A Statistical Programme Assignment Model

Jonas Staghøj
Michael Svarer
Michael Rosholm

DISCUSSION PAPER SERIES

I Z A

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

A Statistical Programme Assignment Model

Jonas Staghøj
University of Aarhus

Michael Svarer
University of Aarhus,
CAM and IZA

Michael Rosholm
University of Aarhus,
AKF and IZA

ABSTRACT

A Statistical Programme Assignment Model[*]

When treatment effects of active labour market programmes are heterogeneous in an observable way across the population, the allocation of the unemployed into different programmes becomes a particularly important issue. In this paper, we present a statistical model designed to improve the present assignment mechanism, which is based on the discretionary choice of case workers. This is done in a duration model context, using the timing-of-events framework to identify causal effects. We compare different assignment mechanisms, and the results suggest that a significant reduction in the average duration of unemployment spells may result if a statistical programme assignment model is introduced. We discuss several issues regarding the implementation of such a system, especially the interplay between the statistical model and case workers.

JEL Classification: J64, J68

Keywords: profiling, targeting, statistical treatment rules, heterogeneous effects

Corresponding author:

Jonas Staghøj
Department of Economics
University of Aarhus
8000 Aarhus C
Denmark
E-mail: jstaghoej@econ.au.dk

1 Introduction

In this paper, we develop a statistical model for estimating heterogeneous effects of active labour market programmes, where the heterogeneity is in terms of observable characteristics of the unemployed worker. The purpose of estimating heterogeneous programme effects is to improve the assignment mechanism of unemployed workers into different active labour market programmes (henceforth ALMPs).

The main motivation for using a statistical model is the possibility to include literally hundreds of thousands of observations in performing inference on previously treated unemployed, and then to use this information in an attempt to predict the future. Even experienced case workers will only meet a very limited number of unemployed, and the case worker may not be able to follow the unemployed over a sufficiently long time period to actually observe the outcome after the programme is completed.

During the last decade the use of ALMPs has increased in many countries, and many politicians see them as a promising way of decreasing unemployment. In Denmark, ALMPs constitute an essential element in the so-called flexicurity system of the labour market. The flexicurity system consists of flexible employment relations in terms of low hiring and firing costs, a generous income replacement scheme in case of job loss, and an active labour market policy. The active labour market policy is essential in two ways. First, it ensures that workers who become unemployed can obtain the qualifications necessary for finding new employment. Secondly, and perhaps more importantly, the active labour market policy tests the availability of unemployed workers and provides an incentive to conduct active job search[1].

In the flexicurity model, where ALMPs are used on a large scale, and where the costs are far from negligible on the government budget (costs of active labour market policies constitute 1.5-2% of GDP, see e.g. various issues of the OECD Employment Outlook), it is essential that they are used effectively. However, evaluation studies conducted across the world generally find that the effects of these programmes are modest and sometimes even negative.[2] One reason could be that ALMPs are not used in an optimal way; i.e. if the effects of ALMPs are heterogeneous across the population, it is possible that the unemployed are allocated to the wrong programmes.

In this paper, we analyze different assignments of unemployed to ALMPs, and the

---

[1]For a more elaborate presentation of the Danish Flexicurity model see Andersen & Svarer (2007).

[2]See e.g. Heckman et al. (1999) and Kluve (2006) for reviews of the evaluations of ALMPs.

results show that the choice of assignment mechanism potentially has a considerable impact on the expected average duration of unemployment. The assignment of unemployed workers to different programmes often involves a large degree of discretion to the case worker. Sometimes deterministic rules state that programmes should be offered to particular groups of unemployed, but the choice of a particular programme is usually made by the case worker (and the unemployed). We argue that this may not necessarily be an optimal way of programme assignment. Instead we propose a statistical programme assignment model, which uses information on past programme effects to suggest how to assign the unemployed to programmes. We estimate a multivariate duration model and use the timing-of-events model (Abbring & van den Berg, 2003) to estimate causal effects by taking selection based on observed and unobserved heterogeneity into account.

Related models are being developed and tested in e.g. Switzerland (SAPS) and Germany (TrEffeR)[3]. Our model differs from these in relation to the econometric strategy. Where SAPS and TrEffeR are based on matching estimators, we use duration analysis. Our choice of model is based on the explicit policy goal stated by the Danish Ministry of Employment, which is to bring the unemployed back to work as quickly as possible.

The paper is organized as follows. In section 2, we first provide a framework for analyzing statistical programme assignment models, and we then discuss the issue of case workers' discretion versus statistical programme assignment to set the stage for the rest of the paper. In section 3, we describe the econometric model, and section 4 contains a description of the data. Section 5 presents estimation results, discusses different ways of presenting the potential outcomes to case workers, and assesses the potential gains from implementing a statistical programme assignment model. Finally, section 6 contains the conclusion and some considerations for future work.

2 A statistical programme assignment model - concepts and considerations

The first subsection discusses a framework for statistical programme assignment models, and the second subsection discusses the interplay between the statistical model and the

---

[3]See Frölich (2006) or Behncke, Frölich & Lechner (2006, 2007) for more details about the SAPS (Statistically Assisted Programme Selection) model and Stephan, Rässler & Schewe (2006) for information about TrEffeR (Treatment Effects and Prediction). For more general description on statistical treatment rules see e.g. Dehejia (2005) and Eberts et al (2002).

case worker.

2.1 Framework for analyzing statistical programme assignment models

The potential scope for statistical assignment models is very broad and includes applications in e.g. finance, medicine, insurance, criminology, marketing and data mining.[4] The main purpose of using a statistical assignment model is to reveal systematic relationships between some observed variables, X, and an outcome variable, Y, and then use this information to provide better assignments of some treatment.[5] Denoting by $\{0, 1, ..., R\}$ the set of possible programmes, we are looking for a mapping of characteristics and outcomes into the set of programmes

$$\{X, Y\} \curvearrowright \{0, 1, ..., R\}$$

In this paper, we will analyze the assignment of ALMPs to unemployed, using a statistical programme assignment model. This is an interesting problem because resource constraints imply the need for a precise targeting of expensive programmes.

Ideally, we would specify a welfare function for the society and let this function guide the implementation of the system. The programmes should then be allocated in a way that maximizes the welfare function. Although such an approach may not be very useful for practical purposes, it can help by structuring our thoughts about how to define specific goals. Ultimately, the maximization of many types of welfare functions can be boiled down to maximizing equity and/or efficiency. Increasing equity should only be the goal if it is the most effective way of increasing the welfare function. A maximin welfare function is an example of a welfare function leading to this choice of goal. Similarly, if improving efficiency is the goal, then the reasoning behind this should be that an increase in the welfare function is most effectively achieved by improving efficiency. In this paper, we consider a goal of maximizing efficiency in the sense of offering programmes to those for whom the expected gain is largest.

---

[4] For applications of profiling in other literature areas see e.g., Auerhahn (1999) for an application in criminology and Gottfredson & Moriarty (2006) for a recent overview, Yeo et al. (2001) for an application about insurance, Shaw et al. (2001) for an application in marketing, and Khan et al. (2001) or Murphy (2005) for applications in medicine.

[5] See also Black, Berger & Smith (2001) for a similar discussion of the theoretical issues concerning the design of statistical assignment mechanisms.

Having defined the goals of the system, we move on to discuss its implementation. The first objective is to define the outcome variable, Y. The stated goals for the system should guide the choice of outcome variable, but even with clearly defined goals, the definition of the outcome variable(s) is not given. The choice of outcome variable may also depend on the available X variables, or rather, on how precisely we are able to relate X to Y. Y is typically a scalar, but it could in principle be an index measuring different aspects related to the goals. In the case of ALMPs, possible Y variables include duration of unemployment, duration of subsequent jobs, average time spent in unemployment over some time period, the wage in the next job or discounted future earnings.

With efficiency being the stated goal in this paper, a measure like the net monetary benefit of participating in a programme would be a good variable to use for the statistical programme assignment model. We do, however, choose the duration of unemployment as the relevant variable and the motivation for this is threefold. Firstly, the duration of unemployment is actually the main focus of the Danish Ministry of Employment. Hence, using this variable will make policy recommendations immediately applicable. Secondly, there is a trade-off between using ideal long-term outcome measures and more readily available short-term outcomes (which are hopefully correlated with long-term outcomes). Shortening the duration of unemployment is at least one way of increasing earnings. The third reason for using duration of unemployment as the outcome variable is that we do not have access to data on the costs of the programmes. We will have to keep this in mind when interpreting the results of the model and notice that we will now be maximizing the effectiveness of the system rather than its efficiency.

2.2 Statistical model vs. case worker discretion

As mentioned in the introduction, current assignment models typically consists of a combination of some initial deterministic screening mechanism (e.g. target groups for ALMPs) and a large degree of discretionary power to case workers. In this section, we discuss the potential contribution of a statistical programme assignment model and its implementation.

The main contribution from a statistical programme assignment model is the possibility to include literally hundreds of thousands of observations in performing inference on previously treated unemployed, and then use this information in an attempt to predict the future. Even experienced case workers will only meet a very limited number of unem-

ployed, and the case worker may not be able to follow the unemployed over a sufficiently long time period to observe the outcome after the programme is completed. Also, it may be impossible for case workers to realize if there actually is an effect from a very small sample of unemployed; if for example the effect is a 10 % increase in an exit rate from unemployment if the exit rate is 3% per week that hence increases to 3.3%. This would be nearly impossible to see without statistical processing of the data.

The first question to address is whether statistical assignment should be seen as an alternative to case workers or as a tool offered to case workers? If implemented as an alternative to case workers, the system would ensure equal treatment of similar people, which may be an important criteria, and it would of course offer considerable scope for cost savings since no case worker salaries would have to be paid.[6] However, in this paper we argue that using the model to equip case workers with additional information is a better idea. Obviously, case workers perform many tasks in addition to assigning the unemployed to ALMPs. Such other tasks include assisting and advising in job search, motivating and monitoring the unemployed etc.. A statistical assignment model would not render case workers redundant, but rather give them the possibility to focus more resources on other tasks. Since the institutions taking care of the assignment of the programmes are often publicly funded, this way of implementing the model might also be seen as a an attempt to compensate for the lack of a natural pricing-mechanism as the model advises the case workers on which programmes are the most valuable for the different types of unemployed.

