

DIFFERENCES IN CHARACTERISTICS BETWEEN UNEMPLOYED  
WITH DIFFERENT SPELLS OF UNEMPLOYMENT<sup>\*</sup>

Hendrik Folmer<sup>1</sup>

and

Jouke van Dijk<sup>2</sup>

Department of General Economics  
Agricultural University

Faculty of Economics  
University of Groningen

Abstract

The causes of differences in the duration of unemployment can be grouped into two categories: the personal characteristics of the unemployed and the structure of the regional economy. In the recent literature the personal characteristics age, education, family status, sex, work-experience, and unemployment benefits to income ratio are mentioned as the most important personal characteristics of the unemployed. The demand for labor is considered to be the relevant aspect of the regional economy. To find out whether the group of personal characteristics or the regional structure is the most important determinant of unemployment duration, the 1979 Labor Force Survey for the Netherlands is analyzed. The relevance of the variables for the Dutch situation is critically discussed and a theoretical model is constructed and empirically verified. Possible problems with regard to "data-mining" are taken into account by means of cross-validation. The results indicate that longer spells of unemployment depend mostly on personal characteristics and that the structure of the regional economy is of minor importance. Finally, some policy recommendations are formulated.

<sup>\*</sup> We gratefully acknowledge to Joost Veenstra for research assistance and to Jules Theeuwes and an anonymous referee for valuable remarks.

---

<sup>1</sup> P.O. Box 8130  
NL-6700 EW Wageningen  
Tel. 08370 - 84255

<sup>2</sup> P.O. Box 800  
NL-9700 AV GRONINGEN  
Tel. 050 - 633765

## 1. Introduction

The explosion of unemployment in most Western countries during about the past decade has left deep traces in the welfare of individuals and their households. Not all individuals, however, have been affected to the same degree. For instance, there are strong variations from one individual to another in income effects, in the number and the duration of spells of unemployment. It is obvious that the analysis of the causes of variation in unemployment is of great importance, inter alia, for the design of labor market and welfare policies.

The variation in the duration of spells of unemployment has been intensively studied. Theoretical and methodological aspects have been intensively discussed. Micro-economic approaches have been found to be most appropriate to study the underlying causes of variation (cf. Lancaster and Nickell, 1980).<sup>1)</sup> Various types of micro models have been developed by, among others, Ehrenberg and Oaxaca (1976), Salant (1977), MacKay and Reid (1972), Lancaster (1979) and Nickell (1979). Lancaster and Nickell (1980), Sider (1985), Heckman and Singer (1985), Narendranathan and Nickell (1985) among others, have dealt thoroughly with methodological aspects of and econometric methods for the estimation of duration models.<sup>2)</sup>

The present paper also deals with the variation in the duration of spells of unemployment. It is, however, primarily empirical in the sense that it tries to identify the main causes of differences in the duration of unemployment in the Netherlands on the basis of the 1979 Labor Force Survey. Two kinds of variables will be distinguished. First, personal characteristics of the unemployed such as age, education, family status, sex, work-experience and the replacement ratio (unemployment benefits to income ratio) and, secondly, variables reflecting the regional demand for labor. One of the aims of this paper is to investigate the relative importance of these two groups of variables. This is not only interesting from a scientific

---

<sup>1)</sup> A common alternative to the micro approach is the macro approach which explains shifts in aggregate unemployment or unemployment duration by changes in aggregate demand and supply side factors such as social security benefit levels (see, among others, Maki and Spindler (1975) and Cubbin and Foley (1977)). The macro approach has been criticized by Lancaster and Nickell (1980) because of the quite different results obtained as a consequence of the weakness of the data and the lack of a sound theoretical basis.

<sup>2)</sup> Recent information about these issues can also be found in a special issue about duration analysis of the Journal of Econometrics vol. 28, Annals 1985-1.



point of view, but has also some important policy implications. If the spell length is primarily determined by personal characteristics, schooling and (re-)training might be appropriate policy measures. If the spell length is mainly influenced by the regional demand for labor, however, migration and demand stimulating policies could be more adequate. (In the sequel schooling, (re-)training, migration and demand policies will simultaneously be denoted as "labor market policy".)

The organization of the remainder of this paper is as follows. In section 2 some micro approaches to the variation in the duration of spells of unemployment are briefly described. In particular, the micro approach applied in the present paper is justified. In section 3 the data base is presented and the conceptual model for the problem at hand (i.e. the a priori expected differences in personal and regional labor market characteristics which lead to variation in unemployment duration) is outlined. The econometric method is briefly described in section 4. The main results of the Dutch case study are described in section 5. The paper ends with a concluding section where some policy recommendations are formulated.

## 2. Main approaches

The usual starting point in modelling the duration of the spell of unemployment is search theory (see, Lippman and McCall, 1976, Joll et al 1983). This theory assumes that unemployed, who are looking for jobs, have sketchy knowledge of suitable vacancies and their associated wages. Therefore, they sample the available job opportunities. The recent literature is dominated by sequential models where in each period an unemployed will be offered a job at some wage with probability  $p$ . The unemployed will accept the job if the wage offer associated with the job offer is larger than his predetermined reservation wage. Otherwise he will go on searching. The reservation wage is primarily determined by the distribution of wage offers, the level of unemployment compensation, the probability of receiving a wage offer in any period and the duration of the spell of unemployment. The probability of leaving unemployment in any period is equal to the product of the probability of receiving a job offer and the probability that this offer is larger than the reservation wage.

