Self-employment grants vs. subsidized employment: Is there a difference in the re-unemployment risk?

by

Kenneth Carling* and Lena Gustafson*

October 7, 1999

Abstract

Self-employment grants and employment subsidies are active labor market programs that aim at helping unemployed workers to escape unemployment by becoming self-employed or being hired at an initially reduced cost for the employer.

In Sweden in the 1990's the participation rate in the self-employment program increased from virtually none to almost same as in the employment subsidy program. The advancement of the self-employment program is likely to be a result of (i) a change in the labor market program policy, and (ii) an increase in the supply of skilled unemployed workers. The justification for the policy change is unclear, however. The literature indicate that a rather specific group of unemployed workers may benefit from self-employment programs; Neither are there any strong reasons to believe in general that self-employment should be preferable to conventional employment through subsidies.

We examine, *ex post*, the justification for the policy change by comparing the post-program duration of employment for the two programs. In addition, we focus in some detail on the outcome for female workers and workers of foreign citizenship. The reason for this is the explicit policy to direct those workers to self-employment.

The data we study are the inflow to the two programs from June 1995 to December 1996. The program participants are followed to March 1999. The data contain detailed spell and background information on 9,043 unemployed workers who participated in the self-employment program and 14,142 who participated in the employment subsidy program.

The second explanation, see (ii), for the increase in self-employment program implies a potentially serious selection problem. We discuss how the selection process may bias the effect estimate in the non-linear duration model that we use. Simulations help us to determine the magnitude of the selection bias in our application. Moreover, we exploit the existing behavioral heterogeneity across labor market offices to reduce the selection bias.

We find that the risk of re-unemployment is more than twice as high for the subsidized employment program compared with the self-employment program. The large positive effect is, however, limited to male and female workers of Swedish origin. We thus conclude that the policy change in general has been successful, though we note that directing immigrant workers to self-employment is unlikely to improve the situation for this group of unfortunate workers on the Swedish labor market.

Keywords: Empirical Bayes methods, Employment duration, Program evaluation, Proxy variables, Selection bias, Simulations

---

[*] *IFAU, Office of Labor Market Policy Evaluation, P.O. Box 513, S-751 20 Uppsala, Sweden.*

We are grateful to Laura Larsson, Henry Ohlsson and Maria Melkersson for their suggestions and comments on earlier drafts. We are also grateful to participants of the IFAU workshop and seminar participants at Department of Economics, Umeå University for providing helpful comments.

1. Introduction

The self-employment program provides a mean for Swedish unemployed to escape unemployment by setting up their own business. To be eligible for such grant it is required that the worker is registered as unemployed and the Labor Market Officials approve the proposed business plan. The self-employment program (hereafter the SEMP-program) can be compared to subsidizing workers' initial spell of employment (SUBE-programs), where in principle the programs differ only in the employer.

In the 1990's, the number, as well as the proportion, of unemployed workers who receive self-employment grants has increased drastically[1]. The increase reflects most likely a change in the labor market policy rather than a shift in the preferences among the unemployed workers. For example, before 1993, the self-employment program was an alternative that was considered only after having participated in other programs. In 1993 the self-employment program was given priority, a policy change that is likely to have raised the participation rate.

Another possible explanation to the rise in SEMP-participation is the increase in unemployment for skilled workers. Figure 1 shows the Swedish (total) unemployment rate and the entering rate to the SEMP-program during the period 1985-1997. The Labor Market Officials have met the increase in the unemployment rate in the 1990s by increasing proportionally the supply of programs, the program-to-unemployment ratio has been about one third over the period 1985-97. The figure shows that the increase in SEMP-participation coincides with the increase in overall unemployment. Consequently the new SEMP-participants come from a larger stock of unemployed. This stock has better skills on the average than previously and we also believe this to be a reason for the increased likelihood of entering self-employment. The idea that a larger stock of unemployed has better skills on the average is for instance supported by the fact that the average education level among the unemployed has risen considerably during the 1990s.[2] Also, the SEMP-program is probably directed to more skilled workers since the

[1] It is estimated that currently 30 per cent of all new born Swedish firms are owned and run by this group of workers (Statistics Sweden, 1998).
[2] See for example Edin and Holmlund, (1994), p 16.

program is more demanding than other programs. If the typical SEMP-participant is more qualified than other program participants, then we may face a serious selection problem. This will be discussed in Section 3.3.

The objective of this paper is to compare the efficiency of the self-employment program with the traditional subsidized employment program and therefore evaluate the policy that has promoted participation in the SEMP-program.

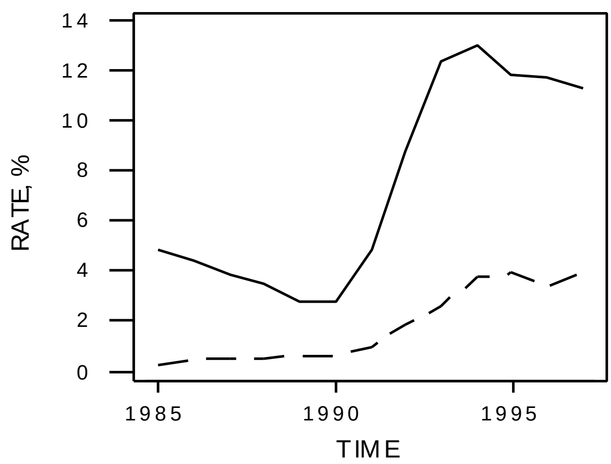

Figure 1. Swedish (total) unemployment rate and the SEMP-program's share of all labor market programs during the period 1985-1997.

Some Swedish experience indicates that subsidized employment programs, as well as other traditional programs, work rather poorly for immigrant workers - a group of workers who have a disproportionate high rate of unemployment. For various reasons it has been argued that the immigrants might perform relatively much better through self-employment programs compared with, e.g., subsidized employment programs. Indeed, it has been an explicit goal of the labor market policy to direct unemployed immigrants to self-employment (Riksdagens Revisorer, 1997). For instance, immigrants may have their grant period extended to 12 months instead of the usual duration of 6 months. We will for this reason in some detail examine whether in fact this has been achieved and if it has had the expected positive result.[3] Women is another group that can have their grant periods' extended to 12 months. We therefore also check the outcome for women.

[3] It is not obvious what definition of *immigrants* should be used. In this study we use foreign citizenship.

We attempt to study the success of the policy regarding self-employment and subsidized employment among unemployed workers by examining the risk of re-unemployment for workers who entered the programs under the period June 1995 to December 1996. Our definition of a successful outcome - not returning to unemployment - is justified by the belief that holding a job is drastically better than being unemployed, whichever type of employment it may be.[4] [5] There are some compelling reasons to believe in such a claim, the foremost important one is that most parts of the Swedish Work fare system presupposes employment, e.g., pensions, subsidized day care, unemployment and health insurance. An alternative approach in this kind of study, is to focus on the program effects on future wages. We had no possibility to study future wages in our data set. Furthermore, the expressed goal of the labor market policy is to reduce the registration periods at the unemployment offices and therefore we stick to this outcome measure.

The relevance of this paper extends beyond the Swedish case as many OECD countries provide the self-employment program to unemployed workers (OECD, 1995). The OECD report shows that the participation rate[6] in self-employment programs ranges from 0.5 to 4 percent, for various OECD countries, with a median rate of approximately 2 percent. The question is however, if unemployed workers would benefit more by being hired by an external employer than being self-employed.

