

**SPECIFICATION TESTS BASED ON THE HETEROGENEOUS GENERALIZED
GAMMA MODEL OF DURATION: WITH AN APPLICATION TO KENNAN'S
STRIKE DATA**

SANJIV JAGGIA

SUMMARY

In this paper, tests for neglected heterogeneity and functional form misspecification of some commonly used parametric distributions are derived within a heterogeneous generalized gamma model. It is argued that the conventional test of heterogeneity may not be valid when the underlying hazard function is misspecified. Hence, if the estimated hazard function is deemed restrictive, tests for functional form misspecification should accompany any test of heterogeneity. An empirical illustration based on Kennan's (1985) model of strikes is used to show that incorrect inferences may be drawn, as in a number of previous analyses, if the relevant restrictions are not tested jointly.

1. INTRODUCTION

The validity of most separate tests in the econometric literature relies on auxiliary assumptions made in addition to the assumptions being tested. In the presence of more than one source of misspecification, inferences drawn from the outcome of any separate test are distorted. This is especially true when separate tests are correlated. As a consequence, the size and power of such tests are incorrect when the additional auxiliary assumptions are not satisfied by the given data. One solution to such a problem is to compute an omnibus statistic, that tests all the relevant assumptions made within a given model jointly and hence has power against several forms of misspecification. In this paper such joint tests are derived, and it is shown that the application

of joint tests of misspecification, rather than partial or separate tests, is necessary in conducting specification evaluations of parametric duration models.

In the context of parametric duration models it is desirable to have diagnostic tests of the validity of the distributional assumptions since, in the presence of misspecification, estimation by maximum-likelihood methods may lead to inconsistent estimates. Two important sources of misspecification are the functional form of the hazard function and neglected heterogeneity. To date, tests of neglected heterogeneity have been emphasized (see Lancaster, 1983, 1985; Keifer, 1984; Burdett *et al.*, 1985; Lancaster and Chesher, 1985a). The separate heterogeneity test is derived on the assumption that the functional form of the model is correctly specified and, therefore, will not have the right size or power when the maintained assumption is not true. Moreover, Manton, Stallard, and Vaupel (1986) have shown that model estimates may be sensitive to the choice of hazard function. Hence, if the choice of the estimated hazard function is deemed restrictive, then, at the very least, tests for functional form misspecification should accompany any test for heterogeneity.

The problems with the standard separate test can be illustrated by the following simple example. Consider an exponential model, with the density function given by: $f(t) = \mu \exp(-\mu t)$ where $\mu = \exp(X\beta)$. The log-linear form, using $y = \log(t)$, can be written as: $y = -X\beta + W = X\theta + W$, where W has an extreme value distribution with variance = 1.6449. If there is some unobserved heterogeneity in the model, represented by V , then $y = X\theta + V + W$. A test for heterogeneity can be constructed that tests for over-dispersion in the data. The test would detect heterogeneity when the sample variance of the error term is significantly greater than 1.6449.

Now, consider a situation where the underlying model is not exponential but Weibull, with the density function given by: $f(t) = \mu \alpha t^{\alpha-1} \exp(-\mu t^\alpha)$. Here, the log-linear form is given by $y =$

$X\theta/\alpha + W/\alpha$, and $y = X\theta/\alpha + W/\alpha + V$ if there is unobserved heterogeneity. If the same test of heterogeneity for an exponential model is computed, and there is positive duration dependence in the sample ($\alpha > 1$), the effect of *the* variance of V will cancel out with the reduced variance of W/α . The test will pick up no misspecification even though both duration dependence and neglected heterogeneity exist in the sample. The actual positive duration dependence will cancel out with the spurious negative duration dependence induced by neglected heterogeneity. This heuristic argument points out the limitations of the separate tests when multiple misspecifications exist concurrently. The analysis can be extended to situations where the underlying model is more general than the Weibull model.

The rest of the paper is organized as follows: in section 2 a heterogeneous generalized gamma model is considered and used to derive joint score tests for functional form misspecification and neglected heterogeneity for Weibull and exponential models. In section 3 these tests are applied to Kennan's strike data. It is inferred that when a separate heterogeneity test is implemented, Weibull as well as exponential models seem appropriate. However, when this test is applied jointly with the test of functional form, both exponential and Weibull models are found to be inadequate. With the rejection of the joint null hypothesis the possible misspecification could arise either from the functional form of the hazard function or from neglected heterogeneity. To detect if there is neglected heterogeneity in the sample, a flexible, generalized gamma model is estimated and used to test for heterogeneity. The results indicate neglected heterogeneity in the sample. Informal plotting procedures are also applied both for an exploratory analysis of the data and for testing for parametric models. Concluding remarks are contained in section 4.

