The Extended Generalized Gamma Model and its Special Cases: Applications to Modeling Marriage Durations

Gebrenegus Ghilagaber*

Abstract

The paper demonstrates how various parametric models for duration data such as the exponential, Weibull, and lognormal may be embedded in a single framework, and how such competing models may be assessed relative to a more comprehensive one. To illustrate the issues addressed, the survival patterns of marriages among 1203 Swedish men born 1936-1964 are studied by parametric and non-parametric survival methods. In particular, we study the sensitivity of model-choice with respect to level of aggregation of the time variable; and of covariate-effects with respect to the model chosen. In accordance with previous works our empirical results indicate that the choice of a parametric model for the duration variable is affected by the level of time aggregation. In contrast to previous results, however, our analysis shows that estimates of covariate effects are not always robust to distributional assumptions for the duration variable.

1 Introduction

Survival data typically contain information on the date a sample member entered a social state such as marriage, employment, clinical trial; the date the state was left, if at all left; and the value of the next state entered. Further, the data usually contain information on some sociodemographic covariates of the sample member.

*Address for Correspondence: Stockholm University, Department of Statistics, SE-106 91, Stockholm - Sweden. Tel. 46-8-162983, Fax. 46-8-167511, E-mail: Gebre@stat.su.se.

In analyzing data of this nature, interest focuses on examining the effects of covariates on the rate at which the event of interest occurs. Alternatively, one may focus on the duration spent in a state and examine its relationship to the covariates. In the present paper, we describe a number of models for the analysis of right censored survival data. Models for the hazard rate as well as for the duration to occurrence of an event are considered.

The paper has a number of purposes. First, it describes how a number of the parametric models such as the exponential, Weibull, and lognormal may be embedded in a single parametric framework, and how each competing model may be assessed relative to a more comprehensive one. Cox's semi-parametric model whose estimation is based on partial likelihood is also described and used as a standard for comparison of effect estimates. A second purpose is to analyze a real life data using various models and examine the sensitivity of model choice to time aggregation, and of parameter estimates to choice of a particular model. A comparison with previous similar works is also attempted. We illustrate the issues addressed with data on divorce among Swedish men with a view of describing the distributional shape of marriage durations and examining its sociodemographic correlates.

Our empirical results with regard to the parametric family of models shows that the number and type of models rejected in favor of a more comprehensive model depends on the level of time aggregation. More importantly, and in contrast to some previous works, it is demonstrated here that parameter estimates and inference based on rejected models differ from those based on relatively adequate models.

The rest of the paper is organized as follows. In Section 2, we introduce the proportional hazards model for the rate of occurrence of an event of interest. Section 3 is devoted to a discussion of accelerated-failure-time models - models for the duration until the occurrence of the event. We describe the Extended Generalized Gamma model and demonstrate how a number of common parametric models may embedded within this parent model. The models discussed in Sections 2 - 3 are fit to data on family dissolution among Swedish men and the results are described in Section 4. The last section summarizes the paper while our empirical findings are tabulated in an appendix.

2 Hazard-Rate Models: Cox's Proportional Hazards Model

A central concept in the analysis of data representing times to occurrence of some specified event is the hazard function. Such a function, commonly denoted by $\lambda(t)$, is defined as the instantaneous rate at which the event occurs at a specific point t of a (non-negative) time variable T :

$$\lambda(t) = \lim_{\Delta t \longrightarrow 0} \frac{P\left[t < T \leqslant t + \Delta t | T \geq t\right]}{\Delta t} \tag{1}$$

Hazard rates can vary not only over time, but also among individuals within a population. Thus, one objective in the analysis of time-to-event data is to draw inferences about the influence of covariates on the hazard function.

In his influential paper, Cox (1972) proposed a model where a vector of covariates $\mathbf{z}$ affects the hazard function in a multiplicative manner according to

$$\lambda(t|\mathbf{z}) = \lambda_0(t) \exp\left[\mathbf{z}\boldsymbol{\beta}\right] \tag{2}$$

where $\lambda_0(t)$ is an unspecified base-line function of time and $\boldsymbol{\beta}$ is an unknown vector of parameters representing the effect of the covariate $\mathbf{z}$. The factor $exp(\mathbf{z}\boldsymbol{\beta})$ describes the hazard of failure for an individual with vector z **relative** to that of a standard (with $\mathbf{z} = \mathbf{0}$). Details on estimation and tests on $\boldsymbol{\beta}$ may be found in Cox (1975).

3 Accelerated Failure-Time Models

3.1 Introduction

A second class of models, more akin to ordinary linear regression, specifies the covariates to act multiplicatively on failure time itself (or linearly on log-failure time) rather than on the hazard function (as in the proportional hazards model above).

