Order Number 9030419

Specification diagnostics in the presence of multiple misspecifications for parametric duration models

Jaggia, Sanjiv, Ph.D.

Indiana University, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

SPECIFICATION DIAGNOSTICS IN THE PRESENCE OF MULTIPLE MISSPECIFICATIONS FOR PARAMETRIC DURATION MODELS

Sanjiv Jaggia

Submitted to the faculty of the Graduate School
in partial fulfillment of the requirements
for the degree Doctor of Philosophy
in the Department of Economics
Indiana University

May, 1990.

Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements of the degree of Doctor of Philosophy.

Doctoral Committee:

Pravin K. Trivedi, Chairman

Sanghamitra Das

R. Jeffery Green

Thomas J. Kniesner

Date of Oral Examination:

April 20, 1990

Acknowledgements

I would like to express deep gratitude to many individuals who made the completion of this project possible. I will start by thanking the members of my dissertation committee -- Professors Sanghamitra Das, Jeffery Green, Thomas Kniesner and Pravin Trivedi -- for their help and encouragement. I am especially grateful to my dissertation chairman, Professor Pravin Trivedi, for introducing me to duration models which finally became the topic of this dissertation. He has always been extraordinarily generous with his time, spending many hours carefully reading earlier drafts and discussing many issues addressed in this thesis. His extensive comments have greatly improved both the scope of the dissertation and the quality of its exposition. I would also like to thank Professor Essie Maasoumi, who was part of my committee in the early stages of my dissertation. I feel privileged to have had the opportunity to attend the two courses he taught in econometrics that sparked my interest in this area.

I am also indebted to a great friend, Rilla, who is also, coincidentally, my wife. Completion of this dissertation would have been a lot harder without her. I cherished her companionship in this long and, sometimes, excruciating process that she simultaneously went through. Her incessant confidence in me pulled me out of the doldrums that frequently accompanied this process. She has also been instrumental in the better exposition of this dissertation.

I would also like to extend my deepest gratitude to my persevering parents and family who have patiently waited for the completion of my graduate work. Their belief in me has been a constant source of strength. I am especially grateful for the continuing concern and encouragement expressed by my sister Mona.

And last, but not least, I would like to express my thanks to all my friends. Romi and Beena have always been special. Satish, Arindam, Gulten and Mo to name but a few, have made Bloomington into a second home.

Abstract

SPECIFICATION DIAGNOSTICS IN THE PRESENCE OF MULTIPLE MISSPECIFICATIONS FOR PARAMETRIC DURATION MODELS

Sanjiv Jaggia

Parametric models have been used widely in the analysis of duration data. The term duration data is used in the context of the duration of time until the occurrence of some event of interest. Such models have often been used in econometrics to explain why different people spend varying lengths of time in the state of unemployment. Also of interest has been the question of whether the probability of finding a job changes with the length of time a person is unemployed. Duration models are generally estimated using maximum likelihood methods. However, such estimates may be inconsistent in the presence of model misspecification. Consequently, it is desirable to have diagnostic tests that examine the validity of the distributional assumptions made in the parametric duration models.

This thesis is aimed primarily at identifying various sources of misspecification and developing tests for model evaluation. It is emphasised that the conventional approach of testing each parametric restriction in isolation is inconclusive when multiple misspecifications exist concurrently. Incorrect inferences may be drawn in such situations, especially when separate tests are correlated. The need to compute an omnibus statistic is stressed and some joint tests of this form are developed. Such a statistic tests all the relevant assumptions made within a given model jointly and hence has power against several forms of misspecification. Furthermore, modified separate tests are developed that can provide the additional information needed to pin-point the exact error, once the joint null hypothesis is rejected. Finally, some adjustments to standard tests, that are valid in the presence of censored data as well, are suggested in this thesis. An empirical application and extensive Monte Carlo evidence are provided as illustration for all the above mentioned tests.

CONTENTS

CHAPTER 4

CHAPTER 5

CHAPTER 6

CHAPTER 1

INTRODUCTION

Parametric models have been used widely in the analysis of duration data. The term duration data is used in the context of the duration of time until the occurrence of an event. An example of the above is the duration of time spent by an individual in the state of unemployment. However, distortions in the econometric results of such analyses that may be caused by an incorrectly parametrised model have led to a shift of emphasis from the estimation and interpretation of econometric results to the evaluation of such distortions and model testing. The main focus of many studies in econometrics has been the reasons for and the consequences of neglected (unobserved) heterogeneity in the data.[1] Due to the inherently non-linear nature of duration models, parameter estimates may be highly biased if such heterogeneity is ignored. As a result, most of the diagnostic tests proposed have been tests of heterogeneity.[2]

In a different vein, Manton, Stallard and Vaupel (1986) show

[1] see, for example, Salant (1977), Lancaster (1979, 1983, 1985), Lancaster and Nickell (1980), Elbers and Ridder (1982), Heckman and Singer (1982, 1984a, 1984b).

[2] see Lancaster (1983,1985), Lancaster and Chesher (1985a, 1985b), Kiefer (1984, 1985, 1988), Burdett et al. (1985), Jensen (1987) and Sharma (1989).

that model estimates may be very sensitive to the choice of hazard function.[3] As a result, selecting an inappropriate parametric hazard function may be another significant source of specification error and may lead to biased estimates. Therefore, for a thorough model evaluation, one should test for the functional form specification of the hazard function along with any test of heterogeneity.

The following thesis is aimed primarily at identifying various sources of misspecification and testing for them in the context of duration models. Different specification diagnostics for parametric duration models are motivated and developed. The tests include tests of heterogeneity and functional form specification. It is emphasised that the conventional approach of testing each parametric restriction in isolation is inconclusive when multiple misspecifications exist concurrently. Incorrect inferences may be drawn in such situations, especially when separate tests are correlated. The need to compute an omnibus statistic is stressed and some joint tests of this form are developed. Such a statistic tests all the relevant assumptions made within a given model jointly and hence has power against several forms of misspecification. Furthermore, modified separate tests are developed that can provide the additional

[3] see, also, Ridder and Verbaekel (1983) and Trussell and Richards (1985).

information needed to pin-point the exact error, once the joint null hypothesis is rejected. Finally, some adjustments to standard tests, that are valid in the presence of censored data as well, are suggested in this thesis. An empirical application and extensive Monte Carlo evidence are provided as illustration for all the above mentioned tests.

In what follows, a summary of each chapter of the thesis is provided. Chapter 2 contains a brief review of some of the characteristic features of duration models and techniques for the parametric estimation of such models. The Weibull model is used for exposition. The causes and consequences of the problem of heterogeneity are considered. Some of the tests for neglected heterogeneity proposed in the literature are analysed. These tests include the informal graphical methods and the more formal tests that are based on parametric alternatives. It is shown that all the parametric tests examined basically amount to testing the same moment restriction of the generalised error.

A formal diagnostic statistic that is frequently used in testing is based on the integrated hazard function which is a generalised error in the sense of Cox and Snell (1968). Since the generalised error, so defined, has a unit exponential distribution when the model is correctly specified, testing its second moment restriction has formed the basis for model evaluation. Such a test has been interpreted as a score test for heterogeneity for a small variance of the heterogeneity term (Lancaster (1985), Kiefer (1984), Burdett et al (1985)). Chesher (1984) has shown that this test is the same as White's Information Matrix test with respect to the intercept, since neglected heterogeneity causes random variations in the intercept term.

In Chapter 3, it is pointed out that for the above-mentioned test to be a valid heterogeneity test, the hazard rate function has to be correctly specified. More generally, the use of a joint (simultaneous) test of all moment restrictions of the so defined generalised error is preferable to that of a test of the second moment only. Such a test will be a general misspecification test rather than a test of heterogeneity per se. It is also noted that there is no unique way of defining a generalised error. Any non-linear function of the integrated hazard function can be interpreted as a generalised error in the sense of Cox and Snell. For example, if ϵ is defined as the integrated hazard, then $\epsilon_1 = \log(\epsilon)$ has a standard extreme value distribution with well defined moments. Similarly, one can define an infinite number of generalised errors that may be used for the diagnostic testing of the model. However, testing moment restrictions of all such errors will be especially meaningful only if an interpretation can be provided for a particular moment restriction, given that all the other restrictions are satisfied by the data.

Some of the easily understood moment restrictions of generalised errors are considered in this chapter. Some general specification tests are developed along with the heterogeneity test. Score tests of the functional form specification are developed within a fairly flexible generalised gamma distribution. All tests of misspecification are given the tests of conditional moment restriction interpretation. Using the Tauchen (1985) and Newey framework (1985), conditional moment restriction tests are expounded in a general setup. The moment restriction testing framework is used to develop diagnostics for the exponential and Weibull specifications. Kennan's (1985) strike data is used as an empirical illustration. It is inferred that the application of joint tests rather than separate tests is necessary in evaluating the specification of parametric models. It is shown that erroneous conclusions may be reached if separate tests are implemented, as in a number of previous analyses of Kennan's strike data.

A known problem in testing for more than one source of misspecification jointly is that the procedure is incomplete, as the rejection of a joint test is not sufficient to identify the exact source of misspecification. Some information from the separate tests of misspecification is necessary. As the standard separate tests are of little use when such tests are correlated, adjusted score tests, that may provide the additional information needed to pin-point the exact error, are proposed in Chapter 4.

In conducting separate score tests on some given parametric restrictions, one typically assumes the validity of auxiliary restrictions on some nuisance parameters. This method can have misleading consequences if these auxiliary restrictions are not valid. Such restrictions translate into the assumption that the score vector with regard to the nuisance parameters is equal to zero. Instead of simply making this assumption, a test is proposed that is conditional on the realised value of the score vector with respect to the nuisance parameter. With this technique, the relevant score, corresponding to the parameters to be tested, is cleansed of its correlation with the score vector corresponding to the nuisance parameters.

The properties of the suggested adjusted test depend on the choice of the estimate of the nuisance parameter used to evaluate the scores. If any root-N consistent estimate of the nuisance parameter is used, the resulting test is robust and is the same as Neyman's $C(\alpha)$ test. This test is especially useful when it is difficult to obtain maximum likelihood estimates of the nuisance parameters. Based on the preliminary outcome of these tests, an applied researcher can determine if it is worthwhile to continue the study and estimate the model with maximum likelihood methods that ensure efficiency.

The second variant of the adjusted score test is relevant when the root-N consistent estimate of the nuisance parameter is also not available. Here, the scores at the restricted maximum likelihood estimator are evaluated such that the restrictions comprise both the auxiliary restrictions as well as the ones being tested. In other words, instead of estimating the nuisance parameter, auxiliary restrictions are placed on it. This test is valid only under the joint null, however, in this case, the standard separate tests will also be valid. Nevertheless, it has the merit of having been derived under a more general alternative and includes the adjustment factor for correlation between the scores, even though the adjustment is done under the restricted joint null. Therefore, even though both tests are based on the same joint null, the adjusted score tests will contain more information regarding the specific source of misspecification than the standard separate tests.

The above two renditions of the adjusted tests are developed to test for heterogeneity and duration dependence. The $C(\alpha)$ version of the test is based on the consistent estimates obtained from running an ordinary least square regression. Monte Carlo analyses of these tests along with the joint and standard separate tests are provided in this chapter.

A peculiar feature of duration models is that data on durations are seldom complete. It is common for some observations to be censored, typically right censored. In Chapter 5, modifications of diagnostic tests when the data consist of censored observations are considered.[4] The test of heterogeneity is used as an example even though the suggested procedures can also be applied for other specification tests.

In order to implement a score test for heterogeneity, the theoretical information matrix, under the null, has to be evaluated. With censored observations, such a matrix cannot be derived without additional assumptions regarding the censoring mechanism. In this chapter, a distinction is made between data that are Type-1 censored and those that are

[4] Horowitz and Neumann (1989), using Kennan's strike data, show that the testing procedures may lead to erroneous conclusions when data consist of censored observations.

not. With Type-1 censored data, a heterogeneity test based on the theoretical information matrix is derived although it involves evaluating some expressions numerically. A method is suggested that uses only uncensored observations to implement this test when data are not Type-1 censored but are assumed to be censored randomly, involving no length-biased censoring. The method consists of first estimating the parameters of the model using all observations and then using these consistent and efficient estimates to evaluate scores consisting only of complete observations. It is assumed that the fact that an observation is complete has no influence on the duration of the random variable.

Monte carlo analyses of the above tests are carried out in this chapter. Also considered in this chapter are tests based on the observed information matrix. Two candidates considered for this matrix are the sample hessian of the log-likelihood function and the outer product of the sample scores. It is found that the performance of the tests based on the observed information matrix is case sensitive. The information matrix based on the sample hessian is not always positive definite, a phenomenon that makes the test meaningless. Tests based on the outer product of the sample scores are easy to implement but the nominal size of these tests is different from the actual size. The number of times that the test is rejected when the model is correctly specified, is more than the chosen level of significance for many of the Monte Carlo experiments conducted in the chapter. However, with Type-1 censored data, such tests seem to perform reasonably well. The relative performance of a test based on Kiefer's method for approximating the standard error of the relevant score is also examined.

Finally, Chapter 6 contains a summary of the results of this thesis and concluding remarks.

CHAPTER 2

RELEVANT LITERATURE REVIEW WITH SPECIAL EMPHASIS ON SPECIFICATION TESTING

2.1 Introduction

Parametric models have been used widely in the analysis of duration data. The term duration data is used in the context of measuring the duration of time until some event of interest. In econometrics, duration models have been used to study labor market transitions using the data on duration of unemployment spells of individuals. Such models have been especially utilised to explain why different people spend varying lengths of time in the state of unemployment. Also of interest has been the question of whether the probability of finding a job changes with the length of time a person is unemployed. The modelling process is typically initiated by specifying a family of duration distributions, up to a finite number of parameters, to be used for reduced form estimation of the model.

In Section 2.2, characteristic features of duration models, including some definitions, are analysed. Throughout this exposition, unemployment is used purely as an example to describe the modelling process.[1] As the major emphasis of this thesis is on testing for parametric restrictions, some existing testing procedures based on residuals are surveyed. Residuals have often been used to evaluate the parametric restrictions on a given model. However, unlike in linear regression models, there is no natural choice for these residuals in, generally non-linear, duration models. In this section, generalised errors and residuals are defined that may be used for specification testing.

Section 2.3 contains a discussion of the problem of neglected heterogeneity caused primarily by omitted regressors. It is shown that parameter estimates may be biased if such heterogeneity is ignored. Testing procedures for model misspecification, with emphasis on neglected heterogeneity, are considered in Section 2.4. Informal graphical methods are suggested that may be used to evaluate model specification. More formal parametric tests for neglected heterogeneity are also discussed. It is shown that all such parametric tests are based on the generalised residuals defined in Section 2.2. These tests amount to testing for the second moment restriction of the generalised errors and are, thus, interpreted as tests for conditional moment restrictions.

---

[1] Kiefer (1988) provides a survey and various applications of duration models in economics.

2.2 Overview

2.2.1 Definitions

In duration models, it is customary to specify the parametric form of the model in terms of its hazard function. This function specifies the instantaneous rate of escape from the state of unemployment at time t, given that the individual was unemployed up until t. The **hazard function**, h(t), is defined as:

$$h(t) = \lim_{\delta t \to 0} \frac{\Pr(t \leq T \leq t+\delta t \mid T \geq t)}{\delta t}$$

$$= \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \qquad (2.2.1)$$

where f(t) is the **probability density function** and F(t) is the **distribution function** of the random variable, T, which denotes duration of unemployment. S(t) = 1 - F(t) is called the **survivor function**. Other names for the hazard function found in the literature are "force of mortality" and "failure rate".

Since h(t) = - (d/dt) log S(t), the density and survivor functions can be written in terms of the hazard function as:

$$1 - F(t) = S(T) = \exp(-\int h(s)ds) \text{ and:} \qquad (2.2.2)$$

$$f(t) = h(t)\exp(-\int h(s)ds) = h(t)\exp(-\Omega(t)) \qquad (2.2.3)$$

$\Omega(t) = \int h(s)ds$ is the **integrated hazard function.** (2.2.4)

Even though the functions h(t), f(t), F(t) and S(t) provide mathematically equivalent definitions for the distribution of T, some interesting properties of the data, such as duration dependence, are most easily interpreted in terms of the hazard function. **Duration dependence** is said to exist when (d/dt)h(t) does not equal 0. Dependence is positive when $(d/dt)h(t) > 0$ and negative when $(d/dt)h(t) < 0$.

The general **reduced form** of the hazard function used in estimation is $h(t|X) = \Phi(\mu(X),t)$, where $\mu(X)$ is used in accounting for person specific observed heterogeneity by including a list of all the relevant variables, X. The other component, t, is used to explain the dependence of the hazard function on time. In the case of Cox's (1972) **Proportional Hazard Model,**[2] the hazard rate factors into a function of time and a function of all the relevant covariates:

$$h(t|X) = \mu(X)\Phi(t). \qquad (2.2.5)$$

Here $\Phi(t)$ is the **base-line hazard function** and $\mu(X;\beta)$

[2] see Kalbfleish and Prentice (1980) for a good discussion of Proportional Hazard rate models and their estimation, using non-parametric methods.

describes how the base-line hazard varies with the covariates. β is a vector of unknown parameters. $\mu(X)$ is typically taken as $\exp(X\beta)$ to ensure the non-negativity of the hazard function.

The choice of a function for $\Phi(t)$ is arbitrary. For a Weibull model, $\Phi(t) = \alpha t^{\alpha-1}$. The choice of this base-line hazard function allows for positive ($\alpha > 1$), negative ($\alpha < 1$) and no duration dependence ($\alpha = 1$).[3] The hazard, survivor and density functions of the Weibull distribution conditional on the regressors are:

$$h(t|X) = \exp(X\beta)\alpha t^{\alpha-1} \tag{2.2.6}$$

$$S(t|X) = \exp(-\exp(X\beta)t^{\alpha}) \tag{2.2.7}$$

$$f(t|X) = \exp(X\beta)\alpha t^{\alpha-1} \exp(-\exp(X\beta)t^{\alpha}). \tag{2.2.8}$$

2.2.2 Censored Data and Parameter Estimation

A characteristic feature of duration models is that data on durations are seldom complete. It is common for some observations to be **censored**, typically right censored. An observation is said to be **right censored** if the length of the spell of unemployment that is observed is only a portion of the actual length of the spell, provided that the

[3] The Weibull distribution, however, may be restrictive as it allows only monotonically changing hazard rate. A more general hazard function would allow for at least a 'U' or 'inverted U' shaped hazard rate.

observed portion includes the beginning of the spell. This phenomenon typically occurs because of finite observation periods. Some events of interest may not have ended when the data acquisition period ends.

The data available to an econometrician are on the variables t_i , X_i , and C_i , $i = 1...N$. C is an indicator variable, such that:

$$C = \begin{cases} 1 \text{ if a spell is complete} \\ 0 \text{ if a spell is right censored} \end{cases} \quad (2.2.9)$$

Given the assumption of independent censoring, a right censored spell of length t contributes a probability, $P(T>t)$, rather than a density to the likelihood function. If observations on the pair (t,C) are independent, the **log likelihood function** is generally written as:

$$L = \sum_{i=1}^{N} [C_i \log f(t_i | X_i) + (1 - C_i) \log S(t_i | X_i)] \quad (2.2.10)$$

$$= \sum_{i=1}^{N} [C_i \log h(t_i | X_i) + \log S(t_i | X_i)] \quad (2.2.11)$$

and for a Weibull model is:

$$\sum_{i=1}^{N} \left[C_i \{ \log(\alpha) + (\alpha-1)\log(t_i) + X\beta \} - t_i^{\alpha} \exp(X_i\beta) \right]. \quad (2.2.12)$$

The log-likelihood function can be maximised using standard non-linear maximising procedures to obtain estimates of the parameters $\theta = [\alpha, \beta']'$. If the model has been correctly specified, then under mild regularity conditions, $\tilde{\theta}$ is consistent for θ with a probability density function that can be approximated in the usual way by a normal distribution with the precision matrix equal to Fisher's (theoretical) Information Matrix. The theoretical information matrix is derived as minus the expected value of the second derivative (hessian) of the log likelihood function. However, finding this expected value may be problematic sometimes, especially when the data consist of censored observations. Typically, the observed (sample) information matrix is used to find the precision matrix.

2.2.3 Generalised Residuals

Residuals have been widely used to assess the adequacy of linear models. An examination of residuals from a fitted model is an important way of testing the parametric assumptions made in linear regression models.[4] In non-linear models, there may not be any natural or automatic way of defining a residual. Following Cox and Snell (1968), residuals can be defined in a general sense that can be used for diagnostic checks of non-linear models. These

---

[4] see Pagan and Hall (1983) for a survey of residual based tests in linear regression models

generalised residuals are obtained through a set of transformations of the observations that results in a random variable that has a simple and known distribution, if the model is correctly specified.

Consider the following model:

$Y_i = g_i(X_i, \theta, \epsilon_i)$ where $i = 1 \ldots N$.

Here the ith observation on a random variable Y_i is written as a function of:

a) the characteristics of the individual represented by the observed vector X_i,

b) a vector of unknown parameters θ, and

c) an unobserved random variable ϵ_i where ϵ_i 's are i.i.d.

