Wolfgang Hess, Gerhard Tutz, Jan Gertheiss

A Flexible Link Function for Discrete-Time Duration Models

Technical Report Number 155, 2014
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

A Flexible Link Function for Discrete-Time Duration Models*

Wolfgang Hess[†‡] Gerhard Tutz[‡] Jan Gertheiss[§¶]

February 23, 2014

Abstract

This paper proposes a discrete-time hazard regression approach based on the relation between hazard rate models and excess over threshold models, which are frequently encountered in extreme value modelling. The proposed duration model employs a flexible link function and incorporates the grouped-duration analogue of the well-known Cox proportional hazards model and the 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. The simulation results and an empirical analysis of US import durations also show that the choice of link function in discrete hazard models has important implications for the estimation results, and that severe biases in the results can be avoided when using a flexible link function as proposed in this study.

Keywords: Discrete-Time Duration Model, Hazard Rate, Threshold Excess Model, Link Function Estimation, Duration of Trade.

1 Introduction

This study considers the modelling of duration times, that is, the time until the occurrence of some specific event is the variable of primary interest. In theory, time is a continuous

*This paper draws in parts on Working Paper 2009:18, Department of Economics, Lund University. Financial support from the Jan Wallander and Tom Hedelius Foundation under research grant number W2010-0305:1 is gratefully acknowledged.

[†]Corresponding author: Department of Economics and Centre for Economic Demography, Lund University, Sweden. E-mail: Wolfgang.Hess@nek.lu.se.

[‡]Department of Statistics, Ludwig Maximilian University of Munich, Germany.

[§]Department of Animal Sciences, University of Göttingen, Germany.

[¶]Center for Statistics, University of Göttingen, Germany.

variable, and the usual approach for modelling duration times is as follows: Let T_c be a continuous, non-negative random variable denoting the (continuous) duration time. Then, the main tool in modelling duration is the hazard function $\lambda_c(t)$, which at time t is defined by

$$\lambda_c(t) := \lim_{dt \to 0^+} \frac{P(t \leq T_c < t + dt | T_c \geq t)}{dt}, \tag{1}$$

where $P(t \leq T_c < t+dt|T_c \geq t)$ is the probability that duration ends between time points t and $t + dt$, given the duration time is at least t. So the hazard function can be seen as the instantaneous rate of "death" given that the subject or unit of interest "survives" until time t.

In many empirical studies, however, time is observed on a discrete scale, for example, in weeks or months. When analyzing the duration of bilateral trade, for instance, the grouping of duration times is even coarser, and durations are typically measured in years. So let, in general, k intervals be given by $[a_0, a_1), [a_1, a_2), \ldots \ldots, [a_{q-1}, a_q)$, $[a_q, \infty)$, where $q = k - 1$. Discrete time means that $T = t$ is observed if failure occurs within the interval $[a_{t-1}, a_t)$. The corresponding discrete-time hazard rate is then given by

$$\lambda(t) := P(a_{t-1} \leq T_c < a_t | T_c \geq a_{t-1}). \tag{2}$$

Using connections between the hazard function and the so-called survival function $S_c(t) = P(T_c > t)$, it can be shown (see, e.g., Lawless, 1982) that

$$\lambda(t) = 1 - \exp\left\{ - \int_{a_{t-1}}^{a_t} \lambda_c(s)\, ds \right\}. \tag{3}$$

This is the conditional probability that duration ends in the t^{th} interval given the t^{th} interval is reached. For simplicity, let T be the random variable 'discrete time' with possible values $T \in \{1, \ldots, k\}$. That means $T = t$ is observed if failure occurs within the interval $[a_{t-1}, a_t)$, and the discrete hazard function is given by

$$\lambda(t) = P(T = t | T \geq t), \quad t = 1, \ldots, q.$$

With this specification, the model can also be applied to duration data that are intrinsically discrete, i.e., discrete duration data that are not a grouped version of continuous duration times.

Most applications are targeted at modelling and investigating the influence of some covariates on duration times. For doing so, discrete duration models including covariates are typically parameterized as

$$\lambda(t | \boldsymbol{x}_{it}) = F(\gamma_{0t} + \boldsymbol{x}_{it}^T \boldsymbol{\gamma}), \tag{4}$$

where $F(\cdot)$ is a fixed response function, which is assumed to be strictly monotonically increasing. The parameters γ_{0t} represent the baseline hazard, which allows the hazard rate to vary across periods. The contribution of the predictors is captured by the term $\boldsymbol{x}_{it}^T\boldsymbol{\gamma}$, where $\boldsymbol{x}_{it} = (x_{1,it}, \ldots, x_{p,it})^T$ is a p-dimensional column vector of (possibly time-varying) predictors for observation $i \in \{1, \ldots, N\}$, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^T$ are the corresponding regression coefficients.

Since usually there is little *a priori* knowledge about the duration dependence of the hazard rate, it is common to model γ_{0t} in the most flexible way possible by means of period-specific dummy variables. However, also a (flexible) functional specification for γ_{0t} may be chosen to reduce the number of parameters in the model. Since the discrete-time hazard is a conditional probability, the response function $F(\cdot)$ needs to be chosen such that $0 \leq \lambda(t|\boldsymbol{x}_{it}) \leq 1$ for all t. A popular choice is the complementary log-log (cloglog) link function, as, when including period-specific intercepts γ_{0t}, this specification represents the exact grouped-duration analogue of the well-known Cox (1972) proportional hazards (PH) model (see, e.g., Kalbfleisch and Prentice, 1973, or Prentice and Gloeckler, 1978). However, the PH assumption implied by the cloglog model is sometimes not supported by economic theory and empirically questionable, for instance, in labor economics when individual unemployment spells are analyzed (see, e.g., van den Berg, 1990a,b, Blanchard and Diamond, 1994, and McCall, 1994).

Obvious, and also quite popular, alternatives to the cloglog link 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 in this study, the choice of link function is not innocuous in a duration context, as it affects both the estimated covariate effects and the predicted hazards. Therefore, a flexible specification of the link function is proposed here, which can be motivated by the asymptotic distribution of threshold excesses of the underlying continuous duration variable T_c. The hazard model proposed incorporates the well-known cloglog and logit models as special cases, which reduces the choice between these two models to the estimation of a single additional parameter. Besides nesting the two most commonly applied discrete-time hazard models, the model can also produce estimation results that are entirely different from those obtained from the cloglog and logit specifications.

The remainder of the paper is organized as follows. Section 2 introduces the modelling approach, in particular the new type of flexible link function. Section 3 discusses extensions to frailty models and estimation of model parameters. Section 4 evaluates the performance of the proposed model using simulations, Section 5 provides a real data application, and Section 6 concludes.

2 The Duration Model Proposed

An interesting family of response functions, which is a special case of the type IV generalized logistic distribution (see, e.g., Johnson et al., 1995), uses the specification

$$\lambda(t|\boldsymbol{x}_{it}) = F_\xi(\gamma_{0t} + \boldsymbol{x}_{it}^T\boldsymbol{\gamma}), \tag{5}$$

where F_ξ is the distribution function

$$F_\xi(u) = 1 - (1 + \xi \exp\{u\})^{-1/\xi} \tag{6}$$

with shape parameter ξ. For $\xi = 1$ one obtains the logistic distribution function, for the limit $\xi \to 0$ one obtains the cloglog model with $F_0(u) = 1 - \exp\{-\exp\{u\}\}$. Thus, this family comprises the two models that are most widely used in discrete survival modelling, the logistic and the grouped proportional hazards model. The function $F_\xi(u) = 1 - (1 + \xi \exp\{u\})^{-1/\xi}$ is also known as the distribution function of the log-Burr distribution (see Burr, 1942, or Tadikamalla, 1980). The corresponding density is left-skewed for $\xi < 1$ and right-skewed for $\xi > 1$. If $\xi = 1$ it is symmetric, which is well known for the logistic distribution. The family has been considered by Prentice (1975, 1976) in the modelling of binary data and by Hess (2009) in discrete survival modelling. Prentice has shown that ξ can be consistently estimated along with the other parameters by maximum likelihood. A Wald test based on the estimate of ξ can be used to test the parameter within the family of distributions. If the logistic model holds ($\xi = 1$), the asymptotic distribution of $\hat{\xi}$ is normal, more concisely, $N(1, 4(\pi^2 + 3)/(n(\pi^2 - 6)))$, where n denotes the total number of binary observations. In the limiting case, $\xi \to 0$, 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 $\xi^* \sim N\left(0; \pi^2/(n(\pi^2 - 6))\right)$.

