poland, i.e. in an entirely different context than the one referred to by Heckman and Smith (2004).

The presented "road map" of the paper should make clear that we are not surveying all evaluation studies on ALMP in transition countries, like it was partially done for the early years of transition in Lehmann (1995) and for the more mature stages in Betcherman et al. (2004). We also do not conduct a meta study such as Kluve (2006), who provides a quantitative assessment of all available rigorous EU evaluation studies. Our paper is strongly selective, paying homage to our own idiosyncratic tastes but also, as we hope, to those contributions that have really made an impact for our understanding of labor markets in transition countries.[2] Notably, this review of several program evaluation studies also discusses the core methodological challenges that evaluation exercises face.

2. Types of active labor market policies, scope and applicability

In table 1 we present archetypical types of programs in OECD countries and state their purpose in a generic fashion.

The first type, public employment services, is of great importance. Its main purpose is to make the matching of unemployed workers to vacant jobs more efficient. In most labor markets substantial friction in the informational flow can impede job matching: firms are unaware of

[2] Several important studies have been undertaken evaluating the efficacy of ALMP in East Germany. Since we consider East Germany's transition to a market economy a special case distinctly different from the experiences of other countries of Central and Eastern Europe, these studies will not be considered here.

unemployed workers who are willing to take up vacant posts while unemployed workers do not know of the existence of these jobs. By setting up public employment services that reduce these informational inefficiencies matching can be improved, sometimes in a dramatic fashion. The existence of private agencies that engage in "job brokerage" particularly for jobs with relatively high skill levels does not invalidate the important role of public employment services in the matching process.

Training measures are employed in order to attenuate skills mismatch: in many cases, unemployed workers do not have the skills that firms look for, so through re-training and further-training measures this skills mismatch can, at least in principle, be remedied. Training measures are on average among the most costly measures per unemployed worker. Consequently, there is often strong pressure on public employment functionaries to ensure that the targeted persons are relatively successful in finding regular employment and "creaming effects", i.e. the targeting of unemployed workers with above average abilities, are a common practice in many OECD countries (see e.g. Anderson, Burkhauser and Raymond, 1993, Aakvik, Heckman and Vytlacil 2005).

The category "employment incentives" entails wage or job subsidies, as well as start-up incentives to the unemployed. The immediate purpose of these schemes is to increase labor demand. However, all schemes connected with subsidized employment have as a longer-term aim the building or re-building of human capital, a process which is supposed to allow the unemployed to enter regular, i.e. non-subsidized employment relationships. Wage and job subsidies are also meant to attenuate distortions arising from asymmetry of information regarding the productivity of workers. While workers observe their own productivity, firms do not know workers' productivity and often use observed spells of unemployment as a screening device, assuming that unemployed workers are of low productivity. Wage and job subsidies might enable firms to find out about the true productivity of workers during the subsidized period of employment in the firm, such that firms might be willing to hire the unemployed after the subsidy runs out. Hence, following this logic, wage and job subsidies increase the efficiency of the hiring process in the labor market.

Direct job creation and employment by governmental agencies is often considered employment of last resort. Nonetheless, it aims at the increase of labor demand and prevention of individual loss of human capital during (long) spells of unemployment. Its medium-term efficacy in terms of furthering the integration or re-integration of unemployed workers into regular employment, however, is frequently questioned. In order to avoid that public employment jobs crowd out jobs in the private sector, the former type of jobs often has a very low skills content and is thus not conducive to the building or re-building of human capital. As a consequence, workers who participated in a public employment scheme are often stigmatized by employers as of low productivity and not hired into regular jobs.

In most OECD countries youth unemployment is a serious problem, in particular because some school leavers have not acquired sufficient skills to be employable at a wage that generates a living income. Training measures are meant to enhance the skills of these school leavers as are subsidized wage and job schemes.

Finally, measures for the disabled consist in financial incentives for firms to hire disabled workers on a priority basis or in the establishment of employment quotas for this group of workers. These measures are conceived to essentially fight discrimination, i.e. the exclusion of this group from the labor market. As many of the disabled are highly productive this ALMP not only contributes to more social equity but also enhances the efficiency of the labor market.

Figures 1 and 2 show average expenditures on ALMP as a percentage of GDP for the 15 old EU members and the new member states. Inspection of the two figures leads us to infer two important points. First, there is large variation in the amount that countries spend on ALMP, both in the old and the new member states. If we take the EU-15 for example, the biggest spender, Sweden, uses nearly 2% of GDP on ALMP while the low spenders, e.g. Luxemburg and Greece, spend less that 0.5% on these policies. The second point one can immediately make is the limited amount of funds spent on ALMP in the new member states, as the highest spender, Hungary, spends roughly as much as Luxemburg and Greece, the low spenders among the old member states. So, even though unemployment rates in the new member states are of the same magnitude as in the EU-15, and often higher, spending on ALMP is very limited. This is not surprising since transition countries have been confronted with major fiscal problems and have had particularly limited funds for labor market policies. In addition, since income support for the unemployed has priority in the eyes of policy makers and the public, active labor market policies are treated like a residual category.

How applicable are policies, which were developed in mature capitalist economies several decades ago, in transition economies? To answer this question, we first need to briefly recapitulate the main aims of ALMP as they were developed in particular in the United States more than 3 decades ago. During the time of the Vietnam War certain demographic groups, e.g. minority youth in inner cities, experienced high unemployment rates in times of overall full employment. Increasing government spending would have resulted most likely in increasing inflationary pressures without necessarily ensuring jobs for marginal groups. As argued by Tobin (1972), schemes were required that increased the human capital of those individuals poorly prepared for the labor market. Such schemes would be able to “cheat the Phillips Curve”, i.e. to lessen the trade-off between a lowered aggregate unemployment rate and inflation. This macroeconomic rationale of ALMP, to lower unemployment without increasing inflationary pressures stands at the beginning of the development of modern “active” schemes to fight unemployment. Such conceived ALMP measures have implied for OECD countries over the last three decades that these measures target marginal or marginalized groups in the labor market. Marginal groups consist of workers who experience problems of finding regular employment from the beginning of their working life. Marginalized groups, on the other hand, consist of workers who were in regular employment but experience difficulty to flow out of unemployment in times of major contractions and enter the state of long-term unemployment where they might actually loose human capital or might be perceived to do so by employers. This targeting has to be seen in the context that the vast majority of workers are continuously employed and have no unemployment spells over their working life. ALMP measures summarized under b, c, d, and e in table 1 seek to integrate marginal or to re-integrate marginalized groups in the labor market. By raising their human capital in training and subsidized employment schemes, the labor market attachment of such groups is meant to be strengthened and the probability of employment or re-employment boosted.

Labor markets in the transition countries of Central and Eastern Europe (CEE) have had special features for many years after the onset of transition that made them quite different from labor markets in mature OECD countries. We focus on those features that are relevant for the adaptability of ALMP measures. Given a lack of physical capital and a very limited amount of entrepreneurial abilities job creation rates have been small in international perspective in most transition countries of the region. Hence, those workers flowing into unemployment as a consequence of labor shedding during the restructuring process have had great difficulty exiting from that state, leading to a “stagnant unemployment pool” and rising long-term unemployment during the early years of transition (Boeri, 1994). One implication of the “stagnant” nature of the unemployment pool was that many among the unemployed were not necessarily belonging to the marginal or marginalized groups mentioned above but belonged to the core workforce. Consequently, in restructuring transition economies more so than in mature OECD countries we find a significant component of the unemployed – and even of the long-term unemployed – showing strong labor market attachment and possessing a large stock of human capital. This meant that those who participated in an ALMP measure found much tougher competition among the unemployed than in most mature OECD countries.

Adopting ALMP measures in a mechanic fashion by targeting these measures on the least capable among the unemployed might, therefore, be an extremely inefficient way to spend the very scarce resources that transition governments have. In other words, given the relatively high "quality" of the average unemployed even if the human capital of marginal persons is increased, this increase might not be sufficient to enable these persons to compete with those potentially very productive workers who also find themselves in the unemployment pool in transition countries, something that would not occur to the same degree in mature OECD countries. Hence, spending resources in transition countries predominantly on marginal persons, who are above all the unskilled, the older and often female workers, might be wasteful, since their employment or re-employment probability might not be affected by participation in a scheme. These considerations have to be seen in light of the fact that in many transition countries training measures are actually targeted at the best among the unemployed and not at marginal groups (Lehmann, 1995, and Betcherman et al., 2004). These "creaming effects" might, however, also be an expression of inefficiency since the targeted persons might have found regular employment even without participation in such ALMP measures.

The upshot of these theoretical considerations is then that what has worked in mature OECD countries might not work in transition economies. It is, therefore, vital to undertake rigorous evaluations of the efficacy of ALMP programs in the latter group of countries. The very tight budget constraints of transitional governments are a further reason for the heightened importance of program evaluation.

3. Macroeconometric evaluation of ALMP measures

The basic principle underlying the macroeconometric evaluation of ALMP measures is to establish whether such measures lower the overall unemployment rate holding all other determinants of the unemployment rate constant. If ALMP measures are administered on a large scale such an effect should be observable in the data.

For this basic principle to be implemented empirically, flow analysis of aggregate administrative data is used. Underlying this implementation is the idea that inflows into unemployment are relatively constant and that the change in the unemployment rate is mainly determined by the outflow rate from unemployment (Pissarides, 1986).[3] Let U_t , U_{t+1} be the stocks of unemployment at the beginning of periods t and t+1, and let I_t and O_t be the inflows and outflows into and from unemployment during period t. Then by definition,

$$U_{t+1} \equiv U_t + I_t - O_t \qquad (1) .$$

[3] The determination of the stock of unemployment by the outflow rate is a contentious issue. Burgess (1989), e.g., highlights the importance of inflows in the British case.

For presentation purposes we assume a steady state, which allows us to drop time subscripts. With simple arithmetic we arrive at the following equation,

$$U = \frac{I}{o} \qquad (2),$$

where o = O/U is the outflow rate from unemployment. Under the assumption that inflows are relatively constant changes in the stock of unemployment are determined by the outflow rate. In a heuristic fashion, the macroeconometric evaluation of ALMP simply consists in establishing the effect of ALMP measures on the outflow rate from unemployment by estimating the following model:

$$o = f(X_1; X_2) \qquad (3),$$

where $\mathbf{X_1}$ is a vector of variables controlling for the state of the labor market, and $\mathbf{X_2}$ is a vector or a scalar with appropriately measured ALMP measures. Holding the elements of $\mathbf{X_1}$ constant, we want to find out whether the partial $\partial f / \partial X_2$ is positive or zero (it is highly unlikely that it is negative). A positive sign on the partial $\partial f / \partial X_2$ tells us that the ALMP measure(s) has (have) raised the overall outflow rate from unemployment, with substitution effects already netted out. With duration-specific outflow rate models the extent of dead weight loss can also be estimated. The great advantage of macroeconometric evaluation thus consists in the accounting of some of the distortive effects like substitution and dead weight loss effects. The macroeconometric approach essentially establishes general equilibrium effects of ALMP in the labor market, something that cannot be done with microeconometric evaluation methods.