Thus, if T_0 is the time (duration) associated with the baseline distribution corresponding to zero values for the covariates ($\mathbf{z} = \mathbf{0}$), then the accelerated failure time model specifies that if the vector of covariates had been $\mathbf{z}$ ($\mathbf{z} \neq \mathbf{0}$), the event time (duration) would have been

$$T = T_0 exp(\mathbf{z}\boldsymbol{\beta}) \tag{3}$$

or equivalently, that

$$\ln T = \ln T_0 + \mathbf{z}\boldsymbol{\beta} \tag{4}$$

where, as before, T is the vector of failure times, $\mathbf{z}$ is a vector of covariates or independent variables, $\boldsymbol{\beta}$ is a vector of unknown regression parameters. Since covariates alter, by a scale factor, the rate at which an individual traverses the time axis, (3) may be referred to as the accelerated failure time model. Thus, for proportional hazards model (2), the explanatory variables act multiplicatively on the baseline hazard so that their effect is to increase or decrease the hazard relative to $\lambda_0(t)$. For accelerated life models, on the other hand, the explanatory variables act multiplicatively on time to the event so that their effect is to accelerate or decelerate time to failure relative to T_0.

The model in (4) is a linear model with lnT_0 playing the role of an error term with an underlying baseline distribution. Usually, an intercept term α and a scale parameter δ are allowed in the model to give

$$lnT = \alpha + \mathbf{z}\beta + \delta lnT_0 \tag{5}$$

In terms of the original (untransformed) event times, the effect of the intercept term and the scale factor are to scale and power the event time, respectively:

$$T = exp(\alpha + \mathbf{z}\boldsymbol{\beta} + \delta lnT_0) = T_0^{\delta} exp(\alpha) exp(\mathbf{z}\boldsymbol{\beta}) \tag{6}$$

In other words, the effect of covariates in an accelerated failure-time model is to change the scale, but not the location, of a baseline distribution of failure times.

One point that is worth noting at this stage is that the parameterizations in (2) and (3) are different. A positive coefficient in (2) implies an increased hazard (shorter duration) while in (3) it implies longer duration (decreased hazard) relative to that of the baseline (where covariates assume the value of zero).

3.2 The Extended Generalized Gamma Model & its Special Cases

3.2.1 The choice between alternative baseline models

As we saw above the model for the response variable (5) consists of a linear effect composed of the covariables together with a random disturbance term. Such models may be rewritten more explicitly as

$$lnT = z\beta^* + \delta\epsilon \tag{7}$$

in which the intercept is incorporated in the coefficient vector β^* and a more conventional notation is used for the random error term. The distribution of the random error term can be taken from a class of distributions that includes the extreme-value, normal, and logistic distributions, and, by using a log transformation, exponential, Weibull, lognormal, loglogistic and gamma distributions. In general, the distribution may depend on additional shape parameter k.

Embedding competing models in a single parametric framework allows the methods of ordinary parametric inference to be used for discrimination and leads to an assessment of each competing model relative to a more comprehensive one. Stacy (1962) showed that the generalized gamma model can be useful in this regard.

The generalized-gamma model is the distribution of T such that $lnT = z\beta^* + \delta\epsilon$, where the random error term ϵ has the density;

$$f(k,\epsilon) = \frac{1}{\Gamma(k)} \exp\left[k\epsilon - \exp(\epsilon)\right], -\infty < z\beta^* < \infty; -\infty < \epsilon < \infty;\ \delta, k > 0 \tag{8}$$

Prentice (1974) showed that a transformation of the form $w = k^{-\frac{1}{2}}(\epsilon - lnk)$ leads to a standard normal distribution for w as $k \to \infty$. Further, he extended the generalized gamma model by setting $q = k^{-\frac{1}{2}}$ and by allowing the error density at $-q$ to be a reflection, about the origin, of that of q. The parameter $q = k^{-\frac{1}{2}}$ was chosen as the unique power of k that leads to finite, nonzero likelihood derivatives at the lognormal model for T. The final model with parameters $-\infty < z\beta^* < \infty$; $q < \infty$; and $\delta > 0$, can be written as $lnT = z\beta^* + \delta\epsilon$ where the error density function $f(q,\epsilon)$ is

$$f(q,\epsilon) = \begin{cases} \frac{|q|}{\Gamma(q^{-2})}(q^{-2})^{q^{-2}} \exp\left\{q^{-2}\left[q\epsilon - \exp(q\epsilon)\right]\right\}, & q \neq 0 \\ \frac{1}{\sqrt{2\pi}} \exp(-\frac{\epsilon^2}{2}), & q = 0 \end{cases} \tag{9}$$

The distribution of T, when the error term has the density (9), will henceforth, be called the Extended Generalized Gamma (EGG) distribution.

