Regional Studies, Vol. 19.3, pp. 243–256.

Entry of the Unemployed into Employment: Theory, Methodology and Dutch Experience[1]

JOUKE VAN DIJK and HENDRIK FOLMER
Department of Economics, University of Groningen, P.O. Box 800, NL9700 AV Groningen, The Netherlands

(Received November 1983; in revised form May 1984)

VAN DIJK J. and FOLMER H. (1985) Entry of the unemployed into employment: theory, methodology and Dutch experience, *Reg. Studies* **19**, 243–256. This study examines the entry probabilities of the unemployed into the employed labour force. The personal characteristics (age, work experience, education and family status) and the regional characteristics (regional unemployment and average regional income) are theoretically derived as the determinants of the entry probabilities. The theoretical model is empirically tested with data from the Dutch Labour Force Survey 1979 by means of maximum likelihood logit analysis. The results closely conform to the theoretical expectations. On the basis of the analysis some conclusions for labour market policy are formulated.

Labour market Unemployment Policy Logit analysis Entry probabilities The Netherlands

VAN DIJK J. and FOLMER H. (1985) L'entrée des chômeurs dans l'emploi: théorie, méthodologie et expérience néerlandaise, *Reg. Studies* **19**, 243–256. Cette étude approfondit jusqu'à ce quel point il soit probable que les chômeurs puissent se lancer sur le marché du travail. Les caractéristiques personnelles, à savoir âge, expérience du travail, éducation, couche sociale, et les caractéristiques régionales, à savoir chômage régional et revenu régional moyen, sont puisées théoriquement comme déterminants de la probabilité d'entrée. Le modèle théorique est mis à l'épreuve de façon empirique à partir des données provenant de l'enquête néerlandaise sur la main-d'oeuvre menée en 1979 à partir d'une analyse de type 'logit' comportant le maximum de vraisemblance. Les résultats se rapportent nettement aux espérances théoriques. Quelques conclusions qui influeront sur la politique de main-d'oeuvre sont formulées à la suite de cette analyse.

Marché du travail Chômage Politique Analyse de type 'logit' Probabilités d'entrée Pays-Bas

VAN DIJK J. and FOLMER H. (1985) Der Eintritt Arbeitsloser ins Arbeitsverhältnis: Theorie, Methodologie und Erfahrungen in den Niederlanden, *Reg. Studies* **19**, 243–256. Diese Studie untersucht die Wahrscheinlichkeit, mit der Arbeitslose ins Erwerbslebens eintreten. Persönliche Eigenschaften wie Alter, Arbeitserfahrung, Bildung und Familienstand sowie regionale Merkmale, regionale Arbeitslosigkeit und regionales Durchschnittseinkommen werden theoretisch als die bestimmenden Faktoren der Eintrittswahrscheinlichkeit abgeleitet. Das theoretische Modell wird mittels einer grösstmöglichen Wahrscheinlichkeitslogitanalyse an Hand von Angaben der Erhebunen über niederländlische Erwerbstätige im Jahre 1979 einer empirischen Prüfung unterzogen. Die Ergebnisse entsprechen weitgehend den theoretischen Erwartungen. Auf der Grundlage der Analyse werden Folgerungen für die Stellenmarktpolitik formuliert.

Stellenmarkt Arbeitslosigkeit Bestrebungen Logitanalyse Wahrscheinlichkeit des Eintritts Die Niederlande

INTRODUCTION

The worldwide rise in unemployment has been accompanied by a rise in research on the working of the labour market. Several fields in this research area can be distinguished, such as the role of unemployment compensation, the impediments of governmental regulations, the segmentation of labour markets, and the personal and regional determinants of unemployment (see, for instance OECD, 1979; and MADDISON and WILPSTRA, 1982, 1983). It is obvious that insight in these aspects of the working of the labour market is not only of scientific interest, but may also be of importance for the design of labour market policies.

The present paper is part of a more comprehensive study which aims at identifying the role of migration in regional labour markets. The main problem of that study is whether migration implies competition to local job searchers and thus may cause a rise in regional unemployment, or whether migration leads to an optimal fulfilment of key economic positions and thus has a positive effect on the regional economy and on regional employment (see VAN DIJK and BARTELS, 1982, for a broader outline of this project and for a first attempt to answer this question by means of using macro-data for occupational groups). Part of this problem is the investigation of the determinants of the entry of unemployed into the labour force.

Globally speaking, two categories of variables which influence the probability of an individual entering the labour force can be found in the literature: (1) personal characteristics, such as age, sex, marital status, education, acquiring capacities, and the replacement ratio, i.e. the unemployment benefits to income ratio, and (2) characteristics reflecting the situation of the (regional) labour market. The relative importance of various personal and regional characteristics is investigated here on the basis of cross-section micro data for the Netherlands in 1979.

The organization of this paper is as follows. In the next section a theoretical model of the entry of unemployed in the labour force is presented. In the third section the statistical estimation procedure of the entry model is described and in the fourth section the empirical results are discussed. In the final section some conclusions for labour market policy are formulated. The data base and the variables are described in the Appendix.

A THEORETICAL MODEL OF ENTRY INTO EMPLOYMENT

A general framework

In this section a theoretical model of entry in the labour force is introduced which serves as a framework of interpretation for the empirical model. It should be noted that the theoretical model is of a rather general nature and could be refined in various respects. An individual's *employment history* during an observation period $[t_o, \tau]$ can formally be represented as:

$$h[t_0, t_n] = \{t_0, y_0, t_1, y_1, \ldots, t_n, y_n\} \quad (1)$$

where: t_k, $k = 1, 2, \ldots, n$;
$t_n \leqslant \tau$ = the timing of an event, i.e., a change from employment to unemployment or vice versa
y_k, $k = 1, 2, \ldots, n$, represents the individual's employment state entered at each event.

(It should be noted that a subscript i referring to the individual is omitted for the moment). In the present case y_k is a binary variable. However, generalizations to several possible states can easily be made. By means of the *survivor function* the probability for event n occurring *after* time t, given the individual's history from the start of the process at t_o, can be described. It is defined as:

$$S_j(t \mid h[t_0, t_{n-1}]) = P(T_n \geq t \mid h[t_0, t_{n-1}]) \quad (2)$$

where the random variable T_n denotes the timing of the nth event and where $j = y_{n-1}$ at time t_{n-1}, i.e. that the individual is in state y_{n-1} at time t_{n-1}.

The *hazard function*, is defined as the limiting probability that the nth event takes place at time t given the individual's previous history and that the nth event has *not* occurred before t.

$$\begin{aligned} H_j(t \mid h[t_0, t_{n-1}]) &= \lim_{\Delta t \downarrow 0} \frac{S_j(t \mid h[t_0, t_{n-1}]) - S_j(t+\Delta t \mid h[t_0, t_{n-1}])}{S_j(t \mid h[t_0, t_{n-1}])\,\Delta t} \\ &= \frac{-d \log S_j(t \mid h[t_0, t_{n-1}])}{dt} \end{aligned} \quad (3)$$

The complementary natures of the hazard and survivor functions can be derived from:

$$\log S_j(t \mid h[t_o, t_{n-1}]) = -\int_{t_{n-1}}^{t} H_j(v \mid h[t_o, t_{n-1}])\, dv \quad (4)$$

Because of (4) only the hazard function is considered further here. The purpose of our analysis is to describe the hazard function in terms of a set of explanatory variables, denoted here by the vector x:

$$H_j(t \mid h[t_o, t_{n-1}]) = \phi(x) \quad (5)$$

When a population of individuals is considered, only a limited number of systematic explanatory variables can be taken into account and the effects of unobserved variables are represented by a disturbance term ε_i. This gives the following relationship for individual i:

$$H_{ij}(t \mid h[t_0, t_{n-1}]) = \alpha^T x_i + \varepsilon_i \quad (6)$$

It should be noted that the research design described above, which is usually denoted as event-history design, gives ample opportunity to analyse the causal structures of various economic and social processes. (for applications in the field of job change see SØRENSEN and TUMA, 1981; and FELMLEE, 1982). However, the data set to be analysed in this paper does not contain information on the individual's employment history and the timing of his entry into the labour force. Therefore, only a modest version of model (6) will be analysed below.

