

Hess, Wolfgang

Working Paper

A Flexible Hazard Rate Model for Grouped Duration Data

Working Paper, No. 2009:18

Provided in Cooperation with:

Department of Economics, School of Economics and Management, Lund University

Suggested Citation: Hess, Wolfgang (2009) : A Flexible Hazard Rate Model for Grouped Duration Data, Working Paper, No. 2009:18, Lund University, School of Economics and Management, Department of Economics, Lund

This Version is available at:

<https://hdl.handle.net/10419/259977>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

A Flexible Hazard Rate Model for Grouped Duration Data*

Wolfgang Hess[†]

November 10, 2009

Abstract

This paper proposes a discrete-time hazard regression approach based on the interrelation between hazard rate models and excess over threshold models, which are frequently encountered in extreme value modelling. The proposed duration model incorporates a grouped-duration analogue of the well-known Cox proportional hazards model and a proportional odds model as special cases. The theoretical setup of the model is motivated, and simulation results are reported to suggest that it performs well. A numerical example using US unemployment data is also provided.

JEL Classification: C41; J64.

Keywords: Discrete-Time Duration Model, Hazard Rate, Threshold Excess Model, Unemployment Duration.

1 Introduction

The interrelation between binary response regression models and discrete-time duration models has long been known in the literature. In particular, it can be shown that a stacked binary choice model employing a complementary log-log (cloglog) link function and including period-specific intercepts represents the exact grouped-duration analogue of the well-known Cox (1972) proportional hazards (PH) model

*Previous versions of this paper were presented at the 2008 Economic Demography conference in Mölle, at the 2009 Nordic Econometric Meeting in Lund, and at seminars at Lund University and VU University Amsterdam. The author would like to thank conference and seminar participants, and in particular David Edgerton, Peter Jochumzen, Barry Jones, Maarten Lindeboom, Gerard van den Berg, Bas van der Klaauw, Johan Vikström, and Joakim Westerlund for many valuable comments and suggestions. Thank you also to the Jan Wallander and Tom Hedelius foundation for financial support under research grant numbers H2007-0281:1 and P2009-0189:1.

[†]Department of Economics, Lund University, P.O. Box 7082, S-220 07 Lund, Sweden. Tel.: +46 46 222 8653; Fax: +46 46 222 4118; E-mail: Wolfgang.Hess@nek.lu.se.

(see e.g. Kalbfleisch and Prentice, 1973, or Prentice and Gloeckler, 1978). Due to its direct relationship to an underlying continuous-time process, the stacked cloglog model has been widely used in the literature to analyze grouped duration data.

The application of duration models is most widespread in the field of labor economics, where job search models provide the theoretical foundation for so-called reduced form analyses of individual unemployment spells (see van den Berg, 2001, for a comprehensive survey of this literature). In this context the PH assumption implied by the cloglog model is not supported by economic theory (see e.g. van den Berg, 1990a,b, or Blanchard and Diamond, 1994) and empirically questionable (see McCall, 1994). In a discrete-time framework, a natural alternative to the PH model is to estimate sequential logit or probit models with observations stacked across interval durations. This method is very appealing because of its simplicity. However, there are two major drawbacks to this approach. First, the choice of a logit or probit link function for the binary regression model is rather *ad hoc*, and little is known about the underlying continuous-time processes leading to these specifications.¹ Second, the estimated effects of covariates on the hazard rate may differ largely across specifications, even if the overall fit in likelihood terms does not allow to discriminate between the models (see Narendranathan and Stewart, 1993, and Sueyoshi, 1995). In that case the researcher is left with little or no idea of which model to choose.

This paper considers a grouped-duration hazard model where transition probabilities are given by a parametric family of distributions. The proposed hazard model contains both logistic and extreme value regression models as special cases, allowing the choice between logit and cloglog to be reduced to the estimation of a single additional parameter. In biostatistics it has long been common to fit such flexible families of models in connection with binary response regression (e.g. Prentice, 1976, and Aranda-Ordaz, 1981). In the social sciences, the use of such location-scale models in binary choice modelling was first promoted by Nagler (1994). This paper extends this approach to the analysis of grouped duration data. An important contribution of this paper is to show that the class of discrete-time duration models considered can be linked to an underlying continuous-time process. Specifically, the parametric specification of the grouped-duration hazard rate is deduced from the asymptotic distribution of threshold excesses of a continuous duration variable. For ease of exposition, focus will be placed on unemployment durations throughout the paper, although the suggested methods are applicable to many other types of durations that are grouped into intervals.

The remainder of this paper is organized as follows. Section 2 introduces the hazard model and investigates its properties from a theoretical angle. Section 3

¹Both Heckman and Singer (1984) and Lancaster (1990) argue that it is preferable to work in continuous time and translate to discrete time if necessary.

contains simulation experiments analyzing the performance of the proposed model. The simulations focus mainly on the model's ability to discriminate between the widely used cloglog and logit specifications of the hazard rate. Section 4 contains an empirical application using US unemployment data, and Section 5 concludes.

2 The Hazard Rate Model

This section introduces the hazard rate model. A brief introduction of the grouped-duration framework is followed by an extensive discussion of the hazard specification. The properties of the proposed hazard rate model are analyzed at the end of the section.

2.1 The Grouped-Duration Framework

Although unemployment durations are always reported in discrete units (mostly weeks or months) and are therefore integer valued, the predominant part of the econometric literature focuses on continuous-time models in which the durations can take on all values on the positive real line. In continuous time the focal point of the analysis is the instantaneous exit rate out of unemployment which is defined by the hazard function. In standard job search models the hazard rate from unemployment is determined by the joint probability of receiving a job offer and the (conditional) probability of accepting it. Thus, job search models provide a theoretical framework for reduced-form analyses of unemployment durations.

Formally, let T_i be a continuous, non-negative random variable denoting the length of individual i 's unemployment spell. Initially it is assumed that the random variables T_i are independent and identically distributed with common distribution function $G(t)$. Then the hazard of any individual at time t is defined as

$$h(t) := \lim_{dt \rightarrow 0^+} \frac{P(t \leq T_i < t + dt | T_i \geq t)}{dt}. \quad (1)$$

In a discrete-time framework, the core of the analysis is formed by an individual's probability of leaving unemployment in a given time interval $[t_k, t_{k+1})$, $k = 1, 2, \dots, k^{max}$, and $t_1 = 0$, conditional on being unemployed up to the beginning of the interval. This conditional probability is termed the discrete-time hazard rate and formally defined as

$$h_k := P(t_k \leq T_i < t_{k+1} | T_i \geq t_k) = 1 - \exp \left\{ - \int_{t_k}^{t_{k+1}} h(t) dt \right\}. \quad (2)$$

The grouped-duration hazard rate given in (2) is typically parameterized as

$$h_k = F(\gamma_k),$$

where γ_k is a function of (interval) time that allows the hazard rate to vary across periods. Somewhat loosely, γ_k will be referred to as the grouped-duration baseline hazard in the following, although this term will not be formally correct in all instances.² Since there is usually little *a priori* knowledge about the duration dependence of the hazard rate, it is common to model γ_k in the most flexible possible way by means of period-specific dummy variables. However, a (flexible) functional specification for γ_k may also be chosen to reduce the number of parameters in the model. Since the discrete-time hazard is a conditional probability, the function $F(\cdot)$ needs to be chosen such that $0 \leq h_k \leq 1$ for all k . Obvious choices are the cumulative distribution functions of the standard normal or the logistic distribution. The hazard rate can then be estimated using conventional stacked probit or stacked logit regression models. While this approach is very appealing due to its simplicity, it suffers from the drawback that the choice of a stacked probit or logit model is rather *ad hoc*, and little is known about the underlying continuous-time processes leading to these grouped-duration specifications. Moreover, as shown by Sueyoshi (1995), the choice of distributional form for the binary response model is not innocuous in a duration context. Therefore, the following section proposes a flexible functional specification of the discrete-time hazard, which is motivated by the asymptotic distribution of threshold excesses of the underlying individual duration times T_i .

2.2 Functional Specification of the Hazard Rate

The derivation of the hazard specification requires two assumptions about the distribution of duration times, $G(t)$, which is directly linked to the grouped hazard through the relation

$$h_k = \frac{G(t_{k+1}) - G(t_k)}{1 - G(t_k)}.$$

First, it is assumed that $G(t)$ is continuous and has unbounded support on $[0, \infty)$. Second, it is assumed that $G(t)$ belongs to the domain of attraction of any one of the extreme value distributions.³

The requirement that $G(t)$ have unbounded support on the positive real line is needed to ensure that the discrete hazard is “non-defective”, i.e. to ensure that $h_k(z) \rightarrow 1$ as $z \rightarrow \infty$. The second assumption is novel in the context of discrete-time duration analysis. It requires that $G(t)$ belongs to the maximum domain of

²Only if γ_k is modelled as an interval-specific step function, and if an extreme value minimum distribution is chosen to specify $F(\cdot)$, γ_k can be shown to be the exact grouped-duration analogue of the continuous-time Cox (1972) baseline hazard (see e.g. Kalbfleisch and Prentice, 1973, or Prentice and Gloeckler, 1978, for a formal proof).

³Formally, this second assumption requires that there exist sequences of constants $\{a_n\}$ and $\{b_n\}$, with $a_n > 0$ for all n , and a non-degenerate distribution function $H(z)$ such that for $M_n = \max_{1 \leq i \leq n} (T_i)$, $\lim_{n \rightarrow \infty} P((M_n - b_n)/a_n \leq z) = H(z)$ for all z at which $H(z)$ is continuous (see Pickands, 1975).

attraction of some non-degenerate function $H(z)$. Then, as shown by Fisher and Tippett (1928), $H(z)$ necessarily belongs to one of the extreme value distributions, with types I, II, and III widely known as the Gumbel, Fréchet, and Weibull families, respectively. This limit theorem for maxima is similar in scope to the central limit theorem for averages, and valid for the vast majority of common distribution functions. In particular, the theorem applies to the exponential, Weibull, Gamma, log-normal, and Burr distributions which are the commonly encountered parametric specifications in duration modelling. Given these assumptions, a functional specification for the grouped-duration hazard can be derived using well-known results from extreme value theory.

