Applied Economics, 1988, **20**, 1233–1251

Differences in unemployment duration: a regional or a personal problem?

HENDRIK FOLMER * and JOUKE VAN DIJK †

**Department of General Economics, Agricultural University, P.O. Box 8130, NL-6700 EW Wageningen, The Netherlands* and †*Faculty of Economics, University of Groningen, P.O. Box 800, NL-9700 AV Groningen, The Netherlands*

I. INTRODUCTION

The explosion of unemployment in most Western countries during the past decade has left deep traces in the welfare of individuals and their households, although not all individuals 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 in 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 labour market and welfare policies.

The variation in the duration of spells of unemployment has been intensively studied. Theoretical and methodological aspects have been widely discussed. Micro-economic approaches have been found to be most appropriate to study the underlying causes of variation (Lancaster and Nickell, 1980).[1] Various types of micro models have been developed by, among others, Ehrenberg and Oaxaca (1976); MacKay and Reid (1972); Lancaster (1979) and Nickell (1979). Salant (1977); 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.[2]

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 Labour Force Survey. Two kinds of causal variables will be distinguished. First, personal characteristics of the unemployed such as age, education, family status, sex, work-experience

*Author to whom correspondence should be addressed.

[1] 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 (including 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.

[2] Recent information about these issues can also be found in a special issue about duration analysis of the *Journal of Econometrics* **28**, Annals. 1985–1.

0003–6846/88 $03.00+.12

and the replacement ratio (unemployment benefits to income ratio) and, secondly, variables reflecting the regional demand for labour. 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 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 labour however, migration and demand stimulating policies would be more adequate. (In the sequel policies aimed at improving the personal labour market characteristics (such as schooling and (re-)training) and regional policies (such as migration and demand policies) will simultaneously be denoted as 'labour market policy' if there is no need to distinguish between both kinds of policies).

The organization of the remainder of this paper is as follows. In Section II 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 III 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 labour market characteristics which lead to variation in unemployment duration) is outlined. The econometric method is briefly described in Section IV. The main results of the Dutch case study are presented in Section V. The paper ends with a concluding section where some policy recommendations are formulated.

II. MAIN APPROACHES

Before turning to the various approaches to model unemployment duration, it is important to note that the data provided by the Labour 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 analysed in studies which focus on individual disutilities of (long-term) unemployed (Löfgren, 1976 and Salant, 1977). The present study, however, focuses on the identification of target groups for labour market policy. As these policies concern unemployed with spells in progress exceeding a certain critical length, the interrupted instead of the completed spell should be analysed.[3]

The purpose of this study is to identify the personal and regional characteristics of unemployed whose spell of unemployment is likely to *exceed a critical length of 12 months.* This spell length is chosen because for policy purposes in The Netherlands long-term unemployed are generally defined as unemployed with spell lengths exceeding 12 months (see, for instance, Annual Labour Market Report, 1987, SOZAWE, 1987, p. 49). It should be observed that unemployed with unemployment spells shorter than 4 months are usually

[3]It should be observed that there also exist serious methodological problems with the analysis of completed spells when no systematic information on completed spells and concomitant variables is available as is the case in The Netherlands (and many other countries as well). So, the lengths of the completed spells would have to be estimated from the Labour Force Survey (or a similar data base) containing the lengths of interrupted spells. This estimation must be based on various assumptions, e.g. with regard to the functional form of the escape rate density. For instance, Salant (1977, p. 48) assumed this density to be the gamma and adds: 'This distribution is both flexible and convenient, but choosing it is otherwise arbitrary'.

considered to be frictional. With regard to frictional unemployed the usual kind of labour market policies provided by the Labour Exchange (consisting of the provision of information on vacancies and mediating) are considered to be sufficient to promote (re-)entry into employment. In the case of long-term unemployed (structural unemployed), however, the services provided by the Labour Exchange are considered to be insufficient to accomplish (re-)entry. For these cases special labour market policies are designed, e.g. employment at special projects, (re-)schooling, and demand and/or out-migration policies for regions with high proportions of long-term unemployed due to demand deficiency, etc. (see also Section III).