Let us now turn to the vector x of systematic variables. Two main categories of systematic explanatory variables can be distinguished: (1) variables describing personal characteristics; and (2) variables describing area characteristics. Both kinds of variables influence an individual's job search behaviour (described by a job search model) as well as the probability of getting a job offer and being accepted by an employer for a specific job opening (determined by the hiring process of employers).

Determinants of job search behaviour of unemployed

In this section the variables determining the entry of unemployed into the labour force as a result of job search activities are discussed. Following among others JOLL *et al.*, 1983; LIPPMAN and McCALL, 1976; and ROGERSON, 1982, elementary *job search models* assume that each worker knows precisely the pecuniary and non-pecuniary rewards attached to each job, where all relevant jobs are, and where vacancies exist. Each employer knows what rewards other firms are offering for specific tasks, and what each available worker has to offer in terms of total productive capability.

In reality, of course, employers and employees only have sketchy knowledge. Furthermore, search for information leads to both benefits and costs and therefore may not go on until the stage of full information is reached. Search will continue as long as its benefits exceed its costs. It should be noted that the early job search models were non-sequential in nature in the sense that the number of job offers to be investigated by the searcher was assumed to be a constant. The recent literature is dominated by *sequential* models where the choice of when to stop searching is determined by an optimal stopping procedure. The optimal search policy is characterized by the existence of a reservation wage. An individual accepts any job offer for which the wage exceeds the reservation wage; otherwise he keeps on searching (for further details see LIPPMAN and McCALL, 1976; ROGERSON, 1982; and ROGERSON and McKINNON, 1981). Two remarks have to be made in this respect. First, the concept of a reservation wage does not only apply to pecuniary rewards but may also be interpreted in terms of utilities. The same applies to the notion of wage to be used below. Secondly, one's reservation wage is set on the basis of one's expected market value, which is dependent on such variables as age, education, the situation at the labour market, etc.

The information sought by workers can relate to various aspects of the working environment. These aspects can be divided into two categories: those which are observable by *inspection* and those which are observable by *experience* (see JOLL *et al.*, 1983). Characteristics belonging to the first category are: wage offer; working hours; overtime availability; holidays; recreation facilities; etc. Examples of the second category are: promotion prospects; supervision; safety; fringe benefits; etc. Characteristics observable by inspection can be discovered in a straightforward way whereas information on characteristics observable by experience can only be obtained by investigation.

In order to measure the information held by an individual one could think of such items as personal contacts of a job searcher with firms or with intermediate informants; knowledge of job offers in local, regional or national newspapers, etc. However, the data base to be analysed in the empirical section does not contain any information on this kind of item. We expect that differences in information available may be explained by personal characteristics such as age, education, etc. So, these variables are also used as proxies for differences in information available.

Selection of job applicants

The outcomes of the search process are not solely determined by the search activities of workers, but also by the *probability of getting a job offer* during a period. This depends heavily on the hiring behaviour of firms.

MIRON, 1978, distinguishes between two polar views of the hiring process: (1) 'the 'bingo' model in which a job opening is offered to the next applicant who appears; and (2) the 'queueing' model in which firms order applicants according to their competences and construct a waiting list of acceptable job applicants. When a vacancy occurs a job is offered to the first person in the queue. In terms of realism the bingo model is rather unsatisfactory. Firms will not choose randomly from the pool of seekers and the probability of being a successful applicant is not the same for each job seeker. Therefore, the queueing model, which does not make these assumptions, is to be preferred.

In the queueing model a firm selects applicants by means of a quality criterion. In this connection the distinction between characteristics observable by inspection and by experience (JOLL *et al.*, 1983) applies again. To the second category belong performance, reliability, flexibility, punctuality and creativity. Because the characteristics observable by experience are not observable by inspection, a firm tries to identify certain characteristics of job applicants which are controllable, readily observable, communicable, and convey information on quality (JOLL *et al.*, 1983). Characteristics such as education, age, race, work experience, sex and marital status are among the most important criteria; therefore these characteristics are important determinants of the probability of job finding. The optimal search policy of a firm is very similar to that of a job seeker, but is characterized by the existence of a minimally acceptable productivity ('reservation productivity') instead of a reservation wage (LIPPMAN and McCALL, 1976).

Overview of the explanatory variables of the entry-model

In this section the determinants of the entry of unemployed into the labour force are summarized. Furthermore, we will indicate whether a given variable increases or decreases the probability of getting a job. We will also pay attention to variables which at first sight seem relevant in this connection but which are not incorporated in our model and to interactions between variables.

Age. In general, the probability of getting a job is assumed to decrease with age, because the young 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. At first sight, this general picture may not apply to the youngest age-groups. Young people usually lack any work experience (SOZAWE, 1983a) and, before they have found out what work they prefer, they are likely to try out various jobs (job-hopping). Employers may be less inclined to hire very young job applicants, because of their rather uncertain productivity level and the risk that they quit after a short while (VAN BEKKUM, 1983). However, when work experience is controlled for in the model, the general negative relationship between getting a job and age is assumed to hold. This tendency is likely to be strengthened by the lower (minimum) wages for young people.

It should be noted that figures on labour market dynamics from the Ministry of Social Affairs and Employment (see SOZAWE, 1982, 1983) show that young people are over-represented in the flow into unemployment as well as in the outflow. This indicates a high job turnover rate and so a high number of young unemployed persons can go together with high probabilities of getting a job. In the studies of NICKELL, 1979; LANCASTER, 1979; LANCASTER and NICKELL, 1980; and KOOREMANN and RIDDER, 1983, age indeed shows a significant negative effect on the probability of getting a job. However, HERZOG and SCHLOTTMANN, 1982, who examined US data on the probability of being employed in 1970 for employed and unemployed blue-collar workers in 1965, did not find a significant effect of age.

Education. This variable reflects three aspects. First, it contains information on the level of education in terms of years of schooling. Secondly it distinguishes between types of education which in the present study are occupation specific, scientific and more general forms of education. Thirdly, it partly reflects the heterogeneity of labour so that the group specific situation on the labour market is taken into account by means of this variable. The probability of getting a job is expected to increase with the level of education and to be highest for those with occupation specific education. The reason for this is that people with a higher level of education are better equipped for job search. Furthermore, people with a higher level of education or occupation specific education have higher productivity rates.

Family status. We expect family members to have higher probabilities of obtaining a job than singles (i.e. unmarried, widowed and divorced people), because employers value family membership as an indicator of reliability and stability. The expectation is confirmed by NICKELL, 1979; and HERZOG and SCHLOTTMANN, 1982. We assume that, in the typology of the segmented labour markets, family heads apply merely in the primary segment embracing jobs characterized by high earnings, good working conditions, good opportunities for promotion, employment stability and well established and fairly administered rules governing work. Singles, children and especially spouses apply much more than family heads for secondary jobs. The jobs are low paying, offer few fringe benefits and poor working conditions and involve menial and repetitive work. In this segment patterns of employment are unstable, opportunities for promotion are few and firm management is often arbitrary and unfair (see for instance JOLL *et al.*, 1983 for further details on segmented labour markets). One of the characteristics of jobs in this secondary segment of the labour market is a rather high job turnover rate. This implies that the probability of getting a job will be rather high because of the high number of job openings.

Because the outcomes of the countervailing tendencies mentioned above are unknown, the effect of this variable on the probability of getting a job is unknown *a priori.*

Experience. This variable distinguishes between unemployed, who look for jobs for the first time and unemployed, who lost their jobs. On the one hand one might expect the latter, who have work experience, to have the best probabilities of finding a job. On the other hand, the fact that an individual has been employed may influence the level of his reservation wage. This will at least be set equal to his unemployment benefit. It is possible that he gets a wage offer below his reservation wage because part of his work experience may be firm specific and not of full utility for other employers. Under these circumstances the job offer may be unacceptable to him. Therefore, it is uncertain whether experience has a positive or negative effect.

