Vocational Training:

Does it speed up the Transition Rate out of Unemployment?[1]

Bart Cockx[2] and Isabelle Bardoulat

June 22, 2000

Abstract:

In this paper we estimate, for the 1989-93 period in Belgium, the effect of vocational classroom training on the rate of transition out of unemployment. We show that rationing of the demand for training increases the unemployment duration of non-participants, an effect neglected in programme evaluations. We propose a “control function” estimator accounting for variable treatment effects. In the absence of interaction effects between explanatory variables this estimator identifies treatment effects free from selection bias. A natural experiment induces exogenous sub-regional variation in the training supply. This provides over-identifying restrictions that cannot be rejected. During participation, the transition rate decreases by 27%. Afterwards it increases by 62%. Making training available for a broader population would, however, reduce the effectiveness of the programme.

Keywords: Vocational training, evaluation, transition models, job-search.

JEL classification numbers: C41, J24, J64 and J68.

---

[1] This project has been financed by the following grants: SSTC n° DB/10/035, PAI P4 /01 and Tournesol n° 98.067. We thank the Office National de l’Emploi and the Office communautaire et Régional de la Formation Professionnelle et de l’Emploi for providing the data. This research project has especially benefited from the support of and the discussions with Michel Mouchart and Bruno Van der Linden. We thank Jan van Ours, the participants of the conference on the Econometric Evaluation of Active Labour Market Policies in Europe, June 25-26, 1999 in Mannheim and the participants of the research seminars of the Laboratoire de microéconométrie du CREST in Paris and at the Tinbergen Institute in Amsterdam for their comments.

1. Introduction

Since public authorities have realised that macro-economic policies no longer suffice to cope with high and persistent unemployment, active labour market policies have been increasingly used in OECD countries (OECD 1993). Unemployment persistence is especially important in Belgium. The share of long-term[3] unemployment in Belgium is among the highest in the OECD. During the 1983-1996 period this share has only been (slightly) below 60% in the 1991-93 recession (OCDE 1997, p.8). Despite this persistent unemployment, we observe that Belgian has spent a higher share of its GDP on active labour market policies than other countries of the European Union (EU) on average. In 1995 this share amounted to 1.4% for Belgium, as compared to 1.1% for the UE (OCDE 1997, p.99). The high share of expenditures, oriented to direct job creation in the public and the non-profit sector, largely explains this finding. For, if we exclude expenditures on direct job creation, the figures on shares drop to 0.8% for Belgium and 0.9% for the EU. Moreover, if we consider the outlays on training schemes for unemployed workers or workers at risk of loosing their jobs, Belgium spends only 0.16% of its GDP, whereas this share is 0.27% for the EU on average.

In this paper we study whether the Belgian authorities are likely to reduce unemployment by increasing their expenditures on vocational classroom training programmes for unemployed workers. We provide an answer to this question by investigating whether, during the 1989-93 period, training has increased the transition rate out of unemployment in Wallonia, the French speaking region in the South of Belgium[4]. We will distinguish between the impact of training during participation and afterwards. Moreover, we explicitly account for the variability of the impact across trainees. In particular, typically only unemployed workers with the highest returns to training enrol. Consequently, we find that increasing the coverage of the programme to a broader population, such as recently encouraged by the European Union (EU) in Luxembourg (Commission of the EU 1997), is likely to be much less effective than the current programme.

We cannot, however, provide a definite answer to the above-mentioned question and this for two reasons. First, this study focuses on the short run by only considering programme effects during the unemployment spell in which the worker is trained. The long-term effects, *i.e.* the impact on the duration and the incidence of unemployment after this first

---

[2] IRES and Department of Economics, Université Catholique de Louvain, Place Montesquieu, 3, B-1348 Louvain-la-Neuve, Belgium. E-mail: cockx@ires.ucl.ac.be.

[3] Defined as being unemployed for more than one year.

[4] Plasman (1993, 94) and Bollens and Nicaise (1994) also studied this relationship. The former finds that training decreases the transition rate out of unemployment, while the latter authors report shorter unemployment spells for participants in Flanders, the region in the North of Belgium. These studies suffer from serious methodological shortcomings, however (see Cockx 1999 for a discussion). Cockx

unemployment spell is not studied in this paper. Second, even if we estimate that training reduces the length of unemployment spells, the effectiveness of such a policy will eventually depend on the importance of spill-over effects: displacement (jobs created by the programme are at the expense of other jobs), substitution (jobs created for a certain category of workers replace jobs for other categories), and tax effects (the effects of taxation required to finance the programme on the unemployment rate) (Calmfors 1994). Heckman, Lalonde and Smith (1998) argue that these general equilibrium effects are likely to be important.[5]

Even if we do not study the above mentioned general equilibrium effects of training, we do account for an indirect impact of the training scheme on non-participants neglected in the evaluation literature. The rationing of the demand for participation in the voluntary programme under consideration induces this indirect effect. We show, on the basis of a standard job-search model[6], that increasing the availability of training slots will both increase the number of applicants and the reservation wage of these candidate trainees. This will reduce, among non-participants, the average rate at which unemployment is left for employment.

In Europe, the literature (see Heckman, Lalonde and Smith 1999 for a survey), generally finds that participation in classroom training increases the chances of being employed and decreases the risk of unemployment[7], at least if the training content is sufficiently large[8]. For East Germany, and Norway we find, however, mixed evidence[9]. If we consider the effects of classroom training on wages and earnings, these are often reported to be insignificant or even negative. Taking the relatively small amount of time spent in training,

---

et al. (1998) find that vocational training for unemployed workers increases job *tenure*, but the effect is statistically insignificant.

[5] Van der Linden (1997), on the basis of a questionnaire submitted to employers, confirms this allegation for active labour market policies in Belgium.

[6] In a similar spirit, Davidson and Woodbury (1993) analyse in a job-search framework the indirect effect of a reemployment bonus scheme for workers entitled to Unemployment Insurance on workers who are ineligible for the bonus.

[7] See Bonnal, Fougère and Sérandon (1997) for France, de Koning et al. (1991) and Ridder (1986) for the Netherlands, Zweimuller and Winter-Ebmer (1996) for Austria, Breen (1991) for Ireland, Björklund (1993, 1994) for Sweden, Firth et al. (1999) for the U.K., Puhani (1998) for Poland, and Lubyova and van Ours (1998) for the Slovak Republic, although this programme mainly consisted of on-the-job training rather than in a classroom. Note that the Youth Training Scheme in the U.K., analysed by Dolton et al. (1994), should rather be considered as a work experience scheme with little training content.

[8] Magnac (1998), for instance, finds no noticeable effects of training in France. However, he includes work experience programmes in his definition of training: these are found to have a negative impact on employment probabilities (see Bonnal et al. 1997).

[9] Kraus et al. (1997) find that classroom training in East Germany increases the hazard from non-employment to stable employment. Lechner (1999) finds no significant impact of participation on the unemployment rate. For Norway, Torp et al. (1993) find an insignificant effect on the employment rate, becoming negative focusing on those who complete training. Torp (1994) finds that classroom training has positive employment effects for short courses (5-10 weeks) and long courses (more than 30 weeks), but negative effects for courses of intermediate length. Aakvik (1999) finds positive effects. Flaws in the methods used to correct for the selection bias may be the source of this conflicting evidence.

some weeks or at most some months, into account, the latter result may not come as a surprise. Compare it to the annual rate of return to education, which is of the order of 10% (Ashenfelter and Rouse 1995).

The literature on the evaluation of training programmes in Europe has only recently started to accumulate significantly. Given that some of these studies may not have adequately corrected for selection bias, we compare these findings to the U.S. findings. In the U.S. there exists a vast literature evaluating the impact of public sector sponsored classroom training programmes (see Lalonde 1995, Friedlander, Greenberg and Robins 1997, and Heckman et al. 1999 for recent surveys). In contrast to the European studies, the most common outcome of interest is earnings, rather than employment or unemployment. In general, this literature finds significantly small, but significantly positive effects on earnings of adults, especially for women. No training programme has been found effective for youth, however. Studies that decompose these earnings gains in its components usually find that they result more from increased weeks of employment than from increased hourly wages or increased hours per week for those who are employed. This is consistent with the findings of the European studies. Eberwein, Ham and Lalonde (1997) report that participation in classroom training increases the frequency of employment, but not the length of employment spells.

Each micro-econometric evaluation study of a labour market policy needs dealing with potential biases of the programme effect due to selective participation. In this paper, we rely on administrative data with a small number of explanatory variables. We cannot therefore rely on matching estimators as recently advocated in the literature (Dehejia and Wahba 1999, Heckman, Ichimura and Todd 1997, Lechner 1999, and Brodaty et al. 1999 in the context of transition models). In this study, we follow a quasi-experimental approach (see Meyer 1995, Angrist and Krueger 1999). In particular, we follow Cockx and Ridder (2000) who show that the selection bias may be eliminated by aggregation if the variation of the participation rate at the aggregate level is exogenous (see also Wald 1940 and Angrist 1991). We will proceed in two steps. First, we will assume that this exogeneity is only assured after controlling for the correlation of the participation rate with other explanatory variables. In this case, like for a differences-in-differences estimator, identification hinges upon the absence of interaction effects. In a second step we will exploit a natural experiment generating sub-regional variation in the participation rates in training that is that and argue that is unrelated to the outcome variable, i.e. the transition rate out of unemployment. We may then ignore to control for sub-regional effects. These over-identifying restrictions allow us to test for the validity of the identifying assumptions.

Exogenous variation allows the average effect of treatment on the treated, the *selected average treatment effect* (SATE[10]) to be identified non-parametrically, if the effect of training is constant over the population. If the treatment effect varies over trainees, this exogenous variation will only identify SATE if the programme effect does not influence the decision to participate in training (Heckman and Robb 1986 and Heckman 1997). However, the participation in the programme under consideration is voluntary and application costly. The expected returns to training need therefore exceed the application costs for a worker to participate in the programme. This clearly violates the above-mentioned condition. Under these circumstances we rather identify the local average treatment effect (LATE), i.e. the average impact of training for those unemployed workers who would be induced to participate if they moved from a sub-region with low training capacity to a high one (Imbens and Angrist 1994, Angrist, Imbens and Rubin 1996).

In order to identify SATE rather than LATE, we follow a strategy such as proposed by Heckman and Honoré (1990). The distribution of the treatment effects can be identified if participation depends only on the gain of the programme, such as in Roy's model (1951). We therefore model the participation decision in training within a standard job-search framework and derive conditions under which this assumption is satisfied. Heckman and Honoré (1990) show that with sufficient price variation across different markets of "occupational choice" one can identify the population skill distribution from aggregate data. In this paper, since the participation in training is rationed, variation in the sub-regional training capacities induces sub-regional variation in the returns to training, i.e. in "prices". We show that this variation allows us to identify the distribution of treatment effects non-parametrically. In the empirical application we will impose a parametric distribution, because there is insufficient variation in the training capacity to estimate the distribution non-parametrically. We experiment, however, with different parametric specifications. This parameterisation allows us not only to estimate SATE, but also to estimate the *average treatment effect* (ATE) of all unemployed workers.

In the following section we describe the institutional framework. In Section 3 we present a formal job search model predicting the participation decision in training and discuss the identification of SATE. In Section 4 we describe the data and present the statistical model. Section 5 discusses the estimation results. A final section concludes.

[10] See Angrist and Imbens (1991).

2. The Institutional Framework

In Wallonia the regional public employment agency, FOREM, is the main operator of vocational classroom training for adult unemployed workers[11]. For this purpose, it has a network of training centres at its disposal widely dispersed over the territory of the Walloon region. FOREM provides vocational training for a wide range of professions of the secondary (FP II) and tertiary sector (FP III). Apart from vocational training, it also offers services information, vocational guidance and work experience to a more disadvantaged population of unemployed workers in its Centres of Reception, and of Guidance and Socio-Professional Initiation (C.A./C.O.I.S.P). During the 1989-93 period, about two thirds of the available training slots are of the vocational type, each sector being equally important. In terms of hours of training, the share of vocational training increases to 85%[12]. The level at which vocational training is offered varies, but the majority of the programmes consists in the development of basic skills required in particular vocations. The data that were used for the evaluation study did not allow distinguishing between the different programme types. The estimated effect of participation to training is therefore necessarily an average effect over the wide variety of programmes offered.

What determines participation into these training programmes? In principle no restrictions are imposed. Participation is voluntary and any unemployed worker can apply. However, eligibility requires candidates to pass tests on basic reading and writing when applying for a training programme in the tertiary sector and on calculation in the secondary sector. In addition, they need to pass medical and psycho-technical tests. These tests are uniform over training centres and types. Subsequently, programme administrators select candidates on the basis of some vague criteria as “motivation”. Finally, because there is an excess demand for training, candidates are registered on a waiting list[13] until a new training session of their choice starts off.

Neither training centres, nor administrators or instructors are in any way rewarded on the basis of placement ratios or to other labour market outcomes. Consequently, direct incentives to cream-skim candidates, likely to find jobs even without participation in the programme, are absent. Even if we cannot evaluate for certain the implicit selection criteria of the programme administrators, selection seems therefore rather to be based on criteria that affect the returns to training than on the absolute employability of the worker. In the theoretical model below we will assume that workers with high returns to training are more

---

[11] More details of the institutional framework can be found in OCDE (1997).

[12] In the sequel figures for which no source are mentioned are taken from FOREM (1989-94) or from unpublished document of the employment agency.

[13] We do not have any information on the length of this waiting period.

likely (or at least not less likely) to be selected for participation than workers with low returns.

Workers are not only selected into the programme, however. The costs imposed on the participation in a training programme induce a process of self-selection. Only workers with high returns to training can compensate for these costs and participate. Participation costs are partially reimbursed to the participant. First, a trainee remains entitled to unemployment benefits. Moreover, he/she is paid 40 BEF (= 1 EURO) for each effective training hour and is reimbursed transportation costs. The candidate trainee must, however, bear two main costs: the application costs and the earnings forgone during the period in which the candidate trainee is waiting to enter training. The application costs are the cost of gathering information on the programme, as well as the time, effort and transportation costs of the higher mentioned test and interview. Second, as a consequence of the limited training capacity, the candidate must be ready to wait before he or she can participate in training. This means that participation in training requires the worker to reject job offers during this waiting period.

3. Modelling the Participation Decision in Training and the Identification of Average Treatment Effects

In this section we present a standard job search model incorporating the decision to participate in training. Although a reduced form will be estimated in the empirical analysis, the structural model is not only useful in that it provides prior information regarding the sign of particular parameters, but also in that it clarifies under what conditions the distribution of the treatment effects can be identified. The identification strategy resembles the one followed by Heckman and Honoré (1990) assuming that the selection in the programme depends on its return[14]. We first describe the optimal strategy of the individual unemployed worker. Subsequently, we discuss the identification of SATE and ATE.