Before turning to the various approaches to model unemployment duration, it is important to note that the data provided by the Labor Force Surveys for the Netherlands (and for many other countries as well) consist

of cross-section observations of interrupted spells, i.e. spells in progress at the date of the survey, which are partly through their full length. The interrupted spell has to be distinguished from the completed spell which is usually considered to be superior to the former as an index of economic welfare. (cf. Löfgren, 1976 and Salant, 1977).<sup>3)</sup> In the present study, however, the variation in the duration of completed spells of unemployment will not be studied because the purpose of this study is to identify the personal and regional characteristics of unemployed whose spell of unemployment is likely to exceed a given critical length. This critical length is an important index for the the design of labor market policy. For instance, individuals who have been unemployed for a given period are target groups for (re-)schooling and training programs and regions with high proportions of long-term unemployed due to demand deficiency are favored by regional demand and/or out-migration policies. The upshot of this paragraph is that in the context of labor market policies the interrupted spell of unemployment is more appropriate as an index of policy intervention than the completed spell.

Two types of approaches to model unemployment duration can be distinguished within the framework of search theory. The first consists of analyzing the duration of unemployment directly (see, among others, MacKay and Reid, 1972, Ehrenberg and Oaxaca, 1976). The second focuses on the conditional probability of an individual leaving unemployment in a particular period, given the individual's unemployment duration. Next, these probabilities are used to estimate the expected unemployment duration. The following types of variables are used in these models (see, among others, Burman, 1980):

- a. Characteristics of the individual which do not vary in the short-term and which are not directly influenced by being (un-)employed, for instance, age, health, education, marital status, number of dependents.
- b. Characteristics of the individual which may change with the duration of his unemployment, i.e. unemployment benefits, other family income and the "replacement ratio" (the ratio of the unemployment benefits and the expected labor income).
- c. Regional demand for labor, i.e. the regional unemployment rate or the regional ratio of vacancies to unemployment.
- d. Duration of unemployment.

---

<sup>3)</sup> The density of the completed spells can be derived from the density of the interrupted spells (cf. Löfgren, 1976, Salant, 1977, and Folmer and Van Dijk, 1986).



The inclusion of the duration variable is based on the assumption that the probability of an individual to leave unemployment decreases with the time he or she has been unemployed. Against this train of thought the objection could be raised that, ultimately, the time an individual has already been unemployed is a function of the personal and regional characteristics a. - c. So, the duration variable could be replaced by genuine labor market characteristics. Moreover a spell length exceeding about one year is often viewed as an additional indicator of negatively valued (primarily personal) characteristics. Even in that case, however, the duration variable could be replaced by genuine personal and regional characteristics, though the coefficients might differ from those which apply to short-term unemployed. For these reasons it may be preferable to substitute the duration variable by the genuine explanatory variables a. - c. Using genuine variables makes it also possible to identify potential groups with high risks of relatively long spells of unemployment, because the results are not blurred by the duration variable which implicitly represents genuine labor market characteristics. It is obvious that, for the design of labor market policies, identification of potential risks groups is much more important than an accurate prediction of the expected unemployment duration at which the models including the duration variable are aimed.

An additional advantage of models without the duration variable as an explanatory variable is that a serious econometric problem can be circumvented. As shown by Nickell (1979) and Lancaster and Nickell (1980) the duration variable in combination with unobservable variables leads to specification errors which are more serious than in the case of linear regression with omitted variables which are correlated with the included regressors. This implies that the estimated effect of duration will be biased. Lancaster and Nickell (1980) also showed that in the case of cross-section observations it is not quite well possible to discriminate between genuine time variation and unobserved sample heterogeneity because of identification problems. In the kind of approach advocated in this paper both problems mentioned here can be circumvented because the duration of unemployment is not an explanatory variable.

In the next section the types of variables with regard to which the groups with different spells of unemployment are expected to differ will be discussed.

### 3. Data base and conceptual model

The data which will be analyzed in this paper is micro data from the Labor Force Survey 1979 for the Netherlands<sup>4)</sup>. This 3% sample contains 142,300 cases. Only persons born before 1965 were interviewed. Every interviewee was asked whether he was employed or unemployed at the date of the interview. Everyone who was looking for a job, whether a full-time (40-hours) job or a job for only a few hours a week, was classified as being unemployed. All these unemployed persons were asked how long they had been looking for a job. Besides information on employment status and unemployment duration, information was also gathered about the personal characteristics age, family status, education, work-experience and region of living. (For detailed information see CBS, 1982). In Table 1 we present the distribution of the unemployed over three duration categories, which are relevant for the present analysis.

Table 1. Distribution of unemployed according to the duration of the spell of unemployment in the 1979 Labor Force Survey for the Netherlands.

Duration categories	Absolute	Percentages
< 4 months	1555	25.5
4-11 months	1361	22.4
<u>≥ 12 months</u>	<u>3175</u>	<u>52.1</u>
Total	6090	100.0

On the basis of Table 1 two groups of unemployed can clearly be distinguished. The first group consists of short-term unemployed (< 4 months). We expect that this category consists mainly of individuals who have a high probability of getting (re-)employed (i.e. school-leavers, job-hoppers, frictional unemployed, etc.), and to a lesser extent of potentially long-term unemployed. The second group consists of long-term unemployed (≥ 12 months) who have slighter probabilities of getting (re-)employed. In addition to these two groups a third group will be distinguished,

<sup>4)</sup> The Dutch Labor Force Survey is part of the sample survey which the Census Bureau of the European Community organizes every two year in the member countries. Detailed information about the 1979 survey can be found in CBS (1982).



i.e. the medium-term unemployed (4-11 months). It consists of potentially long-term unemployed and of individuals whose (re-)employment probabilities are in-between those of the short and long-term unemployed for various reasons.

