THE DETERMINANTS OF UNEMPLOYMENT AND JOBSEARCH DURATION IN THE NETHERLANDS

BY

C. KERCKHOFFS, C. DE NEUBOURG AND F. PALM*

1 INTRODUCTION

In The Netherlands, unemployment increased most severely in the period 1980–1983. The years thereafter were a period of recovery. Since 1985, however, more than half of the unemployed persons have been jobless for more than one year, while more than three-quarters spent over six months out of work. The aim of this paper is to contribute to a better understanding of the factors which affect unemployment duration by investigating the dependence of the exit rates for the unemployed on duration and on personal and labour market characteristics, using a proportional hazard model.

Our study investigates the shape of the proportional hazard model using a semi-parametric specification. Correction for unobserved heterogeneity is based on parametric (gamma) and semi-parametric techniques. For both sexes, the hazard rate is found to be nonmonotonic during the first year of unemployment: it first increases during five months, declines sharply thereafter, and stays virtually constant after a year. Commonly used parametric specifications for the baseline hazard rate prove to be too restrictive. Our results show the dangers of incorrect inference that can result from choosing inappropriate functional forms for the duration dependence. The impact of the regressors on the exit probability gives insight into which characteristics make people prone to long-term unemployment. These results generally confirm previous findings. Interestingly, no systematic relation could be found between education level and (unemployed) jobsearch behaviour. Moreover, the results indicate that there is an interesting but puzzling relation between the population size of the place of residence and unemployment duration. The estimates of the coefficients of the regressors are found to be almost insensitive to the specification adopted for the baseline hazard rate or the heterogeneity correction. The analysis is based on data for 1987 which allows us to

* University of Limburg, Faculty of Economics and Business Administration, P.O. Box 616, 6200 MD Maastricht, The Netherlands. We like to thank Simon Kuipers, Theo van de Klundert, Wiji Narendranathan, Mark Stewart and an anonymous referee for helpful comments on an earlier draft of this paper.

De Economist **142**: 21–42, 1994.

analyze how long-term unemployment and jobsearch, incurred in the first half of the eighties, are affected by a recovering labour market. We monitored the exit process during two months observing the complete population of (unemployed) jobsearchers (± 840,000 persons), providing us with data on completed spells of unemployment as well as uncompleted spells. Samples of 20,000 persons for each sex are used in the empirical analysis.

The paper is organized as follows. Section two describes the data, while section three presents the basic model and the semi-parametric specification used in this study. In section four, the results for the shape of the hazard rate are discussed. Section five analyzes the parameters of the regressors for men only. A summary section concludes the paper.

2 THE DATA

Our data set consists of all jobseekers registered as being available for work at the Dutch labour exchange offices on March 31, 1987. Registration is mandatory for persons younger than 57.5 years who receive unemployment benefit. The total number registered at March 31, 1987 was 864,000, of whom we observe 837,880 persons who were available according to the registration (510,792 men and 327,088 women). Not all of them can be considered as unemployed in the usual sense.[1] From April 1 until May 28 1987, the exit process of this population has been monitored.[2] During this period, 86,089 men and 43,483 women left the register (uncensored cases). The remaining persons were still registered on May 29 (censored cases). We observed the exact exit date of the uncensored cases, but it is unknown whether an individual left the register because he or she found a (new) job ((re-)employment) or because he or she withdrew from the labour force *e.g.* for reasons related to disability, retirement or re-entry into full-time education.[3] The proportion of persons leaving the unemployment register in order

1 The relation between the concept of unemployment and registration data is always troublesome (see de Neubourg (1988), chapter 2). Adopting a very strict interpretation of the ILO recommendations, there were on average 622,000 people unemployed in The Netherlands in 1987 (Min. van SOZA (1988), p. 39). We estimate that about 430,000 of them are present in our data set. The remaining unemployed are unobserved in this study, because they are not registered for whatever reason. On the other hand, our data contain a number of persons who do not comply with even the broadest unemployment definition. First of all, some 89,600 persons actually have a job according to the information available at the labour exchange, though still declaring themselves available for a new job (and hence present in our dataset). We are unable to remove these persons. Secondly, there is the problem of 'registration pollution' ('bestandsvervuiling'). Flaws in the registration exist: a number of persons who are unemployed according to the register are, in fact, not readily available for the market or even have a job which has not been reported to the labour exchange.

2 This period has been chosen because we expect the exit behaviour of the unemployed to be least influenced by seasonal effects in these months.

3 Our data set is corrected for those who left the register for administrative reasons (*e.g.* failure to re-apply); they are treated as censored cases. In order to avoid misunderstanding among the readers who are familiar with this type of data, it should be stressed that renewal of registration after

to quit the labour force is estimated between 1 and 29 per cent depending on the source used (Theeuwes, Kerkhofs and Lindeboom, 1990; de Neubourg, 1990; Van den Berg, 1990). We may say that the estimates produced by this study can be interpreted as a proxy for re-employment probabilities. However, the presence of withdrawal may especially affect the estimates of the age-effect and the duration dependence parameters.

Within our original data set, some very detailed categorical variables are available. We recoded these in order to get a manageable number of categories, while removing 3,612 persons either because they were over 64 years old or because of inconsistencies in the data. For all remaining persons in the data set the following information is available:

1. A dummy variable indicating censored *vs.* uncensored cases;
2. Length of the elapsed unemployment spell, measured in days as the difference between March 31, 1987, and the date at which the individual became available for a (new) job[4];
3. For uncensored cases, the length of the residual unemployment spell, measured as the number of days from March 31, 1987, to the exit date[5];
4. Age at the start of the unemployment spell;
5. Sex;
6. Population size of place of residence;
7. Nationality;
8. Education;
9. Occupation;
10. Desired working time;
11. Previous employment experience;
12. Change of profession indicating whether the main profession as registered at the labour office is the same as a person's profession exercised in the last job;
13. Time series of monthly unemployment rates according to province of residence, 1962 – May 1987.

Wage and benefit information is not available and can therefore not be accounted for. All variables are of the categorical type except 2, 3, 4 and 13. Variable 13 is the only one that was not present in the original dataset. It is a general, gender-

expiration is *not* counted as a new unemployment period, but is observed as a prolongation of the first period.

4 The date of availability, which we use to indicate the start of the spell, coincides with the date of registration for most persons. Some, however, could not accept a (new) job until some time after their registration, because they were still at a former job or outside the labour force (*e.g.* at school) at the time they registered.

5 The date of exit in our study is the date at which individuals are removed from the registration. For a number of technical and administrative reasons this date might not be entirely correct, but differences are small.

neutral labour market tension indicator, constructed on the basis of information in successive editions of the Monthly Bulletin of Socio-economic Statistics (published by CBS Voorburg and based on registration data). We were not able to construct similar series for the preferable *UV*-ratio. Finally, for estimation purposes random samples of 20,000 persons were drawn for both sexes separately.