3.1. The Optimal Strategy of a Worker

Consider a standard job search model in which risk neutral unemployed workers search sequentially for a job. Job offers arrive according to an exogenous rate l_0 and offers

[14] See Heckman, Smith and Clements (1997) for other identification strategies.

are random drawings from a wage offer distribution $F(w)$. The worker accepts all wage offers exceeding the endogenously determined reservation wage, x_0. He continues job search otherwise. Once a job is accepted it will be held forever at the same wage. The instantaneous utility of the job search is equal to the unemployment benefit net of (exogenous) search costs, z. Let r and V_u, respectively, denote the discount rate and the lifetime utility of a job searcher. Then, in a stationary environment, the familiar returns-to-assets representation of Bellman's equation (see e.g. Mortensen 1986) is:

$$(3.1) \qquad rV_u = \max_{x_0}\left\{ z + l_0 \int_{x_0}^{\infty} \left(\frac{w}{r} - V_u \right) dF(w) \right\}$$

It can be easily shown that the optimal reservation wage is set at a level such that the lifetime utility of continuing search equals the lifetime utility of accepting a job at the reservation wage: $V_u = x_0^*/r$. The equilibrium condition for the reservation wage is therefore

$$(3.2) \qquad x_0^* = z + \frac{h_0}{r} E\left[w - x_0^* \middle| w \geq x_0^* \right]$$

where $h_0 = l_0 \overline{F}(x_0^*)$ is the hazard rate out of unemployment and $\overline{F}(.) \equiv 1 - F(.)$.

We now introduce a complication by allowing unemployed workers to participate in a training programme. In doing so, we assume that the participant in a training programme remains unemployed. This is consistent with the point of view of labour economists that a worker is unemployed until he/she is employed or leaves the labour market (cf. Lubyova and van Ours 1998, p.9)[15]. We identify four stages. In a first stage the unemployed worker decides whether he/she applies for participation in the training programme. Subsequently, if administrators retain the candidate trainee, the latter cannot enter training immediately, because the demand for training exceeds the number of available training slots. In this stage the worker must wait until a training slot becomes available. In the third stage, the worker is trained. Finally, if the worker didn't find a job meanwhile, he/she enters the post-training unemployment state. We solve by backward induction, first considering the last stage.

[15] Other researchers, Gritz (1993), Ham and Lalonde (1996), Bonnal et al. (1997) and Eberwein et al. (1997), treat training as a separate state distinct from unemployment. This distinction matters if the transition rate out of unemployment exhibits duration dependence, because with this assumption trainees who return to unemployment enter a fresh spell of unemployment, whereas, with the assumption in this research, they pursue the initial interrupted unemployment spell. This difference is not important when studying the long-term impacts, as the above-mentioned authors, but it is when concentrating on the short-run impacts.

In the post-training state the worker can fully benefit from the returns to training. We can distinguish between short-run and long run returns. In the short-run, training increases the job arrival rate and therefore the probability of leaving unemployment. In the long run, training may also increase the level of the wage offers and decrease the incidence and length of subsequent unemployment spells. In the review of the literature we reported that classroom training increases the employment probability and not so much the earnings or the length of employment spells. We therefore will consider the short run effect of training[16].

The returns to training, Δ, vary between unemployed workers. G(.) denotes the distribution of these returns for any population of workers entering unemployment[17]. Workers know their own returns. As we concentrate on the short run effect of training, the returns to training will just affect the job arrival rate. In fact, we assume that training increases this rate by a factor $\exp(\Delta)$.

Consider now the post-training state. If V_p and x_p denote, respectively, the lifetime utility and the reservation wage of an unemployed ex-trainee, then the following equation defines the equilibrium condition of the reservation wage for a worker with a return Δ:

(3.3) $$x_p^* = rV_p^* = z + \frac{h_{pe}}{r} E\left[w - x_p^* \middle| w \geq x_p^*\right]$$

where $h_{pe} = l_0 \exp(\Delta)\overline{F}\left(x_p^*\right)$ is the hazard rate out of unemployment after participation in a training programme.

The proportional impact of training on the transition rate out of unemployment is

(3.4) $$\begin{aligned} a_p(\Delta) &\equiv \ln\left[l_0 \exp(\Delta)\overline{F}\left(x_p^*(\Delta)\right)\right] - \ln\left[l_0 \overline{F}\left(x_0^*\right)\right] \\ &= \Delta + \ln\left[\overline{F}\left(x_p^*(\Delta)\right)\right] - \ln\left[\overline{F}\left(x_0^*\right)\right] \end{aligned}$$

This can in principle be negative, because an increase in the job arrival rate makes a trainee choosier in his job acceptance behaviour, i.e. $x_p^* > x_0^*$[18].

In the participation stage a trainee can continue searching for a job. We assume that participation in a training programme does not affect the search effectiveness of a worker.

---

[16] In fact, a positive long run effect of training tends to reduce the hazard rate out of unemployment. For, the trainee will anticipate this positive impact and be more selective in his/her job acceptance behaviour. In the empirical analysis below we find that the hazard increases in the post-training state, implying that, if there is a long run impact, it is certainly not dominant.

[17] As a consequence of sorting, the form of this distribution will evolve with unemployment duration.

[18] See Proposition A1 in the Appendix for proof of this.

The job arrival rate remains therefore l_0[19]. The transition intensity from training to employment is therefore

(3.5) $$h_{te} = l_0\overline{F}\left(x_t^*\right)$$

If V_t and x_t^* denote, respectively, the lifetime utility and the reservation wage of a trainee, then the equilibrium condition in the training stage is

(3.6) $$x_t^* = rV_t^* = z_t + \frac{h_{te}}{r} E\left[w - x_t^* \middle| w \geq x_t^*\right] + \frac{h_{tp}}{r}\left(x_p^* - x_t^*\right)$$

where z_t denotes the instantaneous utility of a trainee and $h_{tp} = g$ the exogenous transition intensity of a trainee back to unemployment, i.e. to the post-training state. A trainee remains entitled to unemployment benefits, but he also incurs costs such as the costs of effort, time and transportation. Part of these costs is reimbursed: the monetary transportation costs plus an allowance of 40 BEF (about 1 EURO) per trained hour to compensate for the costs time and effort. A participant is therefore likely to have a lower instantaneous utility than a non-participant, i.e. we expect $z_t \leq z$. The qualitative conclusions in which we are interested do not depend, however, on the level of z_t.

Consider now the proportional impact of participation in training on the transition rate out of unemployment:

(3.7) $$a_t(\Delta) = \ln\left[\overline{F}\left(x_t^*\right)\right] - \ln\left[\overline{F}\left(x_0^*\right)\right] < 0$$

Since one can show that $x_t^* > x_0^*$[20], a worker participating in training is always less likely to leave unemployment.

As to mimic the institutional environment as closely as possible, we assume that the demand for training is rationed by a limited supply of training slots. This means that a worker, whose application has been retained, must wait before he can participate in training. During this waiting period, the worker may continue searching for jobs. However, as shown below, he is less likely to accept job offers. If V_w and x_w denote, respectively, the lifetime utility

[19] One could argue that a participant cannot search as effectively. This would only reinforce the qualitative conclusions.
[20] See Proposition A.2 in the Appendix for proof of this.

and the reservation wage of a worker waiting to participate in training, then the equilibrium condition is

$$(3.8) \qquad x_w^* = rV_w^* = z + \frac{h_{we}}{r} E\left[w - x_w^* \middle| w \geq x_w^*\right] + \frac{h_{wt}}{r}\left(x_t^* - x_w^*\right)$$

where $h_{we} = l_0 \bar{F}\left(x_w^*\right)$ is the transition intensity from waiting to employment and h_{wt} is the transition intensity from waiting to participation in training. In a stationary environment, the latter transition intensity is implicitly defined by the following equality:

$$(3.9) \qquad h_{wt} p_w \equiv \bar{h}_t p_t$$

where $\bar{h}_{t.} = \bar{h}_{te} + h_{tp}$ is the average rate at which training is left for employment or the post-training state. p_t and p_w are respectively the number of training slots available and the number of unemployed workers waiting to enter a training programme, both in proportion to the active population. In a stationary environment, the number of workers entering training, the left-hand side of the equality in (3.9), must be equal to the number of workers leaving the training programme, the right-hand side of the equality in (3.9). Rationing implies that p_t is exogenous and depends on the budget allocated to training. On the other hand, p_w is endogenous and depends eventually on the level of p_t [21].

Finally, we consider the application stage. We assume that the application occurs instantaneously at the beginning of the unemployment spell. Once the worker has been rejected access to training, he does not re-apply. In reality, the application takes time and one should consider the job acceptance behaviour of the worker during this application period. However, this behaviour resembles the one of a worker waiting to enter training. We therefore ignore this aspect of the application stage.

A worker applies for training if the utility of applying, V_a, exceeds the utility of not doing so, $V_0^* = x_0^*/r$. The utility of applying is

$$(3.10) \qquad V_a(\Delta) = -c(\Delta) + p(\Delta)\frac{x_w^*}{r} + [1 - p(\Delta)]\frac{x_0^*}{r}$$

[21] See Lemma A.1 in the Appendix.

where $p(\Delta)>0$ is the probability that a worker with return Δ is selected for participation in training and $c(\Delta)$ is the application cost. The latter refers to the direct costs related to the preparation and the presentation of the test of basic abilities, and to the interview by the programme administrator. These costs include costs of effort and time, as well as monetary transportation costs. We assume that $c^{'}(\Delta) \leq 0$, reflecting that the test requires less effort and preparation time for workers with high returns. Moreover, $p^{'}(\Delta) \geq 0$, reflecting that workers with higher returns will be more successful in both the test and the interview.

Since $\partial x_w^* / \partial \Delta > 0$[22], V_a increases with the returns in training if $x_w^* > x_0^*$, i.e. $\partial V_a / \partial \Delta \geq 0$. Therefore there exists a unique threshold return, Δ^*, beyond which workers will apply. At this threshold an unemployed worker is indifferent between applying or not for training: $V_0^* = x_0^* / r = V_a^*$. If we insert this in (3.10) we obtain the following implicit equation for Δ^*:

$$(3.11) \quad x_w^*\left(\Delta^*\right) = x_0^* + r\frac{c\left(\Delta^*\right)}{p\left(\Delta^*\right)}$$

It follows that the reservation wage of the marginal worker who is waiting to enter training, $x_w^*\left(\Delta^*\right)$, is higher than the reservation wage, x_0^* of a worker who doesn't apply for training. Moreover, since $\partial x_w^* / \partial \Delta \geq 0$, this is true for all $\Delta \geq \Delta^*$. Hence, the speed at which unemployment is left is lower during the waiting period. The worker has become choosier, because if he/she accepts a job offer he/she foregoes the expected benefits of training:

$$(3.12) \quad a_w(\Delta) = \ln\left[l_0 \bar{F}\left[x_w^*(\Delta)\right]\right] - \ln\left[l_0 \bar{F}\left(x_0^*\right)\right] = \ln\left[\bar{F}\left[x_w^*(\Delta)\right]\right] - \ln\left[\bar{F}\left(x_0^*\right)\right] < 0$$

3.2. Identification of SATE and ATE

Up to this point we only considered the individual specific returns to training. These individual returns cannot be identified from the data. In this section we therefore discuss how we can identify the average treatment effects, SATE and ATE. SATE is the average effect of the treatment, conditional on being treated, i.e. $E\left[a_j(\Delta) \middle| participation\ in\ j\right]$ for $j = t, p, w$. ATE, by contrast, is the unconditional average treatment effect if all unemployed workers were to participate, i.e. $E[a_j(\Delta)]$, $j = t, p, w$. We show that variation in the number of treated as a proportion of the number of unemployed workers, i.e. the variation in p_j, identifies the distribution of the treatment effects, $H_j(.)$, in state j = t, p, w non-parametrically. If this distribution is identified, we can calculate expectations with respect to it and therefore identify SATE and ATE. We discuss consecutively identification of SATE and ATE for the states t, p and w.

We already mentioned that the demand for training is rationed by the supply of available training slots. Moreover, we argue below that the training capacity, i.e. p_t, varies exogenously between sub-regions. We now show that we can exploit this variation to identify $H_t(.)$ non-parametrically.

Proposition 1: Variation of p_t over the interval [0,1] identifies $H_t(.)$ non-parametrically, if and only if

$$\frac{\partial h_{wt}}{\partial \Delta^*}\left(x_t^* - x_0^*\right) > -\left(r + h_{w.}\right)\left(x_w^* - x_0^*\right)\left(\frac{p'(\Delta^*)}{p(\Delta^*)} - \frac{c'(\Delta^*)}{c(\Delta^*)}\right) - h_{wt}\frac{\partial x_t^*}{\partial \Delta^*} \quad \text{(P.1)}$$

Proof:

From (3.7) we have that

$$\text{(3.13)} \quad \frac{\partial a_t(\Delta)}{\partial \Delta} = -\frac{\overline{F}\left(x_0^*\right)}{\overline{F}\left(x_t^*\right)} f\left(x_t^*\right)\frac{\partial x_t^*}{\partial \Delta} < 0$$

where the inequality sign follows from $\partial x_t^* / \partial \Delta > 0$ [23]. It follows that $a_t(\Delta)$ is monotonically decreasing in ?. If and only if condition (P.1) is satisfied, we also have that $\partial \Delta^* / \partial p_t < 0$ [24]. Since these functions are monotonic their inverse exist. We may therefore write

$$\text{(3.14)} \quad p_t = \Delta^{*-1}\left\{a_t^{-1}\left[a_t\left(\Delta^*(p_t)\right)\right]\right\} \equiv H_t\left(a_t^*\right)$$

[22] See Proposition A.3 in the Appendix for proof of this.
[23] See (A.4) in Appendix A.
[24] See Lemma A.2 in the Appendix.

where $a_t^* \equiv a_t\left(\Delta^*(p_t)\right)$. Since $\Delta^{*-1}(.)$ and $a_t^{-1}(.)$ are both monotonically decreasing, $H_t(.)$ is monotonically increasing. Moreover, $0 \leq H_t\left(a_t^*\right) \leq 1$, because $0 \leq p_t \leq 1$ by definition and since $a_t(.)$ is a monotonically decreasing function, a_t^* is the upper bound of a_t for workers participating in training. This means that $H_t(.)$ defines the distribution function of a_t. Moreover, by varying p_t we can identify its form non-parametrically ?.