As can be seen from the lower part of (9) the EGG model reduces to the standard normal distribution for ϵ when the shape parameter q is equal to zero. Accordingly, T will have a log-normal distribution. When the shape parameter q equals 1, (9) reduces to

$$f(q,\epsilon) = \exp\left\{\epsilon - \exp(\epsilon)\right\}, \qquad -\infty < \epsilon < \infty \tag{10}$$

which is the standard (type 1) extreme-value distribution. As lnT is a linear function of ϵ, it has the same (extreme-value) distribution as ϵ. Hence $T = exp(z\beta^* + \delta\epsilon)$ will have a Weibull distribution. If $q = 1$ and $\delta = 1$, then T has the exponential distribution as a special case of the Weibull distribution. The case of $q = -1$ corresponds to extreme maximum-value distribution for lnT. This, in turn, corresponds to reciprocal-Weibull distribution for T. The case of $\delta = 1$ and $q > 0$ is also of interest. Farewell and Prentice (1977) argue that this gives the ordinary gamma distribution for T, though, in accordance with Bergström and Edin (1992) and Bergström, Engvall, and Wallerstedt(1994; 1997), this does not hold in our case illustration. Consequently, we shall label this special case ($\delta = 1$, $q > 0$) the 'gamma' distribution in our illustrative example.

Thus, five models for T are included as special cases of the EGG model. Another model of interest, though not a special case of the EGG model, is the loglogistic model. A loglogistic distribution is the distribution of T such that $logT$ follows a logistic distribution. Description and applications of the loglogistic model may be found in Bacon (1993), Diekmann (1992), Little, Adams, and Anderson (1994), Nandram (1989), and Singh, Lee, and George (1988).

3.2.2 Estimation

The practical estimation of (7) proceeds as follows. Consider survival times of n individuals $t_1, t_2, ..., t_n$ and p covariates $z_1, z_2, ..., z_p$. Let d_i take value

0 if t_i is a censoring time and value 1 if t_i represents an event time. The log-likelihood function $ln(lnt; z\beta^*, \delta, q)$, assuming a noninformative censoring mechanism, will then be proportional to

$$\sum_{i=1}^{n} d_i \left[\ln f(q, \epsilon) - \ln \delta\right] + \sum_{i=1}^{n} (1 - d_i) \ln S(q, \epsilon_i) \tag{11}$$

where $f(\epsilon, q)$ is given by the EGG model (9), $S(\epsilon, q)$ is the corresponding survivor function, and $\epsilon_i = \frac{y_i - z_i\beta}{\delta}$..

At each of several $q-$values the maximum likelihood estimates $\left[\widehat{\beta}^*(q), \widehat{\delta}(q)\right]$ are obtained by using the Newton-Raphson method to solve the normal equations arising from (11). Standard errors of coefficients may be obtained from the information matrix as usual.

Allison (1995) contains a discussion on how most of these models may be estimated with the SAS software.

4 Illustration: The distributional shape of Marriage Durations

The aim of the present illustration is to fit the various models discussed above to a demographic data set in order to study the distributional shape of marriage durations and discriminate between special-case models. Related questions concern the dependence of inference on the regression coefficient on the corresponding values of the shape and scale parameters in (9); and of model choice on the level of time aggregation . For previous similar studies or general discussion on the model, see, Farewell and Prentice (1977), Addison and Portugal (1987, 1992), Corak (1993), Addison and Fox (1993), Bergström and Edin (1992), Bergström, Engvall, Wallerstedt (1994; 1997), Swaim and Podgursky (1992; 1994), Brannäs and Roonqvist (1994), Thomas (1996), and Tahai and Meyer (1999).

4.1 The data

The data set providing the basis for the following analysis comes from the 1985 mail survey of Swedish men which was conducted by Statistics Sweden. A simple random sample of men was drawn from each of the five-year cohorts born in 1936-40, 1941-45, 1946-50, 1951-55, 1956-60 as well as from the four-year cohort born in 1961-64.

From each man who responded the survey obtained retrospective information on the community in which he grew up, his current occupation, education, leisure time and financial situation at the time of the survey, his previous marital and cohabitational history, present family situation, and on attitudes and future plans on fatherhood and children. A total of 3171 males responded.

By the survey time (April 1985), 665 men were still single and had no children; 132 men had one or more child outside a union; 648 had initiated a family through formal marriage; while 1434 had initiated a family through cohabitation. Some of the cohabitations had been legitimatized into formal marriages before the survey time. Those who were never married before the date of the survey were excluded. Further, some of the married men were excluded from analysis because they either had incomplete information on some important covariates or had little information to contribute to the phenomenon being studied. This left us with 1203 usable records for our particular purposes. These were men with complete information, and were ever married before the time of the survey and were either divorced or still in their first marriage at the time of the survey.