The paper is organized as follows. In Section 2 we give a background to the Swedish active labor market policy, and discuss in some detail the setting for the two competing programs and show the historical evolution of the programs. In section 3 we present the statistical methodology and describe the data used in the empirical analyses. We also examine the determinants of the choice of program to participate in and, finally, we dis-

---

We are well aware that immigrants may receive Swedish citizenship, but due to limited data we could not discriminate between these citizens.

[4] A positive secondary effect of the self-employment program would arise if the self-employed hired other unemployed workers. We are abstracting from this effect because of some evidence that the effect has historically been small in the Swedish case (Riksdagens Revisorer, 1997) (US studies have found both a positive and a negligible effect, Vroman, 1997).

[5] Commonly the firm survival rate is reported (OECD, 1995). It seems, however, reasonable to consider the transition from self-employment to regular employment as a successful outcome, thus shifting the focus from firm survival to the attachment to the self-employment/employment state.

[6] The rate is calculated as the number of individuals receiving the self-employment grant divided by the number of individuals receiving regular UI benefits.

cuss selectivity issues. Section 4 gives the results and section 5 ends the paper by a discussion of side effects and results.

2. Additional background information and the institutional setting

Sweden has traditionally attempted to fight unemployment by active labor market policy. A cornerstone in the policy has been the National Labor Market Board (AMS), and its local offices, which has had many measures to its disposal to reduce unemployment, e.g. labor market training. The offices also monitor the unemployment insurance system. Consequently, detailed records of all unemployed have been kept[7]. Furthermore, it is mandatory for employers to report vacant positions to the office, and hence information on job-matches ought to be well recorded.[8]

It appeared for many years that the Swedish policy was successful in keeping a low unemployment rate, however, in the early 1990's unemployment rose. In the end of the 1990's the unemployment rate still remains high by Swedish standards. The question is whether labor market programs have succeeded in reducing unemployment. Given the empirical facts presented in Figure 1, one might be tempted to answer 'no' to this question. However, many would attribute the failure in reducing the unemployment rate to the lack of demand for labor caused by a drastic reduction in governmental spending, (see Modigliani et al, 1998). We will not focus on the question whether labor market programs in general are effective. Instead, we will compare the programs self-employment grants and subsidized employment and measure their relative effectiveness.

It is sometimes argued that unemployed are the least likely in the labor force to succeed as self-employed. On the other hand, US evaluations of self-employment programs to dislocated workers show that the program can be quite successful, at least for the specific subgroups of unemployed workers that the program historically has been targeted to (Katz, Stanley, and Kruger, 1998 and Vroman, 1997).[9] US evaluation studies based on experimental data and data from 'natural experiments' also suggest that subsidized em-

---

[7] A validation study by Statistics Sweden (1993) shows that more than 90 percent of those who self-reported as unemployed in the labor force surveys in August-October 1992 also registered at the public employment agency.

[8] About 15 per cent of the workers is lost to follow up. It is estimated that about 50 per cent of those actually was lost because of employment (Bring and Carling, 1998).

ployment programs are effective if they are targeted at specific groups of unemployed workers. There are also some non-experimental European studies of subsidized employment programs. The German study by Kraus et al. (1997) and the British study by Payne et al. (1996) both indicate positive results for the programs. In the German study, the studied outcome is non-employment hazard and the British evaluation studies the employment rate.

Table 1. Background information about the self-employment program (SEMP) and the subsidized employment program (SUBE).

Program:	SEMP	SUBE
Unemployment compensation:		
Eligible to UB	80% of earlier income, max 580 SEK/day	Collective agreement
Eligible to CA	240 SEK/day	Collective agreement
Not eligible	103 SEK/day	Collective agreement
Type of activity/employment: %[a]		
retail trade	32	23
consulting services	23	18
personal & cultural services	20	13
manufacturing	10	22
construction	10	15
Program cost[b]		
cost/participant & month	10.300 SEK	8.400 SEK
Stock of participants[b]	10,000	15,000
Duration of program, months[c]	6	4-6

Note: (a) Åtgärdsundersökningen 1998. (b) The figures correspond to the fiscal year 1994/95. (c) Some statements from the unemployment offices are that the duration of the SUBE-program sometimes depends on whether the employer seems serious in his intentions of a post-program employment or not. If not, i.e. if the office suspects that the employment will not continue after the subsidized period, the program duration might be shorter. For women and immigrants participating in the SEMP-program, the program period may be extended with 6 months.

To be eligible for the SEMP- grant it is required the worker is registered as unemployed and that the labor market officials approve the proposed business plan. The grant is provided for six months and is the equivalent of 80 per cent of the pre-unemployment

[9] Wong et al (1998) find an overall positive effect of the Canadian self-employment program, but the effect varies over subgroups of participants.

earning if the person is entitled to unemployment benefit. As is shown in Table 1, the maximum amount of unemployment benefit is 580 SEK per day and if the person is entitled to cash assistance, CA, he receives 240 SEK per day and if he is not entitled to any unemployment income (UI) he receives 103 SEK per day.

Concerning the SUBE-program, it is also required the worker is registered as unemployed and the program is approximately of the same duration as the self-employment program. Participants in the SUBE-program are paid according to the collective agreement. The employer is compensated with a maximum of 350 SEK per day or at most 50 % of the wage cost. If the person is only entitled to CA or is not entitled at all, the reimbursement can be higher for those participating in the SUBE-program than for those in the SEMP-program. On the other hand, the SEMP-grant is of course a minimum income since the business may generate some additional income. We believe it is a reasonable approximation to say that the participants of both programs will earn about the same.

For the SEMP-program, the majority types of activity for which the firms were registered are retail trade (32 %), consulting services (23%), personal and cultural services (20%) manufacturing (10%) and construction (10%). For the SUBE-program, most jobs are in the private sector and the largest fraction is found in retail trade, (23%), thereafter follows manufacturing, (22%) and consulting services (18%).[10]

The government expenditure for the SEMP-program was about 10.300 SEK per participant and month, while for the SUBE-program the equivalent cost was 8.400 SEK.[11] Table 1 provides an overview of the key features of the two programs.

Figure 2 shows the entering rate to the self-employment program. The rate is calculated as the number of workers entering the program, during a given year, divided by the sum of the workers entering the self-employment program and the subsidized employment program during the same year. As time has progressed, the self-employment program has increased its share from zero to about 40% of these two programs and the question is whether this is a well justified progress.[12]

[10] Åtgärdsundersökningen 1998.
[11] The figures correspond to the fiscal year 1994/95.
[12] The participation in the SEMP and SUBE programs adds up to about 10 % of the total labor market program participation.

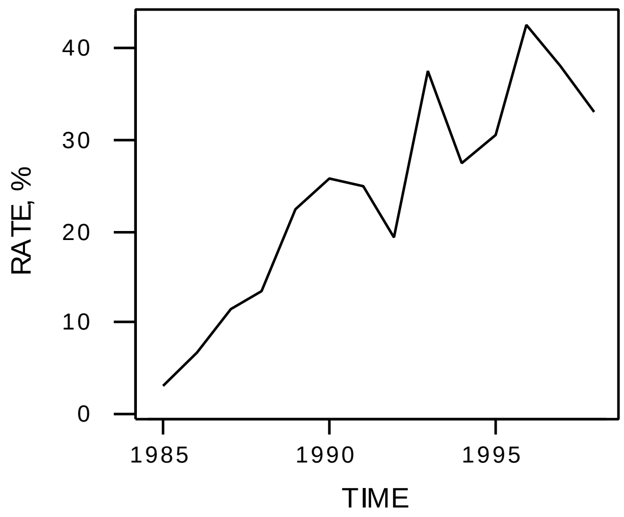

Figure 2. Entering rate to the self-employment program, calculated as the number of workers entering the program, during a given year, divided by the sum of the workers entering the self-employment program and the subsidized employment program during the same year. The figure shows the rate for the period 1985 to 1997.