Intuitively, in a region with a high training capacity, the expected utility of application to a training programme is higher. Therefore the threshold value Δ^* beyond which an unemployed worker applies for training is lower than in a region with a low training capacity. In fact, the training capacity p_t is the probability that trainees have returns above this threshold value. Since participants with high returns have less interest to leave training prematurely for employment ($\partial a_t(\Delta)/\partial\Delta < 0$), p_t is also the probability that the effect of participation is below $a_t\left(\Delta^*\right)$, i.e. p_t is the distribution function of the effect of participation evaluated at $a_t\left(\Delta^*\right)$. By varying p_t we can therefore identify the distribution function of the effects.

If the returns to training increase it seems more attractive to enter training. However, there is one factor that makes entry more difficult: workers with high returns will set the reservation wage higher during participation in the training programme. Consequently, the participation state is left less rapidly for employment and therefore, in a stationary environment and for a fixed training capacity p_t, the rate at which training is entered is lower. This effect reduces the attractiveness of training. Condition (P.1) ensures that the latter effect does not dominate. It ensures that if the training capacity p_t increases, it is always more attractive for workers with lower returns to apply for training: $\partial\Delta^*/\partial p_t < 0$. Such a monotonic relationship is required to identify the distribution of the treatment effects from p_t.

In the empirical application below p_t varies over a limited range. We cannot therefore identify the full distribution of the treatment effects non-parametrically. This does not matter if the estimation of SATE is our sole concern. However, given the limited variability in p_t, we will impose a parametric form on this distribution. Such a parametric choice allows identifying ATE as well. We will test whether our findings are sensitive to the chosen parameterisation.

For the benchmark model in the empirical analysis below, we assume that $|a_t(\Delta)|$ has an exponential distribution with mean $|\overline{a}_t|$[25]. We take the absolute value, because, during participation, the impact of training on the transition rate out of unemployment is negative (see (3.7)). Consequently,

$$\text{(3.15)}\quad \Pr\left[a_t \le a_t^*\right] = \exp\left[-a_t^*/\overline{a}_t\right]$$

and therefore, by (3.14)

$$\text{(3.16)}\quad a_t^* = -\overline{a}_t \ln(p_t)$$

SATE is then

$$\text{(3.17)}\quad E\left[a_t(\Delta)\middle| participation\ in\ t\right] = -\frac{\int_{-\infty}^{-\overline{a}_t \ln(p_t)} a \exp\left[-a/\overline{a}_t\right] da}{\overline{a}_t p_t} = \overline{a}_t\left[1 - \ln(p_t)\right]$$

The exogenous variation of p_t across sub-regions identifies SATE and consequently ATE: $\overline{a}_t \equiv E\left[a_t(\Delta)\right]$.

By similar arguments we can identify the distribution of post-programme impacts and the corresponding SATE and ATE. However, as can be seen in the proposition below, identification requires an additional assumption.

Proposition 2: If condition (P.1) is satisfied and if $\partial a_p(\Delta)/\partial\Delta > 0$, the variation of p_p over the interval [0,1] identifies $H_p(.)$ non-parametrically.

Proof:

Since in a stationary environment the number of workers entering the post-training state must be equal to the number of workers leaving it, we have that

$$\text{(3.18)}\quad h_{tp} p_t = \overline{h}_{pe} p_p$$

[25] See Appendix D for the derivation of SATE and ATE if the distribution of the treatment effects is normal or uniform.

where $\bar{h}_{pe}$ is the average hazard to employment in the post-training state. Consequently,

$$\text{(3.19)} \quad \frac{\partial \Delta^*}{\partial p_p} = \frac{\partial \Delta^* / \partial p_t}{\partial p_p / \partial p_t} = \frac{\partial \Delta^*}{\partial p_t} \frac{h_{tp}}{\bar{h}_{pe}} < 0$$

where the inequality follows from condition (P.1)[26]. We can therefore write:

$$\text{(3.20)} \quad p_p = \Delta^{*-1}\left\{a_p^{-1}\left[a_p\left(\Delta^*(p_p)\right)\right]\right\} \equiv \bar{H}_p\left(a_p^*\right)$$

where $a_p^* \equiv a_p\left(\Delta^*(p_p)\right)$. Since $\Delta^{*-1}(.)$ is monotonically decreasing and $a_p^{-1}(.)$ is monotonically increasing by assumption, $\bar{H}_p(.)$ is monotonically decreasing. Moreover, $0 \le H_p\left(a_p^*\right) \le 1$, because $0 \le p_p \le 1$ by definition and since $a_p(.)$ is a monotonically increasing function, a_p^* is the lower bound of a_p for unemployed workers having participated in training. This means that $\bar{H}_p(.)$ defines the survivor function of a_p. Moreover, by varying p_p we can identify its form non-parametrically ? .

If the returns are exponentially distributed, SATE can then be shown to equal to

$$\text{(3.21)} \quad E\left[a_p(\Delta) \mid \textit{participation in } p\right] = \bar{a}_p\left[1 - \ln(p_p)\right]$$

where $\bar{a}_p \equiv E\left[a_p(\Delta)\right]$ defines ATE.

A problem in (3.21) is that the variation in p_p is not exogenous to the transition rate out of unemployment, even if p_t is. For, p_p will decrease as the transition rate out of unemployment, $\bar{h}_{pe}$, increases (see (3.18)). However, p_t is a good predictor of p_p and suggests using it as an instrument for p_p. This is what we will do in the empirical application.

Finally, we consider the pre-programme effect. A problem we face in the empirical analysis is that we cannot distinguish between non-participants who don't apply for training

[26] See Lemma A.2 in the Appendix for proof of this.

and those who are waiting to enter the programme. Since we do not know the proportion of workers waiting, we cannot identify the average pre-programme effect as above.

Consider the aggregate transition rate out of unemployment of non-participants at the time of entry in unemployment:

$$h_{ne} = h_0 - \int_{\Delta^*}^{\infty} g(\Delta)p(\Delta)[h_0 - h_{we}(\Delta)]d\Delta \tag{3.22}$$

Since $x_w^* > x_0^*$, we have that $h_0 > h_{we}(\Delta), (\forall \Delta : \Delta \geq \Delta^*)$ and therefore $h_{ne} < h_0$. If in the empirical analysis we assume that $h_{ne} = h_0$, we therefore over-estimate the higher mentioned treatment effect. Because only small proportion (2.1% of all observed unemployment spells) of unemployed workers is trained, this bias is, however unlikely to be important. In principle, we can test whether the negative pre-participation programme effect is important. For, if condition (P.1) is satisfied, h_{ne} decreases with p_t:

$$\frac{\partial h_{ne}}{\partial p_t} = p(\Delta^*)g(\Delta^*)[h_0 - h_{we}(\Delta^*)]\frac{\partial \Delta^*}{\partial p_t} - \int_{\Delta^*}^{\infty} p(\Delta)g(\Delta)l_0 f(x_w^*)\frac{\partial x_w^*(\Delta, p_t)}{\partial p_t}d\Delta < 0 \tag{3.23}$$

For, it can be shown that $\partial\Delta^*/\partial p_t < 0$ and $\partial x_w^*(\Delta, p_t)/\partial p_t > 0$ [27]. The aggregate transition rate out of unemployment decreases with the proportion of trainees[28], because the larger number of training slots increases the expected returns to training. This in turn increases therefore both, the number of candidate trainees (the first term) and the reservation wage of those who are already waiting for participation (the second term). In the empirical analysis we will therefore allow h_{ne} to depend on p_t.

---

[27] See Lemma A.1 and A.2 in the Appendix for proof of this.

[28] We only derived the effect on the aggregate transition rate at the time of entry. If the individual effects enter proportionally, this effect will tend to zero as duration increases (see Lancaster 1990, p.65).

4. The Data and the Statistical Model

4.1. The Data

We analyse administrative data on unemployment spells beginning[29] between May 1989 and March 1993 in Wallonia, the French speaking Region of Belgium. We consider all officially registered full-time unemployed workers. On the one hand, 23% of these spells consist of school-leavers entitled to unemployment benefits after a waiting period of 6 months. The remainder of these spells are of workers entitled to benefits through their past work experience[30]. In the sequel, the school-leavers will be referred to as the 'young' and the other group as the 'old'. In principle, the entitlement duration is indefinite in Belgium. There is one important exception to this rule[31]. Cohabitants, who are not a head of a family, can loose entitlements on grounds of "excessive" unemployment duration. This duration is, however, well beyond the maximal duration, i.e. 28 months, considered in the empirical analysis below.

The data are grouped into monthly intervals. However, the recorded transitions of school-leavers are not reliable throughout the waiting period, i.e. throughout the first 6 months. We therefore need to group the first 6 months in a single interval. Moreover, for purposes of a symmetric treatment, we also group the data of the 'old' unemployed workers in a similar way. Transitions within a monthly interval are recorded, but the timing is not. Unfortunately, as a consequence of administrative errors there is a proliferation of movements within the month, rendering this information unreliable. Another flaw of the data is that the destination for which unemployment is left is not known. We therefore cannot determine whether a worker leaves unemployment for a job or whether he/she has left the labour force. This can also partly explain why we observe a high proportion of very short unemployment spells in the data. For, an unemployed worker who is ill, is not entitled to unemployment benefits, but to sickness benefits provided by another administration.

In order to avoid the measurement of spurious exits and exits out of the labour force, we follow Plasman (1993,1994) by imposing that an exit should last more than two months to be recorded as such. Moreover, we only consider workers younger than fifty at the start of the

[29] We do not sample workers who are unemployed at the beginning of the observation period. As the date at which their spell began is not known, these spells are left censored. It is well known that the analysis of left censored spells would complicate the statistical analysis substantially.

[30] Workers are entitled to unemployment benefits if they have been employed for at least 75 days within a prescribed period prior to their claim if younger than 18 years, and up to 600 days if older than 50 years (Van Langendonck 1991, p.450)

[31] Leaving out administrative sanctions, such as imposed on workers refusing job offers or declaring false information.

unemployment spell. In that way, we eliminate a large share of the withdrawals from the labour force. Besides, for obvious reasons, older workers hardly participate in training.

We can identify the months in which a worker participated in a training programme organised or subsidised by the Walloon employment agency, FOREM. We do not know, however, whether the trainee completed the programme successfully or whether he/she dropped out. Neither are we informed on the nature of training offered.

Unemployed workers may also participate in training programmes organised on-the-job or by other institutions not subsidised by FOREM. Unemployed workers participating in such programmes are assimilated with non-participants. This contaminates our "control group". However, in view of the marginal number of workers involved in such programmes as a proportion of the total number of unemployed (2.1% in 1991), this will not significantly bias our results.

Table 1 summarises some descriptive statistics on the population retained. The population consists of 1,361,660 spells. One individual may have experienced several periods of unemployment during the observation period. This means that there haven't been as many individuals entering unemployment during this period. 23% of this population is 'young'. For both, the young and the old, Table 1 provides statistics with respect to all spells and with respect to spells during which some time was spent in training. According to these statistics trainees are higher educated. The average age of trainees is not very different from any unemployed worker. Among the old, males are more likely to participate in training. Among the young, it is rather the women who are over-represented. Old workers who have been unemployed previously also have a slightly higher likelihood of programme participation.

INSERT TABLE 1 APPROXIMATELY HERE

The administration of FOREM is decentralised in 11 sub-regional departments. For purposes of the empirical analysis it is important to group these sub-regions in ones that are sufficiently homogenous with respect to the mix of training schemes offered. We therefore only report statistics on the 7 re-grouped sub-regions.

Median unemployment duration is 3 months, both for the young and the old. This is not very long, in view of Belgium's reputation of being one of the OECD countries with the highest share of long-term unemployment. However, the latter statistics are based on stock samples suffering from the well known "length bias" (Salant 1977). Median unemployment duration is much longer if we consider spells with some time in training: 9 months for the old and 12 months for the young. This may have several explanations. First, very few workers enter a training programme immediately when entering unemployment. The median duration until a worker enters a training programme is 3 months for the old and 6 months for the young. This may be the consequence of both, the rationing of the demand for training and the time required for an unemployed worker to be convinced that training may increase his/her job finding rate. Second, if the transition out of unemployment exhibits negative duration dependence, trainees will experience longer spells as a consequence of the elapsed unemployment duration at the moment of programme entry. Finally, trainees may be a selective sub-population of the unemployed. These explanations will be disentangled in the empirical analysis below. Note that the median length of a training spell is 2 months.

4.2. The Statistical Model

We now develop the statistical model on the basis of which we will estimate the effect of participation in training on the transition intensity out of unemployment. We choose Minimum Chi-Square (MCS) as the method of estimation (cf. Cockx 1997 and Cockx and Ridder 2000). This method has two main advantages. First, it does not require strong assumptions on the duration dependence[32] and the distribution of unobserved heterogeneity, so that estimates are insensitive to specification errors. Second, since it transforms the transition model to a linear regression model, we can use simple methods of aggregation to correct for the selection on unobservables. Cockx and Ridder (2000) develop a grouping/IV estimator within this framework. We propose a control function estimator (Heckman and Robb 1986, p. 172).

The MCS method requires a grouping of the data not only by duration, but also by explanatory variables. A drawback of this method is that it is difficult to take fixed unobserved individual heterogeneity within the groups into account. In order to allow for heterogeneity in the returns to training, we therefore assume that these returns are not fixed, but can vary over an unemployment spell. At the beginning of each duration interval defined below an unemployed worker draws a return from this distribution independently from the drawings in the previous duration intervals.

In order to reduce biases induced by within-group heterogeneity, we choose, in the benchmark model, to analyse only spells with an elapsed duration of 6 months or more. As such, the sorting process has already rendered groups much more homogenous. Afterwards we will compare the benchmark model with one in which data on the first 6 months are included. We will also test whether within-group heterogeneity can indeed be neglected in the benchmark model.

There is another reason to analyse only spells with a certain elapsed duration. We argued in the theoretical model, if the returns to training are heterogeneous, that only workers

[32] The baseline transition intensities are assumed to be piece-wise constant.

with the highest returns participate in training. The reason is that there exist a number of barriers in the application process implying that entering a training programme is uncertain, costly and takes time[33]. Nevertheless, we observe that within the group of the "old" unemployed workers, there are workers who are trained immediately after entering unemployment. These participants in training cannot have followed the same application procedure or they must have done so while they were still employed. For instance, these trainees may be displaced workers for whom the participation in training may have been a condition negotiated by unions to accept a collective dismissal of employees in the framework of the restructuring of a firm. However, if such an alternative selection procedure applies to these types of workers, the average programme effects are not necessarily the same.