The dependent variable is the rate at which the event of marital dissolution occurs or the duration of marriage depending on whether the model under consideration is a hazard-rate model or a duration model. In the latter case, the duration measures the number of months from entry into first marriage to the time of divorce or the survey date, whichever comes first. Apart from the time variable, the following six categorical explanatory variables were considered in the analyses:

z_1 - Birth cohort (1936-1945, 1946-1960).
z_2 - Disruption status of family of origin (Intact, Disrupted).
z_3 - Age at marriage ($<$ 25 yrs., $>$ 25 yrs.).
z_4 - Social class at survey time (Skill./Unsk., White C., Farm/Self Emp.).
z_5 - Educational level at survey time (Prim., Sec., Univ.).
z_6 - Mode of entry into marriage (Direct, After cohabiting).

These variables are among those considered to be correlated with the event of family formation, family dissolution, or both in previous analyses of the same data set. Summary statistics for the data is given in Table 1. Thus, the 1203 married men analyzed here consisted of 689 (57%) who married after cohabiting for some period (at least one month) and 514 (43%) men who married with no previous cohabitation. 699 (58%) were born between 1936 and 1945 while the rest 504 (42%) came from the younger cohort born between 1946 and 1960. Similarly, 1028 (85%) came from intact parental home while the rest 175 (15%) came from disrupted families.

4.2 Results

4.2.1 An overall view

By the time of the survey, 137 (11%) of the sample members had dissolved families while the rest 1066 (89%) were still in union and were, therefore, considered as censored. Of the 137 dissolved families 116 were preceded by cohabitation while the rest 21 come from direct marriages. In other words 17% of the marriages preceded by cohabitation were dissolved by the time of the survey while the corresponding figure for marriages not preceded by cohabitation was only 4%.

4.2.2 Covariate effects

In Table 2 we report results of fitting models (2) and (9) to our original data set (when duration variable was measured in months). The baseline categories (not represented by a coefficient) were cohort born in 1936-1945, from intact parental home, married below age 25, skilled or unskilled worker, with primary education, and married with no previous cohabitation. The estimated coefficients represent effects of the respective levels of a factor relative the corresponding baseline level (where covariates assume the value of zero). Estimates given in the last column of the table (based on (2)) measure the effect of the covariates on the hazard of marriage dissolution, while those on the other columns measure effect of the covariates on marriage duration. As we explained earlier, the two models follow different parameterizations. A positive coefficient in the last column implies an increased hazard (shorter duration) while according to (3) or (5) it implies longer duration (smaller hazard) relative to that of the baseline.

According to Table 2, the factors that considerably increase the hazard of dissolution (decrease the marriage duration) are coming from disrupted parental family, having a university level education, and having cohabitational experience before marriage. Being a white collar employee has an effect to the opposite direction, while the rest of the factors have only marginal effects. The above results are reported by most models though the reciprocal Weibull, lognormal, and loglogistic models differ in the direction of the effects of some (less significant) factors. In section 4.2.3, we shall see that some of these latter models are rejected in favor of a more comprehensive model.

Tables 3 and 4 report estimates of models (2) and (9) when the duration variable was aggregated in years and five-years, respectively. The overall picture with regard to parameter estimates is not much different from what we reported for the data based on months. Differences in scale and shape parameters shall be addressed later.

In the Cox model (2) the relative hazards are obtained by exponentiating the estimates reported in the table. Thus, if we had analyzed our original (monthly) data with the Cox model (2), then the hazard of divorce of a secondary-level educated man relative to that with only primary-level education is (see last column of Table 2) given by $exp(-0.061) = 0.941$. The corresponding figure for a university-level educated man is $exp(0.386) = 1.471$.

High relative hazards are shown for some categories. The characteristics that increase the hazard of marital dissolution by say more than 30% percent of the baseline category are coming from a disrupted parental home, having a university-level education and the experience of premarital cohabitation. The strongest predictor of the fate of a marriage is whether or not a man cohabited before marriage. A typical man from the former group is at a hazard of dissolution which is about 6 times a man married without cohabiting. Such a strong effect has been consistently captured by all models we have fitted.