3 REDUCED-FORM ESTIMATION OF DURATION MODELS

Unemployment duration is commonly modelled using the exit or hazard rate, *i.e.* the conditional probability density function (pdf) of leaving unemployment, given that one has been unemployed for a specific period. Let $f(\mathbf{x}, t)$ be the (unconditional) pdf and $F(\mathbf{x}, t)$ the cumulative density function of the random variable unemployment duration t, for an individual with characteristics $\mathbf{x}$. The hazard function $\Theta(\mathbf{x}, t)$ can be expressed as

$$\Theta(\mathbf{x}, t) = f(\mathbf{x}, t)/\{1 - F(\mathbf{x}, t)\}\,. \tag{1}$$

The denominator of (1) is the survival function $S(\mathbf{x}, t) = exp\{-I(\mathbf{x}, t)\}$ with $I(\mathbf{x}, t) = \int_0^t \Theta(\mathbf{x}, s)\,ds$ being the so-called integrated hazard.

Search theory views the hazard rate as the product of the probability (density) of receiving a job offer in a given period, and the probability (density) that such an offer will be accepted by the unemployed. The offer probability or arrival rate is a function of the overall tension on the labour market, and of personal characteristics that are regarded by the employers as indicators of a jobseeker's productivity. The acceptance probability depends on the individual reservation wage (see McKenna (1985) for a detailed survey). Several factors may cause the hazard rate to change during an unemployment spell, implying duration dependence. Duration dependence is positive if the hazard rises with duration ($\delta\Theta/\delta t > 0$) and negative if it falls ($\delta\Theta/\delta t < 0$).

Although hazard functions link up with economic theory and with intuition (Kiefer (1988) pp. 648–649 offers an elaboration of this point), neither theory nor intuition provide much guidance in choosing functional forms. Proportional hazard models are most commonly used in this context:

$$\Theta^*(\mathbf{x}, u, t) = u * \Theta(\mathbf{x}, t)\,, \quad u > 0\,, \tag{2}$$

where u is a random variable with pdf $g(u)$ which accounts for unobserved heterogeneity. The hazard function $\Theta(\mathbf{x}, t)$ factorizes into the effects of the regressors $\mathbf{x}$, given by $exp(\mathbf{x}\beta + \gamma)$ to ensure nonnegativity, and that of duration t, $\alpha(t)$. Along the lines of Meyer (1988) (see also Han and Hausman (1990), Jensen (1990) and Stewart and Narendranathan (1990)), we use a *semi-parametric* piecewise constant specification $\alpha(t)$ for the baseline hazard. For all individuals, the time profile of the hazard function is independent of $\mathbf{x}$, and completely determined by the baseline hazard $\alpha(t)$. We can interpret $\boldsymbol{\beta}$ as a vector of (semi-)elasticities with respect to $\mathbf{x}$. The flexibility of the specification can be increased by estimating proportional hazard functions for specific subsets of the population. For this reason, we estimated our models separately for both sexes. Unobserved heterogeneity u occurs when not all relevant variables are observed and therefore cannot be incorporated in $\mathbf{x}$. Amongst other things, it leads to a bias towards negative duration dependence (see, *e.g.*, Heckman and Singer (1984a)). Given a specification for the pdf $g(u)$ of unobserved heterogeneity, we can construct a survival function

$$S(\mathbf{x}, t) = \int_0^{\infty} g(u)\, exp\left\{- \int_0^t \Theta^*(\mathbf{x}, u, s)\, \mathrm{d}s\right\} \mathrm{d}u\,. \tag{3}$$

By differentiating (3), an expression can be found for $f(\mathbf{x}, t)$. The expressions for f and S are used in the likelihood function. The implementation of the estimation procedure is described in Appendix I.

4 ESTIMATION OF THE BASELINE HAZARD AND HETEROGENEITY DISTRIBUTION

In this section, we present the results for both sexes using the semi-parametric baseline hazard rate without heterogeneity correction, with gamma correction and with discrete correction, respectively. We also briefly compare the results with those for three parametric specifications, the Weibull, Gompertz and Nickell (1979a, b) baseline hazard specifications which were found to be too restrictive and therefore are not reported here. Notice that the Nickell specification which is obtained by including a quadratic term in t in the Gompertz form, can account for nonmonotonic duration dependence. Estimates for the Gompertz gamma specification can be found in section 5 where the results for the regressor parameters are discussed.

4.1 *Semi-parametric Baseline Hazard Rates*

The specification presented in section 3 amounts to a piecewise constant hazard rate. Ideally, one would choose the time unit of observation as the common step length. In that case, the approach is closely related to Cox' partial likelihood technique, which is completely non-parametric with regard to the baseline hazard (compare Han and Hausman (1990)). However, the sample information is not sufficiently rich to estimate a baseline parameter for each day (which is our time unit of observation). Instead, we defined 22 duration intervals, gradually broadening them as the number of observations declines (see the duration distributions in Appendix II). After six years, the hazard is assumed to be constant.[6]

6 Artificial right-censoring of all unemployment durations beyond 6 years without estimating any additional baseline parameters would be another option: compare Meyer (1988).

TABLE 1 – MAIN ESTIMATION RESULTS FOR SEMI-PARAMETRIC BASELINE HAZARD RATE

Heterogeneity correction	Males			Females		
	no	gamma	discrete	no	gamma	discrete
σ^2		0.289 (5.63)			0.198 (1.89)	
J			3			3
N	59	60	63	61	62	65
log-lik.	– 21324.5	– 21300.5	– 21295.7	– 18462.5	– 18460.2	– 18459.4

t-ratios in parentheses
σ^2 variance gamma distribution
J number of mass points discrete distribution
N number of independent parameters of the model

In Table 1, the results of applying this semi-parametric baseline hazard without heterogeneity correction, with gamma correction[7] and with discrete correction are reported.

Even without heterogeneity correction, for the semi-parametric specifications for both sexes we obtained a much higher log-likelihood value than for any of the parametric specifications mentioned above. Introducing the gamma correction leads to further improvement, though σ^2 is only marginally significant for females. Though the parametric baseline specifications are not strictly nested in the semi-parametric specification, the latter is expected to be sufficiently flexible to closely approximate any parametric baseline specification. When we compare the differences in log-likelihood values with the appropriate χ^2 cut-off points, even for $\alpha = 0.5\%$, all parametric baseline rates are clearly rejected in favour of the semi-parametric baseline hazard with gamma correction. For the parametric specifications, the highest log-likelihood value was – 21400.3 for males and – 18491.0 for females. Finally, for the regressor parameters, the adoption of a parametric *versus* a semi-parametric baseline is found to be almost immaterial. Adopting a discrete heterogeneity correction does not yield any real further improvement upon the semi-parametric baseline specification with gamma correction. The baseline shape as well as the regressor parameters remain very similar, and the log-likelihood values increase only moderately. It seems that the gamma heterogeneity correction is reasonably well in accordance with the sample information.