The benefits for individuals without work experience cannot be based on past earnings. Their benefits are on a level equal to or below the lowest unemployment benefits. It should be noted that some groups of unemployed, such as married women, have had no benefits at all. So, the reservation wage of unemployed without work experience is assumed to be lower than that of unemployed with work experience. The variable experience also reflects differences in the replacement ratio among individuals. As will be outlined at the end of this section this is the only way differences in the replacement ratio are taken into account here.

Because of the counterworking tendencies mentioned above, the sign of the experience variable is not clear *a priori*: experience increases the probabilities for an individual who has been employed, but this may be offset by the higher level of his reservation wage.

Regional unemployment. It is expected that in regions with relatively high unemployment the probability of getting a job will be smaller than in regions with relatively low unemployment, because in the former kind of regions the number of competitors will be greater for each job opening.

The unemployment rate may also negatively affect searching by means of the additional- and discouraged-worker effects (see, among others, JOLL *et al.*, 1983; and VAN DER VEEN, 1982) and thus the probability of getting a job. Finally, unemployment may be seen as an indicator of the economic situation in a region. It should be noted that one would prefer a variable like the ratio of vacancies to unemployment in order to measure job competition but adequate vacancy data are not available in our data set.

Instead of a measure of total unemployment, NICKELL, 1979, uses an occupation specific (three digit) measure of unemployment. It is not clear to us if the latter is preferable. The effects of the occupation specific unemployment variable are partly reflected by the education and skill variables. Furthermore, the frequently occurring possibility of occupational mobility is ignored by such a refined unemployment variable. It is also thinkable that search behaviour is more influenced by general unemployment than by group-specific unemployment situations (SCHIFFEL and GOLDSTONE, 1976).

With our data the computation of group specific unemployment rates is possible, but for the above mentioned reasons only a general unemployment measure is operationalized in this study.

The empirical investigations by LANCASTER, 1979; LANCASTER and NICKELL, 1980; and KOOREMAN and RIDDER, 1983, indeed show a significant decreasing effect of regional unemployment rates on the probability of getting a job. Most striking in these studies is the fact that both age and unemployment show significant (negative) effects on the entry probability. However, the effect of the individual characteristic age is much greater than the regional characteristic unemployment.

Regional income. We assume the probability of getting a job to be higher in regions with an average income level above the national level. First, the probability of getting a job offer which exceeds a given reservation wage will be greater in high income regions, because there are likely to be more jobs with high wages and thus more job offers with high wages. Secondly, the average regional income level may also be seen as a general indicator of economic well-being. As mentioned above this aspect is also partly reflected by means of the unemployment variable.

Replacement ratio and migration. We will now pay attention to variables which are *not* explicitly taken into account in our analysis, i.e. the replacement ratio and migration. We start with the replacement ratio (unemployment benefits to income ratio), which is generally considered to be an important explanatory variable (see LANCASTER, 1979; NICKELL, 1979; LANCASTER and NICKELL, 1980; and LIPPMAN and MCCALL, 1976). The main reason for the omission is that the data base does not contain any information on this variable. At first sight the omission of the replacement ratio from the investigation may seem unwarranted because it is well-known that the omission of important explanatory variables may lead to specification errors. However, in the present analysis of the Dutch Labour Force Survey, omitting the replacement ratio is not likely to lead to specification errors, because the only important difference in this ratio is between unemployed individuals, who look for a job for the first time, and individuals who became unemployed after having been employed. This can be seen as follows.

In general, the benefits for individuals *without* work experience are set on a level equal to or below the lowest unemployment benefits. Everyone who has been employed for sixty-five days receives an unemployment benefit of 80% of his last earnings during the first six months. The next two years this percentage falls to 75%, except for a married woman who is not the main wage-earner. They receive only an unemployment benefit during the first six months. After corrections for taxes and social premiums the replacement ratio ranges from 0·993 for minimum wage earners to 0·792 for those earning three times the minimum wage during the first half year. For the next two years these figures are 0·985 and 0·751 (WIEBRENS, 1981). It should be noted that, compared to LANCASTER, 1979, the possible variation in the replacement ratio among unemployed who have been employed is very small in the Dutch case. Consequently the difference in the replacement ratio is very small *within* the two categories of unemployed persons (on the one hand, those who have been employed before and on the other hand those without work experience who look for a job for the first time). But there may be a substantial difference *between* the two categories. Unfortunately there is no information about the replacement ratio in the survey. However, the differences between both categories are, as mentioned before, crudely reflected by means of the variable work experience.

In addition to the considerations given above there is further evidence from empirical research for the weak relationship between entry in the labour force and the replacement ratio in The Netherlands (see SIDDRÉ and THEEWES, 1977; IVA, 1977, cited in SOZAWE, 1982; and BRON *et al.*, 1983, cited in SOZAWE, 1983b, who examined Dutch data for

1979). In these studies no statistically significant effect of the level of unemployment benefits on the duration of unemployment was found.

In some studies, e.g. HERZOG and SCHLOTTMAN, 1982, migration is also used as a proxy for information acquiring, in particular, for job search outside of the region of residence. However, this feature of migration only applies to speculative migrants who move first and then look for jobs (post-migration search). It certainly does not apply to contracted migrants who move with jobs in hand (see SILVERS, 1977; ROGERSON and MACKINNON, 1981). In the latter case migration is a *consequence* of getting a job instead of a cause.

The data set to be analysed does not allow a distinction to be made between speculative and contracted migration. Furthermore, pure speculative job related migration hardly occurs in modern developed countries (see FLOWERDEW, 1983). This applies especially to the Netherlands because of the relatively small distances and the slight differences in employment opportunities between regions. Another reason why migration cannot be used as a proxy for information acquiring is that it only relates to people who actually move. People who did acquire information outside of their own region, but did not succeed in getting a job, or who accepted a job for which no change of residence was necessary, are not identified by this variable. Furthermore, migration may also be induced by factors that are not related to job search. In that case a measured change of residence is an invalid indicator for information acquiring outside of their own region.

For the reasons mentioned above, information acquiring, both inside and outside of their own region will be assumed to be satisfactorily operationalized by such variables as age and education. Therefore, migration will not be used as an explanatory variable of the entry into the labour force, although a migration variable is available in the data set.

The foregoing brings us to the conclusion that the omission of the replacement ratio and migration does not lead to a serious mis-specification problem in the present study.

Interactions. Until now we have considered the effects of the explanatory variables on the probability of getting a job separately. Besides these so-called main effects, attention has also to be paid to the *interaction* effects between variables. Possible important interactions can easily be derived from the descriptions of the individual variables given above. The following first-order interactions will in the first instance be included into the model:

age—work experience
family status—age
family status—education
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 interactions other than the ones mentioned. Furthermore, when significant first-order interaction effects have been found higher-order interactions will be considered.

ESTIMATION AND TESTING PROBLEMS OF THE ENTRY MODEL

In this section the statistical aspects of the entry model are described. Furthermore, attention is paid to the problems which follow from the fact that one and the same sample (the 1979 Labour Force Survey) is used for both finding a model that fits the data at hand well (in particular the detection of 'significant' interaction effects) and testing purposes. We will start with a description of the entry model.

In the preceding section, the hazard function has been used to describe the probability of an individual entering into the labour force given his employment history. In the present application the general model (6) has to be simplified because no information is available about the individual's history and the timing of his entering (see Appendix). Therefore, only the occurrence or non-occurrence of an individual entering into the labour force during the year between April 1978 and April 1979 as a function of personal and regional characteristics is considered here. Model (6) is now specified as follows (see also GOLDBERGER, 1964; and MADDALA, 1983). Let y_i^* be an unobservable variable defined by the relationship:

$$y_i^* = \alpha^T x_i + \varepsilon_i \qquad (7)$$

with $E(\varepsilon_i) = 0$

Furthermore, a dummy variable y defined by:

$$y_i = 1 \quad \text{if } y_i^* > 0 \qquad (8a)$$

$$y_i = 0 \quad \text{otherwise} \qquad (8b)$$

is observed. The situation $y_i = 1$ corresponds to the individual being employed at the moment of observation and the situation $y_i = 0$ to being unemployed at the time mentioned.