There is a fundamental question: Are the two programs comparable? The idea of running a business on the one hand, and being an employee on the other, the differences seem more significant then the similarities. This is a general problem when it comes to evaluating labor market programs. However, when looking at how the two activities are implemented as programs, we find that the comparability grows. The duration of the programs are the same, the reimbursements are approximately the same and most importantly, there is the fact that both programs involve employment that is intended to last even after the program period ends. In this view, we argue that if the SEMP-program is to be compared to any program, the SUBE-program is the most natural choice.

Another fundamental question is whether participants in the SEMP-program are comparable to the participants in the SUBE-program. Tables 2a-2b give background information about the participants. Most striking is the difference in educational level and work experience between the two groups. Such differences might lead to selection bias in the empirical analysis and the issue of the next section is to explore how this bias can be made miniscule.

Table 2a. Descriptive statistics for the participants in the self-employment program. The variables used in the empirical analysis enters the table.

Variable	Mean	St Dev	Min	Q1	Median	Q3	Max
Age (in years)	36.8	11.6	18	27	36	46	66
Gender (Female=1)	.352	.478					
Disability	.062	.241					
Historical spell info							
Earlier program participation[a]	.552	.247					
Dur. of unemp. prior to program, days/100	5.11	5.29	0	.980	2.98	7.86	40.6
Educ. and work experience							
Compulsory school	.278	.201					
Upper sec. (Max 2 years)	.231	.178					
Upper sec. (> 2 years)	.238	.181					
University (Max 2 years)	.070	.065					
University (> 2 years)	.183	.150					
No work experience	.143	.122					
Some work experience	.128	.111					
Long work experience	.730	.197					
Unemployment compensation							
Eligible to UI	.329	.221					
Eligible to CA	.081	.075					
Not eligible	.590	.242					
ln(Wage)[b]	3.83	3.26	0	0	6.31	6.60	8.47
Daily allowance, SEK/day	305	239	0	0	383	563	580
Citizenship:							
Swedish	.917	.076					
Western Europe	.032	.031					
Eastern Europe	.021	.020					
Asia, Africa, South America	.031	.030					
Local labor market structure							
ER[c]	.094	.063	0	.054	.080	.109	.340
Herfindahl's index[c]	.076	.015	.053	.065	.073	.082	.181

Note: (a) Earlier program participation =1 if the person has only been in the job-seeker categories 11-14 before participating in the actual program. (b) ln(wage) : the wage variable comes from the event table AKSTAT. It is expressed as the logarithm of the daily wage in SEK and refers to the wage the person had during the period preceding the unemployment period. (c) See Section 3.3.

Table 2b. Descriptive statistics for the participants in the subsidized employment program. The variables used in the empirical analysis enters the table.

Variable	Mean	St Dev	Min	Q1	Median	Q3	Max
Age (in years)	37.1	11.5	18	27	37	46	66
Gender (Female=1)	.323	.468					
Disability	.082	.274					
<u>Historical spell info</u>							
Earlier program participation[a].	.587	.242					
Dur. of unemp. prior to program, days/100	6.53	5.37	0	2.04	5.03	10.1	59.4
<u>Educ. and work experience</u>							
Compulsory school	.279	.201					
Upper sec. (Max 2 years)	.227	.176					
Upper sec. (> 2 years)	.373	.234					
University (Max 2 years)	.040	.038					
University (> 2 years)	.081	.074					
No work experience	.317	.217					
Some work experience	.217	.170					
Long work experience	.466	.249					
<u>Unemployment compensation</u>							
Eligible to UI	.370	.233					
Eligible to CA	.142	.123					
Not eligible	.488	.250					
ln(Wage)[b]	3.12	3.24	0	0	0	6.46	8.40
Daily allowance, SEK/day	265	229	0	0	245	494	580
<u>Citizenship:</u>							
Swedish	.821	.147					
Western Europé	.039	.038					
Eastern Europé	.103	.092					
Asia, Africa, South America	.037	.036					
<u>Local labor market structure</u>							
ER[c]	.090	.059	0	.049	.080	.109	.340
Herfindahl's index[c]	.078	.014	.053	.067	.074	.083	.181

Note: (a) Earlier program participation =1 if the person has only been in the job-seeker categories 11-14 before participating in the actual program. (b) ln(wage) : the wage variable comes from the event table AKSTAT. It is expressed as the logarithm of the daily wage in SEK and refers to the wage the person had during the period preceding the unemployment period. (c) See Section 3.3.

3. Methods and data

In this section, the method and the data we use to study the effect of the SEMP-program relative to the SUBE-program will be discussed. We also derive the determinants of the program participation and define a variable, which will serve as a proxy for unobserved factors. The intent is to reduce the potential selection bias.

3.1 Methods

The construction of the programs implies that a spell of employment will proceed the program. The question is to what extent the programs are successful in providing long-term employment. The worst case would be that the programs are immediately followed by re-unemployment.[13]

Let the random variable T_e be the employment duration, be that also employment at other employers than the one of the program, until the individual returns to unemployment. Let also D take on unity if the individual participated in the SEMP-program and zero if the individual participated in the SUBE-program. In general, the key parameter to identify, assuming a homogeneous effect, is

$$\text{(1)} \qquad \Pr[t_e / D],$$

which sometimes is simplified in empirical analysis to be the expected value of the distribution, $E[t_e / D]$. We prefer however to consider the hazard rate, $l[t_e / D]$, which is equally general to (1) since the hazard has a simple relation to the probability distribution. The relation is

$$\text{(2)} \qquad l[t_e / D] = -\frac{\partial \ln(1 - \Pr[t_e / D])}{\partial t_e}.$$

[13] Interrupted program spells are uncommon, perhaps reflecting the fact that neither the employee nor the employer has any incentives to terminate the employment prior to its end. Hence, we will discard information on the duration of the programs.

We prefer the hazard function because of its natural appeal in studies of duration phenomena. The relative effect of participating in the SEMP-program could be determined as

$$(3) \qquad \frac{l\left[t_e / D=1\right]}{l\left[t_e / D=0\right]}=a\left(t_e\right)$$

and if this ratio is constant for all values of t_e, it would be sufficient to identify $a \equiv a\left(t_e\right)$.

It is however likely that there are other factors that determines T_e and that the programs are heterogeneous with respect to these factors. An appropriate analysis requires that these factors are controlled for and we therefore extend the hazard in (2) to include such factors:

$$(4) \qquad l\left[t_e / D, x, y, z, v\right]=a\left(t_e\right) l_0\left(t_e\right) m(x, y, z, v ; \Omega)$$

where $l_0\left(t_e\right)$ is a baseline hazard and $m(.)$ is some function which links the factors to the duration variable. We think of x as attributes of the individual, y as regional factors that may effect the duration, z as structural features of the local labor market, and v[14] as residual factors that we can not control for. Finally, Ω is a vector of parameters associated to the factors.