The indicator variable of the training status is time varying. This means that we need to make assumptions on the evolution of the indicator variable within the duration intervals of the grouped data. We follow the solution proposed by Cockx and Ridder (2000). Define the following three states:

1. Unemployment without participation in training
2. Unemployment with participation in training
3. Out of unemployment

If one is willing to assume that the base-line transition intensities are constant within the intervals (as in Prentice and Gloeckler 1979) and that, within a duration interval, there is at most one transition between the states defined above, then Cockx and Ridder show that the treatment of a time-varying indicator variable boils down to jointly analysing two competing risks models, one with origin state 1 and destination states 2 and 3, and another one in which state 2 is the origin state and states 1 and 3 the destinations. State 3 is an absorbing state, because we can no longer observe a worker who left unemployment. In the sequel we will denote origin and destination states by superscripts u and v respectively.

The MCS method requires the data to be grouped in homogeneous cells. We define these cells by crossing four criteria: the elapsed unemployment duration, k[34], the sub-region, m, the eligibility criterion to unemployment benefits ('old' or 'young'), s, and the training status, (u,t). The training status can be one of the following three:

1. (u,t)=(1,0): no participation in training, neither presently or in the past
2. (u,t)=(2,1): presently participating in training

[33] For simplicitly, we did not model that the application process takes time, but we argued that it does in reality.
[34] The data are assumed homogeneous over calendar time.

3. (u,t)=(1,1): past participation in training

The data are grouped in K+1 intervals $[t_0, t_1), ..., [t_{k-1}, t_k), ...[t_{K-1}, t_K), [t_K, t_{K+1})$ of (possibly unequal) length $\Delta_k = t_k - t_{k-1}$ with $t_0 = 0$ and $t_{K+1} = \infty$. We assume the following proportional specification of the transition intensities between the states:

$$\text{(4.1)} \qquad \log\left(h_{kmts}^{uv}\right) = g_k^v + b_m^v + y_s^v + l^{uv} p_{km1s}^2 + \bar{a}_t^{uv}\left[1 - \log\left(p_{kmts}^u\right)\right] + e_{kmts}^{uv}$$

where $u = 1, 2;\ v = 1, 2, 3 \neq u;\ k = 1, 2, ..., K;\ m = 1, 2, ..., M;\ t = 0, 1;\ s = o, y$. The duration effect g_k^v, the sub-regional effect b_m^v and the effect of the eligibility status y_s^v depend only on the destination state v. We normalise the type effects and set $b_1^v = y_o^v = 0$.

In order to allow for heterogeneity in the treament effects, we allow, as derived in Section 3, the transition intensities depend on p_{km1s}^u, $u = 1,2$, i.e. on the proportion[35] of training slots during participation (u=2) and after participation (u=1) in segment (k,m,s). First, in Section 3.1 we predicted that unemployed workers waiting to enter a training programme increase their reservation wage. This causes the aggregate transition rate out of unemployment for non-participants to decrease with the training capacity (see (3.23)): $l^{13} < 0$. By definition, $l^{2v} = 0, v = 1,3$. Second, if we assume that the impacts of training are exponentially distributed in the population, (3.14) and (3.21) provide expressions of SATE, respectively during and after participation in training. We have inserted these expressions in (4.1). By definition, $\bar{a}_0^{12} = \bar{a}_0^{21} = 0$ and by normalisation, $\bar{a}_0^{13} = 0$. The parameters of interest are the average effect of training in the population during participation $\bar{a}_1^{23}$ and the average post-training effect in the population a_1^{13}.

Finally, e_{kmts}^{uv} are the unobserved group effects. Note that these reflect the between-group heterogeneity and not the within-group heterogeneity. Selection on unobservables means that these unobserved group effects are correlated with the indicator variables of participation in training.

Under these assumptions the transition probability P_{kmts}^{uv}, i.e. the conditional probability of a transition of an individual of type (m,t,s) from state u to state v in duration interval k, is equal to

[35] This proportion is calculated to the total number of unemployed workers and not to the active population.

(4.2) $$P_{kmts}^{uv} = \frac{h_{kmts}^{uv}}{\sum_{w=1,w\neq u}^{3} h_{kmts}^{uw}} \left[1 - \exp\left(\sum_{w=1,w\neq u}^{3} h_{kmts}^{uw} \Delta_k\right)\right], \quad u = 1, 2 \; and \; v = 1, 2\; 3 \neq u$$

This expression maps transition intensities to transition probabilities. The inverse mapping from transition probabilities to transition intensities is

(4.3) $$z_{kmts}^{uv} \equiv \log\left[\frac{-P_{kmts}^{uv} \log\left(P_{kmts}^{uu}\right)}{\Delta_k \left(1 - P_{kmts}^{uu}\right)}\right] = \log\left(h_{kmts}^{uv}\right), \quad u = 1, 2 \; and \; v = 1, 2\; 3 \neq u$$

with $P_{kmts}^{uu} \equiv 1 - \sum_{w=1,w\neq u}^{3} P_{kmts}^{uw}$ is the complementary probability of staying in the origin state u during the kth interval.

Let r_{kmts}^{u} be the number of individuals of type (m,t,s) in state u at the start of duration interval k, i.e. at time t_{k-1}, and let q_{kmts}^{uv} of these individuals make a transition from state u to v in the k-th duration interval, then we estimate P_{kmts}^{uv} by

(4.4) $$\hat{P}_{kmts}^{uv} = \frac{q_{kmts}^{uv}}{r_{kmts}^{uv}}, \quad u = 1, 2 \; and \; v = 1, 2\; 3 \neq u$$

Upon substitution of these estimates in the middle expression of (4.3), we obtain $\hat{z}_{kmts}^{uv}$. If in (4.3) we replace the transition probabilities by their estimates, then the second equality does not hold exactly. However, a Taylor series expansion around P_{kmts}^{uv} yields

(4.4) $$\hat{z}_{kmts}^{uv} = \log\left(h_{kmts}^{uv}\right) + w_{kmts}^{uv}$$

with $w_{kmts}^{uv} \equiv \sum_{w=1,w\neq u}^{3} b_{kmts}^{uvw}\left(\hat{P}_{kmts}^{uw} - P_{kmts}^{uw}\right)$[36]. We can ignore the remainder of the Taylor expansion as it can be shown that its probability limit, for r_{kmts}^{u} tending to infinity, converges to zero at a faster rate than w_{kmts}^{uv} (Amemiya 1985, p.276-77). Hence, the omission of the remainder does not affect the consistency of the estimator nor its asymptotic distribution.

[36] b_{kmts}^{uvw} is defined in Appendix C.1.

Conditional on r^u_{kmts} the random vector of the number of transitions of individuals of type (m,t,s) from u to states $v \neq u$ in the k-th interval has a multinomial distribution with parameters r^u_{kmts} and P^u_{kmts}, which is the vector of transition probabilities from state u to states $v \neq u$ in interval k for type (m,t,s). The errors of the system of regression equations (4.4) have mean zero, but are heteroskedastic and correlated.[37]

Upon substitution of (4.1) in (4.4) we obtain a linear regression model. More precisely, we obtain for each origin state $u = 1, 2$ two regression equations that correspond to the two destination states $v = 1, 2, 3 \neq u$

(4.5) $$\hat{z}^{uv}_{kmts} = g^{v}_{k} + b^{v}_{m} + y^{v}_{s} + l^{uv} p^{2}_{km1s} + \bar{a}^{uv}_{t}\left[1 - \log\left(p^{u}_{kmts}\right)\right] + e^{uv}_{kmts} + w^{uv}_{kmts},$$

$u = 1, 2$ *and* $v = 1, 2\ 3 \neq u$. In the sequel we use the notation $u^{uv}_{kmts} = e^{uv}_{kmts} + w^{uv}_{kmts}$.

We now proceed in three steps. First, we assume that the unobserved group effects, e^{uv}_{kmts}, are not correlated with the training status, i.e. we allow only for selection on observables. Next, we allow for the endogeneity of the proportion of workers in the post-training state, p^{1}_{km1s} (cf. (3.18)). Finally, we propose a control estimator controlling for selection on unobservables.

The benchmark assumption is that the unobserved group effects are not correlated with the training status:

(4.6) $$E\left(e^{uv}_{kmts} \middle| k, m, s\right) = 0$$

The unobserved variables corresponding to different duration classes, types, and origin states are assumed to be independent. The unobserved variables corresponding to the different destination states may be correlated.

(4.7) $$E\left(e^{uv}_{kmts} e^{u'w}_{k'm't's'}\right) = d_{kk'} d_{mm'} d_{ss'} d_{uu'} d_{tt'} s^{uvw}$$

where $d_{..}$ is the Kronecker delta. Note that the disturbances w^{uv}_{kmts} have the same pattern of zero correlation. Hence, for the origin state u and for duration interval and type (k,m,t,s), the distribution of the 2-vector disturbances of the two regression equations (4.5) with

[37] The variance-covariance matrix is given in Appendix C.1.

$v = 1, 2\ 3 \neq u$, $u^{u}_{kmts} \equiv \left[u^{uv}_{kmts}\ u^{uw}_{kmts}\right]$, which is the sum of the unobserved group effect and the approximation error, has a variance-covariance matrix with typical element $s^{uvw} + s^{uvw}_{kmts}$, with $v,\ w = 1, 2\ 3 \neq u$. We denote this 2x2 matrix by V^{u}_{kmts}. This completes the specification of the regression equations.

The regression equations are heteroskedastic and have correlated disturbances. The parameters in regression equation (4.5) can therefore be efficiently estimated by Generalised Least Squares (GLS). We follow the two-step estimation procedure proposed by Amemiya and Nold (1975) for a logit model. Because we have grouped data, we can use a c^{2}-goodness-of-fit test to evaluate the specification of the model. This statistic can be interpreted as a GMM test for the null hypothesis of error-regressor orthogonality and therefore of selectivity bias in the training impact.[38]

In Section 3.2. we argued that the proportion of workers in the post-training state is endogenous. In the regression equation (4.5) we therefore instrument $\left[1 - \log\left(p^{1}_{km1s}\right)\right]$ by $\left[1 - \log\left(p^{2}_{km1s}\right)\right]$, i.e. by replacing this proportion by the fraction actually participating in training. The estimator is then a Generalised Instrumental Variable (GIV) estimator instead of a GLS.

We now allow for the unobserved group effects e^{u3}_{kmts} for destination state 3 to be correlated with the training status. Without loss of generality, we can write the unobserved group effects for destination state 3 in the following way:

$$(4.8) \qquad e^{u3}_{kmts} = h_{k} + V_{m} + k_{s} + r^{u}_{t} + z^{u3}_{kmts}$$

where we assume

$$(4.9) \qquad E\left(z^{u3}_{kmts}\right) = 0, \qquad E\left(z^{u3}_{kmts} z^{u'3}_{k'm't's'}\right) = d_{kk'} d_{mm'} d_{ss'} d_{uu'} d_{tt'} \tilde{s}^{uvw}$$

This means that the regressors of Equation (4.5) may depend on the unobserved group effects in duration interval k, h_{k}, in sub-region m, V_{m}, of eligibility type s, k_{s} and of the training status (u,t), r^{u}_{t} , but not on their interaction z^{u3}_{kmts}.

The first three terms in (4.8) can be neglected by adding them to the corresponding regressor coefficients, i.e. g^{v}_{k}, b^{v}_{m}, y^{v}_{s} respectively. This will bias the estimators of these

---

[38] The details of the estimation procedure can be found in the Appendix C.2 and in Cockx and Ridder (2000).

coefficients, but since these are not the parameters of interest, this does not matter. Aggregation of the unobserved group effects over the training status and taking expectations therefore yields

$$\text{(4.10)} \quad \sum_{t=0,1} \left[f^{1}_{kmts} E\left(e^{13}_{kmts}\right)\right] + f^{2}_{km1s} E\left(e^{23}_{km1s}\right) = 0$$

where f^{u}_{kmts} is, for any given (k,m,s), the population fraction of spells of training type (u,t). This equation provides the identifying restriction of the treatment effects. Note that it uses (4.9), implying that there may not be any systematic interaction effects. It allows us to express the mean bias of non-participants in training in terms of the mean bias of participants:

$$\text{(4.11)} \quad E\left(e^{13}_{km0s} \middle| k, m, s\right) \equiv r^{1}_{0} = -\frac{f^{1}_{km1s}}{f^{1}_{km0s}} r^{1}_{1} - \frac{f^{2}_{km1s}}{f^{1}_{km0s}} r^{2}_{1}$$

for all (k,m,s). Because we observe all unemployment spells starting between May 1989 and March 1993, the population fractions f^{u}_{kmts} are known exactly. The aggregation in (4.10) therefore does not result in an errors-in-variables problem as in Deaton (1985). If we impose the KxMx2 restrictions implied by (4.11) on the data, we can estimate r^{u}_{t}. The mean selection bias r^{u}_{t} identified as such, is a control function which, inserted in the regression equation (4.5), purges the equation of the covariance between e^{u3}_{kmts} and $\bar{a}^{u3}_{t}$:

$$\text{(4.12)} \quad \hat{z}^{uv}_{kmts} = g^{v}_{k} + b^{v}_{m} + y^{v}_{s} + l^{uv} p^{2}_{km1s} + \bar{a}^{uv}_{t}\left[1 - \log\left(p^{u}_{kmts}\right)\right] + r^{u}_{t} + z^{uv}_{kmts} + w^{uv}_{kmts},$$

$u = 1, 2 \; and \; v = 1, 2 \; 3 \neq u$. The control function estimator of the parameters of (4.12), restricted by (4.11) and taking the endogeneity of $\left[1 - \log\left(p^{1}_{km1s}\right)\right]$ into account, is therefore a GIV estimator. The feasible GIV is a two step estimator. The estimation procedure is the one proposed for the estimation of (4.5), where r^{u}_{t} are additional parameters and where we replace the OLS estimator in the first stage by an IV estimator. In the benchmark model we will impose the additional restriction that, conditional on the observable variables (k,m,s), the bias of participating trainees is on average equal to the post-training bias: $r^{2}_{1} = r^{1}_{1}$. [39]

[39] Such a restriction cannot be imposed on the grouping/IV-estimator of Cockx and Ridder (2000).

Finally, in the beginning of this sub-section we mentioned that the control function estimator of the impacts of training is biased if within-group unobserved heterogeneity is important. Within-group heterogeneity will generate an interaction between duration and the other explanatory variables. For, negative duration dependence induced by sorting is more important for heterogeneous groups with a high average propensity to leave unemployment than for groups with a low average propensity[40]. In a proportional hazard specification such an interaction is absent. Consequently, if within-group heterogeneity is important, groups with a high propensity to leave unemployment will systematically leave unemployment at a lower rate than the one predicted by the proportional hazard model. Groups with a low propensity will leave a higher rate than predicted. These deviations should therefore generate, for any given group, positive auto-correlation in the residuals over duration.