The statistical assignment model systematically analyzes information for many individuals and identifies past average effects of different programmes for different types of individuals, but the case worker may observe some information about the specific individual, which the statistical model cannot take into account. Motivation or ability are typical examples. These unobserved variables are important to take into account, and if they are sufficiently important for estimating the true relationships between X and Y variables, then case workers may actually do a better job than the statistical model. Lechner & Smith (2007), however, provide evidence suggesting that case workers are not very good at predicting effects, or interpreting the results differently, they are not seeking to maximize the expected outcomes. They show that case workers do no better than a

---

[6]We do not attempt to say more about fairness in this paper, since the focus is on efficiency. There is a related economic literature on racial profiling where fairness issues are discussed more thoroughly see e.g. Persico (2002), Persico & Todd (2005), and Harcourt (2005).

random assignment mechanism. Similar results are found in Bell & Orr (2002) where case workers are also shown to have difficulties in identifying those with the largest gains. In this paper, we also find that case workers do not act in accordance with the statistical predictions, and in terms of the models' predictions they actually perform worse than a random assignment mechanism.

When considering a combination of a statistical assignment model and case worker discretion, we need to think about the optimal way of implementation. A good statistical model provides case workers with new and relevant information, so the important ability of the system becomes the partial degree of explanation it can bring. A good case worker identifies unobservable characteristics and uses these to modify the recommendations of the statistical model. Thus, the case workers could sort the unemployed into groups of somewhat homogeneous individuals in terms of unobservables; for example by identifying the group of individuals who have problems beside unemployment (e.g. psychiatric, health or abuse problems). The statistical predictions would probably not be very accurate for a person with an alcoholic abuse or similar problems not observed in the data, but for relatively mainstream unemployed the system may do a good job in assisting the case workers.

An obvious, and perfectly valuable, objection often mentioned as an argument against a statistical programme assignment model is that it cannot estimate the outcomes with sufficient precision. However, case workers also make mistakes, and providing case workers with additional information (*not* replacing them with a statistical model) should lead to more precision. This objection may be one possible explanation of the latest results from the randomized experiment of implementing the SAPS model in Switzerland (Behncke, Frölich & Lechner (2007)). The main result from this experiment is that case workers do not voluntarily use the information from a statistical model, and the authors hence conclude that stronger incentives are needed for case workers to comply with information from statistical models. Adequate incentive schemes should hence be carefully designed in order to find the optimal implementation of the statistical model somewhere in between completely voluntary and completely mandatory.

3 Econometric Model

We estimate a model intended to maximize the effectiveness of programmes in terms of their ability to decrease the duration of unemployment. To operationalize this, we

predict the future outcome for each individual, conditional on participation in any of the possible programmes. Subsequently, the difference in the predicted outcomes compared to the option of *no programme* is calculated, and the programme with the best effect (*no training* if all impacts are negative) is chosen.

For each individual, we define the potential outcomes[7] as

$$Y_0, Y_1, ..., Y_R$$

where $\{0, 1, ..., R\}$ is the set of possible programmes, denoting 0 as the outcome in the case of *no programme*. By constructing these hypothetical outcomes, we are able to analyze various features of the distribution of conditional outcomes, and hence also the expected impact compared to *no programme*. The fundamental evaluation problem is that for each individual we observe at most one of the potential outcomes. We need additional identifying assumptions to estimate all other potential outcomes, and since the assignment of unemployed into programmes is not random, we have to distinguish the causal programme effects from selection effects.

One possible identifying assumption used when estimating the Swiss SAPS or German TrEffeR model is the Conditional Independence Assumption (CIA), which can be stated as

$$Y_0, Y_1, ..., Y_R \amalg D | X \quad \forall x \in \chi$$

where $D \in \{0, 1, ..., R\}$ is a dummy variable indicating which programme the individual is assigned to, and χ is the relevant set of characteristics. With this assumption, we can estimate the counterfactuals conditional on X. If the assignment of programmes is random, this assumption is clearly fulfilled (when the support conditions are satisfied). Proper use of this identifying assumption will typically require a very rich data set containing detailed data on individual characteristics as well as market specific informations. For the present analysis, we actually have access to quite detailed data, so it might be reasonable to assume that the CIA is fulfilled and proceed by employing a matching estimator for duration outcomes.

However, if there exists some unobserved variables which influences the selection process as well as the potential outcomes, the CIA approach will result in biased estimates. Moreover, in the Danish Flexicurity model, where programme participation becomes mandatory after some time, the construction of a matched comparison group

---

[7]See Rubin (1974).

for those participating in a programme after say 12 months of unemployment, becomes suspiciously close to selection on the dependent variable. Those who do not participate in a programme in the current unemployment spell, but have more than 12 months of unemployment, are likely to a) have found employment a short time after the 12 months (otherwise they would have ended up in a programme), or b) have unfavorable unobserved characteristics (rendering them unfit for programme participation). Hence, a model adequate for describing the dynamic process of selection into programmes (e.g. competing risks duration models for entry into different types of programmes) as well as dynamic sample selection (through explicit modelling of the exit rate from unemployment) is preferred in the Danish case. As described in the following section, we therefore use the timing-of-events model developed by Abbring & van den Berg (2003) to estimate the potential outcomes. This allows for unobserved heterogeneity, and corrects automatically for dynamic selection bias.

Finally, we assume that the Stable Unit Treatment Value Assumption (SUTVA) is fulfilled, which means that the potential outcomes for each individual does not depend on the treatment of other individuals.[8] This implies that we ignore possible general equilibrium effects and restrict ourselves to a partial equilibrium analysis. Of course, this assumption is particularly questionable in the Danish case of large scale mandatory programmes, but this issue is left for future research.

3.1 The Timing-of-Events Duration Model

The stochastic variable of interest is the duration of unemployment, $T_u \in (0, \infty)$. To model the selection process into programmes, we define another stochastic variable, $T_p \in (0, \infty)$, as the duration until assignment to ALMP. If $T_p < T_u$, the individual is assigned to an ALMP, and if $T_p = T_u$ then T_p is right censored and the individual has not participated in ALMP before leaving unemployment. T_p is the minimum of four latent durations until assignment into one of the 4 types of programmes, $T_{p_1}, T_{p_2}, T_{p_3}, T_{p_4}$.[9] We only evaluate each individual's first ALMP during an unemployment spell, so if we observe the individual in a second programme, the unemployment spell is treated as right-censored at the time of the start of the second programme.

The central concept in duration analysis is the hazard functions, so to estimate the

---

[8]Rubin (1980).

[9]The types of programmes are described in the data section.

model we have to specify these. We model the hazard functions as Mixed Proportional Hazards, which means that the hazards are modelled as the product of a baseline hazard, $\lambda(t)$, depending on time, and a scaling function, $\phi(x_t, v)$, depending on the observed characteristics, x_t, and possibly also some unobserved component, ν,[10] that is,

$$\theta(t|x, v) = \lambda(t) \cdot \phi(x_t, v)$$

In the data, we observe when an individual is assigned to one of the 4 types of ALMPs, and using this information, we construct two sets of time-varying indicator variables; one set of indicators for participating in a programme of type j, $d^1_{j,t}$, and another set of indicators for having completed programme j, $d^2_{j,t}$. We chose the functional form $\phi(x_t, v) = \exp(x_t\beta + v_u)$ for the scaling function and write the hazard function out of unemployment as

$$\theta_u[t_u|x_t, d^1_{j,t}, d^2_{j,t}v_u] = \lambda_u(t) \exp[x_t\beta_u + \sum_{j=1}^{4} \left(d^1_{j,t}\gamma^1_j + d^2_{j,t}\gamma^2_j\right) + v_u] \tag{1}$$

FIGURE 1: LOCKING-IN AND POST PROGRAMME EFFECTS

---
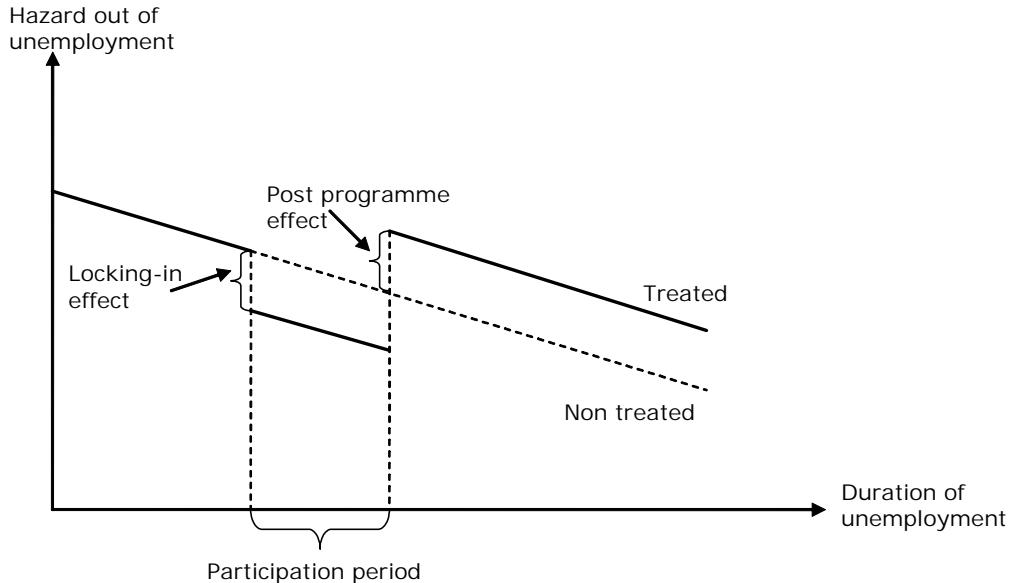
[10] As a result of the time-varying covariates, this scaling function is actually time-varying, but we use the conventional notation and denote it as a scaling function.