As first shown by Pickands (1975), the generalized Pareto distribution arises as a limiting distribution for excesses over thresholds, if the parent distribution is continuous and belongs to the domain of attraction of an extreme value distribution. Specifically, for any random variable T_i with distribution function, $G(t)$, fulfilling the two assumptions above, and for a given large threshold τ , the conditional distribution of T_i given that $T_i \geq t_k \geq \tau$, which can be expressed as $P(t_k \leq T_i < t_k + z | T_i \geq t_k)$, is very nearly of the form

$$1 - \left(1 + \frac{\xi z}{\sigma(t_k)}\right)^{-1/\xi}.$$

The above expression describes the generalized Pareto distribution with scale parameter $\sigma > 0$ and shape parameter ξ . With $G(t)$ having unbounded support on the positive real line, it holds that $\xi \geq 0$ and $0 < z < \infty$. While σ depends on the threshold level t_k , it can be shown that ξ is constant for all t_k above a level τ at which the asymptotic motivation for the generalized Pareto distribution is valid (see e.g. Leadbetter *et al.*, 1983, or Embrechts *et al.*, 1997, for extensive surveys on the generalized Pareto distribution). Setting $z = 1$ leads – under the common assumption of equal unit interval length – to the discrete-time hazard rate representation

$$h_k = P(t_k \leq T_i < t_k + 1 | T_i \geq t_k) = 1 - \left(1 + \frac{\xi}{\sigma(t_k)}\right)^{-1/\xi}. \quad (3)$$

So far, the specification in (3) does not explicitly allow for individual variation in the hazard rate. A main objective of econometric duration analysis is, however, to estimate the effects of explanatory variables on exit probabilities. As in standard extreme value models, explanatory variables can be incorporated into the model by writing

$$h_{ik} = 1 - \left(1 + \frac{\xi}{\sigma(\mathbf{x}'_{ik}\boldsymbol{\beta} + \gamma_k)}\right)^{-1/\xi}$$

for some positive-valued function $\sigma(\cdot)$ and some vector of (possibly time-varying)

covariates \mathbf{x}_{ik} .⁴ In view of the requirement $\sigma > 0$, one natural formulation is $\sigma(\mathbf{x}'_{ik}\boldsymbol{\beta} + \gamma_k) = \exp\{-(\mathbf{x}'_{ik}\boldsymbol{\beta} + \gamma_k)\}$, resulting in the hazard rate specification

$$h_{ik} = 1 - \left(1 + \frac{\xi}{\exp\{-(\mathbf{x}'_{ik}\boldsymbol{\beta} + \gamma_k)\}}\right)^{-1/\xi}. \quad (4)$$

Due to its relation to the generalized Pareto distribution, the discrete-time duration model resulting from the functional specification in (4) will be referred to as the *Pareto hazard model* in the following.

Estimates for the parameters $\boldsymbol{\beta}$, γ_k , and ξ can be obtained as in conventional binary choice models for panel data. The resulting log-likelihood function for a sample of n individuals is

$$\ln \mathcal{L} = \sum_{i=1}^n \sum_{k=1}^{k_i} [y_{ik} \ln(h_{ik}) + (1 - y_{ik}) \ln(1 - h_{ik})], \quad (5)$$

where h_{ik} is specified as in (4), y_{ik} is a binary indicator variable taking the value one if individual i 's unemployment spell is observed to cease during the k^{th} time interval, and zero otherwise, and k_i denotes the last interval where individual i is observed.⁵ As opposed to conventional hazard specifications such as cloglog or logit, the Pareto hazard rate contains a shape parameter, ξ , which makes the model considerably more flexible with respect to the imposed effects of covariates on exit probabilities. In the following, the properties of the Pareto hazard model are investigated in more detail.

2.3 Properties of the Pareto Hazard Model

The Pareto hazard model specified in (4) contains the two most commonly applied discrete-time hazard models (cloglog and logit) as special cases. To see this, note that

$$\lim_{\xi \rightarrow 0} \left(1 - \left(1 + \frac{\xi}{\exp\{-(\mathbf{x}'_{ik}\boldsymbol{\beta} + \gamma_k)\}}\right)^{-1/\xi}\right) = 1 - \exp\{-\exp\{\mathbf{x}'_{ik}\boldsymbol{\beta} + \gamma_k\}\},$$

which is then the definition of h_{ik} for $\xi = 0$. The case $\xi = 1$ yields the logistic hazard rate

$$\frac{1}{1 + \exp\{-(\mathbf{x}'_{ik}\boldsymbol{\beta} + \gamma_k)\}}.$$

⁴A detailed discussion of covariate modelling in the context of threshold excess models is provided by Davison and Smith (1990).

⁵To obtain consistent parameter estimates from this log-likelihood, spells must be independent, censoring must occur only at interval boundaries, and censoring must not provide any information about T_i beyond that available in the covariates (see e.g. Allison, 1982, Singer and Willett, 1993, or Jenkins, 1995, for excellent surveys on the derivation of the likelihood function).

Due to its direct relationship to an underlying continuous-time process, the cloglog model has been widely used in the literature to analyze grouped duration data. Indeed, the cloglog model is sometimes regarded as a more sophisticated alternative to conventional probit- or logit-type models with observations pooled across interval durations (Diamond and Hausman, 1984). But due to its familiarity and simplicity of computation, the logit model has received increasing attention in both applied and theoretical work. In unemployment duration studies, the choice of the logit specification is also sometimes motivated by the fact that it does not imply the theoretically unsupported proportional hazards assumption (e.g. Steiner, 2001). However, the logit model implies the rather restrictive proportional odds assumption which is also lacking a solid theoretical foundation. Since the Pareto model nests both the cloglog and the logit model, the choice between these specifications can be reduced to the estimation of the shape parameter ξ . Moreover, the Pareto model does neither impose proportional hazards nor proportional odds and allows for covariate effects that are rather different from both the cloglog and the logit model.

The functional specification of h_{ik} in equation (4) is given by the cumulative distribution function of the log-Burr distribution (see Burr, 1942, or Tadikamalla, 1980). Its density function is left-skewed if $\xi < 1$, right-skewed if $\xi > 1$, and symmetric in the intermediate case $\xi = 1$. The log-Burr distribution is a special case of the type IV generalized logistic distribution (see e.g. Johnson *et al.*, 1995), which has been proposed by Prentice (1975, 1976) for discrimination between parametric specifications in a binary response model context.⁶ Prentice has shown that ξ can be consistently estimated along with the structural parameters by plugging (4) into (5) and maximizing the log-likelihood function. The maximum likelihood estimate $\hat{\xi}$ can then be used to test the suitability of submodels relative to (4) by means of a simple Wald test. In particular, in the logistic case, $\xi = 1$, the asymptotic distribution of $\hat{\xi}$ is normal with mean 1 and variance $4(\pi^2 + 3)/(N(\pi^2 - 6))$, where N denotes the total number of observations (person periods) in the sample. At the boundary, $\xi = 0$, the asymptotic distribution of $\hat{\xi}$ is equal to the distribution of a random variable defined as

$$\xi^{trunc} = \begin{cases} \xi^* & \text{if } \xi^* \geq 0 \\ 0 & \text{if } \xi^* < 0, \end{cases}$$

where $\xi^* \sim N(0; \pi^2/(N(\pi^2 - 6)))$.

The conventional approach to analyzing grouped duration data is to choose between a cloglog, logit, or (less often) probit specification of the hazard rate, where the researcher has little *a priori* information about the adequacy of the different specifications for the data under consideration. However, the choice of distributional

⁶Kalbfleisch and Prentice (1980) consider this distribution for the parametric specification of continuous-time duration models.

form for the binary response model is not innocuous in a duration context, and will have important implications for the effects of covariates on exit probabilities (see e.g. Narendranathan and Stewart, 1993, or Sueyoshi, 1995). For discrete-time models, a natural measure of the effects of explanatory variables is derived by examining the derivative of the log hazard rate with respect to an arbitrary outcome at various interval durations,

$$\frac{\partial \ln h(\mathbf{x}'\boldsymbol{\beta} + \gamma_k)}{\partial x_m} = \beta_m \frac{h'(\mathbf{x}'\boldsymbol{\beta} + \gamma_k)}{h(\mathbf{x}'\boldsymbol{\beta} + \gamma_k)},$$

where h' denotes the first derivative of h . This expression provides a measure of the proportionate change in the hazard rate resulting from changes in the explanatory variables. In continuous time, the assumption of proportional hazards would ensure that the derivative of the log hazard rate remains constant over time. This result, however, does not hold for the discrete-time PH model which does not exhibit proportional interval hazards.⁷ The relevant issue in discrete-time duration analysis is rather whether the logarithmic derivative of the hazard is roughly constant across intervals.