In the literature on job search, the duration of unemployment and the (re-)entry of unemployed into employment, the following labor market characteristics have been identified as important discriminating variables between unemployed with different spells of unemployment. (Detailed information about the various variables in the data set can be found in Appendix A.)

## PERSONAL CHARACTERISTICS

### Age

The probability of getting a job is found to decrease with age, because the younger age groups with the most up to date formal schooling are more flexible and have a relatively long payback period for the employer on investments in on-the-job training. The Ministry of Social Affairs and Employment (see SOZAWÉ, 1982, 1983) finds that young people are over-represented in both the in- and outflow into and out of unemployment, which indicates high job turn-over rates and relatively high probabilities for short spells of unemployment. On the basis of these considerations an over-representation of the elder age groups among the unemployed with relatively long spells of unemployment is expected.

### Education

This variable contains information on the level and type of education. It is assumed that individuals with high and occupation-specific education are dominant in the group of short-term unemployed because they have higher productivity rates and are better equipped for job search.

### Family status

With respect to this variable two counteracting forces have to be taken into account. On the one hand, employers prefer family heads and, to a lesser extent, spouses and other family members to singles, because they value family membership as an indicator for reliability and stability (cf. Nickell, 1979). Therefore, the former are likely to have shorter spells of unemployment than singles. On the other hand, singles, spouses and other family members apply more often than family heads for jobs in the secondary labor market. This segment is characterized by frequently occurring short spells of (un-)employment and high (re-)entry probabilities. Because of the

contradictory tendencies differences in unemployment duration between the four categories of family status are not a priori clear.

It should be noted that the combination of the variables sex and marital status, for which data is available, is quite similar to family status. Because of their close correspondence, the use of both family status and sex-marital status should be avoided. In the empirical part, however, the combination will be tried as an alternative for family status.

#### Work-experience

This variable distinguishes between unemployed, who look for jobs for the first time, and unemployed who lost their jobs. The latter, who have work-experience, might be preferred by employers because their human capital may lead to higher productivity. However, they may also have higher reservation wages, which makes them less attractive to employers. The outcome of these two opposing tendencies is not a priori clear.

#### Replacement ratio

The replacement ratio, defined as the unemployment benefits to potential income ratio, is generally considered to be an important explanatory variable in (re-)entry and duration studies. The reason for this is that it is assumed that individuals, who have a high replacement ratio, have the opportunity to prolong their search periods, which may result in longer spells of unemployment than are strictly necessary.

It is important to remark that for the Dutch situation no significant effects have been found (Van Opstal and Theeuwes, 1986; Bron et al, 1983). A possible explanation may be in the organization of the Dutch social security system. In the Dutch situation, for a large number of unemployed the unemployment benefits are about 85 to 95% of the net income previously earned. Moreover, the percentage is nearly constant during the first two and a half years for most unemployed.<sup>5)</sup> So, differences in individual spells of unem-

-----  
<sup>5)</sup> There are some exceptions to the general regulations with regard to unemployment benefits. We will briefly discuss these exceptions and indicate the ways in which they can be captured. First, spouses who have been unemployed for over six months are not entitled to unemployment benefits. The same applies to spouses, who enter the labor market for the first time and do not find a job. Secondly, for other categories of unemployed than spouses, who have not been previously employed, the unemployment benefits are equal to the minimum subsistence benefits. The same applies to those who have been employed for less than 6 months before they became unemployed and for those who have been unemployed for more than two and a half years. Thirdly, there exists a maximum for the unemployment benefits so that for the higher income groups the benefits can be substantially lower than 85% of the income (continue on next page)



ployment may hardly be explained by the "variation" in the replacement ratio, though there might be an overall increase in spell length.

In addition to the peculiarities of the Dutch situation there are some explanations for the slight effect of the replacement ratio on the increase in the duration of unemployment which apply in general (see also de Neubourg, 1985). First, prolongation of the spell may lead employers to the impression that the unemployed concerned have less favorable labor market characteristics, which is likely to cause a substantial decrease in their re-entry probabilities (see also section 2). Secondly, being unemployed and living from unemployment benefits usually implies a decrease in social status. This will stimulate unemployed to keep the spell of unemployment as short as possible. Furthermore, "paid work" itself has a positive utility because it is highly-valued in most Western societies. So, awareness of the deterioration of the re-entry probability and social norms contradict simple economic theory, which predicts that unemployed will choose to stay or become unemployed when income maintenance is (partly) guaranteed by social security benefits. The upshot of the considerations outlined above is that unemployed will usually try to avoid the disutility of the loss of reputation and accept the first job offer they get. This hypothesis was empirically confirmed by Clark and Summers (1979) who found that 90% of the unemployed did accept the first job offer.

From this it follows that in general the replacement ratio is not an unambiguous labor market characteristic to discriminate between unemployed with different spells of unemployment. Moreover, in the Dutch situation it is an inappropriate discriminating variable because it is more or less constant for a substantial proportion of unemployed during a period of two and a half years.

-----  
<sup>5</sup> (continued)

previously earned. Finally, those who quit voluntarily are not entitled. It is obvious that all these categories may be inclined to keep the spell of unemployment as short as possible and thus are likely to be overrepresented among the short-term unemployed. The various categories distinguished above are taken into account in the following ways. Spouses and those who enter the labor market for the first time are captured by means of the variables family status and work-experience. With regard to the other categories, i.e. unemployed with a potential income in the right tail of the income distribution, those who became unemployed after having been employed for less than 6 months, those who had been unemployed for more than two and a half years and those who quit voluntarily, no information is available in the data set. These categories, however, are very small and may be neglected.