As illustrated in Figure 1, the parameter γ_j^1 captures the effect on the hazard, while participating in programme j and γ_j^2 captures the effect of having finished programme j. Both effects influence the hazard rate, but at different time intervals. We would typically expect γ_j^1 to be negative if there is a 'locking-in effect' while participating, since less time is available for searching for a job. A positive γ_j^2 indicates that the hazard rate increases after completion of the programme.

It is not necessarily clear that a negative locking-in effect is a bad thing. If programmes are designed to enhance valuable skills for the unemployed, it may be optimal for the unemployed to complete the programme. However, in this case the post programme effects should be large. Later, we discuss how to obtain the combined effect by computing the expected duration of unemployment.

We ignore any effects of ALMPs that might kick in before programmes start, i.e. threat effects (see e.g., Black, Smith, Berger & Noel (2003) and Rosholm & Svarer (2004)). These *ex-ante* effects are of course important in terms of evaluation of the gross effects of ALMPs, but not necessarily so important when it comes to assigning individuals to specific programmes. Nevertheless, the *no training* option will be affected by the threat effect. For a detailed discussion, see Rosholm & Svarer (2004). This potential bias is ignored in the following.

We specify the baseline hazards as $\lambda(t) = \exp(\lambda_m(t))$ where m refers to a specific interval on the timeline. We divide the timeline into $M = 12$ intervals measured in weeks. The hazard into programmes is defined as the sum of all 4 programme hazards,

$$\theta_p(t_p|x_t, v_p) = \sum_{j=1}^{4} \theta_{p_j}(t_{p_j}|x_t, v_{p_j}) \tag{2}$$

We allow for unobserved heterogeneity by introducing the stochastic variables V_u, and $V_p = (V_{p_1}, V_{p_2}, V_{p_3}, V_{p_4})$ as unobserved variables, which are allowed to have direct effects on the hazards into employment and programmes.[11] We restrict the distributions of the unobserved variables to be discrete with two mass-points, and allow the distributions of V_u and V_p to be correlated. This way of introducing unobserved heterogeneity is based on Heckman & Singer (1984). Van den Berg (2001) writes that "... a consensus has emerged that multi-spell data allow for reliable inference that is robust with respect to

---

[11]It is well known that problems with unobserved heterogeneity are particularly important to handle in duration models. In contrast to usual regression models, even unobserved heterogeneity which is uncorrelated with the covariates implies biased results in duration models if not taken properly into account.

the specification of the unobserved heterogeneity distribution."[12] The reason for this is, that with multi-spell data the identification does not lie completely on the proportionality assumption when we assume the unobserved heterogeneity term to be constant over time for each individual. Since we observe multiple spells for many individuals in our event history data, we argue that the discrete distribution for unobserved heterogeneity is applicable in this context.

The unobservable components in the 4 different programme hazards are assumed to be perfectly correlated. This assumption can easily be relaxed to allow for different selection on unobservables into the different programmes, but the perfect correlation restriction simplifies the estimation process[13]. The correlation between V_u, and V_p is important, because this is the way this procedure allows for selection on unobservables without a resulting bias in the estimated effects. Therefore, this correlation is allowed to vary freely in the estimation process.

Defining a non-censoring indicator, C_i, equal to 1 if the duration spell of unemployment for individual i is uncensored, we can construct the likelihood function for individual i with K unemployment spells as

$$
\begin{aligned}
L_i(v_u, v_p) &= \prod_{k=1}^{K} \theta_p[t_{pk}|x_{tk}, v_p]^{1[t_{pk}<t_{uk}]} \cdot \theta_u[t_u|x_t, d^1_{j,t_u}, d^2_{j,t_u} v_u]^{C_i} \\
&\quad \cdot \exp\left[-\int_0^{t_{pk}} \theta_p[s|x_s, v_p]ds - \int_0^{t_{uk}} \theta_u[t|x_t, d^1_{j,t}, d^2_{j,t} v_u]dt\right]
\end{aligned}
\tag{3}
$$

Summing over the support of the discrete distributions for the unobservables, and summing over the sample of individuals, we can construct the complete likelihood function as

$$
\prod_{i=1}^{N} \sum_{m=1}^{4} P_m \cdot L_i(v_u, v_{p_j})
\tag{4}
$$

where P_m are the associated probabilities of the mass-points. The parameters to be estimated are $\Psi = \{(\beta_u, \lambda_u), (\beta_j, \lambda_j), (v_u, v_{p_j}, P_m)\}, \; j, m = 1, 2, 3, 4$.

---

[12]Van den Berg (2001) suggests ways to use economic restrictions to provide a more robust analysis, and Gaure, Røed & Zhang (2007) show in a Monte Carlo analysis that the method proposed by Heckman & Singer (1984) is quite precise and robust.

[13]In terms of practical implementation it is often quite difficult to estimate a completely unrestricted correlation structure. Several of the mass points collapse to the same values, and several of the associated probabilities converge to zero. We therefore choose to present the perfect correlation structure, but note that in an actual policy implementation, a more thorough estimation process would be recommendable.

3.2 Estimation of heterogeneous treatment effects

We estimate heterogeneous effects by allowing the effects to depend on the observable characteristics, and we assume that all heterogeneity is captured in this way. Conditional on observables, the effects are assumed homogeneous, and hence we do not have to distinguish between the average treatment effect on the treated (ATET) and the average treatment effect (ATE) as long as we condition on the observable characteristics.[14] This is convenient since we are estimating ATET, but we are actually more interested in ATE, because we want to predict all potential outcomes for every unemployed individual. In principle, effects can be allowed to be heterogeneous after conditioning on observable characteristics if individuals either do not know their own heterogeneous effects or do not act upon this information. However, since these assumptions appear to be quite restrictive, we maintain the assumption of homogeneous effects conditional on observable characteristics.[15]

To estimate heterogeneous effects of the programmes, we augment the set of characteristics by including interaction terms between a subset of the characteristics, x_t^s, and the programme indicators, $d_{j,t}^1$ and $d_{j,t}^2$. This implies that the effect of a particular programme is allowed to depend on these characteristics. Apart from a larger set of parameters, the estimation procedure is as before, and the hazard function out of unemployment can be written as

$$\theta_u[t_u|x_t, v_u] = \lambda_u(t) \exp\left(x_t\beta_u + \sum_{j=1}^{4} \left[\left(\begin{array}{cc} 1 & x_t^s \end{array} \right) \cdot d_{j,t}^1 \cdot \gamma_j^1 + \left(\begin{array}{cc} 1 & x_t^s \end{array} \right) \cdot d_{j,t}^2 \cdot \gamma_j^2 \right] + v_u \right)$$

(5)

where, if x_t^s is a $1 \times K$ vector of characteristics, $\left(\begin{array}{cc} 1 & x_t^s \end{array} \right) \cdot d_{j,t}^1$ is a $1 \times (K+1)$ vector, and γ_j^1 is now a $(K+1) \times 1$ parameter vector, and similarly for γ_j^2.

3.3 Identification

The exact identifying assumptions of the timing-of-events model are stated in Abbring & van den Berg (2003), and the two main assumptions are 1) *proportional hazards*[16] and

---

[14] See Heckman et al. (1999) for a detailed discussion.

[15] For more on heterogeneous treatment effects, see for example Heckman, Urzua & Vytlacil (2006) or Heckman, Smith & Clemens (1997).

[16] As already noted the presence of multiple spells decreases the dependence on this part of the assumption. Futhermore, Brinch (2007) has recently shown that the proportionality assumption is in fact not needed for identification provided that we observe a sufficient amount of variation in covariates over

2) *no anticipation.* The proportionality assumption is needed in order to identify the unobserved heterogeneity terms. To see how this works, note that the distribution of unobservables will change over time because those unemployed with higher values of the unobserved variables will leave unemployment at a faster rate. Next, when the unobserved heterogeneity enters the hazard rate as a multiplicative term, it has a larger absolute effect on those individuals with a higher value of the rest of the scaling function, $\exp(x_t\beta)$. Combining these observations, this means that the distribution of unobservables among the individuals with high $x\beta$ will change more rapidly over time than for those individuals with low $x\beta$. However, this introduces an apparent nonproportionality since the time profile supposed to be captured by the baseline hazard $\lambda(t)$ now looks different for different kinds of $x\beta$ individuals. That is, we observe an apparent interaction between the scaling function, $\exp(x_t\beta)$, and the baseline hazard $\lambda(t)$, and since we have assumed the hazard to be proportional in these terms, the only way we can capture the observed nonproportionality is by introducing the unobserved variables, and this is how the distribution of these is identified. The *no anticipation* assumption states that individuals are not allowed to know in advance the precise time at which they will be assigned to a programme. If they had this information, they would be able to adjust their behavior even before actually starting in the programme, which would invalidate our identification strategy. They are, however, allowed to know the distribution of time until assignment to treatment. We argue that the *no anticipation* assumption is reasonable in this analysis because programme participation is typically not planned a long time before actually taking place (typically only a few weeks before).

The intuition behind the identifying strategy in the timing-of-events approach is to use exogenous variation in the time until the unemployed are assigned into programmes. This strategy is well-suited for an evaluation of ALMPs in Denmark, because we observe a lot of variation in the time until individuals are assigned to a programme. Some unemployed are assigned into a programme very early in their unemployment spell, and if there is exogenous variation in the timing of the assignments, we can use similar unemployed, not yet assigned to a programme, as the relevant counterfactuals. We model the selection processes into programmes conditional on observed and unobserved variables, and argue that some of the remaining variation is exogenous. This exogenous variation may emerge for several reasons. The most important reasons are supply constraints at the local unemployment office influencing amount of time before an unemployed can be sent

---

time and across observations.

into a programme without directly affecting the exit rate out of unemployment, and the fact that assignment into ALMPs is often planned at meetings between the case worker and the unemployed, and these meetings occur with some exogenous variation for several practical reasons.