We therefore follow Cockx and Ridder's (2000) suggestion and calculate a Durbin-Watson (d) test statistic that accounts for the panel structure of the data. If e^{uv}_{kmts} denote the weighted residuals corresponding to one of the estimated regression models, then

$$(4.13)\quad d = \frac{\sum_{u=1}^{2}\sum_{v=1\neq u}^{3}\sum_{m=1}^{M}\sum_{t=0}^{1}\sum_{s=o}^{y}\sum_{k=3}^{K}\left(e^{uv}_{kmts} - e^{uv}_{k-1mts}\right)^2}{\sum_{u=1}^{2}\sum_{v=1\neq u}^{3}\sum_{m=1}^{M}\sum_{t=0}^{1}\sum_{s=o}^{y}\sum_{k=2}^{K}\left(e^{uv}_{kmts}\right)^2}$$

Asymptotically we have that $d \sim N\left(2, 4/\left(12M(K-1)\right)\right)$.

4.3. Over-Identifying Restrictions: Exogenous Sub-Regional Variation in the Training Supply

In regression equation (4.12) the treatment effects are identified on the assumption that, after aggregation, no systematic interactions between the treatment and the other explanatory variables exist. In this subsection we argue that we can exploit a natural experiment to test for this assumption. This natural experiment generates sub-regional variation in the participation rates that is not, directly or indirectly, caused by the sub-regional variation in the transition rates out of unemployment. Consequently, we need not control for the sub-regional effects in the regression, i.e. we may impose the over-identifying restrictions $b^3_m = 0$ for m = 2, …M.

[40] Abbring, van den Berg and van Ours (1994) contrast an economic upturn, when workers have a high propensity to leave unemployment, to a downturn. They prove that unobserved heterogeneity indeed generates a more rapid decline of the hazard in an upturn than in a downturn.

These restrictions enhance the efficiency of the estimator without bias if there is indeed no systematic correlation between the sub-regional participation rate in training and the rate at which both trainees and non-trainees leave unemployment in a region. The validity of this assumption can be checked by a Hausman (1978) test.

The natural experiment is justified in two steps. First, we argue that the participation rate is completely determined by supply. Second, we claim that the sub-regional variation in training capacity is exogenous. The rationing of demand for training by the supply is an important stage in our argument. For, suppose that the demand were not rationed. Then, if the demand for training is correlated with the transition rate out of unemployment, variations in the transition rate imply variations in the participation rate in training. If demand exceeds supply, variations in the transition rate will only affect the length of the queue of workers waiting for training and not the participation rate.

Interviews of employees of FOREM revealed that hardly any training programme had problems in filling the available training slots during the 1989-1993 period. Objective statistics on the importance of rationing are only partially available. We do have information on the importance of rationing of vocational training for secondary sector jobs. These represent roughly one third of the available supply of training slots. For these programmes FOREM (1991) reports the average number of workers, who have applied and are waiting either for a decision in the application procedure, or for a training slot to become available. If we report these figures in proportion to the average number of trainees entering a programme in the secondary sector each month, then we obtain a ratio of 4.1 in 1989, 5.4 in 1990, and 7.6 in 1991. This confirms that rationing is important. Note that the demand for training seems to increase significantly when labour market conditions worsen.

As a consequence of rationing, variation in the budget of the local training centres determines variation in participation rates. We mentioned higher that this budget is in no direct way related to placement ratios and other labour market outcomes. Interviews of employees of the employment agency, FOREM, confirm that the sub-regional variation in training budgets does not depend on objective criteria, but rather reflects the balance of political power of the respective sub-regions. There is one major exception to this rule. In Arlon the number of training slots in proportion to the number of unemployed is much higher than in other sub-regions (see Table 1). This is a consequence of the economic restructuring of this region. For, in response to the crisis in the steel industry, the region has benefited from important subsidies of the European Structural Funds, among which subsidies to develop training initiatives. We will therefore test below whether the results are sensitive to setting the regional effect of Arlon to zero, i.e. to $b_6^3 = 0$.

Even if there is no direct effect, the sub-regional level of the transition rate out of unemployment could have indirectly influenced the number of available training slots. In order to preclude this possibility, we examined whether variations in training capacities over time were in any way systematically related to the sub-regional state of the labour market. We first considered this question for the 1989-1993 observation period. In a second step, we analysed this relationship for the 1980-1988 period. This is important, because the levels in the observation period of the variables under consideration are largely determined by the past. From this analysis[41] we conclude that there is no indication of a systematic relationship between the regional training capacity and unemployment over time.

5. The Estimation Results

5.1 The Benchmark Model

In Table 2 we report the estimation results of the benchmark model in which we did not correct for selection on unobservables. Durations are measured in months. For the duration intervals we have chosen {[6,8), [8,10), [10,12), [12,15), [15,18), [18,22), [22,28)}. Note that we have left out the first duration interval for reasons mentioned in sub-section 4.3.

INSERT TABLE 2 APPROXIMATELY HERE

We find negative estimates for the variances of the unobserved group effects for the transitions from origin state 2 to the two destination states (v = 1, 3). Following Parks (1980, p.299, footnote 5), we set these (co-)variances equal to zero. We only find unobserved group effects for the first origin state (u=1).

If these unobserved group effects are correlated with the training status, the estimates of the training impact are biased. However, according to the GMM test the null hypothesis of regressor-error orthogonality cannot be rejected at a significance level of 16%. We will verify below whether this claim is confirmed when we allow for selection on unobservables.

The impacts of training are the following. While participating in training the average unemployed worker reduces his/her transition rate out of unemployment by 7% (= 1 - exp(-.07)). After he/she completed training the transition rate increases by 11% (= exp(.104) – 1). Both effects are very significantly different from zero. However, these effects are

[41] See Appendix B for a more detailed discussion.

extrapolations for the population (=ATE), on the assumption that the treatment effects are exponentially distributed. In reality only a small minority of the unemployed workers participate in training: 2.1% of the spells starting in the 1989-93 period. The average effect of training on the treated (SATE) is therefore more reliable. We calculate that during participation the transition intensity decreases by 27%[42], while the post-training effect increases this transition rate by 62%[43]. These results suggest that training programmes can already yield sizeable positive effects in the short-run, within the unemployment spell in which one is trained.

In the theoretical analysis we predicted that a training programme could lengthen the spell of unemployment of non-trainees. In particular, if demand for training is rationed, sub-regions with a higher training capacity should have both, a higher proportion of unemployed workers waiting to participate in training and a higher reluctance of these workers to accept job offers. The estimated coefficient, l^{13}, has the correct sign, but is estimated very imprecisely. The data do therefore not allow evaluating the importance of this adverse impact.

Finally, on the basis of the Durbin-Watson statistic we cannot reject the hypothesis that duration is positively auto-correlated. We suggested that this is sign of within-group heterogeneity. In that case, the training effects are likely to be biased downwards in absolute value (see Lancaster 1990, p.65). A model that takes heterogeneity into account is matter for further research, however.

INSERT TABLE 3 APPROXIMATELY HERE

In the first column of Table 3 we report the results of the GIV estimator, correcting for selection on unobservable between-group effects. We only provide information on the parameters of interest, the mean selection bias, and the GMM statistic[44]. We did not report the other estimates, because the point estimates of all the coefficients are nearly identical. The coefficient corresponding to the mean selection bias, $r_1^1 = r_1^2$, is small and insignificant, indicating that there is no significant selection on unobservables. This is in accordance with the finding that the preceding model, reported in Table 2, could not be rejected on the basis of the GMM test statistic. Nevertheless, according to the Hausman test statistic (=19.48), distributed c^2 with two degrees of freedom, the two impact effects of

[42] This corresponds to an average coefficient equal to -.313. The smallest impact in the treated population is -.377 and the largest is -.250.
[43] This corresponds to an average coefficient equal to .480. The smallest impact in the treated population is .340 and the largest is .702.
[44] The complete results can be obtained from the authors on request.

training, $\bar{a}_1^{u3}, u = 1,2$, of the asymptotically more efficient benchmark model reported in Table 2 are jointly to be rejected against the consistent GIV estimates (P-value=0.000).

The standard errors of coefficients measuring the effect of training increase considerably. The participation effect is no longer significantly different of zero. The post-training impact remains significantly positive. The relative imprecision of the estimates calls for a sensitivity analysis of these results.

5.2 Sensitivity Analysis

In Section 4.3 we argued that the training capacity varies exogenously between the sub-regions. The second column of Table 3 reports the GIV estimates setting the regional effects to zero: $b_m^3 = 0$ for m = 2, …M. According to the Hausman test statistic (= 21.44) these over-identifying restrictions should be rejected. However, we argued that the natural experiment might not apply to the sub-region Arlon. The third column of Table 3 shows how the results are affected if all, but the regional effect of Arlon, are set to zero. The point estimates are now much closer to the ones of benchmark GIV model. The Hausman test statistic (= 2.62) allows no longer a rejection of equality of the impact effects of training at a P-value of 27%. This provides support for the validity of our identifying assumptions.

INSERT TABLE 4 APPROXIMATELY HERE

In column (4) of Table 3 we allow the mean bias to be different during and post participation: $r_1^1 \neq r_1^2$. We mentioned that such a specification is very likely to encounter problems of identification. These problems are reflected in the size of the standard errors. The post-training effect is remarkably stable. The effect during participation is much less negative. According to the Hausman test statistic we cannot, however, reject (P-value = 1) that these effects are equal to the ones reported for the benchmark GIV model.

We estimated the benchmark model including the first duration interval and allowing for the effect of training to be different during this first interval. The findings, reported in column (5) of Table 3, point to an effect of training that is indeed significantly different in the first duration interval. The other parameter estimates are hardly affected. The model is by and large rejected on the basis of the GMM test statistic. Its value is equal to 666.74 for 536 degrees of freedom. This rejection is likely to be induced by within-group unobserved heterogeneity, inducing considerable non-proportionality in the first duration interval.

In Table 4 we report estimates allowing for the treatment effects to vary with duration classes and with the eligibility status. The young participants post-programme effect is smaller and the one during participation more negative. During participation the effect does not vary significantly with duration, apart from the last duration interval in which it the effect is no longer negative. The post-programme effect is significantly higher from the fifth duration class onwards. Nevertheless, we must treat these results cautiously. For, on the basis of the GMM test statistic, we must reject both models against the saturated model at conventional significance levels. This suggests that these explanatory variables be correlated with unobserved group effects. On the other hand, noteworthy is that we do not observe the absolute value of these impacts to trend down towards zero with duration. This suggests that the within-group unobserved heterogeneity is unimportant (Lancaster 1990, p.64-65). However, if the treatment effect increases with duration, this may conceal the downward trend.

INSERT TABLE 5 APPROXIMATELY HERE

We also tested whether it is important to instrument the proportion of post-trainees in the control function estimator. If we base the method of estimation on Generalised Least Squares (GLS) rather than on GIV, then column (6) of Table 3 shows that the training effects drop significantly. The mean selection bias pointing to significant creaming is in line with these findings. According to the Hausman (1978) test statistic we should reject at a 3.6% level of significance the null hypothesis that the proportion of post-trainees is exogenous. We conclude that it is essential to base the analysis on an IV estimator. Unfortunately, this considerably reduces the precision of the parameter estimates.

Up to this point we have assumed a variable treatment effect in the population. We may question whether a constant treatment effect model yields very different results. Moreover, we test whether the findings are sensitive to the parametric specification of the distribution of the treatment effect.

A constant treatment effect model should bias SATE. For, from the theoretical model in Section 3 we derived that SATE during (after) participation in training is positively (negatively) correlated with the sub-regional training capacity. Ignoring this correlation will invalidate our identifying restrictions of the treatment effects (4.10): the aggregate sub-regional unobserved group effects are correlated with the training status.

In column (2) of Table 5 we report the estimates of the constant treatment effect model. In contrast, with the variable treatment model, the point estimate of the mean selection bias is large and negative. Consequently, in line with expectations, SATE's of both treatments are larger than those of the benchmark model reported in column (1). Note that the standard errors also increase significantly. This is another indication that the treatment effects are indeed variable. On the other hand, on the basis of the GMM test-statistic, the model with a constant treatment effect fits better than the benchmark. We do not know why this is so. However, note that in neither model the mean selection bias is significantly different from zero. If we restrict it to be zero, the benchmark model specification is to be chosen. The weighted sum of squared residuals (WSSR) is then 527.73 as compared to 529.25 for the constant treatment model (not reported in the Tables).

INSERT TABLE 6 APPROXIMATELY HERE

We assumed until now that the treatment effects follow an exponentially distribution. We now consider two alternative distributions. First, assume a uniform distribution over the interval $[a^-,0]$ during participation and over $[0,a^+]$ post participation. In Appendix D.1 we derive how this affects our regression specification. The estimation results can be found in column (3) of Table 5. These resemble very much those of the constant-treatment-effect model. This is because the uniform distribution allows for much less variation then the exponential distribution. For, p^u_{kmts} varies much less then $\log\left(p^u_{kmts}\right)$ does. The discussion on the constant treatment effect model therefore equally applies here.

Assume now that the treatment effect is normally distributed with mean $\overline{\alpha}_t^{u3}$ and standard error q_t^{u3}. We derive in Appendix D.2 how this affects the specification of the variable treatment effects. Column (4) of Table 5 reports the findings of this model. The point estimates are very different. The mean selection bias suggests now that there is considerable cream skimming in the selection process. The estimated effects of training are now both negative. However, the standard errors of the parameters of interest are so large that the values implied by the benchmark model lie well within the conventional confidence intervals. These huge standard errors suggest that we impose too flexible a model on the available data. Consequently, even if the GMM statistic suggests a considerable improvement in the model specification, we cannot reject the estimated treatment effects of the benchmark model against those found in the model assuming a Normal distribution of the treatment effects. If we impose a zero selection bias, the WSSR of the model with Normally distributed treatment effects is 531.98 as compared to 527.73 for the exponential model. In that case the Normal distribution would be rejected against the exponential.

6. Conclusion

In this paper we studied whether the Belgian authorities are likely to reduce unemployment by increasing the expenditures on vocational classroom training programmes for unemployed workers. On the basis of data for the 1989-93 period in Wallonia, the French-speaking region of Belgium, we find that participation in training has on average reduced the rate of transition out of unemployment by 27%. However, if the worker is still unemployed after participation in the programme, the rate at which unemployment is left is 62% higher than in the absence of training. On the basis of these findings, we simulate that the median unemployment duration[45] of an unemployed worker, who entered training after 6 months and who was trained during 2 months[46], decreases from 21 months to 16 months. Training is therefore found to considerably speed up the transition rate out of unemployment. However, the analysis warns that the returns to training are rapidly diminishing and that increasing the coverage of the programme is unlikely to produce as favourable results. Moreover, we derived on the basis of a job-search model that the programme induces non-participants to postpone their exit out of unemployment. The data did not, however, allow estimating the importance of this effect with sufficient precision.