The upshot of this section is that because of the policy interest in spells in progress exceeding the critical length of 12 months, the interrupted spell of unemployment is more appropriate as an index of policy intervention than the completed spell in the present study.

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). Two types of approaches can be distinguished within this framework. The first consists of analysing the duration of unemployment directly (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 both kinds of models (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 dependants.

b. Characteristics of the individual which may change with the individuals' duration unemployment, i.e. unemployment benefits, other family income and the 'replacement ratio' (the ratio of the unemployment benefits to the expected labour income).

c. Regional demand for labour, i.e. the regional unemployment rate or the regional ratio of vacancies to unemployment.

d. Duration of unemployment.

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 labour market characteristics. Moreover, a spell length exceeding 1 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 has the additional advantage that it makes it easier to identify potential groups with high risks of relatively long spells of unemployment, because the results are not blurred by the duration variable (which, as argued above, implicitly represents genuine labour market characteristics). It is obvious that, for the design of labour 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 primarily 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.

The question may be raised how unobservable explanatory variables, such as health, ambition, 'street' knowledge could be taken into account. In as far as these variables are systematic they often go together with the observable systematic variables (a)–(c) (e.g. health with age, ambition with education, etc.). So, the observable systematic variables also represent the systematic unobservable variables. In as far as they are unsystematic the unobservables are taken into account by the disturbance term.

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. Moreover, the data base will be described.

III. DATA BASE AND CONCEPTUAL MODEL

The data which will be analysed in this paper are micro data from the 1979 Labour Force Survey for The Netherlands.[4] 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 the three duration categories distinguished in the survey.

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

[4]The Dutch Labour Force Survey is part of the sample survey which the Census Bureau of the European Community organizes every 2 years in the member countries. Detailed information about the 1979 survey can be found in CBS (1982). When the present study was carried out, the Labour Force Surveys for more recent years were not yet available. It should be observed, however, that 1979 is rather typical for the regional unemployment situation in The Netherlands for the past decade. Although the unemployment rates for all provinces have substantially increased from 1979 through to 1987, no substantial changes in the regional distribution of unemployment have occurred (SOZAWE 1979, 1986 and 1987).

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

Duration categories	Absolute	Percentages
<4 months	1555	25.5
4–11 months	1361	22.4
⩾12 months	3175	52.1
Total	6090	100.0

long-term unemployed. The second group consists of long-term unemployed (⩾12 months) who have considerably slighter probabilities of getting (re-)employed than the short-term unemployed. In addition to these two groups a third group will be distinguished, 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. With regard to the latter it is important to observe that in times of recessions and increasing overall unemployment rates (as in The Netherlands in 1979) the average search period for frictional unemployed may increase with several months. Therefore, we did not choose for a subdivision of the unemployed in two groups, *viz.* short versus long-term unemployed, but for the subdivision in three groups. The empirical results will show whether or not the intermediate group has to be viewed as frictional short-term unemployed or as long-term unemployed.

In the literature on job search, the duration of unemployment and the (re-)entry of unemployed into employment, the following labour 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 SOZAWE, 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 (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 labour 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 are available, is quite similar to family status. Because of their close correspondence, the use of both family status and sex-martial 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 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 2.5 years for most unemployed.[5] So, differences in individual spells of unemployment may hardly be explained by the 'variation' in the replacement ratio, though there might be an overall increase in spell length.

[5]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 6 months are not entitled to unemployment benefits. The same applies to spouses, who enter the labour 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 2.5 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 previously earned. Finally, those who left 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 over-represented among the short-term unemployed.

The various categories distinguished above are taken into account in the following ways. Spouses and those who enter the labour 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 2.5 years and those who left voluntarily, no information is available in the data set. These categories, however, are very small and may be neglected.