A possible problem with the timing-of-events approach in this context could be, that unemployed in Denmark are required to participate in some kind of ALMP after a certain period of unemployment. After one year of what we label as open unemployment, the unemployed enter the so-called active period in which they must participate in a programme for 75% of the time. But apparently this is not enforced too strictly by the labour market authorities, so even for longer unemployment durations, we still have some individuals which can be used as counterfactuals. In this case where the dynamic selection process is explicitly modelled and allowed to depend on unobservables, this is not a problem.

4 Data

4.1 Institutional settings in the Danish labour market

Denmark has a two-tiered system for unemployed workers. Most workers in Denmark - around 80% - are members of an unemployment insurance (UI) fund. These individuals have - upon the fulfilment of a few conditions - the right to receive UI benefits corresponding to 90% of the previous wage with an upper limit of approximately 1800 Euro per month. UI benefit payments are heavily subsidized by the state, which finances around 70% of total payments.

Unemployed workers without UI benefit eligibility may instead receive social assistance (SA) benefits. While non-insured workers only make up around 20% of the workforce, they make up a much larger fraction of the unemployed as the group typically consists of workers with a low attachment to the labour market. Hence, they are more often unemployed, and on average they are unemployed for longer periods. Social assistance benefits are means tested, but the amount is typically below the UI benefit level. Social assistance is administered by the municipal authorities.

Unemployed workers receiving any of the two types of benefits are required to search for a job and to be fully available for the labour market. If they do not fulfil these requirements, they run the risk of being sanctioned, so they cannot immediately turn down job offers or programme assignments made by the authorities. In the observation

period which we consider, i.e. 1998-2003, those who stay unemployed for more than a year enters what is called the 'active period'. Before that period, they may also participate in programmes, and many do so, but naturally, the scale is smaller.

4.2   The sample and the variables

The data we use is a 10% sample of the adult population in Denmark followed in the period 1988-2003, and it is constructed by merging information from several Danish administrative registers maintained by Statistics Denmark. We can follow each individual on a weekly basis and observe transitions between different labour market states. This individual labour market history is then combined with socioeconomic variables in order to get a very detailed data set suited for the analysis.

We define the population of interest as insured unemployed men aged 25-55, entering unemployment in the period from 1998 to 2003. This subsample is chosen because there are special rules related to individuals below 25 as well as those above 55. The time window is chosen to get a period without major reforms in the regulation of the labour market. We distinguish temporary unemployment from ordinary unemployment by defining temporary unemployment as unemployment spells where the individual returns to a job within the same firm. This is done because former analyses have shown that temporary unemployment spells are of a different nature than ordinary unemployment.[17] We only look at ordinary unemployment spells.

In the following section, we provide a brief description of the data used in the analysis. The dependent variable is unemployment duration measured in weeks.

The explanatory variables are the following: We include YEAR dummies to capture major business cycle and other calendar time effects. To capture local labour market effects, we include the LOCAL UNEMPLOYMENT RATE at the county level.[18] A dummy for each COUNTY is also used.

Individual background variables: We use a dummy for having CHILDREN as well as a dummy for having CHILDREN LESS THAN 7 YEARS OLD and for being MARRIED. Dummies for different AGE groups are included: 25-29, 30-39, 40-49 (reference), and above 50. A dummy for being member of a UNION is included, and dummies for membership of different types of UI FUNDS are included. The UI funds are catego-

---

[17]See Jensen & Svarer (2003).

[18]In the considered time period, Denmark consisted of 14 counties and we included a dummy for the municipality of Copenhagen as well.

rized into 9 different groups: Building, Production, Technology, Trade, Service, Academic, Others and Self-employed. The reference group is Metal. We also add a dummy for the individual coming from a NON-OECD country. Years of working EXPERIENCE in the labour market is included along with the hourly WAGE in the last observed job. Also, from this wage we construct the UI REPLACEMENT RATE. Considering education, we include dummies for NO FURTHER EDUCATION, FURTHER EDUCATION, and the reference group is VOCATIONAL EDUCATION. We have constructed a variable with the number of WEEKS REMAINING IN OPEN UNEMPLOYMENT to account for the fact that after the period of open unemployment, the unemployed enters the so-called active period and is more likely to be assigned to ALMP. Former studies have shown an increase in the hazard out of unemployment in the weeks preceding this transition into a new period (Geerdsen, 2006).

ALMP variables:

1. PRIVATE JOB TRAINING: The individual is employed in a temporary job in the private sector where the employer receives a subsidy. The duration of these programmes is typically 6-9 months.

2. PUBLIC JOB TRAINING: The individual is employed in a temporary job in the public sector and the duration of these programmes is typically 6-12 months.

3. CLASSROOM TRAINING: The individual participates in some kind of classroom training which includes short courses as well as ordinary education. It typically lasts only a few months.

4. OTHER TRAINING: A somewhat mixed group of all programmes which cannot be categorized into the other categories. Typical programmes in this group includes job search assistance, programmes of competence detection, individualized job training etc.

These quite detailed variables concerning the history of the individuals and the state of local labour markets, should allow us to identify the selection process into programmes and important heterogeneity in the impacts of the programmes. But just as important as the variables we are able to include in the analysis are the variables we may be missing. First of all, we only have access to very broad measures of the different types of ALMP. Although there are trade-offs when deciding on how finely to define different groups, the

4 groups we have are almost certainly too broad. For the moment, however, it is all we have. Variables like motivation and ability are unobserved to the econometrician, but we might be able to improve the quality of the data by including information supplied by the case workers. This is done in the SAPS project in Switzerland, where the statistical model incorporates several variables with case workers' ratings of different skills of the unemployed. Unfortunately, this information is not available in the Danish data.

4.3 Descriptive data analysis

In this section, we present some descriptive statistics to get an impression of the data. As seen in Table 1, the sample of men in the considered age group consists of 29,221 individuals with a total of 58,673 unemployment spells beginning during the observation period. 8,578 are observed to be assigned to a programme, of which *classroom training* is by far the largest programme.

TABLE 1: DESCRIPTION OF THE SAMPLE

	Observations		
	Number	In percent	Ave. duration in weeks
Men aged 25-55	29,221		
Unemployment spells	58,673		
Participates in ALMP	8,578	100.00%	
Private Job Training	834	9.72%	22.1
Public Job Training	777	9.06%	35.4
Classroom Training	5,790	67.50%	15.8
Other Training	1,177	13.72%	15.2

There are a few remarkable differences between the participants in the 4 programmes.[19] Particularly, those assigned to *public job training* are generally older, less educated, had lower wages in their last job, and have experienced a longer period of unemployment before programme participation than participants in other programmes. This fits quite well with the perception that *public job training* is a 'last resort'. When estimating the model, we will be able to look deeper into the selection processes, and to some extent also be able to see whether there is selection on unobservables as well. The average duration of the programmes shows that *private* and particularly *public job training* have longer durations than *classroom training* and *other training*.

---

[19]Descriptive statistics are available on request from the authors.

As stated when describing the econometric model, it is a key identifying assumption that we observe some (exogenous) variation in the time until being assigned to a programme. Taking a closer look at the hazard rates into different programmes in Figure 2, we see that there is indeed a lot of variation in these durations. The entry rates are relatively stable for the first 50 weeks of unemployment and increase as the unemployed enters the active period.

FIGURE 2: KAPLAN-MEIER HAZARD RATES FROM UNEMPLOYMENT TO ALMPs
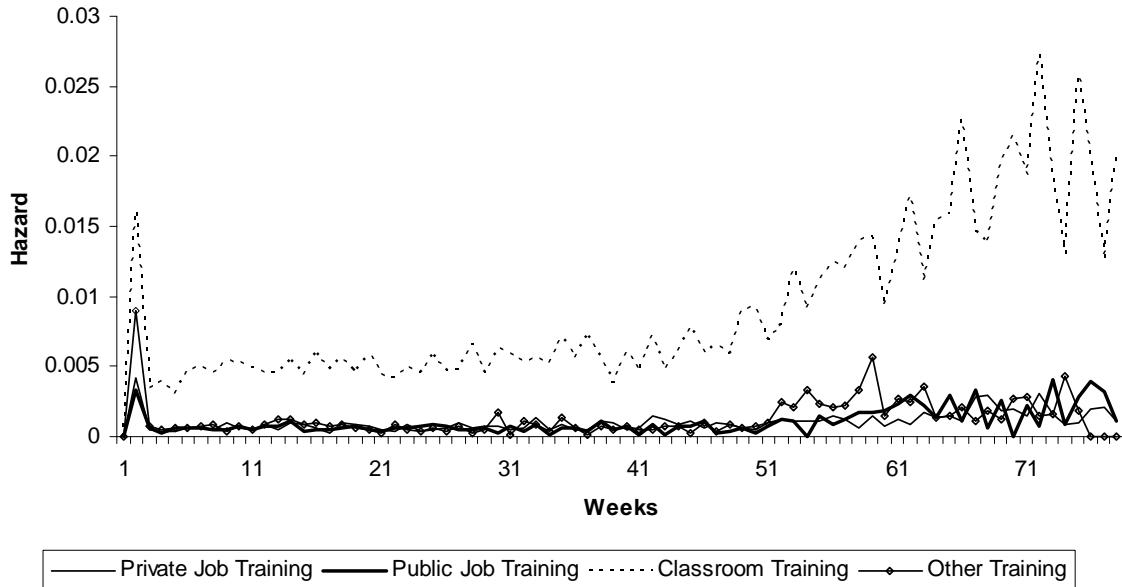
5 Results

To save space, we only present the results that are based on the model with heterogeneous treatment effects, and we only present the most relevant parameters. Remaining results are available upon request.