These estimates of the training effects are conservative, because we found that within group heterogeneity remains important. Moreover, the long-run effects, the effects of training on the incidence and duration of subsequent employment spells and on wages and/or earnings are not included. On the other hand, the effects the analysis ignored general equilibrium effects of training that could significantly downsize the reported impacts.

The paper also dealt with the problem of selective participation. To this purpose, the analysis was based on the Minimum Chi-Square method (Cockx 1997). The transition data are then grouped and can be analysed within a linear framework. This is convenient, for, if participation in training is exogenous at an aggregate level, it allows eliminating the bias induced by endogenous participation, by simple aggregation and by assuming the absence of interaction effects. A drawback of the method is that it does not allow for within-group heterogeneity[47]. It is well known that this neglect biases the estimator downward (Lancaster 1990, p. 64-65). In order to minimise this bias we analyse unemployment spells only after an elapsed duration of 6 months. After this duration the sorting process has rendered the groups of unemployed workers more homogeneous. Nevertheless, specification testing suggests that

[45] Conditional on the elapsed duration of 6 months.
[46] Which corresponds to the median training duration (see Table 1).
[47] The distribution of the training effect is identified by between-group variation. In the empirical study we assume that the within-group variation of the training effect is negligible once one has controlled for the between-group variation.

within-group heterogeneity is still important. A fruitful avenue of future research is therefore to explicitly account for this. In mixed multiple spell transition models exogenous variation in the participation in training is not required to identify the impact of training (see Gritz 1993, Bonnal et al. 1997, Abbring et al. 1998 and Lubyova and van Ours 1998 for continuous time models, and Card and Sullivan 1988 and Magnac 1998 for discrete time models). This would shed some light on the reliability of our findings.

In this study some evidence for selection on unobservables was found: according to the Hausman test statistic the impact effects of training are significantly different from the model that only corrects for selection on observables. This finding was found to be robust to an alternative identifying assumption. A natural experiment generated the exogenous sub-regional variation in the participation rates in training. The over-identifying restrictions implied by this natural experiment could not be rejected.

The aggregation procedure identifies LATE. However, because treatment effects are variable and correlated with the sub-regional training capacities, this estimator does not identify SATE, unless it explicitly accounts for this variability. On the basis of the theoretical model of the participation decision, we explicitly stated the identifying assumptions to account for this variability in the empirical analysis. Even if these assumptions do not require a parametric specification of the distribution of treatment effects, we did impose parametric forms in the empirical application. We experimented with several forms, but we concluded that the data were not sufficiently rich to test these choices against each other.

Appendix

A. Modelling the Participation Decision in Training

Proposition A1: $\forall\Delta \geq 0 : x_p^* > x_0^*$

Proof:

Suppose $x_0^* \geq x_p^*$. Then from (3.2) and (3.3) we obtain

(A.1)
$$r\left(x_0^* - x_p^*\right) = -l_0\left\{\left[\exp(\Delta)-1\right]\int_{x_p^*}^{\infty}\left(w - x_p^*\right)dF(w) + \int_{x_p^*}^{x_0}\left(w - x_p^*\right)dF(w) + \left(x_0^* - x_p^*\right)\overline{F}\left(x_0^*\right)\right\} < 0$$

This contradicts the initial assumption ?.

Proposition A.2: $x_0^* < x_w^* < x_t^*$

Proof:

From (3.2) and (3.8) it follows that

(A.2)
$$r\,x_w^* = r\,x_0^* - l_0\left[\int_{x_0^*}^{x_w^*}\left(w - x_0^*\right)dF(w) + \left(x_w^* - x_0^*\right)\overline{F}\left(x_w^*\right)\right] + h_{wt}\left(x_t^* - x_w^*\right)$$

Suppose $x_t^* \leq x_w^*$. Then, since $x_w^* > x_0^*$ by (3.11), (A.2) cannot hold, a contradiction ?.

Proposition A.3: $\partial x_w^* / \partial\Delta > 0$

Proof:

From (3.3) it follows that

(A.3)
$$\frac{\partial x_p^*}{\partial\Delta} = \frac{r\left(x_p^* - z\right)}{r + h_{pe}} > 0\,.$$

Deriving (3.6) with respect to Δ yields:

(A.4)
$$\frac{\partial x_t^*}{\partial\Delta} = \frac{g\,\dfrac{\partial x_p^*}{\partial\Delta}}{\left(r + h_{te} + g\right)} > 0.$$

From (3.8) we derive:

(A.5) $$\frac{\partial x_w^*}{\partial \Delta} = \frac{h_{wt} \dfrac{\partial x_t}{\partial \Delta}}{(r + h_{w.})} > 0 \text{ ?}.$$

Lemma A.1: $\partial x_w^* / \partial p_t > 0$

If for any given Δ^*, we differentiate (3.8) partially with respect to p_t and use (3.9), we obtain:

(A.6) $$\frac{\partial x_w^*}{\partial p_t} = \frac{\bar{h}_{t.}}{p_w (r + h_{w.})} \left(1 - \frac{\partial p_w}{\partial p_t} \frac{p_t}{p_w} \right).$$

Hence, the lemma is proven, if we can show that $\partial p_w / \partial p_t < 0$. This requires an expression of p_w as a function of p_t. We can find this by using the condition that in a stationary environment the number of workers entering a state must equal the number of workers leaving it. First, the number of workers entering the state 'w', i.e. the number of workers entering unemployment that are selected for training, is equal to the number of workers leaving state 'w':

(A.7) $$\bar{p} h_{eu} p_e = \left(h_{wt} + \bar{h}_{we} \right) p_w$$

where $\bar{p} \equiv \int_{\Delta^*}^{\infty} p(\Delta) g(\Delta) d\Delta$ is the average probability of selection, h_{eu} is the average hazard from employment into unemployment[48] and p_e is the number of workers employed as a proportion of the active population. Second, the number of workers entering training that neither apply nor are selected to participate in training must be equal to the number of non-trainees leaving unemployment:

(A.8) $$(1 - \bar{p}) h_{eu} p_e = h_0 p_0$$

where p_0 is the proportion of unemployed workers not being (or waiting to be) trained. The third equation is given by (3.9) in the text. Finally, the number of workers entering the post-training state must be equal to the number of workers leaving the post-training state for employment:

(A.9) $$h_{tp} p_t = \bar{h}_{pe} p_p$$

where $\bar{h}_{pe}$ is the average transition intensity from the post-training state to employment and p_p is the proportion of workers in the post-training state.

From (3.9) we obtain an expression for h_{wt}. If we insert this in (A.7) and add this equation to (A.8), we obtain:

(A.10) $$h_{eu} p_e = \bar{h}_t p_t + \bar{h}_{we} p_w + h_0 p_0$$

---

[48] For simplicity, we assume that this hazard does not depend on ?.

If the size of the active population is fixed, we may write:

(A.11) $p_e = 1 - p_0 - p_w - p_t - p_p$

Therefore, using (A.9) to replace p_p and (A.11) to replace p_e, (A.8) and (A.10) define a set of equations determining the equilibrium values of p_0 and p_w as a function of exogenous variables:

(A.12) $$\begin{bmatrix} h_0 + (1-\bar{p})h_{eu} & (1-\bar{p})h_{eu} \\ h_0 + h_{eu} & \bar{h}_{we} + h_{eu} \end{bmatrix} \begin{bmatrix} p_0 \\ p_w \end{bmatrix} = \begin{bmatrix} (1-\bar{p})h_{eu}\left[1 - \left(1 + \frac{h_{tp}}{\bar{h}_{pe}}\right)p_t\right] \\ h_{eu} - \left[\bar{h}_{t.} + h_{eu}\left(1 + \frac{h_{tp}}{\bar{h}_{pe}}\right)\right]p_t \end{bmatrix} \equiv \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

Solving for this set of equations we obtain:

(A.13) $$\begin{bmatrix} p_0 \\ p_w \end{bmatrix} = \begin{bmatrix} (\bar{h}_{we} + h_{eu})/D & -(1-\bar{p})h_{eu}/D \\ -(h_0 + h_{eu})/D & [h_0 + (1-\bar{p})h_{eu}]/D \end{bmatrix} \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$$

where

(A.14) $D \equiv h_{we}[h_0 + (1-\bar{p})h_{eu}] + \bar{p}h_0 h_{eu} > 0$

Partial differentiation in (A.13) then yields:

(A.15) $$\frac{\partial p_w}{\partial p_t} = -\frac{\bar{p}h_{eu}h_0\left(1 + h_{tp}/\bar{h}_{pe}\right) + [h_0 + (1-\bar{p})h_{eu}]\bar{h}_{t.}}{D} < 0 \ ?.$$

Lemma A.2: $\partial\Delta^* / \partial p_t < 0$ if and only if condition (P.1) is satisfied.

Differentiating (3.11) partially w.r.t. p_t yields:

(A.16) $$\frac{\partial\Delta^*}{\partial p_t} = \frac{-\dfrac{\partial x_w^*}{\partial p_t}}{\left[\dfrac{\partial x_w^*}{\partial\Delta^*} + (x_w^* - x_0^*)\left(\dfrac{p'(\Delta^*)}{p(\Delta^*)} - \dfrac{c'(\Delta^*)}{c(\Delta^*)}\right)\right]}.$$

Subsequently, by differentiating (3.8) partially w.r.t. Δ^* one obtains:

(A.17) $$\frac{\partial x_w^*}{\partial\Delta^*} = \frac{\left[\partial h_{wt}/\partial\Delta^* \left(x_t^* - x_w^*\right) + h_{wt}\, \partial x_t^*/\partial\Delta^*\right]}{(r + h_{w.})}$$

If we insert this expression in (A.16), use Lemma A.1 and Proposition A.2, then indeed $\partial\Delta^* / \partial p_t < 0$ if and only if condition (P.1) is satisfied ? .

B. Exogeneity of the Sub-Regional Variation in the Training Capacity

During the 1989-1993 FOREM's real expenditures on training were at first constant until 1990, but then dropped significantly by 18% in 1991. In 1992 they remained at this low level, but in 1993 they increase by 10%, as a consequence of the introduction of the counselling programme PAC in March. Despite these reductions in outlays, we observe in our data that the average number of unemployed workers entering training each month increases during the observation period. If we set the index of this number at 100 in 1989[49], it remains at the same level in 1990, rises to 105 in 1991, 110 in 1992 and 124 during the first three months of 1993.

This apparent contradiction in the evolution of the training capacity can be explained in the following way. First, from April 1991 onwards, vocational training became fee-based to employed workers (cf. Section 2). This induced a substitution of employed trainees for unemployed. Second, an increasing number of training slots is provided in external training centres, set up with the help of employers. Even if this number represents only a minor fraction of the total (5.5% in 1990), its growth has been very significant. From an index of 100 in 1990 to 149 in 1991, 181 in 1992 and 240 in 1993. To the extent that these employers partly financed the investment and operation costs of these centres, this allowed FOREM to increase training capacity without burdening the budget. Finally, from 1992 onwards the training programmes have been redesigned. They evolved from a unique relatively long programme to a sequence of targeted programmes of shorter duration. This increases the number of workers entering training each month without increasing the number of training hours supplied.

We now examine whether this increase in the number of workers entering training varied between sub-regions. If the increase was common to all sub-regions, then the sub-regional variation in the transition rates out of unemployment cannot explain the evolution of training capacity across sub-regions in the 1989-1993 period. To test this, we calculated for each year the number of trainees entering training in each sub-region as a proportion of all entries in Wallonia and examined the evolution of these proportions over time. The increase in the number of trainees is common if this evolution is constant across sub-regions. We find a relative stability in these proportions, apart from Liège and Verviers for which these proportions decline from 1990 onwards. We test this formally in a weighted[50] regression of these proportions on indicator variables for the sub-regions. The weighted sum of squared residuals (WSSR) of this regression is distributed Chi-Square with 28 degrees of freedom. Its value is equal to 55.8. We can therefore reject the hypothesis that the evolutions are indeed proportional at a significance level of 0.1%. Alternatively, if we allow for interactions between the indicator variable for Liège and Verviers and the dummy variables for calendar time, this statistic drops to 14.9. With 24 degrees of freedom the hypothesis of proportionality can therefore not be rejected for the other 6 sub-regions.

We did not find any satisfying explanation of this divergent evolution for Liège and Verviers. However, this evolution does in no way point to a causal relationship between the transition rate out of unemployment and the training capacity. For, during the 1989-1993 period, the sub-region of Liège and Verviers occupies a median position in terms of the

[49] In 1989, on average 1959 unemployed workers enter training each month.
[50] The weights for each year being equal to the inverse of the total number of entrants in training in Wallonia.

propensity to be trained. This means that there cannot have been any systematic negative relationship between the relatively high transition rate out of unemployment[51].

We now test whether, for the 1980-1988 period, we find any systematic correlation between the growth rate in the training capacity of a sub-region and the unemployment rate. The unemployment rate is taken as a proxy variable, because figures on the transition rate out of unemployment are lacking for this period. For reasons of data availability we also need to proxy the variable regarding training capacity. Data for the whole period are only available for the number of unemployed workers having completed a training programme in the secondary or tertiary sector. Regarding the total number of terminated training spells, including C.A./C.O.I.S.P., data are missing for the 1984-1986 sub-period. We tested whether conclusions are sensitive to the proxy variable chosen for the analysis.

From this analysis we retain the following. First, the correlation coefficients between the growth rate of the number of trainees in the secondary and tertiary sector and the unemployment rate are not systematically positive or negative. Second, the absolute values of the correlation coefficients are always small (less then .35) and insignificant, except for the years 1983-84 and 1987-88. The corresponding correlation coefficients (standard errors) are respectively -.58 (.17) and -.61 (.14). However, when the sub-region Mons is excluded from the analysis, these values, -.32 (.53) and -.53 (.28), become insignificantly different from zero at a 5% level of significance. Moreover, if we include the trainees in C.A./C.O.I.S.P. to calculate the growth rate of trainees in 1987-88, the correlation coefficient drops to an even smaller insignificant level: -.39 (.38). On the other hand, on the basis of this “correct” growth rate we now find a significant positive correlation for 1982-83: .63 (.12). This value is, however, generated by an outlier. Without this observation, the sub-region of Mons again, the correlation coefficient drops to -.08 (.87). The correlation coefficient for the other two available years 1980-81 and 1981-82 are small and insignificant: respectively -.04 (.93) and -.12 (.80).