If each equation $Y_i = g_i(X_i, \theta, \epsilon_i)$ has a unique solution for ϵ_i, such that $\epsilon_i = h_i(Y_i, X_i, \theta)$, then all the ϵ_i 's are defined as the **generalised errors** of the model in the sense of Cox and Snell. If θ is replaced by its maximum likelihood estimate, the **generalised residuals** are obtained as:

$$\tilde{\epsilon}_i = h_i \, (Y_i, X_i, \tilde{\theta}) \qquad (2.2.13)$$

Note that in a simple linear regression model, $Y_i = X_i'\theta + \epsilon_i$ where the ϵ_i 's are the errors, and

$\tilde{\epsilon}_i = Y_i - X_i'\tilde{\theta}$ are the residuals.

In the context of duration models, the survivor and the density functions are written as:

$$S(t|X) = \exp[-\Omega(t|X)] \quad \text{and:}$$

$$f(t|X) = h(t|X)\exp[-\Omega(t|X)].$$

Let $\epsilon = \Omega(t|X)$. The Jacobian for this transformation is:

$$|J| = \frac{dt}{d\epsilon} = \frac{1}{d\Omega(t|X)/dt} = \frac{1}{h(t|X)}.$$

Therefore, the density of ϵ is given by:

$$h(t|X)\exp(-\epsilon)\frac{1}{h(t|X)} = \exp(-\epsilon) \qquad (2.2.14)$$

which is unit exponential.

Given independent durations, the ϵ_i's defined as the integrated hazard functions are generalised errors in the sense of Cox and Snell. For a Weibull model,

$$\epsilon = \exp(X\beta)t^{\alpha}. \qquad (2.2.15)$$

As before, parameters are replaced by their maximum likelihood estimates to obtain generalised residuals. Notice that the choice of a generalised error ϵ, given by (2.2.15), is by no means unique. Any non-linear transformation of ϵ is also a generalised error in the sense of Cox and Snell. However, most of the specification diagnostics suggested in the literature are based on ϵ.

2.3 Unobserved Heterogeneity

In some instances, individuals in a population have differing hazard functions. This phenomenon is referred to as heterogeneity. The individual specific, **observed heterogeneity** can be controlled for by including a list of covariates and conditioning the hazard function on them. A problem arises, however, when some relevant variables are omitted such that the included regressors do not "sufficiently" control for the heterogeneity. Omissions of this type may be encountered when relevant variables are unobservable. Examples of such variables are motivation, ability, spunk etc. Heterogeneity can also arise due to measurement errors in either the covariates, X, or duration time, t.[5] Thus, after conditioning hazard functions on the included regressors, they may still differ across individuals due to some **unobserved heterogeneity.** This remaining heterogeneity may be incorporated into the model by rewriting $\mu(X)$ such that:

$$\mu(X) = \exp(X\beta+U) = V\exp(X\beta) \qquad (2.3.1)$$

where V contains the effect of omitted regressors. Then for a Weibull model:

$$f(t|X,V) = V\exp(X\beta)\alpha t^{\alpha-1}\exp(-V\exp(X\beta)t^{\alpha}). \qquad (2.3.2)$$

[5] Lancaster (1983,1985) discusses different interpretations for unobserved heterogeneity.

2.3.1 Consequences of Unobserved Heterogeneity

Given only cross-sectional observations on N individuals, all the parameters α, β, $V_1 \ldots V_N$, will, in general, not be identifiable. A common solution is to make V a random effect. Let the distribution of V be represented by the density function, p(V). Then the survivor function, given X and V is $S(t|X,V) = \exp(-V\,\Omega(t|X))$. However, given data only on X, one needs to model the distribution of T conditional only on X and this is done by integrating the survivor function with respect to the distribution of V. Thus, conditional only on X,

$$S(t|X) = \int \exp(-V\,\Omega(t|X))\ p(V)dV \qquad (2.3.3)$$

$$= E_v[\exp(-V\,\Omega(t|X))]. \qquad (2.3.4)$$

Similarly, one can derive expressions for the hazard and density functions conditional only on X.

In the simple linear regression model, neglected heterogeneity often leads to an inaccurate interpretation of econometric results. This problem in linear regression analysis becomes serious if the omitted regressors are correlated with those which are included in the regression equation. This, however, is not true when the underlying model is non-linear. It is shown that results from duration models may be misleading even when V is distributed independently of X and t.

In the Weibull model,

$$S(t|X,V) = \exp(-V\mu t^{\alpha}); \; \mu = \exp(X\beta), \text{ and:} \tag{2.3.5}$$

$$S(t|X) = \int \exp(-V\mu t^{\alpha}) \, p(V)dV. \tag{2.3.6}$$

The hazard function conditional on X alone can be derived as:

$$h(t|X) = - \frac{\delta}{\delta t} \log S(t|X)$$

$$= \mu\alpha t^{\alpha-1} \int \frac{V \exp(-V\mu t^{\alpha}) \; p(V)dV}{S(t|X)}$$

$$= \mu\alpha t^{\alpha-1} \, E(V|T\geq t). \tag{2.3.7}$$

Since $E(V|T\geq t)$ is the average of V over the survivors at time t, it must decrease with time as people with higher values of V tend to leave the unemployed state first.[6] This sorting out effect, due to the presence of heterogeneity leads, to a downward biased estimate of duration dependence. More formally, the average of V over time can be written as:

$$E(V|T\geq t) = \int \frac{V \exp(-V\mu t^{\alpha}) \; p(V)dV}{S(t|X)}.$$

[6] Salant (1977), Lancaster and Nickell (1980) etc. discuss this problem of identification in duration models.

Therefore,

$$\frac{\delta}{\delta t} E(V|T\geq t) = -\mu\alpha t^{\alpha-1} \left[\int \frac{V^2 \exp(-V\mu t^{\alpha})\ p(V)dV}{S(t|X)} \right]$$

$$+ \mu\alpha t^{\alpha-1} \left[\int \frac{V \exp(-V\mu t^{\alpha})\ p(V)dV}{S(t|X)} \right]^2$$

$$= -\mu\alpha t^{\alpha-1}\ \mathrm{Var}[V|T\geq t] < 0 \qquad (2.3.8)$$

Thus, neglecting heterogeneity results in an estimated hazard rate that is falling faster or rising more slowly than the actual hazard rate. Furthermore:

$$\frac{\delta}{\delta X_j} E(V|T\geq t) = -\beta_j\ \mu t^{\alpha}\ \mathrm{Var}[V|T\geq t\ ] < 0. \qquad (2.3.9)$$

If there is no unobserved heterogeneity (V is known), then $\log h(t|X,V) = \log(\alpha t^{\alpha-1}) + \log(\mu) + \log(V)$.

Therefore,

$$\frac{\delta}{\delta X_j} \log h(t|X,V) = \beta_j. \qquad (2.3.10)$$

Thus, the proportional impact of changes in X on the hazard function is constant over time, a property of the proportional hazard model. However, with some unobserved heterogeneity,

$$\log h(t|X) = \log(\alpha t^{\alpha-1}) + \log \mu + \log E(V|T\geq t)$$

and:

$$\frac{\delta}{\delta X_j} \log h(t|X) = \beta_j - \frac{\beta_j \mu t^{\alpha} \, Var[V|T\geq t]}{E(V|T\geq t)}$$

$$= \beta_j [1 - \frac{\mu t^{\alpha} \, Var[V|T\geq t]]}{E[V|T\geq t]} \,. \qquad (2.3.11)$$

As a result, the proportional impact of changes in X on the hazard function besides being diminished is also dependent on t and is no longer is of the proportional hazard type.[7] Thus, estimates derived from the model may be misleading even when variables included in the model are not correlated with variables that have been excluded. As a result, a major emphasis of diagnostic testing in duration models has been on testing for neglected heterogeneity.

[7] Lancaster and Nickell (1980) discuss the consequences of unobserved heterogeneity for models that are more general than a Weibull model.

2.4 Testing for Heterogeneity

2.4.1 Non-Parametric Graphical Analysis

The generalised residuals defined above can be used to test the specification of a duration model. For a correctly specified model, the residuals should behave approximately like a random sample taken from a unit exponential distribution. Graphical plots of the generalised residuals are used for informal specification checks of parametric models. As ϵ has a unit exponential distribution under the null, its survivor function, $S(\epsilon) = \exp(-\epsilon)$. Therefore, $-\log s(\epsilon) = \Omega(\epsilon) = \epsilon$ and the estimated integrated hazard function for the residuals is often compared with a 45° line to test the specification of a given model.

If the observations are uncensored, an empirical survivor function of the generalised residuals can be computed as: $\hat{S}(\tilde{\epsilon})$, where $\hat{S}(\tilde{\epsilon}) = N^{-1}$(Number of sample observations $\geq \epsilon$) and minus the logarithm of the survivor function is plotted against the residuals. If the model is correctly specified, the scatter plot should cluster around a 45° line through the origin.[8] Note that such graphical checks are general specification tests, not necessarily directed at testing for neglected heterogeneity.

---

[8] see Lawless (1982), Lancaster and Chesher (1985b), Kiefer (1988) etc. for details.

In Figure 2.1 and 2.2, such scatter plots are depicted. Two computer simulated samples, each consisting of 200 uncensored observations, are generated. The first sample consists of Weibull variates and the second consists of variates derived from a Weibull distribution with some neglected heterogeneity, and hence are not Weibull variates.[9] The maximum likelihood estimates of the parameters, from both sample sets, are obtained using the Weibull model. Using (2.2.15), the generalised residuals are obtained as:

$$\tilde{\epsilon} = \exp(X\tilde{\beta})t^{\tilde{\alpha}}.$$

These residuals are further used to construct the above mentioned graphs to check the adequacy of the Weibull specification. Figures 2.1 and 2.2 represent graphs for a correctly and incorrectly specified models respectively. From Figure 2.1, it seems that the model is correctly specified. The departure from the 45° line in Figure 2.2 is obvious, indicating model misspecification.

In general, however, data consist of both complete and censored observations and thus the residuals have to be suitably adjusted to incorporate censored observations. In the case of right censored observations, the observed

---

[9] Such a mixture distribution is generated using μ=exp(Xß+U) where U is the heterogeneity term. Here, a random draw for U is taken from a normal distribution with var(U) = 2.

duration is $t = \min[T,L]$, implying that an observation is complete only if it is less than the censoring time, L. If the observation is longer, it gets censored at L. Note that $\epsilon(t)$ is no longer distributed unit exponentially. However, as $\epsilon(T)$ has a unit exponential distribution when the model is correctly specified, the following can be derived using the memory-less property of an exponentially distributed random variable:

$$E(\epsilon(T)|T>L) = \epsilon(L) + E(\epsilon(T)) = 1 + \epsilon(L). \qquad (2.4.1)$$

Therefore, the residual can be redefined as:

$$\tilde{e}(t) = \begin{cases} \tilde{\epsilon}(t) \text{ if uncensored} \\ \tilde{\epsilon}(t)+1 \text{ if censored.} \end{cases} \qquad (2.4.2)$$

Even though the modified generalised errors, e(t), do not have a unit exponential distribution, they still have a unit mean. Thus, graphical plots using the modified generalised residuals given by (2.4.2) can still be used to check the adequacy of the model for a moderate amount of censoring. The two samples, mentioned above, are again used to examine the performance of the graphical procedures with censored data. With both data sets, about 22 percent of the observations are artificially censored by fixing the censoring times, L, appropriately. Again, maximum likelihood estimates are obtained and the survivor function of the modified generalised residuals is used for plotting.

Figures 2.3 and 2.4 represent graphs for correctly and incorrectly specified models respectively. With graphical procedures some degree of caution should be exercised in interpreting the results, especially when data are censored.[10] From Figure 2.3, one may infer that a Weibull model is inappropriate even though data are actually generated from a Weibull distribution. When the model is actually misspecified, departure from the straight line is some what more pronounced (Figure 2.4).

Alternatively, as $\epsilon(t)$ is a positive function of t, one can use $\epsilon(t) = \min(\epsilon(T), \epsilon(L))$ directly to plot graphs for model evaluation. If the model is correctly specified, sample observations on $\epsilon(t)$ constitute a random sample drawn from a right censored unit exponential distribution. The Kaplan-Meier procedure can be used directly to compute the survivor function of the generalised residuals, $\epsilon(t)$, along with the indicator function, C, denoting censoring. Figures 2.5 and 2.6 are used to illustrate this method of graphical study of misspecification. As before, Figure 2.5 is based on the Weibull model and Figure 2.6 depicts the results when the underlying model is not Weibull. Here, graphs perform reasonably well. The departure from the 45° line is

[10] Horowitz and Neumann (1989), using Kennan's strike data, show that graphical methods may lead to erroneous conclusions, when data are censored. Lancaster and Chesher (1985) examine some graphs using computer simulated models.

apparently greater in Figure 2.6 as compared with Figure 2.5.

2.4.2 Score Test

Generalised residuals can also be used for more formal diagnostic checks. A score test of heterogeneity can be constructed that tests for the zero variance of the heterogeneity term. Score or the Lagrange Multiplier (LM) tests have recently gained popularity in econometrics as they are based on estimation of the null model that incorporates parametric restrictions.[11] These restricted models are easy to estimate, in many cases, as opposed to the Wald or the Likelihood Ratio tests for which the alternative model has to be estimated.

In order to implement a score test of heterogeneity, the probability density function, $f(t|X)$ has to be specified. As mentioned earlier, the unconditional survivor function is $S(t|X) = E_v[\exp(-\epsilon V)]$ where ϵ is the generalised error. One way to obtain $S(t|X)$ is to take expectation with respect to a specified parametric mixing distribution for V. As economic theory provides no information regarding the choice of any mixing distribution, an arbitrarily chosen

[11] see Breusch and Pagan (1980), Engle (1982, 1984), Bera and McKenzie (1986) and Godfrey (1988) etc. for an exposition and the required regularity conditions.

distribution may lead to distorted results.[12]

Alternatively, one can derive a score test that does not depend on any parametric mixing distribution. Given a constant term in X, E(V) can be set to equal 1 without loss of generality. Let σ^2 represent the variance of V. For small σ^2, $S(t|X)$ can be approximated for by a second order Taylor series expansion of $\exp(-\epsilon V)$ around the unit mean of V as follows:

$$S(t|X) = E_v\left[\exp(-\epsilon) + (V - 1)\left[\frac{d}{dV}\exp(-V\epsilon)\right] + \frac{1}{2}(V-1)^2\left[\frac{d^2}{dV^2}\exp(-V\epsilon)\right]\right]_{V=1}$$

$$= \exp(-\epsilon) - \epsilon\exp(-\epsilon)E_v(V-1) + 1/2\epsilon^2\exp(-\epsilon)E_v(V-1)^2$$

$$= \exp(-\epsilon)\left[1 + \frac{\sigma^2\epsilon^2}{2}\right]$$

$$= S(t|X,V=1)\left[1 + \frac{\sigma^2\epsilon^2}{2}\right]. \qquad (2.4.3)$$

The density and the hazard functions can similarly be derived as:

$$f(t|X) = f(t|X,V=1)\left[1 + \frac{\sigma^2}{2}(\epsilon^2 - 2\epsilon)\right] \text{ and:} \qquad (2.4.4)$$

---

[12] Heckman and Singer (1984a) show that parameter estimates may be very sensitive to the choice of the mixing distribution.

$$h(t|X) = h(t|X,V=1)\ [1 - \frac{\sigma^2 \epsilon}{1+.5\ \sigma^2 \epsilon^2}]. \tag{2.4.5}$$

By using the preceding functions, a score test of whether $\sigma^2 = 0$ can be easily be implemented as a test for unobserved heterogeneity. The log likelihood function, given by (2.2.10), using (2.4.3) and (2.4.4), can be written as:

$$L = \sum_{i=1}^{N} \Big[C[\log f(t|X,V=1) + \log(1 + \frac{\sigma^2}{2}(\epsilon^2 - 2\epsilon)] + (1 - C)[\log S(t|X,V=1) + \log(1+ \frac{\sigma^2 \epsilon^2}{2})] \Big]. \tag{2.4.6}$$

The mean score evaluated at $\sigma^2 = 0$ is:

$$\frac{1}{N} \frac{\delta L}{\delta \sigma^2}\Big|_{\sigma^2=0} = \frac{1}{2N} \sum \Big[C(\epsilon^2 - 2\epsilon) + (1 - C)\epsilon^2 \Big]$$

$$= \frac{1}{2N} \sum[\epsilon^2 - 2C\epsilon]. \tag{2.4.7}$$

Kiefer (1984), Burdett et al (1985) derive a similar mean score by using a 'U' representation of the heterogeneity term where $\mu(X) = \exp(X\beta + U) = V \exp(X\beta)$. Given a constant term in X, the unconditional functions can be approximated by using a Taylor's series expansion of the density function around the zero mean of U as follows:

$$f(t|X) = f(t|X,U=0)\ [1 + \sigma^2/2\ (1 - 3\epsilon + \epsilon^2)]. \qquad (2.4.8)$$

$$S(t|X) = S(t|X,U=0)\ [1 + \sigma^2/2\ (\epsilon^2 - \epsilon)]. \qquad (2.4.9)$$

Using these functions in the likelihood, the mean score is:

$$\frac{1}{N}\frac{\delta L}{\delta \sigma^2}\bigg|_{\sigma^2=0} = \frac{1}{2N}\ \Sigma\Big[C(1 - 3\epsilon + \epsilon^2) + (1 - C)(\epsilon^2 - \epsilon)\Big]$$

$$= \frac{1}{2N}\ \Sigma[\epsilon^2 - 2C\epsilon + C - \epsilon]. \qquad (2.4.10)$$

As $\Sigma(C - \epsilon) = \delta L/\delta \beta_0$ is set equal to zero to maximise the likelihood function, the two approximations to the density functions result in identical diagnostics used in testing for heterogeneity.[13]

If all the observations are complete, the mean score is $(1/2N)\ \Sigma(\epsilon^2 - 2\epsilon)$. A score test is $(1/2N)\ \Sigma(\tilde{\epsilon}^2 - 2\tilde{\epsilon})$ divided by its asymptotic standard error where maximum likelihood estimates have been substituted into $\tilde{\epsilon}$. Lancaster (1985) derives the variance of the mean score equal to 2.55063N for a Weibull model, using the theoretical information matrix. However, with censored data, Lancaster's test cannot be used as the theoretical information matrix cannot be derived without exactly specifying the censoring mechanism.

[13] Jensen (1987) and Sharma (1989) make a similar remark when data consist of uncensored observations only.

Burdett et al.(1985) on the other hand, following the testing procedure proposed by Kiefer (1984,1985), use sample information to calculate the variance of the mean score. For example, let the mean score $(1/N)(\delta L/\delta\sigma^2)$, evaluated at $\sigma^2=0$, be $S = (1/N)\Sigma s_i$. The suggested variance of the mean score is $(1/N^2)\Sigma(si - s_m)$ where s_m is the sample mean of s. As the information matrix, using a Weibull specification, is not block diagonal, the covariance between the elements of the score vector, $\delta L/\delta\sigma^2$ and $\delta L/\delta\Theta$ is ignored where $\Theta = (\alpha\ \beta')'$. As a result, the proposed variance of the test statistic is over-estimated and thus would result in the under-rejection of the null hypothesis of no heterogeneity. Kiefer's (1985) claim that such a testing procedure is conservative in the sense that it leads to more rejections, cannot be true. This approach, however, can be used even when the data consist of some censored observations.

One possible way to improve on the computed variance of the mean score is to use the observed information matrix but include the effect of the off diagonal terms in the matrix. When the model is correctly specified, the theoretical information matrix equality holds and, therefore, two candidates for the observed information matrix are the sample hessian of the log-likelihood function and the outer product of the sample scores. If the outer product form of the observed information matrix is used, the resulting score test can be implemented easily. Let D be a (Nxk) matrix consisting of N sample observations of a (kx1) score vector. Therefore, the observed information matrix is given by D'D. The score test can easily be implemented by running an artificial regression of a vector of ones on D.[14] The test statistic can be calculated as NR^2 where R^2 is the uncentered coefficient of determination derived from this artificial regression.

2.4.3 Information Matrix test

The variation in the intercept term in X can also be viewed as having been caused by neglected heterogeneity. A score test of the hypothesis that the intercept has zero variance can be used as a test for neglected heterogeneity. Chesher (1984)[15] has shown that a score test whose form does not depend on the form of the heterogeneity distribution is the same as the information matrix test of White (1982).
The information matrix test is based on the principle that given the hypothesis of no misspecification, the information matrix can be expressed in either the hessian or the outer product form. For a correctly specified model:

[14] Godfrey and Wickens (1981) first suggested this easily implementable version. See, also, Davidson an MacKinnon (1983).

[15] see, also, Cox (1983).

$$E\left[\frac{\delta^2 \log L}{\delta\theta_o \delta\theta_o'}\right] = -E\left[\frac{\delta \log L}{\delta\theta_o}\right]\left[\frac{\delta \log L}{\delta\theta_o}\right]' \qquad (2.4.11)$$

where θ_o is the true value of the parameter vector. Significant differences between the two can be interpreted as being caused by model misspecification. The test can be implemented by using the sample analogue of moments as follows:

$$dn(\tilde{\theta}) = \frac{1}{N}\Sigma\frac{\delta^2 \log L}{\delta\theta\delta\theta'}\Bigg|_{\theta=\tilde{\theta}} + \frac{1}{N}\Sigma\left[\frac{\delta \log L}{\delta\theta}\right]\left[\frac{\delta \log L}{\delta\theta}\right]'\Bigg|_{\theta=\tilde{\theta}} \qquad (2.4.12)$$

where (~) represents evaluation at the maximum likelihood estimates. The test is conducted to see if vector given by (2.4.12) is significantly different from zero.