2. SPECIFICATION TESTS

2.1. Generalized Gamma Distribution

In order to avoid distortions arising from a restricted choice of a parametric duration distribution, it is proposed that tests be based on a three-parameter generalized gamma distribution (g.g.d.). The density function of a g.g.d. is:

$$f(t | X) = \mu^k \alpha t^{\alpha k - 1} \exp(-\mu t^\alpha) / \Gamma(k) \quad (1)$$

μ is taken as $\exp(X\beta) = \exp(\beta_0 + X_1\beta_1)$ to ensure its non-negativity; X is a vector of explanatory variables.

This distribution encompasses all of the most frequently used parametric distributions, such as exponential ($\alpha = k = 1$), Weibull ($k = 1$), gamma ($\alpha = 1$) and log-normal ($k \rightarrow \infty$), and it also accommodates other non-monotone hazards (see Lawless, 1982). This property makes g.g.d. a useful distribution to discriminate between such alternate models, in addition to its use as a flexible duration distribution itself.

Pereira (1978), using Cox's (1961, 1962) approach for testing non-tested hypotheses, develops tests to discriminate between log-normal, gamma, Weibull and exponential models. This approach, however, becomes intractable when heterogeneity is allowed for. Alternatively, the encompassing approach is suggested as a means of discriminating between the above-mentioned parametric models. This point is elucidated in the context of a heterogeneous generalized gamma distribution which can be specialized to the above models with or without heterogeneity. Even though it is difficult to estimate parameters of this distribution, score tests can be easily implemented since only the null model needs to be estimated for such tests.

Given some unobserved multiplicative heterogeneity, represented by V , the distribution (1) conditional on V can be written as:

$$f(t | X, V) = V^k \mu^k \alpha t^{\alpha k - 1} \exp(-V \mu t^\alpha) / \Gamma(k) \quad (2)$$

As V is not observable, the unconditional distribution can be derived by integrating (2) with respect to the distribution of V . If V has a finite mean,¹ $E(V)$ can be set to equal 1 without loss of generality, given that X includes a constant term. Further, for a small variance of the heterogeneity term denoted by σ^2 , the density function can be approximated by a second-order Taylor series expansion around the unit mean of V as follows:

$$f(t | X, V) \approx f(t | X, V = 1) + (V - 1) \left[\frac{\partial}{\partial V} f(t | X, V) \right]_{V=1} + \frac{1}{2} (V - 1)^2 \left[\frac{\partial^2}{\partial V^2} f(t | X, V) \right]_{V=1} \quad (3)$$

and:

$$f(t | X) = E_v[f(t | X, V)] \approx f(t | X, V = 1) [1 + (\sigma^2/2) \{k(k-1) - 2k\mu t^\alpha + (\mu t^\alpha)^2\}] \quad (4)$$

2.2. Joint and Partial Score Tests for a Weibull Model

The hypothesis that the given distribution is Weibull can be tested using a score test where the null is specified as:

$$H_0: \sigma^2 = 0 \text{ and } k = 1 \quad (5)$$

For the approximate density in (4), the log-likelihood function is:

$$L = \Sigma [k \log(\mu) + \log(\alpha) + (\alpha k - 1) \log(t) - \mu t^\alpha - \log(\Gamma(k)) + \log[1 + (\sigma^2/2) \{k(k-1) - 2k\mu t^\alpha + (\mu t^\alpha)^2\}]] \quad (6)$$

Let $\theta = (\theta_1, \theta_2)'$ where:

$$\theta_1 = (\sigma^2 k) \text{ and } \theta_2 = (\beta_0 \beta_1 \alpha) \quad (7)$$

Let $s(\theta)$ and $I(\theta)$ denote the score vector and the information matrix respectively. The elements

¹ Some distributional assumptions on V are needed for the identification of duration models. For instance, if the conditional distribution is Weibull, and V belongs to a particular family of stable distributions with no finite mean, the unconditional distribution is also Weibull (Hougaard, 1986). See also Elbers and Ridder (1982) and Heckman and Singer (1984) for the identification of proportional hazard models.