It is obviously impossible to characterize the duration dependence for the semi-parametric baseline hazard using some convenient single measure. Instead,

7 Meyer (1988) also uses a gamma correction, as do Han and Hausman (1990) (but only for their single risk models), Stewart and Narendranathan (1990) use a normal distribution, whereas Jensen (1990) does not correct for unobserved heterogeneity at all.

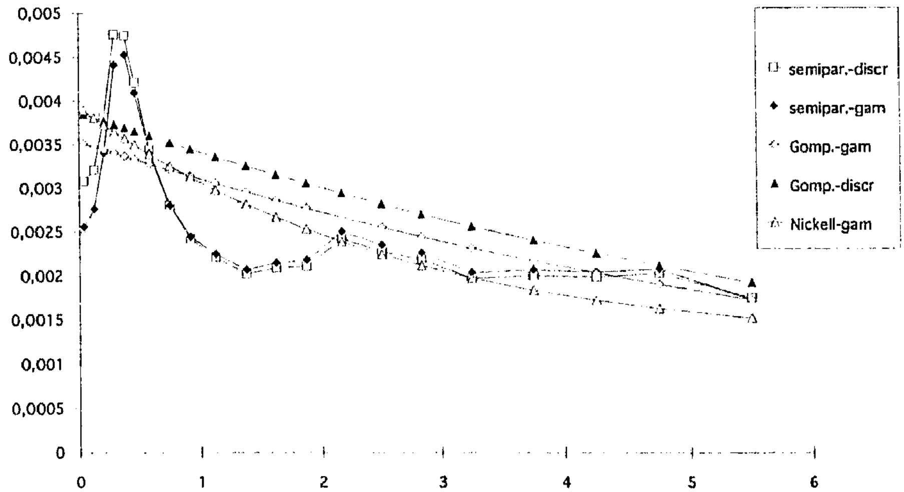

Figure 1 – Hazard time path for the average man

Figures 1 and 2 show (for males and females, respectively) the hazard time path for the five most interesting heterogeneity-corrected models, that is the Gompertz and semi-parametric baseline models, both with gamma and discrete heterogeneity correction, and the Nickell model with gamma correction. In constructing these graphs, we have put all regressors at their sample averages[8] to get maximum comparability of the different functional forms. Hence the figures represent hazard time paths for average men and women. Furthermore, we smoothed the semi-parametric baseline picture by computing three-interval moving averages. The y-axis represents the probability per day of an average man or woman to leave unemployment. Duration expressed in number of years is given on the x-axis. The choice of the day as a basis explains why the hazard rates are low.

For both sexes, the semi-parametric baseline results show a clearcut non-monotonicity. During the first four months of unemployment, the hazard rises quickly and then stabilizes for a short period. From the sixth month onwards until the end of the first year, however, it declines fairly fast. A period of considerable positive duration dependence is hence followed by a period of negative dependence. After one year of unemployment, there is no longer any clear tendency for the hazard to change as duration proceeds. If anything, it declines very slowly (this tendency is most visible for women). For men, the more or less stable hazard value after the twelfth month is *below* the hazard rate during the first month, for women it is *above* the initial hazard. Hence, for women, the tendency towards negative

8 For age, we used the median instead of the average values for both sexes (25 and 29 years for females and males, respectively). For unemployment, we used the average nationwide Dutch unemployment rate in April–May 1987 (14.2%).

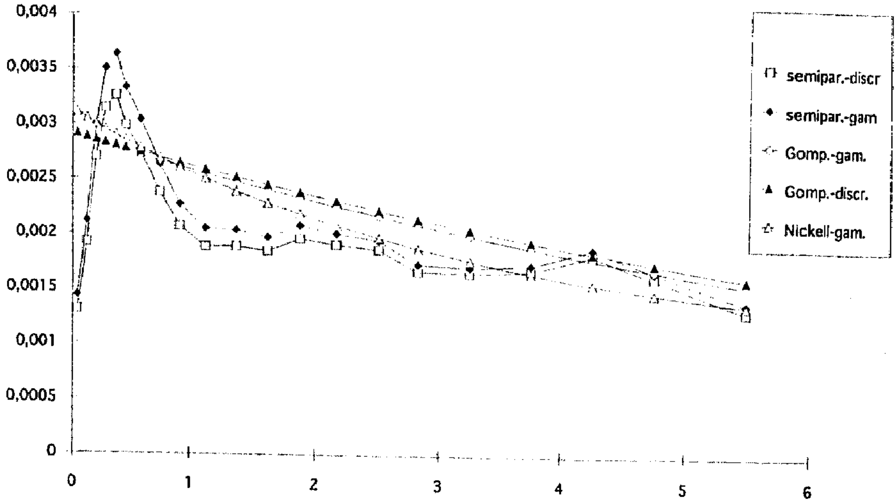

Figure 2 – Hazard time path for the average woman

duration dependence after the fifth month seems a bit smaller than for men. This is consistent with the picture emerging from the Gompertz and Nickell models. Finally, both sexes show two small hazard peaks at later dates, around month 24 and month 54.

The nonmonotonic picture offered by the semi-parametric baseline results links up with the duration distribution observed in the dataset (see Appendix II). *All* the parametric baselines are found to be misspecified. The graphs indicate clearly, how the early peak in the nonmonotonic semi-parametric baseline translates into monotonically negative duration dependence when the Gompertz and Nickell forms are imposed upon the data. Not even this Nickell form would have been able to yield the specific nonmonotonicity observed in the graph. Though it would have been able to capture the parabolic pattern during the first year correctly, for higher durations it would have implied a continuously declining hazard, which is inconsistent with the data (compare Figure 1b in Nickell (1979b)). The Weibull specification (not shown in the graphs) yields positive duration dependence, so it essentially follows the rising hazard during the first half year.

4.2 *Interpreting the Time Shape of the Hazard Rate*

How can we relate these results to economic theory? For both sexes, it seems that frictional unemployment and information effects are important at the beginning of an unemployment spell, with the hazard starting low and rising very rapidly within a few months. It is likely that people remain unemployed during the first few months because it simply takes time to apply for a job and get hired. Moreover, it is plausible that jobseekers get better informed about their labour market prospects as unemployment lengthens and therefore adjust their reservation wage during the first months: this leads to an increasing hazard. After the fifth month, the effects towards negative duration dependence (declining attractiveness for employers, lower search intensity, changing appreciation of leisure time) start to dominate, and after about one year, these effects are more or less exhausted.