The most general form of the test involves comparing all the N(N+1)/2 distinct elements of $dn(\tilde{\theta})$ with zero. However, attention may be confined to any subset of the distinct elements, in particular, the intercept. Given $\mu(X) = \beta_o + X_1\beta_1$ in the log-likelihood function,

$$L = \Sigma[C\log f(t|X) + (1 - C)\log S(t|X)],$$

$$\delta L/\delta\beta_o = \Sigma(C - \epsilon).$$

This implies that $(C - \tilde{\epsilon}) = 0$ as $\delta L/\delta\beta_o$ is set to zero to get to get maximum likelihood estimates. Furthermore,

$$\delta^2 L/\delta \beta_o^2 = \Sigma(-\epsilon) \text{ and:}$$

$$[\delta L/\delta \beta_o]^2 = \Sigma(C - \epsilon)^2.$$

Therefore, $(1/N)\Sigma[\delta^2 L/\delta \beta_o^2] + (1/N)\Sigma[\delta L/\delta \beta_o]^2$ at the maximum likelihood estimates is:

$$= 1/N \ \Sigma[(C - \tilde{\epsilon})^2 - \tilde{\epsilon}]$$

$$= 1/N \ \Sigma[\tilde{\epsilon}^2 - 2C\tilde{\epsilon}]$$

which is the same as the numerator of the score test derived before.

From the above result it can be seen that the test for heterogeneity can be conducted using White's information matrix test. Chesher (1983) and Lancaster (1984) have suggested a simple and easily implementable NR^2 version of the Information Matrix test. This test is carried out by running an ordinary least squares regression. The lhs variable for this regression is simply a vector of ones. The rhs variables are:

a) $[\delta L/\delta \beta_o]^2 + \delta^2 L/\delta \beta_o^2$ and:

b) the scores of all the parameters of the model.

For a Weibull model, the rhs variables are:

$$\frac{\delta L}{\delta \beta_j} = (C - \epsilon)\ X_j \qquad j = 0,\ldots K\ , \qquad (2.4.13)$$

$$\frac{\delta L}{\delta \alpha} = \frac{C}{\alpha} + \log(t)\ (C - \epsilon),\ \text{and:} \qquad (2.4.14)$$

$$[\delta L/\delta \beta_o]^2 + \delta^2 L/\delta \beta_o^2 = (C - \epsilon)^2 - \epsilon. \qquad (2.4.15)$$

The asymptotic $X^2(1)$ statistic can be computed by multiplying N by the uncentered R^2 from the above mentioned artificial regression. Note that this implementable version is algebraically equivalent to the improved Kiefer's version suggested earlier.

2.4.4 Conditional Moment Restriction Test

As seen earlier, all suggested tests of heterogeneity in the literature are based on the generalised residual, $\tilde{\epsilon}$. As ϵ has a unit exponential distribution when the model is correctly specified, testing its second moment restriction has been the basis of tests of heterogeneity. The quantity that is equated with zero is:

$$(1/2N)\ \Sigma[\epsilon^2 - 2C\epsilon]$$

$$= (1/2N)\ \Sigma[\epsilon^2 - 2\epsilon],\ \text{if all the observations are complete}$$

$$= (1/2N)\ \Sigma[\epsilon^2 - 2],\ \text{as } \Sigma\ \epsilon/N = 1$$

$$= (1/2)\ (s^2 - 1) \qquad (2.4.16)$$

where s^2 is the sample variance of the generalised residual.

Thus, the score test or White's information test of no heterogeneity amounts to testing the second moment restriction of ϵ, namely that $Var(\epsilon) = E(\epsilon-1)^2 = 1$.

If some observations are censored, then:

$$(1/2N)\ \Sigma[\epsilon^2 - 2C\epsilon] = (1/2N)\ \Sigma[(e-1)^2 - C]$$

$$= (1/2)\ [s_1^2 - \Sigma(C/N)] \qquad (2.4.17)$$

where $e = \epsilon+1-C$ and s_1^2 is the sample variance of e. This result can be interpreted as testing the second moment restriction of the adjusted generalised residual, e, namely that $Var(e) = E(e-1) = \pi^*$ where π^* is the expected probability of censoring.[16]

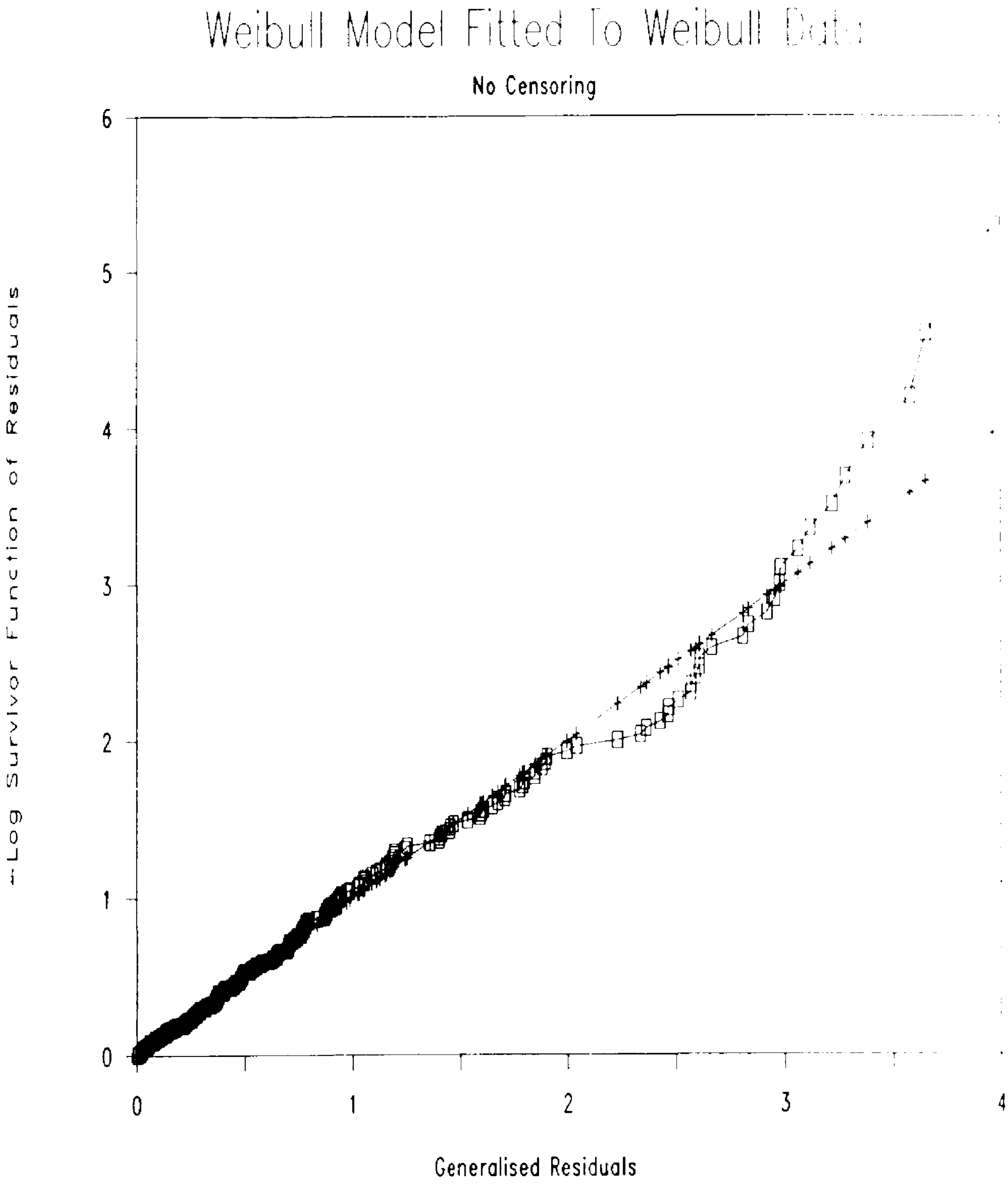
Thus, heterogeneity tests can be considered conditional moment tests as studied by Tauchen (1985) and Newey (1985).[17] An easily implementable rendition of the conditional moment tests exists which can be computed by running an artificial regression. An ordinary least square regression is run where the lhs variable is simply the difference between the theoretical and the predicted moment from the probability model and the rhs variables comprise a constant and all the scores of the model. Testing if a

[16] see Lancaster and Chesher (1985a, 1985b).

[17] Pagan and Vella (1989) provide a good discussion of such tests.

particular moment restriction is satisfied is equivalent to performing a t test for a non zero intercept. A major advantage of this procedure, based on an artificial regression, is that it gives the user a detailed information on the statistical significance of each moment restriction separately, instead of the joint significance of all moment restrictions (Tauchen (1985)). This information, however cannot be useful when moment restrictions are correlated. Validity of a particular moment restriction will depend on the validity of the other if the two are correlated. The joint test of all restrictions is implemented by testing for a non-zero intercept in a SUR regression model.

FIGURE 2.1



Weibull Model Fitted To Weibull Data
No Censoring
-Log Survivor Function of Residuals
Generalised Residuals
0
1
2
3
4
5
6
0
1
2
3
4

FIGURE 2.2

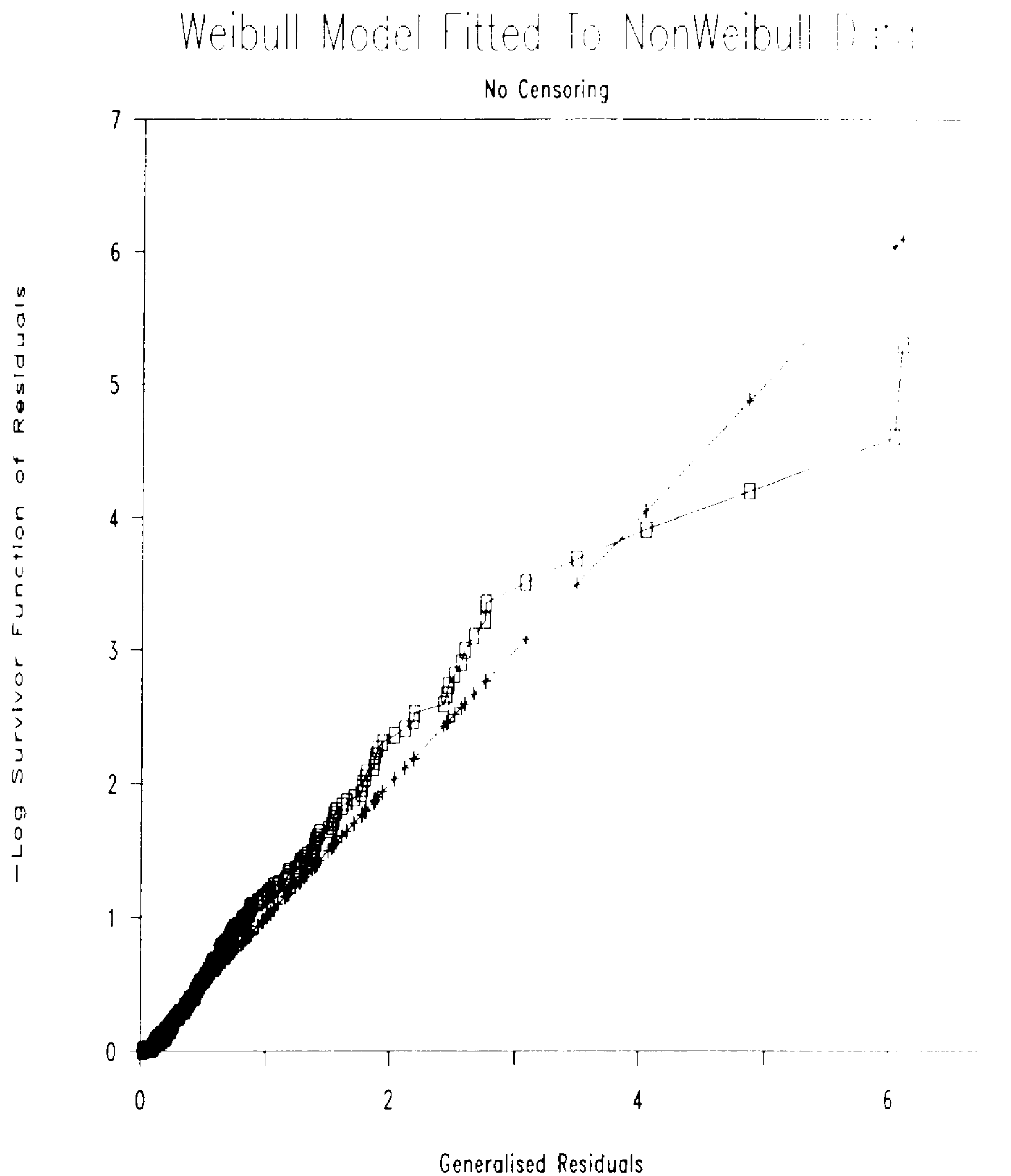

Weibull Model Fitted To NonWeibull
No Censoring
7
6
5
4
3
2
1
0
0
2
4
6
-Log Survivor Function of Residuals
Generalised Residuals

FIGURE 2.3

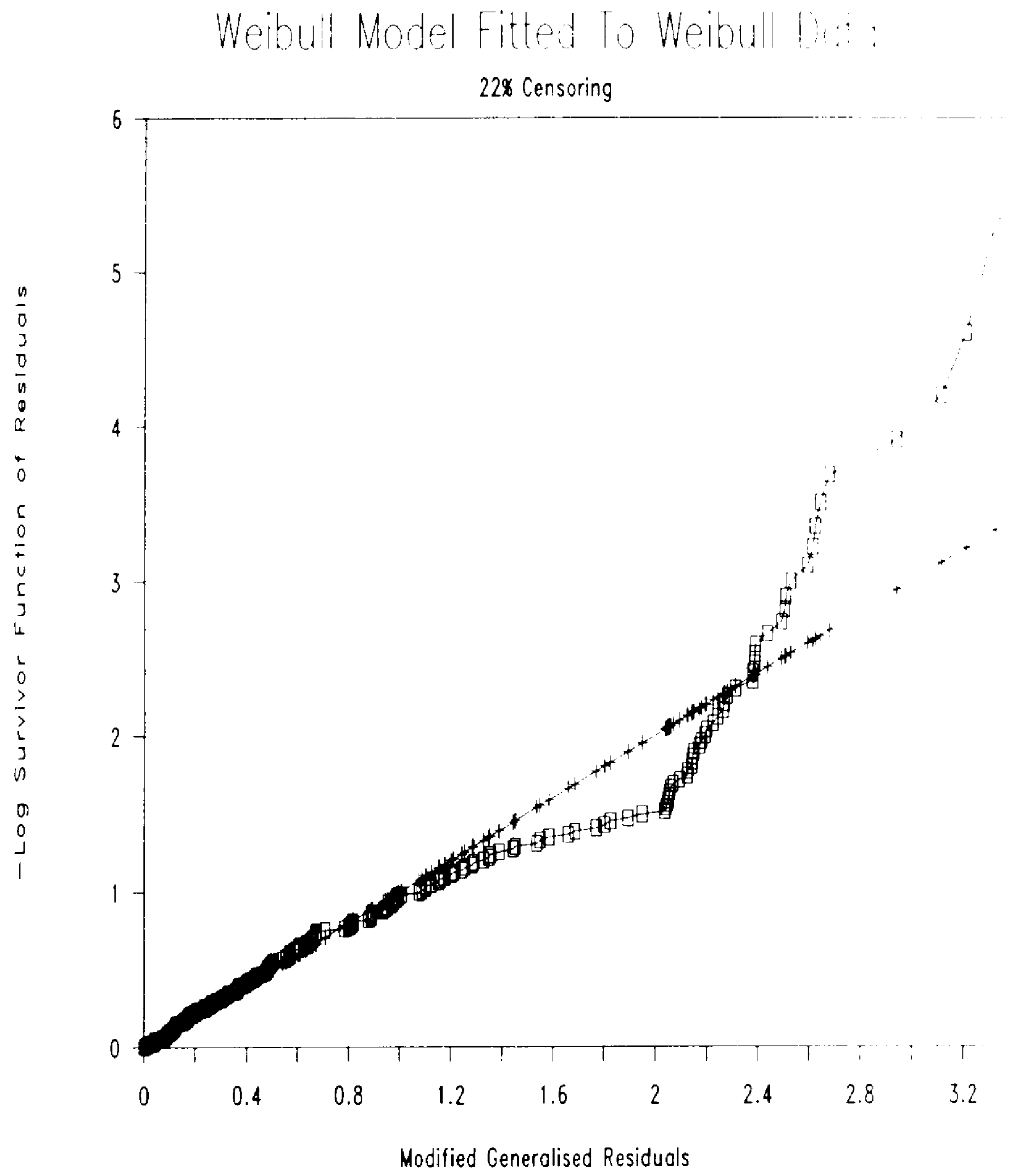

Weibull Model Fitted To Weibull Data
22% Censoring
6
5
4
3
2
1
0
0
0.4
0.8
1.2
1.6
2
2.4
2.8
3.2
Modified Generalised Residuals
-Log Survivor Function of Residuals

FIGURE 2.4

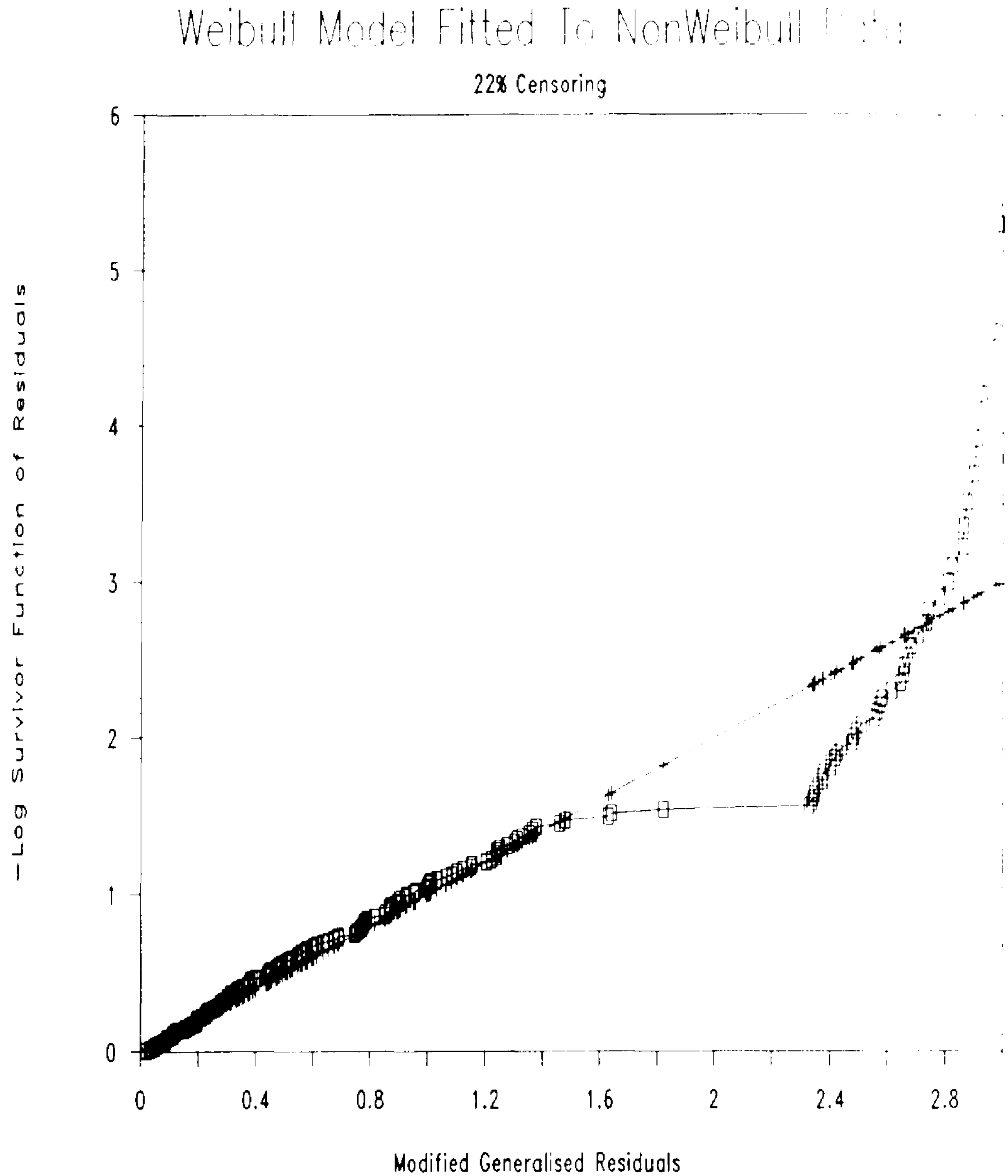

Weibull Model Fitted To NonWeibull
22% Censoring
6
5
4
3
2
1
0
0
0.4
0.8
1.2
1.6
2
2.4
2.8
Modified Generalised Residuals
-Log Survivor Function of Residuals