2.1 An Underlying Continuous-Time Process

The choice of the log-Burr distribution as the response function can be motivated by the asymptotic distribution of threshold excesses of the continuous duration variable T_c. The derivation of the hazard specification requires two assumptions about the cumulative distribution function of T_c, $G(t)$, which is directly linked to the grouped hazard through the relation

$$\lambda(t) = \frac{G(a_t) - G(a_{t-1})}{1 - G(a_{t-1})}.$$

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. Formally, this second assumption requires that there are 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 the maximum of N independent duration times, $M_N = \max_{1\le i\le N}(T_c^i)$, $\lim_{N\to\infty} P\left((M_N - b_N)/a_N \le z\right) = H(z)$ for all z at which $H(z)$ is continuous (see Pickands, 1975).

The requirement that $G(t)$ has unbounded support on the positive real line is needed to ensure that $\xi \ge 0$, which, in turn, ensures that $\lambda(t|\boldsymbol{x}_{it}) \le 1$. In principle, one could allow ξ to be negative, but this would require restrictions on the parameters γ_{0t} and $\boldsymbol{\gamma}$. For example, $\xi = -1$ would yield the discrete proportional hazards model $\lambda(t|\boldsymbol{x}_{it}) = \exp\{\gamma_{0t} + \boldsymbol{x}_{it}^T\boldsymbol{\gamma}\}$.[1] This hazard specification, however, requires the restriction $\gamma_{0t} + \boldsymbol{x}_{it}^T\boldsymbol{\gamma} \le 0$ to rule out hazard rates that are larger than one.

The second assumption requires that $G(t)$ belongs to the maximum domain of 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_c with distribution function $G(t)$, fulfilling the two assumptions above, and for a given large threshold τ, the conditional distribution of T_c given that $T_c \ge a_{t-1} \ge \tau$, which can be expressed as $P(a_{t-1} \le T_c < a_{t-1} + z|T_c \ge a_{t-1})$, is approximately of the form

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

The expression above 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 $\xi \ge 0$ and $0 < z < \infty$ (see, e.g., Coles, 2001). While σ is a function of the threshold level a_{t-1}, it can be shown that ξ is constant for all a_{t-1} above a level τ at which the asymptotic motivation for the generalized Pareto distribution is valid (see,

[1] Strictly speaking, this model exhibits proportional interval hazards only if the predictors in $\boldsymbol{x}$ do not vary over duration time t. Note also that the cloglog model, although being the grouped-duration analogue of the Cox (1972) proportional hazards model, does not exhibit proportional interval hazards (see, e.g., Sueyoshi, 1995).

e.g., Leadbetter et al., 1983, and 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

$$\lambda(t) = P(a_{t-1} \leq T_c < a_t | T_c \geq a_{t-1}) = 1 - \left(1 + \frac{\xi}{\sigma(a_{t-1})}\right)^{-1/\xi}. \tag{7}$$

As in standard extreme value models, explanatory variables can be incorporated into the model by writing

$$\lambda(t|\boldsymbol{x}_{it}) = 1 - \left(1 + \frac{\xi}{\sigma(\gamma_{0t} + \boldsymbol{x}_{it}^T \boldsymbol{\gamma})}\right)^{-1/\xi}$$

for some positive-valued function $\sigma(\cdot)$.[2] In view of the requirement $\sigma > 0$, a natural formulation is $\sigma(\gamma_{0t} + \boldsymbol{x}_{it}^T \boldsymbol{\gamma}) = \exp\{-(\gamma_{0t} + \boldsymbol{x}_{it}^T \boldsymbol{\gamma})\}$, resulting in the hazard rate specification

$$\lambda(t|\boldsymbol{x}_{it}) = 1 - \left(1 + \xi \exp\{\gamma_{0t} + \boldsymbol{x}_{it}^T \boldsymbol{\gamma}\}\right)^{-1/\xi}. \tag{8}$$

The right-hand side of (8) describes the cumulative distribution function of the log-Burr distribution given in (6) with argument $u = \gamma_{0t} + \boldsymbol{x}_{it}^T \boldsymbol{\gamma}$. Due to its relation to the generalized Pareto distribution, the discrete-time duration model resulting from the functional specification in (8) will from now on be referred to as the *Pareto hazard model*.

2.2 Illustration of the Impact of Response Functions

This section provides a brief illustration of the importance of the response function for estimation results. The results shown are based on a simulated data set consisting of 5000 individual duration times. The data generating process (DGP) used to create the simulated data set employs the log-Burr distribution as the response function. Specifically, the true hazard rates are given by

$$\lambda(t|\boldsymbol{x}_{it}) = 1 - (1 + \xi \exp\{\gamma_{0t} + x_{1,i}\gamma_1 + x_{2,i}\gamma_2\})^{-1/\xi},$$

where $\gamma_1 = \gamma_2 = 1$, and $\xi = 5$. The (time-invariant) variables $x_{1,i}$ and $x_{2,i}$ are generated as independent random draws from a normal distribution with zero mean and unit variance and a demeaned Gamma distribution with unit variance, respectively. The baseline hazard is specified as $\gamma_{0t} = -\ln(t)$. All durations exceeding $t = 12$ periods are artificially right-censored.

To illustrate the impact of the choice of response function on estimation results, four different hazard models are used to analyze the simulated data: Pareto models with $\xi = 0$

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

(cloglog), $\xi = 1$ (logit), and $\xi = 5$, and a probit model which is not nested in the class of Pareto models. In all the models estimated, the true set of explanatory variables is used, and the baseline hazard is modeled flexibly by means of duration dummies for each discrete time interval $t \in (1, \ldots, 12)$. Table 1 provides an overview of the impact of the response functions on the estimated covariate effects.

The first two rows of Table 1 show the estimates of γ_1 and γ_2 obtained from the four different hazard models. To enable a meaningful comparison of the estimated coefficients, which in these models are identified only up to a scale factor, the estimates are standardized using the conversion factors proposed by Amemiya (1981). Using these conversion factors, the coefficient estimates are weighted by a factor $F'(\bar{u})$, where $F'(\cdot)$ is the first derivative of the response function used, and $\bar{u}$ is the mean of the index function $u = \gamma_{0t} + \boldsymbol{x}_{it}^T\boldsymbol{\gamma}$ at estimated values of γ. When comparing results obtained from response functions that differ vastly in skewness, as is the case here, standardization by means of Amemiya conversion factors appears to work better than the frequently used standardization based on the variances of the response functions. However, coefficient standardization in this type of models is a delicate issue, and hazard ratios will therefore be used as well when analyzing the strength of covariate effects.

TABLE 1: *Covariate Effects Obtained from Different Response Functions*

	Estimated models			
	Cloglog	Logit	Pareto ($\xi = 5$)	Probit
$\widetilde{\gamma}_1$	0.713	0.815	0.953	0.843
$\widetilde{\gamma}_2$	0.590	0.738	0.967	0.788
$\widetilde{\gamma}_1/\widetilde{\gamma}_2$ ($\equiv \gamma_1/\gamma_2$)	1.209	1.104	0.986	1.070
Hazard Ratio at $t = 1$ (x_1)	1.566	1.643	1.659	1.626
Hazard Ratio at $t = 12$ (x_1)	1.625	1.784	2.177	1.968

Note: The values for $\widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$ are standardized using the conversion factors proposed by Amemiya (1981). The hazard ratios are calculated for an increase in x_1 from zero to one, keeping x_2 fixed at its expected value of zero.

The results shown in Table 1 indicate that the covariate effects are quite accurately estimated when the correct response function is used, and that they are substantially underestimated when any of the misspecified response functions is used. The third row of Table 1 shows the ratios of the estimated covariate effects, and the results indicate that the *relative* effects of explanatory variables are also biased when the response function is misspecified. Lastly, rows four and five show estimated hazard ratios at the shortest and longest duration observed. The hazard ratios are calculated for an increase in x_1 from zero to one, keeping x_2 fixed at its expected value of zero. When the response function is misspecified, the estimated hazard ratios are smaller than their counterparts obtained from the correct specification. This confirms the above result that the choice of response function affects the strength of the estimated covariate effects. Moreover, the (relative) difference in the hazard ratios at $t = 1$ and $t = 12$ varies substantially between the models. For example, when using the cloglog model, the hazard ratio increases only by about 4%. This is well in line with the notion that the cloglog model is the grouped-duration analogue of Cox's proportional hazards model. The Pareto model with $\xi = 5$ and the probit model, however, are decidedly non-proportional, and the estimated hazard ratios increase by about 31% and 21%, respectively.

Figure 1 illustrates the impact of response functions on the estimated hazards. The figure shows the estimated hazard rates obtained from the cloglog model relative to the true hazard rates generated by the Pareto model with $\xi = 5$. Specifically, the true hazard rates are grouped into percentiles, and for each group the corresponding estimated hazards are summarized by means of box plots. The figure shows that the hazard estimates obtained from the cloglog model are substantially biased. Small and large hazard rates are overestimated, whereas medium-sized hazard rates are underestimated.