These considerations are related to re-employment only. However, withdrawal from the labour market is important in our dataset and influences duration dependence. Ideally, one would estimate a competing risk model, with 're-employment' and 'withdrawal' as the distinct risks, but for reason of lack of data, we had to use a single risk model. The effects of the explanatory variables and duration dependence may differ considerably for these two risks. For example, it is a well-documented feature of the Dutch labour market that withdrawal becomes relatively more important as an unemployment spell proceeds.[9] In our single risk model, however, the different features of both underlying risks are lumped together. It is therefore very likely that the pure re-employment probability declines faster than indicated by our figures, and will probably decline beyond the twelve month point mentioned above. Finally, the fact that for females the negative duration dependence after the fifth month seems a bit smaller than for males may be related to the fact that for women, withdrawal is more important than for men.

One problem remains. In The Netherlands, when a person finds a job or leaves the register for whatever further reason, he/she can either actively terminate his/her registration or simply fail to renew it (however, note 3 still applies; this means that the data are corrected for administrative failure of renewal). This last option is chosen quite commonly. For most persons, registration must be renewed after every three months. Persons who do not renew their registration are then removed approximately two weeks later, that is somewhere in the fourth month of registration (or in the seventh, the tenth *etc.*). This registration effect may be responsible for part of the nonmonotonicity in the figures above. In fact, for both sexes the hazard reaches its maximum precisely in the fourth month. Hence this high value will in part reflect people who actually found a job in one of the first three months. However, the high hazard values in the fifth and sixth months cannot possibly be explained by this registration effect, so despite this effect nonmonotonicity must be present in the exit process.

Our results are at variance with conclusions based on parametric specifications for the baseline hazard rate. In some instances it is difficult to compare our results for the baseline hazard with the existing studies for The Netherlands, because these studies use either aggregate data (Kooreman and Ridder (1983)) or individual data of the first half of the eighties only (Van Opstal and Theeuwes (1986), Groot and Ter Huurne (1988, 1989), Van Opstal and Van der Pol (1990), Groot (1990), Theeuwes, Kerkhofs and Lindeboom (1990), Gorter and Hoogteijling (1990) and

9 See Min. van SOZA (1986), p. 28. In 1985, among those unemployed for less than 1 year, 73% of the exits was because of a new job. Among those unemployed for more than 4 years, this figure was only 45%.

and Hoogteijling (1990) and Lindeboom and Theeuwes (1991)). The analyses by Groot and Ter Huurne (1988) and Van Opstal and Theeuwes (1986) are limited to specific age-groups, while others focused on the impact of unemployment benefits and the duration of benefit entitlement upon search duration (Lindeboom and Theeuwes (1991) and Van den Berg (1990)). For the purpose of comparison, Van Opstal and Van der Pol (1990) are the most interesting. These authors also find a nonmonotonic pattern using a piecewise constant hazard rate, with peaks at 1 and 3–4 months. After 6 months, the piecewise constant hazard rate follows the Gompertz curve fairly closely. Correcting for unobserved heterogeneity along the same lines as in this paper does not have any significant impact on their results.

Finally, our graphs might suggest that the female hazard rate is systematically below the male hazard rate. However, since the average man and the average woman differ in characteristics, the difference is not necessarily due to a pure sex effect. But even if we correct for these differences the female hazard rate is indeed significantly below that for comparable men.

5 RESULTS WITH REGARD TO THE REGRESSOR VARIABLES FOR MALES

In Table 2, the complete estimation results for males are reported for the Gompertz gamma and semi-parametric discrete models. These very different specifications are presented together, in order to highlight the robustness of the regressor parameters with respect to the choice of specification.

Before discussing the regressor parameters one at a time, a brief remark on their precise interpretation is in order. As mentioned in section 2, we distinguish two continuous regressors, age and unemployment. Neither of these is entered in log form. Hence, if age increases by one year or unemployment increases by one percentage point, the hazard changes by about $100 * \beta_i \%$, where β_i is the age or unemployment parameter. The vast majority of the regressors, however, is of the categorical type. The treatment of such variables usually amounts to selecting a reference category, defining dummy variables indicating the other categories and then estimating the effects of these other categories in deviation of the reference category. The resulting coefficients indicate the difference between any category and the reference category: likewise, the resulting t-ratios indicate whether any category differs significantly from the reference category. Especially this last property is undesirable, since it implies that the pattern of significance depends on the (essentially arbitrary) selection of a reference category. Instead, we have estimated coefficients and t-ratios which can be interpreted as deviations from the average, by imposing the following restriction on the coefficients of the J_k categories of a categorial variable k

$$\sum_{i=1}^{J_k} \beta_{ik} * f_{ik} = 0 \, . \tag{4}$$

In equation (4), f_{ik} represents the frequency of category i of variable k and β_{ik} its coefficient. The necessary recodings have been executed, using for both sexes their own frequencies as measured in the original dataset. As a result, our estimates of β_{ik} must be interpreted as follows: if a coefficient of, say, -0.08 appears, then the members of category i of variable k have a hazard which is about 8% lower than average. Likewise, a t-ratio of 2.40 implies that the behavior of this category differs significantly from the average.[10]

A monotonic and highly significant negative age effect is obtained. For each extra year, the hazard rate decreases by some 2.5%. There is a half-life value of about 28 years, which means that if two men differ in age by 28 years but are identical in all other respects, the younger one will have a hazard rate which is twice as large as that of the older person. These results are in the median range of those obtained in other studies.

Potential reasons for a negative age effect on one's re-employment probability are well-known (perception of low productivity and reluctance to engage in investments in human capital from the employer's side, less geographical and professional mobility on the unemployed's part). But apart from re-employment, withdrawal from the labour force appears as a way to end an unemployment spell in our data set. It is plausible that the tendency for withdrawal increases with age, *e.g.* because of the smaller re-employment probability for the elderly, and because of certain opportunities and incentives provided by the Dutch social security system. Though withdrawal might in this way mitigate or even reverse a negative age effect, the addition of a quadratic age term proved useless.

The variable 'population size of place of residence,' intended to proxy rural *vs.* urban environments, yields significant and remarkable results. A completely monotonic relationship between population size and hazard rate emerges. Inhabitants of very small villages have a hazard rate which is almost three times as large as that of an inhabitant of Amsterdam, the largest city of The Netherlands.