FIGURE 2.5

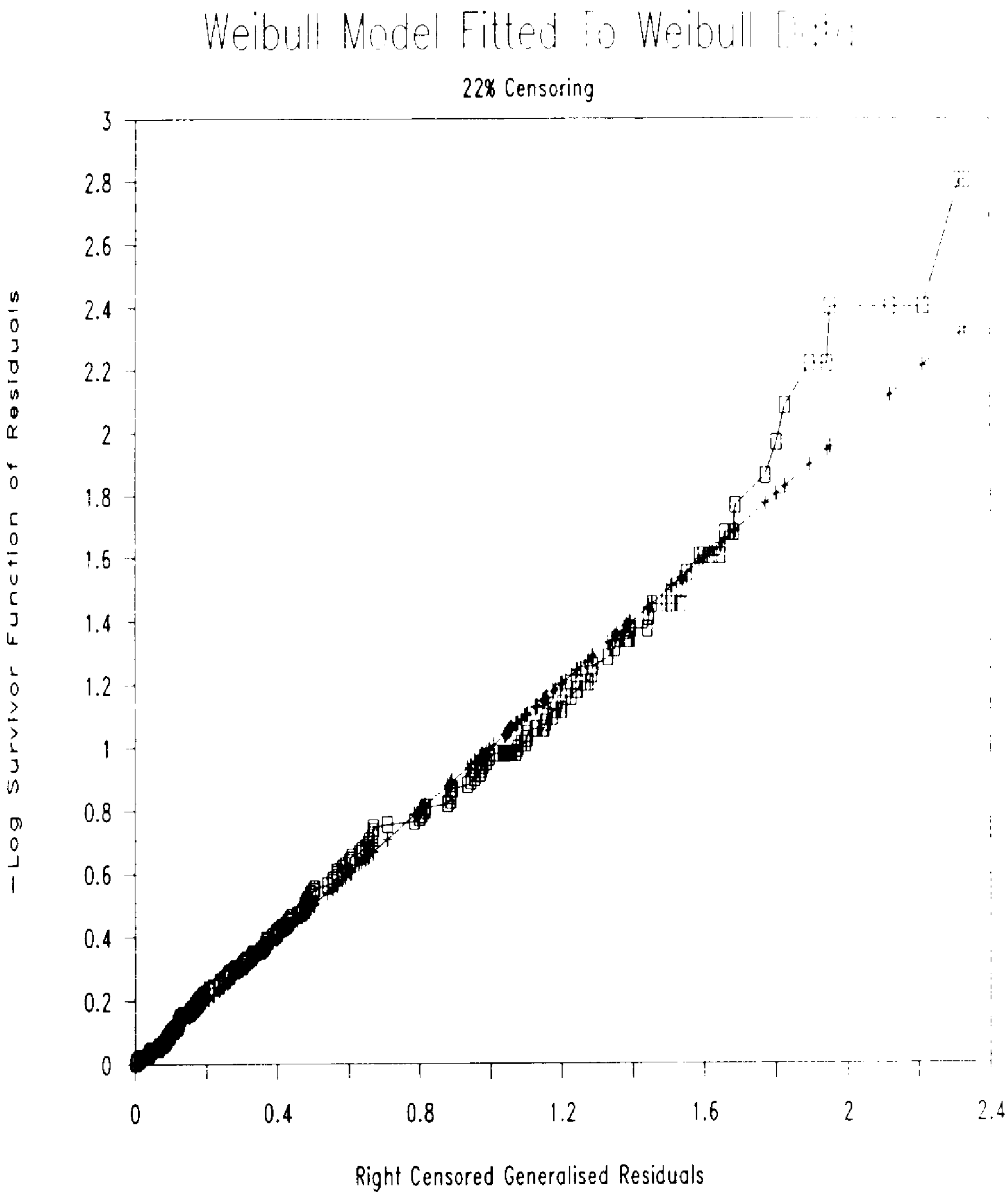
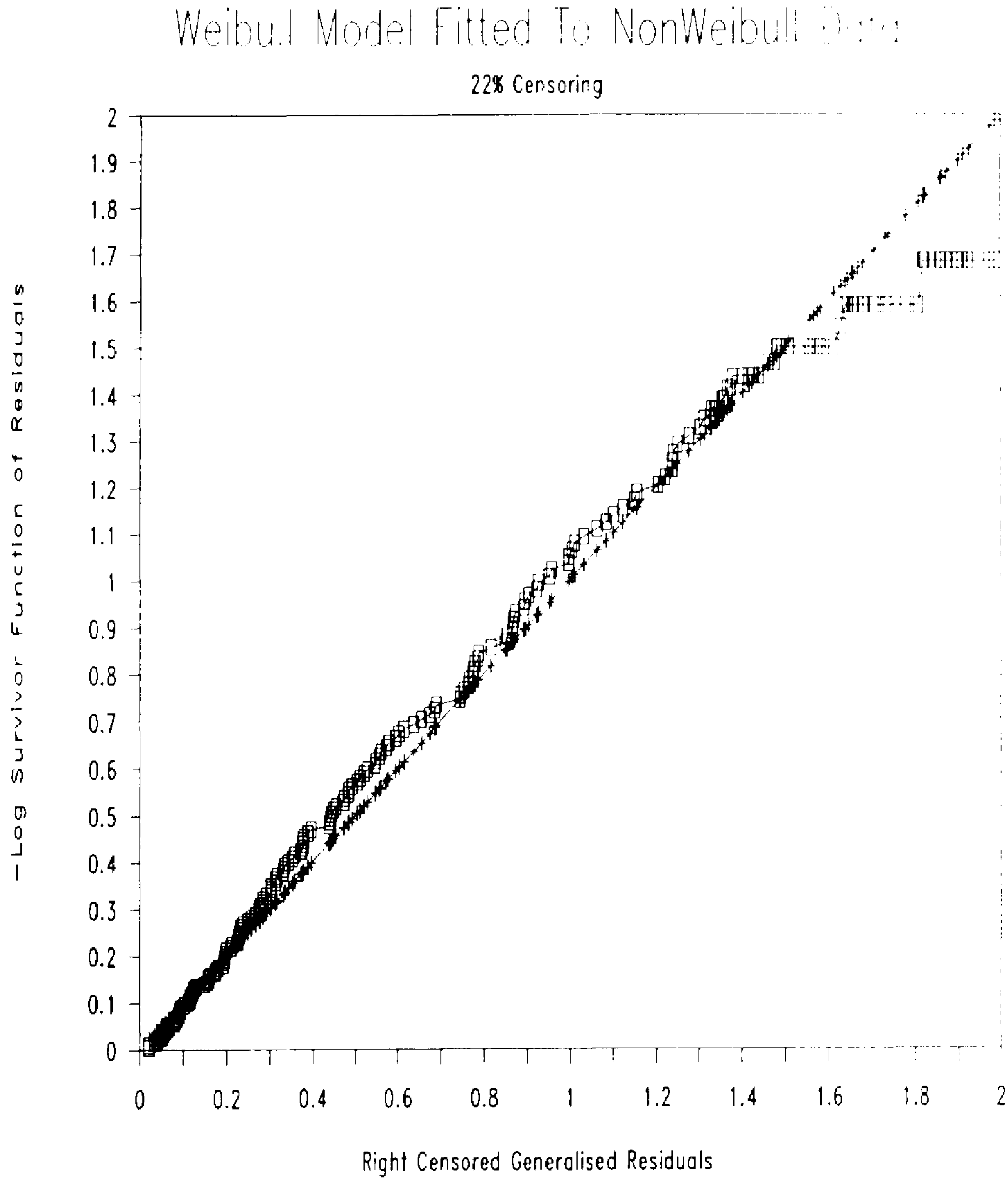

Weibull Model Fitted To Weibull Data
22% Censoring
-Log Survivor Function of Residuals
3
2.8
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.4
0.8
1.2
1.6
2
2.4
Right Censored Generalised Residuals

FIGURE 2.6


Weibull Model Fitted To NonWeibull Data
22% Censoring
-Log Survivor Function of Residuals
Right Censored Generalised Residuals
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

CHAPTER 3

GENERAL SPECIFICATION TESTS OF MOMENT RESTRICTIONS: WITH AN APPLICATION TO KENNAN'S STRIKE DATA

3.1 Introduction

In this chapter it is shown that the application of joint tests of misspecification, rather than partial or separate tests, is necessary in conducting specification tests of parametric duration models. The validity of most of the popular separate tests in the econometric literature relies on some auxiliary assumptions in addition to the ones being tested. For example, a test of heteroscedasticity, generally relies on the auxiliary assumption that there is no serial correlation in the data. In the presence of more than one source of misspecification, inferences drawn from the outcome of any separate test are distorted, especially 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 assumptions made within a given model jointly, and hence has power against several forms of misspecification.[1]

In this chapter, such joint tests are considered for Weibull and exponential duration models. An empirical illustration of the tests based on Kennan's (1985) model of strikes is provided. It is shown that even when none of the separate tests is able to detect misspecification, the joint test indicates ample evidence of misspecification. This is a concrete example of the advantages of omnibus test procedures since they avoid the incorrect inferences drawn in a number of previous analyses of the Kennan strike data.

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 emphasised.[2] The separate heterogeneity test is derived on the assumption that the

---

[1] Bera and Jarque (1982) follow such a procedure in the case of a linear regression model. Moon (1988) considers joint score tests for skewness and heteroscedasticity in a binary logit model.

[2] see Lancaster (1983,1985), Kiefer (1984), Burdett et al. (1985), Lancaster and Chesher (1985a).

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) and Trussell and Richards (1985) have shown that model estimates may be sensitive to the choice of the underlying 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 of heterogeneity.

As mentioned in Chapter 2, the test of heterogeneity is interpreted as testing for the second moment restriction of the integrate hazard function, ϵ, which is the generalised error in the sense of Cox and Snell (1968). However, such a test is a valid separate heterogeneity test only if the conditional duration distribution is correctly specified. The second moment restriction of ϵ cannot be evaluated in isolation when other sources of misspecification exist concurrently. More generally, a joint test of all moment restrictions of ϵ should be used rather than a test of the second moment only. An apparent limitation of such a strategy is that higher order moments may not be estimated accurately from a given sample. Moreover, testing for all higher order moments of ϵ implies a test for general misspecification and thus rejecting the null would not automatically direct one to a better respecified model.

Alternatively, tests are proposed in this chapter that are based on a specified parametric alternative. This point is elucidated in the context of a heterogeneous generalised gamma distribution which can be specialised to exponential, Weibull, gamma and log-normal duration models, with or without heterogeneity. In Section 3.2, a heterogeneous generalised gamma duration model is considered and used to derive joint and separate score tests for functional form misspecification and neglected heterogeneity for a Weibull model. Similar tests are derived for an exponential model. In Section 3.3, these tests are applied to Kennan's strike data. It is inferred that when any separate is implemented, Weibull as well as exponential models seem appropriate. However, when a joint test is applied, both exponential and Weibull models are found to be inadequate. Informal plotting procedures are also applied both for an expository analysis of the data and for testing for parametric models.

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 the source of misspecification, a sequential parametric test procedure of a "general-to-specific" kind is proposed and carried out. A flexible, generalised gamma model is estimated and used to test for heterogeneity in Section 3.4. Similarly, heterogeneity is allowed for through the use of a gamma representation for the distribution of the heterogeneity term in the Weibull and exponential models and is used to test for functional form misspecification. The results indicate neglected heterogeneity in the sample and a non-monotone hazard function for the strike duration which is in accordance with Kennan's inference.

The tests of misspecification are given the tests of conditional moment restriction interpretation in Section 3.5. It is inferred that all tests of misspecification can be interpreted as the tests for moment restrictions of the appropriately defined generalised errors. Using the Newey (1985) and Tauchen (1985) framework, the procedure for such tests is expounded in a general setup. In Section 3.6, tests of moment restrictions for Weibull and exponential models are developed. The tests of higher order moment restrictions of ϵ are applied to Kennan's strike data. It is seen that even though the tests of the second and third moment restrictions of ϵ do not spot misspecification, the results change when the fourth moment restriction is added on to the second and third moment restriction. Section 3.7 contains the concluding comments.

3.2 Score Tests For Misspecification

3.2.1 Generalised 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 generalised gamma distribution (g.g.d.). The density function of a g.g.d. is:

$$f(t|X) = \mu^k \alpha t^{\alpha k-1} \exp(-\mu t^\alpha)/\Gamma(k). \qquad (3.2.1)$$

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.

Pereira (1978), using Cox's (1961, 1962) approach for testing non-nested hypotheses, develops tests to discriminate between log-normal, gamma, Weibull and exponential models. This approach, however, becomes intractable when heterogeneity is allowed for.[3] 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 generalised gamma distribution which can be specialised to the above models with or without heterogeneity. Even though it is difficult to estimate parameters of this distribution, score tests can be easily

[3] Heckman and Walker (1987) suggest goodness of fit criterion, among other criteria, to validate competing non-nested models.

implemented since only the null model needs to be estimated for such tests.

Given some unobserved multiplicative heterogeneity, represented by V, the distribution (3.2.1) conditional on V can be written as:

$$f(t|X,V) = V^k \mu^k \alpha t^{\alpha k-1} \exp(-V\mu t^\alpha)/\Gamma(k). \qquad (3.2.2)$$

As V is not observable, the unconditional distribution can be derived by integrating (3.2.2) with respect to the distribution of V. Given a constant term in X, E(V) can be set to equal 1 without loss of generality. 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:

$$f(t|X,V) = f(t|X,V=1) + (V-1)\left[\frac{\delta}{\delta V} f(t|X,V)\right]_{V=1}$$

$$+ \frac{1}{2}(V-1)^2\left[\frac{\delta^2}{\delta V^2} f(t|X,V)\right]_{V=1}. \qquad (3.2.3)$$

and:

$$f(t|X) = E_V[f(t|X,V)]$$

$$= f(t|X,V=1)[1 + (\sigma^2/2)(k(k-1) - 2k\mu t^\alpha + (\mu t^\alpha)^2)]. \qquad (3.2.4)$$

Note that the unconditional density function, given by (3.2.4), does not depend upon any parametric representation of the heterogeneity distribution.

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:

$$H_o: \sigma^2=0 \text{ and } k=1. \qquad (3.2.5)$$

For the approximate density in (3.2.4), the log likelihood function is:

$$L = \Sigma \Big[k\log(\mu) + \log(\alpha) + (\alpha k - 1)\log(t) - \mu t^{\alpha} - \log \Gamma(k)$$

$$+ \log[1 + (\sigma^2/2)\{k(k-1) - 2k\mu t^{\alpha} + (\mu t^{\alpha})^2\}]\Big]. \qquad (3.2.6)$$

Let $\Theta = (\Theta_1{}' \ \Theta_2{}')'$ where:

$$\Theta_1{}' = (\sigma^2 \ \ k) \text{ and } \Theta_2{}' = (\beta_o \ \beta_1{}' \ \alpha). \qquad (3.2.7)$$

Let $s(\Theta)$ and $I(\Theta)$ denote the score vector and the information matrix respectively. The elements of the relevant score vector, $s_1(\Theta_o)$, evaluated under the null hypothesis are:

$$\left.\frac{dL}{d\sigma^2}\right|_{H_o} = \frac{1}{2}\Sigma(\epsilon^2 - 2\epsilon) = s_{11}(\Theta_o). \qquad (3.2.8)$$

$$\left.\frac{dL}{dk}\right|_{H_o} = \Sigma[\log(\mu) + \alpha\log(t) - \phi(1)]$$

$$= \Sigma[\log(\epsilon) - \phi(1)] = s_{12}(\Theta_o). \qquad (3.2.9)$$

where $\epsilon = \mu t^{\alpha}$ 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:

$$LM_{hf} = s_1'(\tilde{\theta}_o) I^{11}(\tilde{\theta}_o) s_1(\tilde{\theta}_o) \qquad (3.2.10)$$

where (~) denotes that the quantities have been evaluated at the restricted maximum likelihood estimate of the parameter vector, θ, and $I^{11} = [I_{11} - I_{12}(I_{22})^{-1}I_{21}]^{-1}$ is the partitioned inverse of $I(\theta)$. The given test statistic has a chi-square distribution under H_o with degrees of freedom determined by the number of restrictions imposed. The partitioned inverse based on (3.2.5) can be derived as (see Appendix 3A):

$$[I^{11}(\theta_o)]^{-1} = N \begin{bmatrix} 1-q & q-1/2 \\ q-1/2 & \phi'(2)-q \end{bmatrix} \qquad (3.2.11)$$

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

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

$$LM_h = s_{11}' \, [N(1-p)]^{-1} \, s_{11} \qquad (3.2.12)$$

which is similar to the test proposed by Lancaster (1985).

Similarly, a partial test for functional form misspecification, given no neglected heterogeneity, is:

$$LM_f = s_{12}' \ [N(\phi'(2)-p)]^{-1} \ s_{12}. \qquad (3.2.13)$$

From (3.2.11), 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 generalised gamma model, can be tested using:

$$H_o\colon \ \sigma^2=0, \ \alpha=1, \ k=1. \qquad (3.2.14)$$

Implementing the notation given above, with $\theta_1' = (\sigma^2 \ \ \alpha \ \ k)$, the elements of the relevant score vector, $s_1(\theta_o)$, are:

$$\left.\frac{dL}{d\sigma^2}\right|_{H_o} = \frac{1}{2}\Sigma[\epsilon^2 - 2\epsilon] = s_{11}(\theta_o). \qquad (3.2.15)$$

$$\left.\frac{dL}{d\alpha}\right|_{H_o} = \Sigma[1 + (1-\epsilon)\log(t)] = s_{12}(\theta_o). \qquad (3.2.16)$$

$$\left.\frac{dL}{dk}\right|_{H_o} = \Sigma[\log(\epsilon) - \phi(1)] = s_{13}(\Theta_o). \qquad (3.2.17)$$

The test will be based on the following expression for the relevant partition of the information matrix evaluated under (3.2.14):

$$[I^{11}(\Theta_o)]^{-1} = N \begin{bmatrix} 1 & -1 & -1/2 \\ -1 & \phi'(1) & 1 \\ -1/2 & 1 & \phi'(2) \end{bmatrix}. \qquad (3.2.18)$$

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

$$LM_{\sigma^2} = s_{11}'[N]^{-1}s_{11} \qquad (3.2.19)$$

$$LM_\alpha = s_{12}'[N\phi'(1)]^{-1}s_{12} \qquad (3.2.20)$$

$$LM_k = s_{13}'[N\phi'(2)]^{-1}s_{13}. \qquad (3.2.21)$$

Similarly, using the appropriate elements from (3.2.18), 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 (3.2.18), 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.[4]

[4] In Chapter 5 of this thesis, various ways of implementing score tests, in the presence censored observations, are discussed.

3.3 Analysis of Strike Data

3.3.1 Background

In order to illustrate specification tests, data on duration of the contract strikes in U.S. 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 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 duration of completed strikes measured in days and the corresponding value of INDP.

3.3.2 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[5]. 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 homogenous samples for X below and above the mean. The empirical integrated hazard function, $\Omega(t)$, for the two samples is derived. The

[5] see Lancaster and Chesher (1985b), Kiefer (1988), Lawless (1982) and Chapter 2 of this thesis for details.

estimated $\Omega(t)$ = minus log of the estimated sample survivor function, $S(t)$, where:

$$\hat{S}(t) = N^{-1}(\text{Number of sample observations} \geq t).$$

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 3.1 and 3.2, 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 sub-samples.

Graphical plots are also used to ascertain if a particular parametric model is adequate. For a correctly specified parametric duration model, the generalised 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 generalised residuals is obtained. If the model is correctly specified, the scatter plot of this estimate against the generalised residuals should cluster around a 45° line through the origin.

Such scatter plots for the exponential, Weibull and generalised gamma models are plotted in Figures 3.3, 3.4 and 3.5. 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 3.3, that the exponential model is inadequate, the less restrictive models graphed in Figures 3.4 and 3.5 offer no improvement. As a preliminary look at Figures 3.1 and 3.2 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.

3.3.3 Parametric Specification Analysis

All the previous conjectures made using informal graphical plots are here formalised with parametric tests. The specification tests, described in Section 3.2, implemented on the exponential model and the Weibull model are reported in Table 3.1 and Table 3.2 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.

From Table 3.1, it is seen that none of the partial tests detect 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 (1989)). However, the joint null hypothesis of $\sigma^2=0$ and $\alpha=1$ is also not rejected. In fact, none of the two restriction tests suggest 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 generalised gamma distribution, does indicate that the exponential model is inadequate.

The results from the estimated Weibull model, shown in Table 3.2, 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.

This analysis suggests that in order to detect violation of any parametric assumption, no additional auxiliary assumptions should be made in estimation. For example, one has to allow for heterogeneity in estimation to test for functional form misspecification. Similarly, one has to estimate a fairly general hazard model in order to have a valid test of heterogeneity. This will provide the useful information about the desirable direction for the respecification of the model and is considered in the next section.

3.4. Detecting the Source of Misspecification

3.4.1 Functional Form Misspecification

A known limitation of joint testing is that a significant joint test does not indicate the nature of the required respecification of the model. In order to test for functional form misspecification, one may proceed as follows: Consider the Weibull model, with the multiplicative heterogeneity term, V, in the density function:[6]

$$f(t|X,V) = V\mu\alpha t^{\alpha-1} \exp(-V\mu t^{\alpha}). \qquad (3.4.1)$$

Several authors (Lancaster (1979), Vaupel et al.(1979) etc.) have used the gamma distribution as a convenient mixing distribution for V.[7] If V has a gamma distribution with a unit mean and variance of 1/Θ, the density function conditional only on the observed X is:

$$f(t|X) = \mu\alpha t^{\alpha-1} \left[\frac{\Theta}{\Theta+\mu t^{\alpha}}\right]^{\Theta+1}. \qquad (3.4.2)$$

---

[6] Jaggia and Thosar (1989) argue that V should not be thought of only as a proxy for omitted regressors. More generally, V captures some intrinsic randomness in the model. The mixing distribution is used not only to compensate for some omitted regressors but also to correct for an overly restrictive individual hazard function.