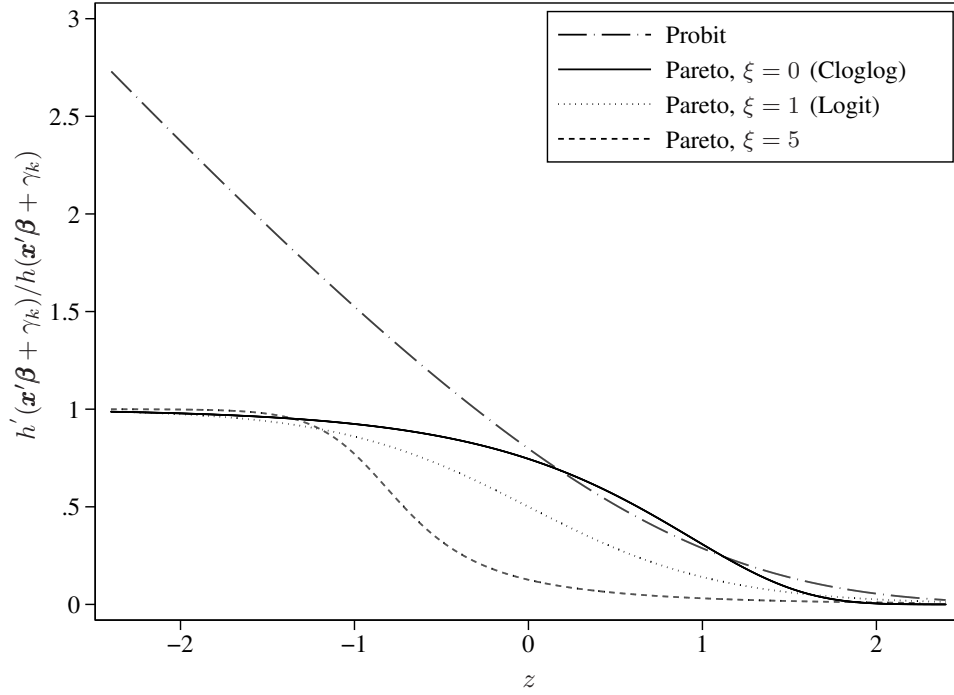
For various functional specifications, Figure 1 plots the logarithmic derivative of the hazard rate against standardized values of the aggregator function. These standardized values are calculated as $z = (\mathbf{x}'\boldsymbol{\beta} + \gamma_k)/\sigma - \mu/\sigma$, where μ and σ denote the mean and standard deviation of the respective distribution functions. For reasons of comparison, Figure 1 also includes the probit hazard, although this functional specification is not nested in the Pareto hazard model. The degree of non-proportionality inherent in the different hazard models is indicated by the (absolute) value of the respective slopes.⁸ From Figure 1 it is apparent that the slopes for the different specifications differ considerably at various values of z . In particular, the slope of the probit curve differs considerably from the other specifications for small values of z . Thus, the conventional wisdom regarding the similarity of probit and logit models does not extend to the evaluation of proportionality in the discrete hazard, nor does it apply to the probit and cloglog specifications. As a result, for duration data exhibiting low to moderate exit rates, the discrete-time hazard rate models employing a probit specification will tend to depart far more from proportionality than logit models, which in turn will be slightly less proportional than cloglog models. Logit and cloglog models diverge most significantly at the center of the distributions. Since the Pareto hazard model nests both the logit and the cloglog model, it can exhibit

⁷A model with proportional interval hazards would be obtained from the Pareto model with $\xi = -1$, if the class of binary response models considered could be extended in a regular manner to include negative ξ . This possibility has not been pursued.

⁸Remember that the grouped-duration baseline hazard is contained in the (standardized) aggregator function. Thus, for given values of \mathbf{x} , changes in z are solely driven by changes in the period-specific constants γ_k .

identical proportionality effects over the whole range of the aggregator function. It can also mimic the proportionality effects of the probit hazard rate over a wide range of z . For low values of z , however, the Pareto model does not perform any better in copying the non-proportionality of the probit model than do the logit or cloglog model. In fact, for very low values of z , the Pareto model exhibits almost perfectly proportional interval hazards, irrespective of the value of ξ .

Figure 1: Proportionate Change in the Hazard Rate for Various Functional Specifications



This suggests that the Pareto hazard model should only be applied if the grouping of the duration times is sufficiently coarse to ensure that the discrete-time exit probabilities are not altogether close to zero. With sufficiently high interval hazards, the proportionality effect implied by the Pareto hazard model depends on the shape parameter ξ , where larger values of ξ are associated with a higher degree of non-proportionality of the interval hazards.

3 Simulation Experiments

This section evaluates the performance of the Pareto hazard model using simulations. The simulations focus mainly on the estimation of the shape parameter, ξ , of the Pareto hazard model. Precise estimation of ξ is crucial for discriminating between submodels such as cloglog and logit. For comparison, models with prespecified values

of ξ (cloglog and logit) are estimated, and the effects of misspecifying the functional form of the hazard are analyzed.

3.1 The Data Generating Process

The performance of the Pareto hazard model is investigated in a variety of settings. These settings vary by (a) the functional specification of the hazard rate; (b) the sample size used in estimation; (c) the censoring schemes applied for generating the binary outcome data; and (d) the explanatory variables affecting the hazard rate.

In setting up the data generating process (DGP), the true hazard rates are calibrated to resemble those typically observed in data on unemployment spells measured in months. For the United States, Sider (1985) estimates aggregate monthly probabilities of remaining unemployed over the period 1969 to 1982. For the first month of unemployment, he reports continuation probabilities ranging from 41 to 59 percent, thereafter increasing to about 90 to 95 percent for month 10 onwards. This implies monthly hazard rates of 0.41 to 0.59 for the first month of unemployment, thereafter declining to about 0.05 to 0.10 by month 10. Using this as a benchmark, the DGPs are calibrated to generate hazards with about half the sample exiting during the first month, and that are decreasing with duration.

For all simulations, the DGP considered employs individual hazard rates of the Pareto form, i.e.

$$h_{ik} = 1 - \left(1 + \frac{\xi}{\exp\{-z_{ik}\}} \right)^{-1/\xi},$$

where ξ takes on the values 1 and 5.⁹ The index is specified as

$$z_{ik} = \mathbf{x}'_{ik}\boldsymbol{\beta} + \gamma f(k),$$

where the two terms represent individual-specific heterogeneity in the hazard rate and duration dependence, respectively. For all simulations, negative duration dependence is assumed, where $f(k) = \ln(k)$, and γ is set equal to -1 . Since variation in the explanatory variables is a key determinant for the precise estimation of ξ , two different specifications with $\text{Var}(\mathbf{x}'_{ik}\boldsymbol{\beta}) = 1$ and $\text{Var}(\mathbf{x}'_{ik}\boldsymbol{\beta}) = 2$ are compared. The respective DGPs employ either time-invariant covariates, time-varying covariates, or both. Finally, two different censoring schemes are used for generating the binary outcome data. In Scheme I (low censoring) all ongoing spells are censored after 12 months. In addition to this type I censoring, Scheme II (high censoring) also involves

⁹The specification with $\xi = 5$ was chosen to analyze situations where transition probabilities are given by a heavily skewed distribution. The specification with $\xi = 1$ was chosen since it constitutes the important special case of a logistic hazard rate. Results of the second special case of the Pareto hazard rate ($\xi = 0$) are not presented here, since they do not provide additional insights. Simulation results for this specification are available from the author upon request.

random censoring, thereby increasing the total fraction of censored spells. For each of these DGPs, 100 random samples of size 500, 1 000, and 2 000 are generated and analyzed using different estimation models.

3.2 Monte Carlo Exercises and Estimation Models

The Monte Carlo simulations are organized in three main sets. In the first set, Monte Carlo exercise A, focus is placed on the estimation of ξ under various circumstances. In particular, the data are generated using a sequential logit model (i.e. a sequential Pareto model with $\xi = 1$) including a single time-invariant covariate. The individual heterogeneity in the model is specified as

$$\mathbf{x}'_{ik}\boldsymbol{\beta} = x_i\beta,$$

where β is fixed at 1, and $x_i \sim IIDN(0, \sigma_x)$. Two different cases with $\sigma_x^2 = 1$ and $\sigma_x^2 = 2$ are considered. Moreover, two different censoring schemes (Scheme I and Scheme II described above) are applied. This produces four different DGPs for simulation exercise A. For each of these DGPs, a logit and a Pareto hazard model are estimated. Thus, the performance of the Pareto model can be evaluated in comparison to the “true” hazard model. To evaluate the effects of a flexible baseline specification on the estimation of β and ξ , all models are estimated using both a nonparametric dummy baseline and a logarithmic baseline hazard.

The second set of simulations, Monte Carlo exercise B, is concerned with the consequences of misspecifying the functional form of the hazard rate for the estimated effects of covariates on exit probabilities. In simulation exercise B, the data are generated using a sequential logit model with individual heterogeneity specified as

$$\mathbf{x}'_{ik}\boldsymbol{\beta} = x_{1,ik}\beta_1 + x_{2,ik}\beta_2,$$

where β_1 and β_2 are fixed at 1. The variables $x_{1,ik}$ and $x_{2,ik}$ are generated as independent random draws from a normal distribution with zero mean and variance 0.25 and a demeaned Gamma distribution with variance 0.75, respectively. Three different simulations for the explanatory variables are considered:

1. Two time-invariant variables that are iid across individuals.
2. Two time-varying variables that are iid across individuals and time.
3. Two independent variables; one constant over time, and one defined as the sum of a time-invariant variable and a time-varying one.

The first simulation represents an extreme case, where all explanatory variables are time-constant and thus correlated across time intervals. The second simulation represents the opposite extreme case, where all explanatory variables are time-varying

and iid across individuals and time. Both scenarios are interesting from a theoretical point of view, but they are not very likely to occur in practice. Thus, the third simulation represents an intermediate case, where the explanatory variables are a mixture of time-invariant and time-varying variables that are correlated across time intervals. For all simulations, logit, cloglog, probit, and Pareto models with nonparametric baseline specifications are estimated, and the results are compared.

The third set of simulations, Monte Carlo exercise C, is organized in the same way as exercise B, except that exit probabilities are not logistic but calculated using the generalized log-Burr distribution with $\xi = 5$. The aim of this exercise is to analyze the consequences of misspecifying the functional form of the hazard rate when the “true” exit probabilities are not given by a symmetric distribution (as in the logit case) but a heavily (right-)skewed distribution.

For all simulation exercises, the computations were performed in STATA. For the Pareto hazard model, the maximum likelihood estimation was carried out using STATA’s method *ml dl*. For maximizing the Pareto likelihood function, a Davidon-Fletcher-Powell algorithm was employed, where the starting values were obtained from a conventional logit specification. All other models were estimated using STATA’s built-in commands.