From (7) and (8a) it follows that:

$$P(y_i = 1) = P(\varepsilon_i > -\alpha^T x_i) = 1 - F(-\alpha^T x_i) \qquad (9)$$

where P denotes the probability and F is the cumulative distribution function for ε. It should be noted that y follows a Bernoulli distribution with $P(y_i = 1)$ given by (9) and $P(y_i = 0) = F(-\alpha^T x_i)$. From (9) the likelihood function L can be derived as:

$$L = \prod_{y_i=1} [1 - F(-\alpha^T x_i)] \prod_{y_i=0} F(-\alpha^T x_i) \qquad (10)$$

If the cumulative distribution of ε is the *logistic* distribution the model is called the logit model. Under the assumption of the logistic distribution with the scale parameter fixed at 1 (9) reads as:

$$1-F(-\alpha^T x_i)=\frac{\exp(\alpha^T x_i)}{1+\exp(\alpha^T x_i)} \qquad (11a)$$

similarly:

$$F(-\alpha^T x_i)=\frac{1}{1+\exp(\alpha^T x_i)} \qquad (11b)$$

Furthermore, (10) becomes:

$$L=\prod_{i=1}^{n}\left\{\frac{\exp(\alpha^T x_i)}{1+\exp(\alpha^T x_i)}\right\}^{y_i}\left\{\frac{1}{1+\exp(\alpha^T x_i)}\right\}^{1-y_i} \qquad (12)$$

where n is the sample size.

Maximizing (12) (or its logarithm) with respect to α gives the maximum likelihood estimates of α. For this purpose use can be made of, among others, the computer package GLIM (BAKER and NELDER, 1978). This program also produces an estimate of the asymptotic variance-covariance matrix of the maximum likelihood estimator of α.

It should be noted that several alternatives to the logit model exist among which the linear probability model and the probit model are most frequently used. However, the linear probability model has the defect that the conditional expectation $E(y_i|x_i)$, which is interpreted as the probability that the event will occur given x_i is not constrained to [0,1]. Furthermore, in the univariate dichotomous case a simple transformation of the logit estimates gives results which are almost equal to probit estimates, except in cases where observations are heavily concentrated at the tails (AMEMIYA, 1981). So, in the present case, the logit and probit models are statistically hardly distinguishable. We want to remark that MADDALA, 1983, gives suggestions to compare both types of models, e.g. by means of the calculated sum of squared deviations from predicted probabilities. In order to make the model estimates to be presented in the following section easily understandable, a few remarks have to be made here about estimation and testing of logit models in general and about the GLIM package in particular. It should be noted that these remarks also apply to other types of models, such as the probit model.

The first remark relates to *aliasing*. When the $(n \times p)$ design matrix X, where n is the number of observations and p the number of explanatory x 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 intrinsic, for example, when both an overall mean and effects for all levels of an explanatory variable are included in the model; or extrinsic, for example, when one explanatory variable is a linear combination of other explanatory variables, or when there are no observations for some level of an explanatory variable. Linear dependency is taken into account in the GLIM framework by assigning zero values to the aliased parameters until a maximal independent subset is obtained. It should be noted that in the case of aliasing as a consequence of having both the overall mean and coefficients for all levels of an explanatory variable in the model, the coefficient of the first level of the explanatory variable concerned is put equal to zero. This means that the coefficients estimated, which are all differences on the log-odds scale, are expressed versus the omitted level. So, the omitted level is used as a reference level.

The GLIM system also gives measures of goodness-of-fit, referring to the correspondence between the hypothesized model and the data at hand. One measure is the *scaled deviance*, defined as:

$$SD=-2\log\frac{\lambda_h}{\lambda_s} \qquad (13)$$

where λ_h and λ_s are the likelihood functions of the hypothesized and the saturated models, respectively. The saturated model contains all possible linearly independent parameters, whereas in the hypothesized model a set of parameters is restricted. Under quite general regularity conditions SD is asymptotically distributed as a χ^2 variable with degrees of freedom equal to $(t_s - t_h)$, where t_s and t_h are the numbers of independent parameters in the saturated and hypothesized models (see, among others, RAO, 1965; KENDALL and STUART 1967).

It should be noted that in a small data set the scaled deviance may be considerably less than its expected value. On the other hand, in very large samples with data on counts or proportions the scaled deviance is likely to be larger than expected, even for models with a good fit (see BAKER and NELDER, 1978). This implies that the scaled deviance is not the most appropriate test statistic for the evaluation of the overall goodness-of-fit of the whole model in the present analysis where the sample size n = 1,934. Under these circumstances other measures of goodness-of-fit should be used, e.g. the Pearson χ^2, which is also given by the GLIM program. The scaled deviance, however, is a useful statistic to test sequences of nested hypotheses. The difference in values of the scaled deviances for the models under comparison is asymptotically distributed as a χ^2 variable, with degrees of freedom equal to the number of independent restrictions, i.e. to $(t_i - t_j)$, where i is the less restrictive model. So, if a model contains the linearly independent parameters $\alpha_1, \alpha_2, \ldots, \alpha_p, \alpha_p + {}_1, \ldots, \alpha_r$ and if we want to test the hypothesis that the first p parameters are all equal to zero then the difference in scaled deviances of the two models is asymptotically ${\chi^2}_p$ distributed. It should be noted that this χ^2 test on differences can also be applied on single parameters, i.e. on hypotheses of the form $H_o: \alpha_i = 0$ versus $H_a: \alpha_i \neq 0$. Furthermore, this test is to be preferred to t-tests where parameter estimates are compared with their standard errors (see BAKER and NELDER, 1978). It should be noted that the χ^2 test on differences is independent of the fact whether or not the less restrictive hypothesis is true (LEHMANN, 1959). So, it is possible to test whether a given model is 'worse' than a less restrictive 'bad' model. Furthermore, when a sequence of nested hypotheses is tested, one should start with the less restrictive hypothesis of the sequence (MALINVAUD, 1970).

Not only may the inclusion of a given variable into the model be tested, but also the reduction of the number of levels in a given variable. In the second case the hypothesis, that the coefficients associated with two or more different levels of a given variable are the same, is tested. Coefficients which are hypothesized to be equal may be constrained by combining the corresponding levels. The test statistic for this problem is constructed as follows. First the difference in scaled deviances of the model with all levels and the model with a reduced number of levels is calculated and divided by the appropriate number of degrees of freedom. The scaled deviance thus obtained is called the scaled mean deviance. Next the ratio of this scaled mean deviance to the scaled mean deviance of the full model is calculated. Because both components of this ratio are χ^2-variates divided by their degrees of freedom this ratio is approximately distributed as a F-variable with appropriate degrees of freedom, if the χ^2 variates are independent. Because $t^2_n \sim F^1_n$ the combination of two levels only can be tested by means of Student's t distribution.

As mentioned in the introductory remarks to this section two forms of model judgement have to be distinguished: genuine hypothesis testing (considered above); and assessment of the fit of a tentative model to the data at hand. In the case of genuine hypothesis testing an *a priori* hypothesis concerning the model structure is available. The plausibility of this hypothesis is tested on the basis of the sample data according to the rules of standard statistical decision theory. In the case of assessment of model fit an *a priori* hypothesis is not available. On the contrary, the purpose of this analysis is to explore the data in order to find an appropriate model (see FOLMER, 1984b). In practical research both forms of model judgement are frequently intertwined. The same data, that gave birth to the model ultimately chosen, are also used for testing purposes. Intertwining both forms, usually denoted by such fanciful expressions as data mining, data grubbing, fishing or torturing the data, makes the greatest use possible of any and all idiosyncrasies of the data. Therefore, the goodness-of-model-fit to the sample data is likely to be greater than the fit to the population.