[7] Hougaard (1984, 1986) shows that inverse gaussian distribution, like gamma, has all the desirable properties of a mixing distibution though it is not as widely used. He also suggests a more general distribution that specialises, among others, to gamma and inverse gaussian distributions.

The log-likelihood function, using (3.4.2), can be maximised in the usual way. To test for a functional form misspecification in the generalised gamma family with gamma heterogeneity, one may test for k=1. The heterogeneous generalised gamma distribution, given by (3.2.2), conditional on V is:

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

Given that V is gamma distributed with unit mean, the unconditional density is:

$$f(t|X) = \frac{\mu^k \alpha t^{\alpha k-1}\, \theta^\theta\, \Gamma(\theta+k)}{(\theta+\mu t^\alpha)^{\theta+k}\, \Gamma(k)\Gamma(\theta)} \tag{3.4.3}$$

and the log-likelihood function and the efficient scores are, respectively,

$$L = \Sigma\Big[k\log(\mu) + \log(\alpha) + (\alpha k-1)\log(t) + \theta\log(\theta) + \log\Gamma(\theta+k) - (\theta+k)\log(\theta+\mu t^\alpha) - \log\Gamma(k) - \log\Gamma(\theta)\Big] \tag{3.4.4}$$

and:

$$\left.\frac{dL}{dk}\right|_{H_0} = \Sigma\Big[\log(\mu) + \alpha\log(t) + \phi(\theta+1) - \phi(1) - \log(\theta+\mu t^\alpha)\Big]$$

$$= \Sigma[\log(\epsilon) + \phi(\theta+1) - \phi(1) - \log(\theta+\epsilon)]. \tag{3.4.5}$$

If an exponential model with a gamma heterogeneity distribution is estimated, the extra restriction, $\alpha=1$, can similarly be tested.

These tests of functional form misspecification are applied to the heterogeneous exponential model and the Weibull model, as shown in Table 3.3 and 3.4 respectively. The test statistics are based on the observed information matrix computed as the outer product of the sample scores. From the tables, it can be inferred that the hazard function is neither constant nor monotonic, which is consistent with Kennan's findings. The next logical step is to determine if there is neglected heterogeneity in the sample.

3.4.2 Neglected Heterogeneity

Given that even the Weibull hazard specification is not appropriate for the hazard function, one can estimate a generalised gamma model and test for neglected heterogeneity. Generalised gamma models are known to have convergence problems, especially when the parameter, k is large. The model has to be reparametrised to obtain the maximum likelihood estimates.[8] However, no convergence problems were encountered with the given strike data, even with the original parametrisation. A possible explanation for this result is the large sample size and the fact that

---

[8] see Lawless (1980), and the references therein, for an explanation of why such models do not converge and for possible modifications that would allow for their estimation.

the estimated value of k is small.

A score test of $\sigma^2=0$ may be based on (3.2.6) which does not depend upon any parametric representation of the heterogeneity distribution. Using (3.2.6), the appropriate score is:

$$\left.\frac{dL}{d\sigma^2}\right|_{H_0} = \Sigma \frac{1}{2} \left[k(k-1) - 2k\mu t^{\alpha} + (\mu t^{\alpha})^2 \right]. \qquad (3.4.6)$$

The score test of neglected heterogeneity in a generalised gamma model is also based on the observed information matrix, computed as the outer product of the sample scores. From Table 3.5, It can be concluded that there is evidence of unobserved heterogeneity in the sample. It is interesting to note that a partial test of heterogeneity detects misspecification only when a fairly general duration distribution is used in estimation.[9]

The results of the parameter estimates under alternate model specifications are presented in Table 3.6. 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 generalised gamma model, it is found that both α and k are not

---

[9] Convergence problems in estimation were encountered when a g.g.d. with gamma heterogeneity model was attempted.

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 maximised log-likelihood values of the null and the alternative models, suggests that the generalised gamma distribution is not an improvement over the Weibull or the exponential model. However, unlike the cases of the exponential and Weibull models, when a generalised 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.

When a generalised 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 Kennan.

3.5 Conditional Moment Tests

3.5.1 Interpretation

As seen earlier, the test of heterogeneity is based on the generalised residual, $\tilde{\epsilon}$. Since ϵ has a unit exponential distribution when the model is correctly specified, testing its second moment restriction has been the basis of tests of heterogeneity. When all observations are complete, the quantity that is equated with zero is:

$$1/2N\ \Sigma[\tilde{\epsilon}^2 - 2\tilde{\epsilon}]$$

$$= 1/2N\ \Sigma[\tilde{\epsilon}^2 - 2], \text{ as } \Sigma\tilde{\epsilon}/N = 1 \text{ at the ML estimates.}$$

$$= 1/2\ (s^2 - 1)$$

where s^2 is the sample variance of the generalised residual. Thus, the score test or White's information test of no heterogeneity amounts to testing the second moment restriction of ϵ, namely that $Var(\epsilon) = E(\epsilon - 1)^2 = 1$.

However, one cannot test for the second moment restriction in isolation as neglected heterogeneity may not be the only source of misspecification in the model. Therefore, it is necessary to test for the second order moment restriction along with the higher order moment restrictions of ϵ in order to test the specification of a model. Alternatively, it is noted that there is no unique way of defining a generalised error. Any non-linear transformation of the integrated hazard function can be interpreted as a generalised error in the sense of Cox and Snell. Testing moment restrictions of all such errors can be used to check model adequacy. For example, if ϵ is defined as the integrated error, then $\epsilon_1 = \log(\epsilon)$ has a standard extreme value distribution with well defined moments, such as $E(\epsilon_1) = \phi(1) = -.5772$ and $Var(\epsilon_1) = \phi'(1) = \pi^2/6$. Similarly, one can define $\epsilon_2 = \epsilon_1 \log(t)$ as yet another generalised error, and so on. There is an infinite number of moment restrictions that can be thus evaluated. However, testing moment restrictions of all such errors will be meaningful only if an interpretation can be provided for the situation in which a particular moment restriction is not satisfied, given that all the other restrictions are met by the given data.

From (3.2.8), one can see that the score test of k=1 is equivalent to testing for the first moment restriction of $\epsilon_1 = \log(\epsilon)$. If all relevant moment restrictions can be thus identified, an omnibus test can be implemented using the conditional moment restrictions framework studied by Tauchen (1985) and Newey (1985) and, more recently, White (1987).[10] With the Tauchen and Newey framework, the asymptotic

---

[10] Pagan and Vella (1989) give a good exposition of such tests and their applications to cross section data.

distribution of all such moment restrictions can be derived in a general setup and can be used to construct a joint test to check their validity.

3.5.2 The Tauchen-Newey Framework

Tauchen-Newey tests are based on the auxiliary criterion function, $m(y,\theta)$ that has a mean of zero with the specified probability distribution under the null hypothesis. This is similar to specifying a moment restriction that is expected to be satisfied if the model is correctly specified. The test is based on the magnitude of the following statistic:

$$\tilde{\tau} = \frac{1}{N} \Sigma\, m(y,\tilde{\theta}) \qquad (3.5.1)$$

where (~) denotes that the quantities have been evaluated at the maximum likelihood estimates of θ. m is a (qx1) vector implying q moment restrictions. The statistic $\tilde{\tau}$ is useful for testing the specification of the model as it converges almost surely to zero when the model is correctly specified and to a non-zero quantity when the specification is incorrect (Tauchen (1985)). Intuitively, when the model is correctly specified, the maximum likelihood estimate of θ converges to the true parameter value θ_0, and $\tilde{\tau}$ converges to the expectation of the auxiliary function which by construction is zero. Under the alternative hypothesis, the estimated θ converges to the quantity $\theta^*$and the expectation of $m(y,\theta^*)$ under the alternative probability model, in general, is non-zero.

The test is based on the statistic $\tilde{\tau}$ that has a well defined asymptotic distribution under the null. Using the notation of Pagan and Vella (1989)[11], the asymptotic distribution of estimated τ can be specified as:

$$N^{\frac{1}{2}} \tilde{\tau} \quad \sim \quad n(0,\ AVA') \qquad (3.5.2)$$

where:

$$A = [I_{qq} \quad JH^{-1}]. \qquad (3.5.3)$$

I_{qq} is a (qxq) identity matrix, where q represents the number of conditional moments being tested.

J is a (qxk) matrix = $E(\delta m/\delta\theta)$.

H is a (kxk) matrix = $-E(\delta^2 l_i/\delta\theta\delta\theta')$.

$$V = \begin{bmatrix} V_{mm} & V_{md} \\ V_{dm} & V_{dd} \end{bmatrix}.$$

V_{mm} is a (qxq) matrix = $E(mm')$,

---

[11] see Tauchen (1985), for example, for the regularity conditions.

V_{md} is a (qxk) matrix = E(md'),

V_{dd} is a (kxk) matrix = E(dd'), where
$d = \delta l_i/\delta\theta$ is the score vector.

Therefore,

$$AVA' = V_{mm} + JH^{-1}V_{dm} + V_{md}H^{-1}J' + JH^{-1}J'. \qquad (3.5.4)$$

Moreover, the generalized information equality holds under the null hypothesis, implying that $J = -V_{md}$. This further simplification results in:

$$AVA' = V_{mm} - V_{md}H^{-1}V_{dm}. \qquad (3.5.5)$$

The conditional moment test can, thus, be implemented as:

$$N\ \tilde{\tau}'[V_{mm} - V_{md}H^{-1}V_{dm}]^{-1}\ \tilde{\tau}. \qquad (3.5.6)$$

Furthermore, if the expectations cannot be obtained analytically, sample moments can be used in place of the population moments to implement the test. The quantities in the variance of the relevant statistic can be estimated as:

$$V_{mm} = \frac{1}{N}\ \Sigma\ mm' = \frac{1}{N}\ M'M.$$

$$V_{md} = \frac{1}{N}\ \Sigma\ md' = \frac{1}{N}\ M'D.$$

$$H = -\frac{1}{N}\ \Sigma(\delta^2 l/\delta\theta\delta\theta') = \frac{1}{N}\ \Sigma\ dd' = \frac{1}{N}\ D'D.$$

Here M is an (Nxq) matrix specifying N sample observations of a (qx1) vector with q moment restrictions. Similarly, D is an (nxk) matrix specifying N sample observations of a (kx1) score vector. With these quantities, the test is computed as:

$$N^2\ \tilde{\tau}'\ [M'M - M'D(D'D)^{-1}D'M]^{-1}\ \tilde{\tau}.$$

$$= i'M\ [M'M - M'D(D'D)^{-1}\ D'M]^{-1}\ M'i \qquad (3.5.7)$$

where i is a (Nx1) vector of units.

This rendition of the test can easily be implemented by running an artificial linear regression. Consider a regression of i on Z where Z = [D M]. The NR^2 derived from this artificial regression, where R^2 is the uncentered coefficient of determination, is given by:

$$i'Z(Z'Z)^{-1}Z'i.$$

As $i'D = \delta L/\delta\theta = 0$ by the first order restriction of the maximum likelihood estimation, this term is equal to:

$$[0\ \ i'M] \begin{bmatrix} D'D & D'M \\ M'D & M'M \end{bmatrix}^{-1} [0\ \ i'M]'$$

$$= i'M\ [M'M - M'D(D'D)^{-1}D'M]^{-1}\ M'i$$

which is the same as (3.5.7). Therefore, the Tauchen-Newey type moment restriction test can easily be implemented by computing the NR^2 from the above mentioned artificial regression.

Another readily implementable version of the conditional moment test also exists. This test consists of running an OLS where the lhs variable is simply the difference between the theoretical and the predicted moment from the probability model. The rhs variables comprise a constant and all the scores of the model. Testing if a particular moment restriction is satisfied is equivalent to performing a t test for a non-zero intercept. A major advantage of this procedure is that it gives the user detailed information on the statistical significance of each moment restriction separately, instead of the joint significance of all moment restrictions (Tauchen (1985)). This information, however, cannot be useful when moment restrictions are correlated such that V_{mm} is not block-diagonal. The validity of any particular moment restriction will depend on the validity of the other if the two are correlated. Thus, when V_{mm} is not block-diagonal, the first step taken should be the implementation of a joint test which is achieved by testing for a non-zero intercept in a SUR regression model.

3.6 Moment Restriction Tests for Duration Models

3.6.1 Test Against a Parametric Alternative

In order to test for both neglected heterogeneity and functional form misspecification within a generalised gamma distribution, the set of moment restrictions for a Weibull model are:

$$m(y,\theta) = \begin{bmatrix} m_1(y,\theta) \\ m_2(y,\theta) \end{bmatrix} = \begin{bmatrix} \epsilon^2 - 2 \\ \epsilon_1 - \phi(1) \end{bmatrix}. \qquad (3.6.1)$$

Using the notation defined above, the test of the moment restrictions, specified above, can be implemented as:

$$N\ \tilde{\tau}'\ [V_{mm} - V_{md}H^{-1}V_{dm}]^{-1}\ \tilde{\tau}.$$

where $\tilde{\tau} = (1/N)\ \Sigma m(y,\tilde{\theta})$, and (see Appendix 3B)

$$V_{mm} - V_{md}H^{-1}V_{dm} = \begin{bmatrix} 4-4q & -1+2q \\ -1+2q & \phi'(2)-q \end{bmatrix}. \qquad (3.6.2)$$

Note that this test is the same as the score test developed within the generalised gamma model.

3.6.2 Test based on an Unspecified Alternative

Tests based on specified alternatives have a limitation in that they restrict the alternative and hence may not have good properties when the specified alternative is also incorrect. For instance, if the restricted alternative is a heterogeneous Weibull model, the heterogeneity test will not be valid test if the restricted alternative is incorrect. Here, it is suggested that a joint test of all moment restrictions of ϵ should be used rather than a test of the second moment only. Such a test has the merit of being based on an unspecified alternative and thus is not restrictive in any way. Given any parametric model, the integrated hazard function has a unit exponential distribution under the null, and hence its moment restrictions must be satisfied if the model is correctly specified. Here it should be pointed out that there is no statistical justification why the higher order moment restrictions of ϵ should be preferred to that of any other choice of a generalised error. Furthermore, higher order moments of ϵ may not be estimated efficiently from a given sample. However, it is argued that if the first few moment restrictions are satisfied, the higher order moment restrictions are expected to hold as well. One can resort to higher orders if the sample size is sufficiently large.

Kiefer (1985) and Sharma (1989) develop score tests for the exponential and Weibull models respectively. The tests they derive are based on Laguerre polynomials and amount to testing for moment restrictions of higher orders of and functions of generalised residuals, ϵ. Here, a more transparent version of these tests is given that consists of computing the integrated hazard function, ϵ, and testing its moment restrictions using the Tauchen-Newey framework. The advantage of this form is that it can be used for any parametric distribution, not restricting to the exponential and Weibull distributions only.

In order to test for the first four moment restriction of ϵ, the auxiliary criterion function is:

$$m(y,\Theta) = \begin{bmatrix} m_1(y,\Theta) \\ m_2(y,\Theta) \\ m_3(y,\Theta) \end{bmatrix} = \begin{bmatrix} \epsilon^2 - 2 \\ \epsilon^3 - 6 \\ \epsilon^4 - 24 \end{bmatrix}. \qquad (3.6.3)$$

Therefore, in order to implement the test the following can be derived (see Appendix 3B):

$$V_{mm} - V_{md}H^{-1}V_{dm} = \begin{bmatrix} 4-4q & 36-30q & 288-208q \\ 36-30q & 360-225q & 3168-1560q \\ 288-208q & 3168-1560q & 30528-10816q \end{bmatrix}. \qquad (3.6.4)$$

Tests for an exponential specification can similarly be derived. In order to test for the same moment restrictions of $\epsilon=\mu t$, for an exponential specification, the variance-covariance matrix of the moment restriction is:

$$V_{mm} - V_{md}H^{-1}V_{dm} = \begin{bmatrix} 4 & 36 & 288 \\ 36 & 360 & 3168 \\ 288 & 3168 & 30528 \end{bmatrix}. \qquad (3.6.5)$$

Tests of moment restrictions of various orders of ϵ are applied to Kennan's Strike data (see Table 3.6 and 3.7). It is inferred that even though the tests for lower order moment restrictions do not imply any misspecification, results change when higher order moment tests are implemented. For example, when the second, third and fourth moment restrictions are tested jointly, the test results in the rejection of a Weibull specification.

All the above mentioned tests are derived by taking expectations of the relevant quantities under the null. As mentioned earlier, if it is not possible to take expectations analytically, sample moments can be used to derive the test. This method has the added advantage of computational ease. Such a test is especially appealing when the data consist of censored observations and it is not possible to obtain expectations without specifying the censoring mechanism. However, since in the given sample all observations are complete, tests based on the sample moments are not carried out.

3.7 Conclusion

In this chapter it has been shown that a 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 generalised 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 is presented which provides evidence that incorrect inferences can be drawn due to the use of partial 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.

Moreover, tests of misspecification are interpreted as being tests of conditional moment restrictions. Conditional Moment restriction tests, developed by Tauchen and Newey, are discussed and such tests are developed in the context of exponential and Weibull models. Again, it is inferred that misspecification is detected only when the joint test is implemented on the higher order moment restrictions of ϵ.

TABLE 3.1

SCORE TEST RESULTS FOR AN EXPONENTIAL MODEL

Restrictions	Test Statistic	p-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 3.2

SCORE TEST RESULTS FOR A WEIBULL MODEL

Restrictions	Test Statistic	p-Value
$\sigma^2 = 0$	0.014199	0.9051
$k = 1$	2.009156	0.1564
$\sigma^2=0,\ k=1$	11.870698	0.0026

TABLE 3.3

SCORE TEST RESULTS FOR AN EXPONENTIAL-GAMMA MODEL

Restrictions	Test Statistic	p-Value
$\alpha = 1$	0.353771	0.5520
$k = 1$	0.010174	0.9197
$\alpha=1,\ k=1$	47.390425	0.0000

TABLE 3.4

SCORE TEST RESULTS FOR A WEIBULL-GAMMA MODEL

Restrictions	Test Statistic	p-Value
$k = 1$	51.398681	0.0000

TABLE 3.5

SCORE TEST RESULTS FOR A GENERALSIED GAMMA MODEL

Restrictions	Test Statistic	p-Value
$\sigma^2 = 0$	60.056180	0.0000

TABLE 3.6

PARAMETER ESTIMATES UNDER ALTERNATIVE MODEL SPECIFICATIONS

Variable	Exponential	Weibull	Generalised 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			1.7125 (0.7975)
Log-Lik. Value	-2698.420	-2698.20	-2696.619

* Standard Errors in parenthesis.