These findings for Mons are reasons for concern. They suggest a correlation between the growth rate of the number of trainees and the unemployment rate in 1983-84 and 1987-88. If both the training capacity and the transition rate from unemployment display inertia, such a relationship would invalidate our hypothesis of exogenous sub-regional variation of the participation rates. Even if this correlation is no longer significant (or even reversed) when we include trainees in C.A./C.O.I.S.P., the findings could reflect a causal link between the unemployment rate and the *type* of the training scheme offered. However, since the natural experiment provides us with over-identifying restrictions, we can test whether the participation rate in training is truly exogenous for Mons. On the basis of this test (see Section 5), we conclude that there is no reason to exclude the data on Mons from the analysis.

C. The Statistical Model

C.1. The system of SUR equations (4.4)

Below (4.4) b_{kmts}^{uvw} is defined as follows:

$$\text{(C.1)} \qquad b_{kmts}^{uvw} = \frac{1}{P_{kmts}^{uu} - 1} - \frac{1}{P_{kmts}^{uu} \log\left(P_{kmts}^{uu}\right)} + d_{vw} \frac{1}{P_{kmts}^{uv}}$$

and d_{vw} the Kronecker delta. If we order the errors in (4.4) for $v \neq u$ in a vector, we have for all k, m, s, t, and u=1, 2

[51] The transition rate out of unemployment of this sub-region is the second highest in the 1989-1993 period.

(C.2) $E\left(w^{u}_{kmts}\right) = 0$.

The variance-covariance matrix of the vector w^{u}_{kmts} has typical element

$$s^{uvw}_{kmts} = E\left[w^{uv}_{kmts} w^{uw}_{kmts}\right] =$$

$$\text{(C.3)}\quad E\left\{ \left[\begin{array}{l} \left(b^{u}_{kmts}\right)^2 \left(\sum_{x=1,x\neq u}^{3} P^{ux}_{kmts}\left(1-P^{ux}_{kmts}\right) - \sum_{x=1,x\neq u}^{3} \sum_{y=1,y\neq x}^{3} P^{ux}_{kmts} P^{uy}_{kmts} \right) \\ +\, 2b^{u}_{kmts} P^{uu}_{kmts} + d_{vw}\left(1-P^{uv}_{kmts}\right) / P^{uv}_{kmts} - \left(1-d_{vw}\right) \end{array} \right] \Big/ r^{u}_{kmts} \right\}$$

with $v,\, w = 1,\, 2,\, 3 \neq u,\;\; b^{u}_{kmts} = \dfrac{1}{P^{uu}_{kmts} - 1} - \dfrac{1}{P^{uu}_{kmts} \log\left(P^{uu}_{kmts}\right)}$, and where the expectation is taken with respect to the distribution of r^{u}_{kmts}. A consistent estimate is obtained by omitting the expectation and replacing the transition probabilities by their estimates.

C.2. Estimation of the SUR system (4.5)

Under assumption (4.6), we can estimate the parameters of the equations in (4.5) by OLS, or more efficiently by GLS. If we order the regressions by m, k, t, s, u, v, in this order, we obtain a block-diagonal variance-covariance matrix of the vector of disturbances with the diagonal blocks being 2x2 matrices V^{u}_{kmts}.

The feasible GLS procedure consists of two steps. In the first step we use OLS to estimate the parameters of the equations in (4.5). Next we estimate the 2x2 variance-covariance matrix of unobserved group effects, e^{uv}_{kmts}, by

$$\text{(C.4)}\quad \hat{s}^{1vw} = \frac{1}{4M(K-1)} \left\{ \sum_{m=1}^{M} \sum_{k=2}^{K} \sum_{t=0}^{1} \sum_{s=o}^{y} \left(\hat{u}^{1v}_{kmts} \hat{u}^{1w}_{kmts} - \hat{s}^{1vw}_{kmts} \right) \right\}$$

for origin state u=1, and by

$$\text{(C.5)}\quad \hat{s}^{2vw} = \frac{1}{2M(K-1)} \left\{ \sum_{m=1}^{M} \sum_{k=2}^{K} \sum_{s=o}^{y} \left(\hat{u}^{1v}_{km1s} \hat{u}^{1w}_{km1s} - \hat{s}^{1vw}_{km1s} \right) \right\}$$

for origin state u=2, where $\hat{u}^{uv}_{kmts}$ is the OLS residual of the regression (4.5), and $\hat{s}^{uvw}_{kmts}$ the estimate of (B.3). substitution of this consistent estimate gives a consistent estimate of V^{u}_{kmts}, which is used in the second step of the feasible GLS procedure.

Let $\hat{u}$ be the vector of residuals of the regression equation (4.5) ordered by k, m, t, s, u, v, in that order. Let the estimated variance-covariance matrix of disturbances of (4.5) be denoted by $\hat{V}$. This is a block diagonal matrix, which simplifies the computation of its inverse. Under the assumption that the model is correctly specified, we have that $\hat{u}'\hat{V}^{-1}\hat{u}$ follows a c^2-distribution with 12(K-1)M-3(K-1)-3(M-1)-3-2x2-1 degrees of freedom (= the number of cells minus the number of estimated parameters).As the regression equations can be seen as a set of 12(K-1)M moment conditions, this statistic can be interpreted as a GMM test for the null hypothesis of regressor-error orthogonality.

D. Various Assumptions with respect to the variable treatment effect

D.1. The Uniform distribution

Assume that, for workers in the post-training state, $a_p(\Delta)$ is uniformly distributed over the interval $[0, a^+]$. The unconditional mean of the post-training effect is then $E[a_p(\Delta)] \equiv \bar{a}_p = a^+/2$. Then from (3.20), we have

(D.1) $$p_p = \int_{a_p(\Delta^*)}^{2\bar{a}_p} \frac{1}{a^+} da = 1 - \frac{a_p(\Delta^*)}{2\bar{a}_p}$$

and therefore

(D.2) $$a_p(\Delta^*) = 2\bar{a}_p(1 - p_p)$$

The average treatment effect of the treated is then

(D.3) $$E[a_p(\Delta) | \Delta \geq \Delta^*, \textit{participation in } p] = \frac{1}{p_p} \int_{2\bar{a}_p(1-p_p)}^{2\bar{a}_p} \frac{a}{2\bar{a}_p} da = \bar{a}_p(2 - p_p)$$

Similarly, if we assume that $a_t(\Delta)$ is uniformly distributed over the interval $[a^-, 0]$, then we can derive that

(D.4) $$E[a_t(\Delta) | \Delta \leq \Delta^*, \textit{participation in } t] = \bar{a}_t(2 - p_t)$$

If we assume a uniform distribution, we may therefore replace $\bar{a}_t^{u3}[1 - \log(p_{kmts}^u)]$ in (4.12) by $\bar{a}_t^{u3}[2 - p_{kmts}^u]$.

D.2. The Normal distribution

Assume that, for workers in the post-training state, $a_p(\Delta)$ has a Normal distribution with mean $\bar{a}_p$ and variance q_p^2. Then from (3.20), we have

(D.5) $$p_p = 1 - \Phi\left(\frac{a_p(\Delta^*) - \bar{a}_p}{q_p}\right),$$

where $\Phi(.)$ denotes the standard Normal distribution. The average treatment effect of the treated is

(D.6)
$$E[a_p(\Delta) | \Delta \geq \Delta^*, \textit{participation in } p] = \frac{1}{p_p q_p \sqrt{2pi}} \int_{a_p(\Delta^*)}^{+\infty} a \exp\left[-(a - \bar{a}_p)^2 / 2q_p^2\right] da$$

If we change variables such that $a = \bar{a}_p + q_p x$ and $da = q_p dx$, and if we use that by (D.5), $\left[a_p(\Delta^*) - \bar{a}_p\right]/q_p = \Phi^{-1}(1-p_p)$, we obtain

(D.7)
$$E\left[a_p(\Delta) \middle| \Delta \geq \Delta^*,\ participation\ in\ p\right] = \frac{1}{p_p\sqrt{2pi}}\int_{\Phi^{-1}(1-p_p)}^{+\infty}(\bar{a}_p + q_p x)\exp\left[-x^2/2\right]dx$$
$$= \bar{a}_p + \frac{q_p}{p_p}\frac{\exp\left[-\left[\Phi^{-1}(1-p_p)\right]^2/2\right]}{\sqrt{2pi}}$$

Similarly, if we assume that $a_t(\Delta)$ has a Normal distribution with mean $\bar{a}_t$ and variance q_t^2, we can derive

(D.8)
$$E\left[a_t(\Delta) \middle| \Delta \leq \Delta^*,\ participation\ in\ t\right] = \bar{a}_t - \frac{q_t}{p_t}\frac{\exp\left[-\left[\Phi^{-1}(p_t)\right]^2/2\right]}{\sqrt{2pi}}$$

References

Aakvik, Arild (1999), "Assessing the Effects of Labour Market Training in Norway", *mimeo*, prepared for the ZEW Conference on Econometric Evaluations of Active Labour Market Policies in Europe, Mannheim, June 25-26.

Abbring, J.H., G.J. van den Berg, and J.C. van Ours (1994), "The anatomy of unemployment dynamics", working paper n°9437, Tinbergen Institute.

Abbring, J. H., G.J. van den Berg and J.C. van Ours (1998), "The Effect of Unemployment Insurance Sanctions on the Transition Rate from Unemployment to Employment", *mimeo*, Amsterdam : Free University.

Amemiya Takeshi (1985), *Advanced Econometrics*, Oxford: Basil Blackwell.

Amemiya, Takeshi and Frederick Nold (1975), "A Modified Logit Model", *The Review of Economics and Statistics*, 57, 255-57.

Angrist, Joshua D. (1991), "Grouped-data estimation and testing simple labor-supply models", *Journal of Econometrics*, 47, 243-66.

Angrist, J. and G. Imbens (1991), "Sources of Identifying Information in Evaluation Models", NBER Technical Working Paper No. 117, December.

Angrist, Joshua. D., Guido W. Imbens and Donald B. Rubin (1996), "Identification of Causal Effects Using Instrumental Variables", *Journal of the American Statistical Association*, 91 (434), 444-55.

Angrist and Krueger (1999),…

Ashenfelter, O. and C. Rouse (1995), "Schooling, intelligence, and income in America: Cracks in the Bell Curve", unpublished mimeograph, Princeton University.

Björklund, A. (1993), "The Swedish experience", in Jensen, K. and P.K. Madsen (eds.), *Measuring Labour Market Measures,* Copenhagen: Ministry of Labour, 243-63.

Björklund, A. (1994), "Evaluations of Swedish Labor Market Policy", *International Journal of Manpower,* 15(5, part 2), 16-31.

Bollens J. and Nicaise I. (1994), "The Medium-Term Impact of Vocational Training on Employment and Unemployment Duration", paper prepared for the European Association of Labour Economists Conference, Warsaw, September.

Bonnal, Liliane, Denis Fougère and Anne Sérandon (1997), "Evaluating the Impact of French Employment Policies on Individual Labour Market Histories", *The Review of Economic Studies,* 64, 683-713.

Breen, R. (1991), "Assessing the Effectiveness of Training and Temporary Employment Schemes: Some Results from the Youth Market", *The Economic and Social Review,* 22(3), 177-98.

Brodaty, Thomas, Bruno Crépon and Denis Fougère (1999), "Using Matching Estimators to Evaluate Alternative Youth Employment Programmes: Evidence from France, 1986-1988", Handout prepared for the ZEW Conference on Econometric Evaluations of Active Labour Market Policies in Europe, Mannheim, June 25-26.

Calmfors, L. (1994), "Active Labor market policy and unemployment – A framework for the analysis of crucial design features," *OECD Economic Studies*, 22(1), 7-47.

Card D. and D. Sullivan (1988), "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment", *Econometrica,* 56(3): 497-530.

Cockx, Bart (1997), "Analysis of Transition Data by the Minimum Chi-Square Method. An Application to Welfare Spells in Belgium", *The Review of Economics and Statistics,* 79(3), 392-405.

Cockx, Bart (1999), "Les formations professionnelles du FOREM accélèrent-elles la sortie du chômage en Wallonie ?", *Etude spéciale du Service d'Analyse économique de l'IRES,* septembre, Louvain-la-Neuve: IRES, Université catholique de Louvain, 105-124.

Cockx, Bart and Geert Ridder (2000), "Social Employment of Welfare Recipients in Belgium: an Evaluation", mimeo, revision of Discussion Paper n° 9618, Institut de recherches économiques et sociales, Université Catholique de Louvain, Louvain-la-Neuve, *forthcoming in The Economic Journal*.

Cockx, Bart, Bruno Van der Linden, and Adel Karaa (1998), "Active labour market policies and job tenure", *Oxford Economic Papers,* 50, 685-708.

Commission of the European Union (1997), *Conclusions de la Présidence – Luxembourg, 20/21 novembre 1997. Les "lignes directrices" pour 1998,* Bruxelles.

Davidson, C. and S. Woodbury (1993), "The displacement effect of reemployment bonus programs", *Journal of labor Economics,* 11(4), 575-605.

Deaton, A. (1985), "Panel Data from Time Series of Cross-sections", *Journal of Econometrics,* 30, 109-26.

Dehejia, Rajeev H. and Sadek Wahba (1999), "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs", *The Journal of the American Statistical Society*, 94 (448), December, 1053-62.

de Koning J., M. Koss, and A. Verkaik (1991), "A quasi-experimental evaluation of the vocational training centre for adults", *Environment and Planning C: Government and policy,* 9, 143-53.

Dolton, Peter J., Gerald H. Makespace, and John G. Treble (1994), "The Youth Training Scheme and the School-to-Work Transition", *Oxford Economic Papers,* 46(4), 629-57.

Eberwein, C., J. Ham and R. Lalonde (1997), "The impact of classroom training on the employment histories of disadvantaged women: evidence from experimental data", *Review of Economic Studies,* 64(4), 655-82.

Firth, David, Clive Payne and Joan Payne (1999), "Efficacy of programmes for the unemployed: discrete time modelling of duration data from a matched comparison study", *Journal of the Royal Statistical Society A,* 162, Part 1, 111-20.

Flinn, C.J. and J.J. Heckman (1982), “Models for the Analysis of Labor Force Dynamics”, in Bassman, R. and G. Rhodes (eds.), *Advances in Econometrics,* vol 1, 35-95.

Friedlander, D., D. Greenberg and P. Robins (1997), "Evaluating government training programs for the economically disadvantaged", *Journal of Economic Literature,* 35(4), 1809-55.

FOREM (1989-1994), Rapport Annuel, Bruxelles.

Gritz, M. (1993), "The impact of training on the frequency and duration of employment", *Journal of Econometrics,* 57(1-3), 21-51.

Hausman, J. A. (1978), "Specification Tests in Econometrics", *Econometrica,* 46, 1251-72.

Heckman, James J. (1997), "Instrumental Variables. A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations", *Journal of Human Resources,* 32(3), 441-62.