The working horse of our heuristic outflow rate model is the "augmented matching function", i.e. the usual matching function whose arguments are augmented by variables representing ALMP measures. We now derive a class of simple theoretical models of the "augmented matching function", borrowing partially from Lehmann (1993).

Let O be the number of people leaving unemployment during a period, U and V be the stocks of unemployed and vacant jobs at the beginning of the period. To account for shifts in the Beveridge Curve, i.e. for a changing quality of the stock of unemployment, we introduce a search effectiveness index, s. We define s the average search effectiveness of the unemployed at a given point in time[4], when ALMP measures meant to enhance search effectiveness are absent. Also let

$$s^* = s(1+\gamma M), \quad where\ 0 \le s^* \le 1\,,\ M = \sum_{i=1}^{m} \beta_i E_i\ and \sum \beta_i = 1 \quad (4)$$

M is the weighted sum of those ALMP measures which do not directly create additional vacancies, but are meant to increase the search effectiveness of the unemployed. On a priori grounds we can assume that γ is non-negative, i.e. that these ALMP measures should not lower the average search effectiveness of the unemployed. We then postulate that the number of people leaving unemployment into regular jobs is mainly determined by V and the search effective part of the stock of unemployment, i.e.

$$O = f(V, s^*U),\ with\ f_1, f_2 > 0 \qquad (5)$$

Two points need to be made about this outflow function. First, as long as outflows are into employment and not into inactivity, our outflow function is approximately equivalent to the aggregate matching function as presented e.g. in Blanchard and Diamond (1989). In a Western context, researchers have often restricted their attention to males, with the conjecture in mind that male outflows from unemployment have employment as their destination state (e.g. Pissarides and Haskel, 1987, Jackman and Layard, 1991, and Lehmann, 1993). The administrative data in transition countries often report outflows into regular jobs as well as general outflows. When the former type of outflow data is used it is easy to argue that equation (5) is the equivalent of a matching function. .Second, matching models are often criticized that they neglect the competition for jobs between the unemployed and the employed (e.g., Burgess, 1989). While this criticism has merit, it is not very relevant in our context, where we want to analyze the *additional* effects of ALMP measures on outflows from unemployment. Casual evidence tells us that *the unemployed (and certainly the long-term unemployed) who are helped by ALMP measures* do not compete directly with the employed. Also, on a conceptual level, we have maintained that ALMP measures in mature OECD countries are particularly targeted at marginal or marginalized groups in the labor market. It is clearly not very likely that such groups are competing directly with the employed for jobs. Our outflow function essentially allows us to determine, whether ceteris paribus the hiring of the unemployed has been improved by ALMP measures.

Since we do not know a priori which returns to scale apply in transition countries[5], we log-linearize equation (5) and arrive at the following equation:

[4] The average search effectiveness s crucially depends on the duration structure of unemployment and thus varies over time.

[5] In large mature capitalist economies, constant returns to scale are often assumed (see e.g. Jackman and Layard, 1991)

$$\ln O = \alpha_0 + \alpha_1 \ln V + \alpha_2 \ln U + \alpha_2 \ln[s(1+\gamma M)] \qquad (6)$$

For small γM, $\ln(1+\gamma M) \approx \gamma M$, we get:

$$\ln O \approx \alpha_0 + \alpha_1 \ln V + \alpha_2 \ln U + \alpha_3 \ln s + \alpha_4 M \qquad (7)$$

Adding a white noise error term, a time trend and seasonal dummies, we arrive at an empirical equation that we can estimate:

$$\ln O \approx \alpha_0 + \alpha_1 \ln V + \alpha_2 \ln U + \alpha_3 \ln s + \alpha_4 M + \alpha_5 t + \sum_{j=2}^{4} d_j + \varepsilon \qquad (8).$$

Most early papers that empirically evaluated ALMP in transition countries have empirical equations similar to equation (8): outflows out of unemployment or outflows from unemployment into regular employment are assumed to be mainly determined by the stocks of vacancies and unemployment, by a measure of the search effectiveness of the unemployed and by appropriate measures of ALMP. These latter measures are either stocks of participants or expenditures on ALMP schemes. The search effectiveness of the unemployed essentially is a conceptual device that accounts for the duration structure of the unemployment stock. The basic idea behind this concept is that for a given level of vacancies the larger the share of persons with long unemployment spells the lower the number of matches. A simple measure to account for the duration structure of the unemployment stock is a dummy for the long-term unemployed, which is used e.g. by Pissarides and Haskel (1987). More sophisticated measures of the search effectiveness are used in Jackman and Layard (1991) and in Lehmann (1993), but are hard to construct in an early transition context where the available data only have a short span.

It is this time series limitation of aggregate data at the national level that causes major problem in transition economies when evaluating ALMP with the help of macroeconometric methods. In mature market economies there exist long time series on unemployment, vacancies and ALMP measures; for example, in Britain we have quarterly administrative data at the national level that go back to the 1960s. As long as these time series are stationary, simple OLS estimation gives consistent estimates of the coefficients on the right-hand-side variables as long as the used time structure of these variables guarantees that they are predetermined.

In transition economies, administrative data aggregated at the national level only provide a few data points and cannot be used for meaningful estimations. Instead, researchers use regional panel data that have a relatively large number of observational units (N) and are high frequency, i.e.

monthly or quarterly data. The main problem with such high frequency regional data is the endogeneity of ALMP measures: in regions in which during a given month unemployment rises, i.e. where outflow rates fall and/or inflows rise, policy makers might increase the share of expenditures on ALMP for this particular month in these regions. While the overall financial allocation for ALMP for the whole year per region is determined at the beginning of the year, policy makers often have substantial discretion in allocating these funds over the various months. By reacting to a falling outflow rate or rising inflows with increased spending they confound the effect of expenditure of ALMP on the outflow rate, thus rendering the coefficient estimates on the ALMP measures inconsistent. An allocation rule of yearly predetermined funds that allows a redistribution of monthly expenditures across the year, while creating an endogeneity bias, at the same time provides the basis for finding a valid instrument as pointed out by Boeri and Burda (1996) and Boeri (1997). The predetermined amount of yearly regional expenditures can be assumed to be highly correlated with monthly or quarterly regional expenditures and not to be correlated with the error term. Boeri (1997), using monthly regional panel data for the Czech Republic, Hungary, Poland and the Slovak Republic, finds for three of the countries a positive effect of ALMP on outflow rates from unemployment, demonstrating at the same time that the OLS coefficient estimates are larger than the estimates derived with IV.[6]

The macroeconometric evaluation of ALMP has somewhat fallen from grace as micro data became available in many transition countries, even though this approach is in principle able to establish the net effect of an ALMP measure, when substitution effects have been netted out, and is able to detect dead weight loss. There are several reasons why this happened. First, even relatively sophisticated estimation techniques as, e.g., employed in Boeri and Burda (1996) and Boeri (1997) could not completely hide the fact that the administrative high frequency regional data had major drawbacks. Second, it is always difficult to tease out a significant correlation between unemployment rates and the level of ALMP expenditures across countries as figures 3 and 4 show for the EU-15 and the new member states. Third, the main impulses given to the evaluation literature in the last twenty years, associated above all with the name of James Heckman, have been of a microeconometric nature.

[6]Boeri shows that with his high frequency data increased inflows into unemployment are be positively correlated with higher ALMP expenditures as well as with higher outflow rates, thus producing an upward bias of OLS estimates. The scenario of increased ALMP expenditures due to a decrease in outflow rates mentioned in the text would instead result in a downward bias.

The microeconometric evaluation problem and "first generation" papers on the evaluation of ALMP in transition countries

The microeconometric evaluation of ALMP is interested in the impact of program participation on post-treatment labor market outcomes. The post-treatment outcomes considered are a) variables capturing labor market status, such as (re-) employment probability or the probability to leave the unemployment register, but sometimes also b) earnings. The conceptual challenge of measuring a causal impact of the active labor market program lies in the comparison of the realized outcome of persons who have participated in the program with the *hypothetical* labor market outcomes that these individuals would have realized if they had not participated in the scheme. This approach requires establishing a credible counterfactual, a methodological challenge often referred to as the "evaluation problem". This core methodological aspect of evaluating ALMPs, which we briefly reiterate here, along with potential solutions, have been discussed extensively in the literature (cf., for instance, Heckman, LaLonde, Smith 1999, Blundell and Costas-Dias 2000, Kluve and Schmidt 2002, and many others).

Why do we have a problem in establishing a credible counterfactual? Consider a binary treatment variable D, which indicates treatment participation or its absence, and let the outcome variable $Y=Y_0$ if D=0 and $Y=Y_1$ if D=1. For a particular individual the observed value is:

$$Y = DY_1 + (1-D)Y_0 \qquad (9).$$

The unit level effect, on the other hand, $\Delta = Y_1 - Y_0$, is never directly observable, since we cannot observe the same person participating in a program and not participating in it. Because individual level effects cannot be observed, research on program evaluation has focused on average treatment effects. The most commonly used evaluation parameter is the average treatment effect of the treated, ATET:

$$E(\Delta \mid D=1) = E(Y_1 - Y_0 \mid D=1) = E(Y_1 \mid D=1) - E(Y_0 \mid D=1) \qquad (10)$$

This parameter consists of the difference between the average outcome of participants in the participation state and the average outcome of participants in the non-participation state. Clearly, this last expression is the desired counterfactual. The vast literature on the microeconometric evaluation of ALMP is therefore essentially concerned with the construction of control groups such that:

$$E(Y_0 \mid D=1) \approx E(Y_0 \mid D=0) \qquad (11).$$

This allows replacing the average outcome of participants in the non-participation state (the counterfactual) with the average outcome of non-participants in the non-participation state, which is observable. Expression (11) is most likely to hold when there are – on average – no observable or unobservable differences between the participants (treatment group) and non-participants (control group). In general, randomized controlled trials (RCTs) are the most straightforward and most convincing research design to construct the desired counterfactual.