TABLE 3.7

MOMENT RESTRICTIONS TEST RESULTS FOR AN EXPONENTIAL MODEL

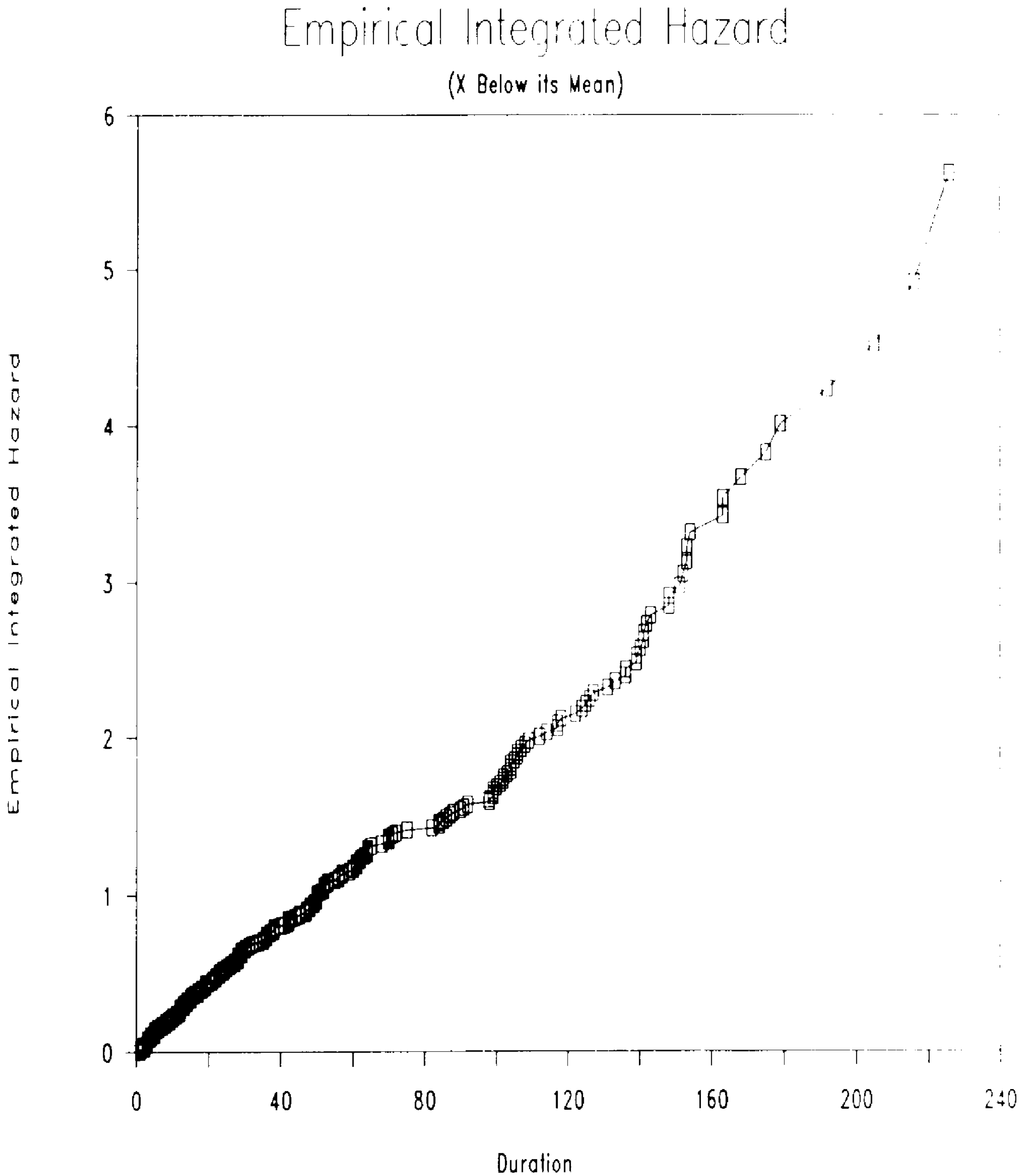
Restrictions	Test Statistic	p-Value
$\epsilon^2=2$	0.156109	0.6928
$\epsilon^2=2$, $\epsilon^3=6$	4.724089	0.0942
$\epsilon^2=2$, $\epsilon^3=6$, $\epsilon^4=24$	7.388818	0.0605

TABLE 3.8

MOMENT RESTRICTIONS TEST RESULTS FOR A WEIBULL MODEL

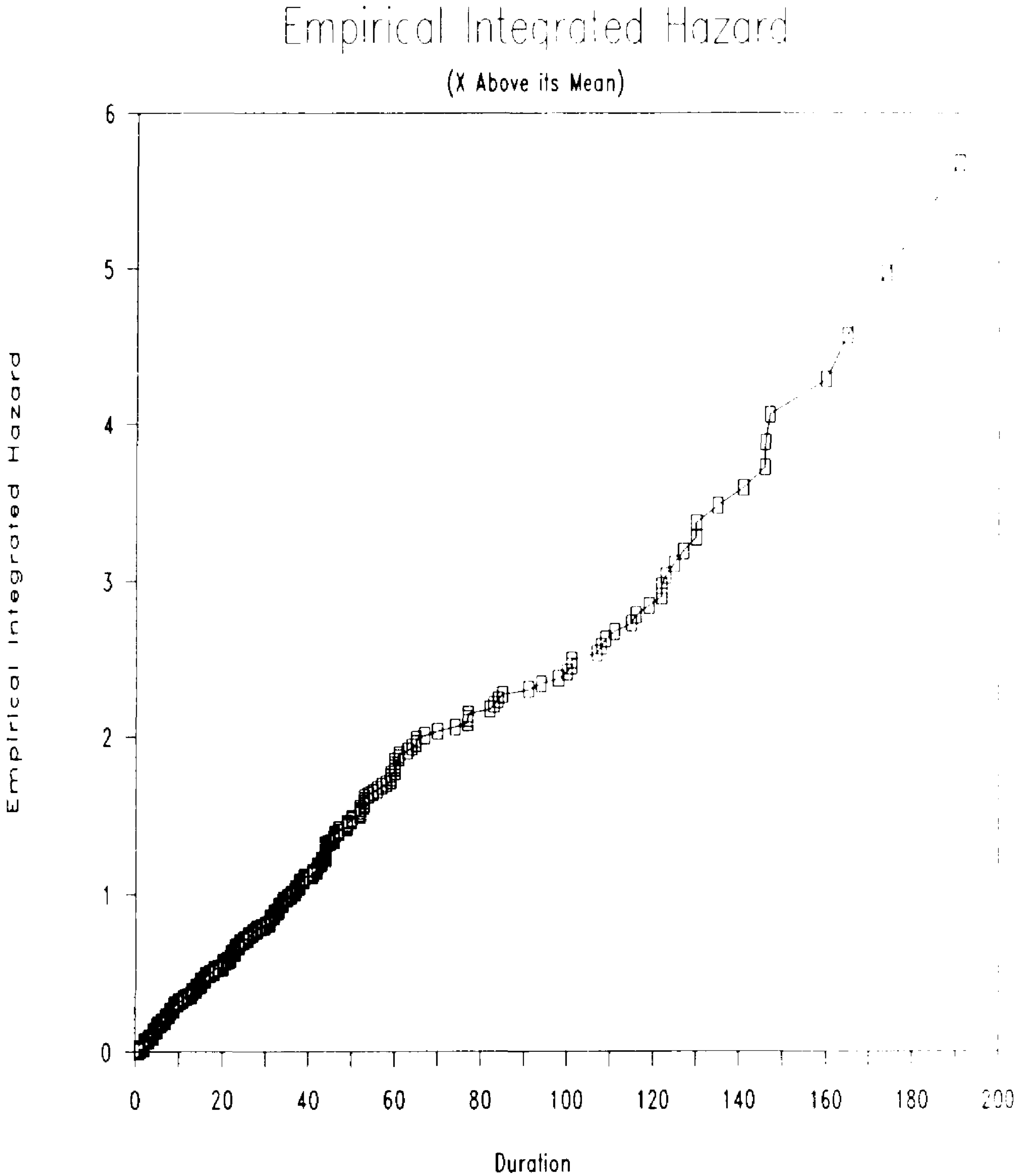
Restrictions	Test Statistic	p-Value
$\epsilon^2=2$	0.014199	0.9051
$\epsilon^2=2$, $\epsilon^3=6$	4.488741	0.1060
$\epsilon^2=2$, $\epsilon^3=6$, $\epsilon^4=24$	10.727937	0.0133

FIGURE 3.1



Empirical Integrated Hazard
(X Below its Mean)
Empirical Integrated Hazard
Duration
0
1
2
3
4
5
6
0
40
80
120
160
200
240

FIGURE 3.2



Empirical Integrated Hazard
(X Above its Mean)
Empirical Integrated Hazard
0
1
2
3
4
5
6
0
20
40
60
80
100
120
140
160
180
200
Duration

FIGURE 3.3

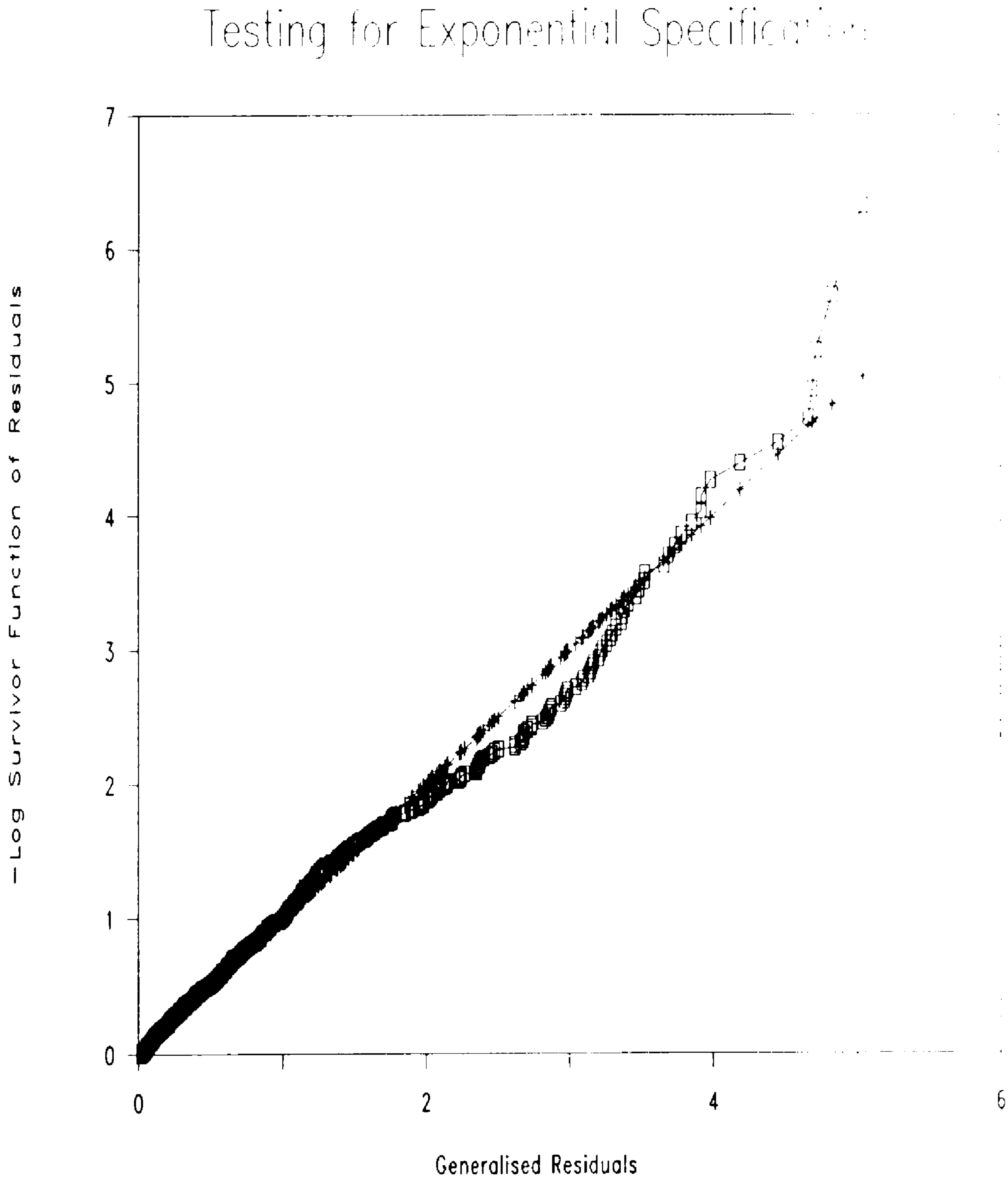

Testing for Exponential Specific
7
6
5
4
3
2
1
0
0
2
4
6
-Log Survivor Function of Residuals
Generalised Residuals

FIGURE 3.4

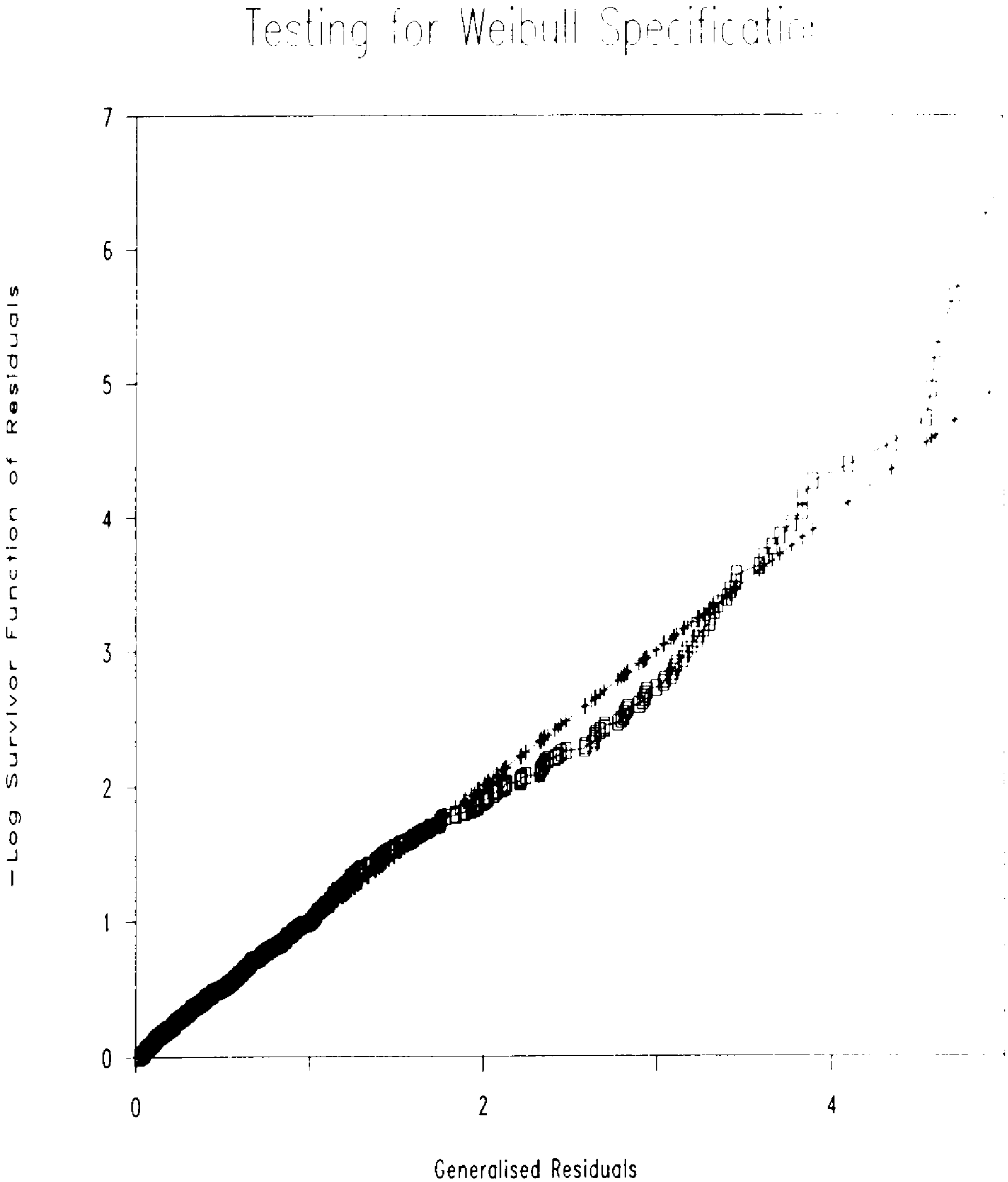
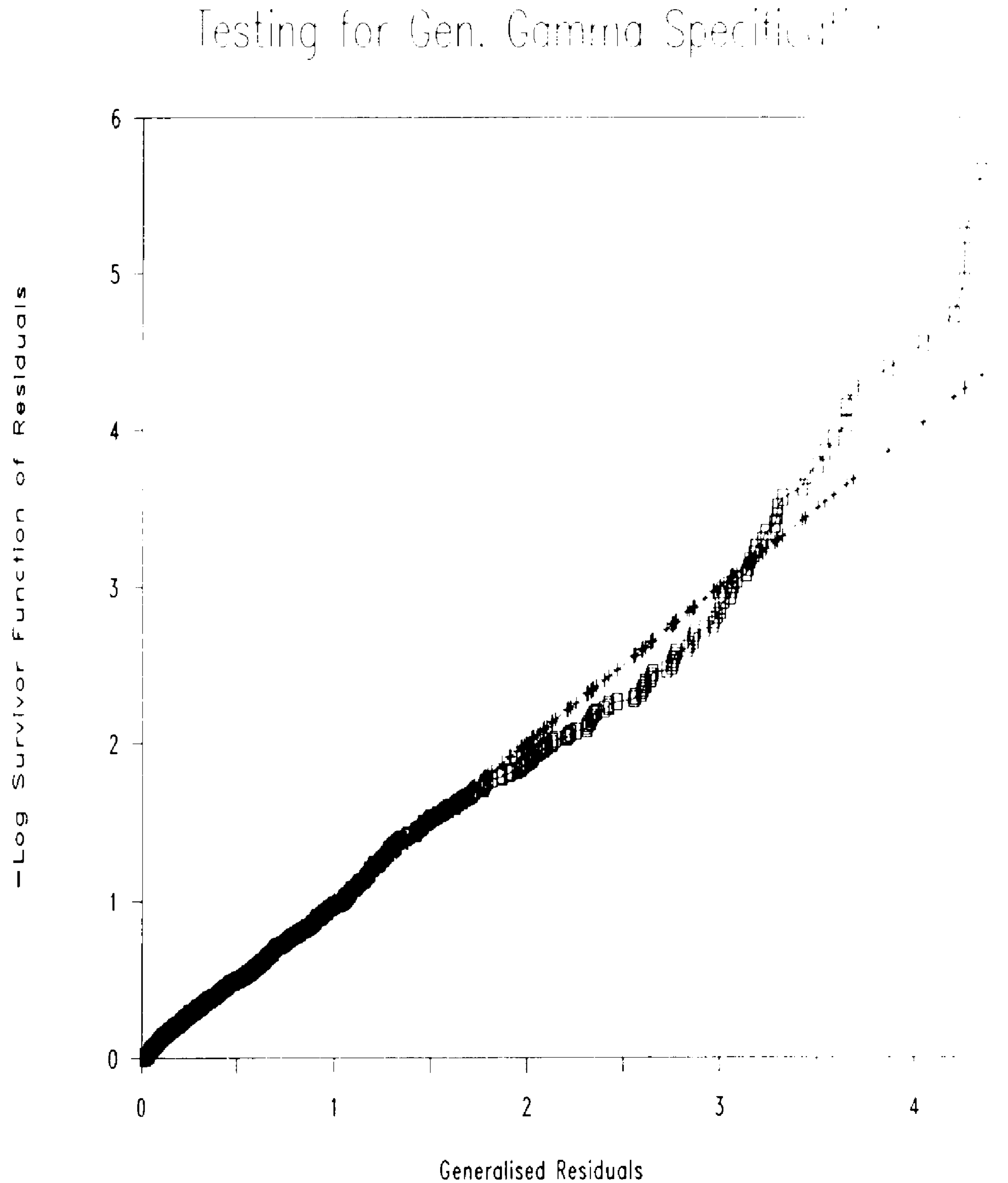

Testing for Weibull Specificatio
-Log Survivor Function of Residuals
7
6
5
4
3
2
1
0
0
2
4
Generalised Residuals

FIGURE 3.5


Testing for Gen. Gamma Specifi...
6
5
4
3
2
1
0
0
1
2
3
4
-Log Survivor Function of Residuals
Generalised Residuals

Appendix 3A

In this appendix, the theoretical information matrix required to implement the score tests for a Weibull specification is derived. The likelihood function, given by (3.2.6), based on the heterogeneous generalised gamma function is:

$$L = \sum_{i=1}^{N} l_i \quad \text{where}$$

$$l_i = k\log(\mu_i) + \log(\alpha) + (\alpha k - 1)\log(t_i) - \epsilon_i - \log\Gamma(k) + \log[1 + (\sigma^2/2)\{k(k-1) - 2k\epsilon_i + (\epsilon_i)^2\}]. \qquad (3A.1)$$

Here $\epsilon = \mu t^\alpha$ and $\mu = \exp(\beta_o + X_1\beta_1)$. It is assumed, without loss of generality, that $E(X_1) = 0$ and $E(X_1X_1') = \Omega$ where Ω is a is non-singular (kxk) matrix. Therefore:

$E(X) = [1 \quad 0_k]$ and:

$$E(XX') = \begin{bmatrix} 1 & 0 \\ 0 & \Omega \end{bmatrix}. \qquad (3A.2)$$

The information matrix under the null can be derived using the fact that when the model is correctly specified, ϵ has a unit exponential distribution. This result implies that:

$$E(\epsilon^j) = j!. \qquad (3A.3)$$

Furthermore, the following results can easily be derived:

$$E[\log(\epsilon)\epsilon^{r-1} / \Gamma(r)] = \phi(r) \qquad (3A.4)$$

$$E[(\log(\epsilon))^2\epsilon^{r-1} / \Gamma(r)] = \phi'(r) + (\phi'(r))^2 \tag{3A.5}$$

$$\phi(z+1) = \phi(z) + 1/z \tag{3A.6}$$

$$\phi'(z+1) = \phi'(z) - 1/z^2 \tag{3A.7}$$

where for $r > 0$, $\phi(r)$ is the digamma function $\delta\log\Gamma(r)/\delta r$ and $\phi'(r)$ is the trigamma function $\delta^2\log\Gamma(r)/\delta r^2$ (see Lawless (1982)).

Using (3A.2)-(3A.7), the components of the information matrix, evaluated at $\sigma^2=0$ and $k=1$, can be derived as:

$$-E(\delta^2 l_i/\delta\sigma^2\delta\sigma^2) = (1/4)E(4\epsilon^2 + \epsilon^4 - 4\epsilon^3) = 2. \tag{3A.8}$$

$$-E(\delta^2 l_i/\delta\sigma^2\delta k) = (-1/2)E(1 - 2\epsilon) = 1/2. \tag{3A.9}$$

$$-E(\delta^2 l_i/\delta\sigma^2\delta\beta') = E(\epsilon - \epsilon^2)\ E(X') = [-1 \quad 0_k]. \tag{3A.10}$$

$$-E(\delta^2 l_i/\delta\sigma^2\delta\alpha) = E((\epsilon - \epsilon^2)\log(t))$$

$$= (1/\alpha)\ (\beta_o - \phi(2) - 1). \tag{3A.11}$$

$$-E(\delta^2 l_i/\delta k^2) = E(\phi'(1)) = \phi'(1). \tag{3A.12}$$

$$-E(\delta^2 l_i/\delta k\delta\beta') = E(-X') = [-1 \quad 0_k]. \tag{3A.13}$$

$$-E(\delta^2 l_i/\delta k\delta\alpha) = E(-\log(t)) = (1/\alpha)(\beta_o - \phi(1)). \tag{3A.14}$$

$$-E(\delta^2 l_i/\delta\alpha^2) = (1/\alpha^2) + E(\epsilon\log(t)^2)$$

$$= (1/\alpha^2)[1 + \phi'(2) + \beta_1'\Omega\beta_1 + (\beta_o - \phi(2))^2]. \tag{3A.15}$$

$$-E(\delta^2 l_i/\delta\alpha\delta\beta') = E(\epsilon\log(t))E(X')$$

$$= (1/\alpha)[\phi(2)-\beta_o \quad -\beta_1'\Omega]. \tag{3A.16}$$

$$-E(\delta^2 l_i/\delta\beta\delta\beta') = E(\epsilon)E(XX') = \begin{bmatrix} 1 & 0 \\ 0 & \Omega \end{bmatrix}. \tag{3A.17}$$

Using (3A.8) - (3A.17), the components of the information matrix, with $\theta_1' = (\sigma^2 \ k)$ and $\theta_2' = (\beta_o \ \beta_1' \ \alpha)$, can be derived as follows:

$$\frac{1}{N} I_{11} = \begin{bmatrix} 2 & 1/2 \\ 1/2 & \phi'(1) \end{bmatrix}. \tag{3A.18}$$

$$\frac{1}{N} I_{12} = \begin{bmatrix} -1 & 0 & -1/\alpha(m+1) \\ -1 & 0 & -1/\alpha(m+1) \end{bmatrix} \tag{3A.19}$$

where $m = \phi(2)-\beta_o$ and N is the sample size.