of the relevant score vector, $s_1(\theta_0)$, evaluated under the null hypothesis are:

$$\left. \frac{\partial L}{\partial \sigma^2} \right|_{H_0} = \frac{1}{2} \Sigma [\varepsilon^2 - 2\varepsilon] = s_{11}(\theta_0) \quad (8)$$

$$\left. \frac{\partial L}{\partial k} \right|_{H_0} = \Sigma [\log(\varepsilon) - \phi(1)] = s_{12}(\theta_0) \quad (9)$$

$$\varepsilon = \mu t^\alpha \text{ and } \phi(r) = \frac{d \log \Gamma(r)}{dr}$$

is the digamma function and

$$\phi'(r) = \frac{d^2 \log \Gamma(r)}{dr^2}$$

is the trigamma function.

The joint score test of the null hypothesis is:

$$LM_{hf} = s_1'(\tilde{\theta}_0) I^{11}(\tilde{\theta}_0) s_1(\tilde{\theta}_0) \quad (10)$$

where $(\tilde{\cdot})$ denotes that the quantities have been evaluated at the restricted maximum-likelihood estimate of the parameter vector, θ , and $I^{11} = [J_{11} - J_{12}(J_{22})^{-1}J_{21}]^{-1}$ is the partitioned inverse of $I(\theta)$. The given test statistic has a chi-square distribution under H_0 with two degrees of freedom. The partitioned inverse based on (5) can be derived as:

$$[I^{11}(\theta_0)]^{-1} = N \begin{bmatrix} 1-p & p-1/2 \\ p-1/2 & \phi'(2)-p \end{bmatrix} \quad (11)$$

where $p = 1/\phi'(1)$ and N is the sample size.

The partial test for heterogeneity, LM_h , under the assumption that the hazard function is correctly specified is:

$$LM_h = s_{11}' [N(1-p)]^{-1} s_{11} \quad (12)$$

which is similar to the test proposed by Lancaster (1985). Similarly, a partial test for functional form misspecification, given no neglected heterogeneity, is:

$$LM_f = s_{12}' [N(\phi'(2)-p)]^{-1} s_{12}. \quad (13)$$

From (11), it can be seen that the relevant portion of the information matrix is not block diagonal and thus the partial tests are not independent. Due to the non-zero correlation between the two tests, the nominal size and power of any partial test will be affected by the presence of the other source of misspecification that is ignored. Therefore, results of partial tests can be misleading when both sources of misspecification exist.

2.3. Joint and Partial Score Tests for an Exponential Model

Analogous to the above procedure, the exponential specification in the context of a heterogeneous generalized gamma model can be tested using:

$$H_0: \sigma^2 = 0, \quad \alpha = 1, \quad k = 1 \quad (14)$$

Implementing the notation given above, with $\theta_i' = (\sigma^2 \alpha k)$, the elements of the relevant score

vector, $s_i(\theta_0)$, are:

$$\left. \frac{\partial L}{\partial \sigma^2} \right|_{H_0} = \frac{1}{2} \Sigma [\varepsilon^2 - 2\varepsilon] = s_{11}(\theta_0) \quad (15)$$

$$\left. \frac{\partial L}{\partial \alpha} \right|_{H_0} = \Sigma [1 + (1 - \varepsilon) \log(t)] = s_{12}(\theta_0) \quad (16)$$

$$\left. \frac{\partial L}{\partial k} \right|_{H_0} = \Sigma [\log(\varepsilon) - \phi(1)] = s_{13}(\theta_0) \quad (17)$$

where $\varepsilon = \mu t$. The test will be based on the following expression for the relevant partition of the information matrix evaluated under (14):

$$[I^{11}(\theta_0)]^{-1} = N \begin{bmatrix} 1 & -1 & -1/2 \\ -1 & \phi'(1) & 1 \\ -1/2 & 1 & \phi'(2) \end{bmatrix} \quad (18)$$

The partial tests of $\sigma^2 = 0, \alpha = 1$ and $k = 1$ represented by $LM_{\sigma^2}, LM_{\alpha}, LM_k$ respectively, can similarly be derived as:

$$LM_{\sigma^2} = s'_{11} [N]^{-1} s_{11} \quad (19)$$

$$LM_{\alpha} = s'_{12} [N\phi'(1)]^{-1} s_{12} \quad (20)$$

$$LM_k = s'_{13} [N\phi'(2)]^{-1} s_{13} \quad (21)$$

Similarly, using the appropriate elements from (18), tests of two restrictions can be derived.