3.3 Results

This section presents the results of the Monte Carlo exercises A to C. The results of exercise A are reported in Table 4. The depicted results are the averages and standard deviations over 100 replications for the parameters β (the true value of which is 1), γ (with a true value of -1), and ξ (with true value equal to 1). The type of DGP used to generate the data is specified by column. More precisely, four different types of DGPs are considered: sequential logit with $\text{Var}(x_i\beta) = 1$ or $\text{Var}(x_i\beta) = 2$ and with a low or high fraction of censored spells. The low fraction of censored spells involves only type I censoring, leading to a censoring proportion of about 16% if $\text{Var}(x_i\beta) = 1$ and about 20% if $\text{Var}(x_i\beta) = 2$. In addition to the type I censoring, a high fraction of censored spells also involves random censoring. The resulting proportions of censored spells are approximately 24% if $\text{Var}(x_i\beta) = 1$ and about 27% if $\text{Var}(x_i\beta) = 2$. The sample sizes (500, 1 000, or 2 000) and the type of estimation model used are given by row.

The structural parameter β is estimated very accurately for all sample sizes and all estimated models considered. Not surprisingly, increasing the sample size reduces the standard deviation of the estimates. Increasing the proportion of censored spells does not seem to distort the estimates of β in any of the estimated models, but does lead to increased standard deviations. The increased standard deviations are not surprising, since censoring reduces the sample size in terms of person periods.

A lower variance of $x_i\beta$ also increases uncertainty in the estimation of β . Finally, whether the baseline hazard is estimated using the true logarithmic specification or a nonparametric step function does not seem to affect the estimation results.

When fitting the Pareto model to the different generated data sets, the estimation of the shape parameter ξ is of particular interest. In general, ξ is very accurately estimated, but with rather large standard deviations in small samples. Especially with a sample size of only 500 observations and a high fraction of censored spells, the estimate of ξ exhibits a large variance. However, increasing the sample size or the variance of $x_i\beta$ greatly reduces variation in the estimation of ξ . With 2000 observations, the effect of censoring on the standard deviation has almost vanished. For samples of 1000 or more observations, the estimated values of ξ are significantly different from 0 for all DGPs considered. Applying the true logarithmic specification of the baseline hazard instead of period-specific constants slightly reduces uncertainty in the estimation of ξ . This effect, however, is only modest. Finally, when comparing the β -estimates between the estimated logit and Pareto models, standard deviations obtained from the latter ones are approximately twice as large. However, standard deviations for the β - and γ -estimates obtained from the Pareto hazard models have to be interpreted with caution. Since the scale of the structural parameters depends on the shape parameter ξ , uncertainty in the estimation of the shape leads to a higher variation in the absolute value of the structural parameters.

To summarize, the Pareto hazard model performs well under all circumstances considered. In order to be able to effectively discriminate between the logit and the cloglog model, a sample size of 1000 or more observations is desirable. A high variation in the explanatory variables across individuals also enhances a precise estimation of the shape parameter ξ .

The results of Monte Carlo exercises B and C are presented in Tables 5 and 6. The results reported are the averages and standard deviations over 100 replications for the covariate coefficients, β_1 and β_2 (both with a true value of 1), and their ratio β_1/β_2 . The DGPs used to generate the simulated data are specified by column. More precisely, 6 different DGPs are considered: a sequential logit model (Table 5) and a sequential Pareto model with $\xi = 5$ (Table 6), where the covariates included in the models are either time-invariant, time-varying, or a mixture of both. In all cases, the variance of $\mathbf{x}'_{ik}\boldsymbol{\beta}$ is fixed at 1, and the censoring mechanism involves type I censoring only. The sample size and the type of estimation model used are specified by row.

The results in Table 5 indicate that choosing a cloglog specification of the hazard rate when the true one is logit leads to biased estimates of the structural parameters. For all DGPs and sample sizes, the ratio β_1/β_2 is greater than 1. While increasing the sample size reduces the standard deviation, the distortion in the estimated parameter ratio is persistent. The distortion is largest in the extreme case of two time-invariant

covariates.¹⁰ The Pareto model performs as well as the true logit specification in estimating the parameter ratio. This holds both for the average estimated ratios and the respective standard deviations. In terms of estimating the parameter ratio, the probit model also performs quite well with only a modest underestimation of the true value in most cases. However, as illustrated in Figure 4, the probit model exhibits a much higher degree of non-proportionality and generally larger predicted effects of the covariates on exit probabilities. Figure 4 shows the relative change in the hazard, associated with a change in the explanatory variable x_1 , at the period-specific lower, middle, and upper quartiles of covariates and for all durations. The depicted values are averages over 100 replications of simulation exercise B with time-invariant covariates and a sample size of 1 000. As expected from the theoretical discussion in Section 2.3, the predicted degree of non-proportionality (i.e. the slope of the depicted curves) is largest for the probit model and smallest for the cloglog model, and the logit model constitutes the intermediate case. Moreover, the predicted change in the hazard differs substantially across specifications, especially if hazard rates are low, i.e. at longer durations and at the lower quartile of covariates. The predicted change in the hazard is largest for the probit specification and smallest for the cloglog specification, with the logit model again being the intermediate case. These results are in line with the empirical findings of Narendranathan and Stewart (1993). The relative change in the hazard predicted by the Pareto hazard model is practically indistinguishable from the relative change in the hazard predicted by the true logit specification. With respect to the predicted hazard rates, the four specifications do not differ substantially. To economize on space, the respective figures are not reported in this paper but are available from the author upon request.

Monte Carlo exercise C differs from exercise B only in so far as the true exit probabilities are given by a highly right-skewed log-Burr distribution (with $\xi = 5$) instead of the symmetric logistic distribution. The results of exercise C are reported in Table 6. As compared to exercise B, the cloglog model results in a noticeably larger bias in the estimated ratio β_1/β_2 . Again, increasing the sample size does not reduce the bias. The logit and probit specifications exhibit considerably less bias in the estimated parameter ratio than the cloglog model, with the probit model being closest to the true value for all DGPs and sample sizes. As in exercise B, the Pareto model performs well in estimating the parameter ratio. A striking feature of the Pareto estimates are the large standard deviations of the parameter estimates obtained from the 500 observations sample. This is due to the fact that, as ξ gets

¹⁰In a recent simulation study, Nicoletti and Rondinelli (2009) find that misspecifying the functional form of the hazard rate causes only a proportional rescaling of the covariate coefficients. However, Nicoletti and Rondinelli (2009) consider only the special case of normally distributed covariates, and their findings do not seem to extend to different setups.

larger, the likelihood function becomes increasingly flat with respect to ξ . This leads to an increased uncertainty in the estimation of ξ , especially if sample sizes are small. Since the scale of the parameter estimates obtained from the Pareto model depends on the value of ξ , an imprecise estimation of the shape manifests itself in large standard deviations of the parameter estimates. For obvious reasons, this scale effect does not affect the standard deviation of the estimated ratio β_1/β_2 . For the various estimation models, Figure 5 shows the estimated relative change in the hazard associated with a change in the variable x_1 . As expected from the theoretical discussion in Section 2.3, the Pareto hazard model exhibits a large degree of non-proportionality if the shape parameter ξ is large. For the data at hand, the Pareto model exhibits a considerably higher degree of non-proportionality than the cloglog and logit specifications. With respect to the slope of the depicted curves, the Pareto model is similar to the decidedly non-proportional probit specification. However, for the data considered, even the probit model underestimates the change in the covariate effects over time.

To sum up the results of Monte Carlo exercises B and C, misspecifying the functional form of the hazard rate is found to have two important negative consequences. First, estimated relative effects of covariates on exit probabilities are biased. This bias is more pronounced for time-invariant explanatory variables than for time varying ones which are independent both across individuals and time. Second, the estimated effect of a single explanatory variable on the relative change in the hazard is sensitive to the functional specification of the hazard rate. Due to its flexibility, the Pareto hazard model can incorporate a wide range of proportionality effects.

To summarize, the results from Monte Carlo exercises A-C indicate that the Pareto hazard model performs well even in small samples. It is shown that, for data typically encountered in unemployment studies, the Pareto hazard model can be applied to effectively discriminate between the commonly applied cloglog and logit specifications of the hazard rate. Moreover, the Pareto model is sufficiently flexible to allow for a wide range of proportionality effects other than those implied by the conventional cloglog and logit specifications.

4 An Empirical Example

To examine the performance of the Pareto hazard model in practice, a sample of unemployment duration data taken from the Panel Study of Income Dynamics (PSID) is analyzed in the following.¹¹ The focus of the analysis is placed on the effectiveness of the Pareto model in discriminating between submodels such as the cloglog and

¹¹More extensive analyses of this sample can be found in papers by Katz (1986), Han and Hausman (1990), and Sueyoshi (1995).

the logit. In addition, models with different (fixed) values of ξ are estimated on the same data to investigate the implications of the functional form assumptions for the effects of covariates on the hazard rate.

Table 1: Variable definition and descriptive statistics of the PSID unemployment spell sample ($n = 1\,055$)

Variable	Description	Mean	S.D.
<i>Demographic and economic variables</i>			
Age	Age of individual in years	33.154	10.607
Female	Indicator for female	0.167	0.373
Schooling	Years of schooling	11.341	2.170
Non-white	Indicator for non-white	0.506	0.500
Dependents	Number of dependents in household	3.038	1.639
UI receipt	Indicator for receipt of UI during spell	0.636	0.481
Married	Indicator for marital status in previous year	0.632	0.482
Unemployment	Local rate of unemployment	7.701	2.551
Wife employed	Indicator for wife's employment status	0.342	0.475
Homeowner	Indicator for homeownership	0.439	0.496
<i>Industry dummies</i>			
Equipment	Indicator for transportation equipment	0.118	0.322
Metals	Indicator for metals	0.058	0.234
Durables	Indicator for other durable goods manufacturing (excluding metals)	0.123	0.329
Non-durables	Indicator for non-durable goods manufacturing	0.133	0.339
Trade	Indicator for wholesale or retail trade	0.103	0.305
Transportation	Indicator for transportation or public utilities	0.080	0.271
Mining	Indicator for mining or agriculture	0.034	0.182
Service	Indicator for services	0.172	0.377
Construction	Indicator for construction	0.180	0.384
<i>Occupation dummies</i>			
Laborer	Indicator for laborer or operative	0.508	0.500
Craft	Indicator for craftsman	0.228	0.416
Clerical	Indicator for clerical, services, sales	0.186	0.389
Manager	Indicator for manager	0.045	0.206
Professional	Indicator for professional or technical worker	0.039	0.139

Source: Author's calculation from PSID sample and Katz (1986).