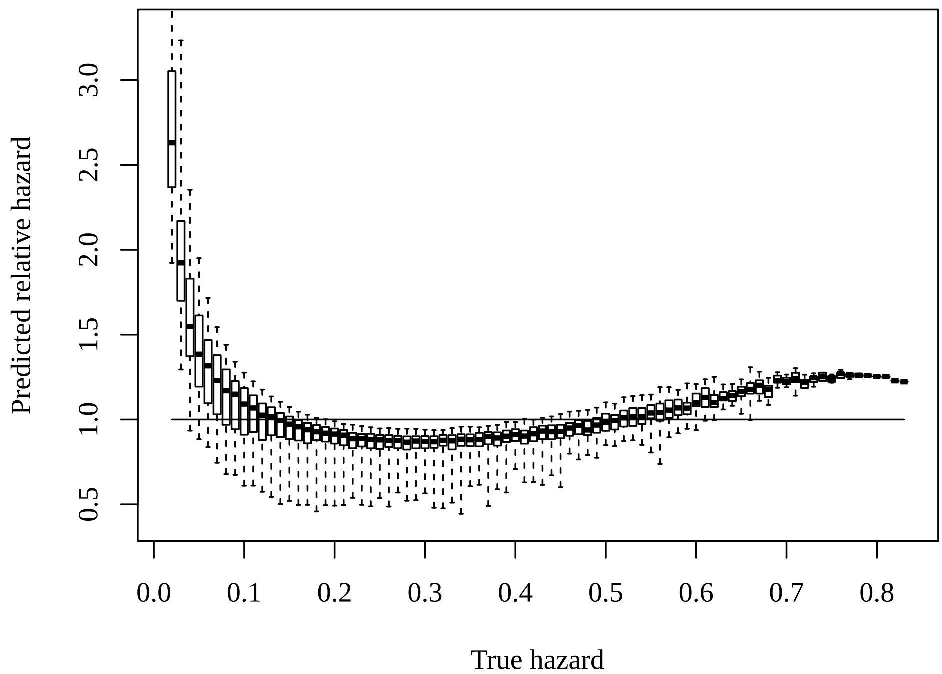

FIGURE 1: *Predicted hazards obtained from a misspecified link function; true model: Pareto ($\xi = 5$); estimated model: cloglog.*

In summary, this illustration suggests that a misspecification of the response function in discrete-time duration models has the following negative impacts on the estimation results. First, the estimated effects of a single covariate, as measured by standardized coefficients and hazard ratios, are biased. Second, the estimated relative effects of two covariates are biased. Third, the imposed degree of proportionality in the hazard rates is incorrect. Fourth and last, the estimated individual hazard rates are biased. An in-depth analysis of these four implications of the choice of the response function is provided in Sections 4 and 5.

3 Extension to Frailty Models and Estimation

3.1 Frailty Models

The basic model (4) does not account for potential unobserved heterogeneity among individuals. As the neglect of unobserved heterogeneity may lead to severe bias in the estimated hazard function (see, e.g., Salant, 1977, Vaupel et al., 1979, and Vaupel and Yashin, 1985, for early discussions of this phenomenon), it is important to allow for unobserved sources of variation in the hazard. An extended model, which includes unobserved heterogeneity, is the so-called frailty model. It assumes for the i^{th} observation the hazard rate

$$\lambda(t|\boldsymbol{x}_{it}, b_i) = F(b_i + \gamma_{0t} + \boldsymbol{x}_{it}^T\boldsymbol{\gamma}), \tag{9}$$

where b_i is a random effect that is assumed to follow a fixed distribution with density $f(\cdot)$, typically chosen as the normal distribution. It has sometimes been argued in the duration literature that a misspecification of the heterogeneity distribution may severely bias the estimation results (see, e.g., Heckman and Singer, 1984, for continuous time and Baker and Melino, 2000, for discrete time). However, Nicoletti and Rondinelli (2010) have shown in an extensive simulation study that using normal random effects in discrete-time models works well, even if the true heterogeneity distribution is not Gaussian. This finding is supported by several empirical studies (see, e.g., Trussell and Richards, 1985, Meyer, 1990, and Dolton and van der Klaauw, 1995). Moreover, the misspecification biases reported by Baker and Melino (2000) have later been shown to be incorrect (see Mroz and Zayats, 2008). This suggests that using normal random effects is a sensible approach when estimating discrete-time duration models.

3.2 Estimation Including Censoring

In the modelling of survival data censoring is a phenomenon that has to be expected. In the case of right censoring, which is considered here, it is only known that T exceeds a certain value but the exact value is not known. Let C_i denote the censoring time and T_i

the exact failure time for observation i. In random censoring it is assumed that T_i and C_i are independent random variables. The observed time is given by $t_i = \min(T_i, C_i)$ as the minimum of survival time T_i and censoring time C_i. It is often useful to introduce an indicator variable for censoring given by

$$\delta_i = \begin{cases} 1 & \text{if } T_i \leq C_i, \\ 0 & \text{if } T_i > C_i, \end{cases}$$

where it is implicitly assumed that censoring occurs at the end of the interval.

Under random censoring the probability of observing (t_i, δ_i) is given by

$$P(t_i, \delta_i|\boldsymbol{x}_i, b_i) = P(T_i = t_i)^{\delta_i} P(T_i > t_i)^{1-\delta_i} P(C_i \geq t_i)^{\delta_i} P(C_i = t_i)^{1-\delta_i},$$

where $\boldsymbol{x}_i = (\boldsymbol{x}_{1,i}, \ldots, \boldsymbol{x}_{p,i})^T$ and $\boldsymbol{x}_{l,i} = (x_{l,i1}, \ldots, x_{l,it_i})$ $(l = 1, \ldots, p)$. It should be noted that the probability is defined given the random effect b_i, which is suppressed on the right hand side of the equation. In the simple survival model without heterogeneity b_i is omitted and the probability is $P(t_i, \delta_i|\boldsymbol{x}_i)$.

If one assumes that the the censoring contributions do not depend on the parameters that determine the survival time (noninformative censoring in the sense of Kalbfleisch and Prentice, 1980), one can separate the factor $c_i = P(C_i \geq t_i)^{\delta_i} P(C_i = t_i)^{1-\delta_i}$ to obtain the simpler form

$$P(t_i, \delta_i|\boldsymbol{x}_i, b_i) = c_i P(T_i = t_i)^{\delta_i} P(T_i > t_i)^{1-\delta_i}.$$

An important tool in discrete survival is that the probability and therefore the corresponding likelihood can be rewritten by using sequences of binary data (see, e.g., Allison, 1982, Singer and Willett, 1993, and Jenkins, 1995, for excellent surveys on the derivation of this likelihood function). By defining for a non-censored observation ($\delta_i = 1$) the sequence $(y_{i1}, \ldots, y_{it_i}) = (0, \ldots, 0, 1)$ and for a censored ($\delta_i = 0$) the sequence $(y_{i1}, \ldots, y_{it_i}) = (0, \ldots, 0)$, the probability (omitting c_i) can be written as

$$P(t_i, \delta_i|\boldsymbol{x}_i, b_i) = \prod_{s=1}^{t_i} \lambda(s|\boldsymbol{x}_i)^{y_{is}} (1 - \lambda(s|\boldsymbol{x}_i))^{1-y_{is}}. \qquad (10)$$

After construction of an appropriate design matrix, the model can be fitted by using software for binary response models (see, e.g., Fahrmeir and Tutz, 1994, and Tutz, 2012).

Alternative estimation procedures are needed for the frailty model. Then, the unconditional probability is given by

$$P(t_i, \delta_i|\boldsymbol{x}_i) = \int P(t_i, \delta_i|\boldsymbol{x}_i, b_i) f(b_i) db_i,$$

and, therefore, by

$$P(t_i, \delta_i|\boldsymbol{x}_i) = \int P(T_i = t_i)^{\delta_i} P(T_i > t_i)^{1-\delta_i} f(b_i) db_i,$$

one obtains

$$P(t_i, \delta_i|\boldsymbol{x}_i) = \int \prod_{s=1}^{t_i} \lambda(s|\boldsymbol{x}_i)^{y_{is}}(1-\lambda(s|\boldsymbol{x}_i))^{1-y_{is}} f(b_i)db_i. \tag{11}$$

This is the unconditional probability of a random effects model for structured binary data.

A practical way to estimate the Pareto hazard model with frailty is to carry out the estimation at a grid of fixed values for the shape parameter ξ. Estimation can then be performed using standard software that allows for user-defined link functions in generalized linear mixed models, such as the R software package `gamlss.mx` (Stasinopoulos and Rigby, 2012). Inference regarding ξ can then be based on the profile log-likelihood $l_p(\xi)$ and its asymptotic χ^2-approximation (see, e.g., Koenker and Yoon, 2009). A $(1-\alpha)$-confidence interval for ξ is thus given by

$$I = \{\xi | 2(l_p(\hat{\xi}) - l_p(\xi)) \leq \chi^2_{1,1-\alpha}\},$$

where $\chi^2_{1,1-\alpha}$ is the $(1-\alpha)$-quantile of the χ^2-distribution with one degree of freedom.