Heckman, James J. and B. Honoré (1990), "The empirical content of the Roy model", *Econometrica*, 58(5), 1121-49.

Heckman, James J., H. Ichimura, J. Smith and P. Todd (1997), “Matching as an econometric evaluation estimator: evidence from evaluating az job training program”, *Review of Economic Studies,* 64(4), 605-54.

Heckman, James J., R. J. Lalonde and J.A. Smith (1999), "The Economics and Econometrics of Active Labor Market Programs", in Ashenfelter, O. and D. Card (eds.), *The Handbook of Labor Economics, Volume III…..*

Heckman, James J. and Richard Robb, Jr. (1985), "Alternative Methods for Evaluating the Impact of Interventions", in Heckman, James J. and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data,* Econometric Society Monograph series, New-York: Cambridge University Press.

Heckman, J.J., J. Smith and N. Clements (1997), "Making the most out of programme evaluations and social experiments: accounting for heterogeneity in programme impacts", *Review of Economic Studies*, 64(4), 487-535.

Heckman, J.J., J.A. Smith and Taber (1996), "What Do Bureaucrats Do? The Effects of Performance Standards and Bureaucratic Preferences on Acceptance into the JTPA Program", in Libecap, Gary D. (ed.), *Reinventing Government and the Problem of Bureaucracy, Advances in the Study of Entrepreneurshi, Innovation, and Economic Growth, Volume 7,*Greenwich, Connecticut: JAI Press Inc.

Imbens Guido W. and Joshua D. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica,* 62(2), 467-75.

Kraus, F., P. Puhani and V; Steiner (1997), "Employment effects of publicly financed training programs – The East German experience", Discussion Paper No. 97-33, Centre for European Economic Research (ZEW).

Lalonde, R. (1995), "The Promise of Public Sector-Sponsered Training Programs", *Journal of Economic Perspectives,* 9(2), Spring, 149-68.

Lancaster, Tony (1990), *The Econometric Analysis of Transition Data,* Cambridge: Cambridge University Press.

Lechner, M. (1999), “Earnings and Employment Effects of Continuous Off-the-Job Training in East-Germany after Unification”, *Journal of Business and Economic Statistics,* 17, 74-90.

Lubyova, Martina and Jan C. van Ours (1998), "Effects of Active Labour Market Programmes on the Transition Rate from Unemployment into Regular Jobs in the Slovak Republic", CentER discussion paper, No. 98127, Tilburg University.

Magnac, T. (1998), "State Dependence and Heterogeneity in Youth Employment Histories", CREST Working Paper n°9747.

Meyer, Bruce D. (1995), "Natural and Quasi-Experiments in Economics", *Journal of Business and Economic Statistics,* 13(2), April, 151-61.

Mortensen, Dale T. (1986), "Job search and labor market analysis", in O. Ashenfelter and R. Layard (eds.), *Hanbook of Labor Economics*, Amsterdam: North Holland, 849-919.

OECD (1993), *Employment Outlook*, Paris.

OECD (1997), *Le service publique de l'emploi. Belgique*, Paris.

Parks, Richard W. (1980), "On the Estimation of Multinomial Logit Models with Relative Frequency Data", *The Journal of Econometrics,* 13, 293-303.

Plasman (1993), "Estimation de Durée de Chômage et Rôle des Politiques d'Emploi", in Point d'Appui Emploi Formation (ed.), *Chômage de Longue Durée – Comment en parler? Comment le combattre?*, dossier 4, Bruxelles.

Plasman (1994), *Les Politiques du Marché du Travail: Analyse et Comparaisons européennes; Procédure d'Evaluation (micro et macroéconomique); Evaluation des Politiques de Resorption du Chômage et des Politiques du Temps de Travail en Belgique,* Thèse de Doctorat, Université Libre de Bruxelles, Bruxelles.

Prentice, R. L. and L.A. Gloeckler (1978), "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data", *Biometrics,* 34, 57-67.
Puhani, Patrick A. (1998), "Advantage through Training? A Microeconometric Evaluation of the Employment Effects of Active Labour Market Programmes in Poland", *mimeo,* Centre for European Economic Research (ZEW), Mannheim.

Ridder, Geert (1986), "An event history approach to the evaluation of training, recruitment, and employment programmes", *Journal of Applied Econometrics,* 11, 109-26.

Roy, A.D. (1951), "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers (New Series)*, 3, 135-46.

Salant, S.W. (1977) "Search Theory and Duration Data: A Theory of Sorts", *Quaterly Journal of Economics*, 91, 39-57.

Torp, Hege (1994), "The Impact of Training on Employment: Assessing a Norwegian Labour Market Programme", *Scandinavian Journal of Economics,* 96(4), 531-50.

Torp, H., O. Raaum, E. Heraes, And H. Goldstein (1993), "The first Norwegian experiment", in Jensen, K. and P.K. Madsen (eds.), *Measuring Labour Market Measures,* Copenhagen: Ministry of Labour, 97-140.

Van der Linden, B.(1997), "Effets des formations professionnelles et des aides à l'embauche: exploitation d'une enquête auprès des employeurs", Economie et Prévision, 130, 113-30.

Van Langendonck, J. (1991), *Handboek Sociale Zekerheid,* Acco, Leuven.

Wald, A. (1940), "The fitting of straight lines if both variables are subject to error", *Annals of Mathematical Statistics,* 11, 284-300.

Zweimüller, Josef and Rudolf Winter-Ebmer (1996), "Manpower Training Programmes and Employment Stability", *Economica,* 63, 113-30.

Table 1.: Population Characteristics

	Kind of unemployment spells			
	Old		young	
	All Spells	Spells with some time in training	All Spells	Spells with some time in training
Eligibility Type				
Old	76.98%	81.58%	0%	0%
Sex				
Male	46,25%	55,35%	42,19%	39,19%
Age group				
<=20 years	5,99%	5,94%	60,23%	49,75%
21-25 years	31,28%	29,77%	38,84%	48,83%
26-30 years	23,98%	25,28%	0,94%	1,42%
31-40 years	26,15%	29,74%	-	-
41-50 years	12,61%	9,27%	-	-
Average age (years)	29,8	29,5	20,3	20,8
1st unemployment registration				
Yes	15,77%	11,60%	90,27%	90,82%
Educational attainment				
Primary (6years)	31,33%	22,47%	11,75%	8,04%
Lower Secondary (9 years)	29,56%	26,74%	29,02%	20,95%
Higher Secondary (12 years)	22,21%	26,13%	36,34%	45,12%
Higher Education (>12 years)	11,05%	16,48%	18,86%	24,02%
Other or unknown	5,85%	8,19%	4,04%	1,87%
Sub-regional department				
Nivelles and La Louvière (m=1)	17,00%	13,97%	16,10%	17,20%
Charleroi (m=2)	16,79%	14,19%	18,82%	12,28%
Mons (m=3)	8,87%	7,29%	8,91%	9,92%
Mouscron and Tournai (m=4)	9,43%	12,36%	8,93%	11,94%
Liège and Verviers (m=5)	27,90%	27,06%	25,84%	24,41%
Arlon (m=6)	5,45%	9,48%	5,99%	12,11%
Namur and huy (m=7)	14,56%	15,64%	15,41%	12,13%
Median Duration in months				
Unemployment Duration	3	9	3	12
Duration until Training	-	3	-	6
Time Spent on Training	-	2	-	2

Table 2: Estimates of the Benchmark Model –No Correction for Selection on Unobservables
(standard error in parenthesis)

	Destination state v		
	1: Unemployed not participating	**2: Unemployed participating**	**3: Out of Unemployment**
Duration interval (months)			
6-8 $\left(g_2^v\right)$	-2.173*	-5.951*	-3.094*
	(.041)	(0.127)	(.061)
8-10 $\left(g_3^v - g_2^v\right)$	.365*	.154*	.132*
	(.040)	(.075)	(.041)
10-12 $\left(g_4^v - g_2^v\right)$	.258*	-.100	-.154*
	(.043)	(.078)	(.042)
12-15 $\left(g_5^v - g_2^v\right)$	.552*	.013	.044
	(.040)	(.075)	(.041)
15-18 $\left(g_6^v - g_2^v\right)$	.497*	-.177*	-.202*
	(.043)	(.076)	(.042)
18-22 $\left(g_7^v - g_2^v\right)$	.688*	-.273*	-.125*
	(.045)	(.077)	(.043)
22-28 $\left(g_8^v - g_2^v\right)$	.686*	-.586*	-.364*
	(.056)	(.092)	(.051)
Sub-region [1]			
Charleroi $\left(b_2^v - b_1^v\right)$	-.072*	.143*	-.120*
	(.043)	(.077)	(.044)
Mons $\left(b_3^v - b_1^v\right)$	-.050	.200*	-.081*
	(.050)	(.081)	(.046)
Mouscron and Tournai $\left(b_4^v - b_1^v\right)$	-.093*	.296*	.030
	(.045)	(.092)	(.046)
Liège and Verviers $\left(b_5^v - b_1^v\right)$	-.082*	.161*	.040
	(.038)	(.073)	(.040)
Arlon $\left(b_6^v - b_1^v\right)$	.245*	.451*	.327*
	(.046)	(.151)	(.058)
Namur and Huy $\left(b_7^v - b_1^v\right)$	.057	.331*	.045
	(.042)	(.076)	(.043)
Eligibility type			
Young $\left(y_y^v\right)$	.101*	.227*	.241*
	(.028)	(.045)	(.024)
Impact of Training			
Before participation $\left(l^{1v}\right)$		4.782	-.242
		(4.188)	(1.681)
During participation $\left(\varpi_1^{23}\right)$		-	-.070*
			(.014)
Post participation $\left(\varpi_1^{1v}\right)$		2.324*	.104*
		(.043)	(.006)
Unobserved group effects			
ς^{1vv} $\left(\varsigma^{1vw}\right)$	-	.0404 (.0034)	.0189 (.0034)
ς^{2vv} $\left(\varsigma^{2vw}\right)$	0 (0)	-	0 (0)
WSSR		527.73	
(degrees of freedom)		(499)	
P-value		.180	
Durbin-Watson statistic (d)		1.2597	
(number of cells)		(546)	

1 The reference sub-region is Nivelles and La Louvière
* significant at the 5% level

<u>Table 3:</u> Testing Over-Identifying Restrictions
(standard errors in parentheses)

	(1) The Benchmark GIV	**(2) Without region dummies $\forall i$: $b_i^3 = 0$**	**(3) Without region dummies except for Arlon b_6^3**	**(4) $r_1^1 \neq r_1^2$**	**(5) Including k=1**	**(6) GLS**
Impact of Training						
Before participation $\left(l^{13}\right)$	-.416	8.548*	2.034	-1.358	1.707	1.344
	(1.899)	(1.112)	(1.500)	(2.645)	(1.686)	(1.708)
During participation $\left(\varpi_1^{23}\right)$	-.058	-.120*	-.081	-.010	-.047	-.164*
	(.053)	(.053)	(.050)	(.107)	(.052)	(.034)
During participation for k=1	-	-	-	-	.068	-
					(.048)	
Post participation $\left(\varpi_1^{13}\right)$	.118*	.001	.083	.090	.117	-.000
	(.058)	(.052)	(.053)	(.069)	(.057)	(.035)
Mean selection bias						
During participation (r_1^2)	-.048	.450*	.111	-.290	-.058	.450*
	(.243)	(.219)	(.220)	(.527)	(.240)	(.149)
Post participation $\left(r_1^1\right)$	-.048	.450*	.111	.084	-.058	.450*
	(.243)	(.219)	(.220)	(.300)	(.240)	(.149)
WSSR	530.84	531.62	548.49	522.65	666.74	531.58
(degrees of freedom)	(498)	(504)	(503)	(497)	(536)	(498)
P-value	.149	.191	.079	.206	.000	.144
Hausman test[1]: $c^2(2)$	19.48	21.44	2.62	.00	2.23	6.66
P-value	.000	.000	.270	1.000	.327	.036

*significant at the 5% level

1 Tests whether the asymptotically more efficient estimators of the 2 impact effects of training, ϖ_1^{u3} of model (2), (3), (5) and (6) are jointly to be rejected against the consistent estimators of the corresponding parameters in the baseline model (1). The statistic in column (4) tests whether the asymptotically more efficient benchmark model (1) is to be rejected against the consistent estimator of model (4). The statistic in column (1) tests whether the model without correction for selection bias of Table 2 is to be rejected against model (1).

Table 4: The Variation of the Treatment Effect over duration and Eligibility Status.
(standard errors in parentheses)

	s=y	s=o						r_1^u	WSSR (DF) P-value
During participation (a_{1s}^{23})	-.117*	-.062							
	(.049)	(.047)						.039	547.5
Post participation (a_{1s}^{13})	.068	.114*						(.215)	(495)
	(.051)	(.052)							.051
	k=2	k=3†	k=4†	k=5†	k=6†	k=7†	k=8†	r_1^u	WSSR (DF) P-value
During participation (a_{1k}^{23})	-.076	.005	.012	.036	.016	.039	.083*		
	(.048)	(.020)	(.022)	(.020)	(.023)	(.025)	(.031)	-.056	566.7
Post participation (a_{1k}^{13})	.079	.013	.031	.048*	.071*	.074*	.081*	(.217)	(480)
	(.044)	(.019)	(.021)	(.021)	(.024)	(.026)	(.027)		.004

† Difference with respect to the reference participant
* Significant at the 5% level

Table 5: Various Assumptions on the Variable Treatment Effect.
(standard errors in parentheses)

	(1) Benchmark	(2) Constant Treatment	(3) Uniform Distribution	(4) Normal Distribution
Impact of Training				
Before participation $\left(l^{1v}\right)$	-.416 (1.899)	-1.686 (1.937)	-1.687 (1.937)	.287 (2.836)
During participation $\left(\alpha_1^{23}\right)$				
ATE	-.058 (.053)	-.047 (.576)	-.023 (.267)	-0.835 (1.669)
SATE	-.262 (.240)	-.047 (.576)	-.045 (.525)	-.962 (1.288)
Standard error $\left(q_1^{23}\right)$	-	-	-	.056 (.346)
Post participation $\left(\alpha_1^{1v}\right)$				
ATE	.118* (.058)	.789 (.598)	.402 (.280)	-1.449 (2.107)
SATE	.548* (.269)	.789 (.598)	.790 (.550)	-.172 (1.352)
Standard error $\left(q_1^{13}\right)$	-	-	-	.555 (.469)
Mean selection bias $r_1^1 = r_1^2$	-.048 (.243)	-.289 (.560)	-.290 (.515)	.628 (1.100)
WSSR (degrees of freedom)	530.84 (498)	529.18 (498)	528.84 (498)	516.80 (496)
P-value	.149	.161	.164	.258

* significant at the 5% level