The data used in the analysis are taken from the waves 14 and 15 of the PSID (years 1980 and 1981) and consist of observations on 1055 unemployment spells

accounting for a total of approximately 13 300 person-period responses. In addition to the spell durations, the data provide information on a variety of demographic and economic characteristics for each individual and an indicator of whether a spell ended via a new job, recall, or by censoring.¹² A brief description of the explanatory variables used in the analysis and descriptive statistics are given in Table 1.

Table 2: Estimated Effects of Selected Covariates on Job Recall Hazards for Various Specifications of the Discrete Hazard Function

	Pareto hazard ($\xi = 5$)	Logit hazard ($\xi = 1$)	Cloglog hazard ($\xi = 0$)	Pareto hazard ($\xi > 0$)
Age	0.0162 (0.0054)	0.0148 (0.0045)	0.0144 (0.0043)	0.0144 (0.0043)
Female	-0.0040 (0.1814)	-0.0141 (0.1608)	-0.0159 (0.1554)	-0.0159 (0.1554)
Schooling	-0.0362 (0.0278)	-0.0301 (0.0234)	-0.0286 (0.0221)	-0.0286 (0.0221)
Non-white	-0.2925 (0.1170)	-0.2508 (0.1012)	-0.2397 (0.0971)	-0.2397 (0.0971)
Dependents	-0.0110 (0.0366)	-0.0042 (0.0308)	-0.0023 (0.0292)	-0.0023 (0.0292)
UI receipt	-0.2171 (0.1163)	-0.1947 (0.1003)	-0.1882 (0.0961)	-0.1882 (0.0961)
Married	0.0755 (0.1759)	0.0505 (0.1526)	0.0446 (0.1465)	0.0446 (0.1464)
Unemployment	-0.0070 (0.0203)	-0.0067 (0.0173)	-0.0067 (0.0165)	-0.0067 (0.0165)
Wife employed	0.1647 (0.1252)	0.1345 (0.1055)	0.1268 (0.1004)	0.1268 (0.1003)
Homeowner	0.4939 (0.1190)	0.4113 (0.1026)	0.3905 (0.0985)	0.3905 (0.0985)
Industry dummies	yes	yes	yes	yes
Occupation dummies	yes	yes	yes	yes
Shape ξ	5	1	0	0.0000 (0.0071)
Log-likelihood	-2170.14	-2168.37	-2167.89	-2167.89

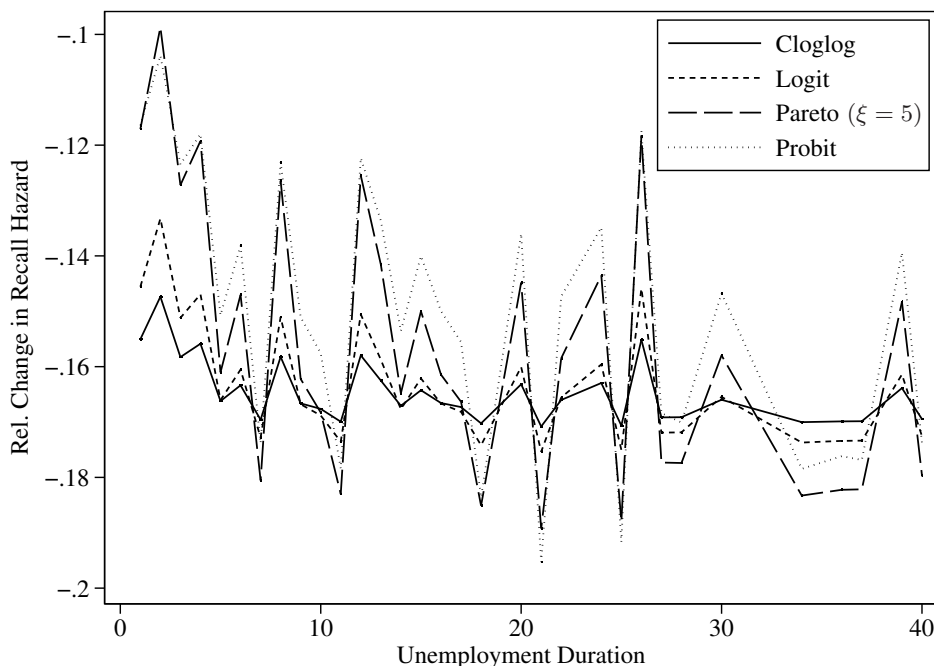
Note: Standard errors in parentheses. For each of the four models, the baseline hazard is modelled nonparametrically using duration dummies. All estimates are derived from 1 055 unemployment spells corresponding to 13 246 person periods.

Table 2 presents parameter estimates from four different specifications of the recall hazard. Following Sueyoshi (1995), exit from unemployment via a new job is consid-

¹²A detailed description of the data set is provided by Katz (1986). The data are available for download at <http://qed.econ.queensu.ca/jae>.

ered an independent competing risk, and the corresponding durations are treated as censored outcomes. All durations exceeding 40 weeks are artificially right-censored. In all model specifications the baseline hazard is estimated nonparametrically using period-specific dummy variables. The first three columns of Table 2 show estimates from the Pareto hazard model with ξ being fixed at different values. In particular, columns two and three correspond to the logit and cloglog models of Table II in Sueyoshi (1995). The last column of Table 2 contains estimates from the Pareto hazard model, where ξ is estimated along with the structural parameters.

Figure 2: Relative Change in the Recall Hazard Associated with a Change in the *UI Receipt* Indicator Variable



Overall, the different model specifications yield quite similar results. Due to differences in scale, models with larger values of ξ yield slightly larger parameter estimates (and standard errors) in absolute values. The differences, however, are not striking, and none of the predicted covariate effects changes sign across specifications. The fit of the models in likelihood terms is very similar across specifications. In particular, the log-likelihood values obtained from the cloglog and logit specifications are almost identical. It would thus be premature to prefer one model over the others on the basis of a simple likelihood criterion. However, despite their similarities at first sight, the model specifications are quite different with respect to the imposed effect of explanatory variables on the hazard rate. This is illustrated in Figure 2.

Figure 2 depicts the relative change in the recall hazard due to a change in the receipt of unemployment insurance for a representative individual (married white male, 30 years old, 12 years of education, home owner, wife employed, 3 dependents, laborer in metals industry, 7% regional unemployment rate). Since the receipt of unemployment insurance is merely indicated by a dummy variable, the proportionate change in the recall hazard cannot be calculated in the usual manner as the derivative of the log hazard rate at various interval durations. Instead, the relative change in the hazard associated with a *discrete* change in the *UI receipt* indicator is computed for various durations and hazard specifications.¹³ Specifically, the effect of receiving unemployment benefits is calculated as the relative change in the hazard when the *UI receipt* indicator is first set to 0 and then to 1, i.e.

$$\frac{h(UI = 1) - h(UI = 0)}{h(UI = 0)}.$$

The differences across model specifications are striking. As expected, the PH assumption imposed by the cloglog model results in covariate effects with low variability over time. The logit model exhibits somewhat greater non-proportionality, whereas the Pareto model with ξ fixed at 5 predicts largely non-proportional effects of the *UI receipt* indicator. In fact, the latter specification is very similar to the decidedly non-proportional probit model.

Table 3: Descriptive Statistics for the Relative Change in the Recall Hazard Associated with the *UI Receipt* Indicator

	Cloglog	Logit	Pareto ($\xi = 5$)	Probit
Mean	-0.165	-0.163	-0.157	-0.151
Standard deviation	0.006	0.011	0.025	0.024
Minimum	-0.171	-0.175	-0.189	-0.195
Maximum	-0.147	-0.133	-0.099	-0.104

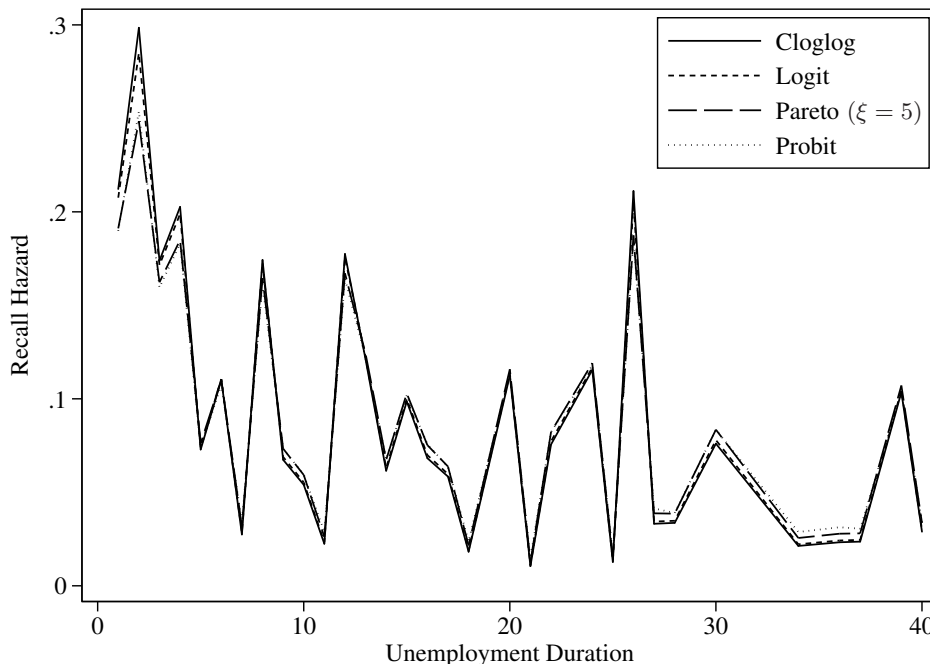
Note: The relative change is computed for a representative individual (married white male, 30 years old, 12 years of education, home owner, wife employed, 3 dependents, laborer in metals industry, 7% regional unemployment rate).