4 Simulation Study

This section evaluates the importance of the response function $F(\cdot)$ used to parameterize the discrete hazard rate by means of simulations. Throughout this section it is assumed that the true response function is of the log-Burr form specified in (6). The focus lies mainly on the effects of misspecifying the functional form of the hazard and on evaluating the performance of the flexible Pareto hazard model in this context. In particular, scenarios are studied where the true response function is heavily right-skewed (i.e., ξ is substantially larger than one) and the estimated model employs a symmetric or left-skewed response function, as is the case for the commonly used logit, probit, and cloglog models. For comparison, results obtained from the correct model specification and from the flexible Pareto specification with an unspecified value of ξ are presented. Estimation is performed using the R software package `gamlss.mx` (Stasinopoulos and Rigby, 2012).

4.1 The Data Generating Process

In setting up the DGP, the true hazard rates are calibrated to resemble those typically observed in data on country-level trade. Empirical studies on the duration of country-level trade have found that bilateral trade relationships are surprisingly short-lived. Typically, the hazard that a trade relationship ceases is about 50% for the first year and steadily declining thereafter (see, e.g., Besedeš and Prusa, 2006a,b, Brenton et al., 2010, and Hess and Persson, 2011). Using this as a benchmark, the DGPs are calibrated to generate hazards with about half the sample exiting at $t = 1$, and that are decreasing with duration.

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

$$\lambda(t|\boldsymbol{x}_{it}, b_i) = 1 - (1 + \xi \exp\{b_i + \gamma_{0t} + x_{1,i}\gamma_1 + x_{2,i}\gamma_2\})^{-1/\xi} .$$

As in the illustrative example of Section 2.2, the (time-invariant) variables $x_{1,i}$ and $x_{2,i}$ are generated as independent random draws from a normal distribution with zero mean and unit variance and a demeaned Gamma distribution with unit variance, respectively. The baseline hazard is specified as $\gamma_{0t} = -\ln(t)$. All durations exceeding $t = 12$ periods are artificially right-censored.

The simulation experiments are then organized in three different sets. In the first set, *Simulation Experiment A*, models without frailty are considered, i.e., $b_i = 0$ for all i. The remaining parameters are fixed at $\gamma_1 = \gamma_2 = 1$ and $\xi = 5$. In the second set, *Simulation Experiment B*, models including frailty are considered, where the individual effects b_i are independent draws from a normal distribution with zero mean and standard deviation $\sigma = 0.5$. The remaining parameters are fixed at $\gamma_1 = 1$, $\gamma_2 = 0$, and $\xi = 5$. The purpose of these two experiments is twofold. One aim is to thoroughly analyze the effects of a misspecified response function in a setting where the true response function is extremely right-skewed and thus rather different from the left-skewed and symmetric response functions of the commonly used cloglog, logit, and probit models. Another aim is to investigate the performance of the Pareto hazard model, which employs a flexible response function. The third set, *Simulation Experiment C*, has the same setup as Experiment B, except that ξ is fixed at a value of one instead of five. The purpose of this experiment is solely to analyze the inference regarding ξ in the presence of unobserved heterogeneity.[3] The value of ξ is fixed at one in this simulation experiment because it is of particular relevance to analyze the Pareto model's capacity of discriminating between the important logit and cloglog specifications.

4.2 Results

This section presents the results obtained from the three simulation experiments described above. In Experiment A and B, five different hazard models are used to analyze the simulated data: Pareto models with $\xi = 0$ (cloglog), $\xi = 1$ (logit), and $\xi = 5$, a Pareto model with an unspecified value of ξ, and a probit model. In Experiment C, a Pareto model with an unspecified value of ξ is estimated. In all the models estimated, the true set of explanatory variables is used. Specifically, in Simulation Experiment A, the models estimated include the covariates $x_{1,i}$ and $x_{2,i}$ and no random intercepts, and in

[3]For the simple model without frailty, Hess (2009) has shown in an extensive simulation study that the shape parameter ξ can be reliably estimated under a variety of circumstances covering variations in sample size, the proportion of censored spells, and the types of explanatory variables included in the model.

Experiment B and C, the models estimated include the covariate $x_{1,i}$ as well as Gaussian random intercepts. The baseline hazard is always modelled in a flexible way by means of duration dummies for each discrete time interval $t \in (1, \ldots, 12)$.

Simulation Experiment A

The upper part of Table 2 shows the estimation results obtained from Simulation Experiment A. The depicted results are the averages and standard deviations over 100 replications of the simulation exercise. The first three rows of Table 2 show the estimates of γ_1 and γ_2 obtained from the five different hazard models. The estimates are standardized using the conversion factors proposed by Amemiya (1981). The number of simulated durations is 1000, resulting in an average number of binary observations of about 5400. The results show that the covariate effects are rather accurately estimated when the hazard model is correctly specified, i.e., when the Pareto hazard model with $\xi = 5$ is used. When the cloglog, logit, or probit model is used for estimation, the covariate effects are substantially underestimated. When the logit or probit model is used, the standardized coefficients are around 10% to 20% smaller than their true value of one. When the cloglog model is used, the standardized coefficients are up to almost 40% smaller. The relative effects of the two covariates are also biased when the response function is misspecified, and again, the relative bias is similar for the logit and probit models and approximately twice as large for the cloglog model.[4] When the flexible Pareto model with an unspecified value of ξ is used, the covariate effects are rather accurately estimated, and the mean estimates differ hardly from those obtained from the correct specification. However, due to the additional uncertainty introduced when ξ is left unspecified, the estimates exhibit slightly larger standard deviations.

Rows four to six of Table 2 contain the same information as the first three rows, but this time the estimates are based on 5000 simulated durations, resulting in an average of approximately 27000 binary observations. With such a large sample size, the covariate effects are very precisely estimated when the true hazard model is used. In this case, there is virtually no bias in the estimated coefficients, and the corresponding standard deviations are very small. When the hazard rate is misspecified, however, the increase in sample size does nothing to improve the bias in the estimated (relative) covariate effects. Again, the results obtained from the flexible Pareto model are almost as accurate as their counterparts obtained from the correct specification.

[4] The result that a misspecification of the response function in discrete-time hazard models causes bias in the estimated relative covariate effects contradicts the results of Nicoletti and Rondinelli (2010). These authors find that misspecifying the functional form of the hazard rate causes only a proportional rescaling of the covariate coefficients. However, Nicoletti and Rondinelli (2010) consider only the special case of two normally distributed covariates, and their findings do not seem to extend to different setups.

TABLE 2: *Comparing Covariate Effects Across Model Specifications*

n		Cloglog	Logit	Pareto ($\xi = 5$)	Pareto ($\xi = \widehat{\xi}$)	Probit
		Estimated models				
		Simulation Experiment A: models without frailty				
1000	$\widetilde{\gamma}_1$	0.772 (0.056)	0.870 (0.060)	1.010 (0.069)	1.021 (0.099)	0.896 (0.059)
	$\widetilde{\gamma}_2$	0.628 (0.062)	0.787 (0.066)	1.028 (0.085)	1.038 (0.097)	0.838 (0.068)
	$\widetilde{\gamma}_1/\widetilde{\gamma}_2$ ($\equiv \gamma_1/\gamma_2$)	1.239 (0.131)	1.112 (0.102)	0.987 (0.090)	0.987 (0.093)	1.075 (0.096)
5000	$\widetilde{\gamma}_1$	0.761 (0.026)	0.861 (0.026)	1.000 (0.030)	0.998 (0.043)	0.887 (0.026)
	$\widetilde{\gamma}_2$	0.601 (0.029)	0.766 (0.031)	1.001 (0.040)	1.000 (0.044)	0.815 (0.031)
	$\widetilde{\gamma}_1/\widetilde{\gamma}_2$ ($\equiv \gamma_1/\gamma_2$)	1.269 (0.070)	1.126 (0.053)	1.000 (0.044)	0.999 (0.046)	1.091 (0.050)
		Simulation Experiment B: models with frailty				
1000	γ_1	0.507 (0.033)	0.618 (0.035)	1.031 (0.096)	1.082 (0.229)	0.349 (0.021)
	σ	0.078 (0.108)	0.055 (0.075)	0.526 (0.349)	0.590 (0.499)	0.056 (0.072)
	HR ($t = 1$)	1.581 (0.052)	1.655 (0.057)	1.751 (0.121)	1.756 (0.151)	1.648 (0.066)
	HR ($t = 12$)	1.642 (0.052)	1.803 (0.058)	2.352 (0.157)	2.380 (0.275)	2.067 (0.130)
5000	γ_1	0.498 (0.014)	0.611 (0.015)	0.994 (0.043)	1.051 (0.127)	0.342 (0.008)
	σ	0.025 (0.031)	0.015 (0.005)	0.442 (0.198)	0.553 (0.305)	0.016 (0.016)
	HR ($t = 1$)	1.567 (0.020)	1.644 (0.023)	1.709 (0.046)	1.716 (0.050)	1.630 (0.023)
	HR ($t = 12$)	1.628 (0.021)	1.792 (0.027)	2.321 (0.083)	2.374 (0.148)	2.042 (0.053)

Note: The results shown are mean values over 100 replications of the simulation experiments. Standard deviations are given in parentheses. The values for $\widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$ are standardized using the conversion factors proposed by Amemiya (1981). The hazard ratios are calculated for an increase in x_1 from zero to one, keeping b_i fixed at its expected value of zero.

