ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
Lönnrotinkatu 4 B 00120 Helsinki Finland Tel. 601322 Telefax 601753

# Keskusteluaiheita - Discussion papers

No. 352

Juha Kettunen*

# HETEROGENEITY IN UNEMPLOYMENT DURATION MODELS

* I am grateful to Andrew Chesher and Richard Dunn for valuable comments and the Yrjö Jahnsson Foundation for financial support. The first versions of this paper were presented at ETLA, LAMA and The Finnish Statistical Society.



ISSN 0781-6847 31.01.1991

**KETTUNEN**, Juha, **HETEROGENEITY IN UNEMPLOYMENT DURATION MODELS.** Helsinki : ETLA, Elinkeinoelämän Tutkimuslaitos, The Research Institute of the Finnish Economy, 1991. 22 p. (Keskusteluaiheita, Discussion Papers, ISSN 0781-6847; no. 352).

**ABSTRACT:** This paper is concerned with the specification of duration models and the effects of omitted variables, which may cause bias to parameter estimates. Weibull models with gamma and mass point heterogeneity are estimated using Finnish unemployment duration data. Graphical examination of residuals derived for the heterogeneity models show that the discrete mass point mixing distribution is better than the continuous gamma distribution to rectify the model misspecification.

**KEY WORDS:** Heterogeneity, Unemployment duration.

## I. Introduction

In this paper two methods for incorporating individual heterogeneity into Weibull duration models of unemployment spells are considered using Finnish unemployment data. It is well known that omitted variables cause bias to parameter estimates if duration models are not controlled for omitted variables (Lancaster, 1979, Nickell, 1979 a,b, Lancaster and Nickell, 1980). Especially the shape of the hazard function of leaving unemployment over duration time is considered in this paper. A Weibull model applied to the data produces a decreasing hazard function, but controlling for heterogeneity implies an increasing hazard function, which is in concordance with the standard search theory (e.g. Kiefer, Neumann, 1989) for a number of reasons. Hence the correction for heterogeneity and model specification tests are particularly important.

The improvement of model specification, when introducing heterogeneity, is shown using a graphical procedure based on examination of residuals derived for the heterogeneity models. The discrete mass point mixing distribution is shown to provide a better pattern of heterogeneity than the continuous gamma heterogeneity.

The paper is organized as follows. A Weibull model of unemployment duration is estimated in section II with the allowance for gamma heterogeneity. Mass point models are estimated in section III. The residuals of these models are derived, and a graphical examination based on the residuals are presented in section IV. Finally, section V concludes the study.

## II. Gamma heterogeneity

In this section an approach to the incorporating of gamma heterogeneity into duration models is described and the integrated hazard for graphical examination of residuals is derived. However, the properties of a basic Weibull model are studied first.

Let us consider independent pairs of independent random variables T and Z, where T is the duration variable of primary interest and Z is a censoring variable. A censoring time or a duration time is observed, $t = \min(T, Z)$, with the censoring indicator, $\bar{c} = 1$ if $T \geq Z$ and $\bar{c} = 0$ otherwise. An indicator of a completed spell of unemployment is defined as $c = 1 - \bar{c}$.

The survivor function for T is equal to one minus the distribution function of the duration variable, which can be written as

$$S(t) = e^{-I(t)} \tag{1}$$

and the density function can be written as

$$f(t) = h(t)e^{-I(t)} \tag{2}$$

for $t \geq 0$. $I(t)$ is the integrated hazard

$$I(t) = \int_0^t h(\tau)d\tau. \tag{3}$$

The likelihood contribution of an individual can be written in view of the above definitions as

$$\ell = h(t)^c e^{-I(t)}. \tag{4}$$

The Weibull model is a versatile family of duration distributions in view of its interpretation and its flexibility for empirical fit, and it has been widely used in applications of duration models to unemployment spells. The hazard function is as follows

$$h(t) = \alpha t^{\alpha-1} e^{x\beta}, \tag{5}$$

where x is a vector of explanatory variables for an individual and $\beta$ is a vector of structural parameters. $\alpha$ is the shape parameter. If $\alpha > 1$, the hazard function is increasing in duration time and it is said that there is *positive duration dependence*. If $\alpha = 1$, the hazard function is constant and the distribution of unemployment spells is exponential. If $\alpha < 1$, the hazard function is decreasing in time and it is said that there is *negative duration dependence*. The integrated hazard can be written as

$$I(t) = t^{\alpha}e^{x\beta}. \qquad (6)$$

To estimate the unknown parameters, the hazard function (5) and the integrated hazard (6) are substituted into the likelihood function (4).

It is inevitable that in an econometric analysis relevant variables will be omitted, either because they are unmeasurable or because their importance is unsuspected. Even if the omitted variables are uncorrelated with those which are included in the model the parameters will be biased. The usual method for incorporating heterogeneity is to assume a parametric functional form for the pattern of heterogeneity. The gamma mixing distribution has been chosen because it is analytically simple to use and it provides quite a flexible model for the distribution of the heterogeneity component.

Lancaster (1979) found that the estimated falling hazard function represents, at least in part, merely the effect of unrecognized heterogeneity of the sample individuals, i.e. omitted variables. He introduced regressors into the model one at a time and each time found that the parameter estimates increased. Rather than being an estimate of a behavioral parameter, $\alpha$ is, at least in part, merely an index of specification error. The more significant regressors are included, the larger it becomes. It may be shown under fairly general conditions that the coefficients of explanatory variables are then biased towards zero (Lancaster and Nickell, 1980). Therefore, we may expect the parameters of the model to increase in absolute value when the effects of omitted variables are taken into account.

The method of correcting for gamma heterogeneity has been widely used during the 80's in duration models. [e.g.