For example, a test of functional form misspecification for an exponential model would imply testing for $\alpha = 1$ and $k = 1$ jointly. Using (18), such a test is easily implementable.

One further comment needs to be made regarding the computation of these tests in the presence of censored observations. With censored observations the theoretical information matrix needed to implement the score tests cannot be derived without additional information regarding the censoring mechanism. However, the tests can be based on the observed information matrix_ Efron and Hinkley (1978), more generally, recommend the use of the observed information matrix as it is closer to the data than the corresponding expected (theoretical) information matrix. Two possible candidates for this matrix are the sample

hessian of the log-likelihood function and the outer product of the sample scores. Some deterioration in the performance of such tests is expected partly because of the loss of information due to censoring and partially due to the use of the sample information matrix to implement them.²

3. ANALYSIS OF STRIKE DATA

3.1. Background

In order to illustrate specification tests, data on duration of the contract strikes in US manufacturing industries, as reported by Kennan (1985), are analysed. Kennan studies the effect of business cycles on strike durations for the period 1968 through 1976. A proxy for cyclical effects, X , is formed by taking the residual from the regression of the logarithm of industrial production in manufacturing (INDP) on time, time squared, and monthly dummy variables. The data consist of 566 observations on the duration of completed strikes measured in days and the corresponding value of INDP.

3.2. Graphical Analysis

Graphical procedures are often employed in duration models both for an exploratory analysis and for testing for the parametric specification of a given model (see Lancaster and Chesher, 1985b; Kiefer, 1988; Lawless, 1982, etc., for details). Empirical plots, using observations grouped by the levels of the covariates to achieve homogeneity, can be used to suggest the shape of the underlying hazard function. An attempt is made here to create two such homogeneous samples for X below and above the mean. The empirical integrated hazard function, $H(t)$, for the two samples is derived. The estimated $H(t) = \text{minus log of the sample}$

² Jaggia (1990) carries out a Monte-Carlo analysis of tests for heterogeneity in the exponential and Weibull models with various types of censoring. The results of this study are mixed. Such tests perform well under certain types of censoring.

survivor function, $S(t)$, where $S(t) = N^{-1}$ (number of sample observations $\geq t$). The plot of the integrated hazard function can suggest the shape of the underlying hazard function. If the integrated hazard is linear it represents a constant hazard, implying an exponential model. A convex integrated hazard implies an increasing hazard and a concave integrated hazard implies a decreasing hazard. From Figures 1 and 2 the underlying hazard function seems neither constant nor monotonic. This observation may have resulted from the fact that the above procedure to suggest the appropriate, underlying, functional form has not worked due to neglected heterogeneity in the sample. Grouping observations by the levels of the observed covariates may not have resulted in homogeneity in the distinct subsamples.

[Insert Figures 1-5]

Graphical plots are also used to ascertain if a particular parametric model is adequate. For a correctly specified parametric duration model the generalized residuals, defined as the integrated hazard function, should behave approximately like a random sample taken from a unit exponential distribution. A product-limit estimate of the integrated hazard function of the generalized residuals is obtained. If the model is correctly specified, the scatterplot of this estimate against the generalized residuals should cluster around a 45° line through the origin.

Such scatterplots for the exponential, Weibull and generalized gamma models are plotted in Figures 3, 4 and 5. It is observed that the departure from the 45° line is almost identical in all plots. One problem with informal graphical procedures is that some degree of subjectivity is involved in interpreting the results. If one were to infer, from Figure 3, that the exponential model is inadequate, the less restrictive models graphed in Figures 4 and 5 offer no improvement. As a preliminary look at Figures 1 and 2 suggests that the hazard function is neither constant nor monotone, one expects the plots to show a substantial improvement when

the g.g.d. is used to model the hazard function. This is obviously not the case_ The problem may be that not only is the underlying hazard function non-monotone, but there is also some neglected heterogeneity in the sample, causing misleading results.