In this connection it is worthwhile mentioning briefly the results of LOVELL, 1983. He shows that the probability of a Type I error for a data miner, who uncovers t-statistics that appear significant at the 5% level by running a large number of alternative regressions on the same data, is actually much greater than the claimed 5%. Furthermore, by way of simulation experiments he investigates the problem whether data mining is likely to uncover those candidate variables that actually generated the data. Three strategies are considered: stepwise regression, maximizing $\bar{R}^2$ and maximizing minimum t-values in absolute value. The first two procedures are rather successful in identifying the correct or related variables, whereas the third procedure has a very poor performance. So, the author correctly concludes: 'It is ironic that the data mining procedure that is most likely to produce regression results that appear impressive in terms of the customary criteria is also likely to be the most misleading in terms of what it asserts about the underlying process generating the data under study'. However, the claims of significance achieved by stepwise regression and maximising $\bar{R}^2$ are exaggerated. Therefore, the author gives a 'rule of thumb' for deflating the claims of significance.

Several procedures, which describe how to proceed in the absence of a tightly structured theory, have been developed. Well-known are Bayesian strategies (GAVER and GEISEL, 1974), replication (KISH and FRANKEL, 1974), and cross-validation techniques (MOSTELLER and TUKEY, 1977; GEISER, 1974, 1975; STONE, 1974; and MCCARTHY, 1976). In this connection we would also like to draw attention to the prospects of bootstrap procedures (PLATT, 1982).

In the present study a *simple* form of cross-validation will be used. The sample, consisting of 1,934 elements, will randomly be halved. The model described in the preceding section will be estimated using one set of observations. Furthermore, some plausible interactions not mentioned above will be tried out on the same set of data. Next, the model ultimately chosen will be estimated using the other set of observations. If both sets of estimates do not differ substantially, i.e. are within a range of ± 2 standard errors, the model will be estimated using the total set of observations. These estimates will be used as the ultimate estimates. In the case of substantial differences between the estimates based on the two halved sets, more sophisticated methods of estimating the accuracy of the estimators will be applied.

The main reason for using this simple form of cross-validation is that in the present study a quite well structured theory is available. The model described in the preceding section is based on theoretical considerations, most of which have quite frequently been tested empirically. In that sense this study is a replication itself. Furthermore, the sample size is such that halving has little effect on the quality of the estimators, so that there is no need, for example, for balanced sampling techniques (MCCARTHY, 1976) for efficiency reasons.

Finally, we want to mention that a drawback of forms of cross-validation, where one single sample is divided, is that the test sample is usually much more like the selection sample than is typical of the population (MOSTELLER and TUKEY, 1977). We will end this section by a brief description of the model selection procedure. As mentioned above, the goodness-of-fit of a model is indicated by its χ^2 value. Consequently the difference in goodness-of-fit between candidate models can be investigated by calculating the difference in χ^2 values, where the model with the lowest value has the better fit. So, a variable whose preliminary incorporation into the model leads to a substantial drop in χ^2 value, i.e. a drop larger than the accompanying loss of degrees of freedom, is worthwhile being definitely included into the model. A variable having a drop in χ^2 value equal to or smaller than the loss of degrees of freedom should be left out of the model (see, for instance, FOLMER, 1984a). By screening all candidate variables systematically, for instance by means of the tables advocated by GOODMAN, 1978, the model which best fits the data can be formed.

EMPIRICAL RESULTS

In this section the empirical results of the investigation will be described. First, attention will be paid to the identification of an appropriate model along the lines described in the preceding section. Next the ultimate model will extensively be discussed. As mentioned earlier, the sample of 1,934 observations was randomly halved and a logit model, including the variables discussed in an earlier section was estimated. The explanatory variables (see the Appendix for more detailed information) are:

AGE: age
EDUC: education level
FAMS: family status
WEXP: work experience
UNEM: regional unemployment
INCO: regional income

It should be noted that the interactions age-work experience, family status-age, family status-education and work experience-education were included *a priori* into the model. Furthermore, several other interactions between candidate variables were tried out. However, for none of the interactions a substantial decrease in χ^2 value was obtained. Therefore the model given in Table 1 (column a) was ultimately chosen. As mentioned in an earlier section, the identified model was re-estimated on the basis of the second, unused data set. The results are given in table 1 (column b). From both tables we may conclude that there are no substantial differences between both sets of estimates. The same result was obtained by means of covariance analysis.

The identified model was estimated, therefore, on the basis of the total data set. The maximum likelihood estimates for each level of all main effects of the ultimate model are summarized in Table 2. Before paying attention to the results we recall that the parameters estimated are differences on the log-odds scale. They are expressed versus the omitted levels *AGE*(1), *EDUC*(1), *WEXP*(1), *FAMS*(1), *UNEM*(1) and *INCO*(1). For this group, which is used as a reference group, the entry probability is represented by the grand mean. It should be noted that odds and probabilities are related as:

$$\text{probability} = \frac{\text{odds}}{1 + \text{odds}} \quad (\text{GOODMAN, 1978}) \qquad (14)$$

In the following, the results will be discussed in terms of probabilities.

Before turning to the individual variables we will pay attention to the overall fit by means of the Pearson χ^2. This test-statistic is equal to 1,919 with degrees of freedom equal to 1,916. Using the approximation formula:

$$U_\alpha = \tfrac{1}{2}(Z_\alpha + \sqrt{2n-1})^2 \qquad (15)$$

where Z_α is the αth quantile of the cumulative normal distribution, the probability level defined as the

Table 1. Maximum likelihood estimates with the ultimate logit model for both halves of the data set[1]

	a. Estimates for the first half	b. Estimates for the second half
Grand mean	−0·547 (0·368)	−0·104 (0·353)
AGE(2)	−0·364 (0·267)	−0·588 (0·254)
AGE(3)	−0·701 (0·289)	−0·838 (0·278)
AGE(4)	−1·061 (0·324)	−1·356 (0·323)
AGE(5)	−2·342 (0·575)	−1·737 (0·480)
AGE(6)	−7·857 (4·007)	−3·029 (0·605)
EDUC(2)	0·403 (0·301)	0·440 (0·267)
EDUC(3)	0·523 (0·186)	0·151 (0·195)
EDUC(4)	0·754 (0·356)	0·130 (0·380)
EDUC(5)	0·942 (0·217)	0·963 (0·207)
EDUC(6)	1·164 (0·332)	0·973 (0·309)
EDUC(7)	0·168 (0·491)	0·493 (0·457)
WEXP(2)	0·761 (0·206)	0·544 (0·195)
FAMS(2)	0·551 (0·242)	0·177 (0·235)
FAMS(3)	1·021 (0·313)	0·690 (0·298)
FAMS(4)	0·105 (0·258)	0·049 (0·254)
UNEM(2)	−0·282 (0·164)	−0·473 (0·160)
INCO(2)	0·256 (0·167)	0·276 (0·162)
Number of observations	967	967
Degrees of freedom	949	949
Scaled deviance	1132	1122
Pearson χ^2	933	961

Note: 1. Estimated coefficients are differences in the log-odds scale. Standard errors are given in brackets.

Table 2. Maximum likelihood estimates with the ultimate logit model for the total data set[1]

Grand mean	−0·197 (0·253)
AGE(2)	−0·493 (0·183)
AGE(3)	−0·777 (0·199)
AGE(4)	−1·212 (0·227)
AGE(5)	−2·019 (0·364)
AGE(6)	−3·559 (0·551)
EDUC(2)	0·435 (0·198)
EDUC(3)	0·347 (0·133)
EDUC(4)	0·470 (0·255)
EDUC(5)	0·952 (0·149)
EDUC(6)	1·054 (0·225)
EDUC(7)	0·325 (0·335)
WEXP(2)	0·640 (0·140)
FAMS(2)	0·361 (0·167)
FAMS(3)	0·854 (0·214)
FAMS(4)	0·016 (0·179)
UNEM(2)	−0·384 (0·114)
INCO(2)	0·257 (0·116)
Number of observations	1934
Degrees of freedom	1916
Scaled deviance	2307
Pearson χ^2	1919

Note: 1. Estimated coefficients are differences in the log-odds scale. Standard errors are given in brackets.

probability of obtaining a sample value, i.e. Pearson χ^2 value, as extreme as the one actually obtained if the postulated model is true, is found to be 50%.