It is necessary to put additional structure to (4) before proceeding to the final empirical analysis. We begin by considering the parameter $a\left(t_e\right)$. In figure 3, the empirical hazards for the two programs are shown. It can be noted that the hazard for the SEMP-program is monotonically decreasing, whereas the hazard for the SUBE-program rises initially before starting to fall. We will, nevertheless, impose the restriction that $a=a\left(t_e\right)$ neglecting the initial departure from a time-constant ratio.[15]

[14] It is our hope that the heterogeneity arising from v is small, though we can not emprirically determine whether it is small or not. In Section 3.3 we discuss how a proxy for v can be obtained.
[15] Some exploratory analyses where we accounted for a time-varying ratio confirmed that the imposed restriction works reasonably well as an approximation.

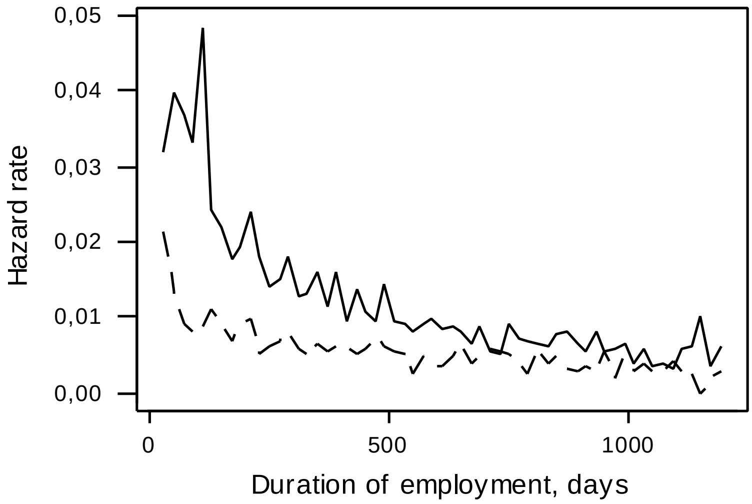

Figure 3. Empirical hazard for the duration of employment, conditional upon survival the first 20 days. The solid line shows the hazard rate for participants in the SUBE-program and the dashed line the rate for the participants in the SEMP-program.

The second step is to disclose the functional form of $m(.)$. Duration data and duration models are difficult in this sense.[16] We prefer to perform the preliminary data analysis on the complete observations, i.e. for those the actual duration was observed. Selecting the covariates in Table 2a-2b and linking these to the duration variable has been done with the help of exploratory tools for ordinal and categorical variables and a number of non-parametric regressions for the continuous variables.[17] This decision has been a compromise between the desire to reach effective linking and to restrict the number of parameters. After these two steps we can specify (4) as

(5) $$l\left[t_e / D, x, y, z, v\right] = a\ l_0\left(t_e\right)\exp\left(x * \mathrm{B} + y * \mathrm{A} + z * \Lambda\right) g(v)$$

[16] Altman and de Stavola (1994) provide a careful discussion on available techniques.

[17] Exploratory tools for ordinal and categorical variables are treated by Hoaglin, Mosteller, and Tukey (1985). For literature on non-parametric regressions, see Cleveland (1979), Cleveland, Devlin, and Grosse (1988), and Härdle (1990).

where B, A, and Λ are parameters associated to the three types of factors and an asterisk, *, denotes the transformed version of the factors. This model, which is of the semi-parametric Proportional Hazard type,[18] can now be estimated by use of the principle of Maximum Likelihood, provided that v is eliminated.

3.2 Data

A comparison of the two programs requires a sample containing spell information as well as background information on the participants. Moreover, it is necessary to obtain information on the subsequent unemployment spells. To achieve this, we rely on several merged databases maintained by the Labor Market Board. From these databases we have event history information on the transitions between the states of unemployment, labor market training and employment, as well as individual specific information of demographic and financial type, and weekly unemployment earnings.

We have collected information on all individuals who commenced one of the two programs in the period June 1995 to December 1996, provided that they were in their first registration period at the unemployment office. We follow these individuals until they once again register as unemployed at labor market office or, at the most, until March 1999. The spell information is retrieved from a database maintained by the Labor Market Board, known as HÄNDEL. We use the database to determine, in addition to background information on the individual, the unemployment duration prior to the date the program commenced and to determine the duration of the program, as well as the duration of the employment that succeeds the program.

Information on working hours, unemployment benefits, and income prior to unemployment is obtained by additional merging to data maintained by officials for the unemployment insurance system, a database known as AKSTAT.

After deleting aberrant observations we have data consisting of 23,185 individuals, where 14,142 participated in the subsidized employment program and 9,043 in the self-employment program. Table 2a-2b gives sample statistics for the program participants.

Due to the sampling scheme, it is inevitable that some observations will be incompletely observed. This happen if the individual is still employed at the time we select the

[18] See Meyer (1990), Narendranathan and Stewart (1993), and Carling et al. (1996) for earlier applica-

observations. For those individuals we use the fact that their spells were still ongoing at the time of censorship.[19]

3.3 Selection issues: What determines the program participation?

During the period under study, 39 % of the individuals enrolled the SEMP-program and 61 % the SUBE-program. According to what rule where these individuals assigned to one of the two programs? This question needs to be examined for, at least, two reasons; firstly, there is presumably a selection on basis of attributes of the individuals which might mask the relative performance of the programs, and secondly a heterogeneous selection process across the local labor market offices might point at administrative inefficiencies. A problem arises if the administrators fail to predict the outcome of the programs or fail to act on basis of this prediction or both. Heterogeneity across offices might show that some offices are less efficient in predicting and acting on the outcome.

3.3.1 The selection bias in a non-linear duration models

Heckman, LaLonde, and Smith (1998) give an extensive presentation of the potential selection problem in evaluation studies. Assume a linear outcome model with a homogenous effect, supposedly capturing the effect of moving an individual from the non-participation state to participation state. The effect will, in general, be over-estimated if the participation decision is driven by unobserved factors that also are positively related to the outcome.

The prediction for linear models does, however, not carry over to our non-linear model. Rewrite the hazard in (5) so that $V \in (0,\infty)$ represents the combined effect of all unobserved factors and $x^*, y^*, z^* \in X$,

(5a) $$l\left[t_e / D, X, V\right] = \exp(a D)\, l_0(t_e) \exp(Xb)\, V .$$

It can be shown, if V is independent of X and D that (see Lancaster, 1990)

---

tions of this model.

[19] In principle, it is possible to distinguish between, for instance, full and part-time unemployment, hence suggesting multiple destinations after the employment spell. We do not discriminate between these destinations since the vast majority of individuals returned to full unemployment.

(5b) $$P\lim \hat{a} = a\left[1-\exp(Xb + a\,D)\Lambda_o(t)\frac{Var(V|T\ge t, X, D)}{E(V|T\ge t, X, D)}\right].$$

Thus, in absence of an unobserved selection process, the effect will be underestimated, unless all factors that determine the duration are controlled for.

It is clear that the risk of over-estimating the effect, in case the participation decision is driven by unobserved factors that also are positively related to the outcome, will partly be decreased by the tendency to underestimate the effect in hazard models. Moreover, the distribution of V will in general be right skewed, implying positive dependence between the variance and the expectation, which, in turn, would render an even more conservative estimate.