It is a well-documented fact that since the beginning of the eighties, Dutch unemployment has become increasingly concentrated in the four largest cities and provincial capitals (see, *e.g.*, Min. van SOZA (Ministry of Social Affairs) (1987a), pp. 56–58). That this phenomenon applies not only to the unemployment rate but also to its duration is in itself hardly surprising. The remarkable point is that, even after correction for a number of heterogeneity sources (age, nationality, *etc.*) the effect is still large, significant and completely monotonic.

Though this result may be in part a statistical artefact (*e.g.*, due to the omission of variables correlated with population), one may offer some plausible theoretical explanations for an independent population effect. In The Netherlands, the majority of unemployed owe a new job not to the intermediation of the labour

10 For each variable, one category cannot be estimated directly because of perfect multicollinearity. We arbitrarily selected such a category, of which the coefficient was later on calculated by means of equation (4), using the estimates of the other categories. Whenever possible, we also calculated its t-ratio (for variables with more than three categories, this is quite tedious and was left undone).

TABLE 2 – ESTIMATION RESULTS FOR MALES, SEMI-PARAMETRIC-DISCRETE *VS.* GOMPERTZ-GAMMA MODELS

	Semi-par.-discrete		Gompertz-gamma	
	par.	*t*-ratio	par.	*t*-ratio
REGRESSOR PARAMETERS				
age	– 0.023	– 8.11	– 0.026	– 8.83
population size place of residence				
0 < 10.000	0.466	6.49	0.518	6.63
10 < 25.000	0.322	6.57	0.362	6.92
25 < 50.000	0.212	–*	0.201	–*
50 < 100.000	– 0.010	– 0.20	– 0.015	– 0.26
100 < 200.000	– 0.140	– 2.49	– 0.142	– 2.31
Utrecht/Rotterdam/The Hague	– 0.422	– 6.31	– 0.461	– 6.37
Amsterdam	– 0.483	– 5.24	– 0.515	– 5.15
nationality				
Dutch	0.022	2.10	0.027	2.34
EC[11]	0.176	0.76	0.154	0.74
other countries	– 0.215	– 2.73	– 0.251	– 2.67
education				
primary school	– 0.131	– 2.50	– 0.149	– 2.61
lower secondary education				
general-dropout	– 0.128	– 1.13	– 0.138	– 1.10
general	– 0.011	– 0.15	– 0.013	– 0.16
vocational-dropout	– 0.088	– 1.42	– 0.082	– 1.20
vocational	0.169	4.35	0.193	4.46
intermediate secondary education				
general-dropout	– 0.356	– 1.64	– 0.385	– 1.60
general	0.117	1.26	0.122	1.17
vocational	0.043	0.51	0.042	0.44
higher vocational education	– 0.053	–*	– 0.101	–*
bachelor	0.314	0.90	0.318	0.92
academic	0.038	0.23	0.057	0.31
desired working time				
full-time (≥20h/week)	0.009	2.93	0.010	2.73
part-time (<20h/week)	– 0.530	– 2.93	– 0.541	– 2.73
main occupation				
agriculture	0.304	–*	0.353	–*
mining	– 0.521	– 0.83	– 0.610	– 0.90
metal	– 0.036	– 0.63	– 0.055	– 0.87
glass/pottery	0.763	0.95	0.187	0.16

11 EC, except Greece, Spain and Portugal.

TABLE 2 (CONTINUED)

	Semi-par.-discrete		Gompertz-gamma	
	par.	*t*-ratio	par.	*t*-ratio
construction/wood	0.788	13.10	0.942	14.86
printing/paper	− 0.329	− 1.55	− 0.401	− 1.66
leather/rubber	− 0.778	− 1.21	− 0.965	− 1.37
textiles	0.333	0.84	0.359	0.84
nourishment/chemicals	− 0.000	− 0.00	− 0.003	− 0.02
catering/housekeeping	0.183	1.93	0.179	1.70
trade/office	− 0.186	− 3.17	− 0.224	− 3.45
education/care	− 0.430	− 4.08	− 0.505	− 4.39
transportation/traffic	0.244	2.81	0.235	2.53
general purpose	− 0.333	− 6.15	− 0.374	− 6.39
to be determined	− 0.397	− 3.66	− 0.419	− 3.56
previous employment				
no	− 0.154	− 3.89	− 0.223	− 5.20
yes	0.070	3.89	0.101	5.20
change of profession				
previous job = main prof.	0.102	5.52	0.119	5.86
previous job ≠ main prof.	− 0.255	− 5.52	− 0.299	− 5.86
provincial unemployment rate	− 0.039	− 3.21	− 0.041	− 3.20
BASELINE-RELATED PARAMETERS				
time-varying constant				
month 1	− 4.174	− 18.56		
month 2	− 4.372	− 21.20		
month 3	− 3.850	− 18.94		
month 4	− 3.659	− 17.54		
month 5	− 3.708	− 17.07		
month 6	− 3.860	− 17.26		
months 7–8	− 4.019	− 17.87		
months 9–10	− 4.312	− 18.37		
months 11–12	− 4.474	− 18.22		
months 13–15	− 4.458	− 18.00		
months 16–18	− 4.583	− 17.87		
months 19–21	− 4.737	− 17.89		
months 22–24	− 4.365	− 16.51		
months 25–28	− 4.550	− 16.67		
months 29–32	− 4.321	− 15.97		
months 33–36	− 4.554	− 15.88		
months 37–42	− 4.666	− 16.37		
months 43–48	− 4.631	− 16.07		
months 49–54	− 4.508	− 15.49		

TABLE 2 (CONTINUED)

		Semi-par.-discrete		Gompertz-gamma	
		par.	t-ratio	par.	t-ratio
months 55–60		−4.687	−15.31		
year 6		−4.560	−15.32		
year 7+		−4.950	−14.54		
constant				−3.817	−18.39
alpha				−0.00035	−4.75
HETEROGENEITY-RELATED PARAMETERS					
mass point 1	probability	0.825	–*		
	value	0.	–*		
mass point 2	probability	0.167	2.57		
	value	−1.032	−6.95		
mass point 3	probability	0.008	1.05		
	value	−2.253	−5.41		
gamma variance				0.296	10.00
VALUE LOG-LOKELIHOOD		−21295.7		−21408.0	

* implicitly estimated, t-ratio not calculated (compare equation (4) and note 10)

exchange, but to personal contacts with friends, family members *etc.*, *i.e.* their social network. It is plausible, that such networks operate more effectively in small communities than in the anonymous environment of a large city. A second explanation equally relies on sociological economics. The labour market behavior of the unemployed is not only related to pecuniary phenomena, but also to 'social customs' or norms, especially the norm that, in principle, a man should work in order to receive an income. It is possible, that in smaller communities this traditional work-norm is still more vigorous than in larger communities. In the larger cities of The Netherlands, there are quarters in which a majority of the inhabitants is either unemployed or disabled (to a large extent hidden unemployed). When there is such a large concentration of non-working people, the work-norm may be eroded and the social pressure to find a new job may be much smaller than elsewhere (see De Neubourg (1992)). In such a situation, after some time people may adapt quite strongly to their unemployed status, a status which may not be regarded as undesirable or deviant by themselves or their environment.