Kooreman and Ridder (1983), Newman and McCulloch (1984), Narendranathan, Nickell and Stern (1985) and Engström and Löfgren (1987)]. However, the model specification has not been examined in these studies.

Suppose the individuals of the sample differ to some degree with respect to some unobservable variable, say, motivation v. Each individual has his own v and hence his own hazard function h(t). Lancaster using data from stock of unemployed persons assumed that these hazard functions have a gamma distribution. The conditional hazard function in a Weibull model allowing for gamma heterogeneity is

$$h(t|v) = v\alpha t^{\alpha} e^{x\beta}, \qquad (9a)$$

where v has a gamma density

$$g(v) = \frac{\epsilon^{\mu}}{\Gamma(\mu)} v^{\mu-1} e^{-\epsilon v} \quad \text{with } \Gamma(\mu) = \int_0^{\infty} w^{\mu-1} e^{-w} dw. \qquad (9b)$$

The expected value of the heterogeneity component $E(v) = \mu/\epsilon$ is normalized to one by setting $\epsilon=\mu$ and its variance, i.e. $\sigma^2 = 1/\mu$, is estimated. Integrating the survivor function over the assumed mixing distribution gives a closed form for the survivor function with gamma heterogeneity. Differentiation gives the corresponding density and the marginal hazard function, not conditional on v, can be written as

$$h(t) = \alpha t^{\alpha-1} e^{x\beta}[1 + \sigma^2 t^{\alpha} e^{x\beta}]^{-1}. \qquad (10)$$

Integrating (10) from nought to t gives the needed integrated hazard

$$I(t) = 1/\sigma^2 \log[1 + \sigma^2 t^{\alpha} e^{x\beta}]. \qquad (11)$$

I(t) has a unit exponential distribution, as will be seen in section IV. The hazard function (10) and the integrated hazard (11) are substituted into the likelihood function (4) to estimate the unknown parameters.

The data of 2077 Finnish unemployed persons has been taken from the register of the Ministry of Labour. The sample has been taken from the flow into unemployment during the year 1985 and the individuals have then been followed until the end of 1986. 40 % of the observations are right censored. The description of the variables of the models are in the Appendix of this paper and reference for further details regarding the data should be made to Kettunen (1989). The models have been estimated using the SAS/IML (1985) programming language using the Berndt, Hall, Hall and Hausman (BHHH) (1974) algorithm, which requires the analytic first derivatives of the log likelihood function with respect to the parameters to be estimated.

The results of the estimations are in Table 1. The constant of the model, where the effect of omitted variables is captured, decreases and the absolute values of the statistically significant parameter estimates increase in most cases when gamma heterogeneity is introduced into the model as was expected. The basic Weibull model produces a decreasing hazard function, but the shape parameter of the Weibull model with gamma heterogeneity takes a value larger than one indicating increasing hazard functions for the individuals. The sample hazard function with gamma mixing distribution is increasing at the beginning of unemployment, but later on it turns into a decreasing function.

Table 1. *Gamma heterogeneity in a Weibull model*

Dependent variable: The length of the spell of unemployment

(A) A Weibull model
(B) A Weibull model with gamma heterogeneity

Std.errors in parentheses

| | (A) | (B) |
|---|---|---|
| Shape parameter | 0.861 | 1.201 |
| | (0.020) | (0.058) |
| Variance of heterogeneity | | 1.045 |
| | | (0.162) |
| Constant | -1.478 | -1.157 |
| | (0.136) | (0.210) |
| Children | -0.004 | -0.020 |
| | (0.050) | (0.080) |
| Married | 0.170 | 0.135 |
| | (0.065) | (0.101) |
| Sex | -0.007 | -0.066 |
| | (0.056) | (0.090) |
| Age | -0.042 | -0.057 |
| | (0.003) | (0.005) |
| Level of education | 0.064 | 0.035 |
| | (0.058) | (0.095) |
| Training for employment | 0.176 | 0.321 |
| | (0.072) | (0.119) |
| Member of UI fund | 0.213 | 0.364 |
| | (0.060) | (0.096) |
| Came from schooling | 0.291 | 0.375 |
| | (0.078) | (0.130) |
| Came from house work | -0.711 | -0.892 |
| | (0.124) | (0.176) |
| Regional demand | 0.168 | 0.353 |
| | (0.238) | (0.338) |
| Occupational demand | 0.641 | -0.098 |
| | (0.600) | (0.963) |
| Taxable assets | 1.021 | 0.822 |
| | (1.080) | (1.379) |
| Replacement ratio | -1.223 | -2.243 |
| | (0.150) | (0.261) |
| Log likelihood | -4962.5 | -4920.6 |

## III. Mass point heterogeneity

In this section a mass point approach to the incorporation of heterogeneity into duration models is described. A method to estimate a discrete mixing distribution is described and integrated hazards for graphical examination of residuals are derived. The main method for incorporating heterogeneity has been to assume a parametric functional form for the pattern of heterogeneity. Heckman and Singer, who propose a discrete pattern of heterogeneity (1984 a,b), have shown that estimates of the structural parameters may be sensitive with respect to the parametric forms assumed for heterogeneity. Furthermore there are a limited number of tractable forms for mixing distributions available.

The approach dispensing with the need to specify a parametric distribution for the heterogeneity component has its origins in the work of Kiefer and Wolfowitz (1956), who showed that a nonparametric characterization of the heterogeneity distribution ensures consistent estimation of simultaneously estimated structural parameters. Further work on the properties of mass point mixing distributions has been carried out by Simar (1976), Laird (1978), Lindsay (1983 a,b) and Heckman and Singer (1984 a,b). Applications of mass point approach in the context of discrete choice models have been presented by Davies and Crouchley (1984), Dunn, Reader and Wrigley (1987), Davies (1987) and Card and Sullivan (1988). Applications to duration models have been presented by Brännäs (1986 a,b), Trussell and Richards (1987) and Ham and Rea (1987).