With $\xi = 5$, the variability in the predicted effect of the *UI receipt* indicator – as measured by the standard deviation of the relative change in the recall hazard over

¹³There are two reasons why the variable *UI receipt* was chosen to illustrate the effect of covariates on the hazard rate. First, as opposed to most of the other explanatory variables employed in the regression, the effect of unemployment benefits on the hazard is significant (at least at the 10% level) for all model specifications. Second, the effect of unemployment benefits is of high political relevance and frequently analyzed in both theoretical work (e.g. van den Berg, 1990a) and empirical studies (e.g. Narendranathan and Stewart, 1993).

time – is about four times larger than under the PH assumption (0.025 vs. 0.006). Looking at the lowest and highest benefit effects, the Pareto specification with $\xi = 5$ predicts that the receipt of unemployment benefits will decrease the recall hazard by about 10 percent in the second week and by roughly 19 percent in week 21. This can be compared to the corresponding effects predicted by the grouped-duration PH model, which are approximately 15 and 17 percent, respectively. Further summary statistics are presented in Table 3.

Figure 3: Predicted Recall Hazard for a Representative Individual



Finally, Figure 3 shows the predicted recall hazard function for a representative individual (married white male, 30 years old, 12 years of education, home owner, wife employed, 3 dependents, laborer in metals industry, 7% regional unemployment rate, no unemployment benefits) obtained from the four different specifications of the discrete hazard rate. Overall, the absolute differences in the predicted hazard across model specifications are not striking. This is, however, not surprising, since the weekly hazard rates are generally rather low. For short durations with relatively high hazard rates, some noticeable differences between model specifications exist. In particular, the predicted hazard rates for the second week of unemployment range from approximately 0.25 for the Pareto model with $\xi = 5$ to about 0.30 for the cloglog specification. Surprisingly, even with regard to the predicted hazard rates, the probit model yields similar results as the highly skewed Pareto specification, whereas the

logit model is close to the grouped-duration PH specification.

To summarize, the findings of this empirical example indicate that the Pareto hazard model is sufficiently flexible to allow for a wide range of proportionality patterns for the effects of explanatory variables on exit probabilities. It incorporates the cloglog and logit models as special cases, and can produce results that are similar to the probit specification both in terms of predicted hazard rates and the imposed effects of covariates on the hazard rate. The choice between different model specifications can thus be reduced to the estimation of a single additional parameter. For the data considered in this empirical example, estimation results from the Pareto model lend support to the PH assumption. As presented in the last column of Table 2, the estimate for ξ is virtually zero, with a very low standard error of 0.0071. Formally, the null hypothesis $\xi = 0$ can be tested against the one-sided alternative $\xi > 0$ on the basis of the asymptotic normality of $\hat{\xi}$, where the probability that $\hat{\xi} \leq 0$ is amassed at 0 (see Prentice, 1975). For the data at hand, the asymptotic results imply rejection of the null hypothesis at the 5% significance level if $\hat{\xi} > 1.64 \times 0.0071$. Clearly, this is not the case, and the PH assumption cannot be rejected for the data sample under consideration.

5 Concluding Remarks

This paper introduces a new hazard rate model for grouped duration data, which includes the widely applied cloglog and logit hazard models as special cases. The paper provides new insights into the interrelation between continuous-time duration processes and grouped-duration hazard specifications. In particular, it is shown that the class of discrete-time hazard models considered can be linked to the asymptotic distribution of threshold excesses of an underlying continuous duration variable.

The performance of the proposed Pareto hazard model is investigated using simulated and empirical data. It is shown that the model works well under a variety of circumstances covering variations in sample size, the proportion of censored spells, and the types of explanatory variables included. It is also shown that the estimated effects of covariates on exit probabilities are very sensitive to the functional specification of the hazard rate. Thus, the choice between cloglog, logit, and probit hazard models is not innocuous in a duration context. Since the Pareto hazard model contains an additional shape parameter, it is very flexible with respect to the imposed effects of explanatory variables on the hazard. The Pareto hazard model nests both the cloglog and the logit model, which reduces the choice between these models to the estimation of a single additional parameter. Moreover, the Pareto model is also capable of exhibiting covariate effects that are very similar to the decidedly non-proportional probit hazard model.

Throughout the paper it is implicitly assumed that all individual variation in the hazard rate can be observed. This assumption, however, is unlikely to hold in practice, and accounting for unobserved heterogeneity may well be important. In particular, it is well-known that the neglect of unobserved heterogeneity may cause spurious negative duration dependence of the estimated baseline hazard (see e.g. Lancaster, 1979, Nickell, 1979, or Vaupel and Yashin, 1985). The negative consequences of neglecting latent heterogeneity may then be more severe than effects of misspecifying the functional form of the discrete-time hazard rate. It is thus advisable to test for the presence of unobserved factors before applying the Pareto hazard model. Evaluating the effect of unobserved heterogeneity on the estimation of ξ and establishing a way to account for latent factors in the Pareto hazard model should therefore be seen as a fruitful area for future research.

Appendix: Simulation Results

Table 4: Parameter Estimates from Monte Carlo Exercise A. DGP: Logit with Negative Duration Dependence

		Var($x_i\beta$) = 1			Var($x_i\beta$) = 2	
Estimated model	n	Fraction of censored spells				
		low		high		
Logit with log. baseline	500	β	0.994 (0.076)	1.020 (0.096)	0.999 (0.066)	1.024 (0.082)
		γ	-0.995 (0.088)	-0.991 (0.111)	-0.997 (0.097)	-0.987 (0.121)
	1 000	β	1.006 (0.058)	1.005 (0.070)	1.009 (0.049)	1.010 (0.056)
		γ	-0.991 (0.068)	-1.002 (0.077)	-0.989 (0.073)	-0.996 (0.083)
	2 000	β	1.006 (0.039)	1.001 (0.044)	1.001 (0.037)	1.002 (0.035)
		γ	-0.999 (0.040)	-0.997 (0.057)	-0.997 (0.041)	-0.998 (0.058)
Logit with dummy baseline	500	β	0.997 (0.075)	1.022 (0.096)	1.001 (0.066)	1.026 (0.082)
	1 000	β	1.007 (0.058)	1.006 (0.070)	1.010 (0.049)	1.011 (0.056)
	2 000	β	1.007 (0.039)	1.002 (0.043)	1.002 (0.037)	1.003 (0.035)
Pareto with log. baseline	500	β	0.996 (0.148)	1.051 (0.207)	1.004 (0.128)	1.027 (0.170)
		γ	-0.995 (0.125)	-1.018 (0.194)	-0.998 (0.115)	-0.988 (0.161)
		ξ	0.996 (0.533)	1.126 (0.839)	1.008 (0.382)	0.999 (0.569)
	1 000	β	1.015 (0.105)	1.023 (0.133)	1.013 (0.082)	1.019 (0.116)
		γ	-0.997 (0.088)	-1.015 (0.117)	-0.992 (0.087)	-1.001 (0.107)
		ξ	1.039 (0.378)	1.066 (0.450)	1.016 (0.286)	1.027 (0.367)
	2 000	β	1.013 (0.082)	1.005 (0.094)	1.003 (0.058)	1.001 (0.066)
		γ	-1.004 (0.068)	-0.999 (0.079)	-0.999 (0.058)	-0.997 (0.070)
		ξ	1.029 (0.319)	1.012 (0.324)	1.010 (0.219)	0.994 (0.228)
Pareto with dummy baseline	500	β	1.002 (0.151)	1.072 (0.236)	1.006 (0.136)	1.035 (0.186)
		ξ	1.018 (0.578)	1.186 (0.937)	1.007 (0.420)	1.017 (0.626)
	1 000	β	1.029 (0.116)	1.026 (0.142)	1.022 (0.090)	1.026 (0.124)
		ξ	1.091 (0.413)	1.073 (0.494)	1.047 (0.324)	1.047 (0.416)
	2 000	β	1.014 (0.087)	1.007 (0.097)	1.003 (0.060)	1.003 (0.071)
		ξ	1.028 (0.334)	1.016 (0.352)	1.005 (0.228)	0.997 (0.251)

Note: The displayed parameter estimates are average values based on 100 replications of the simulation experiment (standard deviations in parentheses). The true parameter values are 1 for β and ξ , and -1 for γ .

Table 5: Parameter Estimates from Monte Carlo Exercise B. DGP: Logit with Negative Duration Dependence