Let us now turn to the individual variables. The estimates for *EDUC*, which are in conformity with our theoretical expectations, show that in general people with a higher and occupation-specific education have relatively high entry probabilities. There are hardly differences in the entry probabilities among those with general education at the lower-medium and upper-medium level and occupation specific education at the lower medium level. The highest entry chances are for those with occupation specific education at an upper medium or high level. Most striking are the very low chances for academics. The estimated parameter hardly differs from the lowest level of education. A possible explanation for this unexpected result may be that in general the unemployment rate is very low for this group. The unemployed are concentrated in relatively few professions with very bad labour market opportunities. Consequently the duration of unemployment is very long (DBAG, 1981; and BRONNEMAN-HELMERS, 1984).

The estimated parameters for *AGE* show increasing negative values for the older age groups, which means that the elderly have a relatively low chance of finding a job. The possible deviation from this pattern mentioned in an earlier section i.e. low chances of getting a job for young people, does not occur. When this had been the case the sign of the parameters estimated for *AGE* (2), and maybe also for *AGE* (3), would have been positive.

The estimated parameter for *WEXP* indicates that work experience is not an advantage. People looking for a job for the first time have higher probabilities of getting a job. As mentioned above a possible explanation for this is that their reservation wages are relatively low.

The estimates of the parameters representing the main effect for family status indicate that singles have the lowest entry probabilities. The estimated parameter for children living with their parents does not significantly differ from that of singles. Family heads and spouses have the highest probabilities of getting a job, maybe because of the reliability characteristic. The entry chances for spouses exceed that of family heads, probably because of the higher turnover rates for jobs filled by spouses. Apparently singles and children possess more undesirable employee characteristics than family heads and spouses do.

The estimates for the regional characteristics *UNEM* and *INCO* confirm our expectations. When the situation on the labour market is characterized as better than the national situation the probability of getting a job is higher. The results are in confirmity with results found by other researchers.

The variable *INCO* shows the expected sign: higher average regional income levels lead to higher probabilities of being successful in job finding. The result for *INCO* cannot be compared with other empirical work, because as far as we know this variable has never been used in these kinds of models. It is obvious that the influence of the regional characteristics on the probability of obtaining a job are much smaller than the influences of the personal characteristics. As mentioned earlier, these findings are upheld by other empirical work. Finally, we remark that the results for the regional characteristics hardly change when they are included as continuous variables instead of as dichotomized variables.

As a concluding remark it may be stated that different sets of personal and regional characteristics may imply widely different entry probabilities. Neglecting sample errors, the most extreme results are:

1. The smallest probability of getting a job ($p = 0{\cdot}02$, log-odds $= -3{\cdot}9$) is an individual who is single, belongs to the oldest age group, has the lowest educational level, has been employed before and lives in a region with above average unemployment and with an average income below the national level;
2. The highest probability of getting a job ($p = 0{\cdot}93$, log-odds $= 2{\cdot}6$) is a spouse who is young, highly occupation specific educated, without work experi-

ence and lives in a region with a relatively low unemployment rate and an average income level above the national level.

CONCLUSIONS

The main finding of the preceding section is that the entry probabilities differ widely for unemployed with different labour market characteristics. This finding has some important consequences for labour market policy which will be described.

The Social Economic Council (SER), the most important advisory council of the Dutch government in social and economic affairs, formulated the *stimulation of full and full-fledged employment* as one of the five major goals of Dutch social-economic policy (SER, 1956). This goal is strived after by two kinds of instruments. Firstly, *employment policy* of the Ministry of Economic Affairs which tries to stimulate the demand for labour in order to reduce Keynesian unemployment. Secondly, *labour market policy* of the Ministry of Social Affairs and Employment, which focuses mainly on the supply of labour. It is obvious that the results of the present study relate especially to the design of this kind of policy. The labour market policy tries to improve the efficiency and equity of the matching process on regional, sectoral and occupational labour (sub)markets.

Three types of instruments are usually distinguished in the framework of labour market policy (see, for instance, VAN DER VEGT *et al.*, 1983, and SER, 1980). Firstly, schooling measures, which intend to improve or change skills, are used to reduce qualitative discrepancies in the labour markets. Secondly, temporary wage subsidies are applied to stimulate the entry of unemployed with low entry probabilities. Thirdly, additional employment is created in governmental institutions and non-profit organizations for those who cannot be favoured by the two types of instruments mentioned previously. Each year these three instruments concern 60,000 or 70,000 persons and the costs amount to about Dfl. 600 millions.

In a recent study by VAN DER VEGT *et al.*, 1983, the effects of instruments of labour market policy are evaluated in groups with different unemployment rates. The analysis is restricted to the Dutch metropolitan areas. One of the conclusions of this study is that, although some instruments are rather successful, unemployed with low probabilities of getting a job hardly benefit from these measures. This is partly due to the fact that unemployed with low entry probabilities are selected on the basis of invalid labour market characteristics. The results of the present study make clear what groups of unemployed have relatively low chances of finding a job and thus on which groups the labour market policy should focus, i.e. the elderly, singles and children, lower educated and those who are now unemployed after having been employed. Furthermore, it shows the influence of regional factors on entry probabilities which advocates a regionally differentiated labour market policy.

We want to end this section and this paper with a word of caution. The present analysis is based on 1979 data only. No information is available about the stability over time of the patterns in entry probabilities described here. Furthermore, since 1979 unemployment has increased sharply. This means that the entry probabilities in general have gone down. However, it may also have led to changes in the ratios of entry probabilities for different groups of unemployed. Further evidence on these presumptions can be obtained when the Labour Force Surveys 1981 and 1983 become available.

NOTE

1. This research project is supported by a grant for region relevant research from the State University of Groningen.

APPENDIX: the data base

Every two years the Census Bureau of the European Community organizes a sample survey among the potential labour force in the member countries. The data we used are micro data from the Arbeidskrachtentelling (Labour Force Survey) 1979 for The Netherlands. The total sample for The Netherlands contains 142,330 cases with individual information. Only persons born before 1965 are questioned. Besides questions about the present situation (i.e. March until May 1979) some information was also gathered about the situation exactly one year before.

For our analysis we selected those individuals, said to be employed in 1978, and who were still in the labour force in 1979. Persons who were students or in military service in 1978 or 1979 have been excluded. For our purpose 1,934 individuals were suited. To each case a prior weight is added to correct for selective nonresponse and for regional differences in the sample fraction. For details about the data see *CBS Arbeidskrachtentelling 1979* (1982), Staatsuitgeverij, The Hague.

The variables

Personal characteristics measured at the individual level.

*ES*79: employment status in 1979.

*ES*79(0): unemployed in 1979 (1,175 cases, 60·8%)

*ES*79(1): employed in 1979 (758 cases, 39·2%)

Note that of the unemployed in 1979 1,013 persons (86.6%) had been unemployed at least one year. The remaining individuals were again unemployed after having been employed during a short time between the points of time of observation.