After this heuristic introduction of the evaluation problem, we shall proceed to a more rigorous treatment of the involved issues, discussing a set of evaluation studies in a transition context that can be broadly labeled "first generation" and "second generation" papers. From the set of "first generation" studies contained in the symposium edited by Boeri and Lehmann (1999), we will discuss the article by Vodopivec (1999), who applies sample selection correction methods. Then we will proceed to present a selection of "second generation" contributions that have been conducted in recent years.

Before we turn to the discussion of the individual studies on the microeconometric evaluation of ALMP in transition countries, we briefly discuss the rationale of this approach. The way we posed the microeconometric evaluation problem makes it clear that these studies cannot establish the overall effect of an ALMP measure. In other words, the effectiveness of a measure at the individual level can never tell us whether this measure, e.g., lowers the unemployment rate in a labor market. Taking training for example, even if an individual's chances of finding a regular job has increased through his or her participation in a training measure, this person might just "jump the queue" of those waiting to enter regular employment. So, substitution effects might render the effect neutral as far as the overall outcome in a labor market is concerned. On the other hand, if an effective training measure is applied on a large scale, this might increase the effective labor supply and thus lower equilibrium wages – or lower upward wage pressure – and result in more employment. The crucial point regarding microeconometric evaluation is, however, that this type of evaluation can only tell us how the individual fares due to participation in a scheme, while it cannot ascertain whether there are beneficial effects for the labor market as a whole. Nonetheless, effectiveness at the individual level is of course a necessary condition for a program to work at the aggregate level at all. Hence, it is vital for policy makers in transition countries, who face very tight budgets, to understand which programs are (in)effective at the individual level, and for what reason.

Vodopivec (1999) on Slovenia

The paper by Vodopivec (1999) is a good example of an informative, rigorous microeconometric evaluation of an ALMP scheme in early transition. He looks at the effectiveness of the Slovenian public works program covering the years 1994 to 1996. His main interest lies in the question whether participation in the scheme increases the chance of finding a regular job. During the analyzed period, the Slovenian public works program is different from public works in other transition countries at least in three regards. First, the human capital content of the offered job is higher on average than in public works jobs in other transition countries; second, many of these jobs in Slovenia have a duration of one year, and are thus substantially longer than elsewhere. Third, given the two mentioned characteristics, it is not surprising that the educational composition of the participants is different in Slovenia, with many more relatively educated persons than we find among the public works participants in other transition countries.

The available data ensure that the first two sources of bias mentioned above, which often exist in evaluation studies, are not an issue.[7] So, we can concentrate on the methodological issue how the author deals with the third source of bias caused by possible selection into the program based on unobservable characteristics.

The dependent variable of interest is $EXIT_{ni}$, which shows individual i's labor market status after searching n months for a job. For those who participated in the public works program, the start of the searching time was set by Vodopivec to zero at the moment when they finished their participation in the public works program. For those who did not participate in public works, the start of the searching time coincided with the registration at the employment office. The variable $EXIT_{ni}$ can take on three values: 0, if after *n* months the individual is still unemployed; 1, if after *n* months she is employed; and 2, if after *n* months she is out of labor force.

The individual's labor force status after n months of job search is modeled as:

$$EXIT_{ni} = \mathbf{X_i \beta_1} + PW_i\beta_2 + \varepsilon_i \quad (12)$$

where $\mathbf{X_i}$ - is a vector of personal characteristics (gender, ethnicity, and age) and human capital characteristics (education, work experience, health condition), PW_i is a dummy representing past participation in public works ($PW_i = 1$ if an individual participated in public works, 0 otherwise), and $\boldsymbol{\beta_1}$ is a vector of parameters and β_2 a parameter to be estimated. By assumption, $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma_\varepsilon^2$.

[7] The reader is referred to the detailed description in Vodopivec (1999).

We might get biased estimates of the impact of public works on chances to find a job if there is a problem of selection. Individuals opting to participate in public works may differ from those opting not to in many aspects, some of which may be unobservable. If these unobservable characteristics also affect the job prospects of individuals, then equation (12) is misspecified and the estimated coefficient β_2 biased. Clearly, this bias can be negative or positive. Vodopivec proposes Heckman's two-stage procedure to remedy the selection problem. In the first stage, an equation of participation in public works is estimated, with regressors derived from the selection process described by Vodopivec in the paper. The outcome of that stage is a new variable (the inverse Mills ratio, λ), to be used as one of the regressors in the second stage, that is, in the estimation of equation of exit from unemployment.

The participation equation takes the following form:

$$PW_i = \mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2} + u_i \qquad (13)$$

where $\mathbf{X_i}$ are personal and human capital variables, and $\mathbf{Z_i}$, factors which capture criteria for selection for public works (number of dependents, for example). This estimation produces a new variable – the inverse Mills ratio - $\lambda_i = \phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2})/\Phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2})$, for participants of the public works, and $\lambda_i = -\,\phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2})/(1 - \Phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2}))$, for non-participants, where $\phi(.)$ and $\Phi(.)$ are standard normal and cumulative standard normal distributions.

Under the assumption that ε_i and u_i are distributed as a bivariate normal with correlation coefficient ρ, for participants we get the following conditional expectation,

$$E(EXIT_{ni} \mid PW_i =1) = \mathbf{X_i}\boldsymbol{\beta_1} + \beta_2 + E(\varepsilon_i \mid PW_i =1) = \mathbf{X_i}\boldsymbol{\beta_1} + \beta_2 + \rho\sigma_\varepsilon \{\phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2})/\Phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2})\},$$

while for non-participants we have:

$$E(EXIT_{ni} \mid PW_i =0) = X_i\beta_1 + E(\varepsilon_i \mid PW_i =0) = \mathbf{X_i}\boldsymbol{\beta_1} + \rho\sigma_\varepsilon \{-\,\phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2})/(1 - \Phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2}))\}.$$

The difference in the conditional expected value of EXIT between the participants and non-participants is thus:

$$E(EXIT_{ni} \mid PW_i =1) - E(EXIT_{ni} \mid PW_i =0) = \beta_2 + \sigma\rho_\varepsilon \{\phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2})/\Phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2})\,(1 - \Phi(\mathbf{X_i}\,\boldsymbol{\gamma_1} + \mathbf{Z_i}\boldsymbol{\gamma_2}))\} \quad (14).$$

By including the selectivity correction term in the estimation of equation (12), the bias presented by the second term of the right-hand-side of equation (14) is purged from the estimates. Program selection rules are used by Vodopivec to produce an instrument identifying the selection equation. The procedure outlined here is rather conventional and has two major potential problems. On the one hand, the joint normality of ε_i and u_i are just assumed and this assumption might be questionable in many cases. On the other hand, the author does not seem to control for the changing macroeconomic environment typical for a transition economy, something that can be done e.g. with a "moving window" technique employed in the paper by Kluve et al. (2008) and discussed below.

The policy relevant evidence of Vodopivec' paper can be briefly summarized. Immediately following participation in the scheme participants have a higher likelihood of finding a regular job than non-participants, but this positive effect disappears as participants continue to linger on in unemployment. A full discussion of this result can be found in the paper. Our focus in this section has been to highlight the evaluation problem and to demonstrate a "classic" method dealing with this problem.

Some second generation models and their contribution to the literature

Micklewright and Nagy (2005) on Hungary

Randomized controlled trials (RTCs) to evaluate ALMP are rare in Europe and virtually absent in transition countries. Such experimental studies consist in the randomization of the control and treatment group, i.e. treated persons and members of the control group are randomly assigned. This assignment process is completely beyond the workers' control but also does not discriminate as to who will receive treatment. As long as there is no "contamination" of the two randomly created groups during the treatment (e.g. a control person switching to treatment), i.e. as long as the researcher has considerable control over the delivery and the individual compliance with the program, experimental studies provide the most convincing solution to the evaluation problem. If sample sizes are large, randomization ensures a complete balancing of observable and unobservable characteristics and thus makes the groups of treated and controls truly comparable, which in turn implies that differences in outcome variables can be attributed to the programs under evaluation.

The one and only RCT in a transition economy, to our knowledge, is the paper by Micklewright and Nagy (2005). This study investigates whether job search monitoring has an impact on the unemployment benefit duration and outflow from unemployment in the Hungarian labor market. The authors' experiment is rather modest in that it randomly divides benefit claimants into treated persons, who are invited to visit the employment office every 3 weeks and who are asked intensively about their job search, and into control persons, who have to come to the employment office every 3 months and who are not asked any questions related to job search. This experiment lasted 4 ½ months, which implies that members of the treatment group made a maximum of 4 visits to the employment office. Since the benefit claimants in the treatment were not aware of the consequences of their search behavior, the authors assume that the treatment should boost efforts to exit unemployment for employment or refrain persons in the treatment group from further claiming unemployment benefits, i.e. increase outflows to inactivity. The authors make the important point that increased search effort might not translate into larger outflow rates from unemployment into employment because of weak labor demand; even when the heavily monitored benefit claimants lower their reservation wages in a substantial fashion, job offers might still not arrive.

This experimental study nicely demonstrates the point that when assignment to treatment is random[8] relatively simple econometric techniques can convincingly establish a causal effect. After the authors have shown that benefit exhaustion and the ending of the experiment constitute for both the treatment and the control group nearly two thirds of all exits from the unemployment benefit register, they compare Kaplan-Meier estimates of survival rates in the benefit register of the treated and the controls. For the full sample there are no statistically significant differences in these survival rates, as demonstrated by the performed log rank tests. However, when the sample is split into women who are younger than 30 years, women 30 years and older, and men, the results are different. The older female group shows clear differences in the survival rates of the treatment and the control group. The estimates of a simple hazard rate model confirm this result, as only older women among the treated have a higher hazard ratio and also substantially higher exit rates to regular jobs. The authors in a final step estimate two hazard rate models, the first with marital status interacted with the treatment dummy and the second with local labor market conditions proxied by the local unemployment rate interacted with the treatment dummy. Their estimates show that only married older women experience a statistically significant treatment effect. Unsurprisingly, this treatment effect is smaller in regions where local labor market conditions are worse.