Also, the inverse of I_{22} can be found as:

$$\frac{1}{N}(I_{22})^{-1} = \begin{bmatrix} 1+m^2 q & -mq\beta_1' & -\alpha mq \\ -mq\beta_1 & \Omega^{-1}+q\beta_1\beta_1' & \alpha q\beta_1 \\ -\alpha mq & \alpha q\beta_1' & \alpha^2 q \end{bmatrix} \tag{3A.20}$$

where $q = 1/\phi'(1)$.

Therefore, the partitioned inverse needed to compute the score test can be found as:

$$[I^{11}]^{-1} = N \begin{bmatrix} 1-q & q-1/2 \\ q-1/2 & \phi'(2)-q \end{bmatrix}. \tag{3A.21}$$

The information matrix needed to compute the score tests for an exponential specification can similarly be derived.

Appendix 3B

In this appendix, conditional moment restriction tests are derived where the variances of the moment restrictions are based on their expected values evaluated under the Weibull specification. The results specified in (3A.2) to (3A.7) are again used to derive these expressions. Let the set of moment restrictions be:

$$m(y,\theta) = \begin{bmatrix} m_1(y,\theta) \\ m_2(y,\theta) \end{bmatrix} = \begin{bmatrix} \epsilon^2 - 2 \\ \epsilon_1 - \phi(1) \end{bmatrix}.$$

where $\epsilon = \mu t^{\alpha}$ and $\epsilon_1 = \log(\epsilon)$. Using the notation defined in Section 5, the relevant components of the variance-covariance matrix can be derived as follows:

$$\mathrm{Var}(m_1) = E(\epsilon^4 + 4 - 4\epsilon^2) = 20. \tag{3B.1}$$

$$\mathrm{Var}(m_2) = E(\ln(\epsilon))^2 - (\phi(1))^2 = \phi'(1). \tag{3B.2}$$

$$\mathrm{cov}(m_1 m_2) = E(\epsilon^2 \ln(\epsilon)) - 2\phi(1) = 3. \tag{3B.3}$$

Therefore,

$$V_{mm} = \begin{bmatrix} 20 & 3 \\ 3 & \phi'(1) \end{bmatrix}. \tag{3B.4}$$

Also:

$$d = \begin{bmatrix} \delta l_i/\delta \beta_o \\ \delta l_i/\delta \beta_1 \\ \delta l_i/\delta \alpha \end{bmatrix} = \begin{bmatrix} (1-\epsilon) \\ (1-\epsilon)X_1 \\ (1-\epsilon)\ln(t) + 1/\alpha \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}. \tag{3B.5}$$

Using the above, the following can be derived:

$$E(m_1d_1) = E((\epsilon^2-2)(1-\epsilon)) = -4. \tag{3B.6}$$

$$E(m_1d_2) = E((\epsilon^2-2)(1-\epsilon)) \; E(X_1) = 0_k. \tag{3B.7}$$

$$E(m_1d_3) = E((\epsilon^2-2)(1-\epsilon)\ln(t)) = -(4/\alpha)(m+1/2). \tag{3B.8}$$

$$E(m_2d_1) = E((\ln(\epsilon)-\phi(1))(1-\epsilon)) = -1. \tag{3B.9}$$

$$E(m_2d_2) = E((\ln(\epsilon)-\phi(1))(1-\epsilon))E(X_1) = 0_k. \tag{3B.10}$$

$$E(m_2d_3) = E((\ln(\epsilon)-\phi(1))(1-\epsilon)\ln(t)) = (1/\alpha)(1-m). \tag{3B.11}$$

These expressions are used to derive:

$$V_{md} = \begin{bmatrix} -4 & 0 & -\frac{4}{\alpha}(m+.5) \\ -1 & 0 & -\frac{1}{\alpha}(m-1) \end{bmatrix}. \tag{3B.12}$$

Moreover, the inverse of $H = -E(\delta^2 l_i/\delta\theta\delta\theta')$ is:

$$H^{-1} = \begin{bmatrix} 1+m^2 q & -mq\beta_1' & -\alpha mq \\ -mq\beta_1 & \Omega^{-1}+q\beta_1\beta_1' & \alpha q\beta_1 \\ -\alpha mq & \alpha q\beta_1' & \alpha^2 q \end{bmatrix}. \tag{3B.13}$$

The test of the above specified moment restrictions can be implemented as:

$$N\ \tilde{\tau}'[V_{mm} - V_{md}H^{-1}V_{dm}]^{-1}\ \tilde{\tau}.$$

where $\tilde{\tau} = (1/N)\ \Sigma m(y,\tilde{\theta})$, and:

$$V_{mm} - V_{md}H^{-1}V_{dm} = \begin{bmatrix} 4-4q & -1+2q \\ -1+2q & \phi'(1)-1-q \end{bmatrix}. \qquad (3B.14)$$

Tests of higher order moments of ϵ that are based on an unspecified alternative can similarly be derived.

$$\text{Let } m(y,\theta) = \begin{bmatrix} m_1(y,\theta) \\ m_2(y,\theta) \\ m_3(y,\theta) \end{bmatrix} = \begin{bmatrix} \epsilon^2 - 2 \\ \epsilon^3 - 6 \\ \epsilon^4 - 24 \end{bmatrix}.$$

The relevant components of the variance-covariance matrix are:

$$Var(m_2) = E(\epsilon^6 + 36 - 12\epsilon^3) = 684. \qquad (3B.15)$$

$$Var(m_3) = E(\epsilon^8 + 576 - 48\epsilon^4) = 39744. \qquad (3B.16)$$

$$Cov(m_1m_2) = E(\epsilon^5 - 6\epsilon^2 - 2\epsilon^3 + 12) = 108. \qquad (3B.17)$$

$$Cov(m_1m_3) = E(\epsilon^6 - 24\epsilon^2 - 2\epsilon^4 + 48) = 672. \qquad (3B.18)$$

$$Cov(m_2m_3) = E(\epsilon^7 - 24\epsilon^3 - 6\epsilon^4 + 144) = 4896. \qquad (3B.19)$$

$$E(m_2d_1) = E((\epsilon^3-6)(1-\epsilon)) = -18. \tag{3B.20}$$

$$E(m_2d_2) = E((\epsilon^3-6)(1-\epsilon))E(X_1) = 0_k. \tag{3B.21}$$

$$E(m_2d_3) = E((\epsilon^3-6)(1-\epsilon)\ln(t)) = -(18/\alpha)(m+5/6). \tag{3B.22}$$

$$E(m_3d_1) = E((\epsilon^4-24)(1-\epsilon)) = -96. \tag{3B.23}$$

$$E(m_3d_2) = E((\epsilon^4-24)(1-\epsilon))E(X_1) = 0_k. \tag{3B.24}$$

$$E(m_3d_3) = E((\epsilon^4-24)(1-\epsilon)\ln(t)) = -(96/\alpha)(m+104/96). \tag{3B.25}$$

These expressions, together with the ones derived earlier, can be used to compute the following:

$$V_{mm} = \begin{bmatrix} 20 & 108 & 672 \\ 108 & 684 & 4896 \\ 672 & 4896 & 38744 \end{bmatrix}. \tag{3B.26}$$

$$V_{md} = \begin{bmatrix} -4 & 0 & -\frac{4}{\alpha}(m+1/2) \\ -18 & 0 & -\frac{18}{\alpha}(m+5/6) \\ -96 & 0 & -\frac{96}{\alpha}(m+104/96) \end{bmatrix}. \tag{3B.27}$$

Therefore, to implement the test the following can be derived:

$$V_{mm} - V_{md}H^{-1}V_{dm} = \begin{bmatrix} 4-4q & 36-30q & 288-208q \\ 36-30q & 360-225q & 3168-1560q \\ 288-208q & 3168-1560q & 30528-10816q \end{bmatrix}. \quad (3B.28)$$

Conditional moment tests for an exponential model can similarly be derived.

CHAPTER 4

ADJUSTED SCORE TESTS FOR DURATION DEPENDENCE AND UNOBSERVED HETEROGENEITY

4.1 Introduction

The validity of separate separate tests of a particular parametric restriction depends on the validity of additional restrictions not being tested at that step. As a result, this chapter is devoted to a discussion of testing for sources of misspecification individually when several such sources exist concurrently. An alternative method to separate tests, is to start with a joint test, of all relevant restrictions, that has power aginst several forms of misspecification.[1] The non-centrality parameter of such a test will be at least as large as that of any separate test. However, the asymptotic power of a joint test may be lower due to higher degrees of freedom. Moreover, simply rejecting the joint null offers no information to respecify the model appropriately. For this reason, some additional information from separate tests is needed.

The standard separate tests which have been proposed to date may be unhelpful in this regard when these tests are asymptotically correlated with each other. In this chapter,

[1] see Chapter 3 of this thesis.

adjusted separate tests are proposed that may be used to gain some additional insight regarding the source of misspecification. Three types of score-based tests for testing the hypotheses of no neglected heterogeneity and no duration dependence are comparared in the context of the heterogeneous Weibull model. These tests are the separate and joint tests for two parametric hypotheses, and the conditional or adjusted score test for testing a single separate hypothesis.

Let $L(\theta)$ be the likelihood function where $\theta = (\theta_1, \theta_2, \theta_3)'$ is the set of parameters. Without loss of generality, let H_{o1}: $\theta_1 = 0$ and H_{o2}: $\theta_2 = 0$ be the two separate hypotheses and H_{o12}: $\theta_1 = \theta_2 = 0$ be the joint hypothesis. Further, suppose that τ_1, τ_2 and τ_{12} are score tests for H_{o1}, H_{o2} and H_{o12} respectively. The objective is to carry out a specification search for selecting the most appropriate of the models considered.

The three possible approaches are examined as below. In the first, H_{o1} or H_{o2} are tested for separately, assuming the truth of the other. The model is generalised if the test is significant.[2] A difficulty with this approach arises when τ_1 and τ_2 are stochastically dependent. Each separate test

[2] Wooldridge (1989b) refers to such a procedure as the "bottom-up" approach.

will in this case offer no information regarding the particular misspecification being tested for. As a result, the separate tests should only be claimed as general misspecification tests and not directed at any particular misspecification. However, under this interpretation a significant test can not suggest direction for respecification. Further, this approach may lead to a model which is either over- or under-parametrised, a common presumption being that over-parametrisation is more likely.

A second approach is to test the joint null and hence, the possible correlation between the separate tests is dealt with directly. A more general model may be inferred if τ_{12} is significant. However, if only a subset of the joint hypothesis is false, this approach will lead to over-parametrisation. Further, as shown by an example later in the chapter, the joint test may have low power for some parameter configurations against certain local alternatives.

A third approach is to test each of the two separate hypotheses without assuming the truth of the complementary hypothesis. That is, separate tests are constructed which allow for the dependence of the test on nuisance parameters. Two methods can be used in order to implement this test. Firstly, a more general model than implied by the joint hypothesis can be fitted as in the least squares regression based Neyman's $C(\alpha)$ test in Section 4.5 or, as in Section 4.4 of the chapter, the scores estimated under the joint null can be "adjusted" for the nuisance parameter. In each case the score test will be based on quantities that are refered to as conditional or adjusted scores. This approach is asymptotically equivalent to separate tests when the joint null is true. Further, the example of the heterogeneous Weibull model considered in this chapter shows that the adjustments to the separate tests are easy to compute and lead to better tests than the other two approaches.

The rest of the chapter is organised as follows. In Section 4.2, a simple but general exposition of variants of the conditional score approach is provided and contrasted with the standard approach. A joint score test for duration dependence and neglected heterogeneity is derived within a heterogeneous Weibull model and is contained in Section 4.3. The non-null distribution and the conditions under which the test has a low power are considered. The analysis is further extended to derive the tests within a heterogeneous generalised gamma distribution. In Section 4.4 the approach of Section 4.2 specialised and used to develop conditional score tests for heterogeneity and duration dependence based on the restricted maximum likelihood estimates derived under the joint null hypothesis. In Section 4.5, the robust conditional test for heterogeneity is suggested that uses a variant based on least squares estimates of the Weibull regression model. This latter approach is essentially an application of the Neyman's $C(\alpha)$ principle. Rather than restricting nuisance parameters to their hypothesised values, consistent, though inefficient, estimates are used to implement the OLS-based $C(\alpha)$ test. As a result, the test is robust. A Monte Carlo analysis of the conditional score tests along with the joint and standard separate tests is reported in Section 4.6. Finally, Section 4.7 contains the summary and concluding comments.

4.2 Conditional (Adjusted) Score Tests

In conducting separate score tests of subsets of parametric restrictions, one typically assumes the validity of other auxiliary restrictions. This method can have misleading consequences if the auxiliary restrictions are not valid. In duration models it is common to assume that the functional form of the model is correctly specified when testing for neglected heterogeneity. The heterogeneity test may be correlated with the test of functional form specification and hence may lead to erroneous conclusions when the functional form of the model is incorrect. In contrast, conditional (adjusted) score tests suggested in this section do not suffer from this limitation as a correction is made to compensate for the effect of the violation of the auxiliary assumption.

Let $\theta = [\theta_1' \ \theta_2']'$ be the vector of k parameters to be estimated, and θ_1 and θ_2 have k_1 and k_2 elements respectively. Let L denote the log-likelihood function and $s_{1(\theta)} = dL/d\theta_1$ and $s_2(\theta) = dL/d\theta_2$ denote the score vectors. Let $I(\theta) = -E[d^2L \ / \ d\theta d\theta']$ denote the information matrix and let $s(\theta_0)$ and $I(\theta_0)$ denote the score vector and the information matrix respectively, evaluated under the null hypothesis. It is known that under the standard regularity conditions, the score vector has a multivariate normal distribution,[3] that is:

$$N^{-1/2}\begin{bmatrix} s_1(\theta_o) \\ s_2(\theta_o) \end{bmatrix} \sim n\left[0, \quad \frac{1}{N}\, I(\theta_o)\right] \qquad (4.2.1)$$

where:

$I(\theta) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$ and N denotes the sample size.

Let the null hypothesis be:

$$H_o : \theta_1 = \theta_{1o}. \qquad (4.2.2)$$

The score test of the null hypothesis is:

$$LM = s_1'(\tilde{\theta}_o) I^{11}(\tilde{\theta}_o) s_1(\tilde{\theta}_o) \qquad (4.2.3)$$

where (~) denotes that the quantities have been evaluated at the restricted MLE of θ and $I^{11} = [I_{11} - I_{12}(I_{22})^{-1}I_{21}]^{-1}$ is the partitioned inverse of $I(\theta)$. The given test statistic has a chi-square distribution under H_0 with k_1 degrees of freedom. The random variable, given by (4.2.3), has a chi-square distribution because it is based on the following marginal distribution result:

[3] see Breusch and Pagan (1980), Engle (1982, 1984), Godfrey (1988) etc..

$$s_1(\theta_o) \sim n[0, \quad (I^{11})^{-1}]. \tag{4.2.4}$$

In order to implement diagnostic tests for misspecification, one can also use the results of the conditional normal distribution where the random variable $s_1(\theta)$ is conditional on the realised value of $s_2(\theta) = c_2$. Using (4.2.1), the following can be derived as:

$$(s_1|s_2=c_2) \sim n[I_{12}(I_{22})^{-1}c_2, \quad (I^{11})^{-1}]. \tag{4.2.5}$$

Define the conditional score as the quantity:

$$s_1^* = (s_1|s_2=c_2) - I_{12}(I_{22})^{-1}c_2 \tag{4.2.6}$$

where the second term on the right-hand side is the "adjustment" whose effect may be interpreted as purging $s_1(\theta)$ of the correlation with $s_2(\theta)$. The adjusted score s_1^* $= s_1$ if $I_{12} = 0$. The adjusted score may also be interpreted as the residual from the regression of s_1 on s_2.

The conditional separate score test of H_o is based on the quadratic form:

$$LM^c(\theta) = s_1^{*\prime}(\theta_o) \; I^{11}(\theta_o) \; s_1^*(\theta_o). \tag{4.2.7}$$

The quadratic form (4.2.7) may be interpreted as a conditional test generating equation, the value and the properties of the test being dependent on the choice of the estimator for the nuisance parameter θ_2. Consider three possible choices. First, if θ_2 is estimated without restrictions, then $c_2 = 0$ by the first-order condition of maximising the likelihood function, namely $\delta L/\delta\theta_2 = 0$. Consequently, the conditional and unconditional scores will be identical. If θ_2 or some component of it is restricted, then, in general, c_2 does not equal 0, but is approximately zero if the restriction is satisfied. Second, if the previous alternative is computationally demanding but the root-N consistent estimate of θ_2, denoted by θ^n, is easy to obtain, then the test statistic may be evaluated at that point. The test will be asymptotically equivalent to the first alternative and will not depend on the distribution of θ^n.

The third alternative is relevant when the root-N consistent estimate of θ_2 is also not available. Here (4.2.7) is evaluated at the restricted maximum likelihood estimator, where the restrictions comprise the auxiliary restrictions in addition to the ones being tested. This test is only valid under the joint null, however, in this case, the separate tests will also be valid. Nevertheless, it is cheap to compute this test along with the joint test, and it has the merit of having been derived under a more general alternative. Moreover, it includes the adjustment factor for correlation between the scores, even though the adjustment is done under the restricted joint null. Unlike the first two alternatives, the properties of this test will depend on θ_2.

The conditional score test of the null hypothesis based on θ^n is defined by:

$$LM^c(\theta^n) = s_1^{*\prime}(\theta^n)\ I^{11}(\theta^n)\ s_1^*(\theta^n). \tag{4.2.8}$$

The test defined by (4.2.8) is known as the Neyman $C(\alpha)$ test.[4] Traditionally the test has been motivated by computational considerations since the root-N consistent estimate of θ under the null is often easier to obtain than the maximum likelihood estimate required for the score test. Hence, the test is sometimes dubbed the "pseudo-score test".

Another variant of the conditional score test, described above, is derived as follows:

$$LM^c(\tilde{\theta}) = s_1^{*\prime}(\tilde{\theta})\ I^{11}(\tilde{\theta})\ s_1^*(\tilde{\theta}). \tag{4.2.9}$$

To reiterate, this test is based on the restricted maximum likelihood estimates derived under the joint null hypothesis. The test is algebraically similar to the $C(\alpha)$ test and has the same properties as $LM^c(\theta^n)$ under the joint null. The use of this test is motivated by the fact that

[4] The $C(\alpha)$ test is discussed in the statistical literature by Neyman (1959) and Moran (1970). For references in the econometrics literature see Breusch and Pagan (1980), Engle (1984), Holly (1987), and more recently Wooldridge (1989a).

some nuisance parameters may not be easy to estimate using maximum likelihood methods. Hence, these parameters are restricted to some hypothesised value using auxiliary restrictions. However, even though the nuisance parameters are restricted, the correlation of the score corresponding to the nuisance parameter with the relevant score of the parameter that is being tested is allowed for. These ajusted score tests, therefore, are expected to outperform the standard separate score tests when the auxiliary restrictions are not valid in a given model.

4.3 Joint Tests of Heterogeneity and Duration Dependence

The problem of distinguishing between genuine duration dependence and the spurious duration dependence induced by neglected heterogeneity is discussed in Salant (1977), Lancaster (1979) etc.[5] Even if every individual case in the sample has constant hazard, implying no duration dependence, neglected heterogeneity induces the appearance of a declining hazard rate. This problem of identification results in distortions of the standard separate tests. It can be seen in Jensen's (1987) Monte Carlo evidence that positive duration dependence is negated by the spurious negative duration dependence due to neglected heterogeneity. As a result, the standard separate tests pick up no misspecification even though both sources of misspecification exist. Moreover, a test of duration dependence is also sensitive to any evidence of neglected heterogeneity and vice-versa. Based on the outcome of this test, one may generalise the model to incorporate duration dependence when, in reality, the problem is of neglected heterogeneity.

There exists a non-zero correlation between tests of heterogeneity and duration dependence that results in distortions in the standard separate score tests. In this section, joint score tests based on a locally heterogeneous

[5] see chapter 2 and 3 for details.

Weibull model that allows for both heterogeneity and duration dependence are developed.[6] The analysis is further extended to consider a more general heterogeneous generalised gamma distribution. For both cases, an analysis of the non-centrality parameter of the joint test is provided.