Tables 2 and 3 contain estimates of the coefficients to the interactions between the programme participation dummies, $d_{j,t}^1$ and $d_{j,t}^2$, $j = 1, 2, 3, 4$, and the most important observed characteristics. First, it should be noticed that it is not straightforward to interpret the coefficients in the tables since the model is non-linear. We therefore calculate expected durations in the next subsection. However, it is obvious from the tables that the

estimated effects are certainly heterogeneous. For example, it appears that programmes are more effective when the local unemployment rate is low, and that immigrants from Non-OECD countries have more favorable effects than the reference group (natives and immigrants from OECD countries). Second, the estimated standard errors are quite large for some parameters. This points to a requirement for large amounts of data (e.g. 100 % rather than 10 %) when estimating models with heterogeneous effects.[20]

TABLE 2: HETEROGENEOUS EFFECTS - INTERACTIONS WITH LOCKING-IN EFFECT

Variable	Priv. Job Tr. Est.	Priv. Job Tr. Std.	Publ. Job Tr. Est.	Publ. Job Tr. Std.	Class. Tr. Est.	Class. Tr. Std.	Other Tr. Est.	Other Tr. Std.
Constant	-0.309	0.367	-0.638	0.462	**-0.673**	0.148	-0.188	0.307
Age over 50	0.072	0.207	**-0.403**	0.198	-0.008	0.097	-0.404	0.217
Non-OECD	0.370	0.313	0.276	0.438	**0.677**	0.130	**0.586**	0.216
Labour Market Exp.	-0.055	0.103	**-0.206**	0.104	0.006	0.044	0.024	0.095
Wage in previous job	0.006	0.163	0.298	0.213	**0.095**	0.020	**0.227**	0.082
Elementary or High school	0.062	0.146	0.210	0.161	**0.274**	0.068	-0.117	0.141
Further Education	0.078	0.218	0.200	0.233	**0.184**	0.082	-0.094	0.170
Local unemployment	0.005	0.037	-0.042	0.040	**-0.067**	0.017	**-0.068**	0.032

Note: Bold coefficients are significant at a 5 % significance level.

---

[20]Danish administrative registers cover the entire population; hence, had these been available to us, we could have estimated the interaction parameters more precisely. This means that in case the model is implemented on a large scale, the results and predictions of the model can be made much more accurate.

Variable	Priv. Job Tr. Est.	Std.	Publ. Job Tr. Est.	Std.	Class. Tr. Est.	Std.	Other Tr. Est.	Std.
Constant	0.143	0.759	0.991	0.755	**0.470**	0.160	0.377	0.493
Age over 50	0.341	0.321	-0.642	0.398	**-0.273**	0.095	-0.312	0.257
Non-OECD	-	-	-	-	0.049	0.166	**0.893**	0.445
Labour Market Exp.	-0.265	0.199	-0.323	0.208	0.036	0.043	**0.284**	0.119
Wage in previous job	-0.202	0.250	0.480	0.330	-0.021	0.037	-0.014	0.171
Elementary or High school	-0.369	0.277	0.477	0.321	-0.095	0.069	0.215	0.199
Further Education	0.383	0.392	-0.003	0.559	-0.114	0.082	**0.516**	0.225
Local unemployment	0.070	0.079	**-0.234**	0.070	**-0.055**	0.018	**-0.158**	0.050

Note: Bold coefficients are significant at a 5 % significance level.

Coefficients to interactions cells with less than 20 observations are not estimated.

5.1 Calculating counterfactual outcomes

In this section, we construct expected unemployment durations conditional on participation in the different programmes. That is, we construct the potential outcomes for each individual. The expected duration of unemployment is a natural way of presenting the results, and facilitates a more clear interpretation compared to just analyzing the coefficients on all the interaction terms. But it also requires us to make additional assumptions. We need to make assumptions regarding the duration of the programmes and the starting time for the programmes. The durations for *private job training* and *public job training* are set to 26 weeks while the durations of *classroom training* and *other training* are set to 16 weeks[21]. With regards to the starting time we focus on two different assumptions each of which provides us with useful information on the effectiveness of the system. First, we consider what happens if we set the starting time of the programme to 0; that is, we assume that the unemployed enter a programme immediately after entering unemployment. This assumption makes it possible to do the analysis for all individuals entering unem-

---

[21]These durations are chosen to reflect realistic durations of the programmes. The average durations shown in Table 1 are observed and not planned durations of the programmes. A strong locking-in effect for a specific type programme willl increase the observed duration. In particular, this is the case for *public job training*. But this has already been incorporated in the model when we estimate the exit rate out of unemployment for those in *public job training*. So to avoid counting the locking-in effect twice we chose to set the programmes for both *private* and *public job training* to 26 weeks (instead of 35 weeks for *public job training*).

ployment and indicates the effects of using the statistical programme assignment model at this point in time.[22] Unfortunately, this choice of starting time does not allow us to compare the statistical model with the current system of case worker assignment because case workers naturally choose *no training* for all unemployed during the first week. For this reason we also consider setting the starting time to the point in time where the unemployed in our sample are in fact assigned to a programme. This allows us to compare case worker assignment with the suggested assignments from the statistical model, but only for those unemployed who eventually are assigned to a programme during the observation period. In the following section we presents the results with the starting time set to $t = 0$, and the analysis with the starting time set to the actual starting times are postponed to section 5.4 where we compare different assignment mechanisms.

By combining the parameters of the model with the observed characteristics and the assumptions on the duration and timing of programmes, we can calculate the expected duration of unemployment conditional on participation in each of the programmes. In practice, this is done by noting that[23]

$$E[T_u|\{x_t, d_{j,t}^1, d_{j,t}^2\}_0^\infty] = \sum_{m=1}^{2}\sum_{n=1}^{2}\Pr(v_{um}, v_{pn})\sum_{k=1}^{\infty}\exp[-\sum_{s=1}^{k}\theta_u(s|x_s, d_{j,s}^1, d_{j,s}^2, v_u, v_p]$$

We have calculated expected values for all the programmes as well as the option of *no training,* and a summary of the results is shown in Table 4, where the starting time of the programmes are set to 0.

---

[22]We have repeated the analysis at various points in time $t =$3,6,9,12 months of unemployment. The findings corroborates the results presented in the rest of the paper and are of course available upon request.

[23]This is most easily seen from the following rewriting $E(T) = \int_0^\infty tf(t)dt = \int_0^\infty(1 - F(t))dt = \int_0^\infty S(t)dt = \int_0^\infty \exp\left[-\int_0^s \theta_u(s|x_s, v_u)ds\right]dt$

TABLE 4: EXPECTED DURATIONS OF UNEMPLOYMENT, T=0

Programme	Mean	Median	Min.	Max	Std. Dev.
No Training	35.1	29.8	6.6	470.2	20.7
Private Job Training	39.4	33.9	7.7	516.2	22.5
Public Job Training	60.8	46.6	1.0	381.4	44.0
Classroom Training	43.7	37.9	9.9	472.0	23.0
Other Training	48.3	39.6	1.7	330.8	31.0

Before moving on to a discussion of different assignment models, we calculate the estimated effects and their standard errors. This is not a trivial task in this case where we first estimate a large number of parameters in the hazard models and then uses these estimates to calculate different estimates. Some kind of bootstrap method where we estimated the model several of times on subsamples of the data would be one possible way to estimate standard errors of the key results. Unfortunately the estimation process in very time-consuming and hence this is not a feasible procedure in this case. The method we applies can however be seen as a short-cut to complete bootstrap where we impose the assumption that the estimated parameters in the hazard model are drawn from a multinormal distribution. From the estimated parameter vector and the corresponding covariance matrix from we can then draw a number of 'simulated' parameter vectors with the estimated mean and covariance matrix. Then, for each of these parameter vectors we can calculate the potential durations and finally compute the standard deviation of the calculated effects over the simulations, and use this as the standard error of the estimated effects. This procedure also allows us to calculate standard errors of the effects of different assignment mechanisms presented in section 5.4.

With the estimated standard errors we can then perform the t-tests for each individual. A summary of this exercise is shown in Table 5.

TABLE 5: SUMMARY OF T-TESTS

Programme	Avg. Effect	Avg. Std.Err.	% Negative	Sig. Pos.	Sig. Neg.
Private Job Training	4.3	9.0	27.4%	6.6%	6.4%
Public Job Training	25.7	13.7	11.3%	54.0%	2.3%
Classroom Training	8.6	2.9	3.0%	86.0%	0.0%
Other Training	13.2	8.4	15.6%	49.0%	2.9%

Note: Effects are calculated as $Y_1 - Y_0$ so a positive effect means that the expected duration of unemployment is prolonged.

Sig. Pos. means that the effect is significantly larger than 0 and correspondingly for Sig.Neg.

The first two columns give an impression of the size of the effects compared to the average standard errors. The third column shows the fraction of negative estimates (decreasing unemployment duration), while the last two columns show the fraction of estimated effects which are significantly positive or significantly negative. For *private job training* 27.4% of the individuals have a negative effect, but only 6.4% of the individuals have a significant negative effect. The other programmes have less negative effects, and almost none of these are significant, and for most individuals the effects are actually significant positive, meaning that the duration in unemployment is prolonged.