AGE: age at 1 January 1979

AGE(1):14–19 years	(235 cases, 12·2%; 133 cases, 56·7%)
AGE(2):20–24 years	(423 cases, 21·9%; 201 cases, 47·6%)
AGE(3):25–39 years	(740 cases, 38·3%; 301 cases, 40·7%)
AGE(4):40–54 years	(356 cases, 18·4%; 105 cases, 29·6%)

AGE(5):55–59 years (86 cases, 4·5%; 13 cases, 15·2%)
AGE(6):60 years or more (93 cases, 4·8%; 4 cases, 4·2%)

FAMS: family status
FAMS(1): a single person (255 cases, 13·2%; 88 cases, 34·3%)
FAMS(2): a family head (873 cases, 45·1%; 394 cases, 33·7%)
FAMS(3): a spouse (181 cases, 9·4%; 100 cases, 55·4%)
FAMS(4): a child (625 cases, 32·3%; 277 cases, 44·3%)

EDUC: education
EDUC(1): low or unknown (847 cases, 43·8%; 257 cases, 30·3%)
EDUC(2): lower medium; gen. (139 cases, 7·2%; 62 cases, 44·8%)
EDUC(3): lower medium; o.s. (408 cases, 21·1%; 161 cases, 39·5%)
EDUC(4): upper medium; gen. (80 cases, 4·1%; 36 cases, 44·6%)
EDUC(5): upper medium; o.s. (298 cases, 15·4%; 155 cases, 52·2%)
EDUC(6): high; o.s. (115 cases, 5·9%; 68 cases, 59·4%)
EDUC(7): high; scientific (48 cases, 2·5%; 19 cases, 40·2%)

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.
Gen. = general education; o.s. = occupation specific education.

WEXP: work experience
WEXP(1): yes (1,528 cases, 79·0%; 531 cases, 34·8%)
WEXP(2): no (406 cases, 21·0%; 237 cases, 58·4%)
yes = unemployed after employment
no = looking for a job for the first time

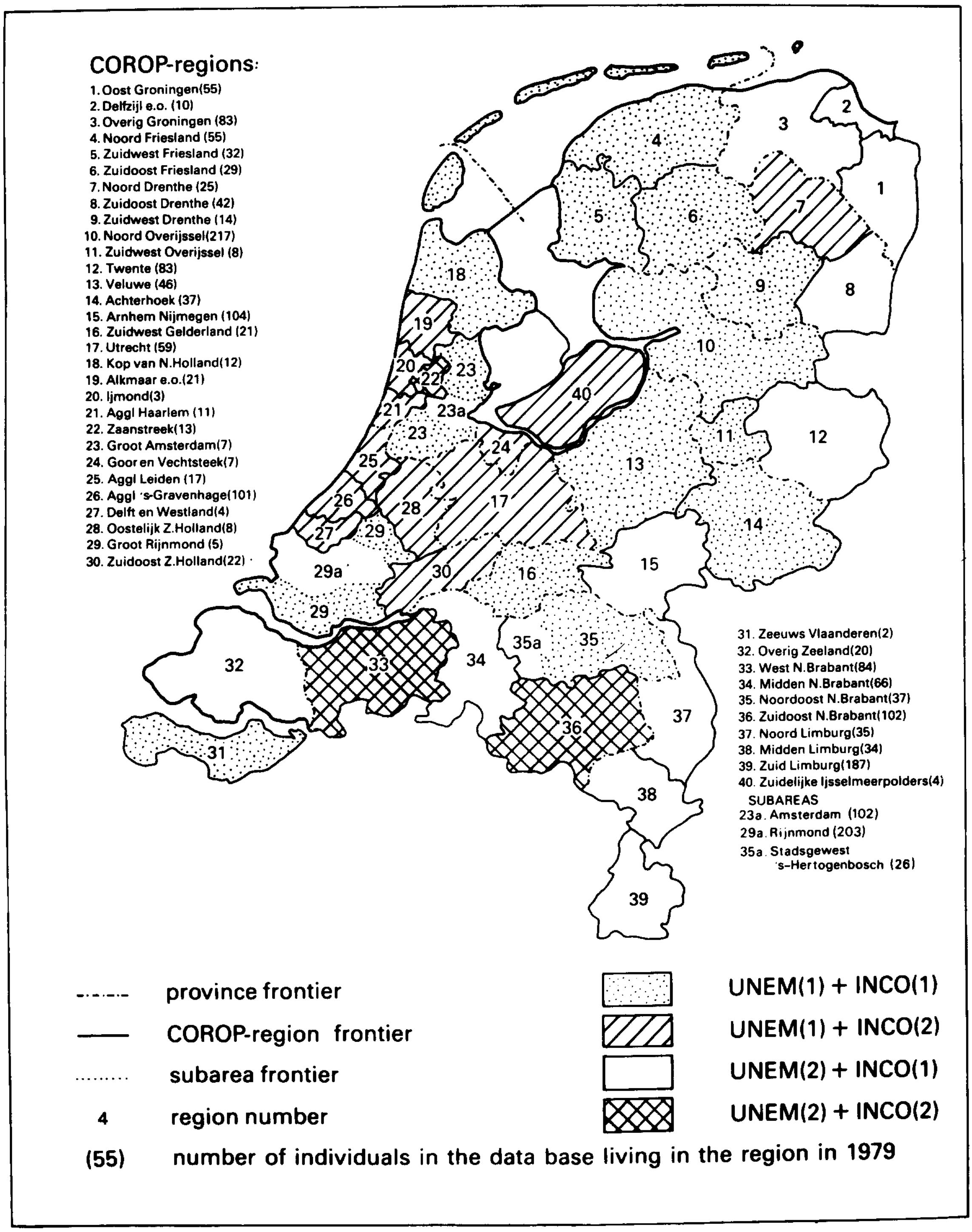

Fig. 1. Regional demarcation of The Netherlands in forty COROP-regions and three subareas

Note: Figures in brackets represent: (1) absolute number of cases in category; (2) % of total number; (5) absolute number of cases in category which were employed in 1979; (4) % employed in 1979 of total number of cases in category.

Regional characteristics are added to the scores of each individual on the basis of the region of residence in 1979. This implies that for migrants only their present location is taken into account. The characteristics of the region of origin may be important too, but we expect that the characteristics of the region of origin mainly influence the decision to move (push-factor), while the probability of getting a job is especially influenced by the characteristics of the present region of residence, which simultaneously acts as a pull-factor for migration. Furthermore, practical problems may arise by adding regional characteristics to persons with residence unknown or abroad in 1978. The scores of each region on the variables are shown in Fig. 1. This figure shows also the regional demarcation of The Netherlands we used and the spatial distribution of the individuals in the sample.

UNEM: regional unemployment in 1979

UNEM(1): regional unemployment rate equal or below the national level (632 cases, 32·7%; 286 cases, 45·2%)

UNEM(2): regional unemployment rate higher than the national level (1,302 cases, 67·3%; 473 cases, 36·3%)

Regional unemployment is computed on the basis of information on unemployment in the same survey as where the individual data are extracted from. However, for the calculation of the unemployment rates in the present case new entrants to the labour force in 1979 are not excluded anymore and also the 1979 information available on 'hidden unemployment' is taken into account for calculating the regional and national unemployment rates.

INCO: regional income in 1978.

INCO(1): average income in the region equal or below the national level (1,325 cases, 68·5%; 490 cases, 37·0%)

INCO(2): average income in the region higher than the national level (609 cases, 31·5%; 268 cases, 44·0%)

The source of this variable is the only one from another source as the sample we used, viz. from the Dutch Central Bureau of Statistics (*CBS Sociale Maandstatistiek* **31,** 42). Data for 1978 are used, because data for 1979 were not available.

REFERENCES

AMEMIYA T. (1981) Qualitative response models: a survey, *J. Econ. Lit.* **19,** 1,483–536.

BAKER R. J. and NELDER J. A. (1978) *The GLIM-system: release 3. Generalised lineair interactive modelling.* Numerical Algorithms Group, Oxford.

BRON J. A. H., VAN GAANS F. and NIEUWENHUIZEN R. D. J. (1983) *Integratie van werkloosheidsregelingen (Integration of Unemployment Regulations).* Onderzoeksreeks no. 7, Instituut voor onderzoek van overheidsuitgaven, The Hague.

BRONNEMAN-HELMERS R. (1984) *De arbeidsmarkt voor hoger opgeleiden (The Labour Market for Higher Educated)* Stukwerk nr. 15, Sociaal en Cultureel Planbureau, Rijswijk.