Estimated model	n		Type of covariates		
			time-invariant	time-varying	mixture
Logit with dummy baseline	500	β_1	1.010 (0.146)	1.034 (0.137)	1.016 (0.145)
		β_2	1.009 (0.119)	1.020 (0.082)	1.018 (0.099)
		β_1/β_2	1.011 (0.166)	1.019 (0.151)	1.003 (0.145)
	1 000	β_1	0.996 (0.108)	1.026 (0.090)	1.004 (0.096)
		β_2	1.002 (0.084)	1.005 (0.058)	1.016 (0.062)
		β_1/β_2	0.999 (0.123)	1.022 (0.099)	0.990 (0.099)
	2 000	β_1	1.004 (0.072)	1.007 (0.066)	1.006 (0.068)
		β_2	1.004 (0.057)	1.012 (0.036)	1.008 (0.051)
		β_1/β_2	1.003 (0.087)	0.996 (0.067)	0.999 (0.070)
Cloglog with dummy baseline	500	β_1	0.775 (0.109)	0.770 (0.103)	0.772 (0.114)
		β_2	0.692 (0.093)	0.707 (0.064)	0.715 (0.070)
		β_1/β_2	1.133 (0.188)	1.097 (0.168)	1.084 (0.159)
	1 000	β_1	0.765 (0.085)	0.768 (0.069)	0.761 (0.078)
		β_2	0.683 (0.065)	0.701 (0.043)	0.717 (0.052)
		β_1/β_2	1.128 (0.143)	1.099 (0.117)	1.065 (0.115)
	2 000	β_1	0.766 (0.059)	0.750 (0.051)	0.765 (0.052)
		β_2	0.678 (0.043)	0.695 (0.035)	0.711 (0.039)
		β_1/β_2	1.133 (0.103)	1.082 (0.080)	1.077 (0.079)
Probit with dummy baseline	500	β_1	0.577 (0.083)	0.604 (0.078)	0.583 (0.082)
		β_2	0.589 (0.069)	0.600 (0.048)	0.591 (0.056)
		β_1/β_2	0.988 (0.163)	1.012 (0.148)	0.992 (0.141)
	1 000	β_1	0.567 (0.061)	0.599 (0.053)	0.577 (0.055)
		β_2	0.585 (0.048)	0.592 (0.033)	0.590 (0.035)
		β_1/β_2	0.974 (0.121)	1.015 (0.098)	0.980 (0.099)
	2 000	β_1	0.572 (0.041)	0.589 (0.039)	0.577 (0.039)
		β_2	0.586 (0.033)	0.595 (0.021)	0.585 (0.029)
		β_1/β_2	0.978 (0.086)	0.990 (0.067)	0.988 (0.071)
Pareto with dummy baseline	500	β_1	1.028 (0.246)	1.060 (0.234)	1.054 (0.191)
		β_2	1.031 (0.269)	1.056 (0.242)	1.071 (0.199)
		β_1/β_2	1.014 (0.166)	1.018 (0.160)	0.998 (0.149)
	1 000	β_1	1.003 (0.167)	1.026 (0.132)	1.015 (0.128)
		β_2	1.015 (0.186)	1.009 (0.129)	1.029 (0.116)
		β_1/β_2	1.000 (0.131)	1.022 (0.102)	0.990 (0.100)
	2 000	β_1	1.018 (0.091)	1.013 (0.100)	1.008 (0.098)
		β_2	1.023 (0.099)	1.019 (0.085)	1.011 (0.091)
		β_1/β_2	0.999 (0.089)	0.996 (0.067)	0.999 (0.071)

Note: The displayed parameter estimates are average values based on 100 replications of the simulation experiment (standard deviations in parentheses). The true parameter values are 1 for β_1 and β_2 . While the estimated parameters are identified only up to scale, the estimated ratio β_1/β_2 needs to equal 1 for all estimated models.

Table 6: Parameter Estimates from Monte Carlo Exercise C. DGP: Pareto ($\xi = 5$)
with Negative Duration Dependence

Estimated model	n		Type of covariates		
			time-invariant	time-varying	mixture
Logit with dummy baseline	500	β_1	0.671 (0.120)	0.659 (0.103)	0.686 (0.116)
		β_2	0.581 (0.079)	0.560 (0.060)	0.611 (0.072)
		β_1/β_2	1.174 (0.260)	1.131 (0.225)	1.134 (0.210)
	1 000	β_1	0.656 (0.096)	0.656 (0.075)	0.669 (0.086)
		β_2	0.593 (0.061)	0.587 (0.050)	0.595 (0.046)
		β_1/β_2	1.113 (0.173)	1.125 (0.165)	1.129 (0.156)
	2 000	β_1	0.681 (0.064)	0.645 (0.052)	0.661 (0.059)
		β_2	0.585 (0.034)	0.586 (0.033)	0.599 (0.036)
		β_1/β_2	1.166 (0.113)	1.104 (0.101)	1.107 (0.108)
Cloglog with dummy baseline	500	β_1	0.573 (0.104)	0.554 (0.091)	0.583 (0.100)
		β_2	0.451 (0.065)	0.460 (0.050)	0.478 (0.060)
		β_1/β_2	1.294 (0.295)	1.220 (0.246)	1.236 (0.240)
	1 000	β_1	0.559 (0.084)	0.551 (0.065)	0.567 (0.075)
		β_2	0.455 (0.051)	0.456 (0.039)	0.459 (0.039)
		β_1/β_2	1.237 (0.202)	1.217 (0.183)	1.241 (0.176)
	2 000	β_1	0.579 (0.057)	0.541 (0.045)	0.559 (0.049)
		β_2	0.450 (0.029)	0.451 (0.026)	0.461 (0.031)
		β_1/β_2	1.291 (0.126)	1.203 (0.112)	1.217 (0.129)
Probit with dummy baseline	500	β_1	0.375 (0.067)	0.374 (0.057)	0.385 (0.064)
		β_2	0.338 (0.045)	0.343 (0.034)	0.354 (0.041)
		β_1/β_2	1.129 (0.251)	1.103 (0.216)	1.099 (0.203)
	1 000	β_1	0.366 (0.052)	0.373 (0.043)	0.376 (0.047)
		β_2	0.345 (0.035)	0.341 (0.029)	0.344 (0.027)
		β_1/β_2	1.068 (0.164)	1.099 (0.160)	1.097 (0.152)
	2 000	β_1	0.381 (0.035)	0.367 (0.030)	0.371 (0.033)
		β_2	0.341 (0.020)	0.341 (0.019)	0.346 (0.021)
		β_1/β_2	1.120 (0.110)	1.079 (0.098)	1.076 (0.107)
Pareto with dummy baseline	500	β_1	1.276 (2.208)	1.330 (1.976)	1.248 (1.419)
		β_2	1.406 (3.551)	1.350 (2.008)	1.380 (2.492)
		β_1/β_2	1.035 (0.243)	1.031 (0.226)	1.008 (0.204)
	1 000	β_1	1.004 (0.202)	1.074 (0.245)	1.089 (0.236)
		β_2	1.054 (0.245)	1.070 (0.257)	1.111 (0.293)
		β_1/β_2	0.972 (0.161)	1.017 (0.151)	1.000 (0.154)
	2 000	β_1	1.047 (0.156)	1.027 (0.144)	1.024 (0.145)
		β_2	1.038 (0.177)	1.035 (0.144)	1.040 (0.160)
		β_1/β_2	1.019 (0.114)	0.995 (0.091)	0.992 (0.106)

Note: The displayed parameter estimates are average values based on 100 replications of the simulation experiment (standard deviations in parentheses). The true parameter values are 1 for β_1 and β_2 . While the estimated parameters are identified only up to scale, the estimated ratio β_1/β_2 needs to equal 1 for all estimated models.

Figure 4: Proportionate Change in the Hazard Rate Associated with a Change in x_1 .
Monte Carlo Exercise B. DGP: Logit with Negative Duration Dependence.

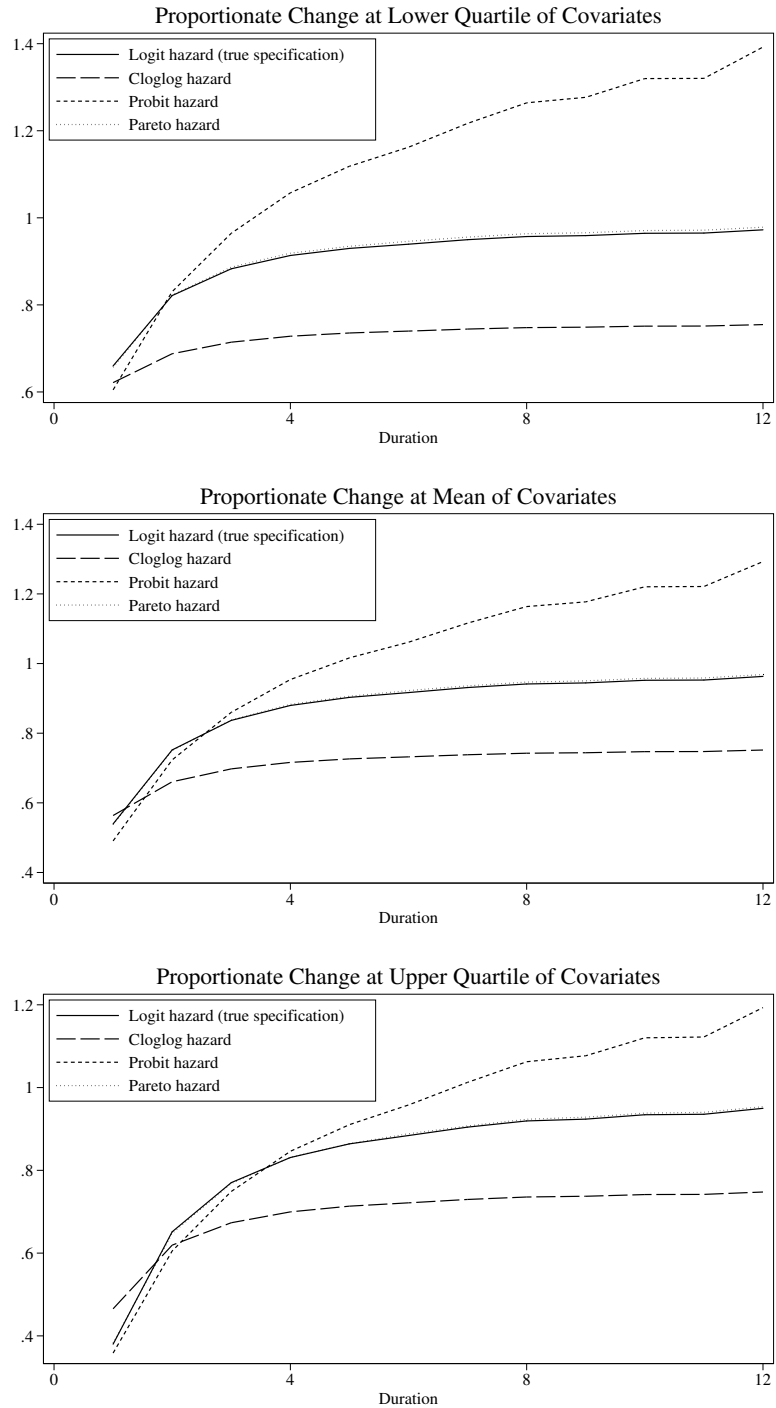
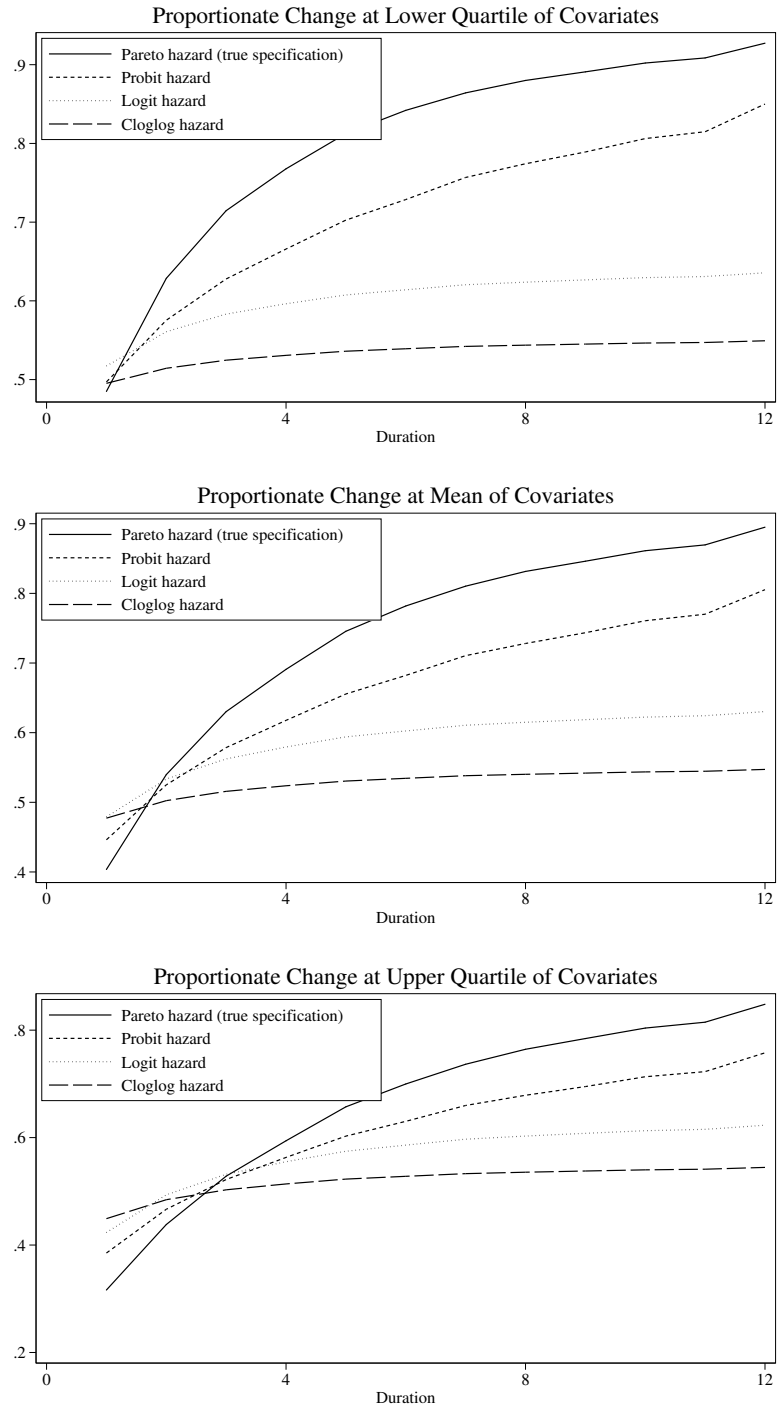