To illustrate the discrete heterogeneity problem with discrete variables, suppose for simplicity that there are in the sample two groups, which have different constant hazard functions $h_1(t) > h_2(t)$ for all $t \geq 0$ and which are not controlled by explanatory variables in the data. At $t=0$ the estimated hazard is the average of the hazards of these groups. The proportion of the low hazard group increases over time and the estimation gives an indication that the hazard function of the individuals is falling when it is in fact constant. The average hazard of the sample is converging asymptotically to the hazard function $h_2(t)$. In the sequel of this section it can be seen that this example comes true with the data in the case of two mass points.

Define the function $f_Q = \int_0^\infty f_u(t)dQ(u)$ to be the mixture density corresponding to a mixing distribution Q. The densities $f_u$ are atomic densities for each value of u. A convex combination of m elements of $f_u$ can be written as $\Sigma\, p_i f_{u_i}$ with restrictions $\Sigma\, p_i = 1$. It is assumed that the density of heterogeneity has a particular functional form, namely the likelihood function has been specified so that there are m types of individuals in the sample not controlled by explanatory variables. The probabilities $p_i$ are the shares of these groups, but it is not possible to distinguish between m types of individuals.

In the case of parametric duration models the mixing likelihood contribution can be written as

$$f_Q = \sum_{i=1}^{m} p_i h_i(t)^c e^{-I_i(t)}, \qquad (12)$$

where $h_i(t) = \alpha t^{\alpha-1} e^{u_i+x\beta}$ and $I_i(t) = t^{\alpha} e^{u_i+x\beta}$ are the atomic hazard functions and integrated hazards. The objective is to estimate the discrete mixing distribution consistently with the atomic densities, a maximizer of the mixture likelihood function $\ell(Q) = \pi\, f_Q$. Maximizing the likelihood function $\ell(Q)$ over Q may be accomplished by maximizing the concave functional $L(f) = \Sigma \log f_Q$. The problem is equivalent to the maximization of a concave function subject to finitely many linear constraints.

To ensure that the probabilities $p_i \in (0,1)$ and sum to one the probabilities associated with each location have been defined using a multinomial logit type of formula

$$p_i = \frac{e^{g_i}}{1 + \sum_{k=1}^{m-1} e^{g_k}}, \quad i = 1,\ldots,m-1, \qquad (13)$$

where $g_k$, $k = 1,\ldots,m-1$ are parameters to be estimated. The probability of the last mass point is $p_m = 1-p_1-p_2-\ldots-p_{m-1}$. By definition $p_1$ equals one when m=1. The parameters $g_k$ work only as a device, and they do not have an interesting economic interpretation in this context.

Locations of mass points are defined as $\exp(u_i)$. The vector of ones has been left out from the explanatory variables. The idea of mass point models is that the constant parameter $\beta_0$ of the basic model is partitioned in m location parameters $u_i$ and each of the location parameters is given a probability $p_i$. In the case where m=1, when there is one location parameter, the parameter $u_1$ is equal to the constant of the basic Weibull model $\beta_0$. Consequently, the likelihood function of mass point models reduces to the likelihood function of the basic Weibull model, and the model with one mass point and the basic Weibull model coincide.

Following Lindsay (1983a) it can be seen that the log likelihood function $L(f) = \Sigma \log f_Q$ is differentiable with the directional derivative of L at $L_{Q_0}$ towards $L_{Q_1}$ being

$$D(u;Q) = \lim_{p\to 0} \{L[(1-p)f_{Q_0} + pf_{Q_1}] - L(f_{Q_0})\}/p$$

$$= \Sigma\ [(f_{Q_1} - f_{Q_0})/f_{Q_0}]$$

$$= \Sigma\ f_{Q_1}/f_{Q_0} - n, \qquad (14)$$

where it will be understood that the summing is over observations. The main idea of the procedure is to increase the number of mass points until $D(u;Q) \le 0$. Then the procedure is stopped and the semiparametric ML estimator is obtained. This procedure is suggested also by Brännäs and Rosenqvist (1988). BHHH algorithm is directly applicable to the constrained problem of maximization over discrete mixtures Q with a fixed number of support points. A simple first order check for a global maximum is to verify that $D''(u^*; Q) \le 0$ at the support points of measure Q. The values of function D of the models with 2,3,4 and 5 mass points are 0.86, 5.97, 1.13 and -3.45 respectively, showing that five mass points are enough to rectify the effect of omitted variables with this data.

In Figure 1 the probabilities of mass points $p_i$ are plotted against the locations $\exp(u_i)$. The mixing distribution seems not to be very far from a gamma distribution. There seems to be a pattern in the way new mass points are located. When the number of mass points are increased, each location in the previous model seems to get new locations on both its sides in the next model. Furthermore, they seem to take less mass than the neighbour mass points in the previous model.

The results of estimations of mass point models are in Table 2. The model with two mass points produces constant hazard functions for the two groups which are not controlled for explanatory variables. Models with three or more mass points produce increasing hazard functions. An increasing hazard function is in concordance with standard search theories. The absolute values of statistically significant parameter estimates increase in most cases when more mass points are introduced into the model, as is to be expected.

Many of the explanatory variables have significant effects on the re-employment probability. Age is a very significant factor. Older people are more likely to have problems in finding jobs. Training for further employment has a significant and positive effect on the re-employment probability. Members of the UI funds, i.e. members of the labour unions in the Finnish system, are often skilled workers and therefore they become employed earlier than the non-members. The persons leaving school or the army usually have no great problems. They leave unemployment clearly earlier than the others. The persons who have come from house work find it very difficult to find a job. The effects of unemployment benefits are measured using the benefit replacement ratio. The benefits decrease significantly the re-employment probability as is expected by the search theoretical models. The number of children, marriage, gender, level of education, demand variables and taxable assets do not have statistically significant effects.