4.2.3 Discrimination among parametric models

Likelihood-Ratio statistics corresponding to various tests for special cases of the EGG model (9) are presented on Table 5. These are used to test whether the corresponding special-case model is adequate relative to the more comprehensive EGG model. Results corresponding to the monthly data of the table show that the reciprocal Weibull and the lognormal models are rejected in favor of the more general EGG model. On the contrary, the Weibull and 'gamma' models are adequate enough compared to the EGG. This is in accordance with the estimated values of the shape and scale parameters under the EGG model The estimates of the shape and scale parameters, as reported in Table 2, are 1.141 and 0.946, respectively. These estimates are closer to the assertions of the Weibull (in which the shape parameter is fixed to 1 with a free scale parameter) and the 'gamma' (in which the scale parameter is fixed to 1 with a free shape parameter). The reciprocal Weibull and the lognormal models, on the other hand, impose a shape parameter of -1 and 0, respectively, with a free scale parameter.

When compared to the Weibull model, the exponential model is also adequate. This is also indicated by the estimated scale parameter (in the Weibull model) of 1.053, which is very close to 1, the value imposed by the exponential model. A point that is worth noting in Table 2 is that the two rejected models (lognormal and reciprocal Weibull) differ from the other members of the family with regard to the signs of estimated effects of the Cohort, Age, and Farm/Self variables and the significance of University level education variable. This is in contrast to previous results by Bergström and Edin (1992) where model rejection was not followed by any substantial change in the estimated parameters.

When the time variable was aggregated in years, the reciprocal Weibull model is, again, rejected in favor of the EGG model. The lognormal model is now not rejected, though the evidence is marginal. The marginality of the evidence is further strengthened by the fact that the model, together with the (strongly) rejected reciprocal Weibull model, yields parameter estimates for the Cohort, Age and Farm/Self variables that are in opposite direction compared to those of the other models which are relatively adequate relative to the EGG model. Furthermore, the estimated shape and scale parameters in Table 3, again, are not in support of the two rejected models.

The exponential model is, again, not rejected (with yearly data) when compared to the Weibull model as could also be inferred from the estimated scale parameter of 0.927 in the Weibull model. The association between the level of aggregation and the number and type of models rejected is in accordance with previous works by Bergström and Edin (1992), though our method and extent of time aggregation differs from theirs.

A different situation arises when the time variable is aggregated at a higher level. According to the last column of Table 5, the only model which is rejected in favor to the EGG model is the Weibull model. Referring to Table 4, the estimated shape parameter when time is aggregated in five-years is -0.174. This is closer to either -1 or 0 than to 1. The results from Tables 4 and 5 are, therefore, consistent. This rejected model again yields estimates for the Cohort, Age and Farm/Self variables, that are in opposite direction to those obtained from the relatively adequate models.

More interesting is the fact that the exponential model is now rejected in favor of the Weibull model. The estimated scale parameter of 0.725, which is relatively far from 1, is again in support of the Weibull. Though our program could not fit the 'gamma' model when time was aggregated in five years, the rejection of the exponential model in favor of the Weibull, together with the estimated scale parameter of 1.491 (in the EGG model) which is reasonably larger than 1, could provide evidence on the possible rejection of the 'gamma' model when compared to the EGG model.

5 Summary and concluding remarks

A natural question arises as to which procedure to use when one is confronted with a specific data-analysis problem. How does one choose among alternative models? As with most statistical methods, it is rather difficult to codify the procedures involved in choice of a model. There are many factors that should legitimately enter the decision and none can be easily quantified. Invariably there is tension among mathematical convenience, theoretical appropriateness, and empirical evidence. Given the wide range of models currently available to the user, it is worth asking whether conclusions are sensitive to the particular statistical model chosen. The answer to this question is unknown until results obtained with one method have been compared to results obtained by another method. Such comparisons have been one of the objectives of the present paper. We, shall therefore, summarize our results with reference to our objectives.

With models of the type we have just been discussing, the key differentiating factor is the way in which the hazard rate depends on time. The first choice would, then be between the Cox regression model in which such dependence is left to be arbitrary, and the other models which postulate some form of dependence.

The exponential model which postulates a constant hazard rate over time, is very attractive from a mathematical and computational point of view. Substantive theory, on the other hand, will usually suggest many reasons why the hazard should change with time. A piece-wise exponential model which assumes a piecewise constant hazard rate, is a flexible alternative in this situation. For this reason, it may be worth using the model as a first approximation. When this fails, one has to resort to the other parametric family of distributions.

Our empirical findings on the parametric models both build on and contradict with previous works in the area. Though our method and extent of aggregation differs from those of previous works, our findings indicate that the choice of a particular model for the duration variable is dependent on the level of aggregation used. More importantly, parameter estimates are not always robust to distributional assumptions, and when the purpose of the analysis is to make inferences about covariate effects, failure to consider the proper distribution may plague the purpose of the analysis.