3.3. Parametric Specification Analysis

All the previous conjectures made using informal graphical plots are here formalized with parametric tests. The specification tests, described in section 2, implemented on the exponential model and the Weibull model are reported in Table 1 and Table 11, respectively. To reiterate, the results presented in the tables are based on the score test principle where the parameters of the alternative hypothesis are not estimated. For instance, in order to implement tests for an exponential specification, the parameters σ^2 , α and k are not estimated.

[Insert Table I and II]

From Table I it is seen that none of the partial tests detects any misspecification in the exponential model. This implies either that the model is correctly specified, or that the joint presence of more than one source of misspecification has some kind of cancellation effect on the partial tests. Pagan and Vella (1988) and Kiefer (1988) have reported specification tests that support a simple exponential model for the same strike data. This is contrary to Kennan's arguments for a non-monotone hazard function. The misleading indication of a good fit of the model may be explained by the cancellation of the effects of true duration dependence with the spurious duration dependence induced by neglected heterogeneity (see also Jaggia and Trivedi, 1990). However, the joint null hypothesis of $\sigma^2 = 0$ and $\alpha = 1$ is also not rejected. In fact, none of the two restriction tests suggests misspecification. This apparent inconsistency may be due to the fact that the hazard function is non-monotone and is not accurately captured by a monotonic Weibull hazard function or gamma hazard function. However, the joint test of three restrictions,

based on a heterogeneous generalized gamma distribution, does indicate that the exponential model is inadequate.

The results from the estimated Weibull model, shown in Table II, also support similar conclusions. The joint test of the two restrictions, $\sigma^2 = 0$ and $k = 1$, is not supported by the given data, even though the two partial tests fail to reject the null hypothesis.

3.4. Detecting Heterogeneity

A known limitation of joint testing is that a significant joint test does not indicate the *nature* of the required respecification of the model. The above analysis suggests that one should allow for heterogeneity in estimation to obtain a valid test for functional form misspecification and vice-versa. This method will provide useful information about the desirable direction to take in respecifying the model.

Given that even the Weibull hazard specification is not appropriate for the hazard function, one can estimate a generalized gamma model and test for neglected heterogeneity. Generalized gamma models are known to have convergence problems, especially when the parameter k is large. The model has to be reparametrized to obtain the maximum-likelihood estimates (see Lawless, 1982). However, no convergence problems were encountered, with the given strike data, even with the original parametrization. A possible explanation for this result is the large sample size and the fact that the estimated value of k is small.

A score test of $\sigma^2 = 0$ may be based on (6), which does not depend upon any parametric representation of the heterogeneity distribution. Using (6), the appropriate score is:

$$\frac{\partial L}{\partial \sigma^2} \Big|_{H_0} = \Sigma \frac{1}{2} [k(k-1) - 2k\mu t^\alpha + (\mu t^\alpha)^2]. \quad (22)$$

The score test based on (22), with the information matrix computed as the outer product of the sample scores, is reported in Table III. It is seen that there is neglected heterogeneity in the

sample. Interestingly, a partial test of heterogeneity detects misspecification only when a fairly general duration distribution is used in estimation.

[Insert Table III and IV]

The results of the parameter estimates under alternate model specifications are presented in Table IV. It is noted that, whenever a more general hazard function is estimated, the additional parameter is found to be insignificant. For example, when a Wald-type test is applied to a generalized gamma model it is found that both α and k are not significantly different from 1. This result implies that an exponential specification is appropriate. Furthermore, the likelihood ratio (LR) test, computed by taking twice the difference between the maximized log-Likelihood values of the null and the alternative models, suggest that the generalized gamma model is not an improvement over the Weibull or the exponential model. However, unlike the cases of the exponential and Weibull models, when a generalized gamma model is estimated, the separate score test of heterogeneity indicates the presence of neglected heterogeneity. The misleading results obtained from using Wald or LR tests can once again all be attributed to the fact that one possible misspecification, in the form of neglected heterogeneity, is being ignored when testing for the functional form specification of the model. This numerical analysis corroborates the earlier findings that graphs show no improvement over the exponential or Weibull models when g.g.d. is used to model the hazard function.

When a generalized gamma model is estimated, the estimates of α and k are found to be 0.71 and 1.71, respectively, implying an inverted 'U'-shaped hazard function (see Glaser, 1980). This result is in contrast to Kennan's finding of a 'U'-shaped hazard. Quite possibly the above estimates of the shape parameters are misleading due to neglected heterogeneity in the sample.