For Dutch men (including those of foreign origin), the hazard rate is significantly higher than for the sample average. For male immigrants ('other countries,' mainly Morocco and Turkey) it is significantly below average, about 20% lower than that for Dutch men. This effect would be even more pronounced if we had a truly ethnic variable (foreigners increasingly assume the Dutch nationality). Males with EC nationality are a small group (1.5% of the male unemployed) so that the effect of this nationality class cannot be determined very well. In much of the existing work on reduced form models of unemployment duration, ethnic or nationality variables appear which yield similar sign patterns (see *e.g.* Lynch (1985) and Meyer (1988), while Han and Hausman (1990) report the opposite result). After correction for a number of heterogeneity sources (especially the relatively low education among migrant workers), foreigners still have a disadvantage on the labour market. It seems that the selection behaviour of employers is among the causes of this disadvantaged position (see Min. van SOZA (1987a), p. 26).

The education variable consists of six levels. Within some of these, we distinguish between general and vocational education. Whenever possible, we indicate whether school has been completed or not (dropouts). Only three categories differ significantly from the average. Men who only attended primary school have a hazard rate which is more than 13% below average. Men with a completed vocational education at the lower level have a hazard rate almost 20% higher than average. These are the two largest groups; each of them accounts for about 26% of all male unemployed. The other categories are much smaller with mostly insignificant effects. Only dropouts from general education at the intermediate level have an effect which is significant at a level of approximately ten per cent. Their hazard is about 35% below average, which is the lowest value found.

As mentioned in section 3, according to search theory, higher educated people have more human capital and are generally expected to have a higher arrival rate *ceteris paribus* than people with a lower education. Nonetheless, the relationship between education level and hazard rate is by no means monotonic. On the other hand, it is very obvious that dropouts are in a disadvantaged position. Their hazard rate is systematically below that of men with the same but completed education (bachelors, to some extent academic dropouts, are the only exception). Obviously, people without certificates from school lack an indicator of their productivity to screening employers, who use formal education as productivity proxy. Another explanation is the potential correlation between the dropout categories and individual motivation.

In this study, part-timers are people seeking work for less than 20 hours per week (this unusually low cut-off value is data-dictated). Their hazard is significantly below that of full-timers (less than one half).

We distinguish 14 professional categories, referring to an individual's main profession as mentioned at the labour exchange. A major problem are the differences in size between the various categories; some are virtually empty. The discrepancies between the two columns of Table 2 are all due to such small categories (*e.g.* there are only 25 glass/pottery workers in our sample). The last six categories are among the larger ones: together they constitute 63% of the unemployed men. They are all significant. Apart from these six, the categories metal and construction/wood are also large, consisting of 13% and 17% of unemployed men, respectively. The group of metal workers shows an average exit behaviour. The extremely large and significant coefficient of the category construction/wood is due to a seasonal effect.[12] The same remark applies to the (very small) agricultural category. The remaining six categories are all smaller than 1.5%. The significantly negative coefficients of the last two categories ('general purpose' and the residual category 'to be determined') can be easily explained. Due to the way in which the labour exchange applies these classifications, they mainly consist of people with little or no specific qualifications.

Men who had a job immediately prior to their registration at the labour exchange, have a hazard rate which is over one-third higher than that for men without such a previous job. This effect is significant. Apparently, the previous employment dummy can be regarded as a proxy for experience, extra human capital *etcetera*, which employers may use as productivity indicators.

The 'change of profession' variable is only relevant for people with previous employment. Men who have not changed profession have a hazard rate which exceeds that of those who did change by almost 50%. It is difficult to explain this significant result, since the interpretation of the variable is unclear. After all, a change of profession can be due to various kinds of circumstances.

An increase in the provincial unemployment rate with 1 percentage point lowers the individual hazard rate significantly by about 4%.[13] Since in many studies a *UV*-ratio is used as a labour market indicator, the scope for comparison is limited, but our result does not seem to lie outside the range of previous ones.

6 CONCLUSIONS

In the preceding sections, we described the estimation results for semi-parametric specifications for the hazard rate without correcting for unobserved heterogeneity and with gamma and semi-parametric correction for unobserved heterogeneity. We also briefly compared them with parametric specifications. The most important conclusions are the following.

Concerning the *model specification*, three points draw attention. First, the dangers of using parametric baseline specifications have been emphasized. If one has a relatively rich data set a semi-parametric specification should be preferred, especially when the 'raw' duration distribution suggests that nonmonotonicity may be present. Applying standard parametric specifications may in such cases lead to paradoxical results, especially for the Weibull specification. Secondly, the

12 Due to the long and cold winter of 1986/1987, the usual outflow of unemployed in this category during the month of March was postponed until April–May 1987. See Min. van SOZA (1987b), p. 1.

13 This interpretation applies to differences between the unemployment rates in separate provinces at a point in time, as well as to changes in the unemployment rate in some province as time proceeds. Meyer (1988) attempts to distinguish between these two aspects of unemployment rate influence on the individual hazard rate. His results suggest that our approach may be too crude.

choice for a specific heterogeneity correction seems much less crucial. Our results support the use of the gamma distribution as a convenient approximation to correct for unobserved heterogeneity. The results with semi-parametric baseline and discrete or gamma heterogeneity correction are superior to all others and hence our discussion of the hazard time path is based on them. Thirdly, the results on the regressor parameters are generally insensitive to the precise model specification.

Concerning the *economic interpretation*, the results for men with respect to the regressor parameters generally confirm our prior expectations. Some striking results are:

- The hazard rate declines significantly and monotonically with age.
- There is an inverse relation between population size of a man's domicile and his hazard rate.
- There is evidence of a significantly disadvantaged labour market position for male migrant workers.
- Men with only primary or unfinished secondary education are in an unfavourable position. Prospects are good for men with a completed lower vocational education.
- Men looking for a very short-term job hardly have any chance on the labour market.
- Having had a job immediately preceding one's unemployment spell significantly raises the hazard.

Both sexes show, after heterogeneity correction, a nonmonotonic duration dependence. The hazard first rises for four months, then from the sixth month onwards it declines until the end of the first year. Thereafter, it does not show any clear tendency to rise or fall anymore.