Finally, it should be observed that in 1987 the Dutch social security system was revised. The main features of the revisions are reductions of the unemployment benefits and shorter entitlement periods (for further details see Vrooran, 1986).

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 favourable labour market characteristics which is likely to cause a substantial decrease in their re-entry probabilities (see also Section II). 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 labour 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 2.5 years. For these reasons the replacement ratio will not be included as a discriminating variable in the empirical part of the paper.

The regional demand for labour

There exists a general consensus that regional variation in the level of demand leads to differentials among regions in unemployment duration (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 (Van Dijk, 1986). For instance, the number of vacancies officially registered by the Labour Exchange usually substantially under-estimates the total number of available vacancies. This follows from the fact that firms are not obliged to register vacancies. Vacancies which can be filled internally or via direct contacts between applicants and employers without the intervention of the Labour Exchange are usually not officially registered.

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 labour market, because both employed and unemployed in the regions where the jobs are available as well as employed and unemployed from other regions, may compete. Furthermore, the unemployment figures are affected by registration errors. Finally, hidden unemployment should be taken into account.

From this it follows that the operationalization of the labour market conditions of the regional economy by means of the number of vacancies and the competition at the labour 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. It should be

observed that the use of this dummy variable has the additional advantage of making 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).

The dummy variables are measures of the regional economy and the regional demand for labour in the sense that they indicate differences between provinces. Variations in the regional economy and the regional demand for labour are *a priori* known to coincide with provinces and aggregates of provinces (e.g. the peripheral North and South (with low unemployment rates) versus the metropolitan West (with high unemployment rates), see, for instance, EZ, 1985). So, if the regional demand for labour, which varies with provinces and groupings of provinces, were an important explanatory variable of the length of spells of unemployment, the dummy variables for provinces of residence and groupings of provinces would turn up to be statistically significant. 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.[6]

This section ends with the remark that the regional supply of labour is usually also considered to influence the duration of unemployment. The regional supply comprises demographic and related factors as education and work-experience. It is well known that individuals with similar labour market characteristics tend to cluster in specific regions (Öberg and Oscarsson, 1979). For instance, individuals with higher levels of education are usually over-represented in urban areas. The possible regional clustering of individuals with similar personal labour 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

In addition to the differences between short-, medium- and long-term unemployed in terms of the separate effects of the labour market characteristics discussed above, differences in terms of joint effects of two or more labour market characteristics should be considered.[7] For example, the duration of unemployment for older workers could be larger because they are over-represented in problem regions where the demand for labour is insufficient to absorb laid off (older) workers. Possible relevant joint effects (or interactions as they are usually called) can easily be derived from the exposure given above. The following first-order interactions will at first instance be included in the model:

provincial dummies * each personal characteristic;
age * work-experience;

[6]It should be observed that if the incidence of unemployment in all three duration categories in the problem regions went up to the same extent (which would imply a change in the regional distribution of the overall unemployment rate) no regional effect would be revealed in the present kind of study. This is because its purpose is to analyse differences between short-, medium- and long-term unemployed instead of differences between regional unemployment rates. However, as argued above, a possible regional effect would most probably result in the concentration of long-term unemployed in the peripheral regions with high overall unemployment rates and of short-term unemployed in the metropolitan regions with low unemployment rates. These kind of effects will certainly be picked up by the present kind of analysis.

[7]See Blalock, H. M. Jr. (1971).

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 on 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, no substantial differences between medium and short-term unemployed may occur in the case of an extension of frictional unemployment to include search periods up to approximately 12 months. As noted above, in times of recessions and increasing unemployment rates the average search period for frictional unemployed is likely to go up.