The intent of our empirical analyses was just to demonstrate the potential application of various techniques for censored life-time data to retrospectively collected demographic survey data, and compare results obtained by different models. As a result no attempt has been made to search for the most parsimonious model representing family dissolution. Neither have we made any effort to address substantive sociological or demographic issues behind our empirical findings. The analyses was also based on combining various levels of many of the covariates with a view to minimize the number of binary variables which are bound to result from larger number of levels. A more thorough analyses could have benefited from including interaction effects between selected factors, for instance, between the Cohort and Mode of Entry factors. As we saw above, certain characteristics of men are associated with high hazards than others. Consequently, groups who are in high-hazard categories on several dimensions can have extremely high hazard. In addition, inclusion of time-varying covariates such as the number of children within a family (parity), which was intentionally ignored in this analysis, could give additional insight into the determinants of marital dissolution.

References

[1] Addison, J. T., and P. Portugal (1987), On the distributional shape of unemployment duration. *Review of Economics and Statistics* **69**, 520-526.

[2] Addison, J. T., and P. Portugal (1992), The distributional shape of unemployment duration - A Reply.. *Review of Economics and Statistics* **74**, 717-721.

[3] Addison, J. T., and Fox, D. A. (1993), Job changing after displacement - A contribution to the advance notice debate. *Southern Economic Journal* **60**, 184-200.

[4] Allison, P. D. (1995), *Survival Analysis using the SAS® System: A Practical Guide.* Cary, NC: SAS Institute Inc.

[5] Bacon, R. W. (1993), A note on the use of the log-logistic functional form for modelling saturation effects. *Oxford Bulletin of Economics and Statistics* **55**, 355-361.

[6] Bergström, R., and P. A. Edin (1992), Time aggregation and the distributional shape of unemployement duration. *Journal of Applied Econometrics* **7**, 5-30.

[7] Bergström, R., Engwall, L., and E. Wallerstedt (1994), Organizational foundations and closures in a regulated environment: Swedish commercial banks 1831-1990. *Scandinavian Journal of Management* **10**, 29-48.

[8] Bergstrom R, Engvall L, Wallerstedt E (1997), The importance of flexible hazard functions in the analysis of organizational survival data - Experiences from a cohort of Swedish commercial banks. *Quality and Quantity* **31**, 15-35.

[9] Brannäs, K., and Rosenqvist, G. (1994), Seminparametric estimation of heterogeneous count data models. *European Journal of Operational Research* **76**, 247-258.

[10] Corak, M. (1993), Cyclical variations in the duration of unemployment spells. *Industrial Relations- Relations Industrielles* **48**, 125-145.

[11] Cox, D. R. (1972), Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society* **B34**, 187-220.

[12] Cox, D. R. (1975), Partial Likelihood. *Biometrika* **62**, 269-276

[13] Diekmann, A. (1992), The log-logistic distribution as a model for social diffusion processes. *Journal of Scientific and Industrial Research* **51**, 285.

[14] Farewell, V. T., and R. L. Prentice (1977), A study of distributional shape in life testing. *Technometrics* **19**, 69-75.

[15] Little, C. L., Adams, M. R., and W. A. Anderson (1994), Application of a log-logistic model to describe the survival of Yersinia enterocolitica at sub-optimal pH and temporature. *International Journal of Food Microbiology* **22**, 63.

[16] Nandram, B. (1989), Discrimination between the complimentary log-log and logistic model for ordinal data. *Communications in Statistics - Theory and Applications* **18**, 2155.

[17] Prentice, R. L. (1974), A log-gamma model and its maximum likelihood estimation. *Biometrika* **61**, 539-544.

[18] Singh, K. P., Lee, C. M. -S., and E. O. George (1988), On generalized loglogistic model for censored survival data. *Biometrical Journal* **7**, 843.

[19] Stacey, E. W. (1962), A generalization of the gamma distribution. *Annals of Mathematical Statistics* **33**, 1187-1192.

[20] Swaim, P., and Podugursky, M. (1992), The distributional shape of unemployment duration - A reconsideration. *Review of Economics and Statistics.* **74**, 712-717.

[21] Swaim, P., and Podgursky, M. (1994), Female labor supply following displacement - a split-population model of labor-force participation and job search. *Journal of Labor Economics* **12**, 640-656.

[22] Tahai and Meyer (1999), A revealed preference study of management journals' direct influences. *Strategic Management Journal* **20**, 279-296.

[23] Thomas, J. M. (1996), An empirical model of sectoral movements by unemployed workers. *Journal of Labor Economics* **14**, 126-153.