## THE REGIONAL DEMAND FOR LABOR

There exists a general consensus that regional variation in the level of demand leads to differentials among regions in unemployment duration (see, among others, Metcalf, 1975, and Green, 1985). The probability of getting a job is positively affected by the number of available jobs and negatively by the competition for jobs. It should be observed that information on the availability of and competition for jobs by means of data about the number of vacancies and the number of unemployed in a region is seriously hampered by deficiencies in the data (see, among others, Van Dijk, 1986). For instance, the number of officially registered vacancies usually substantially under-estimates the total number of available vacancies. This follows from the fact that firms are not obliged to register vacancies. Moreover, vacancies can be filled internally or via direct contacts between applicants and employers without the intervention of the Labor Exchange. The competition for jobs is usually operationalized by means of variables like the regional unemployment rate or the unemployment/vacancy ratio. It should be observed, however, that these indicators only partly reflect the genuine competition for jobs at the regional labor market, because both employed and unemployed in the region where the jobs are available as well as employed and unemployed from other regions may compete. Furthermore, the unemployment figures are affected by hidden unemployment, registration errors, etc.

From this it follows that the operationalization of the labor market conditions of the regional economy by means of the number of vacancies and the competition at the labor market requires detailed information. Unfortunately, this information is not available for the present case study. Therefore, a proxy, viz. the province of living will be used. Moreover, this dummy variable makes it possible to take into account omitted regional variables, such as locational and infrastructural differences and the effects of regional policy, which has been applied in the peripheral provinces Groningen, Friesland, Drenthe and Limburg. On the basis of the above mentioned considerations, we expect the long-term unemployed to be concentrated in the peripheral provinces of Groningen, Friesland, Drenthe and Limburg, especially because of the high unemployment rates in these provinces.

This section ends with the remark that the regional supply of labor is usually also considered to influence the duration of unemployment. The latter comprises demographic and related factors as education and work-



experience. It is well known that employed with similar labor market characteristics tend to cluster in specific regions (see, among others, Oberg and Oscarsson, 1979). For instance, individuals with higher levels of education are usually overrepresented in urban areas. The possible regional clustering of individuals with similar personal labor market characteristics will in the present analysis be taken into account by means of interactions between the relevant personal characteristics and the regional variables (see below).

### INTERACTIONS

The following first-order interactions will at first instance be included in the model:

- provincial dummies \* each personal characteristic;
- age \* work-experience;
- family status \* age;
- family status \* education;
- family status \* work-experience;
- work experience \* education

Several other first-order interactions will be tried out as well, although there is little a priori evidence for the tenability of other interactions than the ones mentioned. When significant first-order interactions have been found higher-order interactions will be considered.

This section ends with the following remarks. The most outspoken differences with respect to the variables described above are expected between short and long-term unemployed. The differences between medium-term unemployed on the one hand and short-term or long-term unemployed at the other hand are likely to be less clear-cut. It may happen that the medium-term unemployed show characteristics which come close to either one of the categories. This may have consequences for the delimitation of the categories. For instance, lack of substantial differences between medium and long-term unemployed can be interpreted that a dichotomy between short-term frictional unemployed on the one hand and long-term structural unemployed at the other is appropriate.

#### 4. Econometric aspects

The problem under study of identifying differences with respect to the various personal and regional variables between the three groups of unemployed can be formulated as the discrimination among three groups on the basis of  $(p-1)$  variables  $x_{ij}$  ( $j = 2, \dots, p$ ) with  $x_{i1}$  an indicator for the constant terms, for  $n$  observations ( $i = 1, \dots, n$ ). Press and Wilson (1978) and Maddala (1983), among others, have shown that this problem can be analyzed by means of logistic regression. In terms of the logistic regression model the problem can formally be represented as follows. Let  $z_{gi}$  be an indicator variable which takes the value 1 if the  $i$ -th case belongs to the  $g$ -th group and 0 otherwise. Moreover, let

$$\phi_{gi} = \Pr (z_{gi} = 1 \mid x_i), \quad g = 0, 1, 2. \quad (1)$$

where  $x_i^T = (x_{i1}, \dots, x_{ip})$

Then

$$\log (\phi_{gi}/\phi_{0i}) = x_i^T \beta_g, \quad g = 1, 2. \quad (2)$$

where  $\beta_g^T = (\beta_{g1}, \dots, \beta_{gp})$  is a vector of  $p$  (unknown parameters.)

Instead of the polychotomous approach (2) the method of pairwise comparisons will be applied here. In this case model (2) is of the form

$$\log (\theta_{gi}/\theta_{0i}) = x_i^T \alpha_g \quad (3)$$

where  $\alpha_g$  is a vector of  $p$  unknown parameters and

$$\theta_{gi} = \Pr (z_{gi} = 1 \mid x_i, z_{0i} + z_{gi} = 1), \quad \theta_{0i} = \Pr (z_{gi} = 0 \mid x_i, z_{0i} + z_{gi} = 1)$$

Begg and Gray (1984) show that  $\beta_g = \alpha_g$  ( $g = 1, \dots, G$ ) and that if maximum likelihood is employed the parameter estimator of the pairwise comparison approach is asymptotically unbiased. Moreover, they show that its asymptotic relative efficiency is generally high. The same applies, though to a lesser extent, to joint tests of parameters from different comparisons. Finally, Wijesinha et al. (1983) has argued that the present method is computationally less cumbersome than the polychotomous approach; that it



facilitates variable selection and that it is more readily available in standard computer packages. For these reasons the method of pairwise comparisons will be applied in this paper. It, will be used in such a way, however, that three groups are mutually compared.