IV. 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} $(j=2, \ldots, p)$ with x_{i1} an indicator for the constant term, for n observations $(i=1, \ldots, n)$. Press and Wilson (1978) and Maddala (1983), among others, have shown that this problem can be analysed 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 ith case belongs to the gth group and 0 otherwise. Moreover, let

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

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

Then

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

where

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

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

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

where α_g is a vector of p unknown parameters and

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

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

Begg and Gray (1984) show that $\beta_g = \alpha_g$ $(g = 1, \ldots, G)$ 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) have 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.

Model 3 will be estimated by means of the computer package GLIM. In order to facilitate the interpretation of the parameter estimates to be presented in the next section we remark that if the $(n \times p)$ design matrix X, where n is the number of observations and p number of explanatory variables, is not of full rank the parameters corresponding to a linearly dependent set of columns of X are called aliased parameters. Aliasing may be extrinsic or intrinsic. In the former case there is linear dependence among columns of X; for instance, if one covariate is a linear combination of other covariates or if there are no observations for some level of a covariate. This kind of aliasing can be handled in the usual way, such as deleting the redundant covariate or level. In the case of intrinsic aliasing, the specification of the linear structure of the population is redundant whatever the design matrix. For instance, if both the overall mean and the effects for all levels of a factor are to be estimated, the sum of the columns of the design matrix for the factor effects is a multiple of the column for the mean. Intrinsic aliasing is a result of the parameterization of the population regardless of the sample obtained. Aliasing as a consequence of having both the overall mean and coefficients for all levels of the explanatory variables in the model is handled by omitting for each explanatory variable one level and by estimating the differences (on the log-odds scale) versus the omitted levels. The omitted values can be viewed as representing a reference group. Its estimated value is the estimate for the overall mean (or grand mean as it is called in the GLIM vocabulary). For further details see Baker and Nelder (1973) and McCullagh and Nelder (1983).

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 will be tried out. 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. Hence, as shown by e.g. Lovell (1983) the probability of a Type I error may actually be much different from the claimed 5% (e.g.). Moreover, there is a substantial *a priori* unknown risk that candidate variables that actually generated the data are not uncovered.[8]

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, pretesting, 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

[8] It should be observed that this risk is also dependent on the selection strategy (such as stepwise regression, maximizing $\bar{R}^2$ and maximizing minimum t-values in absolute value).

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 on 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 the training set. Finally, a model which performs about equally well in both the training and validation sets will ultimately be estimated on the total sample.

V. 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 upon the training set. An interaction was included in the model if this resulted in a drop in the scaled deviance[9] 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.[10] 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[11]. 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 cross-validation results are given in Appendix C. The goodness of fit statistics for both sets (i.e. Pearson χ^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 differnt from zero at the 5% level. Finally, there are no coefficients for which the estimates differ more than twice the minimum standard deviation. Hence, on the basis of the cross-validation criteria outlined in the preceding section the reduced model is acceptable.

[9]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 χ^2 variable. For detailed information on this selection procedure see, Baker and Nelder (1978).

[10]See Appendix B.

[11]It should be observed that even in the training set only two out of eleven provincial dummies (Overÿssel and Limburg) were significant.

Table 2. *Maximum likelihood binary logit estimates of differences between unemployed with different unemployment duration*[a]