This experimental study produces some interesting results that have implications for labor markets in transition in general. With simple econometric techniques it shows nicely that for unemployment benefit claimants general policies that try to boost their search efforts might not translate into improved labor market outcomes. The fact that only married older women experience a treatment effect can be interpreted as showing that only this group among unemployed workers can "afford" to lower their reservation wages enough to exit from unemployment into regular

[8] Random assignment is assured in this experiment since claimants with odd birthdays were assigned to treatment while claimants with even birthdays to the control group.

employment. The average worker, on the other hand, who cannot rely necessarily on other family members' income, finds himself confronted with a weak labor demand that translates into too low wage offers given the income support provided by the unemployment insurance system.

Rodriguez-Planas and Benus (2006) on Romania

The paper by Rodriguez-Planas and Benus (2006) that evaluates 4 Romanian ALMP schemes (job brokerage, self-employment assistance, training and retraining, and public employment) uses a rich data set with many covariates. It is above all this richness of the data that makes this paper interesting, since it allows the construction of a convincing non-experimental control group using statistical matching methods. The average treatment effect of the treated (ATET) can be identified with a matching approach when the conditional independence assumption (CIA) holds. This identification assumption is also called selection-on-observables assumption or "unconfoundedness" assumption (Imbens, 2004). The assumption asserts that conditional on a vector of covariates **X** the assignment mechanism D is independent of potential outcomes Y_0 and Y_1 (see Rubin 1974, 1977). Given this unconfoundedness assumption the ATET is identified, since:

$$E(\Delta|X,D=1) = E(Y_1|\, X,D=1) - E(Y_0|\, X,D=1)$$

$$= E(Y_1|\, X,D=1) - E(Y_0|\, X,D=0).$$

That is, essentially, conditional on the vector **X**, assignment to treatment can be considered random and the average non-participation outcome of the non-participant population can be used to replace the counterfactual non-participation outcome of the participant population. It is in this sense that matching mimics an RCT. In addition to the unconfoundedness assumption, for identification we also need the "common support" assumption for covariates and treatment, i.e. we need $0< Pr(D=1|X)<1$. We will demonstrate the importance of this second assumption below.

As pointed out by Imbens (2004) for the unconfoundedness assumption to hold the researchers need to have control over all observable variables that can influence the outcome of interest as well as the assignment mechanism. To see what this concretely implies let us take a closer look at the data that Rodriguez-Planas and Benus have at their disposal. The data, a random sample of roughly 4000 of all workers who registered in 1999 at Romanian employment offices, were collected at the beginning of 2002. The authors assume that participation in the four ALMP measures was confined to 1999 since the schemes are known to be of short duration. About half of the sample participated in a scheme. During a specialized survey rich information was collected, including retrospective information for the years 2000 and 2001 (when according to the authors participation in the ALMP schemes had already ceased), as well as information on employment and earnings in 1998, i.e. prior to program participation. The authors perform matching on the propensity score, i.e. on the estimated probability of participating in an ALMP scheme. This popular matching approach circumvents the "curse of dimensionality" (i.e. the problem of finding matched treated and control observations when **X** is large) by using a result due to Rosenbaum and Rubin (1983), who show that instead of conditioning on a potentially high-dimensional **X** it is sufficient to condition on the propensity score **P(X)** – a scalar – for unconfoundedness to hold.

In the estimation of the propensity scores ideally all the variables determining program participation and labor market outcomes should enter the set of regressors. The authors moot that the level of education, previous earnings and pre-program labor force status (for the latter factor, see also Kluve et al., 2008, discussed below) are "important factors in determining whether an individual will participate in any program, as well as in which of the programs." In addition, since these factors also have an impact on future labor market outcomes, they should definitely be included when estimating propensity scores. Demographic characteristics like age, gender and marital status also have an influence on future labor market prospects, while the position in the household (e.g., head of household) has an influence on a person's decision to participate in the program.

To account for different local labor market conditions and for different implementation of ALMP measures across counties, the authors also have variables linked to the county of residence, like the unemployment rate and the type of settlement. Finally, to take account of unobserved local aspects related to implementation and utilization of programs as well as to placement practices, the authors include county dummies when estimating propensity scores. Given this impressive list of observables, the assumption of unconfoundedness seems rather plausible. There could, of course, be systematic differences in unobservable characteristics of the treated and of controls. However, having such a rich arsenal of variables it is likely that by balancing observables one also balances the unobservables over the two groups. In addition, the authors when citing motivation and ability as examples of potential unobservable characteristics that might impact on the participation decision and on future labor market outcomes do make the point that earnings in 1998 might proxy for these unobservable characteristics.

The upshot of this discussion is that the use of variables in the matching process needs to be carefully considered, taking recourse to economic theory. When this is done and when one has a rich data set that allows the conditioning on many variables we can be pretty confident that the treatment effect is identified. The authors thus establish convincingly that three measures (job brokerage, self-employment assistance, and training and retraining) are beneficial in that they increase re-employment probabilities and earnings relative to what would have prevailed in the absence of these programs, while public employment are detrimental to participants. A final interesting aspect of this paper consists in the demonstration that program efficacy differs across demographic groups. For example, in the Romanian case job brokerage produces better economic outcomes for younger workers, the short-term unemployed and workers residing in rural areas, while training is beneficial in particular to younger workers. It, therefore, seems in general desirable to evaluate interventions on subsets of participants in order to fully appreciate the effectiveness of these policies.

Bonin and Rinne (2006) on Serbia and Montenegro

Most evaluation studies try to analyze the impact of ALMP measures on “objective” labor market outcomes, i.e. on employment and unemployment rates or wages, for example. The paper by Bonin and Rinne (2006), which evaluates the “Beautiful Serbia” program that was administered to unemployed persons in three cities of Serbia and Montenegro in 2004 and 2005, not only looks at such objective outcomes but also at subjective indicators of self-assessment. This program comprises vocational training and/or temporary employment in construction. In other words, unemployed persons, and in particular long-term unemployed persons, are recruited for vocational training, and the participants in the training measure can then subsequently work temporarily in construction. Many of the unemployed are also hired on these construction jobs without having passed through the training measure.

This study is above all interesting because it looks at the evaluation of ALMPs taking also into account how such programs impact on how people feel about themselves. The main motivation for this approach is to say that the real aim of social policies should be improving how people feel about themselves while labor market outcomes are only a means to reach improved self-assessment by people. The upshot of their estimations shows that the “Beautiful Serbia” program does not lead to better labor market outcomes for participants relative to non-participants but leads to improved self-assessment regarding broader social contacts, better health status and personal qualifications and skills as well as greater chances to find a job.

The authors in our opinion may go a bit far in stipulating improved self-assessment as a sufficient outcome for rendering an ALMP measure socially beneficial. While it is certainly true that happiness of its citizens is the ultimate aim of governmental policy, it is also true that people who experience good labor market outcomes are happier than those who do not. In other words, to achieve sustained feelings of well-being objective labor market outcomes are in all likelihood a necessary condition. In addition, we would argue that given the severe governmental budget constraints of most transition countries – and certainly of Serbia and Montenegro in 2004 and 2005 – one should be allowed to ask how cost effective such programs are. We know from the general evaluation literature that training measures are particularly expensive as are employment programs. This insight leads to the question whether improved social contacts, for example, cannot be achieved with a substantially cheaper social program than a training program. Finally, as we have elaborated at the beginning of this paper, ALMP measures are meant to improve the functioning of the labor market, they are usually not meant to be substitutes for social programs that improve the social inclusion and the health status of marginalized groups.

Besides assessing a different type of program impact, the paper is also of heuristic interest as it nicely develops – in a transition context – the idea what it means to have a "common support" of the covariates of treated and controls given the treatment. Bonin and Rinne do not perform exact matching but match on the estimated propensity score, as detailed above. The upper panel of Figure 5 shows the distributions of the propensity score for participants in both programs and non-participants. Recalling that the propensity score is the probability of participation conditional on the covariates, we can gain three important insights from the figure. First, the characteristics of the participants are distributed quite differently from the characteristics of the non-participants since we find a lot more participants that have characteristics that map into a high probability of participation than non-participants. Second, if we want to compare like with like (i.e. persons with the same characteristics) we need to perform matching with replacement in the upper part of the distribution since there are many more participants than non-participants with comparable characteristics when the estimated probability exceeds 0.55. Third, a few participants have too high estimated probabilities for matching purposes, i.e. none of the non-participants reaches such high estimated probabilities implying, of course, that no one among the non-participants have the characteristics that generate these probabilities. So, these participants are off the common support of the covariates of treated and controls given participation and need, therefore, be excluded from the analysis. These insights are reinforced by the additional two panels, which relate to the two programs separately.

Van Ours (2004) on Slovakia

An important aspect of ALMP evaluation is the potential locking-in effect of subsidized jobs: workers being in subsidized jobs for extended periods of time reduce their search effort for regular employment and thus get essentially locked into unemployment. A study by van Ours (2004) takes advantage of a "natural experiment" in the Slovak Republic to investigate this lock-in effect. In Slovakia, there are two types of temporary employment measures, publicly useful jobs (PUJ) and socially purposeful jobs (SPJ). The latter type puts workers for 24 months in temporary employment while the former had initially a duration of six months. In an earlier study Lubyova and van Ours (1999) find that PUJ has a positive impact on the hazard rate to regular jobs, while SPJ has a negative rate, and this, in spite of the fact that participants in SPJ have on average better human capital characteristics than participants in PUJ. The authors explain this difference in the job finding rate with differing lock-in effects in the two programs.

To prove this assertion van Ours (2004) takes advantage of the fact that the duration of the PUJ measure was increased from 6 to 9 months in 1994 and from 9 to 12 months in 1995, while the duration of the SPJ measure remained constant in these years. Clearly, if lock-in effects explain the differences in the efficacy of the programs, then the increases in the duration of PUJ must have reduced the job finding rate for this program. We have thus a "natural experiment" that enables the researcher to identify the lock-in effects of ALMP programs in Slovakia. While many labor economists who work on transition countries use the term "natural experiment" rather loosely, here we really have such an experiment that sheds light not only on the search behavior in transitional labor markets but on the search behavior in labor markets in general. The empirical results, derived in a careful fashion, confirm the explanation given by Lubyova and van Ours (1999): The increased duration of the PUJ measure lowers the job finding rate in a substantial fashion. Lock-in effects are thus an important aspect of the functioning of ALMP measures, which should be looked at in transition countries as well as beyond.

Kluve, Lehmann, and Schmidt (2008) on Poland

Implementing a "moving window"

A good example for the political usefulness of program evaluation at the micro level is the paper by Kluve et al. (2008) who analyze a training program and a wage subsidies scheme in Poland. The authors find positive effects of training schemes on individual employment probability, while "intervention works" (wage subsidies) result in lower employment rates than would have prevailed if the unemployed participants had not participated in this program. We use this paper to demonstrate how careful exact matching procedures deal with the counterfactual problem and at the same time control for a rapidly changing macroeconomic environment during the early years of transition.