DBAG (1981) *Werkloze academici in het Noorden (Unemployed Academics in the North).* DBAG, Groningen.

FELMLEE D. H. (1982) Women's job mobility processes, *Am. Sociol. Rev.* **67,** 142–50.

FLOWERDEW R. (1983) Job information and migration, paper presented to workshop 'Information in Labour Markets' of the International Geographical Union, Groningen.

FOLMER H. (1984a) *Measurement of effects of regional economic policy—some methodological aspects.* Martinus Nijhoff, The Hague.

FOLMER H. (1984b) Meting van beleidseffecten met tijdreeksanalyse (Measurement of effects of policy by means of time series analysis), *Kwantitatieve Methoden* **13,** 93–108.

GAVER K. M. and GEISEL M. S. (1974) Discriminating among alternative models: Bayesian and non-Bayesian methods, in ZAREMBKA P. (Ed.) *Economic Theory and Mathematical Economics,* Academic Press, New York.

GEISER S. (1974) A predictive approach to the random effect model, *Biometrika* **61,** 101–7.

GEISER S. (1975) The predictive sample re-use method with applications, *J. Am. Statist. Assoc.* **70,** 320–8.

GOLDBERGER A. S. (1964) *Econometric Theory.* Wiley, New York.

GOODMAN L. A. (1978) *Analyzing Qualitative/Categorical data: Log-Linear Models and Latent Structure Analysis.* Addison-Wesley, London.

HERZOG H. W. JR and SCHLOTTMANN A. M. (1982) Labor force mobility in the US: the migration decision, pre- and post-move unemployment and implications for remigration, paper delivered at the ASA/Census conference on forecasting regional population change and its economic determinants and consequences, Airlie, Virginia.

IVA (1977) *Onderzoek onder werklozen*; deel 1. *moeilijkheden bij (her)intreding (Research among Unemployed,* part 1, *(Re-)Entrance Troubles).* IVA, Tilburg.

JOLL C., MCKENNA C., MCNABB R. and SHOREY J. (1983) *Developments in Labour Market Analysis.* George Allen and Unwin, Hemel Hempstead.

KENDALL M. G. and STUART A. (1967) *The Advanced Theory of Statistics,* second ed., vol. 2. Griffin, London.

KOOREMAN P. and RIDDER G. (1983) The effects of age and unemployment percentage on the duration of unemployment, *European Econ. Rev.* **20,** 45–57.

KISH L. and FRANKEL M. R. (1974) Inference from complex samples, *J. Roy. Statist. Soc. B* **36,** 1–37.

LANCASTER T. (1979) Econometric methods for the duration of unemployment, *Econometrica* **47,** 939–56.

LANCASTER T. and NICKELL S. (1980) The analysis of re-employment probabilities for the unemployed, *J. Roy. Statist. Soc. A* **143,** 141–65.

LEHMAN E. L. (1959) *Testing Statistical Hypothesis.* Wiley, New York.

LIPPMAN S. and MCCALL J. (1976) The economics of job search: a survey (part I and II), *Econ. Inquiry* **14,** 155–89 and 347–68.

LOVELL M. C. (1983) Data mining, *Rev. Econ. Statist.* **65,** 1–12.

MADDALA G. S. (1983) *Limited-Dependent and Qualitative Variables in Econometrics.* Econometric Society Monograph in Quantitative Economics, Cambridge University Press, Cambridge.

MADDISON A. and WILPSTRA B. S. (Eds.) (1982) *Unemployment: The European Perspective.* Croom Helm, London.

MADDISON A. and WILPSTRA B. S. (Eds.) (1983) *Unemployment: The Dutch Perspective.* SOZAWE, The Hague.

MCCARTHY P. J. (1976) The use of balanced half-sample replication in cross-validation studies, *J. Am. Statist. Ass.* **71,** 596–604.

MALINVAUD E. (1970) *Statistical Methods of Econometrics.* North-Holland, Amsterdam.

MIRON J. R. (1978) Job-Search Perspectives on Migration Behaviour, *Environ. Plann. A* **10,** 519–35.

MOSTELLER F. and TUKEY J. W. (1977) *Data Analysis and Regression: A Second Course in Statistics.* Addison-Wesley, Reading, Mass.

NELDER J. A. and WEDDERBURN R. W. M. (1972) Generalised linear models, *J. Roy. Statist. Soc. A* **135,** 370–84.

NICKELL S. (1979) Estimating the probability of leaving unemployment, *Econometrica* **47,** 1,249–64.

OECD (1979) *Facing the Future: Mastering the Probable and Managing the Predictable.* OECD, Paris.

PLATT C. A. (1982) Bootstrap stepwise regression, American Statistical Association 1982 proceedings of the Business and Economic Statistics section, pp. 586–9.

RAO R. (1965) *Linear Statistical Inference and its Applications.* Wiley, New York.

ROGERSON P. (1982) Spatial models of search, *Geogr. Analysis* **14,** 217–28.

ROGERSON P. and MACKINNON R. D. (1981) A geographical model of job search, migration and unemployment, *Pap. Reg. Sci. Ass.* **48,** 89–102.

SCHIFFEL D. and GOLDSTONE S. (1976) Employment, induced in-migration and labor market equilibrium, *Growth and Change* **7,** 33–40.

SER (1956) *Onderzoek naar de economische situatie in Nederland (Investigation of the Economic Situation in the Netherlands).* SER, The Hague.

SER (1980) *Interim-nota effectiviteit van Arbeidsvoorzieningsmaatregelen* (Interim Report Effectivity of Labour Market Policy Measures). SER, The Hague.

SIDDRÉ W. and THEEUWES J. (1977) De invloed van werkloosheidsuitkeringen op het werkloosheidspercentage in Nederland (The effect of the unemployment benefits on the unemployment rate in The Netherlands), working paper, Erasmus Universiteit, Rotterdam.

SILVERS A. (1977) Probabilistic income maximizing behaviour in regional migration. *Int. Reg. Sci. Rev.* **2,** 29–40.

SØRENSEN A. B. and TUMA N. B. (1981) Labour market structures and job mobility, in *Research in Social Stratification and Mobility: An Annual Compilation of Research* pp. 67–94. Jai Press, Greenwich.

SOZAWE (1982) *Rapportage arbeidsmarkt 1982 (Labour Market Report 1982).* SOZAWE, The Hague.

SOZAWE (1983a) *De arbeidsmarkt in juli 1983 (The Labour Market in July 1983).* SOZAWE, The Hague.

SOZAWE (1983b) *Rapportage arbeidsmarkt 1983 (Labour Market Report 1983).* SOZAWE, The Hague.

STONE M. (1974) Cross-validitory choice and assessment of statistical predictions, *J. Roy. Statist. Ass. B* **36,** 111–47.

VAN BEKKUM R. (1983) *Jeugdwerkloosheid (Youth-Unemployment).* SMO, The Hague.

VAN DIJK J. and BARTELS C. P. A. (1982) Internal labour migration and its effects on regional labour markets in The Netherlands, paper presented at the 22nd European Congress of the Regional Science Association, Groningen 24–27 August 1982. IEO/RUG Research memorandum no. 109.

VAN DER VEEN A. (1982) Het aanbod van arbeid en de spanning op de arbeidsmarkt (Labour Supply and the Situation on the Labour Market), *Maandschr. Econ.* **46,** 3–16.

VAN DER VEGT C., LUTZ C. H. M., VAN PARIDON W. C. G. M. and RIDDER G. (1983) *De toepassing van arbeidsvoorzieningsmaatregelen op de grootstedelijke arbeidsmarkt (The Application of Labour Market Policy Measures on the Metropolitan Labour Market).* SEO, Amsterdam.

WIEBRENS C. J. (1981) *Inkomen en rondkomen: de financiële positie van huishoudens in Nederland (The Outcome of Income: The Financial Situation of Families in The Nederlands).* Staatsuitgeverij. Serie Sociale en Culturele Studies no. 2 van het Sociaal en Cultureel Planbureau, Staatsuitgeverij, The Hague.