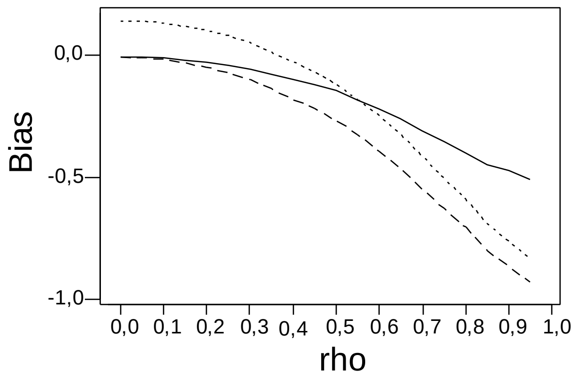

Figure 4. The bias in the effect estimate as a function of the correlation, rho, between the V-factor in the Outcome model and the participation model. The dashed line gives the bias for s=1 and the case where there is no actual effect. The solid line gives the bias for a similar setting, only that s=0.5. The dotted line gives the bias for s=1 and an actual effect equal to -1.

It is not easy to provide, analytically, narrow bounds for the selection bias, so we resort to simulations. In Appendix A, we give the details of the simulation experiment. Figure 4 shows the bias in the effect estimate under various assumptions about the strength of the unobserved selection process in our application.

The simulations show that the bias can be substantial because of non-random selection. We also note that the presence of an actual effect makes the selection bias less problematic. The simulations suggest that a bias in the effect estimate exceeding 25 % is unrealistic, for this to happen it would require the unobserved selection process to be incompatible with most empirical cross-sectional studies in the economic liter ature.

3.3.2 A proxy for the unobserved heterogeneity, V

In the data we have observations from 463 offices (of the 536 offices in the country), though many of these offices are, however, small. Figure 5 shows the variation in the proportion of the SEMP to SUBE participation across offices.[20]

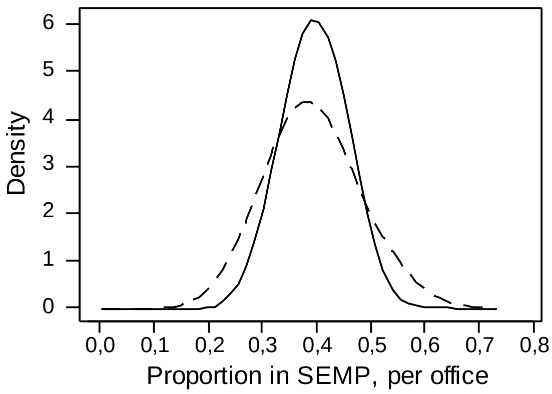

Figure 5. The distribution of the proportion of participants in the SEMP-program per office. The solid line shows the estimated distribution before standardization for factors that might explain the variation, and the dashed line shows the estimated distribution after standardization. The distributions are smoothed by applying a Kernel density est i-mator with the Epanechnikov Kernel and a bandwidth, h=0.15 (see Härdle, 1990).

However, some, or perhaps all, of the variation may occur because of heterogeneity in the individuals registered at the local offices or geographical or structural heterogeneity. The purpose of this analysis is to standardize for these factors so that the office specific variation can be determined.

To do so we apply a logistic regression model to examine the determinants of the program of participation, and to standardize the participation rate . Formally, again, let *D*

[20] The variation is greatest for small offices, presumably because of random variation rather than sy s-tematic variation. To adjust for the random variation and to obtain better estimates of the actual office specific proportion, we have applied the Beta/binomial method in an Empirical Bayes spirit (see Carlin and Louis, 1996). The unequality in size is accounted for by the approach suggested in Louis and Der-Simonian (1982), (see also Andersson, Carling, and Mattsson, 1998).

take on unity if the individual enrolled the SEMP-program and zero otherwise. For the definition of factors we refer to Section 3.1. The model is expressed as,

(6) $$\ln \frac{\Pr[D=1/P_j, x, y, z]}{\Pr[D=0/P_j, x, y, z]} = P_j + m(x, y, z; \Omega),$$

where the discussion in Section 3.1 applies for the $m(.)$ function and its arguments and P_j represents the participation rate in the SEMP-program at office j after standardization. The individual attributes, x, do not need a detailed discussion, we simply note that we have access to many variables that one would expect to be important.

The regional variable, y, is the county. We expect this factor to be important for several reasons. The offices are autonomous but serve under a county administration. At the county level policy recommendations and budget restraint are decided upon, hence one would expect homogeneity of the offices within the county, though not necessarily so between counties. Moreover, the organization and the culture of Swedish counties suggests that commuting within the county is feasible, but between county commuting infeasible. Thus, we expect the local employers to do most of their recruiting from the work force in the county. Likewise, we expect the self-employed to primarily view the county as his home-market (recall from Section 2 that most of the firms were set up in sectors which are local to their character). It can also be noted that some macro-variables, for instance, the unemployment rate varies greatly between counties. This may have an effect on the propensity to offer different types of programs.

The structural variables, z, are difficult to measure and to provide a prior belief about. We consider the Herfindahl's index (Petersen, 1993), per municipality, of the number of employed per sector, as one such variable. It appears likely that the diversity of the local infrastructure may play a role in the propensity to set up a new enterprise. Our belief, although very weak, is that the less diversified the market the more focused is the office on serving the main employers with recruits, hence implying that a high *HFI* ought to be correlated with a low proportion of participants in the SEMP-program. We also consider a measure of the intensity in the job creation and job destruction in the municipality where the participant lives, ER.[21] Again, our priors are weak but perhaps a high intensity might imply a proneness to endure in risky enterprises and thus imply a high proportion of participants in the SEMP-program.

Table 3 reports the estimates of the logistic regression model. There are several fa c-tors increasing the probability of participating in the SEMP-program that concur with previous findings. Among these are UI-eligibility, long previous work experience, higher education and a previously high wage. However, it also seems that women and citizens from Asia etc. are more likely to enroll in the SEMP-program. Recall from Section 1 that women and immigrants may have their program period's extended, i.e., this finding is not surprising and in the next section we will consider more in detail the outcome of the pr o-grams for these two groups.

We also note in Table 3 that the z-factors, which were supposed to measure the diversity and dynamic of the local market, entered the model in line with our priors, though they are unimportant for determining the participation outcome.

In figure 5 the distribution (footnote 20 applies) of the residual office specific propo r-tion, P_j, is shown. It is evident from the figure that a great deal of the variation in the program of participation can be attributed to the variation across offices. We suggest that there are three hypotheses that can be used to explain this variation:

Hypothesis 1. The propensity to offer the SEMP-program is identical in all offices, i.e. $\Pr[D/P] = \Pr[D]$. The variation occurs because of unobserved heterogeneity such that $f\lfloor V/P_j \rfloor \neq f\lfloor V/P_{\bar{j}} \rfloor, j \neq \bar{j}$. The implication is that T_e ought to be positively related to P_j.[22]

Hypothesis 2. There is no difference in unobserved heterogeneity in that $f\lfloor V/P_j \rfloor = f\lfloor V/P_{\bar{j}} \rfloor, \forall j, \bar{j}$. The propensity to offer the SEMP-program varies across offices, i.e. $\Pr[D/P] \neq \Pr[D]$.[23] The implication is that T_e ought to be unrelated to P_j,

---

[21] ER=(C-D) + |(C – D)|, where C is the ratio of the constructed jobs to the labor force in a municipality. D refers to the ratio of destructed jobs, see Andersson (1999).

[22] Provided, of course, that our presupposition that workers with better unobserved attributes are more likely to be in the SEMP-program is correct, otherwise the implication would be the other way round.

[23] This diversity of strategy might be caused by differences in collected anecdotal evidence, motivating the office members to systematically lead the individuals into either the SEMP-program or the SUBE-program according to the preference of the office.

whereas $(T_e|D=1)$ should be negatively related because of a less aggressive screening process in offices with high P_j.