Another econometric aspect that deserves attention here is the explanatory nature of the analysis. As shown above, there is uncertainty about the interactions to be included into the model. In order to find the relevant interactions, various alternative specifications have been tried. Therefore, the ultimate model will have a data-instigated nature. Consequently, the goodness of model fit to the sample data is likely to be greater than to the population (see, among others, Lovell, 1983).

In this paper the consequences of the data-instigated nature of the model will be taken into account by means of cross-validation. That is, the sample at hand is randomly divided into two subsets. One subset, the training set, is used for model selection and fitting purposes such as the examination of outliers, looking for patterns, transforming the data, pre-testing, etc. Next, the model chosen on the basis of the training set is estimated using the second data set, which has not been exercised upon at the previous stage. This set is called the validation set. If the two sets of estimates based on the training and validation sets do not differ substantially one may have confidence that the selected model is close to the "true" model. In this regard we will adopt the following criteria of correspondence between the two sets of estimates. First, both estimates of the confidence interval of a given parameter (i.e. both the one based on the training set and the one based on the validation set) are to indicate that the parameter concerned is significantly different from zero. Moreover, the distance between the two parameter estimates is to be less than twice the minimum of the two standard deviations obtained in the subsamples. Secondly, the goodness of fit statistics based upon each subsample are not to exceed a priori given values. In particular, if an estimate based upon the training set indicates that a given set of parameters is significantly different from zero, whereas the estimate based upon the validation set indicates that it is not, the variable concerned will be deleted from both sets if the deletion does not lead to a substantial deterioration of the goodness of fit statistics in both sets. Finally, a model which performs about equally well

in both the training and validation sets will ultimately be estimated on the total sample. Model (3) has been estimated using computerpackage GLIM.<sup>6)</sup>

## 5. Empirical results

The following estimation procedure was used. First, the basic model made up by the variables family status, education, age, work-experience, and the provincial dummies was estimated on the training set consisting of about 50% of the observations. Next the basic model was extended with the various interactions and was also exercised on the training set. An interaction was included in the model if this resulted in a drop in the scaled deviance<sup>7)</sup> at least as large as the corresponding loss of degrees of freedom.

In the comparison of the most extreme groups, i.e. the short and long term unemployed, the best model in terms of significant parameter estimates turned out to be the basic model extended with the interaction family status \* work-experience<sup>8)</sup>. The estimates on the basis of the validation set, however, deviated from those of the training set with respect to the significance of work-experience, the interaction term, and the provincial dummies. Results obtained with regard to the alternative variables marital status and sex were not satisfactory either. The end result was a reduced model consisting of the variables age, education and family status only. The reduced model was estimated on both the training and the validation set. The main results are given in Appendix B. The goodness of fit statistics for both sets (i.e. Pearson  $\chi^2$  and scaled deviance) are rather similar. Moreover, for each variable at least one coefficient (and the same in both the validation and the training set) is significantly different from zero at the 5% level. Finally, there are no coefficients for which the estimates differ more than twice the minimum standard deviation. Hence, this is on our criteria a satisfactory specification.

-----  
<sup>6)</sup> For detailed methodological background information see McCullagh and Nelder (1983); for information about the computerpackage see Baker and Nelder (1978).

<sup>7)</sup> The scaled deviance is defined as  $-2 \log(h/s)$  where  $h$  and  $s$  are the likelihood functions of the hypothesized and saturated models, respectively. The saturated model contains all possible linearly independent parameters, whereas in the hypothesized model a set of parameters is restricted. The scaled deviance is under quite general regularity conditions asymptotically distributed as a  $\chi^2$  variable. For detailed information on the present selection procedure see, among others, Baker and Nelder (1978).

<sup>8)</sup> Results available upon request.



For the other groups to be compared a similar validation procedure has been applied. In both comparisons the same reduced model as in the comparison of long and short term unemployed was found to be appropriate. In both situations the models performed about equally well in both the training and the validation set.

From table 2 it follows that the overall fit, as measured by the Pearson  $\chi^2$ , is acceptable in all cases. The scaled deviance, however, indicates a less satisfactory fit. It should be noted, however, that this statistic is not very reliable when there is deviation from normality (cf. Baker and Nelder, 1978), which is the case here. On the basis of the Pearson  $\chi^2$  and because no substantial improvement of the scaled deviance could be obtained by adding other variables to the model, we may conclude that the overall fit is satisfactory.

Before discussing the estimation results presented in Table 2, we repeat that no significant changes in scaled deviance for the provincial dummies were obtained. As a further check, a specification was tested in which the regional structure was dichotomized. One group contains the problem regions Groningen, Friesland, Drenthe and Limburg, and the other consists of the remaining provinces. When this dummy variable is included in the ultimately chosen model with the variables age, education and family status no significant coefficients for the regional dummy and an increase in the scaled deviance (instead of a decrease!) were found for both the training and the validation set. Neither were significant interactions with the personal characteristics obtained. This implies that the structure of the regional labor market has no substantial influence on the differences in the duration of interrupted spells of unemployment. This is quite remarkable because it is generally believed that the economic structures in the peripheral provinces Groningen, Friesland, Drenthe and Limburg, which have unemployment rates far above the national average, lead to substantial prolongation of the duration of unemployment. The results of the present analysis contradict this. A possible explanation for the insignificance of the regional dummy, which implies in particular that long-term unemployed are not overrepresented in the peripheral provinces, might be that unemployed whose spells of unemployment exceed a critical limit leave unemployment via employment programs, leave the regional labor force (discouraged worker effect) or are classified as unabled instead of as unemployed. (It should be observed that all these forms are generally denoted as hidden unemployment). So, due to labor market policies and institutional regulations with regard