	Duration of unemployment categories					
	short vs long		short vs medium		medium vs long	
Variables	(1) *est*	(0) *s.e.*	(1) *est*	(0) *s.e.*	(1) *est*	(0) *s.e.*
Grand mean	−0.467	0.163	−0.113	0.186*	−0.350	0.160
age						
14–19	—		—		—	
20–24	−0.532	0.113	−0.234	0.128*	−0.366	0.121
25–39	−0.950	0.125	−0.320	0.142	−0.673	0.128
40–54	−1.671	0.153	−0.498	0.178	−1.149	0.148
55–59	−2.003	0.242	−0.489	0.292*	−1.508	0.217
60 or more	−2.685	0.257	−0.580	0.351*	−2.267	0.250
Education						
low	—		—		—	
lower medium, general	0.453	0.137	0.220	0.152*	0.202	0.132*
lower medium, occu. spec.	0.249	0.104	0.236	0.119*	0.015	0.098*
upper medium, general	0.957	0.170	0.580	0.185	0.430	0.177
upper medium, occu. spec.	0.492	0.112	0.099	0.123*	0.386	0.103
high; occupat. specific	0.409	0.166	0.267	0.174*	0.705	0.140
high; scientific	0.772	0.240	0.092	0.315*	−0.155	0.286*
unknown and undefined	1.110	0.108	1.540	0.133	−0.435	0.130
Family status						
singles	—		—		—	
family heads	0.048	0.129*	0.093	0.155*	−0.045	0.122*
spouses	0.530	0.126	0.379	0.150	0.212	0.122*
other family members	0.044	0.136*	−0.290	0.159*	0.106	0.132*
Goodness of fit statistics[b]						
	train.	valid.	train.	valid.	train.	valid.
Pearson χ^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 unemployed	2365	2365	1458	1457	2268	2267

*Indicates an estimate which does not differ significantly from the grand mean and from zero in the case of the grand mean (5% significance level; two-sided).

[a]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 over-representation of the category coded (1); a negative coefficient over-representation of the category coded (0).

[b]The statistics for the overall fit refer to 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.

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 χ^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 (Baker and Nelder, 1978), which is the case here. On the basis of the Pearson χ^2 and the finding that 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 is made up of the problem regions Groningen, Friesland, Drenthe and Limburg; the other 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 labour 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 over-represented in the peripheral provinces, might be that unemployed whose spells of unemployment exceed a critical limit leave unemployment via employment programs, leave the regional labour 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 labour market policies and institutional regulations with regard 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 over-represented 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 over-represented 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, 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. This is consistent with the theory that in a tight labour market less favourable labour market characteristics (such as lower medium education and scientific education in specific areas such as linguistics and social sciences) goes together with a prolongation of the (frictional) search period. 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 over-represented 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 labour 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. The second reason is more likely because other categories who also frequently opt for jobs in the secondary labour market (i.e. singles and children still living with their parents) are not over-represented in the category of short-term unemployed.

In general the conclusion can be drawn that between the three groups of unemployed there are substantial differences with regard to personal labour market characteristics. Generally speaking, the possession of labour 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. This means that it consists primarily of frictional unemployed with search periods longer than 4 months and to a less extent of potential long-term unemployed.

VI. 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 labour, 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 labour market characteristics. 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 result determined by demand. 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 labour 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 payback period. Early retirement and extra employment programmes 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.

APPENDIX A

Data and variables

	Unemployment duration							
	short		medium		long		total	
Age (at 1979 Jan. 1)								
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%
lower medium, occu. spec.	239	15%	234	18%	598	19%	1071	18%
upper medium, general.	86	6%	65	5%	96	3%	247	4%
upper medium, 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 labour force.
yes = unemployed after having 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%

	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 model selected on the basis of the training set only

Variables	Training set		Validation set	
	est	*s.e.*	*est*	*s.e*
Grand mean	−0.460	0.343*	−0.603	0.347*
Age 14–19	—		—	
20–24	−0.590	0.166	−0.639	0.170
25–39	−1.137	0.183	−0.920	0.188
40–54	−1.747	0.221	−1.728	0.227
55–59	−2.370	0.387	−2.058	0.337
60 or more	−3.004	0.421	−2.571	0.369
Education				
low	—		—	
lower medium, general	0.336	0.198*	0.590	0.195
lower medium, occupation specific	0.108	0.150*	0.409	0.148
upper medium, general	1.152	0.254	0.831	0.239
upper medium, occupation specific	0.475	0.161	0.557	0.160
high, occupation specific	0.463	0.251*	0.413	0.229*
high, scientific	0.735	0.336	0.823	0.349
unknown or undefined	1.250	0.171	1.243	0.170
Family status				
singles	—		—	
family heads	−0.801	0.273	0.420	0.267*
spouses	0.167	0.240*	0.674	0.247
other family members	−0.564	0.259	0.050	0.270*
Work-experience: no	—		—	
yes	−0.925	0.325	0.293	0.304*