The integrated hazards of mass point models need to be derived. The density and survivor functions are obtained from the mixing likelihood contribution (12) by setting $c = 1$ and $c = 0$ respectively. The hazard function $h(t)$ is the ratio of these two functions, i.e. $h(t) = f(t)/S(t)$.

$$f(t) = \sum_{i=1}^{m} p_i h_i(t) e^{-I_i(t)} \tag{15a}$$

$$S(t) = \sum_{i=1}^{m} p_i e^{-I_i(t)} \tag{15b}$$

$$h(t) = \sum_{i=1}^{m} p_i h_i(t) e^{-I_i(t)} / \sum_{i=1}^{m} p_i e^{-I_i(t)} . \tag{15c}$$

Integrating the hazard function gives a rather simple expression for the integrated hazard

$$I(t) = \log\{[\sum_{i=1}^{m} p_i e^{-I_i(t)}]^{-1}\}, \tag{16}$$

which is needed in the graphical examination of residuals. It is based on the fact that $I(t)$ has a unit exponential distribution in the absence of censoring, as will be seen in the next section. Note that if $m=1$ the integrated hazard (16) reduces to the integrated hazard of the basic Weibull model (6).

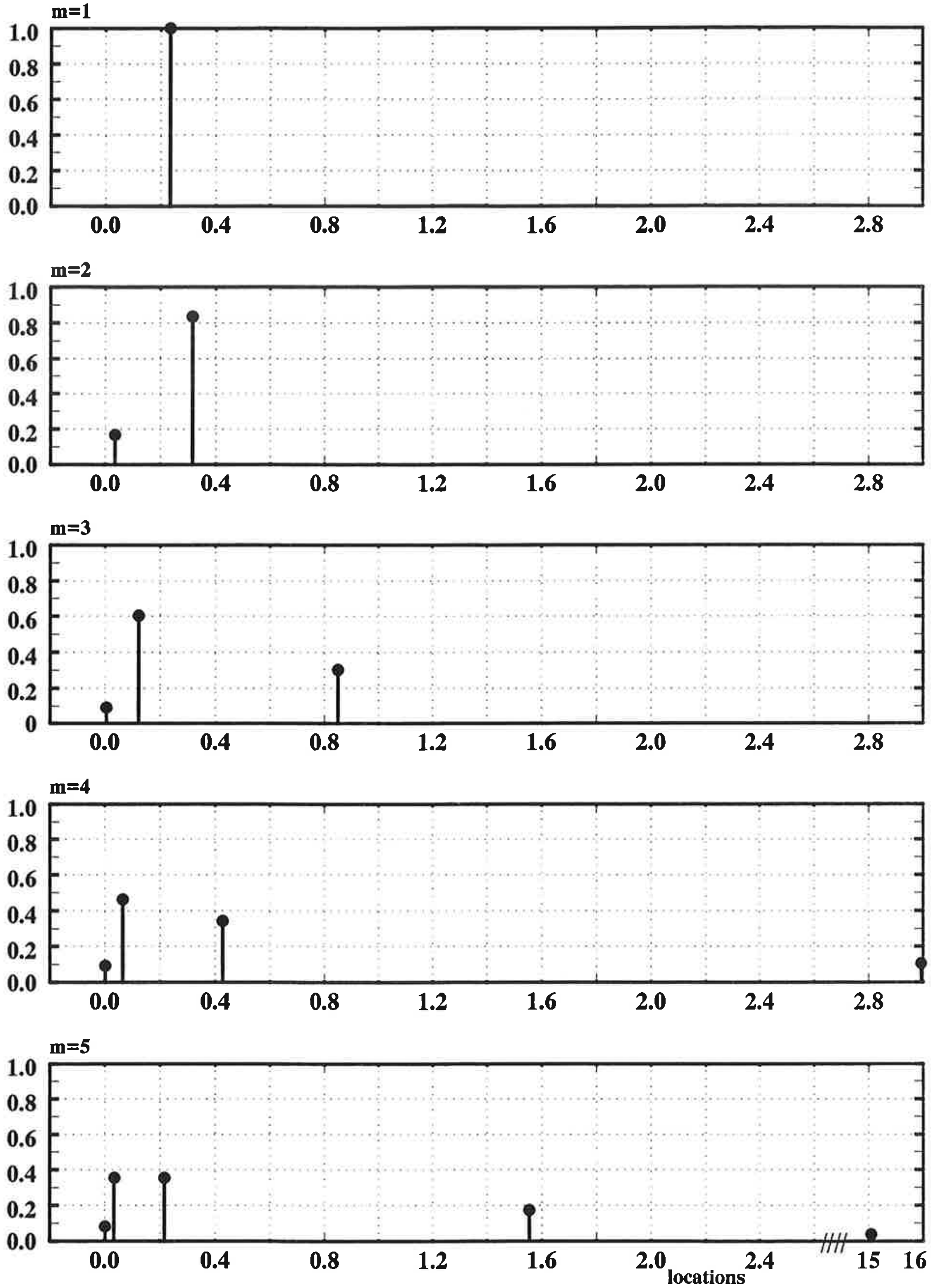


**Fig. 1. Mass point probabilities in a Weibull model**

Table 2. *Mass point heterogeneity in a Weibull model*

Dependent variable: The length of the spell of unemployment

| | Number of mass points m=2 | m=3 | m=4 | m=5 |
|---|---|---|---|---|
| | Std.errors in parentheses | | | |
| Shape parameter | 0.998 | 1.245 | 1.457 | 1.671 |
| | (0.034) | (0.063) | (0.114) | (0.182) |
| Children | -0.004 | -0.035 | -0.020 | -0.034 |
| | (0.059) | (0.080) | (0.095) | (0.108) |
| Married | 0.126 | 0.119 | 0.144 | 0.157 |
| | (0.080) | (0.099) | (0.117) | (0.132) |
| Sex | -0.058 | -0.050 | -0.044 | -0.088 |
| | (0.070) | (0.088) | (0.105) | (0.119) |
| Age | -0.049 | -0.060 | -0.070 | -0.082 |
| | (0.004) | (0.005) | (0.007) | (0.011) |
| Level of education | 0.045 | 0.063 | 0.059 | 0.074 |
| | (0.074) | (0.095) | (0.112) | (0.125) |
| Training for employment | 0.257 | 0.276 | 0.375 | 0.367 |
| | (0.091) | (0.117) | (0.140) | (0.157) |
| Member of UI fund | 0.260 | 0.333 | 0.407 | 0.473 |
| | (0.074) | (0.094) | (0.113) | (0.136) |
| Came from schooling | 0.261 | 0.384 | 0.450 | 0.399 |
| | (0.099) | (0.128) | (0.155) | (0.171) |
| Came from house work | -0.765 | -0.950 | -1.029 | -1.232 |
| | (0.143) | (0.184) | (0.218) | (0.271) |
| Regional demand | 0.221 | 0.396 | 0.542 | 0.432 |
| | (0.274) | (0.348) | (0.408) | (0.464) |
| Occupational demand | 0.233 | 0.038 | -0.529 | 0.020 |
| | (0.736) | (0.932) | (1.129) | (1.268) |
| Taxable assets | 0.781 | 2.166 | 2.124 | 0.553 |
| | (1.176) | (1.546) | (1.930) | (1.877) |
| Replacement ratio | -1.689 | -2.339 | -2.761 | -3.032 |
| | (0.189) | (0.267) | (0.362) | (0.451) |
| $u_1$ | -1.154 | -0.162 | 1.096 | 2.717 |
| | (0.168) | (0.241) | (0.391) | (0.570) |
| $u_2$ | -3.362 | -2.102 | -0.848 | 0.441 |
| | (0.389) | (0.250) | (0.332) | (0.373) |
| $u_3$ | | -5.336 | -2.711 | -1.533 |
| | | (0.781) | (0.376) | (0.414) |
| $u_4$ | | | -6.094 | -3.396 |
| | | | (0.808) | (0.550) |
| $u_5$ | | | | -7.291 |
| | | | | (1.142) |
| $g_1$ | 1.614 | 1.187 | 0.149 | -0.743 |
| | (0.333) | (0.313) | (0.402) | (0.523) |