Figure 2 provides an overview of the estimated hazard rates obtained from four of the five different models. The results obtained from the correct model specification with $\xi = 5$ are omitted since the corresponding hazard estimates are virtually indistinguishable from their counterparts obtained from the flexible Pareto specification. The estimates are based on 5000 simulated durations. For each of the 100 simulated data sets the true individual hazards are grouped into percentiles, and the average values of the corresponding hazard estimates obtained from the four models relative to the true hazards are calculated. These average relative hazard estimates are then again averaged over the 100 replications of the simulation experiment, and the results are plotted in Figure 2.

The figure shows that the individual hazard rates are, on average, very accurately

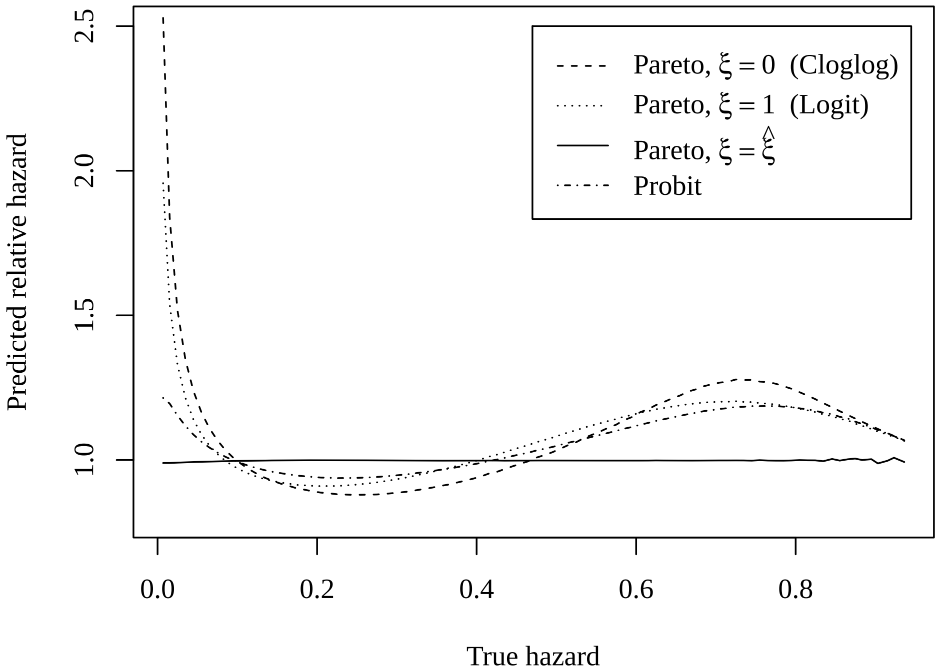

FIGURE 2: *A comparison of predicted hazards obtained from various link functions; average values over 100 replications of Simulation Experiment A; true model: Pareto ($\xi = 5$)*

estimated at all percentiles when the response function is flexibly specified, i.e., when the Pareto hazard model with an unspecified value of ξ is estimated. When the cloglog, logit, or probit model is used for estimation, the average predicted hazards are substantially overestimated at small and large percentiles, and underestimated at mid-sized percentiles. The hazard estimates obtained from the logit and probit models are rather similar over a wide range but differ markedly at extremely low values of the true hazard. This is not surprising, given the similarity of the logistic and the normal distribution, which are both symmetric and exhibit substantial differences only in their tail behavior. The hazard estimates obtained from the cloglog model exhibit a larger bias than the logit and probit estimates. This is likely due to the fact that the cloglog model employs a left-skewed response function, which differs even more from the extremely right-skewed true response function than the symmetric logistic and normal response functions.

Simulation Experiment B

The lower part of Table 2 shows the estimation results obtained from Simulation Experiment B. The depicted results are the averages and standard deviations over 100 replications of the simulation exercise. As opposed to Simulation Experiment A, both the DGP and the models estimated in Simulation Experiment B contain random intercepts. This makes it rather difficult to provide suitably standardized coefficient estimates, since the absolute values of these estimates depend not only on the response function chosen but also on the (possibly incorrectly) estimated variance of the random effects. Thus, the lower part of Table 2 shows estimated hazard ratios in addition to the estimated non-standardized coefficient estimates. The hazard ratios are calculated for an increase in x_1 from zero to one, keeping b_i fixed at its expected value of zero. The corresponding values can then be used instead of standardized coefficients to determine the estimated strength of the covariate effect.

The estimated hazard ratios confirm the findings of Simulation Experiment A that covariate effects are underestimated when a cloglog, logit, or probit model is estimated instead of the true Pareto hazard model with $\xi = 5$. Again, increasing the sample size from 1000 to 5000 renders the mean estimates largely unaffected and does not help improve this bias. Moreover, the estimated hazard ratios reveal another interesting implication of the choice of response function: the models differ substantially with respect to the imposed degree of proportionality, i.e., the (relative) difference in the hazard ratios at $t = 1$ and $t = 12$ varies substantially between the models. When using the cloglog model, the average hazard ratio increases only by about 4%, which is well in line with the notion that the cloglog model is the grouped-duration analogue of Cox's proportional hazards model. The Pareto model with $\xi = 5$ and the probit model, however, are decidedly non-proportional, and the average hazard ratios increase by about 34% and 25%, respectively. The logit model constitutes an intermediate case with an increase of about 9%. Thus, the conventional wisdom regarding the similarity of probit and logit models does not extend to the evaluation of proportionality in the discrete hazard.[5] The results obtained from the flexible Pareto model are again very similar to those obtained from the correct model specification. This suggests that the Pareto model produces reliable results even in the presence of unobserved heterogeneity.

The estimated coefficients shown in the lower part of Table 2 indicate that both γ_1 and σ are rather accurately estimated when the correct model or the flexible Pareto model is used. However, the estimated values of σ reveal rather large standard deviations, in particular when ξ is left unspecified and the sample consists of only 1000 durations. When the number of individual durations is increased to 5000, the estimates of σ become more precise. Although the reported estimates of σ are not standardized, the results shown suggest strongly that the variance of the random effects is incorrectly estimated when the response function is misspecified. All the estimates of σ are very close to zero when the wrong hazard model is used.

[5] This has already been pointed out by Sueyoshi (1995).

Simulation Experiment C

Figure 3 shows the average profile log-likelihood for ξ obtained from 100 replications of Simulation Experiment C. The models estimated include Gaussian random intercepts, and estimation is performed for fixed values of ξ at the regular grid $0, 0.1, \dots, 3.9$. The estimations are based on 5000 individual duration times.

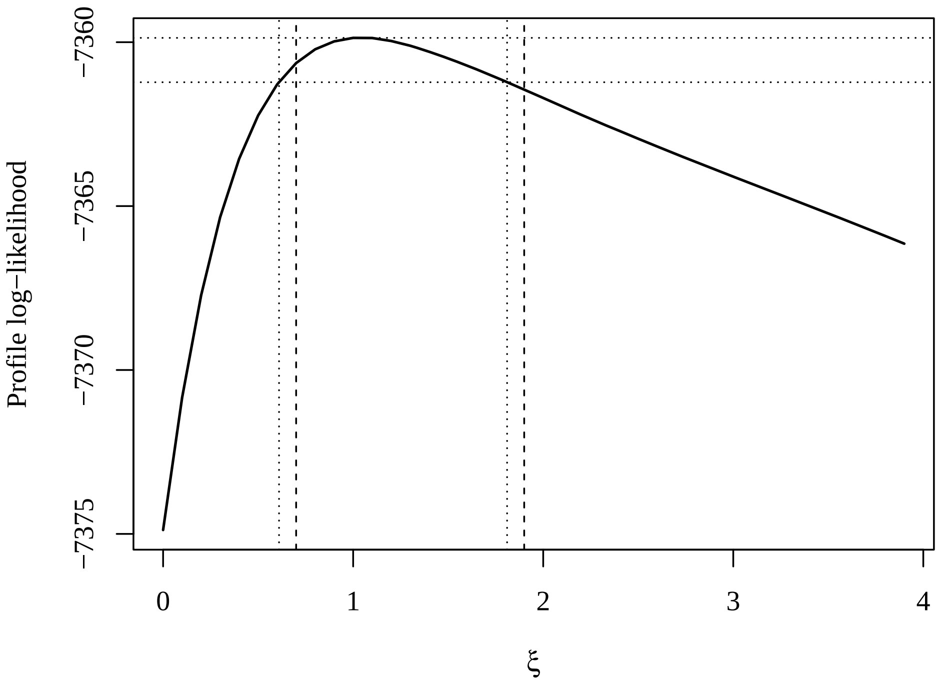

FIGURE 3: *Profile log-likelihood at a grid of ξ-values; average values over 100 replications of Simulation Experiment C; the true value of ξ is one; the maximum of the profile log-likelihood is reached at $\xi = 1$; the dashed vertical lines indicate a simulated 90%-confidence interval; the dotted vertical lines indicate an asymptotic 90%-confidence interval; the difference between the dotted horizontal lines equals $\chi^2_{1,0.9}/2$, where $\chi^2_{1,0.9}$ is the 90%-quantile of the χ^2_1-distribution*