A Tables of Empirical Results

Table 1: Summary statistics across covariats

Covariate	Levels	# of men	Cases (Dissol.)	% cases	Exposure (months)	Case/Exp Rates[1]	Relative Hazards
Cohort	1936-45	699	88	12.59	148348	5.93	1.00
	1946-60	504	49	9.72	53981	9.08	1.53
Family of Origin	Intact	1028	111	10.80	173088	6.41	1.00
	Disrupted	175	26	14.86	29241	8.89	1.39
Age at marriage	< 25	674	79	11.72	128669	6.14	1.00
	25+	529	58	10.96	73660	7.87	1.28
Social Class	Ski/Uns.	423	49	11.58	69221	7.08	1.00
	White C	618	70	11.33	103100	6.79	0.96
	Farm/Self	162	18	11.11	30008	6.00	0.85
Educ.	Prim.	451	52	11.53	84760	6.13	1.00
	Second.	484	47	9.71	76840	6.12	1.00
	Univ	268	38	14.18	40729	9.33	1.52
Mode of entry	Direct	514	21	4.09	108250	1.94	1.00
	After C.	689	116	16.84	94079	12.33	6.36
Total		1203	137	11.39	202329	6.77	-

[1]The Case/Exp. ratios are initial estimates of the baseline intensities when no account is made of the other factors. They are expressed per 10000 exposure months. A rough estimate of the overall dissolution rate (before accounting for differences across covariates) is thus, $137/202329 = 6.77$ dissolutions per every 10000 marriages per month, or equivalently about 81 dissolutions per every 10000 marriages per year.

Table 2: Estimated coefficients and standard errors (below each coefficient) for various parametric models: monthly data[2]

Covariate	EGG	r. W.	Logn	Weibull	'Gamma'	Expon	Logl.	Cox
Intercept	6.308	6.680	6.731	6.342	6.328	6.250	6.164	-
	0.475	0.631	0.558	0.468	0.457	0.422	0.481	-
Scale	0.946	3.340	2.181	1.053	1	1	0.998	-
Parameter	0.396	0.218	0.156	0.084	0	0	0.079	-
Shape	1.141	-1	0	1	1.070	1	-	-
Parameter	0.555	0	0	0	0.104	0	-	-
Cohort	-0.034	0.283	0.135	-0.023	-0.028	-0.052	0.004	-0.090
(1946-60)	0.199	0.267	0.235	0.199	0.196	0.182	0.206	0.190
Family	-0.298	-0.332	-0.372	-0.308	-0.303	-0.293	-0.337	0.297
(Disrupted)	0.230	0.328	0.284	0.232	0.229	0.219	0.245	0.219
Age at Mar.	-0.095	0.340	0.118	-0.082	-0.089	-0.094	-0.045	0.016
(Older)	0.188	0.256	0.223	0.186	0.184	0.176	0.194	0.179
White C.	0.186	0.226	0.225	0.193	0.190	0.186	0.211	-0.187
	0.240	0.302	0.273	0.241	0.240	0.229	0.246	0.229
Farm/Self.	0.051	-0.213	0.034	0.060	0.055	0.059	0.092	-0.057
	0.293	0.386	0.348	0.296	0.295	0.282	0.309	0.282
Second.	0.056	0.032	0.055	0.056	0.057	0.050	0.056	-0.061
	0.224	0.286	0.328	0.227	0.226	0.215	0.233	0.215
Univ.	-0.394	-0.626	-0.555	-0.409	-0.402	-0.392	-0.454	0.386
	0.282	0.370	0.328	0.281	0.279	0.266	0.292	0.266
Cohabited	-1.900	-1.693	-1.823	-1.896	-1.902	-1.810	-1.892	1.760
	0.292	0.296	0.278	0.289	0.291	0.241	0.283	0.242

[2] The abbreviations used for the models (in Tables 2-4) mean as follows: EGG : Extended Generalized Gamma; r.W.: reciprocal Weibull; Logn : Lognormal; ; Expon : Exponential; logl : loglogistic.

Table 3: Estimated coefficients and standard errors (below each coefficient) for various parametric models: Yearly data