| $g_2$ | | 1.879 | 1.323 | 0.769 |
| | | (0.269) | (0.290) | (0.308) |
| $g_3$ | | | 1.622 | 1.482 |
| | | | (0.258) | (0.281) |
| $g_4$ | | | | 1.481 |
| | | | | (0.278) |
| $p_1$ | 0.834 | 0.303 | 0.106 | 0.038 |
| $p_2$ | 0.166 | 0.605 | 0.342 | 0.174 |
| $p_3$ | | 0.092 | 0.461 | 0.354 |
| $p_4$ | | | 0.091 | 0.354 |
| $p_5$ | | | | 0.080 |
| Log likelihood | -4929.0 | -4916.5 | -4913.9 | -4912.7 |

## IV. Graphical examination of residuals

A graphical method to examine model misspecification is described and illustrated in this section. The integrated hazards, i.e. generalized residuals of fitted models, derived in the previous sections are examined. Exponential and Weibull models in the absence of censoring have been studied by Lancaster (1983, 1985). Lancaster and Chesher (1985 a,b) have described the construction of residuals for right censored duration data. In this study their product-limit procedure has been applied to residual definitions (11) and (16) to examine model specification when gamma and mass point heterogeneity have been introduced to the model.

Consider any duration distribution with a hazard function $h(\tau;\ \phi)$ depending upon a parameter vector $\phi$. Then the random variable

$$I(T) = \int_0^T h(\tau;\ \phi)d\tau \qquad (17)$$

has a unit exponential distribution since at any point in time $t > 0$ the survivor function is

$$e^{-I(t)} = P(T > t)$$

$$= P[I(T) > I(t)]. \qquad (18)$$

Thus for every $I(t) \geq 0$, the survivor function $P[I(T) > I(t)] = \exp[I(t)]$, which is the survivor function of unit exponential distribution. The moments of I(T) are $E[I(T)^q] = q!$, q = 1, 2,... The definition of generalized residuals $\hat{I}(T_j)$ in the absence of censoring is given by Cox and Snell (1968) and in the Weibull case the residuals are

$$\hat{I}(T_j) = T_j^{\hat{\alpha}} e^{x_j\hat{\beta}}, \quad j = 1,\ldots,n, \qquad (19)$$

where the $\hat{}$ indicates Maximum Likelihood estimates and n is the size of the sample. Hence, if the opposite of the logarithm of the residual survival function is plotted against the ordered sequence of the residuals, it should give approximately a straight plot on a 45° line through the origin. For graphical plots when a Weibull model with gamma heterogeneity are fitted to data the residuals are

$$\hat{I}(T_j) = 1/\sigma^2 \log[1 + \sigma^2 T_j^{\hat{\alpha}} e^{x_j \hat{\beta}}]. \tag{20}$$

and the residuals for the mass point models are

$$\hat{I}(T_j) = \log\{[\sum_{i=1}^{m} \hat{p}_i e^{-T_j^{\hat{\alpha}} e^{\hat{u}_i + x_j \hat{\beta}}}]^{-1}\}. \tag{21}$$

With censoring t = min(T, Z), where Z is a censoring time. If the model is correct $I(t_j)$, j=1,...,n, approximate a censored random sample from the unit exponential distribution, where the approximation is due to use of the estimated values instead of the true ones. Now I(t) has not a unit Exponential distribution, because its distribution depends on that of the censoring time. However, it is possible to define a set of residuals which do have simple properties under correct specification.