Notice that - in line with the previous literature on short-term effects of ALMPs on individual job finding rates as summarized in e.g. Heckman et al., (1999) and Kluve (2006) - participation in generally ALMPs prolongs unemployment periods. Still, *private job training* is the best programme, and *public job training* is clearly the worst programme overall. There is, however, variation in the estimated effects, so in Figure 3 we have shown histograms for the estimated effects for the programmes.
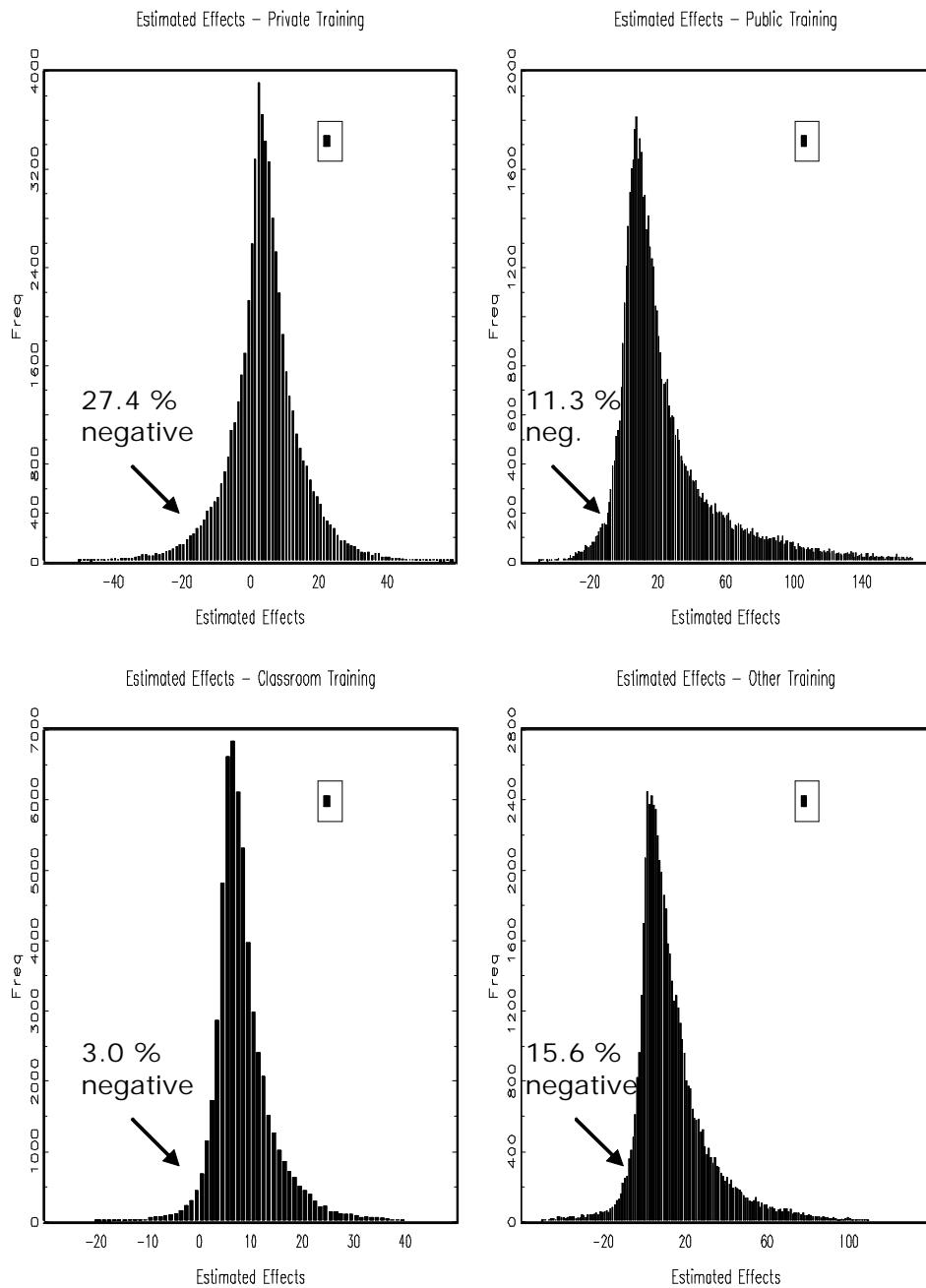
FIGURE 3: HISTOGRAMS OF ESTIMATED EFFECTS

25

A negative effect in the histogram implies that unemployment duration is shortened as a result of participating in the programme when compared to *no training*. It is seen that at least some individuals have negative estimated effects, and so it would be desirable to target the programmes to these individuals and avoid assigning individuals to programmes where the expected unemployment duration is much longer.

5.2   How should we present results to case workers?

For a particular unemployed with certain characteristics, we can now provide the case workers with information on the effectiveness of the different programmes. But how should this information be presented to the case workers? We would of course show the estimated programme effects, but we also needs to find a clear way of expressing the statistical uncertainty of the estimates. We could rank the programmes according to effectiveness, and present standard errors and test statistics, but case workers are not trained in the use of these concepts. Frölich (2006) describes how a method called Multiple Comparisons with the Best (MCB) is used to present results to case workers in the SAPS project in Switzerland.[24] For a given significance level, this method calculates, for each individual, a set of programmes, $\widehat{S}_i$, which contains the best programme with the chosen probability. This set of best programmes may contain only one programme, in which case this programme is significantly better than all the other possible programmes, or it may contain more than one programme, which would allow us to say that the best programme is one of the programmes in the set, but not to point at a particular programme. The method of MCB can also provide a set of programmes which are significantly worse than the best programme, W, and both of these sets might be valuable information to case workers.

This way of presenting the results to case workers seems appropriate when the statistical assignment model is intended as an information tool for case workers, who still have the discretionary power of choosing the programme. The case worker may consider choosing a programme from the set of best programmes, and then take additional information into account - e.g. supply constraints and unobserved characteristics that were not used for prediction - when choosing a particular programme from the set (or even outside it).

MCB also seems appropriate from a statistical point of view. In MCB, the testing is automatically carried out using the joint distribution of the impact estimates. This is not

---

[24] For more details on the usage of MCB in economics see e.g. Horrace & Schmidt (2000).

always what is done when interpreting results like this. Often, t-statistics are calculated to test whether a particular programme effect is significantly larger than 0, or larger than another programme effect. A problem with this test procedure may arise if it is repeated several times because the probability of making a type 1 error, that is, not to reject a false hypothesis, would not be equal to our chosen significance level. If we just test a sufficient amount of parameters, some of them will probably be considered significant, even if they are not. We could then adjust the significance level for the individual t-tests to take into account that we are going to make a certain number of tests. However, using the MCB method, this is automatically taken into account. This latter point may not be very important when we have only 4 types of programmes, but if we could divide the programmes into perhaps 50 categories, then it might be important for the results. Finally, if we want to test all programmes against the *best* programme, we need to include the additional uncertainty stemming from the fact that we do not *ex ante* know the best programme and, hence, which programme to use for comparison.

5.3 Multiple Comparison with the Best - Results

We first present the results for a particular individual and include a traditional t-test analysis for comparison. The upper part of Table 6 shows that *public job training, classroom training* and *other training* all have effects that are significantly different from zero for this individual, and that the 3.0 weeks increase in unemployment for *private job training* is not significant. The lower part of Table 6 shows the results from a MCB analysis for the same individual. Here, *no training* and *private job training* constitute the set of best programmes (shown in bold), while *public job training* is the only programme in the set of worst programmes (shown in italic), which is seen by the fact that the lower bound (MCB L) for the difference from the best programme is greater than zero. *Classroom training* and *other training* is in an intermediate group because neither are in the set of best programmes nor in the set of worst programmes.

TABLE 6: EXAMPLE OF OUTPUT FOR A PARTICULAR INDIVIDUAL

Traditional t-tests analysis			
Programme	Est. Eff.	Std. Err.	p-value
Private Job Training	3.0	3.95	0.221
Public Job Training	**31.5**	9.95	0.001
Classroom Training	**7.6**	1.57	0.000
Other Training	**13.9**	4.66	0.001

MCB analysis				
Programme	Est. Dur.	Diff. from best	MCB L	MCB U
No Training	**24.7**	0.00	0.00	6.25
Private Job Training	**27.7**	3.04	0.00	12.67
Public Job Training	*56.2*	31.50	3.29	55.85
Classroom Training	32.3	7.60	0.00	13.98
Other Training	38.6	13.91	0.00	24.82

Note: Upper part: Bold estimates are significant at a 5% level.
Lower part: Bold estimates are in S and italic estimates are in W.
MCB is calculated using a confidence level at a 95%.

Table 7 shows some summary statistics for the MCB analysis using different confidence levels. First, the cardinality of the set of best programmes reveals that a single programme rarely comes out as the best using a confidence level at 95 %. However, we can often distinguish the 2 or 3 best programmes (14.6 % and 32.7%). Finally, it is often the case that 4 or 5 programmes need to be included in S. If the confidence level is decreased and more mistakes are allowed to be made, we get somewhat more clear cut results with an overall decrease in the cardinality of the set S.

TABLE 7: CARDINALITY OF SET OF BEST PROGRAMMES, S

	Cardinality of set of best programmes, S					
Confidence level	0	1	2	3	4	5
0.95	0.0%	0.5%	14.6%	32.7%	37.1%	15.1%
0.9	0.0%	1.4%	23.5%	36.8%	28.1%	10.3%
0.7	0.0%	8.5%	44.6%	29.5%	13.9%	3.6%

In Table 8, the inclusion of particular programmes in the set of best programmes is analyzed.

TABLE 8: INCLUDED IN SET OF BEST PROGRAMMES, S

	Included in set of best programmes, S				
Confidence level	No Training	Private Job Tr.	Public Job Tr.	Classroom Tr.	Other Tr.
0.95	96.3%	97.7%	71.6%	18.6%	67.4%
0.9	94.4%	93.7%	61.3%	14.6%	58.5%
0.7	89.5%	77.9%	41.6%	8.1%	42.3%

No training and *private job training* are very often in S as expected, but the other programmes are quite often in S as well, except for *classroom training* which is only included in the set for 18.6% of the individuals.

Whether one prefers MCB over the familiar t-statistic is ultimately a matter of taste, but we argue that this may be an easy way of translating the uncertainty to case workers, and it often gives them a set of programmes to choose from.