The figure shows that the average profile log-likelihood reaches its maximum at $\xi = 1$, which equals the true value of the shape parameter. In 90 out of 100 cases, the maximum of the profile log-likelihood was reached for $\xi \in [0.7; 1.9]$. This simulated 90%-confidence interval is depicted by the two dashed vertical lines. If one is willing to consider the function in Figure 3 as a single, representative profile log-likelihood, an asymptotic 90%-confidence interval for ξ would be

$$I = \{\xi | 2(l_p(\hat{\xi}) - l_p(\xi)) \leq \chi^2_{1,0.9}\},$$

where $\chi^2_{1,0.9}$ is the 90%-quantile of the χ^2-distribution with one degree of freedom. This confidence interval is depicted by the two dotted vertical lines. Reassuringly, both intervals are very similar. It is noteworthy that the profile log-likelihood is rather asymmetric. Consequently, the confidence intervals are not symmetric around one, which is the estimated (and the true) value of ξ. However, this is not surprising, given the fact that the hazard specification is closely related to the generalized Pareto distribution. This distribution is frequently used in extreme value analysis, and in this context, it is well-known that the profile log-likelihood for ξ is asymmetric in finite samples (see, e.g., Coles, 2001). The profile log-likelihood is rather steep for small values of ξ. This facilitates discrimination between the cloglog and the logit model, which are important special cases of the Pareto hazard model. However, the profile log-likelihood becomes increasingly flat as ξ increases, and if the true value of ξ is large, larger samples are needed to obtain precise estimates of the shape parameter.

To summarize, the results of this simulation study confirm that the choice of response function in discrete-time duration models has an important effect on estimation results. Specifically, a misspecified response function causes four types of bias. First, the strength of individual covariate effects, as measured by standardized coefficients and hazard ratios, is biased. Second, the relative effects of covariates, as measured by coefficient ratios, is biased. Third, the degree of proportionality, as measured by the change in hazard ratios across duration time, is biased. Fourth and last, the predicted individual hazards are biased. The simulation study also shows that the shape parameter in the Pareto hazard model can be reliably estimated, even in the presence of unobserved heterogeneity, and that the model produces accurate results.

5 Empirical Application

This section analyzes the performance of the Pareto hazard model in an empirical analysis of US import durations. The empirical analysis serves also to investigate whether the choice of response function in discrete-time duration models matters in practice.

5.1 Data and Model Specification

The empirical application employs the same data on US imports as previously analyzed by Besedeš and Prusa (2006b) in their influential seminal study on the duration of trade, and later re-examined by Hess and Persson (2012). The data record annual US imports between 1972 and 1988 from virtually every trading partner in the world, and they also include information on the value of imports, customs collected, and other relevant factors that might affect the duration of trade. The traded products are classified according to the 7-digit Tariff Schedule of the United States (TSUSA), which amounts to a total of some 20 000 products. A *trade relationship* is then defined as a certain product being imported from one specific exporter. A *trade spell* is defined as a period of time with uninterrupted import of a given product from one specific country. These spells of trade constitute the core units of analysis, and the *spell duration* is simply calculated as the number of consecutive years with non-zero imports. The empirical models estimated in this section employ the same set of explanatory variables as used in the original analysis by Besedeš and Prusa (2006b). Specifically, transportation costs, the exporters gross domestic products (GDP), tariff rates, changes in the relative real exchange rate, coefficients of variations of unit values, a multiple spell dummy, and dummies for agricultural good, reference priced products, and homogeneous goods are used to explain the duration of trade. A detailed description of these covariates is provided by Besedeš and Prusa (2006b).

5.2 Results

Eight different hazard models were used to analyze the trade duration data: cloglog, logit, probit, and Pareto models with and without Gaussian random intercepts. Table 3 shows that the results are qualitatively similar for the various estimation procedures. None of the estimated coefficients changes sign across model specifications. While higher transportation costs increase the hazard that a trade relationship ceases, a higher GDP of the trading partner, a higher industry level tariff rate, a real depreciation of the exporting country's currency, and a larger coefficient of variation of unit values decrease the hazard. Higher order spells have an increased hazard and so do trade relationships involving agricultural goods, reference priced products, and homogeneous goods.[6]

Table 3 also shows the estimated ξ-values obtained from the Pareto hazard models. The estimated value of ξ is approximately 4.3 when unobserved heterogeneity is not accounted for (Model 3). This model is estimated using a user-written Stata program based on the maximum likelihood routine `ml d1`. When unobserved heterogeneity is modelled through the inclusion of random intercepts (Model 7), the Pareto model is estimated at a regular grid of fixed ξ-values $(0, 0.1, \ldots, 5.9)$ using the R software package `gamlss.mx` (Stasinopoulos and Rigby, 2012). Figure 4 shows the resulting profile log-likelihood, which reaches its maximum at $\xi = 4.4$. When comparing the maximized log-likelihoods of the various models, which are given in the last row of Table 3, the largest value is found for the Pareto model with frailty. It is noteworthy that the second largest value is reached by the Pareto model without frailty, which outperforms even the cloglog, logit, and probit models that include random effects. This suggests that the Pareto model is best suited to analyze the data at hand. The question arising is then whether the results obtained from the preferred Pareto model differ in a notable

[6]See Besedeš and Prusa (2006b) for a detailed discussion of these results.

TABLE 3: *Estimation Results for 1972-1988 7-Digit TSUSA Data*

	Models without frailty				Models with frailty			
	Cloglog (1)	Logit (2)	Pareto (3)	Probit (4)	Cloglog (5)	Logit (6)	Pareto (7)	Probit (8)
Ad valorem transportation costs (unit=10%)	0.0954	0.1291	0.2297	0.0774	0.1017	0.1356	0.2443	0.0809
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
GDP (unit=$100 billions)	−0.0265	−0.0305	−0.0424	−0.0165	−0.0294	−0.0337	−0.0476	−0.0185
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Tariff rate, 4-digit SITC (unit=1%)	−0.0254	−0.0297	−0.0434	−0.0168	−0.0270	−0.0315	−0.0469	−0.0179
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
%Δ relative real exchange rate (unit=10%)	−0.1246	−0.1466	−0.1894	−0.0627	−0.1316	−0.1526	−0.1996	−0.0659
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Coefficient of variation of unit values	−0.0860	−0.1030	−0.1513	−0.0582	−0.0905	−0.1084	−0.1626	−0.0615
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Multiple spell dummy	0.3662	0.4843	0.8723	0.2956	0.3120	0.4339	0.8256	0.2608
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Agricultural goods (dummy)	0.1169	0.1764	0.3241	0.1045	0.1548	0.2125	0.3783	0.1278
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Reference priced products (dummy)	0.1973	0.2516	0.4090	0.1455	0.2363	0.2903	0.4685	0.1702
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Homogeneous goods (dummy)	0.3095	0.4137	0.7060	0.2441	0.3751	0.4775	0.8039	0.2843
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
ξ	0	1	4.2909		0	1	4.4	
			(0.000)					
σ					0.4643	0.5074	0.7256	0.3093
					(0.000)	(0.000)	(0.000)	(0.000)
Log likelihood	−405 378	−404 571	−403 801	−404 539	−404 978	−404 243	−403 570	−404 259

Note: P-values in parentheses. All models include duration dummies. Models with frailty include exporter-product random effects. The parameter ξ denotes the shape parameter in the *Pareto* model, and ρ denotes the fraction of the error variance that is due to variation in the unobserved individual factors (models with frailty only). A trade relation is defined as an exporter-product combination. The number of observations is given by the total number of years with observed trade for all trade relationships.

way from the results obtained from the conventional cloglog, logit, and probit models. In what follows, this question will be investigated by comparing the various estimation results based on several quantitative indicators. As in the simulation study presented in Section 4, these quantitative indicators comprise the strength of covariate effects (as measured by standardized coefficients and hazard ratios), the relative effects of covariates, the degree of proportionality in the covariate effects, and the estimated hazard rates.