Covariate	EGG	r. W.	Logn	Weibull	'Gamma'	Expon	Logl.	Cox
Intercept	3.723	3.956	3.989	3.698	3.715	3.826	3.533	-
	0.394	0.510	0.474	0.412	0.407	0.422	0.422	-
Scale	0.291	2.706	1.851	0.927	1	1	0.877	-
Parameter	0.024	0.181	0.132	0.074	0	0	0.069	-
Shape	3.377	-1	0	1	0.894	1	-	-
Parameter	0.446	0	0	0	0.100	0	-	-
Cohort	-0.099	0.163	0.063	-0.068	-0.060	-0.028	-0.041	-0.123
(1946-60)	0.168	0.217	0.200	0.175	0.174	0.184	0.181	0.190
Family	-0.242	-0.318	-0.336	-0.276	-0.282	-0.296	-0.303	0.294
(Disrupted)	0.189	0.265	0.240	0.204	0.206	0.219	0.216	0.219
Age at Mar.	-0.135	0.231	0.066	-0.101	-0.091	-0.084	-0.067	-0.001
(Older)	0.155	0.208	0.190	0.164	0.164	0.176	0.171	0.178
White C.	0.151	0.198	0.196	0.169	0.173	0.179	0.185	-0.181
	0.205	0.244	0.232	0.212	0.213	0.228	0.217	0.228
Farm/Self.	0.340	-0.002	0.072	0.057	0.063	0.057	0.083	-0.050
	0.246	0.315	0.297	0.261	0.263	0.282	0.272	0.282
Second.	0.042	0.069	0.049	0.046	0.045	0.055	0.046	-0.063
	0.193	0.232	0.220	0.200	0.201	0.215	0.204	0.215
Univ.	-0.319	-0.511	-0.475	-0.360	-0.371	-0.385	-0.399	0.381
	0.234	0.299	0.278	0.247	0.248	0.266	0.257	0.266
Cohabited	-1.683	-1.462	-1.596	-1.684	-1.672	-1.804	-1.681	1.751
	0.262	0.240	0.237	0.254	0.251	0.241	0.248	0.242

Table 4: Estimated coefficients and standard errors (below each coefficient) for various parametric models: five-year data[3]

Covariate	EGG	r. W.	Logn	Weibull	Expon	Logl.	Cox
Intercept	2.104	1.965	2.105	1.976	2.469	1.832	-
	0.356	0.361	0.353	0.320	0.421	0.328	-
Scale	1.491	1.908	1.379	0.725	1	0.682	-
Parameter	0.373	0.151	0.099	0.057	0	0.053	-
Shape	-0.174	-1	0	1	1	-	-
Parameter	0.574	0	0	0	0	-	-
Cohort	0.030	0.084	0.015	-0.079	0.079	-0.057	-0.285
(1946-60)	0.157	0.152	0.149	0.137	0.184	0.141	0.188
Family	-0.284	-0.293	-0.278	-0.220	-0.296	-0.245	0.288
(Disrupted)	0.181	0.186	0.178	0.159	0.219	0.168	0.219
Age at Mar.	0.345	0.124	0.012	-0.110	-0.041	-0.083	-0.091
(Older)	0.160	0.147	0.141	0.128	0.176	0.133	0.177
White C.	0.130	0.115	0.132	0.122	0.155	0.133	-0.144
	0.173	0.171	0.173	0.166	0.229	0.170	0.228
Farm/Self.	0.070	0.065	0.070	0.044	0.045	0.061	-0.029
	0.223	0.224	0.221	0.204	0.282	0.212	0.281
Second.	0.033	0.030	0.034	0.039	0.073	0.041	-0.085
	0.165	0.163	0.164	0.157	0.216	0.160	0.215
Univ.	-0.381	-0.416	-0.370	-0.279	-0.367	-0.308	0.354
	0.211	0.210	0.207	0.194	0.266	0.201	0.266
Cohabited	-1.216	-1.117	-1.236	-1.329	-1.781	-1.327	1.713
	0.373	0.169	0.177	0.199	0.242	0.193	0.242

[3] When the data was aggregated in five-years the LIFEREG procedure in SAS could not fit the 'gamma' model. The error message 'Domain Error in log arg = 0.00E+00' was reported by SAS. It is worth noting that the factor "Cohabited" has been reported as significant by all models at all levels of aggregation. The EDUC2 factor has been reported as (marginally) significant by the lognormal and reciprocal Weibull models, and for data grouped in five years, also by the EGG model.

Table 5: Hypotheses and corresponding likelihood ratio statistics for testing special cases H_0 against more general models H_1, within the parametric family of models

Hypothesis (H_0)	Model under H_0	Model under H_1	Likelihood Ratio Monthly	Likelihood Ratio Yearly[4]	Likelihood Ratio Five-year[5]
$q = -1$	Recip. Weibull	EGG	31.42	14.22	1.44
$q = 0$	Lognormal	EGG	11.48	2.74	0.10
$q = 1$	Weibull	EGG	0.08	0.44	5.74
$\delta = 1$ given $q > 0$	'Gamma'	EGG	0.02	0.09	-
$\delta = 0$ given $q = 1$	Exponential	Weibull	0.42	0.88	14.84

[4] The test statistics corresponding to $H_0 : q = 1$, and $H_0 : \delta = 0$ are Lagrange Multipliers. Our program could not produce the corresponding likelihood ratio statistics. This does not affect our inference, however, since all tests lead to the same conclusions.

[5] As explained in footnote 3, our program could not fit the 'Gamma' model when the data was aggregated in five-years. We, therefore lack the corresponding test statistic in the table.