Figure 5: Proportionate Change in the Hazard Rate Associated with a Change in x_1 .
Monte Carlo Exercise C. DGP: Pareto ($\xi = 5$) with Negative Duration Dependence.



References

- ALLISON, P. D. (1982), “Discrete-time methods for the analysis of event histories”, *Sociological Methodology*, vol. 13, pp. 61–98.
- ARANDA-ORDAZ, F. J. (1981), “On two families of transformations to additivity for binary response data”, *Biometrika*, vol. 68(2), pp. 357–363.
- BLANCHARD, O. J. and DIAMOND, P. (1994), “Ranking, unemployment duration, and wages”, *Review of Economic Studies*, vol. 61(3), pp. 417–434.
- BURR, I. W. (1942), “Cumulative frequency functions”, *Annals of Mathematical Statistics*, vol. 13(2), pp. 215–232.
- COX, D. R. (1972), “Regression models and life-tables (with discussion)”, *Journal of the Royal Statistical Society, Series B*, vol. 34(2), pp. 187–220.
- DAVISON, A. C. and SMITH, R. L. (1990), “Models for exceedances over high thresholds”, *Journal of the Royal Statistical Society, Series B*, vol. 52(3), pp. 393–442.
- DIAMOND, P. A. and HAUSMAN, J. A. (1984), “The retirement and unemployment behavior of older men”, in H. J. Aaron and G. Burtless (eds.), *Retirement and Economic Behavior*, pp. 97–134, Washington, D.C.: The Brookings Institution.
- EMBRECHTS, P., KLÜPPELBERG, C. and MIKOSCH, T. (1997), *Modelling Extremal Events for Insurance and Finance*, Berlin: Springer.
- FISHER, R. A. and TIPPETT, L. H. C. (1928), “On the estimation of the frequency distributions of the largest or smallest member of a sample”, *Proceedings of the Cambridge Philosophical Society*, vol. 24(2), pp. 180–190.
- HAN, A. and HAUSMAN, J. A. (1990), “Flexible parametric estimation of duration and competing risk models”, *Journal of Applied Econometrics*, vol. 5(1), pp. 1–28.
- HECKMAN, J. J. and SINGER, B. (1984), “Econometric duration analysis”, *Journal of Econometrics*, vol. 24(1-2), pp. 63–132.
- JENKINS, S. P. (1995), “Easy estimation methods for discrete-time duration models”, *Oxford Bulletin of Economics and Statistics*, vol. 57(1), pp. 129–137.
- JOHNSON, N. L., KOTZ, S. and BALAKRISHNAN, N. (1995), *Continuous Univariate Distributions*, vol. 2, 2nd ed., New York: John Wiley & Sons.
- KALBFLEISCH, J. D. and PRENTICE, R. L. (1973), “Marginal likelihoods based on Cox’s regression and life model”, *Biometrika*, vol. 60(2), pp. 267–278.

- KALBFLEISCH, J. D. and PRENTICE, R. L. (1980), *The Statistical Analysis of Failure Time Data*, New York: John Wiley & Sons.
- KATZ, L. (1986), “Layoffs, recall and the duration of unemployment”, NBER working paper No. 1825.
- LANCASTER, T. (1979), “Econometric methods for the duration of unemployment”, *Econometrica*, vol. 47(4), pp. 939–956.
- LANCASTER, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge: Cambridge University Press.
- LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983), *Extremes and Related Properties of Random Sequences and Series*, New York: Springer.
- MCCALL, B. P. (1994), “Testing the proportional hazards assumption in the presence of unmeasured heterogeneity”, *Journal of Applied Econometrics*, vol. 9(3), pp. 321–334.
- NAGLER, J. (1994), “Scobit: an alternative estimator to logit and probit”, *American Journal of Political Science*, vol. 38(1), pp. 230–255.
- NARENDRANATHAN, W. and STEWART, M. B. (1993), “How does the benefit effect vary as unemployment spells lengthen?”, *Journal of Applied Econometrics*, vol. 8(4), pp. 361–381.
- NICKELL, S. (1979), “Estimating the probability of leaving unemployment”, *Econometrica*, vol. 47(5), pp. 1249–1266.
- NICOLETTI, C. and RONDINELLI, C. (2009), “The (mis)specification of discrete duration models with unobserved heterogeneity: a Monte Carlo study”, *Bank of Italy, Temi di discussione (Working papers) No. 705*.
- PICKANDS, J. (1975), “Statistical inference using extreme order statistics”, *The Annals of Statistics*, vol. 3(1), pp. 119–131.
- PRENTICE, R. L. (1975), “Discrimination among some parametric models”, *Biometrika*, vol. 62(3), pp. 607–614.
- PRENTICE, R. L. (1976), “A generalization of the probit and logit methods for dose response curves”, *Biometrics*, vol. 32(4), pp. 761–768.
- PRENTICE, R. L. and GLOECKLER, L. A. (1978), “Regression analysis of grouped survival data with application to breast cancer data”, *Biometrics*, vol. 34(1), pp. 57–67.

- SIDER, H. (1985), “Unemployment duration and incidence: 1968-82”, *American Economic Review*, vol. 75(3), pp. 461–472.
- SINGER, J. D. and WILLETT, J. B. (1993), “It’s about time: using discrete-time survival analysis to study duration and the timing of events”, *Journal of Educational Statistics*, vol. 18(2), pp. 155–195.
- STEINER, V. (2001), “Unemployment persistence in the West German labour market: negative duration dependence or sorting?”, *Oxford Bulletin of Economics and Statistics*, vol. 63(1), pp. 91–113.
- SUEYOSHI, G. T. (1995), “A class of binary response models for grouped duration data”, *Journal of Applied Econometrics*, vol. 10(4), pp. 411–431.
- TADIKAMALLA, P. R. (1980), “A look at the Burr and related distributions”, *International Statistical Review*, vol. 48(3), pp. 337–344.
- VAN DEN BERG, G. J. (1990a), “Nonstationarity in job search theory”, *Review of Economic Studies*, vol. 57(2), pp. 255–277.
- VAN DEN BERG, G. J. (1990b), “Search behaviour, transitions to non-participation and the duration of unemployment”, *The Economic Journal*, vol. 100(402), pp. 842–865.
- VAN DEN BERG, G. J. (2001), “Duration models: specification, identification, and multiple durations”, in J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, pp. 3381–3460, Amsterdam: North-Holland.
- VAUPEL, J. W. and YASHIN, A. I. (1985), “Heterogeneity’s ruses: some surprising effects of selection on population dynamics”, *The American Statistician*, vol. 39(3), pp. 176–185.