Variables	Training set *est*	*s.e.*	Validation set *est*	*s.e*
Province (region)				
Groningen	—		—	
Friesland	−0.689	0.346*	−0.156	0.342*
Drenthe	−0.528	0.356*	0.242	0.332*
Overijssel	−0.766	0.286	0.057	0.278*
Gelderland	−0.138	0.256*	−0.134	0.253*
Utrecht	0.321	0.307*	−0.046	0.283*
Noord-Holland	−0.308	0.237*	−0.057	0.232*
Zuid-Holland	−0.331	0.235*	−0.069	0.228*
Zeeland	0.304	0.380*	0.027	0.368*
Noord-Brabant	−0.422	0.243*	−0.371	0.240*
Limburg	−0.774	0.266	−0.559	0.264
Zuidelijke Ijsselmeerpolders	0.971	0.882*	0.317	0.902*
*Family status * work-experience*				
family status * no work-experience	—		—	
singles, with work-expereince	—		—	
family heads, with work-experience	1.441	0.384	−0.309	0.357*
spouses, with work-experience	0.972	0.407	−0.325	0.409*
other, with work-experience	1.388	0.374	0.157	0.355*
Goodness of fit				
Degrees of freedom	2334		2334	
Pearson χ^2	2347		2359	
Scaled deviance	2534		2611	

*Indicates an estimate which does not differ significantly from the Grand Mean and from zero in the case of the Grand Mean (5% significance level; two-sided).

APPENDIX C

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

Variables	Training set *est*	*s.e.*	Validation set *est*	*s.e*
Grand mean	−0.290	0.227*	−0.585	0.234
Age				
14–19	—		—	
20–24	−0.495	0.159	−0.582	0.161
25–39	−1.063	0.175	−0.867	0.180
40–54	−1.705	0.213	−1.672	0.221
55–59	−2.228	0.381	−1.486	0.331
60 or more	−2.892	0.416	−2.502	0.365

Variables	Training set *est*	*s.e.*	Validation set *est*	*s.e*
Education				
low	—		—	
lower medium, general	0.281	0.195*	0.600	0.192
lower medium, occupation specific	0.191	0.147*	0.411	0.146
upper medium, general	1.088	0.247	0.810	0.237
upper medium, occupation specific	0.422	0.158	0.542	0.158
high, occupation specific	0.396	0.245*	0.418	0.226*
high, scientific	0.757	0.331	0.848	0.345
unknown or undefined	1.103	0.154	1.114	0.153
Family status				
singles	—		—	
family heads	0.122	0.185*	0.203	0.181*
spouses	0.542	0.179	0.489	0.178
other family members	−0.047	0.191*	0.052	0.195*
Goodness of fit				
Degrees of freedom	2349		2349	
Pearson χ^2	2346		2368	
Probability level Pearson χ^2	48%		39%	
Scaled deviance	2582		2631	
Probability level scaled deviance	0%		0%	
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 and from zero in the case of the Grand Mean (5% significance level; two-sided).

ACKNOWLEDGEMENTS

We gratefully acknowledge the computational assistance of Joost Veenstra and the helpful comments of an anonymous referee.

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.

Blalock, H. M. Jr. (1971) Theory building and causal inferences, in *Methodology in Social Research*, (Eds.) H. M. Blalock, Jr. and A. M. Blalock, McGraw Hill, London, pp. 155–98.

Bron, J. A. H., Gaans, F. van and Nieuwenhuizen, R. D. J. (1983) *Integratie van werkloosheidsregelingen.* Instituut voor onderzoek van overheidsuitgaven, s-Gravenhage.

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, 141–65.