The authors use the supplement to the August 1996 wave of the Polish Labor Force Survey (PLFS) that contains detailed individual labor market histories spanning the months from January

1992 to August 1996. They collapse the available information into monthly trinomial sequences representing employment (1), unemployment (2) and inactivity (0). A fourth potential labor market state is treatment, i.e. participation in ALMP. As figure 6 delineates, monthly states are condensed into quarterly states. Before stating the exact matching procedure in a rigorous fashion, let us demonstrate with the help of figure 6 what these procedures entail. In this figure and in what follows training is taken as the exemplary treatment.

Let us look at trainee 1. She is matched to a control person from a pool of non-participants who has the same following characteristics: age, educational attainment, marital status, gender and residence in the same "voivodship" (region) or residence in the capital or in the provinces. The residence variable is to take account of local labor market conditions that are absolutely essential when evaluating ALMP (see, e.g., Heckman et al., 1997). In the example in figure 6 the trainee enters training for two quarters; after these two quarters she is unemployed for one quarter, and employed for two quarters. In other words the post-treatment employment rate averaged over 3 quarters is 2/3. Note that the trainee and control person are not only matched on the above mentioned observable characteristics but also on having the same pre-treatment history of labor market states. The recent literature emphasizes the correlation of outcome before and after intervention and stresses the role of labor force status dynamics in accounting for unobservable characteristics that might determine participation in a program (Heckman and Smith, 1999, 2004).

In fact, Kluve et al. (2008) further investigate this aspect by alternatively matching treated and controls and estimating treatment effects on the basis of the covariates only, excluding the information contained in the pre-treatment labor force status sequences. They are able to show that the labor force status sequences contain essential information to capture the individual (un)employment dynamics leading to program participation, and that disregarding these dynamics would substantially bias treatment effect estimates.

Finally, in figure 6 note that the control person's outcome variable is analyzed for the exact same calendar period as that of the trainee, which takes account of the fact that particularly in a transition economy one needs to compare labor market outcomes during the same interval of the transition cycle. In our case the control person has two quarters of unemployment and 1 quarter of employment, hence an employment rate averaged over the same quarters as that of the trainee amounting to 1/3. Consequently the difference between the two employment rates is the effect of training for trainee 1. Looking at trainee 2 we see the "moving window" structure very nicely. In addition, this trainee has not just one matching control person but three; the effect of training is then calculated by comparing the post-treatment average employment rate of trainee 2 to the average of the three average employment rates of the three control persons. The overall effect is then calculated by taking all pre-treatment history-specific effects and summing them with the appropriate weights applied to each history-specific effect. How these weights look like will be explained in the rigorous derivation of the matching estimator that now follows.

The matching estimator

Denote the state associated with receiving the intervention – training or "intervention works" – with "*1*", and the state associated with not receiving the intervention with "*0*". N_1 is the number of individuals in the intervention sample, with indices $i \in I_1$, while the sample of potential controls consists of N_0 individuals, with indices $i \in I_0$. Receiving the intervention is indicated by the individual indicator variable D_i (*1* = yes, *0* = no). The potential labor market outcomes in post-treatment quarter *q (q = 1, 2, 3)* are denoted by Y_{qi}^1, if individual *i* received treatment, and by Y_{qi}^0, if individual *i* did not receive treatment. These outcomes are defined as multinomials with three possible realizations ("*0*" = out-of-the-labor-force, "*1*" = employed, "*2*" = unemployed).

Only one of the two potential outcomes Y_{qi}^1 and Y_{qi}^0 can be observed for a given individual. This actual outcome is denoted by Y_{qi}. The objective then is to formally construct an estimator of the mean of the unobservable counterfactual outcome E(Y_q^0 |D=1). Since following the quarterly sequence of labor market outcomes might be too detailed for a direct economic interpretation of results, Kluve et al. (2008) condense the available information further and summarize the post-intervention labor market success of each individual i by the individual's average employment rate over the three quarters following the intervention. Using an indicator function 1(.), these employment rate outcomes are $\overline{Y}_i \equiv \frac{1}{3}\sum_q \mathbf{1}(Y_{qi}=1)$, and $\overline{Y}_i^1$ and $\overline{Y}_i^0$, respectively, for employment rates with and without treatment. Observed outcomes for individual i can then be written as

(15) $$\overline{Y}_i = D_i \overline{Y}_i^1 + (1-D_i)\overline{Y}_i^0,$$

and the impact of the intervention on the labor force status of individual i is given by

(16) $$\Delta_i = \overline{Y}_i^1 - \overline{Y}_i^0 .$$

The parameters of interest are weighted population averages over these individual treatment effects, the mean effect of treatment on the treated (ATET) for types of individuals characterized simultaneously by specific sets of characteristics X; and labor force status histories before treatment h,

(17) $$E(\Delta | X, h, D=1) = E(\overline{Y}^1 - \overline{Y}^0 | X, h, D=1) .$$

The ultimate interest then lies in the average treatment effects over the joint support of X and h given D=1,

(18) $$M = \sum_s w_s E(\Delta \mid s, D=1),$$

with s indicating any possible combination of X and h, and w_s representing the corresponding relative frequency in the treatment sample.

How does this matching approach by Kluve et al. (2008) identify the parameters of interest? As detailed above, matching methods can recover the desired counterfactual for a nonexperimental control group if unconfoundedness holds: Within each matched set of individuals, one can estimate the treatment impact on individual i by the difference over sample means, and one can construct an estimate of the overall impact by forming a weighted average over these individual estimates. Matching estimators thereby approximate the virtues of randomization mainly by balancing the distribution of observed attributes across treatment and comparison groups, both by ensuring a common region of support for individuals in the intervention sample and their matched comparisons and by re-weighting the distribution over the common region of support. The central identification assumption is that of mean independence of the labor market status $\overline{Y}_i^0$ and of the treatment indicator Di, given individual observable characteristics. In this specific application these conditioning characteristics are the demographic and regional variables Xi and the pre-treatment history hi, i.e.

(19) $$E(\overline{Y}^0 \mid X, h, D=1) = E(\overline{Y}^0 \mid X, h, D=0) .$$

The authors elaborate in detail how conditioning on both i) socioeconomic characteristics – including, in particular, information on local labor markets – and ii) detailed labor force status sequences with exact alignment of the pre-treatment period lends plausibility to the unconfoundedness assumption, reflecting a meticulous adjustment of the method to the research question and data at hand. Moreover, using the longitudinal structure of the data, labor force status sequences likely reflect relevant unobserved but time-persistent differences, such as e.g. motivation, between treated and untreated individuals.

The matching estimator applies an oversampling exact covariate matching within calipers, allowing for matching-with-replacement. Specifically, for any treatment group history h for which at least one match could be found, Kluve et al. (2008) estimate the impact of the intervention by

(20) $$\hat{M}_h = \frac{1}{N_{1h}} \sum_{i \in I_{1h}} \left[\overline{Y}_i^1 - \sum_{j \in I_{0h} \wedge X_j \in C(X_i)} \frac{1}{n_{i0}} \overline{Y}_j^0 \right] ,$$

where N_{1h} is the number of individuals with history h who receive the intervention ($N_1 = \sum_h N_{1h}$), I_{1h} is the set of indices for these individuals, $C(X_i)$ defines the caliper for individual i's characteristics X_i, and n_{i0} is the number of comparisons with history h who are falling within this caliper, with the set of indices for comparison units with history h being I_{0h}. The overall effect of the intervention is estimated in a last step by calculating a weighted average over the history-specific intervention effects,

$$\hat{M} = \sum_h w_h \hat{M}_h \quad . \tag{21}$$

using the treated units' sample fractions $N_{1h} / \sum_h N_{1h}$ as weights.

Empirical results

Table 2 presents the average treatment effects on the post-intervention employment rate that Kluve et al. (2008) estimate for the training program. The treatment effect estimate for their sample (A) – the one in which no information on labor force status histories is used – is insignificant, while the estimate obtained from sample (B) – including full information – indicates that participation in Training results in an employment rate that is on average nearly 14 percentage points higher than it would have been in the absence of the program. After stratification into the “employed” (“1111”) and unemployed (“2222”) pre-treatment sequences, sample sizes are too small to draw firm conclusions.

Stratifying the sample by time of entry into training shows that the difference in the employment rates of the treated and the controls is unlikely driven by benefit regulations: Those who entered training before January 1st, 1995 were entitled to a full round of benefits if they could not find a regular job after the intervention, while this generous provision was cancelled for training participants entering after this date (details on regulations are discussed in detail in the article). The treatment effect is, however, larger - and statistically significant – for the subsample of entrants of the earlier period.

To illustrate the performance of the algorithm, Kluve et al. report two "raw effects" resulting from simpler matching variants. The first one reports the effect one would estimate on a sample only using the timing structure, i.e. the moving window, but no information on covariates and labor force status sequences. The second one reports the effect obtained from a simple covariate matching, without use of the moving window. The results for training in table 2 show that both of these effects are similar to the one obtained from Sample (A) (and all are insignificant), and that including the labor force status histories makes the difference in revealing the effectiveness of the training scheme.

Table 3 reports the treatment effect estimates for Intervention Works in the same way. On the basis of sample (A), estimates are more negative than the estimate derived from sample (B), reflecting too many “successful” pre-treatment labor market histories in the composition of the control group. Classifying by labor market history, for the “employed” (1111) histories subsample sizes are rather small and the effects not well defined. For the subsample of “unemployed” (2222) histories, which entails almost 80% of total treated and comparison units, the authors find a significantly negative treatment effect close to the full sample effect. This is certainly no surprise, as the estimate of the full sample effect is dominated by the “2222” subsample effect. The "raw effects" also reported in Table 3 illustrate the importance of including the moving window structure, but above all they show again how crucial it is to control for individual labor force status histories.

In finding reasons for the negative treatment effects of Intervention Works, it is sometimes suggested that subsidized jobs are of lower quality, locking the participating workers in a dead end, rather than preparing them for future labor market success. It might also be a stigmatization effect that causes participants of an employment program like Intervention Works to perform worse in the labor market than non-participants. Prospective employers might identify participants as “low productivity workers” and are not willing to accept them into regular jobs. Kluve et al. suggest another explanation, “benefit churning”: Workers with long unemployment spells who have difficulty finding regular employment might be identified by employment office officials and then be chosen for participation in the Intervention Works scheme only such that they re-qualify for another round of benefit payment. This conjecture is in line with the fact that the large majority of Intervention Works participants stay in the program for exactly six months, the time required to renew benefit receipt eligibility.