In the case of right censored observations a product-limit procedure suggested by Lancaster and Chesher (1985 a,b) can be used to estimate the survivor function of residuals and this is distributed unit exponential when the model is correct. Consider the ordered sequence of residuals. The hazard function of residuals can be calculated for each residual corresponding to an uncensored observation at that value as the ratio of the number of residuals with value equal to the particular residual and the number of residuals greater than or equal to it. Let this ratio for the $s^{th}$ ordered uncensored residual be $\hat{h}(\hat{I}_s)$. Then the product-limit estimate of the residual survivor function is

$$\hat{S}(I_j) = \prod_{s=0}^{j-1} [1 - \hat{h}(I_s)], \qquad (22)$$

and minus the logarithm of the residual survivor function is given by

$$- \log \hat{S}(I_j) = - \sum_{s=0}^{j-1} \log[1 - \hat{h}(I_s)]. \qquad (23)$$

The plot of the opposite of the logarithm of the residual survivor function should give a 45° line through the origin in large samples, when the model is right.

The residual plots for the models estimated in sections II and III are in Figure 2. It should be noted that the departure from the 45° line is larger for high values of the residuals. Thus the graphical method reveals best the right tail behaviour of the duration distribution. A plot above (below) the 45° line indicates that the estimated hazard function is too low (high). The behaviour of residuals seems to be slightly better after allowing for gamma heterogeneity. In mass point models no specific distribution assumption is required for the unobserved heterogeneity. Thus the risk of misspecification is reduced. The departure from the 45° line decreases when the number of mass points has been increased and the plots are fairly precisely on the 45° line in the last graphs.

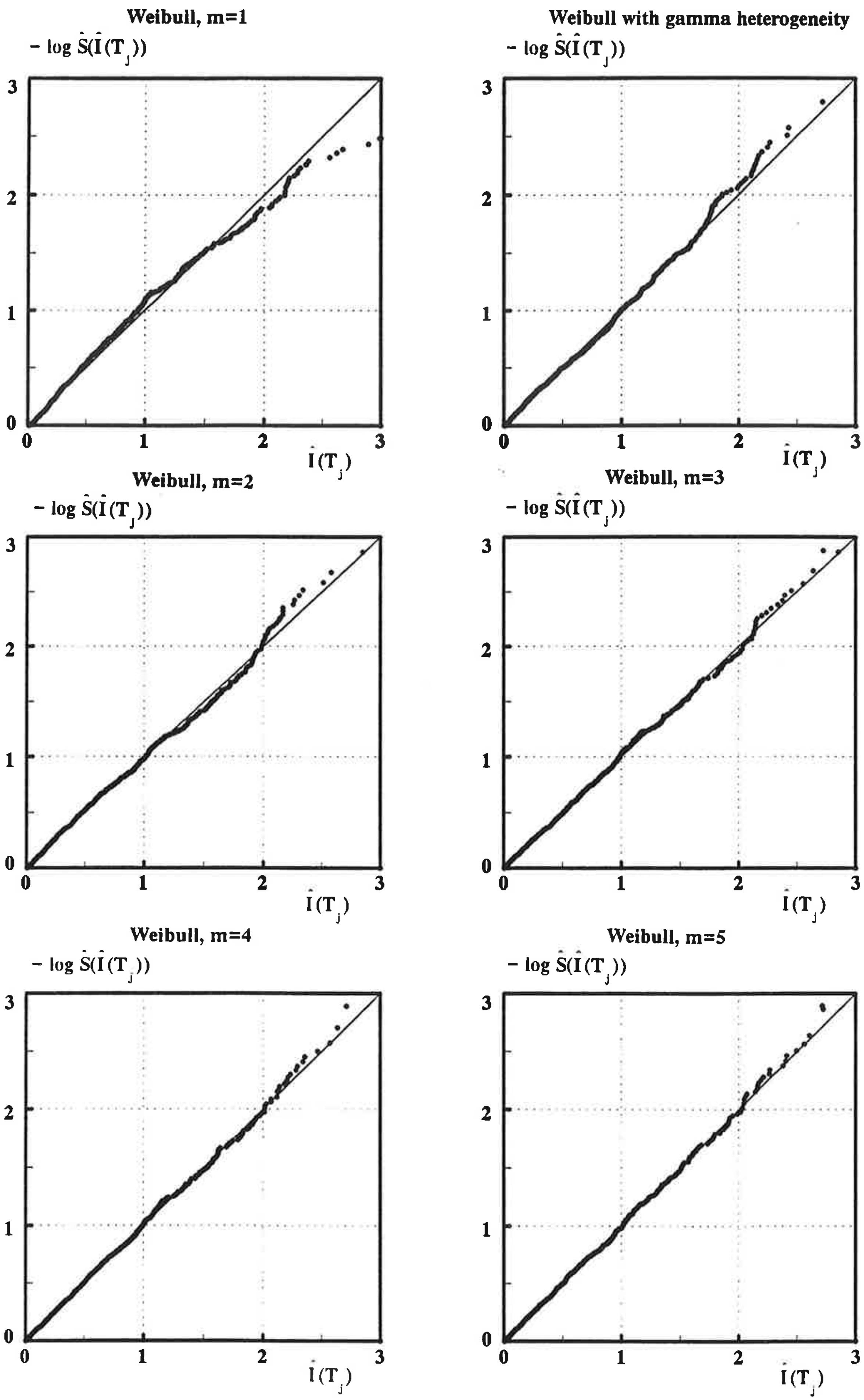


**Fig. 2. Residual plots of duration models**

## V. Conclusions

In this paper the models of unemployment duration allowing for individual heterogeneity were studied. Weibull models with gamma and mass point heterogeneity were estimated using Finnish microeconomic data. In the basic Weibull model the estimated value of the shape parameter was less than one indicating negative duration dependence. However, the parameter estimates of the basic Weibull model were biased. The absolute values of the parameter estimates increased substantially after allowing for heterogeneity. In the Weibull model with gamma heterogeneity and in mass point models with three or more points of support the parameter estimate was larger than one indicating an increasing hazard function for an individual. These results are in concordance with standard search theories.