APPENDIX I

ESTIMATION OF THE SEMI-PARAMETRIC HAZARD RATES

Our data refer to single unemployment spells only and there is no joint entry date into unemployment for the persons in the data set. The appropriate likelihood function for this data type, derived by Lancaster (1979) and used in this study, represents the likelihood of the events in the period during which the exit process of the unemployed is monitored, *i.e.* April–May 1987. In order to avoid arbitrary assumptions concerning the entry process into unemployment, the information contained in the elapsed unemployment periods prior to March 31, is left unused (see Ridder (1984) and Kerckhoffs (1989) for an elaboration of this point). This implies that the contribution of an individual i to the likelihood value is derived, conditional on his or her elapsed spell of unemployment t_i at March 31 (variable 2 in section 2).

For persons who are still unemployed at the end of the observation period with length h, we only know that their duration of unemployment is at least $t_i + h$ days, where $h = 59$ in our application. The probability for this to happen is $S(\mathbf{x}_i, t_i + h)/S(\mathbf{x}_i, t_i)$, given their elapsed spell at March 31. For persons who left unemployment after a period $s_i < h$ (variable 3 in section 2), we know that their duration of unemployment has been exactly $t_i + s_i$ days. Given their elapsed spell on March 31, the pdf for this event is $f(\mathbf{x}_i, t_i + s_i)/S(\mathbf{x}_i, t_i)$. In the absence of heterogeneity, the likelihood function can be expressed in terms of the hazard function

$$L = \prod_{N_u} \Theta(\mathbf{x}_i, t_i + s_i) * exp\left\{ - \int_{t_i}^{t_i + s_i} \Theta(\mathbf{x}_i, s)\, \mathrm{d}s \right\}$$

$$* \quad \prod_{N_c} exp\left\{ - \int_{t_i}^{t_i + h} \Theta(\mathbf{x}_i, s)\, \mathrm{d}s \right\}, \qquad \text{(A.1)}$$

where N_u stands for uncensored cases, while N_c stands for the censored cases. The likelihood function can then be maximized with respect to the parameters of $\Theta(\mathbf{x}, t)$. We used the general maximum likelihood programme GRMAX, which is based on the Newton-Raphson-technique. The maximum likelihood technique yields consistent estimators, provided of course that the model is correctly specified. If one allows for unobserved heterogeneity as in (3) (see section 3), the likelihood function gets more complicated since it involves integration over $g(u)$. The parameters of $g(u)$ are estimated together with the hazard parameters, provided that they are identified (see Elbers and Ridder (1982) and Heckman and Singer (1984a/b) for identification issues).

Following Lancaster (1979), first the continuous parametric gamma pdf is adopted for $g(u)$. The gamma pdf is convenient because it yields a closed form for $S(\mathbf{x}, t)$ and $f(\mathbf{x}, t)$, *e.g.*

$$S(\mathbf{x}, t) = \{1 + \sigma^2 I(\mathbf{x}, t)\}^{-1/\sigma^2}, \qquad \text{(A.2)}$$

where σ^2 is the variance of the distribution (the mean is *a priori* set equal to one, which is not restrictive because we have a constant in Θ). In computing the integrated hazard $I(\mathbf{x}, t)$, appearing in equation (A.2), we took into account the fact that the provincial unemployment rate, one of the regressors in $\mathbf{x}$, varies through time. However, any choice of a parametric pdf $g(u)$ is essentially arbitrary and the hazard parameters may be highly sensitive to this choice, as was forcefully argued by Heckman and Singer (1984a/b). As an alternative, they advocated the use of a discrete non-parametric distribution for $g(u)$. Heckman and Singer (1984b) proved that, even if the true distribution of the unobservables is continuous, the likelihood function for a specific sample has a global maximum at a finite number of mass points, and that this maximum yields a consistent estimate of the hazard parameters, provided that the hazard function $\Theta(\mathbf{x}, t)$ is correctly specified apart from the unobserved heterogeneity. This Non-Parametric Maximum Likelihood Estimator or NPMLE requires maximizing a likelihood function that involves factors like

$$S(\mathbf{x}, t) = \sum_{i=1}^{J} P_i * exp\{-e^{v_i} * I(\mathbf{x}, t)\}, \quad \sum_{i=1}^{J} P_i = 1 . \tag{A.3}$$

Maximization is performed with respect to the hazard parameters, the mass points v_i, the associated probabilities P_i and the number of mass points J. We maximize the likelihood function for a large given number of mass points, using the Newton-Raphson-technique. Exploring the shape of the likelihood function by systematically varying J, the starting values for the v_i's and P_i's, it is possible to locate the NPMLE.

At the start we set $J = 20$ and selected the starting values for the v_i such that the discrete distribution can reasonably well approximate the continuous gamma distribution or any other distribution. For given values of v_i, the associated values of P_i were determined by our algorithm. As starting values for the P_i's we first used those corresponding to a model without heterogeneity correction, *i.e.* $P_1 = 1$, $P_i = 0$ for $i \in \{2, 20\}$, and secondly, values for P_i derived from the previously estimated gamma distribution. For the given values of the v_i's about seventy five per cent of the estimated values of P_i turned out to be zero irrespective of their starting values. In the second stage, these mass points were suppressed and with this much lower number of J, the algorithm was used to jointly estimate all the remaining parameters including the v_i's and P_i's freely.

Alternatively, again for $J = 20$, we fixed the P_i's at values resulting from the gamma distribution and estimated the v_i's freely. In this case clustering occurred, many v_i's assuming the same value. Lumping these together (which obviously resulted in a smaller J), we iterated over all the remaining parameters. Upon convergence, the algorithm reached the same optimum in all the cases that we analyzed.

More detailed information on the parametric specifications for the baseline hazard is given in Kerckhoffs *et al.* (1992).

APPENDIX II

THE DISTRIBUTION OF INCOMPLETE AND COMPLETE DURATIONS FOR MEN AND WOMEN

	Males		Females	
Duration (months)	Incomplete duration	Complete duration	Incomplete duration	Complete duration
0 < 1	1085	73	1116	28
1 < 2	1181	193	1220	139
2 < 3	1197	412	1267	270
3 < 4	1170	466	714	279
4 < 5	1037	417	883	191
5 < 6	833	286	906	180
6 < 8	1296	342	1916	333
8 < 10	979	175	1326	260
10 < 12	755	112	959	154
12 < 15	1047	127	1185	132
15 < 18	886	97	1011	105
18 < 21	788	68	1086	121
21 < 24	652	82	745	82
24 < 28	754	68	797	85
28 < 32	686	76	789	64
32 < 36	563	48	596	55
36 < 42	773	53	670	47
42 < 48	707	51	670	56
48 < 54	795	55	501	33
54 < 60	625	40	461	39
60 < 72	1119	68	602	30
72+	1072	37	580	18
Total	20000	3346	20000	2701

- The table uses the duration intervals defined for the semi-parametric baseline specification.
- Incomplete duration: elapsed unemployment duration at April 1, 1987.
- Complete duration: elapsed unemployment duration at day of exit (for uncensored cases only).