If selection into Intervention Works would indeed depend on the outcome in this manner, it might be problematic to maintain the unconfoundedness assumption. To address this aspect, the authors include an additional covariate indicating whether the individual received benefits in the last month before entering the program. This indicator captures the dynamic of running out of benefits (while remaining unemployed) before program start, and thus controls for benefit exhaustion as a selection criterion. The corresponding results shown at the bottom of Table 3 are similar to the ones obtained without the benefit indicator and continue to point to a generally negative effect of Intervention Works.

Conclusions

With the beginning of economic reform in the formerly centrally planned economies of Central and Eastern Europe (CEE), open unemployment rapidly reached comparable levels to those in Western economies. Governments in the region reacted to this rise by adopting active labor market policies (ALMP) as an important tool in the fight against unemployment. The policies that were adopted had been developed in mature market economies, i.e. in a very different context. We, therefore, present the main stylized facts of labor markets in transition and consider the rationale of applying these policies in such labor markets. The main conclusion of these considerations is that one has to be rather careful when transplanting ALMP measures from labor markets where the bulk of the unemployed consists of marginalized and marginal groups to labor markets where even the core of the labor force can experience prolonged spells of unemployment.

Reviewing the evidence on the efficacy of such policies we present rigorous macroeconometric and microeconometric methods of program evaluation, as they were applied in transition economies. Both these approaches have a raison d'être and should be understood as complementing. Macroeconometric evaluation that uses the "augmented matching function" as its workhorse can help establish the overall effect of a program on the aggregate unemployment rate taking account of distortive effects such as substitution and dead weight loss effects. However, in the early years of transition, when this approach dominated, researchers were confined to using high frequency regional panel data in their estimation of augmented matching functions. Apart from the noisiness of such data, the regional and time dimensions also created endogeneity problems as policy makers might redistribute ALMP expenditures across time and regions in response to observed labor market outcomes.

With the availability of large micro data sets the microeconometric approach to program evaluation has certainly gained the upper hand also in transition countries. One important drawback of micro evaluation studies is the fact that distortive effects at the aggregate level are not accounted for. Nevertheless, it is important to establish program efficacy at the individual level, since at this level a much more detailed analysis of programs can be undertaken than is possible with the macroeconometric approach. Apart from one randomized evaluation study that we discuss all other micro studies in transition countries are of the observational type. The main two technical challenges that such studies have to address are the selection problem and the fast changing macroeconomic environment. We present several studies that seem to get close to solving the selection problem and spend some time to show how the "moving window" employed in our own study controls nicely for a rapidly changing economy. The discussed studies also demonstrate that stratification of participants by e.g. gender, age, educational attainment and duration of unemployment is important when trying to establish how the studied measures affect labor market outcomes.

Which lessons can one draw from the surveyed studies about the efficacy of ALMP? The most promising programs seem to be job brokerage and training and retraining schemes while public works, which are politically popular in many of these countries, have nearly always a negative impact on labor market outcomes, due to either stigmatization of participants in the eyes of potential employers or due to "benefit churning". It is also clear from this paper, though, that the number of rigorous evaluation studies is small and that data collection and evaluation need to be intensified before a final judgment can be made about how well ALMP measures work in a transition context.

References

Aakvik, A., J.J. Heckman, E.J. Vytlacil (2005), Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs, *Journal of Econometrics* 125, 15-51.

Anderson, K.H., R.V. Burkhauser, J.E. Raymond (1993), The effect of creaming on placement rates under the Job Training Partnership Act, *Industrial and Labor Relations Review* 46 (4), 613-624.

Betcherman, G, Olivas, K., Dar, A. (2004), Impacts of active labor market programs: new evidence from evaluations with particular attention to developing and transition countries, Social Protection Discussion Paper 402, The World Bank.

Blanchard, O.J., Diamond, P.A. (1989), The Beveridge Curve, *Brookings Papers on Economic Activity*, 1, 1-60.

Blundell, R., Costas-Dias, M., (2000), Evaluation Methods for Non-experimental Data. *Fiscal Studies* 21, 427-468.

Boeri, T. (1994), Transitional Unemployment, *Economics of Transition* 2, 1, 1-25.

Boeri, T. (1997), Learning from transition economies: assessing labor market policies across Central and Eastern Europe, *Journal of Comparative Economics*, 25, 366-384.

Boeri, T, Burda, M. C. (1996), Active labor market policies, job matching and the Czech miracle, *European Economic Review*, 40, 805-817.

Boeri, T., Lehmann, H. (1999), Introduction: unemployment and labor market policies in transition countries, *Journal of Comparative Economics*, 27, 1-3.

Bonin, H., Rinne, U. (2006), Beautiful Serbia, IZA Discussion Paper 2533.

Burgess, S. M. (1989), A model of competition between unemployed and employed job searchers: an application to the unemployment outflow rate in Britain, University of Bristol, mimeo.

EBRD (2007), *Transition Report 2007: People in Transition*. European Bank for Reconstruction and Development, London, UK.

Eurostat (2008), Labour Market Policy – Expenditure and Participants: Data 2006, Eurostat Statistical Books, Luxembourg: Office for Official Publications of the European Commission.

Eurostat (various years), Europe in Figures: Eurostat Yearbook, Luxembourg: Office for Official Publications of the European Commission.

Heckman, J. J., Ichimura, H., Todd, P.E. (1997), Matching as an econometric evaluation estimator: evidence from evaluating a job training programme, *Review of Economic Studies*, 64, 605-654.

Heckman, J. J., Smith, J. A. (1999), The pre-programme earnings dip and the determinants of participation in a programme: implications for simple programme evaluation strategies, *Economic Journal*, 109, 313-348.

Heckman, J.J., LaLonde, R.J., Smith, J.A. (1999), The economics and econometrics of active labour market programs. In: Ashenfelter, O., Card, D. (Eds.). *Handbook of Labor Economics* 3. Elsevier, Amsterdam.

Heckman, J.J., Smith, J.A. (2004), The determinants of participation in a social program: evidence from a prototypical job program, *Journal of Labor Economics*, 22, 243-298.

Imbens, G. W. (2004), Nonparametric estimation of average treatment effects under exogeneity: A review, *Review of Economics and Statistics*, 86, 4-29.

Jackman, R., Layard, R. (1991), Does long-term unemployment reduce a person's chance of a job? A time series test, *Economica*, 58, 93-106.

Kluve, J., Schmidt, C.M. 2002. Can training and employment subsidies combat European unemployment? *Economic Policy* 35, 409-448.

Kluve, J. (2006), The effectiveness of European active labor market policy, IZA Discussion Paper 2018.

Kluve, J., Lehmann, H., Schmidt C.M. (2008), Disentangling treatment effects of active labor market policies: The role of labor force status sequences, *Labour Economics*.

Leetma, R., Võrk, A. (2004), Evaluation of active labour market programmes in Estonia, unprocessed.

Lehmann, H. (1993), The effectiveness of the Restart Programme and the Enterprise Allowance Scheme, London School of Economics, Centre for Economic Performance Discussion Paper 139.

Lehmann, H. (1995), Active labour market policies in the OECD and in selected transition countries, Economic Policy Working Paper 1502, The World Bank.

Lubyova, M., van Ours, J. (1999), Effects of active labor market programs on the transition rate from unemployment into regular jobs in the Slovak Republic, *Journal of Comparative Economics*, 27, 90-112.

Micklewright, J., Nagy, G. (2005), Job search monitoring and unemployment duration in Hungary: evidence from a randomised control trial, IZA Discussion Paper 1839.

OECD, *OECD Employment Outlook*, OECD: Paris, various issues.

Pissarides, C. A. (1986), Unemployment and vacancies in Britain, *Economic Policy*, 3, 499-559.

Pissarides, C.A., Haskel, J. (1987), Long-term Unemployment, London School of Economics Working Paper 983.

Rodriguez-Planas, N., Benus, J. (2006), Evaluating active labor market programs in Romania, IZA Discussion Paper 2464.

Rosenbaum, P.R. and D.B. Rubin (1983), The central role of the Propensity Score in observational studies for causal effects, *Biometrika* 70, 41-55.

Rubin, D.B. (1974), Estimating causal effects of treatment in randomized and nonrandomized studies, *Journal of Educational Psychology*, 66, 688-701.

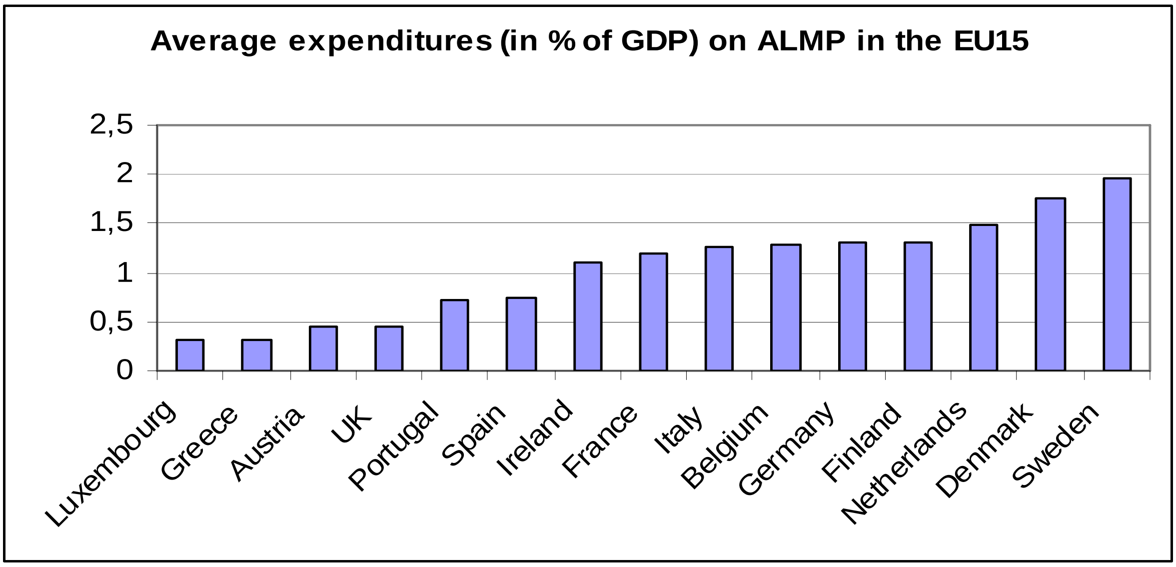
Rubin, D.B. (1977), Assignment to treatment group on the basis of a covariate, *Journal of Educational Statistics*, 2, 1-26.