Table 2. Maximum likelihood binary logit estimates of differences between unemployed with different unemployment duration.<sup>1)</sup>

Variables:	Duration of unemployment categories:					
	short vs long		short vs medium		medium vs long	
	(1)	(0)	(1)	(0)	(1)	(0)
	est.	s.e.	est.	s.e.	est.	s.e.
GRAND MEAN	-.467	.163	-.113	.186*	-.350	.160
AGE:						
14-19	..		..		..	
20-24	-.532	.113	-.234	.128*	-.366	.121
25-39	-.950	.125	-.320	.142	-.673	.128
40-54	-1.671	.153	-.498	.178	-1.149	.148
55-59	-2.003	.242	-.489	.292*	-1.508	.217
60 or more	-2.685	.257	-.580	.351*	-2.267	.250
EDUCATION:						
low	..		..		..	
lower medium;general	.453	.137	.220	.152*	.202	.132*
" " ;occu.spec.	.249	.104	.236	.119*	.015	.098*
upper medium;general	.957	.170	.580	.185	.430	.177
" " ;occu.spec.	.492	.112	.099	.123*	.386	.103
high; occupat. specific	.409	.166	.267	.174*	.705	.140
high; scientific	.772	.240	.092	.315*	-.155	.286*
unknown and undefined	1.110	.108	1.540	.133	-.435	.130
FAMILY STATUS:						
singles	..		..		..	
family heads	.048	.129*	.093	.155*	-.045	.122*
spouses	.530	.126	.379	.150	.212	.122*
other family members	.044	.136*	-.290	.159*	.106	.132*
GOODNESS OF FIT STATISTICS: <sup>2)</sup>						
	train.	valid.	train.	valid.	train.	valid.
Pearson $\chi^2$	2346	2368	1456	1456	2275	2278
Probability level	48%	39%	40%	40%	36%	34%
Scaled deviance	2582	2631	1842	1882	2632	2548
Probability level	0%	0%	0%	0%	0%	0%
Degrees of freedom	2349	2349	1442	1441	2252	2251
Total number of unempl.	2365	2365	1458	1457	2268	2267

\* ) indicates an estimate which does not differ significantly from the grand mean (5% significance level).

<sup>1)</sup> The estimates are differences on the log-odds scale with regard to the grand mean which represents an individual with low education, in the age group 14-19, who is single. A positive coefficient indicates overrepresentation of the category coded (1); a negative coefficient overrepresentation of the category coded (0).

<sup>2)</sup> The statistics for the overall fit refer tot the training and validation sets. The main reason for this is that the total sample sizes in all three cases are very large. As pointed out by Baker and Nelder (1978), for very large samples the statistics are likely to be much larger than expected. Therefore, the statistics for the overall fit obtained in the training and cross-validation sets, which are of moderate sizes, may give a better insight into the fit of the model.



to the registration of unemployed spatial disparities in the duration of unemployment might be removed or hidden.

Let us now turn to the results presented in Table 2. The estimates associated with the variable age confirm the theoretical expectations formulated above in the sense that substantial differences exist between short and long-term unemployed. The older age groups are strongly overrepresented in the latter group. Similar results hold for the medium- and long-term unemployed. Between medium- and short-term unemployed less outspoken differences are found. In particular, the age group 55 and more is equally spread.

The results for the variable education are less unambiguous. The comparison of short- and long-term unemployed shows that those with higher education are overrepresented among the former, which is in agreement with the theoretical hypotheses. Between short- and medium-term unemployed differences occur only with regard to the categories upper medium general and unknown and undefined education. Moreover, between medium- and long-term unemployed no differences can be detected in the categories lower medium and high scientific education. From these results it follows that with regard to education the category of medium-term unemployed is a mixture of short- and long-term unemployed, although it resembles the former group somewhat more than the latter.

The estimates with regard to family status show that differences only occur with regard to the category spouses in the comparisons of the short-term unemployed with both the medium- and long-term unemployed. In both cases the spouses are overrepresented among the short-term unemployed. Two possible reasons for this result can be given. First, as argued above, spouses frequently opt for jobs in the secondary labor market which is characterized by short spells of unemployment. Secondly, a substantial proportion of spouses, who have been unemployed for about 4 months, become discouraged and stop looking for a job. As no concentration of singles and other family members, mainly children still living with their parents, (who also frequently opt for jobs in the secondary labor market) in the category of short-term unemployment occurs, the second reason is more likely.

In general the conclusion can be drawn that between the three groups of unemployed there are substantial differences with regard to personal labor market characteristics. Generally speaking, the possession of labor market characteristics, which are negatively valued by employers, goes together

with an increase in the duration of the spell of unemployment. The differences between long- and short-term unemployed are unambiguous and clear-cut. The group of medium-term unemployed has an intermediate position between these extremes for a large number of variables describing personal characteristics. However, it is somewhat closer to the latter than to the former.