CBS (Central Bureau of Statistics) (1982) *Arbeidskrachtentelling 1979.* Staatsuitgeverij, The Hague.

Clark, K. B. and Summer, L. H. (1979) Labor market dynamics and unemployment: a reconsideration. *Brookings Papers on Economic Activity*, **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**, 128–40.

EZ (Ministry of Economic Affairs) (1985) Nota Regionaal sociaal-economisch beleid 1986–1990. Staatsuitgeverij, The Hague.
Ehrenberg, R. G. and Oaxaca, R. L. (1976) Unemployment insurance, duration of unemployment and subsequent wage gain. *The American Economic Review*, **66**, 5, 754–66.
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; 111–29.
Heckman, J. J. and Singer, B. (1985) Social science duration analysis. In: *Longitudinal Analysis of Labor Market Data*, (Eds.) J. J. Heckman and B. Singer, Cambridge University Press, Cambridge.
Joll, C., McKenna, Chr. and McNabb, R. (1983) *Developments in Labour Market Analysis.* George Allen & Unwin, London.
Lancaster, T. (1979) Econometric methods for the duration of unemployment. *Econometrica* **47**, 4; 939–56.
Lancaster, T. and Nickell, S. (1980) The analysis of re-employment probabilities for the unemployed. *Journal of the Royal Statistical Society A*, **143**, 2; 141–65.
Lippman, S. and McCall, J. (1976) The economics of job search: a survey (Part I and II). *Economic Inquiry*, **14**, 2; 155–89 and 3; 347–68.
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**, 1256–72.
Maddala, G. S. (1983) *Limited-dependent and Qualitative Variables in Econometrics.* Econometric Society Monograph in Quantitative Economics, Cambridge University Press, Cambridge.
Maki, D. and Spindler, Z. A. (1975) The effect of unemployment compensation on the rate of unemployment in Great Britain. *Oxford Economic Papers*, **27**, 440–54.
McCullagh, P. and Nelder, J. A. (1983) *Generalized Linear Models.* Chapman and Hall, London.
Metcalf, D. (1975) Urban unemployment in England. *Economic Journal*, **85**, September: 578–89.
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.* Limburg University, Maastricht: Working paper 85–003.
Nickell, S. (1979) Estimating the probability of leaving unemployment. *Econometrica* **47**, 5; 1249–64.
Öberg, S. and Oscarsson, G. (1979) Regional policy and interregional migration—matching jobs individuals on local labour markets. *Regional Studies*, **2**, 2; 139–63.
Press, S. J. and Wilson, S. (1978) Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association*, **73**, 4; 699–705.
SOZAWE (Ministry of Social Affairs and Employment) (1979) *Rapportage Arbeidsmarkt 1979.* SOZAWE, The Hague.
SOZAWE (1982) *Rapportage Arbeidsmarkt 1982.* SOZAWE, The Hague.
SOZAWE (1983) *Rapportage Arbeidsmarkt 1983.* SOZAWE, The Hague.
SOZAWE (1986) *Rapportage Arbeidsmarkt 1986.* SOZAWE, The Hague.
SOZAWE (1987) *Rapportage Arbeidsmarkt 1987.* SOZAWE, The Hague.
Salant, S. W. (1977) Search theory and duration data: a theory of sorts. *Quarterly Journal of Economics*, **91**, 44–54.
Sider, H. (1985) Unemployment duration and incidence: 1968–82. *American Economic Review*, **75**, 3; 461–72.
Van Dijk, J. (1986) *Migratie en Arbeidsmarkt.* Van Gorcum, Assen.
Van Opstal, R. and Theeuwes, J. (1986) Duration of unemployment in the Dutch youth labour market. *De Economist*, **134**, 3; 351–67.
Vrooman, I. C. (1986) Werklozen nade stelselherziening, *Economisch Statistische Berichtea*, **71**, 3538, 36–40.
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*, **3**, 133–54.