The residuals of estimated heterogeneity models were derived and examined by a graphical method. The conclusion drawn from the graphical examination of residuals is that correction of heterogeneity of duration models is of great importance even with fairly rich and reliable data. The basic Weibull model did not fit very well, but after introducing gamma heterogeneity the model was slightly better specified. However, allowing for mass point heterogeneity made the specification even better. Five mass points were enough to rectify the effects of omitted variables.

**References**


Berndt, E.R. & Hall, B.H. & Hall, R.E. & Hausman, J.A.: Estimation and Inference in Nonlinear Structural Models, *Annals of Economic and Social Measurement 3*, 653-66, 1974.

Brännäs, K.: Small Sample Properties in a Heterogeneous Weibull Model, *Economics Letters 21*, 17-20, 1986a.

Brännäs, K.: On Heterogeneity in Econometric Duration Models, *Sankhya 48*, 284-93, 1986b.

Brännäs, K. & Rosenqvist, G.: Semiparametric Estimation of Heterogenous Count Data Models, *Swedish School of Economics and Business Administration, Working Papers No. 187*, 1988.

Chesher, A.: Testing for Neglected Heterogeneity, *Econometrica 52*, 865-72, 1984.

Card, D. & Sullivan, D.: Measuring the Effect of Subsidized Training Programs on Movements in and out of Employment, *Econometrica 56*, 497-530, 1988.

Cox, D.R. & Snell, E.J.: A General Definition of Residuals, *Journal of the Royal Statistical Society, Series B, 30*, 248-75, 1968.

Davies, R.B.; Mass Point Methods for Dealing with Nuisance Parameters in Longitudinal Studies, in R. Crouchley (ed.), *Longitudinal Data Analysis, Surrey Conference on Sociological Theory and Method 4*, Gower Publishing Company Limited, Avebury, 1987.

Davies, R.B. & Crouchley, R.: Calibrating Longitudinal Models of Residential Mobility and Migration, *Regional Science and Urban Economics, 14*, 231-47, 1984.

Dunn, R. & Reader, S. & Wrigley, N.: A Nonparametric Approach to the Incorporation of Heterogeneity into Repeated Polytomous Choice Models of Urban Shopping Behaviour, *Transportation Research 21A*, 327-43, 1987.

Engström, L. & Löfgren, K.-G.: Disguised and Open Unemployment Intensified Employment Service and Unemployment Durations, *Trade Union Institute for Economic Research, Working Paper Series, No. 39*, Stockholm, 1987.

Ham, J.C. & Rea, S.A.Jr: Unemployment Insurance and Male Unemployment Duration in Canada, *Journal of Labor Economics 5*, 325-53, 1987.

Heckman, J.J. & Singer, B.: Econometric Duration Analysis, *Journal of Econometrics 24*, 63-132, 1984a.

Heckman, J.J. & Singer, B.: A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data, *Econometrica 52*, 271-320, 1984b.

Kaplan, E.L. & Meier, P.: Nonparametric Estimation from Incomplete Observations, *Journal of the American Statistical Association 53*, 457-81, 1958.

Kettunen, J.: Työttömyysturvan vaikutukset työnetsintään, *The Research Institute of the Finnish Economy, C 49*, Helsinki, 1989.

Kiefer, J. & Wolfowitz, J.: Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters, *Annals of Mathematical Statistics 27*, 887-906, 1956.

Kiefer, N.M. & Neumann, G.R.: *Search Models and Applied Labor Economics*, Cambridge University Press, Cambridge, 1989.

Kooreman, P. & Ridder, G.: The Effects of Age and Unemployment Percentage on the Duration of Unemployment, *European Economic Review 20*, 41-57, 1983.

Laird, N.: Nonparametric Maximum Likelihood Estimation of a Mixing Distribution, *Journal of the American Statistical Association* 73, 783-92, 1978.

Lancaster, T.: Econometric Methods for the Duration of Unemployment, *Econometrica 47*, 939-56, 1979.

Lancaster, T.: Generalised Residuals and Heterogeneous Duration Models: The Exponential Case, *Bulletin of Economic Research 35*, 71-85, 1983.

Lancaster, T.: Generalized Residuals and Heterogeneous Duration Models, With Applications to the Weibull Model, *Journal of Econometrics 28*, 155-69, 1985.

Lancaster, T. & Chesher, A.: Residual Analysis for Censored Duration Data, *Economic Letters 18*, 35-8, 1985a.

Lancaster, T. & Chesher, A.: Residuals, Tests and Plots with a Job Matching Illustration, *Annals de L'Insee, No 59/60*, 47-70, 1985b.

Lancaster, T. & Nickell, S.: The Analysis of Re-Employment Probabilities for the Unemployed, *Journal of the Royal Statistical Society, Series A, 143*, 141-65, 1980.

Lindsay, B.G.: The Geometry of Mixture Likelihoods: A General Theory, *The Annals of Statistics 11*, 86-94, 1983a.

Lindsay, B.G.: The Geometry of Mixture Likelihoods, Part II: The Exponential Family, *The Annals of Statistics 11*, 783-92, 1983b.

Narendranathan, W. & Nickell, S. & Stern, J.: Unemployment Benefits Revisited, *The Economic Journal 95*, 307-29, 1985.