Tobin, J. (1972), Inflation and Unemployment, *American Economic Review*, 62, ½, 1-18.
van Ours, J. (2004), The lock-in effect of subsidized jobs, *Journal of Comparative Economics*, 32, 37-55.

Vodopivec, M. (1999), Does the Slovenian public work program increase participants' chances to find a job? *Journal of Comparative Economics*, 27, 113-130.

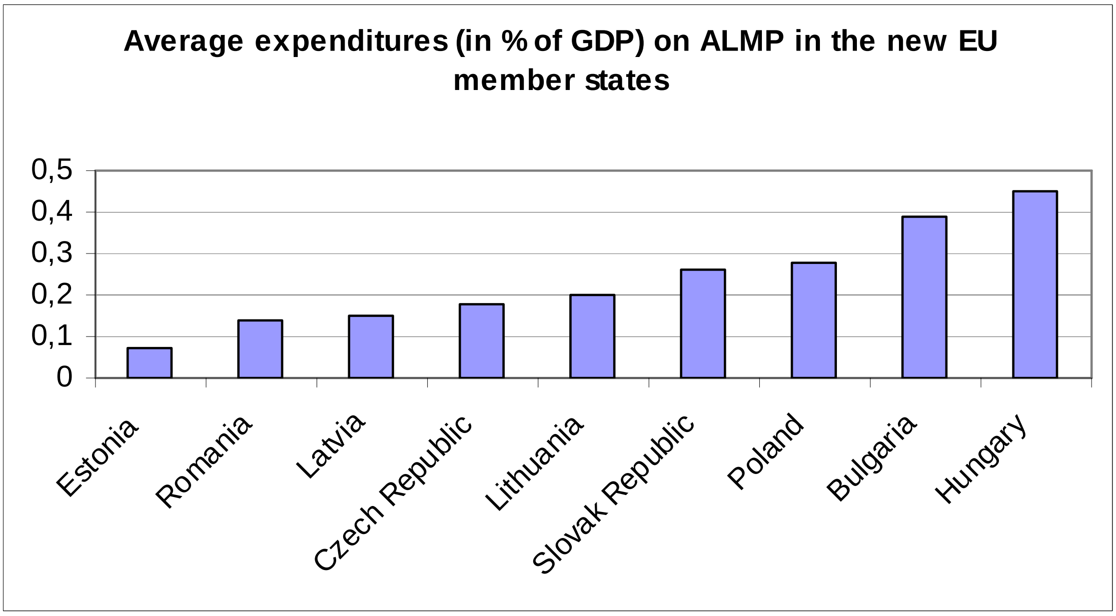
Figures

Figure 1. Scope of ALMP: Expenditures in EU-15



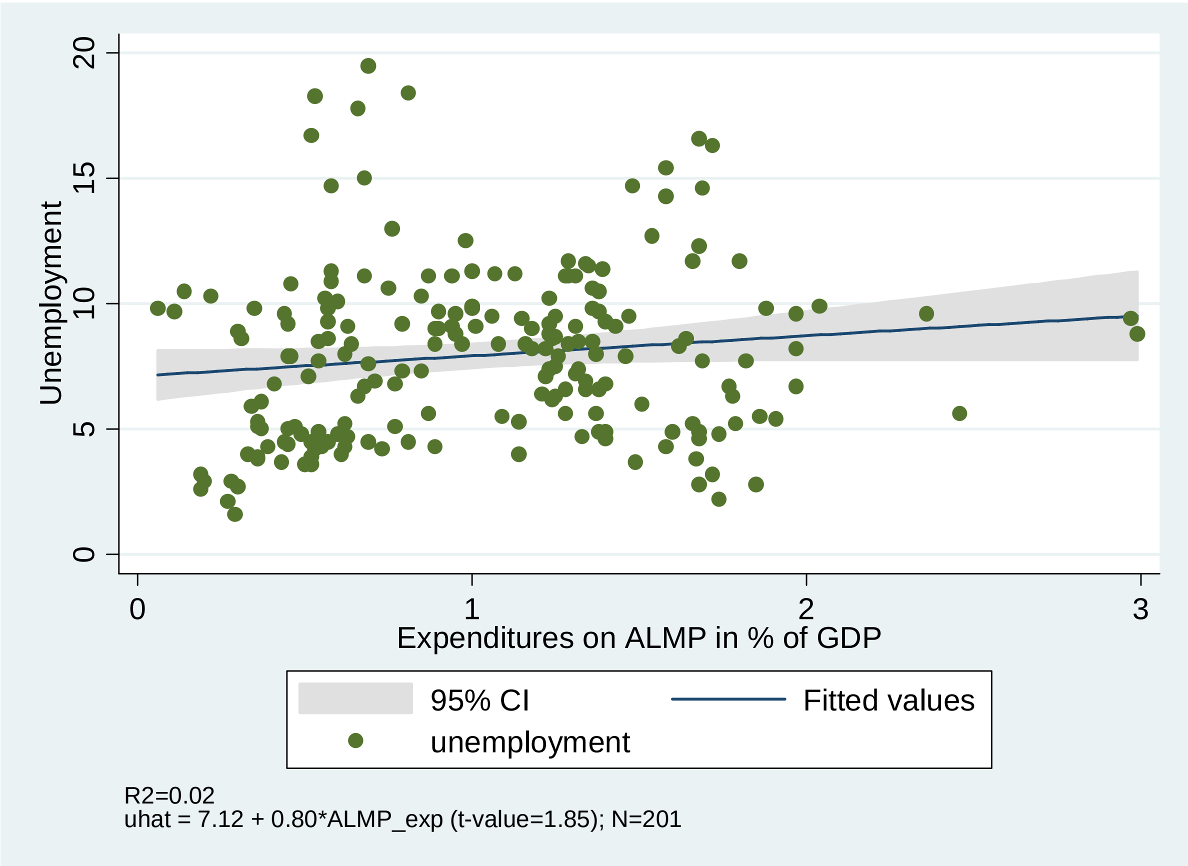
Source: data are from OECD Employment Outlook various issues Notes: Average across 1991-2005 is presented. Active measures include categories 1-7. Since 2002 (Employment Outlook, 2004) there is a change in the definition: active measures include 1-5 categories (2-5 for inflows). For Denmark in 2002, Greece in 2002 and 2003, Italy in 2003, Portugal in 2002 only categories 2-7 are available. The data are missing for Denmark in 2001, Greece in 1999, 2000, 2001, Ireland 1997-2000, 1992 and 1993, Luxembourg 1998-2002, Portugal in 2001, Sweden in 1991, and the UK in 1991.

Figure 2. Scope of ALMP: Expenditures in new EU member states



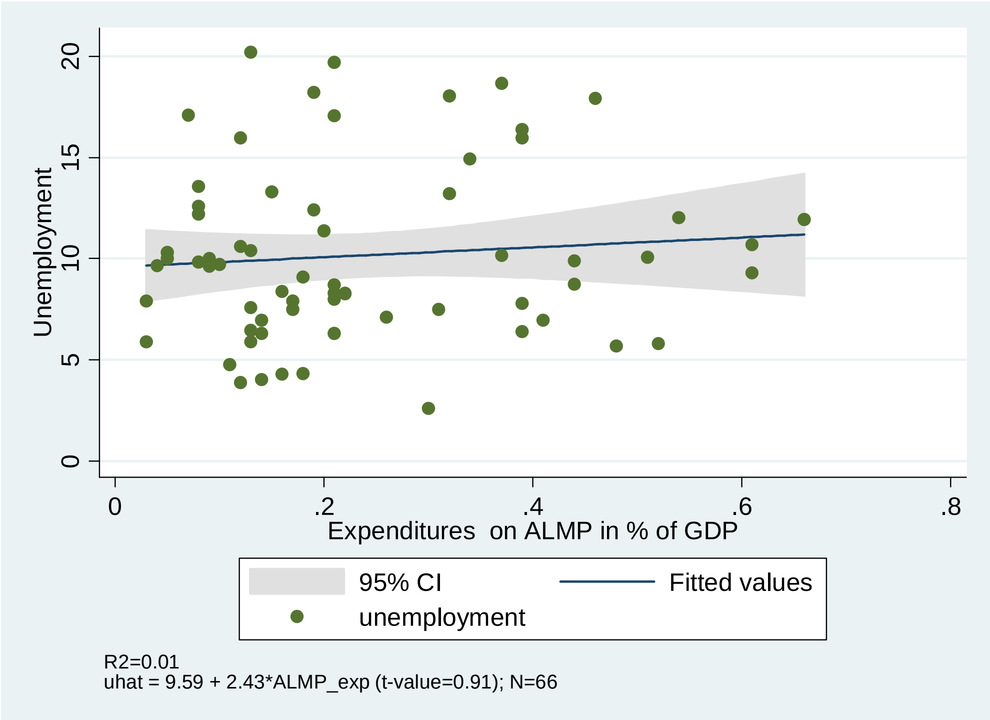
Source: the data for Czech Republic, Hungary, Poland and Slovakia are from OECD Employment Outlook various issues; data for Estonia till 2001 are from Leetmaa and Võrk (2004) and from Estonian Labour Market Board for 2002-2006; data for Latvia, Lithuania, Bulgaria and Romania are from the Eurostat Yearbook various issues (categories 1+2 to 7); data for Bulgaria for 1998 are from Betcherman et al. (2004). Notes: Average across 1992-2006 is presented. In Eurostat data category 1 is missing from 1998 to 2003, and the measure includes categories 2-7 till 2003, and categories 1-7 for 2004-2005. Data for Czech Republic for 2003-2006 are from Eurostat Yearbook (2007) and includes categories 1-7; data for Hungary in 2006 includes categories 1-7, and in 2004 includes categories 2-7; data for Poland for 2002-2004 includes categories 2-7; data for Slovakia in 2006 includes categories 1-7, and in 2004 includes categories 2-7. Data for Hungary in 2003 and 2005, for Poland in 2005, 2006, 1997-2001 and 1992, for Slovakia in 2003, 2005, 1992-1998, for Estonia in 1992-1993, for Latvia, Lithuania and Romania in 2006 and 1992-2002, and for Bulgaria in 2006, 1999-2003 and in 1992-1997 are missing.