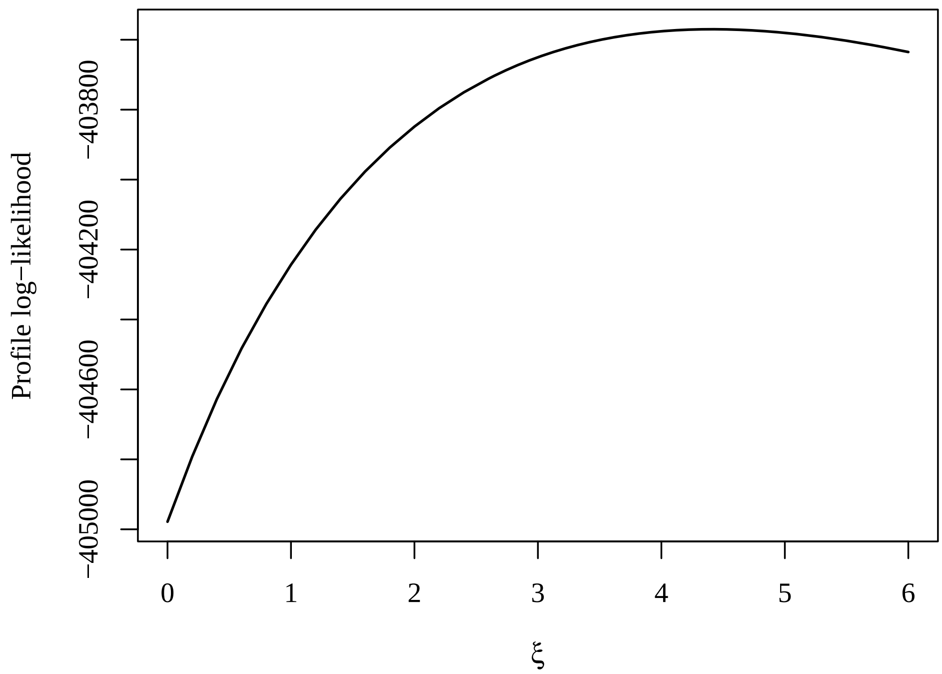

FIGURE 4: *Profile log-likelihood at a grid of ξ-values*

Despite the qualitative similarity of the estimated coefficients shown in Table 3, the estimates obtained from the different model specifications exhibit substantial quantitative differences. The first difference between the models pertains to the estimated strength of covariate effects. To provide a first impression of the estimated covariate effects, Table 4 shows the coefficient estimates from Table 3, standardized by the conversion factors proposed by Amemiya (1981). Since the scale of the estimated coefficients in frailty models depends on both the response function and the variance of the random effects, which makes an appropriate standardization of the coefficients rather difficult, the results in Table 4 are based on the four models without frailty (Models 1–4 in Table 3). The results indicate that the estimated covariate effects obtained from the preferred Pareto model with $\xi \approx 4.3$ are, by and large, larger in absolute terms than their counterparts obtained from the other three models. The estimates obtained from the probit model are always smaller in absolute terms than their counterparts from the Pareto model, the estimates obtained from the cloglog model are smaller for all covariates except 'exchange rate', and the estimates obtained from the logit model are smaller in all but two cases.

TABLE 4: *Standardized Coefficient Estimates*

	Cloglog ($\xi = 0$)	Logit ($\xi = 1$)	Pareto ($\xi \approx 4.3$)	Probit
Transportation costs	0.1482	0.1797	0.2297	0.1886
GDP	−0.0412	−0.0425	−0.0424	−0.0402
Tariff rate	−0.0395	−0.0413	−0.0434	−0.0409
Exchange rate	−0.1936	−0.2041	−0.1894	−0.1528
Coefficient of variation	−0.1336	−0.1434	−0.1513	−0.1418
Multiple spell dummy	0.5690	0.6741	0.8723	0.7202
Agricultural goods	0.1816	0.2455	0.3241	0.2546
Reference priced products	0.3066	0.3502	0.4090	0.3545
Homogeneous goods	0.4809	0.5758	0.7060	0.5947

Note: All models are without frailty (Models 1–4 in Table 3). The coefficient estimates are standardized using the conversion factors proposed by Amemiya (1981).

The results in Table 4 suggest that the estimated average effects of explanatory variables are affected by the response function chosen to specify the discrete hazard. However, when comparing discrete hazard models with different response functions, it is generally better to focus on relative covariate effects and hazard ratios rather than comparing the estimates of the coefficients even after an appropriate conversion. Table 5 therefore shows the relative effect of the two arbitrarily chosen covariates 'transportation costs' and 'exchange rate', as well as the respective hazard ratios at the shortest and longest duration observed. The results in Table 5 are based on the four frailty models (Models 5–8 in Table 3). The hazard ratios are calculated for a unit increase in the respective covariate, keeping all other covariates constant at their mean values and assuming that the individual effect b_i equals its expected value of zero.

For all the four models, the hazard ratio of a unit increase in transportation costs is roughly equal at $t = 1$. At $t = 16$, however, the hazard ratio obtained from the Pareto model is larger than the remaining hazard ratios. This suggests that the estimated average effect of transportation costs is larger in the Pareto model than in the remaining three models. When comparing the sizes of the hazard ratios at $t = 16$, the Pareto model yields the largest value, followed by probit, logit, and cloglog. This is exactly in line with the respective standardized coefficients presented in Table 4. When comparing the hazard ratios of a unit increase in the exchange rate, the results are ambiguous. At $t = 1$ the cloglog and logit specifications yield larger effects than the Pareto and probit specifications, and *vice versa* at $t = 16$. However, since every trade relation lasts for at least one year while only a few last for 16 years, the effects at $t = 1$ can be expected to have a greater impact on the average effect. The cloglog and logit models would thus yield larger average effects of changes in the exchange rate than the Pareto and probit models. This is also in line with the results in Table 4.

TABLE 5: *Comparing Covariate Effects across Model Specifications*

	Cloglog with frailty	Logit with frailty	Pareto ($\xi = 4.4$) with frailty	Probit with frailty
$\gamma^{transport}/\gamma^{exchange}$	-0.772	-0.888	-1.224	-1.227
Hazard Ratio (*transport*) at $t = 1$	1.088	1.097	1.094	1.094
Hazard Ratio (*transport*) at $t = 16$	1.106	1.144	1.269	1.244
Hazard Ratio (*exchange*) at $t = 1$	0.895	0.897	0.923	0.926
Hazard Ratio (*exchange*) at $t = 16$	0.877	0.860	0.822	0.833

Note: The hazard ratios for a unit increase in the relative effects of transportation costs and exchange rate differences are calculated for the shortest and longest duration observed. All other covariates are kept constant at their mean values.

A second aspect associated with the hazard ratios presented in Table 5 is the degree of proportionality imposed by the different hazard models. With the cloglog model, the hazard ratio for transportation costs increases by less than 2% from $t = 1$ to $t = 16$, and the hazard ratio for exchange rate decreases by approximately 2%. In other words, the grouped-duration analogue of the proportional Cox model exhibits almost proportional interval hazards. The Pareto model, however, yields interval hazards that are markedly less proportional. The changes in the hazard ratio are 16% and -9%, respectively. The probit model exhibits changes in the hazard ratio that are very similar to those of the Pareto model, and the logit model constitutes an intermediate case.

Lastly, Table 5 also shows the estimated relative effect of transportation costs and exchange rate. Since the relative effects of covariates are scale-independent, they can be compared directly across different model specifications. The results in Table 5 show that the four models analyzed here differ substantially with respect to the estimated relative covariate effects. The effect of transportation costs is about 22% larger than the effect of exchange rate when the Pareto or the probit model is used for estimation, whereas the logit and the cloglog specification yield an effect that is 11% and, respectively, 23% smaller.

The above results have shown that the choice of response function has important implications for the estimated covariate effects in discrete-time hazard models. Figure 5 shows that the choice of response function also affects the estimated hazard rates. The figure shows the predicted hazards obtained from the four models relative to the predicted hazards obtained from the preferred Pareto model with $\xi = 4.4$. More precisely, the predicted hazards obtained from the Pareto model are grouped into percentiles, and for each percentile the four average relative hazards are shown. As compared to the Pareto model, which serves as the benchmark, the cloglog, logit, and probit models all yield larger predicted hazards at high percentiles, and smaller predicted hazards at mid-sized percentiles. At low percentiles, the cloglog and logit models yield larger hazards, whereas the probit model yields smaller hazards.