Newman, J.L. & McCulloch C.E.: A Hazard Rate Approach to the Timing of Births, *Econometrica 52*, 939-61, 1984.

Nickell, S.: The Effect of Unemployment and Related Benefits on the Duration of Unemployment, *The Economic Journal 89*, 34-49, 1979a.

Nickell, S.: Estimating the Probability of Leaving Unemployment, *Econometrica 47*, 1249-66, 1979b.

*SAS/IML, User's Guide, Version 5 Edition*, SAS Institute Inc., Cary, North Carolina, 1985.

Simar, L.: Maximum Likelihood Estimation of a Compound Poisson Process, *The Annals of Statistics 4*, 1200-9, 1976.

Trussell, J. & Richards, T.: Correcting for Unmeasured Heterogeneity in Hazard Models Using the Heckman-Singer Procedure, *Sociological Methodology*, 242-76, 1985.

## Appendix. Variables of the data

*Duration of unemployment* is calculated in weeks and it is the difference between the date of entry and the date of leaving unemployment. Mean = 15.03.

*Number of children* is the number of unemployed person's children who are younger than 18 years. Mean = 0.23.

*Married* is a dummy variable, 1=yes. Mean = 0.37.

*Sex* is a dummy variable, 1=male. Mean = 0.54.

*Age* is measured in years. Mean = 31.2.

*Level of education* is a dummy variable, 1 = at least 12 years education. The level of education is based on the education code of the Central Statistical Office of Finland. Mean = 0.45.

*Training for employment* is a dummy variable, 1 = The person has got training for further employment. Mean = 0.15.

*Member of UI fund* is a dummy variable, 1 = yes. Mean = 0.42.

*Came from schooling* is a dummy variable, 1 = The person has come from schooling or from the army. Mean = 0.13.

*Came from house work* is a dummy variable, 1 = The person has come from home or elsewhere outside the labour force. Mean = 0.07.

*Regional demand* describes the regional rate of jobs available. It is the number of vacancies divided by the number of job seekers in the area. Mean = 0.10.

*Occupational demand* describes the occupational rate of jobs available in the whole country. It is the number of vacancies divided by the number of job seekers in the occupation group. Mean = 0.12.

*Taxable assets* has been compiled from the tax register and it is measured in millions of marks. Mean = 0.011.

*Replacement ratio* is unemployed persons average replacement ratio of unemployment benefits during the unemployment period after tax. Weekly unemployment benefits after tax have been divided by the weekly income after tax. Mean = 0.17.

ELINKEINOELÄMÄN TUTKIMUSLAITOS (ETLA)
THE RESEARCH INSTITUTE OF THE FINNISH ECONOMY
LÖNNROTINKATU 4 B, SF-00120 HELSINKI

Puh./Tel. (90) 601 322 Telefax (90) 601 753
Int. 358-0-601 322 Int. 358-0-601 753

KESKUSTELUAIHEITA - DISCUSSION PAPERS ISSN 0781-6847

No 326 JEAN MALSOT, Rapport du printemps 1990 - Perspectives à moyen terme pour l'économie européenne (Euroopan keskipitkän aikavälin näkymät). 08.06.1990. 31 p.

No 327 HILKKA TAIMIO, Naisten kotityö ja taloudellinen kasvu Suomessa vuosina 1860-1987, uudelleenarvio. 20.06.1990. 56 s.

No 328 TOM BERGLUND - STAFFAN RINGBOM - LAURA VAJANNE, Pricing Options on a Constrained Currency Index: Some Simulation Results. 28.06.1990. 43 p.

No 329 PIRKKO KASANEN, Energian säästö ympäristöhaittojen vähentämiskeinona, päätöksentekokehikko energian ympäristöhaittojen vähentämiskeinojen vertailuun. 01.07.1990. 41 s.

No 330 TOM BERGLUND - KAJ HEDVALL - EVA LILJEBLOM, Predicting Volatility of Stock Indexes for Option Pricing on a Small Security Market. 01.07.1990. 20 p.

No 331 GEORGE F. RAY, More on Finnish Patenting Activity. 30.07.1990. 9 p.

No 332 KARI ALHO, Odotetun EES-ratkaisun ja Suomen linjan taloudelliset perustelut. 01.08.1990. 10 s.

No 333 TIMO MYLLYNTAUS, The Role of Industry in the Electrification of Finland. 14.08.1990. 35 p.

No 334 RISTO MURTO, The Term Structure and Interest Rates in the Finnish Money Markets - The First Three Years. 17.08.1990. 27 p.

No 335 VEIJO KAITALA - MATTI POHJOLA - OLLI TAHVONEN, An Economic Analysis of Transboundary Air Pollution between Finland and the Soviet Union. 01.10.1990. 23 p.

No 336 TIMO MYLLYNTAUS, Ympäristöhistorian tutkimus Suomessa. 08.10.1990. 35 p.

No 337 KÅRE P. HAGEN - VESA KANNIAINEN, The R&D Effort and Taxation of Capital Income. 15.10.1990. 34 p.

No 338 PEKKA YLÄ-ANTTILA - RAIMO LOVIO, Flexible Production, Industrial Networks and Company Structure - Some Scandinavian Evidence. 25.10.1990. 19 p.

Elinkeinoelämän Tutkimuslaitoksen julkaisemat "Keskusteluaiheet" ovat raportteja alustavista tutkimustuloksista ja väliraportteja tekeillä olevista tutkimuksista. Tässä sarjassa julkaistuja monisteita on rajoitetusti saatavissa ETLAn kirjastosta tai ao. tutkijalta.
Papers in this series are reports on preliminary research results and on studies in progress; they can be obtained, on request, by the author's permission.

E:\sekal\DPjulk.chp/31.01.1991