Figure 3. Relation of unemployment rate and ALMP expenditures – EU 15



Source: data for unemployment rates are from the Eurostat online database. For data sources on ALMP see notes to Figures 1 and 2. Notes: Harmonized unemployment rates, -/+ 25 years, yearly averages. Years: 1991-2005 when available. Data for Austria in 1991, 1992 and Germany in 1991 are missing.

Figure 4. Relation of unemployment rate and ALMP expenditures – new EU member states



Source: data for unemployment rates are from the EBRD Transition Report (2007); for data sources on ALMP see notes to Figures 1 and 2. Note: Years: 1991-2006 when available.

Figure 5 Distribution of propensity score and common support

(source: Bonin and Rinne (2006))

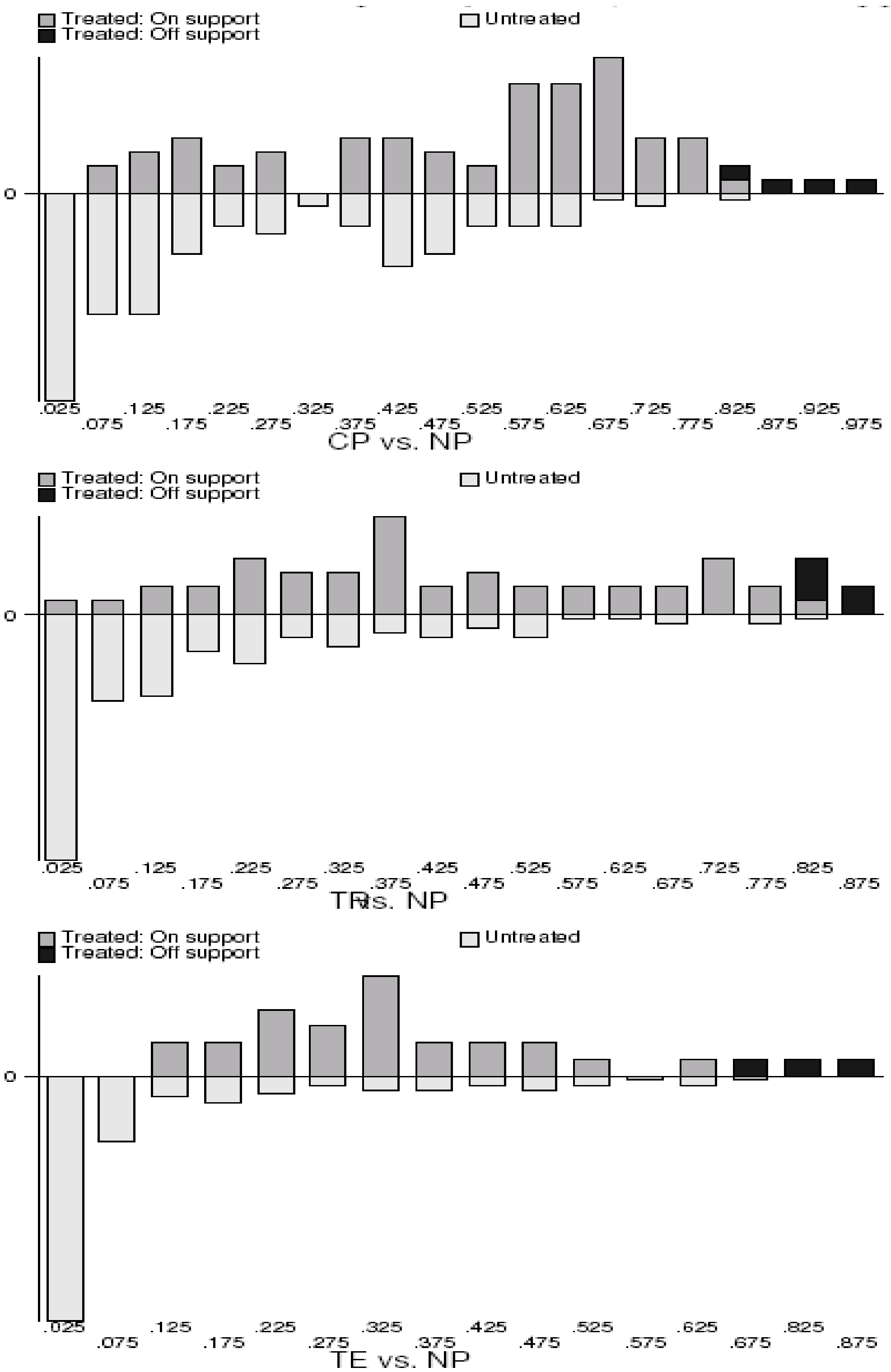
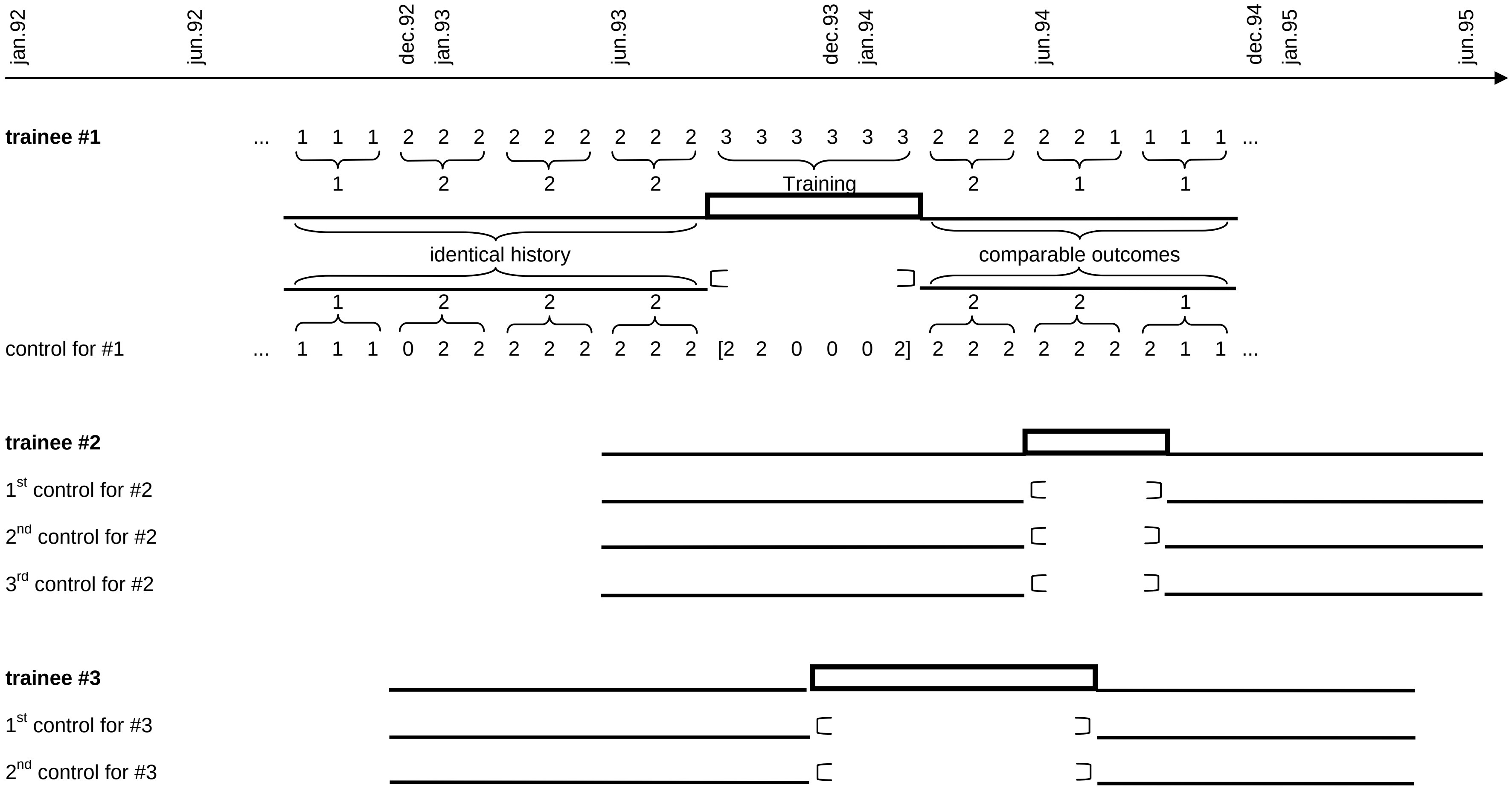


Figure 6 Matching over identical individual labor market histories applying a "moving window"



Source: Adapted from Kluve et al. (2008).

Table 1. Active labor market policies in OECD countries: archetypical types of programs and generic purpose

Type of program	Generic purpose
a. Public employment services ("job brokerage") and administration	Improve matching efficiency
b. Labor market training	Attenuate skill mismatch; human capital accumulation
c. Employment incentives / Start-up incentives	Improve job matching process; increase labor demand
d. Direct job creation / Public sector employment	Increase labor demand; prevent human capital deterioration
e. Youth measures (training and/or subsidized jobs)	See b, c and d.
f. Measures for the disabled	Integrate discriminated persons into the labor market

Note: This classification is based on the commonly used categories that can be found e.g. in Eurostat (2008), OECD (2006), Kluve (2006).

Table 2. Average post-treatment employment rates – Training

	treated units	comparison units	effect	std.error
Sample (A)	114	983	-.048	.049
Sample (B)	87	111	.138	.059
Sample (B) stratified by labor force status history:				
"1111"	24	34	.071	.115
"2222"	32	43	-.077	.103
Sample (B) stratified by program entry date				
Before Jan 1, 1995	55	73	.152	.078
After Jan 1, 1995	32	38	.122	.110
Raw effect (i): No covariates	121	6751	-.027	.046
Raw effect (ii): No moving window	121	6309	-.040	.045

Source: Kluve et al. (2008)

Table 3. Average post-treatment employment rates – Intervention Works

	treated units	comparison units	effect	std.error
Sample (A)	244	1354	-.291	.031
Sample (B)	212	240	-.126	.040
Sample (B) stratified by labor force status history:				
"1111"	16	19	.084	.148
"2222"	168	191	-.150	.045
Raw effect (i): No covariates	275	6757	-.285	.026
Raw effect (ii): No moving window	275	6322	-.312	.030
Additional covariate: benefit receipt				
Sample (A)	242	1152	-.208	.033
Sample (B)	149	243	-.147	.037

Source: Kluve et al. (2008)