Hypothesis 3. There is no difference in unobserved heterogeneity in that $f\lfloor V/P_j \rfloor = f\lfloor V/P_{\bar{j}} \rfloor, \forall j, \bar{j}$. The propensity to offer the SEMP-program is related to the quality in the program implementation. The implication is that T_e ought to be unrelated to P_j, whereas $(T_e|D=1)$ should be positively related.

Whichever hypothesis be true, it is clear that the factor should be conditioned on in the duration analysis. The resulting parameter estimate will give credibility to some of the hypotheses

4. Estimation and results

In this section we report the empirical results. We estimate the model

(7) $$l\left[t_e / D, x, y, z, P\right] = \exp(a\,D)\, l_0(t_e) \exp\left(x^* \mathrm{B} + y^* \mathrm{A} + z^* \Lambda\right) \exp\left(dP_j\right)$$

where α is the effect parameter of interest and P_j will, potentially, serve as a proxy for *V*. By assuming a piece-wise linear hazard in each time-interval, the discrete time hazard can be written as

(8) $$h[t_e / D, x, y, z, P] = 1 - \exp\left\{- \exp\left[a\,D + x^* \mathrm{B} + y^* \mathrm{A} + z^* \Lambda + dP_j + h(t_e)\right]\right\},$$

where $h(t_e) = \ln\left(\int_{t_e}^{t_e+\Delta_k} l_0(w)dw\right)$ and Δ_k is the length of the *k*:th time-interval. We consider intervals of 20 days ranging from zero to 1200 days. The log likelihood function,.

Table 3. Logistic regression model for participation in the self-employment program, (SEMP). St. errors in parentheses.

Variable	Estimates
Age (<30)	
30-39	.082 (.039)
40 -	.063 (.036)
Gender (Female=1)	.222 (.032)
Disability	.038 (.059)
<u>Historical spell information</u>	
Earlier program participation	.132 (.045)
Duration of unemployment prior to program	-.238 (.013)
- squared	.010 (.001)
<u>Education and work experience</u>	
(Compulsory school)	
Upper sec. (Max 2 years)	.033 (.041)
Upper sec. (> 2 years)	-.251 (.041)
University (Max 2 years)	.630 (.070)
University (> 2 years)	.889 (.052)
(No work experience)	
Some experience	.212 (.050)
Great deal of experience	1.09 (.041)
<u>Unemployment compensation</u>	
(Uninsured)	
Eligible to CA	.087 (.058)
Eligible to UB	.352 (.176)
ln(Wage)	-.765 (.091)
ln(Wage)-squared	.121 (.012)
<u>Citizenship</u>	
(Swedish)	
Western Europe	-.249 (.081)
Eastern Europe	-1.30 (.085)
Asia, Africa, South America	.388 (.085)
<u>Local labor market structure</u>	
ER	.516 (.278)
HFI	-.644 (2.89)
-.075	.015 (.059)
.075 - .125	-.123 (.131)
.125 -	.444 (.290)

Note: The county factor also enters the model. The estimates for these 21 counties are not shown due to limited space.

given the model in (8), for a sample of n random observations on T_e and c is

(9)

$$\ln LL(a, \mathrm{B}, \mathrm{A}, \Lambda, d, h) = \sum_{i=1}^{n} \{c_i \ln(1 - \exp\{-\exp[a D_i + x_i * \mathrm{B} + y * \mathrm{A} + z_i * \Lambda + dP_{j,i} + h(t_{e,i})]\})\}$$
$$- \sum_{s=1}^{t_{e,i}} \exp[a D_i + x_i * \mathrm{B} + y * \mathrm{A} + z_i * \Lambda + dP_{j,i} + h(s)]$$

where $c_i = 1$ if the employment was observed to be terminated for a new spell of unemployment and zero otherwise. The function is maximized with respect to its arguments and the resulting estimates of B and d are presented in Table 4a.[24] Excluded from the table are *y*, which is a factor with levels equal to the 21 counties in Sweden, and the structural variables, *z*, which were assumed to determine participation but not employment duration.[25] The results are not surprising, the hazard rate of re-unemployment is lowest for experienced, well-educated, and middle-aged Swedish men.

To identify the effect parameter we consider several models. In the first simple model we control only for program participation. We find in this case that $\hat{a} = -0.927$ implying a much higher risk of re-unemployment for the participants in the SUBE-program (see Table 4b). This effect estimate might however be unreliable due to selection bias. Table 4b shows the effect estimates when the model is stepwise extended to reduce the heterogeneity amongst the participants. The effect estimate decreases to $\hat{a} = -0.752$ in the most extensive model where the office specific heterogeneity has been controlled for. The estimate implies that the hazard rate for re-unemployment is more than 110 percent higher for the SUBE-program than for the SEMP-program.

Table 4a. Hazard regression model for the employment duration.

---

[24] Starting values are obtained from the Approximate Maximum Likelihood method (Carling, 1995), and used in conjunction with the BHHH algorithm (see Carling and Söderberg, 1998).

[25] We have performed some additional sensitivity analysis to confirm the results: We have, one at the time, excluded (i) all workers with reported working disabilities, (ii) workers who had participated in labor market programs prior to the studied one, and (iii) workers who were not entitled to unemployment benefits. The rationale for the latter (iii) is that the uninsured would increase their income relatively more by entering the SUBE-program and that there are reasons to presume that uninsured workers are less inclined to report re-unemployment to the labor market offices that administer the records.

St. errors in parentheses.

Variable	Estimates
Age (<30)	
30-39	-.030 (.033)
40 -	-.034 (.030)
Female	.238 (.027)
Disability	-.018 (.057)
Historical spell info	
Earlier program participation	.138 (.039)
Duration of unemployment prior to program	.093 (.011)
- squared	-.003 (.001)
Education and work experience	
Upper sec. (Max 2 years)	.071 (.036)
Upper sec. (> 2 years)	.031 (.035)
University (Max 2 years)	-.062 (.063)
University (> 2 years)	-.191 (.049)
Some experience	-.017 (.036)
Great deal of experience	-.281 (.033)
Unemployment compensation	
Eligible to CA	.039 (.045)
Eligible to UI	-.192 (.168)
ln(Wage)	.392 (.084)
ln(Wage)-squared	-.057 (.011)
Citizenship	
Western Europé	.087 (.066)
Eastern Europé	-.136 (.053)
Asia, Africa, South America	.163 (.070)
Office specific heterogeneity	
P[a]	-1.16 (.266)

Note: Number of observations in SEMP:9,043 and SUBE: 14,142. The county factor also enters the model. The estimates for these 21 counties are not shown due to limited space. The reference category is a Swedish male of less than 30 years who attended only compulsory school and does not receive une m-ployment income.
(a) See Section 3.3.

Table 4b. Effects estimates after step-wise inclusion of factors. St. errors in parantheses.

	Basic model	Individual specific	Prior duration of unemployment	Wage and UB	Labor market conditions	Local structural factors	Office hetero-geneity
α	-.927	-.824	-.788	-.777	-.775	-.778	-.752
	(.028)	(.030)	(.030)	(.030)	(.030)	(.030)	(.030)

To make our results comparable to the evaluation parameter most often applied in the self-employment literature, we also show the implied survival rates for the two programs. Our almost four-years follow up gives that about 35 % of the self-employed and 60% of the subsidized workers have returned to unemployment four years after the program ended. This figure for the SEMP-program can be compared to previous international findings which are of similar magnitude (OECD, 1995). We have difficulties in relating the figure for the SUBE-program to similar studies but the figure appears high, and r e-mains so even if we leave out the cases where re-unemployment occurred immediately after the end of the program.