## 6. Conclusions and policy implications

In the preceding sections it was shown that differences in the duration of unemployment are mainly caused by personal characteristics and that the structure of the regional economy, viz. the regional demand for labor, is of minor importance. More specifically, if the personal characteristics are taken into account, short-, long- and medium-term unemployed are spatially proportionally distributed. Moreover, the results show that age and education are the most important discriminating labor market characteristics in this case. This is in agreement with the theoretical considerations that these variables are important selection criteria for employers. Although differences have also been found with regard to spouses it is doubtful whether this is a demand determined result. It is more likely that spouses, who have been unemployed for more than 4 months, stop looking for a job. No significant differences could be discerned for previous work-experience. This could imply that the benefits of work-experience are likely to be offset by the disadvantages of a higher reservation wage.

The policy implications of the results obtained in this study are that a labor market policy of (re-)schooling and (re-)training for the less well educated and less experienced below say the age of 40 might be appropriate. It is obvious that this policy is not efficient in the case of older unemployed because of the short pay back period. Early retirement and extra employment programs for these age groups might be more appropriate. These policies may possibly reduce spatial disparities in the variation of the duration of unemployment. A policy of improvement of the matching process is probably only helpful for the short term frictional unemployed. Finally, we conclude that on the basis of our results additional spatially differentiated policy measures aimed at equalizing the provincial differences in the ratios between long, medium and short-term unemployed are not necessary.



## REFERENCES:

- Baker, R.J. and Nelder, J.A., (1978). The GLIM System. Release 3, Generalized Linear Interactive Modelling. Numerical Algorithms Group, Oxford.
- Begg, C.B., and Gray, R., (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. Biometrika 71, 1; 11-18.
- Bron, J.A.H., Gaans, F. van, and Nieuwenhuizen, R.D.J., (1983). Integratie van werkloosheidsregelingen. 's-Gravenhage: Instituut voor onderzoek van overheidsuitgaven.
- Burman, J.P., (1980). Comment on The analysis of re-employment probabilities for the unemployed by T. Lancaster and S. Nickell. Journal of the Royal Statistical Society A 143, 2; 141-165.
- C.B.S., (1982). Arbeidskrachtentelling 1979. The Hague: Staatsuitgeverij.
- Clark, K.B., and Summer, L.H., (1979). Labor market dynamics and unemployment: a reconsideration. Brookings Papers on Economic Activity nr. 1; 13-72.
- Cubbin, J.S., and Foley, K. (1977) The extent of benefit-produced unemployment in Great Britain: some new evidence. Oxford Economic Papers 29; 128-140.
- Ehrenberg, R.G., and Oaxaca, R.L., (1976). Unemployment insurance, duration of unemployment and subsequent wage gain. The American Economic Review 66, 5; 754-766.
- Folmer, H., and Dijk, J. van, (1986). Regional differences in the duration of interrupted and completed spells of unemployment. Paper presented to the European Congress of the Regional Science Association August, 1986, Krakow (Poland).
- Green, A.E., (1985). Unemployment duration in the recession: the local labour market area scale. Regional Studies 19, 2; 111-129.
- Heckman, J.J., and Singer, B., (1985). Social science duration analysis. In: J.J. Heckman and B. Singer (Eds.) Longitudinal Analysis of Labor Market Data. Cambridge: Cambridge University Press.
- Joll, C., McKenna, Chr., and McNabb, R., (1983). Developments in Labour Market Analysis. London: Georg Allen & Unwin.
- Lancaster, T., (1979). Econometric methods for the duration of unemployment. Econometrica 47, 4; 939-956.
- Lancaster, T., and Nickell, S., (1980). The analysis of re-employment probabilities for the unemployed. Journal of the Royal Statistical Society A 143, 2; 141-165.
- Lippman, S., and McCall, J., (1976). The economics of job search: a survey (Part I and II). Economic Inquiry 14, 2; 155-189 and 3; 347-368.
- Löfgren, K.G., (1976). An indirect approach to determine the average duration of unemployment. Papers of the Department of Economics, University of Umeå, Umeå, Sweden.

- Lovell, M.C., (1983). Data mining. Review of Economics and Statistics 65, 1; 1-12.
- MacKay, D.J., and Reid, G.L., (1972). Redundancy, unemployment and manpower policy. Economic Journal 82; 1256-1272.
- Maddala, G.S., (1983). Limited-dependent and Qualitative Variables in Econometrics. Econometric Society Monograph in Quantitative Economics. Cambridge: Cambridge University Press.
- Maki, D., and Spindler, Z.A. (1975) The effect of unemployment compensation on the rate of unemployment in Great Britain. Oxford Economic Papers 27; 440-454.
- McCullagh, P., and Nelder, J.A., (1983). Generalized Linear Models. London: Chapman and Hall.
- Metcalfe, D., (1975). Urban unemployment in England. Economic Journal 85, September; 578-589.
- Narendranathan, W., and Nickell, S., (1985). Modelling the process of job search. Journal of Econometrics 28, Annals-1; 29-49.
- Neubourg, Chr. de, (1985). The world according to Jantje van Leiden or the culture of unemployment: an essay in sociological economies. Maastricht: Limburg University. Working paper 85-003.
- Nickell, S., (1979). Estimating the probability of leaving unemployment. Econometrica 47, 5; 1249-1264.
- Öberg, S., and Oscarsson, G., (1979). Regional policy and interregional migration - matching jobs and individuals on local labour markets. Regional Studies 2, 2; 139-163.
- Press, S.J., and Wilson, S., (1978). Choosing between logistic regression and discriminant analysis. Journal of the American Statistical Association 73, 4; 699-705.
- S.O.Z.A.W.E., (1982). Rapportage Arbeidsmarkt 1982. Den Haag: S.O.Z.A.W.E.
- S.O.Z.A.W.E., (1983). Rapportage Arbeidsmarkt 1983. Den Haag: S.O.Z.A.W.E.
- Salant, S.W., (1977). Search theory and duration data: a theory of sorts. Quarterly Journal of Economics 91; 44-45.
- Sider, H., (1985). Unemployment duration and incidence: 1968-82. American Economic Review 75, 3; 461-472.
- Van Dijk, J. (1986). Migratie en Arbeidsmarkt. Assen: Van Gorcum.
- Van Opstal, R., and Theeuwes, J., (1986). Duration of unemployment in the Dutch youth labour market. De Economist 134, 3; 351-367.
- Wijesinha, A., Begg, C.B., and Funkenstein, H.H. (1983). Methodology for the differential diagnosis of a complex data-set: a case study using data from routine CT-scan examinations. Medical Decision Making nr. 3, 133-154.