The estimate of the regressor coefficient in all models, however, implies that strike durations are countercyclical, as in Kerman.

4. CONCLUSIONS

In this paper it has been shown that the partial test of heterogeneity (functional form) is quite misleading in the presence of functional misspecification (neglected heterogeneity). Score tests for the functional form misspecification of the hazard function, along with neglected heterogeneity, are developed for the Weibull and exponential models. Partial tests are shown to be asymptotically correlated within a heterogeneous generalized gamma model, and thus the nominal size and power of any partial test is affected by the presence of the other misspecification. An empirical illustration based on Kennan's strike data shows evidence of incorrect inferences that are drawn from using partial tests. Impressions obtained from the informal graphs used both for analysing the data and testing for the parametric models are confirmed by numerical results based on partial and joint tests. It is therefore stressed that the first step in model evaluation should always be to implement a joint test as more than one source of misspecification may exist in any given model.

ACKNOWLEDGEMENTS

I would like to thank Professor P. K. Trivedi for his invaluable insights and continuing support and encouragement. I would also like to thank two anonymous referees for their useful comments. All errors are my own.

REFERENCES

Burdett, K., N. M. Kiefer, and S. Sharma (1985). 'Layoffs and duration dependence in a model of turnover', *Journal of Econometrics*, 28, 51-62.

- Cox, D. R. (1961). 'Tests of separate families of hypotheses', in Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California, Berkeley, pp. 105-123.
- Cox, D. R. (1962). 'Further results on test of separate families of hypotheses', *Journal of the Royal Statistical Society, B*, 24, 406-424.
- Elbers, C., and G. Ridder (1982). 'True and spurious duration dependence: the identifiability of the proportional hazard model', *Review of Economics Studies*, 49, 403-409.
- Efron, B., and D. V. Hinkley (1978). 'Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher Information', *Biometrika*, 65, 457-482.
- Glaser, R. E. (1980). 'Bathtub and related failure rate characterizations', *Journal of the American Statistical Association*, 75, 667-672.
- Heckman, J., and B. Singer (1984). 'The identifiability of the proportional hazard model', *Review of Economic Studies*, 51, 231-241.
- Hougaard, P. (1986). 'Survival models for heterogeneous populations derived from stable distributions', *Biometrika*, 73, 387-396.
- Jaggia, S., (1990). 'Specification diagnostics in the presence of multiple misspecifications for parametric duration models', Unpublished Ph.D. dissertation, Indiana University, Bloomington.
- Jaggia, S. and P. K. Trivedi (1990). 'Joint and separate score tests for state dependence and unobserved heterogeneity', Unpublished Indiana University manuscript.
- Kennan, J. (1985). 'The duration of contract strikes in U.S. manufacturing', *Journal of Econometrics*, 28, 5-28.

- Kiefer, N. M. (1984). 'A simple test for heterogeneity in exponential models of duration', *Journal of Labor Economics*, 2, 539-549.
- Kiefer, N. M. (1988). 'Econometric duration data and hazard functions', *Journal of Economic Literature*, 25, 646-679.
- Lancaster, T. (1983). 'Generalised residuals and heterogeneous duration models: the exponential case', *Bulletin of Economic Research*, 35, 71-86.
- Lancaster, T. (1985). 'Generalised residuals and heterogeneous duration models: with applications to the Weibull model', *Journal of Econometrics*, 28, 155-169.
- Lancaster, T., and A. D. Chesher (1985a). 'Residual analysis for censored duration data', *Economics Letters*, 18, 35-38.
- Lancaster, T., and A. D. Chesher (1985b). 'Residuals, tests and plots with a job matching application', *Annales de l'IVSEE*, 59160, 47-70.
- Lawless, J. F. (1982). *The Statistical Models and Methods for Lifetime Data*, Wiley, New York.
- Manton, K. G., E. Stallard and J. W. Vaupel (1986). 'Alternative models for heterogeneity of mortality risks among the aged', *Journal of the American Statistical Association*, 81, 635-644.
- Pagan, A. R., and F. Vella (1988). 'Diagnostic tests for models based on individual data: a survey'. Rochester Center for Economic Research, Working Paper No. 162.
- Pereira, B. de B. (1978). 'Tests and efficiencies of separate regression models', *Biometrika*, 65, 319-327.