REFERENCES

Berg, G.J. van den (1990), 'Search Behaviour, Transitions to Non-participation and the Duration of Unemployment,' *Economic Journal*, 100, pp. 842–865.

Elbers, C. and G. Ridder (1982), 'True and Spurious Duration Dependence: The Identifiability of the Proportional Hazards Model,' *Review of Economic Studies*, 49, pp. 403–410.

Gorter, D. and E. Hoogteijling (1990), 'Duur van het zoeken naar werk, 1985–1987,' Central Bureau of Statistics, BPA nr 5770-90-M1, mimeographed.

Groot, W.N.J. and A.G. ter Huurne (1988), *Re-employment Probabilities of the Dutch Young Unemployed*, IVA, Institute of Social Research of Tilburg University.

Groot, W. and T. ter Huurne (1989), 'Herintredingskansen van werkloze jongeren,' *Mens en Maatschappij*, 64, pp. 364–383.

Groot, W. (1990), 'The Effects of Benefits and Duration Dependence on Re-employment Probabilities,' *Economics Letters*, 32, pp. 371–376.

Han, A. and J.A. Hausman (1990), 'Flexible Parametric Estimation of Duration and Competing Risk Models,' *Journal of Applied Econometrics*, pp. 1–28.

Heckman, J.J. and B. Singer (1984a), 'Econometric Duration Analysis,' *Journal of Econometrics*, 24, pp. 63–132.

Heckman, J.J. and B. Singer (1984b), 'A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data,' *Econometrica*, 52, pp. 271–320.

Jensen, P. (1990), 'Different Escape Routes for Unemployment: Empirical Evidence and Sensitivity Analysis,' Preliminary and incomplete working paper, presented at the second conference of the European Association of Labour Economists, Lund, September 1990.

Kerckhoffs, C. (1989), *De uittreedkans van geregistreerde werklozen in Nederland: Voorjaar 1987*, Unpublished master's thesis, University of Maastricht, The Netherlands.

Kerckhoffs, C., C. de Neubourg and F. Palm (1992), 'The Determinants of Individual Unemployment Duration in The Netherlands,' Research Memorandum 1992, Paper presented at the EEA meeting Cambridge 1991.

Kiefer, N.M. (1988), 'Economic Duration Data and Hazard Functions,' *Journal of Economic Literature*, 26, pp. 646–679.

Kooreman, P. and G. Ridder (1983), 'The Effects of Age and Unemployment Percentage on the Duration of Unemployment,' *European Economic Review*, 20, pp. 41–57.

Lancaster, T. (1979), 'Econometric Models for the Duration of Unemployment,' *Econometrica*, 47, pp. 939–956.

Lindeboom, M. and J. Theeuwes (1991), 'Search, Benefits and Entitlement,' University of Leiden, mimeographed.

Lynch, L.M. (1985), 'State Dependence in Youth Unemployment,' *Journal of Econometrics*, 28, pp. 71–84.

McKenna, C.J. (1985), *Uncertainty and the Labour Market: Recent Developments in Search Theory*, New York.

Meyer, B. (1988), 'Unemployment Insurance and Unemployment Spells,' NBER working paper nr. 2546, Washington D.C.

Ministerie van Sociale Zaken en Werkgelegenheid (1986) (referred to as Min. van SOZA), *Periodieke rapportage arbeidsmarkt 1986*, The Hague.

Min. van SOZA (1987a), *Periodieke rapportage arbeidsmarkt 1987*, The Hague.

Min. van SOZA (1987b), *De arbeidsmarkt in maart 1987*, The Hague.

Min. van SOZA (1987a), *Periodieke rapportage arbeidsmarkt 1988*, The Hague.

Narendranathan, W., S. Nickell and J. Stern (1985), 'Unemployment Benefits Revisited,' *Economic Journal*, 95, pp. 307–329.

Neubourg, C. de (1988), *Unemployment, Labour Slack and Labour Market Accounting*, Amsterdam – New York.

Neubourg, C. de (1990), *Unemployment and Labour Market Flexibility*, ILO, Geneva.

Neubourg, C. de (1992), 'Job Libido and the Culture of Unemployment,' in: R. Coughlin (ed.), *Socio-economic Perspectives 1990*, New York, pp. 61–78.

Nickell, S. (1979a), 'The Effect of Unemployment and Related Benefits on the Duration of Unemployment,' *Economic Journal*, 89, pp. 34–49.

Nickell, S. (1979b), 'Estimating the Probability of Leaving Unemployment,' *Econometrica*, 47, pp. 1249–1266.

Opstal, R. van and J. Theeuwes (1986), 'Duration of Unemployment in the Dutch Labour Market,' *De Economist*, 134, pp. 351–367.

Opstal, R. van and F. van de Pol (1990), 'The Transition from School to Work in 1979–1987 in The Netherlands,' Central Bureau of Statistics, BPA nr. 325-90-M1.

Ridder, G. (1984), 'The Distribution of Single-spell Duration Data,' in: Neumann and Westergaard-Nielsen (eds.), *Studies in Labor Market Dynamics*, Berlin.

Stewart, M.B. and W. Narendranathan (1990), 'How Does the Income Effect Vary as Unemployment Spells Lengthen? Some Semi-parametric Evidence,' Preliminary paper presented at the second conference of the European Association of Labour Economists, Lund, September 1990.

Theeuwes, J., M. Kerkhofs and M. Lindeboom (1990), 'Transition Intensities in the Dutch Labour Market: 1980–1985,' *Applied Economics*, 22, pp. 1043–1061.

Summary

THE DETERMINANTS OF UNEMPLOYMENT AND JOBSEARCH DURATION IN THE NETHERLANDS

This study investigates the determinants of individual unemployment and jobsearch duration and the size and direction of duration dependence in a proportional hazard model. New insights into the shape of the time dependence of unemployment duration are obtained using a semi-parametric baseline hazard specification in combination with a parametric (gamma) heterogeneity correction or with a semi-parametric heterogeneity correction. Registration data on unemployed individuals in The Netherlands in the period April 1987–May 1987 are used. For both sexes, the hazard rate quickly rises during the first five months, but then after the fifth month rapidly declines until the end of the first year of unemployment, not showing any clear tendency thereafter. Commonly used parametric specifications for the baseline hazard prove to be too restrictive. For men we discuss the influences of the regressors.