Table 5a. Interaction analysis: Program and Citizenship. St. errors in parantheses.

Citizenship	SUBE	SEMP	Effect, α
Swedish	0	-.802	-.802
		(.035)	
Western	.009	-.404	-.413
	(.079)	(.141)	
East europé	-.120	.055	.175
	(.057)	(.146)	
Asia, Africa,	.019	-.090	-.109
South Amer.	(.090)	(.136)	

Table 5b. Interaction analysis: Program and Gender. St. errors in parantheses.

Gender	SUBE	SEMP	Effect, α
Male	0	-.713	-.713
		(.038)	
Female	.269	-.545	-.814
	(.032)	(.040)	

Note: Diff in effect = 0.101 (.057).

Consider next the two subsamples: women and foreign citizens. Table 5a shows the e s-timated effect for various foreign citizens with respect to Swedish workers. It can be noted that the large difference between the programs almost exclusively can be attributed to Swedish workers. The difference between the programs appears small for foreign cit i-zens, something caused by foreign citizens in the SEMP-program having a much worse hazard rate than Swedish workers. Table 5b shows a comparison between men and women. The effect estimates appear to be quite similar, the difference in effect estimate is borderline significant but not big enough to rule out the possibility of sele ction bias.

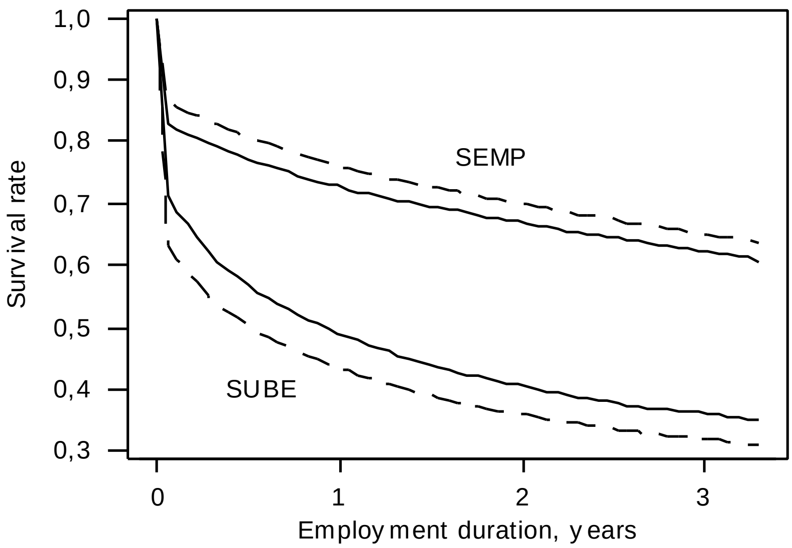

Figure 6. The survival rate of employment for the programs. The solid line gives the rate for a randomly selected worker, the dashed line for the average worker who entered the program.

In the last part of the previous section, we gave three hypotheses indicating how P_j might be related to the hazard rate of re-unemployment. Table 5c shows the estimates of the parameters associated with this variable. The estimate for the interaction term indicates that offices at which the standardized proportion in the SEMP-program is high are associated with long self-employment duration. Hence, there is support for the third hypothesis. However, there is also support for the first hypothesis, i.e. that the composition of unobserved characteristics of program participants varies across offices.

Table 5c. Interaction analysis: Program and office specific proportion in SEMP

Variable	Estimate	St. error
P_j	-0.778	(.312)
$\left(P_j - \overline{P}\right)\times D$	-1.40	(.596)

5. Discussion and conclusions

5.1 Side effects

Previous studies (Riksdagens revisorer 1997) show that the SEMP-program may involve undesired side effects. One such effect arises when the unemployed would have set up the business even without the self-employment grant. This is called "dead weight loss". Since one of the goals with active labor market programs is to reduce the costs involved with people being unemployed, the dead weight loss can be a substantial drawback to this goal. In the study referred to above, between 40-50 percent of the self-employed stated that they would have started their firms without the grant.

In this context, one might question whether we can draw any conclusions from this study, if the objective is to find which of the two programs is the most effective in order to reduce the costs for the public sector? Here, we would like to point out that a similar problem is encountered in the SUBE-program. In this type of program, a displacement effect and a dead weight loss may occur when someone would have been employed at the actual moment even without the subsidy. A survey from the Labor Market Board[26], asked what the participant thought would have happened to his work assignments if there had been no program. The report shows that 48 per cent thought that someone (himself or someone else) would have been employed to perform them.[27] The fact that there is likely to be side effects in both programs makes us believe that this problem can be overlooked, since this study investigates the relative effects of the programs.

5.2 Conclusions

We have shown that the participation rate in the self-employment program (SEMP) has increased between 1985 and 1997, whereas the participation rate in the subsidized employment program (SUBE) remains stable over the period. We argue that this evolution, which also has been observed in many other OECD-countries, is driven by a change, perhaps reflecting an international trend, in the Swedish labor market policy. It comes natural to pose the question whether SEMP is preferable to rival programs such as SUBE, and this is the question we address in this paper.

We claim that it is reasonable to contrast SEMP to SUBE since both programs imply

[26] Labor Market Board Ura 1997:12

[27] In another Labor Market Board report, Uuu 1995:1, the employers are asked whether they would have employed also in the absence of the program and 36 per cent said yes. These 36 per cent in turn, were asked if they would have employed the same person and here 56 per cent answered yes.

an initial spell, with duration of about six month, of subsidized employment, in the first case through self-employment grants. Furthermore, there seems to be no particular differences in the incentives as to participate in either one of the programs (except for those who do not receive UI benefits).

We determine the relative efficiency of the programs by comparing the post-program employment duration until re-unemployment for a large sample of participants in 1995-1996. The potential self-selection problem, inherent in all evaluation studies based on non-experimental data, is attacked by controlling for a large set of confounders and by using the office specific variation in participation rates as an instrument. Furthermore, simulations are carried out to check the sensitivity of the results under various assumptions of the selection process.

We find the risk of becoming unemployed was twice as high in the SUBE-program compared with the SEMP-program. We therefore conclude that the policy change that led to the increase in SEMP-participation was well justified.

The positive result for the SEMP-program is however only valid for workers of Swedish citizenship and therefore, the idea that immigrant workers would perform relatively better through SEMP-programs than other programs is not confirmed in this study.

Finally, the selectivity section in the study showed that there is a great deal of variation across offices concerning program participation. Hence an attempt to further increase the participation rate in the SEMP-program, at least in areas where the program is presently under-represented, might be worthwhile.

References

Altman, D.G., and de Stavola, B.L., (1994), Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates, *Statistics in Medicine*, 13, 301-341.

Andersson, F., (1999), Job flows in Swedish manufacturing, Working paper 1999:X, IFAU, Uppsala.

Andersson, J., Carling, K., and Mattson, S., (1998), Random Ranking of Hospitals is Unsound, *CHANCE*, 3, 34-39.

Bring, J and Carling, K (1998), Attrition and misclassification of drop-outs in the analysis of unemployment duration, Revised for *Journal of Official Statistics*.

Carlin, B.P., and Louis, T.A., (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, New York, Chapman and Hall.