One argument against the MCB-type presentation of the results is caused by the implicit objective function in the MCB analysis which is to identify the best programme with a given probability. This implies that the MCB method has *ceteris paribus* a tendency to include programmes with more uncertainty about the estimated effect in the set of best programmes. This problem arises because the method does not want to leave a possibly best programme outside the set of best programmes, and it is of course a perfectly correct statistical consideration. The problem is that we would prefer to select programmes which are more precisely estimated, and hence uncertainty should be signalled to the case worker. However, since we also present the estimated expected duration, case workers can observe it and easily judge if a programme is only included in the set because of large uncertainty about the estimate.

5.4 Comparing different assignment mechanisms

Having estimated the predicted durations for each individual, we can discuss how to select an assignment mechanism. At any point in time, the case workers can basically assign the unemployed to a programme or assign *no training* in which case a programme can be assigned in a future period. Hence, it is actually not only a question of choosing the best programme, but also a question of when to assign an individual to a programme. Again, to save space we present the results where we calculate expected durations for the potential outcomes at $t = 0$ when an individual enters unemployment.

We first consider the following assignment mechanisms[25]:

- Assigning *no training* to everyone.

- Assigning the *best programme* to everyone.

- Assigning the *worst programme* to everyone.

---

[25] This section follows the procedures suggested in Lechner & Smith (2007), which contains a comparable analysis.

- Assigning a *random programme* to everyone.

These assignment schemes ignore constraints in the availability of different types of programmes, which may be somewhat unrealistic, although the composition of programmes has changed quite dramatically over time. Therefore, we also consider some assignment mechanisms where we have imposed the restriction that the fraction of unemployed assigned to each programme should be as actually observed.[26] Such assignments will not be optimal, because of the imposed constraints, but they may serve as a lower limit of the effects of imposing the assignment mechanisms.

When considering restricted assignment mechanisms, the ordering of the unemployed becomes important, because some programmes may be closed for further assignments when considering how to assign the last unemployed in our sample. To be able to make a fair comparison to the current system, we have chosen to order the sample of unemployed randomly.[27] This should mimic the current system where case workers do not have information about all unemployed who will enter the unemployment office during the relevant period. When using the random ordering of unemployed, we only allow the statistical programme assignment model to use the same information as case workers have access to. We have considered the following restricted assignment mechanisms:

- Assigning the *best programme*, but imposing restrictions on the number of slots in each program. Ordered *randomly*.

- Assigning the *worst programme*, but imposing restrictions on the number of slots in each program. Ordered *randomly*.

- Assigning a *random programme*, but imposing restrictions on the number of slots in each program. Ordered *randomly*.

In Table 9 the average duration of unemployment for each assignment mechanism is shown. The standard errors are estimated in the following way. For each assignment mechanism we fix the chosen assignments and then we consider the consequences of these choices under the simulated alternatives for the potential outcomes as described in section

---

[26]The restrictions of the fractions in each programme are calculated from the actual assignment as seen in Table 1.

[27]Lechner & Smith (2007) also considered ordering the unemployed by "need" or "efficiency". In the Danish context, these orderings would not correspond to a realistic implementation of the assignment model, and hence we have not included these orderings in the results shown.

5.1. Then we can again use the standard deviation over the simulations to approximate the standard errors on the estimates. In the upper part of Table 9, we first show the assignment where all unemployed are assigned to *no training*, *best programme*, *worst programme* or *random programme*. This provides us with an idea of the potential improvements from the system, and it is clearly seen that there are huge differences in the duration of unemployment, confirming the hypothesis that treatment effects are heterogeneous such that a statistical programme assignment model may be valuable. Furthermore, the standard errors seems to be relatively small indicating that the differences between assignment mechanisms are significant.

In the lower part of the table, we consider restricted assignments. The first thing to note is that the imposed restrictions have reduced the difference between the expected durations for the assignment mechanisms, as expected. Still, there is considerable variation across the assignment mechanisms, and the overall reasoning from the results in Table 9 might be as follows: Random (not restricted) assignment implies an average unemployment duration of 45.07 weeks. When we impose the restricted number of slots in each programme the duration goes down by 7.02 [0.82] weeks to 38.05 (random but restricted and ordered randomly). The duration can be decreased by additionally 3.77 [0.32] weeks to 34.27 by allocating to the best programme, but still imposing the restrictions (best but restricted and ordered randomly). And finally, a decrease by 3.05 [0.61] weeks to 31.30 is possible if we assign everyone to the best programme, but then programme availability issues will have to be solved. The most reasonable approximation to the current system is probably the random assignment with restrictions[28]. So depending on the degree to which the number of slots in the programmes are binding an estimate of the overall decrease in the unemployment duration when introducing a statistical model of programme selection would be in the interval 3.77 - 6.82 weeks.

---

[28]We shall elaborate on this issue below.

TABLE 9: COMPARING ASSIGNMENT MECHANISMS

Assignment Mechanism	Estimated duration in weeks	Standard error
No Training	35.11	0.45
Best	31.22	0.72
Worst	66.65	3.30
Random	45.07	1.03
Best but restricted and ordered randomly	34.27	0.46
Worst but restricted and ordered randomly	39.51	0.66
Random but restricted and ordered randomly	38.05	0.51

Assigning everyone to *no training* is almost as good as assigning everyone to the best programme with restrictions, in which case this assignment mechanism seems to be a very easy (and cheap) way to get close to the optimum given the restrictions. But actually this is perhaps not as bad as it might seem at first even if the *no training* to everyone is not a usable option, because then at least an assignment mechanism exists, which does not increase the overall unemployment. This could be important if closing down the entire system is not a feasible option, or if there are large threat effects of programmes leading to shorter unemployment duration for everyone under the system.

The resulting assignments from the proposed assignment mechanisms are shown in Table 10. *No training* is the best programme for more than half of the unemployed in the considered sample, but in contrast to the former results, this actually suggests that many unemployed may do better when assigned to a programme. Notice that the restrictions imposed are almost binding for every assignment mechanisms. A noticeable deviation from this observation is that the assignment to the best programme (both restricted and not restricted) only assigns very few unemployed to *classroom training*.

TABLE 10: RESULTING ASSIGNMENTS USING DIFFERENT ASSIGNMENT MECHANISMS

Assignment Mechanism	No Tr.	Priv. J. Tr.	Publ. J. Tr.	Class. Tr.	Other Tr.
No training	58,673	0	0	0	0
Best	32,714	14,418	4,537	167	6,837
Worst	298	9,665	31,408	4,012	13,290
Random	11,765	11,665	11,816	11,681	11,756
Best, restricted, ordered randomly	51,735	1,675	1,561	1,339	2,363
Worst, restricted, ordered randomly	41,448	1,675	1,561	11,626	2,363
Random, restricted, ordered randomly	41,448	1,675	1,561	11,626	2,363

In the interpretation of the results above, we argued that a reasonable approximation of the current allocation mechanism is random assignments with restrictions. To provide some evidence on this assumption, we now consider the choice situations when case workers actually assigns an unemployed into a programme. Results from this exercise are shown in Table 11.

TABLE 11: ACTUAL CHOICE BY CASE WORKERS

Case workers' choice	Percent	
Best programme	14.9%	
Second best programme	13.7%	
Third best programme	15.5%	
Fourth best programme	19.4%	
Fifth best programme	36.4%	
Expected remaining durations under different mechanisms	Weeks	Standard error
Case workers' choice	45.3	0.70
Best, restricted, ordered randomly	37.4	1.60
Random, restricted, ordered randomly	39.6	1.22

Table 11 shows that case workers actually chose the worst programme most of the time (36.4 %), and the average differences in durations between the actually chosen programme and the best programme is 7.91 [1.42] weeks. Case workers are actually seen to perform worse than random assignment in this case, but since case workers may sometimes be forced (by the rules) to assign unemployed into a programme, we have also analyzed whether they then chose the best available programme. That is, we repeat the analysis in Table 11, but remove the option of assigning unemployed to *no training*. The results are shown in Table 12. This explains part of the difference between case workers' choice and the best choice. The difference between case worker assignment and assignment to the best programme is now 4.87 [0.64] weeks. And case workers are seen to perform almost exactly as good as random assignment in this comparison.

Case workers' choice	Percent	
Best programme	13.7%	
Second best programme	43.6%	
Third best programme	27.7%	
Fourth best programme	15.0%	
Mean durations under different mechanisms	Weeks	Standard error
Case workers' choice	45.3	0.70
Best, restricted, ordered randomly	40.4	0.74
Random, restricted, ordered randomly	45.2	0.92

The reality about the available choice set for case workers is probably something in between the results in Table 11 and 12, which means that they sometimes have the possibility of choosing the option *no training* whereas other times they are effectively forced to chose an actual programme. Hence, the estimated effect of introducing the statistical model at the point in time where case workers have actually assigned unemployed to a programme, is a reduction in the average unemployment duration in the interval 4.87 - 7.91 weeks. Furthermore we take the results in Table 11 and 12 to indicate that random assignment is actually a reasonable (conservative) approximation for case worker assignment.

We should note that when case workers are actually observed to assign unemployed to a programme, they have faced a number of choice situations where they have chosen *no training*, which is very often the best option. Nevertheless, when unemployed are assigned to a programme, they are rarely assigned to the programme which is estimated to be most effective. Some reasons for the poor performance of case workers may be that they are not aware of what is the best programme, or that they have different objectives when assigning the unemployed to programmes; e.g. that they are making the assignment based on Y_0, rather than $Y_r - Y_0$, that is, out of equity rather than efficiency considerations.

6 Conclusion

In this paper, we have discussed a statistical programme assignment model. We have argued that a statistical information tool may be able to improve the assignments of unemployed into ALMPs, and hence improve the effectiveness of these instruments, which is crucial, given the costs of operating these programmes. A multivariate duration model for the duration of unemployment is estimated, and the results indicate that unemployed

do in fact react differently to different programmes, and hence a potential improvement is possible by assigning individuals to the most effective programme. Compared to the current case worker assignment of unemployed to programmes, a statistical programme assignment model could lead to a decrease in the remaining unemployment duration in the range of 4.87 - 7.91 weeks (10.8% - 17.5%). Effects in the same range where found in an analysis of introducing the statistical model at the very first week of unemployment. These large impacts is obtained using a fairly small sample (for the study of heterogeneous effects a much larger sample would be preferred), a few selected explanatory variables in the interaction terms, and a crude grouping of programmes. This suggests to us that the introduction of a statistical model may be worth some serious consideration.