## Appendix A. Data and variables.

AGE: (at 1979 Jan. 1)	UNEMPLOYMENT DURATION:							
	short		medium		long		total	
14-19	560	36%	279	21%	400	13%	1239	20%
20-24	349	22%	299	23%	496	16%	1144	19%
25-39	492	31%	511	39%	1123	35%	2127	35%
40-54	128	8%	178	13%	678	21%	985	16%
55-59	24	2%	36	3%	214	7%	274	5%
60 or more	18	1%	24	2%	280	9%	322	5%
TOTAL	1571	100%	1327	100%	3192	100%	6090	100%
EDUCATION:								
low	322	21%	425	32%	1262	40%	2009	33%
lower medium; general	115	7%	113	9%	227	7%	456	8%
" " ; occu. spec.	239	15%	234	18%	598	19%	1071	18%
upper medium; general.	86	6%	65	5%	96	3%	247	4%
" " ; occu. spec.	198	13%	226	17%	461	14%	885	15%
high; occupation specific	67	4%	114	9%	154	5%	334	6%
high; scientific	32	2%	18	1%	59	2%	108	2%
unknown and undefined	513	33%	132	10%	335	11%	980	16%
A low level of education means less than 7 years of formal schooling, lower medium 7 - 9 years, upper medium 10 - 12 years, and high more than 12 years of formal schooling; those with an undefined education are merely still at school, but already looking for a job.								
WORK-EXPERIENCE:								
no	1091	69%	682	51%	1879	59%	3652	60%
yes	480	31%	645	49%	1313	41%	2438	40%
no = looking for a job for the first time or after having been out of the labor force.								
yes = unemployed after having been employed.								
FAMILY STATUS:								
single	136	8%	141	11%	391	12%	667	11%
family head	293	19%	351	27%	1256	39%	1900	31%
spouses	419	27%	368	28%	747	23%	1534	25%
other family members	723	46%	467	35%	799	25%	1989	33%
PROVINCE: (region)								
Groningen	91	6%	82	6%	159	5%	331	5%
Friesland	46	3%	49	4%	105	3%	200	3%
Drenthe	52	3%	54	4%	91	3%	198	3%
Overijssel	98	6%	91	7%	208	7%	397	7%
Gelderland	174	11%	150	11%	300	9%	624	10%
Utrecht	89	6%	77	6%	162	5%	327	5%
Noord-Holland	301	19%	195	15%	581	18%	1077	18%
Zuid-Holland	334	21%	238	18%	651	20%	1223	20%
Zeeland	45	3%	41	3%	59	2%	145	2%
Noord-Brabant	224	14%	186	14%	516	16%	927	15%
Limburg	111	7%	160	12%	354	11%	626	10%
Zuid. IJsselmeerpolders	6	0%	3	0%	7	0%	16	0%

## APPENDIX A: (cont.)

	UNEMPLOYMENT DURATION:							
	short		medium		long		total	
SEX:								
male	673	43%	642	48%	1744	55%	3059	50%
female	898	57%	685	52%	1448	45%	3031	50%
MARITAL STATUS:								
unmarried	965	61%	727	55%	1517	48%	3208	53%
married	606	39%	600	45%	1675	53%	2882	47%

=====

Appendix B. Estimated differences between short (1) and long (0) term unemployed in the training and validation sets for the ultimately selected model.

-----

Variables:	Training set		Validation set	
	est.	s.e.	est.	s.e.
GRAND MEAN	-.290	-.227*	-.585	.234
AGE:				
14-19	---		---	
20-24	-.495	.159	-.582	.161
25-39	-1.063	.175	-.867	.180
40-54	-1.705	.213	-1.672	.221
55-59	-2.228	.381	-1.486	.331
60 or more	-2.892	.416	-2.502	.365
EDUCATION:				
low	---		---	
lower medium;general	.281	.195*	.600	.192
" " ;occupation specific	.191	.147*	.411	.146
upper medium;general	1.088	.247	.810	.237
" " ;occupation specific	.422	.158	.542	.158
high; occupation specific	.396	.245*	.418	.226*
high; scientific	.757	.331	.848	.345
unknown or undefined	1.103	.154	1.114	.153
FAMILY STATUS:				
singles	---	---	---	---
family heads	.122	.185*	.203	.181*
spouses	.542	.179	.489	.178
other family members	-.047	.191*	.052	.195*

-----

GOODNESS OF FIT STATISTICS:

Pearson $\chi^2$	2346	2368
Probability level	48%	39%
Scaled deviance	2582	2631
Probability level	0%	0%
Degrees of freedom	2349	2349
Number of short term-unemployed	789	766
Number of long term-unemployed	1576	1599
Total number of unemployed	2365	2365

-----

\* indicates an estimate which does not differ significantly from the grand mean (5% significance level).