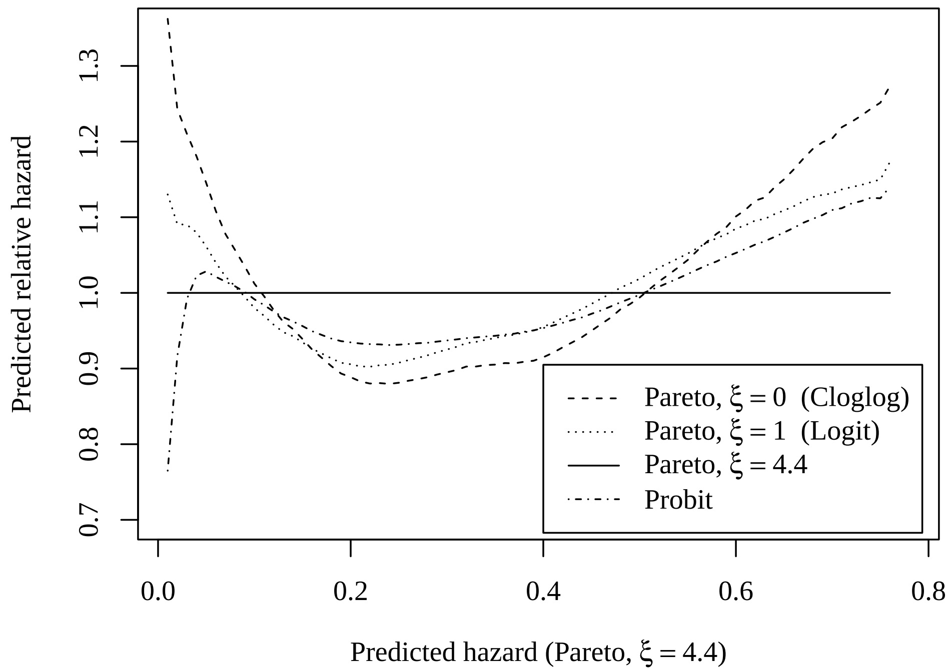

FIGURE 5: *A comparison of predicted hazards obtained from various link functions; average values per percentile; benchmark model: Pareto ($\xi = 4.4$); all the predictions account for individual frailty*

6 Conclusions

This paper introduces a new hazard rate model for discrete duration data. As is well known, discrete-time duration models can be estimated using a conventional binary response regression approach. As shown in this paper, however, the choice of link function used in the binary regression model is not innocuous in a duration context, and has important implications for the estimated covariate effects and the predicted hazards. The model proposed in this paper therefore employs a flexible link function. Specifically, the cumulative distribution function of the log-Burr distribution is proposed for parameterizing the discrete hazard. This distribution function contains a shape parameter, ξ, which can be estimated along with the other coefficients included in the regression model. For $\xi = 1$ one obtains the logit link, and for $\xi = 0$ one obtains the cloglog link. Thus, the hazard model proposed comprises the two models that are most widely used in discrete survival modelling, the proportional odds model and the grouped proportional hazards model. Moreover, 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. This provides new insights into the relation between continuous-time duration processes and discrete-time hazard specifications. Since the asymptotic distribution of threshold excesses is given by the generalized Pareto distribution, the hazard model proposed here is referred to as the *Pareto hazard model*.

Using simulations, it is shown that the shape parameter in the Pareto hazard model can be reliably estimated, even in the presence of unobserved heterogeneity, and that the model produces accurate results. Moreover, it is shown that a misspecified response function causes severe biases in the estimation results. Specifically, the strength of individual covariate effects, the relative effects of covariates, the degree of proportionality, and the predicted individual hazards are biased when using a misspecified response function in the hazard model. Employing a flexible response function may therefore help to avoid severely biased estimation results. An empirical analysis of trade durations confirms the finding that using a flexible response function in discrete-time duration models may be very useful in practice. Specifically, when analyzing the same data on US import durations as previously used by Besedeš and Prusa (2006b) in their influential seminal study on the duration of trade, the Pareto model proves to outperform the conventional cloglog, logit, and probit specifications in terms of model fit. Moreover, the shape parameter ξ in the Pareto model is estimated to be significantly larger than four in this empirical application, and hence the corresponding results differ strongly from their counterparts obtained from the cloglog model with $\xi = 0$ and the logit model with $\xi = 1$. The results, in particular the predicted hazards, differ also from those obtained from the probit model, which is not nested in the class of Pareto hazard models.

References

Allison, P. D. (1982). Discrete-time methods for the analysis of event histories. *Sociological Methodology 13*, 61–98.

Amemiya, T. (1981). Qualitative response models: a survey. *Journal of Economic Literature 19(4)*, 1483–1536.

Baker, M. and A. Melino (2000). Duration dependence and nonparametric heterogeneity: a Monte Carlo study. *Journal of Econometrics 96(2)*, 357–393.

Besedeš, T. and T. J. Prusa (2006a). Ins, outs and the duration of trade. *Canadian Journal of Economics 39(1)*, 266–295.

Besedeš, T. and T. J. Prusa (2006b). Product differentiation and duration of US import trade. *Journal of International Economics 70(2)*, 339–358.

Blanchard, O. J. and P. Diamond (1994). Ranking, unemployment duration, and wages. *Review of Economic Studies 61(3)*, 417–434.

Brenton, P., C. Saborowski, and E. von Uexkull (2010). What explains the low survival rate of developing country export flows? *The World Bank Economic Review 24(3)*, 474–499.

Burr, I. W. (1942). Cumulative frequency functions. *Annals of Mathematical Statistics 13(2)*, 215–232.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B 34(2)*, 187–220.

Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society, Series B 52(3)*, 393–442.

Dolton, P. and W. van der Klaauw (1995). Leaving teaching in the UK: a duration analysis. *The Economic Journal 105(429)*, 431–444.

Embrechts, P., C. Klüppelberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer.

Fahrmeir, L. and G. Tutz (1994). *Multivariate Statistical Modelling based on Generalized Linear Models*. New York: Springer.

Fisher, R. A. and L. H. C. Tippett (1928). On the estimation of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society 24(2)*, 180–190.

Heckman, J. J. and B. Singer (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica 52(2)*, 271–320.

Hess, W. (2009). A flexible hazard rate model for grouped duration data. *Working Paper No. 2009:18, Department of Economics, Lund University.*

Hess, W. and M. Persson (2011). Exploring the duration of EU imports. *Review of World Economics 147(4)*, 665–692.

Hess, W. and M. Persson (2012). The duration of trade revisited. Continuous-time vs. discrete-time hazards. *Empirical Economics 43(3)*, 1083–1107.

Jenkins, S. P. (1995). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics 57(1)*, 129–137.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous Univariate Distributions*, Volume 2, 2nd ed. New York: John Wiley & Sons.

Kalbfleisch, J. D. and R. L. Prentice (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika 60(2)*, 267–278.

Kalbfleisch, J. D. and R. L. Prentice (1980). *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons.

Koenker, R. and J. Yoon (2009). Parametric links for binary choice models: a Fisherian-Bayesian colloquy. *Journal of Econometrics 152(2)*, 120–130.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: Wiley.

Leadbetter, M. R., G. Lindgren, and H. Rootzén (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 9(3)*, 321–334.

Meyer, B. D. (1990). Unemplyment insurance and unemployment spells. *Econometrica 58(4)*, 757–782.

Mroz, T. A. and Y. V. Zayats (2008). Arbitrarily normalized coefficients, information sets, and false reports of "biases" in binary outcome models. *Review of Economics and Statistics 90(3)*, 406–413.

Nicoletti, C. and C. Rondinelli (2010). The (mis)specification of discrete duration models with unobserved heterogeneity: a Monte Carlo study. *Journal of Econometrics 159(1)*, 1–13.

Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics 3(1)*, 119–131.

Prentice, R. L. (1975). Discrimination among some parametric models. *Biometrika 62(3)*, 607–614.

Prentice, R. L. (1976). A generalization of the probit and logit methods for dose response curves. *Biometrics 32(4)*, 761–768.

Prentice, R. L. and L. A. Gloeckler (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics 34(1)*, 57–67.

Salant, S. W. (1977). Search theory and duration data: a theory of sorts. *Quarterly Journal of Economics 91(1)*, 39–57.

Singer, J. D. and J. B. Willett (1993). It's about time: using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics 18(2)*, 155–195.

Stasinopoulos, M. and B. Rigby (2012). *gamlss.mx: A GAMLSS add on package for fitting mixture distributions.* R package version 4.1-0.

Sueyoshi, G. T. (1995). A class of binary response models for grouped duration data. *Journal of Applied Econometrics 10(4)*, 411–431.

Tadikamalla, P. R. (1980). A look at the Burr and related distributions. *International Statistical Review 48(3)*, 337–344.

Trussell, J. and T. Richards (1985). Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure. *Sociological Methodology 15*, 242–276.

Tutz, G. (2012). *Regression for Categorical Data.* Cambridge: Cambridge University Press.

van den Berg, G. J. (1990a). Nonstationarity in job search theory. *Review of Economic Studies 57(2)*, 255–277.

van den Berg, G. J. (1990b). Search behaviour, transitions to non-participation and the duration of unemployment. *The Economic Journal 100(402)*, 842–865.

Vaupel, J. W., K. G. Manton, and E. Stallard (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography 16(3)*, 439–454.

Vaupel, J. W. and A. I. Yashin (1985). Heterogeneity's ruses: some surprising effects of selection on population dynamics. *The American Statistician 39(3)*, 176–185.