We have discussed some issues regarding the actual implementation of such a statistical programme assignment model, but there are still some open questions. One important issue is how to update the model; how do we handle the problem that the model is estimated on the population and then used to assign treatment to the population of unemployed who are going to be part of the estimation sample later on? Problems of this type could probably be reduced by updating the model relatively frequently because then the evolution of parameters in the model would change slowly over time as the model is sequentially updated with fresh unemployment spells. The parameters may change for a number of other reasons, so it would be appropriate to perform the estimation on a rolling observation window.

We have not taken possible general equilibrium effects into account, and these would certainly be important to consider before implementing the model. If the model, for example, suggests to assign a large fraction of unemployed to *private job training*, compared to the fraction in this programme today, this will influence the effects of these programmes, as well as the effects of other programmes. These considerations would also partially be circumvented by frequent re-estimations of the programme effects.

Another important issue not addressed in this paper is the costs of the different programmes. When assigning to programmes, we would like to assign unemployed based on cost-benefit considerations rather than just the effect on unemployment duration. This is not done in the present analysis, but it would be natural to include some measures of programme costs either directly into the identification of best programmes or at least to make cost estimates available to case workers when they are making assignments.

Finally, results from an actual experiment with a comparable model in Switzerland (the SAPS model) has shown that we need to think about how to make appropriate incentives

schemes for using the statistical model, because case workers in their experiment did comply with the suggestions from the model when it was made completely voluntary.

The main conclusion of this paper is that the potential gains from a statistical programme assignment model are large. A finer categorization of programmes and more explanatory variables for the heterogeneous programme effects combined with estimation on the full population of unemployed workers would allow us to identify more heterogeneity and reduce the uncertainty of the estimated effects, and hence make the assignment to the best programmes even more convincing. Based on our results, it would be natural to explore these potential gains further and eventually to conduct a randomized pilot study of the implementation of such a system.

References

[1] Abbring, J. H. & G. J. van den Berg (2003). "The Nonparametric Identification of Treatment Effects in Duration Models", *Econometrica*, Vol. 71, 1491-1517.

[2] Andersen, T.M & M. Svarer (2007). "Flexicurity – Labour Market Performance in Denmark", Forthcoming in *CESifo Economic Studies*.

[3] Auerhahn, K. (1999). "Selective Incapacitation and the Problem of Prediction", *Criminology*, Vol. 37 (4), 703-734.

[4] Behncke, S., M. Frölich & M. Lechner (2006). "Statistical Assistance for Programme Selection - For a Better Targeting of Active Labour Market Policies in Switzerland", *Discussion Paper*, No. 2006-9, Dep. of Eco., University of St. Gallen.

[5] Behncke, S., M. Frölich & M. Lechner (2007). "Targeting Labour Market Programmes: Results from a Randomized Experiment", *IZA Discussion Paper*, No. 3085.

[6] Bell, S. H. and L. L. Orr (2002). "Screening (and creaming?) applicants to job training programs: the AFDC homemaker-home health aide demonstrations", Labour Economics, 9, 279-301.

[7] Black, D., M. Berger & J. Smith (2001). "Evaluating Profiling as a Means of Allocating Government Services", In Lechner, M. and F. Pfeiffer (eds.), *Econometric Evaluation of Active Labour Market Policies*, Heidelberg: Physica, 59-84.

[8] Black, D., M. Plesca, J. Smith and S. Shannon (2003). "Profiling UI Claimants to Allocate Reemployment Services: Evidence and Recommendations for States", Final Report to United States Department of Labour.

[9] Black, D., J. Smith, M. Berger & B. Noel (2003). "Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System", *American Economic Review*, Vol. 93 (4), 1313-1327.

[10] Brinch, C. N. (2007). "Nonparametric Identification of the Mixed Hazards Model with Time-Varying covariates", *Econometric Theory*, Vol. 23, 349-354.

[11] Eberts, R., C. O'Leary and S. Wandner (2002). *Targeting Employment Services.* W. E. Upjohn Institute for Employment Research, Kalamazoo, Michigan.

[12] Frölich, M., M. Lechner and H. Steiger (2004). "Statistically assisted programme selection - International experiences and potential benefits for Switzerland", *Zeitschrift für Volkswirtschaft und Statistik*, Vol. 139 (3), 331-331.

[13] Frölich, M. (2006). "Statistical Treatment Choice: An Application to Active Labour Market Programmes", Forthcoming in *Journal of the American Statistical Association*.

[14] Gaure, S., K. Røed & T. Zhang (2007). "Time and Causality: A Monte Carlo assessment of the timing-of-events approach", *Journal of Econometrics*, Vol. 141, 1159-1195.

[15] Geerdsen, L. (2006). "Is There a Threat Effect of Labour Market Programmes? A Study of ALMP in the Danish UI system", *Economic Journal,* 116, 738-750 .

[16] Gottfredson, S. D. & L. J. Moriarty (2006). "Statistical Risk Assessment: Old Problems and New Applications", *Crime & Delinquency*, Vol. 52, 178-200.

[17] Harcourt, B. E. (2004). "Rethinking Racial Profiling: A Critique of the Economics, Civil Liberties, and Constitutional Literature, and of Criminal Profiling More Generally.", *The University of Chicago Law Review*, Vol. 71 (4), 1275-1381.

[18] Heckman, J., R. Lalonde & J. Smith (1999). "The Economics and Econometrics of ALMP", Handbook of Labour Economics, Vol. 3, North-Holland, Amsterdam.

[19] Heckman, J. & B. Singer (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica*, Vol. 52 (2), 271-320.

[20] Heckman, J., S. Urzua & E. Vytlacil (2006). "Understanding Instrumental Variables in Models with Essential Heterogeneity", *Review of Economics and Statistics*, Vol. 88 (3), 389-432.

[21] Heckman, J., J. Smith & N. Clements (1997). "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts", *The Review of Economic Studies*, Vol. 64 (4), 487-535.

[22] Horrace, W. C. and P. Schmidt (2000). "Multiple Comparisons with the Best, with Economic Applications", *Journal of Applied Econometrics*, Vol. 15, 1-26.

[23] Jensen, P. & M. Svarer (2003). "Short- and Long-Term Unemployment: How do Temporary Layoffs Affect this Distinction?", *Empirical Economics*, Vol. 28, 23-44.

[24] Khan, J., J. Wei, M. Ringnér, L. Saal, M. Ladanyi, F. Westermann, F. Berthold, M. Schwab, C. Antonescu, C. Peterson and P. Meltzer (2001). "Classification and Diagnostic Prediction of Cancers using Gene Expression Profiling and Artificial Neural Networks", *Nature Medicine*, Vol. 7 (6), 673-679.

[25] Kluve, J. (2006). "The Effectiveness of European Active Labor Market Policy", *IZA Discussion Paper*, No. 2018.

[26] Lechner, M. and J. Smith (2007). "What is the Value Added by case workers?", *Labour Economics,* 14 (2), 135-151.

[27] Manski, C. (2004). " Statistical Treatment Rules for Heterogeneous Populations", *Econometrica*, 72 (4), 1221-1246.

[28] Manski, C. (2006). "Profiling: Introduction to the Feature", *The Economic Journal*, Vol. 116.

[29] Murphy, S. A. (2005). "An experimental design for the development of adaptive treatment strategies", *Statistics in Medicine*, Vol. 24, 1455-1481.

[30] Persico, N. (2002). "Racial Profiling, Fairness, and Effectiveness of Policing", *The American Economic Review*, Vol. 92 (5), 1472-1497.

[31] Persico, N. & P. E. Todd (2005). "Passenger Profiling, Imperfect Screening, and Airport Security", *PIER Working Paper 05-005*.

[32] Rosholm, M., J. Staghøj, M. Svarer and B. Hammer (2006). "A Danish Profiling System", *Nationaløkonomisk Tidsskrift*, Vol. 144, 209–229.

[33] Rosholm, M. & M. Svarer (2004): "Estimating the Threat Effect of Active Labour Market Programmes", Working Paper No. 2004–06, Dep. of Economics - University of Aarhus.

[34] Rubin, D. B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, Vol. 66, 688-701.

[35] Rubin, D. B. (1980). "Discussion of "Randomization analysis of experimental data in the Fisher randomization test" by Basu." *Journal of the American Statistical Association*, Vol. 75 (371), 591-593.

[36] Shaw, M., C. Subramaniam, G. Tan and M. Welge (2001). "Knowledge Management and Data Mining for Marketing", *Decision Support Systems*, Vol. 31, 127-137.

[37] Stephan, G, S. Rässler and T. Schewe (2006). "Wirkungsanalyse in der Bundesagentur für Arbeit Konzeption, Datenbasis und ausgewählte Befunde", IAB Discussion paper 07-2006.

[38] Van den Berg, G. (2001). "Duration Models: Specification, Identification and Multiple Durations", in: Heckman, J. & E. Leamer (eds.) *Handbook of Econometrics*, Vol. 5, 3381-3460.

[39] Wandner, S. (2002). "Targeting Employment Services under the Workforce Investment Act", in: Eberts, R., C. O'Leary, S. Wandner (eds.), *Targeting Employment Services*, W. E. Upjohn Institute for Employment Research, Kalamazoo, Michigan.

[40] Yeo, A., K. Smith, R. Willis and M. Brooks (2001). "Clustering Technique for Risk Classification and Prediction of Claim Costs in the Automobile Insurance Industry", *International Journal of Intelligent Systems in Accounting, Finance & Management*, Vol. 10, 39-50.