Carling, K., (1995), Starting values for the iterative maximum likelihood estimator in survival analysis, *Journal of Applied Statistics*, 22, 515-519.

Carling, K., Edin, P-A., Harkman, A., and Holmlund, B., (1996), Unemployment duration, unemployment benefits, and labor market programs in Sweden, *Journal of Public Economics*, 59, 313-334.

Carling, K., and Jacobson, T., (1995), Modeling unemployment duration in a dependent competing risks framework: Identification and estimation, *Lifetime Data Analysis*, 1, 111-122.

Carling, K., and Söderberg, H., (1998), An experimental comparison of gradient methods in econometric duration analysis, *Computational Statistics & Data Analysis*, 27, 83-97.

Cleveland, W.S., (1979), Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, 74, 829-836.

Cleveland, W.S., Devlin, S.J., and Grosse, E., (1988), Regression by local fitting: Methods, properties, and computational algorithms, *Journal of Econometrics*, 37, 87-114.

Edin P-A, Holmlund B, (1994), *Arbetslösheten och arbetsmarknadens funktionssätt*, Ministry of Finance, Stockholm.

Heckman, J., LaLonde, R.J., and Smith, J.A., (1998), *The economics and econometrics* *of active labor market policy*, To appear in Handbook of Labor Economics, Volume III, Ashenfelter, O. and Card, D., ed.

Hoaglin, D.C., Mosteller, F., and Tukey, J.W., (1985*), Exploring data tables, trends, and shapes,* Wiley, New York.

Härdle, W., (1990), *Smoothing techniques: with implementation in S*, Springer-Verlag, New York.

Katz L, Stanley M, and Kreuger A, (1998), *Impacts of employment and training programmes: the American experience, In International Evidence on Employment Policies*, Treasury Working Paper 70, HM Treasury, UK

Kiefer, N.M., (1988), Economic duration data and hazard functions, *Journal of Economic Literature*, 26, 646-679.

Labor Market Board (1995), *Rekryteringsstödets resultat*, Uuu 1995:1, Stockholm.

Labor Market Board (1996), *Starta eget-bidraget – 1995 års uppföljning*, Pfuu 1996:1, Stockholm.

Labor Market Board (1997), *Åtgärdernas undanträngningseffekter. En enkätstudie utifrån deltagarperspektivet*, Ura 1997:12, Stockholm.

Labor Market Board (1998), *Åtgärdsundersökningen 1998*, Apra 1998:3.

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

Louis, T.A., and Dersimonian, R., (1982), *Health statistics based on discrete population groups In Regional Variation in Hospital Use*, D. Rothberg, ed, Boston, D.C. Heath and Co.

Meyer, B., (1990), Unemployment Insurance and Unemployment Spells, *Econometrica*, 58, 757-782.

Modigliani F, Fitoussi J-P, Moro B, Snower D, Solow R, Steinherr A, and Sylos Labini P, (1998), An Economists' Manifesto on Unemployment in the European Union, *Banca Nazionale del Lavoro Quarterly Review*, 206:327-361.

Narendranathan, W., and Stewart, M.B., (1993), Modelling the probability of leaving unemployment: competing risks models with flexible baseline hazard, *Applied Statistics*, 42, 63-83.

OECD, (1995), *Self-employment programmes for the unemployed*: papers and proceedings from a joint US department of labor OECD international conference, OECD working papers 78, Paris.

Petersen, H. C., (1993), *Business and Government*, 4:th ed., HarperCollins, New York.

Riksdagens Revisorer, (1997), *Starta-eget-bidraget* (Self employment grants), Rapport 1996/97:5, Stockholm.

Statistics Sweden, (1993), *A study of the unemployed according to data from the employment agencies and the labor force surveys* (in Swedish), SCB, Stockholm.

Statistics Sweden, (1998) *New started enterprises in Sweden 1996 and 1997*, Statistiska meddelanden, SCB, Örebro.

Wong G, Henson H, Riddell C, (1998), *Earnings Impact of Self-Employment Assistance for the Canadian Unemployed, 1987-1996,* draft, Strategic Evaluation and Monitoring, Human Resources Development Canada.

Vroman W, (1997), *Self Employment Assistance*, Report, The Urban Institute, Washington.

Appendix A: Simulation experiment

The model in (5a) can be written as a regression model (see Carling and Jacobson, 1995)

(A1) $$\ln \Lambda(t_e) = -a D - Xb - \ln V + e \ ,$$

where e follows the Extreme Value distribution with $E[e] = -0.5772$ and $Var[e] = p^2/6$. Thus, there are three sources that produce variation in the duration var i-able. The semi-parametric approach in model (5) prevent us from distinguish the vari a-tion arising from V. However, if we assume a functional form for the base-line hazard, then the three sources can be identified from data. We assume that $T_e | D, X, V$ follows the Weibull distribution. In doing so, we find that 15% of the variation is caused by D and X, 55% by e, and the remaining 30% by V.

We will consider the following Outcome model

(A2) $$\ln T_e = -a D - Xb - \ln V + e \ ,$$

and generate data for this model accordingly

(A3) $$\ln T_e = -a D - N_1(0,0.44) - 1.34 s N_2(0,0.50) + \ln(-\ln U[0,1]),$$

where $N(\ )$ refers to the normal distribution with arguments expectation and variance, and $U[0,1]$ refers to the uniform distribution over zero to one. Censoring is introduced by defining $T = \min(T_e, 1.22E)$ and $C = I(T_e < 1.22E)$, where E is a unity exponential variable. This censoring process gives about 45 % censored observations, which corre-sponds to the censoring degree in our data.

Two things can be noted for this specification, (i) the effect parameter is set to $a = 0,-1$, and (ii) $s = 1$ corresponds to the variance decomposition above, increasing/decreasing s will increase/decrease the association between T_e and the unobserved heterogeneity. We consider $s = 0.5,1$.

The likelihood of entering the SEMP-program will be determined as

$$\text{(A4)} \qquad D^* = Xq + (-\ln V + d), \quad \begin{cases} D = 1, if\ D^* > 0 \\ D = 0, if\ D^* \leq 0 \end{cases}.$$

Decomposing the variance sources, leaving the office specific variable out, in the model in (6), we found that 50% of the variance could be attributed to the observed factors and, thus, 50% of the variation remains unexplained. Generating D^* is done accordingly

$$\text{(A5)} \qquad D^* = -0.25 + N_3(0,0.5) + \left(r^2 N_2(0,0.5) + (1 - r^2) N_4(0,0.5)\right),$$

where r gives the correlation between the unobserved component in the participation decision and the one in the employment outcome equation, $r \in (0,1)$.

Finally, the bias is a function of the three parameters that will be varied in accordance with the statements above. It is defined as $Bias(a, c, r) = P\lim \hat{a} - a$, where $P\lim \hat{a}$ is the maximum likelihood estimate of the effect, obtained for $n = 500{,}000$.

The most interesting results are shown in figure 4. The dashed line shows the bias as a function of r in the case there is no actual effect, i.e. $a = 0$, and $s = 1$. The bias is, as expected increasing with r, though it is of modest size for $r \leq 0.5$. The solid line shows the results for a similar setting, the difference is that s is set to 0.5 and consequently the bias is smaller. The Weibull assumption above is probably producing a too high importance to V and thus $s = 0.5$ might be a more realistic value in our application. Finally, the dotted line shows the bias when there is an actual effect of $a = -1$. Note that for low values of r there is a tendency to underestimate the effect. This tendency is balanced by the selection bias for $r = 0.5$.