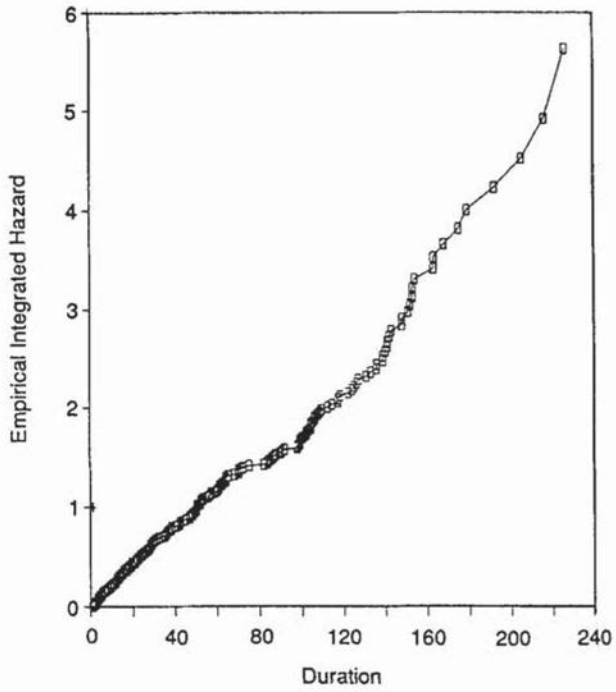


Figure 1. Integrated hazard for X below its mean

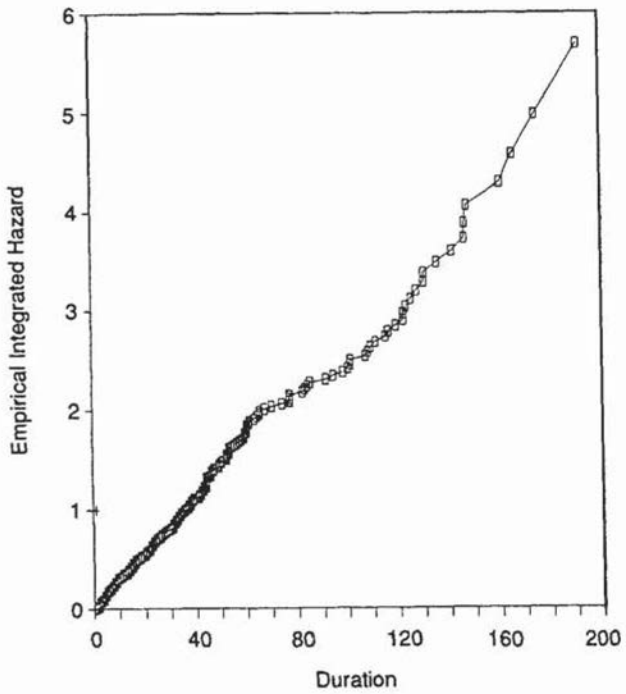


Figure 2. Integrated hazard for X above its mean

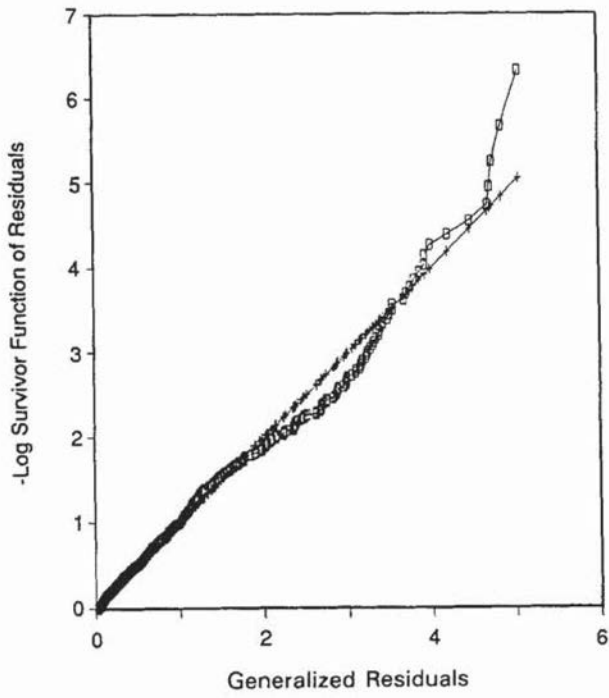


Figure 3. Exponential model

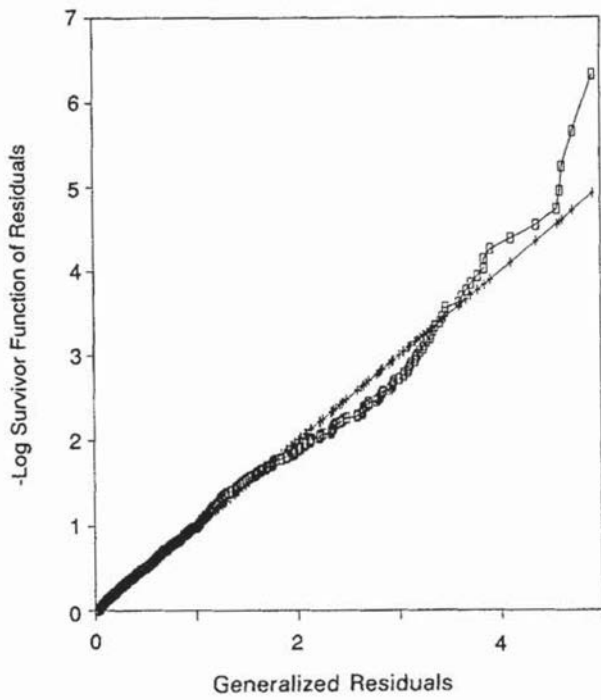


Figure 4. Weibull model

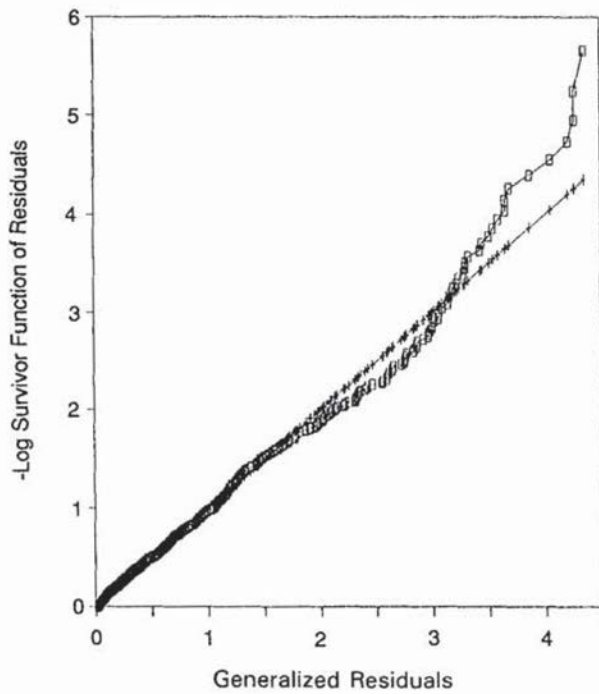


Figure 5. Generalized gamma model

Table I. Score test results for an exponential model

Restrictions	Test statistics	<i>p</i> -value
$\sigma^2 = 0$	0.156109	0.6928
$\alpha = 1$	0.439033	0.5076
$k = 1$	0.092015	0.7616
$\sigma^2 = 0, \alpha = 1$	0.476700	0.7879
$\sigma^2 = 0, k = 1$	0.161478	0.9224
$\alpha = 1, k = 1$	2.455930	0.2929
$\sigma^2 = 0, \alpha = 1, k = 1$	13.445637	0.0038

Table II. Score test results for a Weibull model

Restrictions	Test statistics	<i>p</i> -Value
$\sigma^2 = 0$	0.014199	0.9051
$k = 1$	2.009156	0.1564
$\sigma^2 = 0, k = 1$	11.870698	0.0026

Table III. Score test results for a generalized gamma model

Restriction	Test statistic	<i>p</i> -Value
$\sigma^2 = 0$	60.056180	0.0000

Table IV. Parameter estimates under alternative model specifications

Variable	Exponential	Weibull	Generalized gamma
Constant	-3.7826 (0.0421)	-3.6936 (0.1638)	-2.0585 (1.2168)
X	2.5072 (0.8539)	2.4605 (0.8706)	1.8379 (0.8060)
α		0.9789 (0.0385)	0.7125 (0.1913)
k			(0.7975)
Log-likelihood value	-2698.420	-2698.205	-2696.619

*Standard errors in parentheses.