4.3.1 Heterogeneous Weibull Model

Consider a locally heterogeneous Weibull model with uncensored observations, whose joint likelihood is:

$$L = \Sigma\left[\log(\alpha) + (\alpha - 1)\log(t) + \log(\mu) - \epsilon + \log[1 + \frac{\sigma^2}{2}(\epsilon^2 - 2\epsilon)]\right]. \tag{4.3.1}$$

$$\mu = \exp(X\beta) = \exp(\beta_0 + X_1\beta_1) \text{ and } \epsilon = \mu t^{\alpha}. \tag{4.3.2}$$

Here X_i is a (kx1) vector of exogenous variables, and without loss of generality it is assumed that $X = (1 \quad X_1)$ and $E(X_1) = 0$, implying that:

$$E(X) = [1 \quad 0_k] \text{ and } E(XX') = \begin{bmatrix} 1 & 0 \\ 0 & \Omega \end{bmatrix}. \tag{4.3.3}$$

Also, ϵ is a generalised error in the sense of Cox and Snell and has a unit exponential distribution if the model is correctly specified. The duration dependence parameter is α; $\alpha < 1$ implies negative duration dependence and $\alpha > 1$

[6] see Jaggia and Trivedi (1989).

implies positive duration dependence. The heterogeneity parameter is σ^2; $\sigma^2 > 1$ implies neglected heterogeneity in the model.

The Joint Score Test

The joint test of no neglected heterogeneity and no duration dependence constitutes the null hypothesis:

$$H_o\text{: } \sigma^2=0 \text{ and } \alpha=1. \qquad (4.3.4)$$

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

$\theta_1{}' = (\sigma^2 \ \ \alpha)$ and $\theta_2{}' = (\beta_o \ \ \beta_1{}')$.

The elements of the score vector, $s_1(\theta_o)$, evaluated under the null hypothesis given by (4.3.4) are:

$$\left.\frac{dL}{d\sigma^2}\right|_{H_o} = \frac{1}{2}\Sigma[\epsilon^2 - 2\epsilon] = s_{11}(\theta_o). \qquad (4.3.5)$$

$$\left.\frac{dL}{d\alpha}\right|_{H_o} = \Sigma[1 + (1-\epsilon)\log(t)] = s_{12}(\theta_o). \qquad (4.3.6)$$

Furthermore, the information matrix $I(\theta_o)$, evaluated under (4.3.4), can be derived as:

$$I(\theta_o) = N\begin{bmatrix} 2 & (\beta_o-\phi(2)-1) & -1 & 0 \\ * & I_{22}(\theta_o) & \phi(2)-\beta_o & -\beta_1{}'\Omega \\ * & * & 1 & 0 \\ * & * & * & \Omega \end{bmatrix}. \qquad (4.3.7)$$

where:

$$I_{22}(\Theta_o) = 1 + \phi'(2) + (\phi(2))^2 + \beta_o^2 + \beta_1'\Omega_1\beta_1 - 2\phi(2)\beta_o$$

and $\phi(r) = \dfrac{d\log \Gamma(r)}{dr}$ is the digamma function and

$\phi'(r) = \dfrac{d^2 \log \Gamma(r)}{dr^2}$ is the trigamma function.

The joint score test statistic for heterogeneity and duration dependence is:

$$LM_{hd} = s_1'(\tilde{\Theta}_o)\ I^{11}(\tilde{\Theta}_o)\ s_1(\tilde{\Theta}_o)$$

where, using (4.3.7):

$$(I^{11})^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = N \begin{bmatrix} 1 & -1 \\ -1 & \phi'(1) \end{bmatrix}. \qquad (4.3\ .8)$$

Therefore,

$$I^{11}(\Theta_o) = \frac{(\phi'(1) - 1)^{-1}}{N} \begin{bmatrix} \phi'(1) & 1 \\ 1 & 1 \end{bmatrix}. \qquad (4.3.9)$$

Observe that the matrix in (4.3.9) does not depend upon any unknown parameters.

Non-Centrality Parameter

The fact that spurious duration dependence cancels out genuine duration dependence, has an effect on the power of the joint test. Using (4.3.4), the null hypothesis is:

H_o: $\Theta_1 = \Theta_{1o}$, where:

$$\Theta_1 = \begin{bmatrix} \sigma^2 \\ \alpha \end{bmatrix}, \text{ and } \Theta_{1o} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Let the sequence of alternative hypotheses be given by:

$$H_a: \Theta_{1o} + N^{-1/2}\delta, \text{ where } \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}. \qquad (3.10)$$

Then, asymptotically, LM_{hd}, based on a heterogeneous Weibull distribution, is distributed as $X^2(2,\tau)$[7], where the non-centrality parameter τ is:

$$\tau = \frac{1}{N}\ \delta'(I^{11})^{-1}\delta \qquad (4.3.11)$$

which, after incorporating (4.3.8) is:

$$= \delta_1^2 - 2\delta_1\delta_2 + \delta_2^2\phi'(1). \qquad (4.3.12)$$

It is also noted that:

$$\tau(\alpha = 1) = N(\sigma^2)^2 \text{ and} \qquad (4.3.13)$$

$$\tau(\sigma^2 = 0) = N(\alpha - 1)^2\phi'(1). \qquad (4.3.14)$$

Looking at the negative middle term in (4.3.12), it can be seen that certain configurations of values of σ^2 and α will make the non-centrality parameter small, thus indicating

[7] see Cox and Hinkley (1974)

reduced local power of the joint test. For both δ_1 and δ_2 positive, implying the presence of positive duration dependence as well as neglected heterogeneity in the data, the power of the joint test will be reduced against local alternatives. It is readily seen that the minimum value of τ:

$$\min_{(\delta_2>0)} \tau = \tau(\delta_2 = {}_1/\phi'(1)) = \delta_1{}^2(1 - 1/\phi'(1)) \qquad (4.3.15)$$

may still be high if $\delta_1{}^2$ is large. On the other hand, when $\alpha < 1$, that is, there is negative duration dependence, the power of the joint test is increased. Therefore, the co-existence of positive duration dependence and neglected heterogeneity poses a problem for the joint as well as the standard separate tests since the power of the former is reduced and the latter is invalidated. This problem may lead the researcher to under-parametrise the model. The presence of heterogeneity along with negative duration dependence, however, enhances the power of the joint test.

4.3.2 Heterogeneous Generalised Gamma Distribution

The use of a Weibull distribution, to allow for duration dependence, may not always be appropriate. When the assumption of a constant hazard rate, implied by the exponential distribution, is found to be incorrect, the correct specification may still not be captured by a Weibull model. The exponential assumption can be relaxed by selecting distributions like gamma, Weibull, lognormal etc. Pereira (1978), using Cox's (1961, 1962) approach for testing non-nested hypotheses, develops tests to discriminate between log-normal, gamma, Weibull and exponential models. This approach, however, becomes intractable when heterogeneity is allowed for.

Another means of discriminating between the above-mentioned parametric models is to use the encompassing approach. As seen in Chapter 3, the generalised gamma distribution encompasses all of the above mentioned distributions and also accommodates non-monotonic hazards. Even though it is difficult to estimate parameters of this distribution, it can be used to derive score tests since only the null model needs to be estimated to implement such tests.

Using the results and notation defined above and in Chapter 3, the log-likelihood function of a heterogeneous generalised gamma distribution is:

$$L = \Sigma \left[k\log(\mu) + \log(\alpha) + (\alpha k - 1)\log(t) - \mu t^{\alpha} - \log(\Gamma(K)) + \log[1 + (\sigma^2/2)(\, k(k-1) - 2k\mu t^{\alpha} + (\mu t^{\alpha})^2\,)] \right]. \quad (4.3.16)$$

The Joint Score Test

The joint test of neglected heterogeneity ($\sigma^2=0$) and duration dependence ($\alpha=1$ and $k=1$) implies the null hypothesis:

$$H_o: \sigma^2=0, \alpha=1 \text{ and } k=1. \qquad (4.3.17)$$

The elements of the relevant score vector, $s_1(\theta_o)$, are:

$$\left.\frac{dL}{d\sigma^2}\right|_{H_o} = \frac{1}{2}\Sigma[\epsilon^2 - 2\epsilon] = s_{11}(\theta_o). \qquad (4.3.18)$$

$$\left.\frac{dL}{d\alpha}\right|_{H_o} = \Sigma[1 + (1-\epsilon)\log(t)] = s_{12}(\theta_o). \qquad (4.3.19)$$

$$\left.\frac{dL}{dk}\right|_{H_o} = \Sigma[\log(\epsilon) - \phi(1)] = s_{13}(\theta_o). \qquad (4.3.20)$$

Moreover, the information matrix, $I(\theta_o)$, evaluated under the null hypothesis is:

$$N\begin{bmatrix} 2 & \beta_o-\phi(2)-1 & 1/2 & -1 & 0 \\ * & I_{22}(\theta_o) & \beta_o-\phi(1) & \phi(2)-\beta_o & -\beta_1'\Omega \\ * & * & \phi'(1) & -1 & 0 \\ * & * & * & 1 & 0 \\ * & * & * & * & \Omega \end{bmatrix} \qquad (4.3.21)$$

where, as before:

$$I_{22}(\theta_o) = 1 + \phi'(2) + (\phi(2))^2 + \beta_o^2 + \beta_1'\Omega_1\beta_1 - 2\phi(2)\beta_o.$$

The Joint Score test can be derived using the standard result given by (4.2.3), namely:

$$LM_{hd} = s_1'(\tilde{\theta}_o)\ I^{11}(\tilde{\theta}_o)\ s_1(\tilde{\theta}_o)$$

where:

$$(I^{11})^{-1} = N\begin{bmatrix} 1 & -1 & -1/2 \\ -1 & \phi'(1) & 1 \\ -1/2 & 1 & \phi'(2) \end{bmatrix}. \qquad (4.3.22)$$

Non-Centrality Parameter

In order to derive the non-centrality parameter of the test based on the generalised gamma distribution, let the sequence of local alternatives be:

H_a: $\theta_{1o} + N^{-1/2}\delta$, where:

$$\theta_{1o} = \begin{bmatrix} \sigma^2 = 0 \\ \alpha = 1 \\ k = 1 \end{bmatrix} \text{ and } \delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}.$$

Then, asymptotically, LM_{hd} is distributed as $X^2(3,\tau)$, where the non-centrality parameter τ is:

$$\tau = \frac{1}{N}\, \delta'(I^{11})^{-1}\delta$$

$$= \delta_1^2 + \delta_2^2\phi'(1) + \delta_3^2(\phi'(1)-1) - 2\delta_1\delta_2 - 2\delta_1\delta_3 + 2\delta_2\delta_3. \qquad (4.3.23)$$

The non-centrality parameter of the joint test of three restrictions is at least as great as that of the test of two restrictions. However, the asymptotic power of this test may be smaller due to higher degrees of freedom.

4.4 Adjusted Score Tests for Duration Models

As mentioned earlier, the conditional (adjusted) score tests are used to identify the exact source of misspecification. Due to the existence of non-zero correlation between $s_{11}(\theta)$ and $s_{12}(\theta)$, the separate tests proposed in the literature are of little use. The adjusted score tests are founded on the known distribution of one score conditional on the realised value of the second score. This approach is different from standard test procedures in which the value of the second score is ignored. Thus, for example, consider the distribution of s_{11}, conditional on the realised value of s_{12}, say c_{12}. Using the results from Sections 4.2 and 4.3,

$$(s_{11} | s_{12}=c_{12}) \sim n \left[\frac{\sigma_{12}}{\sigma_{22}} c_{12},\ \sigma_{11} - \frac{(\sigma_{12})^2}{\sigma_{22}} \right]. \qquad (4.4.1)$$

The conditional test of heterogeneity:

$$LM_h^C = \frac{\left[s_{11}(\tilde{\theta}_o) - (\sigma_{12}/\sigma_{22})\ c_{12}(\tilde{\theta}_o) \right]^2}{\sigma_{11} - (\sigma_{12})^2/\sigma_{22}}$$

$$= \frac{\left[s_{11}(\tilde{\theta}_o) + (1/\phi'(1))\ c_{12}(\tilde{\theta}_o) \right]^2}{N(1 + 1/\phi'(1))}. \qquad (4.4.2)$$

Similarly, the conditional duration dependence test is:

$$LM_d^C = \frac{\left[s_{12}(\tilde{\theta}_o) + c_{11}(\tilde{\theta}_o) \right]^2}{N(\phi'(1) - 1)}. \qquad (4.4.3)$$

Relationship Between Adjusted and Standard Tests

By ignoring the correlation between $s_{11}(\theta_o)$ and $s_{12}(\theta_o)$, separate test statistics, as used in the literature, can be derived. The test of heterogeneity is:

$$LM_h = \frac{[s_{11}(\tilde{\theta}_o)]^2}{\sigma_{11}} = \frac{[s_{11}(\tilde{\theta}_o)]^2}{N} \qquad (4.4.4)$$

which is similar to the test proposed by Lancaster (1983). Analogously, a partial test of duration dependence is:

$$LM_d = \frac{[s_{12}(\tilde{\theta}_o)]^2}{\sigma_{12}} = \frac{[s_{12}(\tilde{\theta}_o)]^2}{N\phi'(1)} \qquad (4.4.5)$$

which is the duration dependence test derived by Jensen (1987). Both LM_h and LM_d are quite appropriate under the joint null. The actual size will differ from the nominal significance level if the covariance is non-zero. Block diagonality of $I^{11}(\theta^o)$ is a necessary and sufficient condition for the joint test to be additive in LM_h and LM_d[8], that is, for asymptotic zero correlation between LM_h and LM_d. However, $I^{11}(\theta_o)$ is evidently not block-diagonal. As

[8] Bera and McKenzie (1986)

seen in Chapter 3, tests which ignore this may be misleading.[9] The conditional tests based on the jointly restricted maximum likelihood estimator are also appropriate under the joint null. However, since they are adjusted for possible correlation between $s_{11}(\theta)$ and $s_{12}(\theta)$, they may have better performance under the alternative hypothesis. Usually the joint presence of heterogeneity and duration dependence cannot be a priori ruled out. Hence, the joint score test has obvious advantages over the standard and the conditional separate tests whose nominal size and power will be affected by the presence of the second (ignored) complication and which can be misleading. On the other hand, if the joint null is rejected, one may wish to test the component hypotheses. The conditional score tests suggested above may be more informative than the standard tests in this regard. Further, for some parameter configurations the joint test is likely to have low power.

Finally, the conditional heterogeneity test based on the heterogeneous generalised gamma distribution, under the joint null specified by (4.3.17) can similarly be derived. Here the test will be based on the distribution of s_{11} conditional on the realised value of s_{12} and s_{13}. Using (4.3.22) the test can be easily implemented.

---

[9] see, also, Jaggia (1990) for an empirical illustration.

4.5 OLS Based C(α) Test of Heterogeneity

The distribution of the conditional heterogeneity test developed above, LM^c_h, is independent of nuisance parameters only under the joint null. This test may have better properties for some parameter values than the standard heterogeneity test, LM_h. However, its distribution depends upon unknown nuisance parameters if the auxiliary assumptions regarding those parameters are not satisfied. We desire a more generally valid and robust test whose distribution does not depend upon any nuisance parameters. The OLS based test suggested below is such a test and is asymptotically equivalent to the test based on the maximum likelihood estimates of the Weibull model. This test is based on the root-N consistent estimates of the parameters and not on the distribution of those estimators. The resulting test is robust since the nuisance parameter of duration dependence is estimated, though inefficiently, instead of being restricted to an incorrect hypothesised value.

Since $\theta' = (\theta_1' \ \theta_2')$, let $\theta_1 = \sigma^2$ and $\theta_2' = (\alpha \ \beta_0 \ \beta_1')$. In order to base a test of H_0: $\theta_1 = 0$, root-N consistent estimates of θ_2 are required. These estimates may be obtained through an OLS regression as follows. If t has a Weibull distribution, then using y = log(t) we can write:

$$y = -X\frac{\underline{\beta}}{\alpha} + \frac{\underline{w}}{\alpha} \qquad (4.5.1)$$

$$= X\Gamma + U$$

$$= \Gamma_o + X_1\Gamma_1 + U \qquad (4.5.2)$$

where w has an extreme value distribution with $E(w) = \phi(1) = -.5772$ and $Var(w) = \phi'(1) = 1.6449$, $\Gamma = \beta/\alpha$ an $U = w/\alpha$. Hence,

$$\alpha = sqrt[1.6449 / var(U)]. \qquad (4.5.3)$$

$$\beta_o = - \alpha\Gamma_o - .5772. \qquad (4.5.4)$$

$$\beta_1 = - \alpha\Gamma_1. \qquad (4.5.5)$$

As Γ and $Var(U)$ can be consistently estimated using ordinary least squares, consistent estimates of all the parameters of a Weibull model can be obtained.

To construct the $C(\alpha)$ test of heterogeneity, we use the expression (4.2.8) given as:

$$LM^c(\theta^n) = s_1^{*\prime}(\theta^n) \; I^{11}(\theta^n) \; s_1^*(\theta^n)$$

where $s_1^* = (s_1|s_2{=}c_2) - I_{12}(I_{22})^{-1}c_2$ and θ^n is any root-N consistent estimate of θ. Note that the score, $\delta L/\delta\delta\sigma^2$ is taken conditional on the realised values of the scores corresponding to all the parameters of the model. The test can be implemented by substituting the following expressions into (4.2.8):

$$c_2 = \begin{bmatrix} dL/d\beta \\ dL/d\alpha \end{bmatrix} = \begin{bmatrix} \Sigma[X(1-\epsilon)] \\ \Sigma[y(1-\epsilon) + 1/\alpha] \end{bmatrix}. \tag{4.5.6}$$

$$\frac{1}{N} I_{12} = [-1 \quad 0_k \quad -(m+1)/\alpha]. \tag{4.5.7}$$

$$\frac{1}{N}(I_{22})^{-1} = \begin{bmatrix} 1+m^2 q & -mq\beta_1' & -\alpha mq \\ -mq\beta_1 & \Omega^{-1}+q\beta_1\beta_1' & \alpha q\beta_1 \\ -\alpha mq & \alpha q\beta_1' & \alpha^2 q \end{bmatrix}. \tag{4.5.8}$$

$$I^{11} = N/(1-q). \tag{4.5.9}$$

$$m = \phi(2) - \beta_o; \; q = 1/\phi'(1); \; \epsilon = \mu t^{\alpha}. \tag{4.5.10}$$

The computation of (4.2.8) is achieved by using consistent least squares estimates in conjunction with expressions defined in (4.5.6) to (4.5.9).

4.6 Monte Carlo Experiments

The Monte Carlo experiments are performed to compare the performance of all the tests described in the last three sections. The null hypothesis for all tests is that the observations are drawn randomly from an exponential distribution with no neglected heterogeneity. All tests have been derived in the context of a locally heterogeneous Weibull distribution. The performance of six test statistics is compared. These tests are LM_h, LM_d, LM_{hd}, $LM_h^C(\tilde{\theta})$, $LM_d^C(\tilde{\theta})$ and $LM_h^C(\theta^n)$, all of which are based on the theoretical information matrix. Note that both $LM_h^C(\tilde{\theta})$ and $LM_d^C(\tilde{\theta})$ are evaluated at the restricted maximum likelihood estimates derived under the joint null implied by an exponential distribution. $LM_h{}^c(\theta^n)$, on the other hand, is the $C(\alpha)$ test that is based on the consistent estimates obtained using least squares including the nuisance duration dependence parameter.

4.6.1 Design of Sampling Experiments

The size and power properties of the six test statistics mentioned above are evaluated on the basis of twelve different data generating processes (models). Each simulation experiment is based on 500 replications. The parameters (β_o, β_1) are set at (5.0, 1.0). The variable X_1 is taken as a random draw from a uniform [0,1] distribution and is held fixed for all experiments. The heterogeneity term is denoted by V where log V is a random draw from a normal, $n(0,\sigma^2)$ distribution implying multiplicative log-normal heterogeneity. Once the draw corresponding to a particular value of σ^2 is made, the vector of the heterogeneity term is held fixed, to be used with different values of α and for all replications. This method reduces sample variability between replications and between experiments.

Different combinations of α and σ^2 are used to generate the twelve experiments shown in Table 1. Estimation under the joint null (Model 1) is done using sample sizes, N = 50, 100, 200, and 500. Different sample sizes are chosen to compare the asymptotic and finite sample distributions of the test statistics when the null holds. To make a power comparison, the sample size chosen under the alternative hypotheses (Models 2-12) is 200. The data generation process for Model 1 is exponential and for Models 2-12 is either Weibull ($\alpha \neq 1$) or exponential ($\alpha = 1$) with heterogeneity ($\sigma^2 > 0$) or without heterogeneity ($\sigma^2 = 0$).

<u>4.6.2 Results of Sampling Experiments</u>

For the correctly specified Model 1, the match between the theoretical asymptotic distribution and the actual results is good (see Table 2). The $C(\alpha)$ test given by $LM_h^c(\theta^n)$ appears to have relatively thicker tails compared with the rest especially for N = 50 and 100. For LM_h and LM_{hd} the proportion of rejection of the null hypothesis is smaller than the nominal level of the test, especially when the test is performed at the 10 or 5 percent significance levels. When the model is correctly specified, all tests perform slightly better than the $C(\alpha)$ test.

Table 3 contains the results for Models 2, 3 and 4 which incorporate duration dependence but not unobserved heterogeneity. The data generating process in Model 2 is subject to negative duration dependence (α =0.75) and Models 3 and 4 are subject to positive duration dependence (α = 1.30 and 1.45). For Model 2 LM_h, LM_d and LM_{hd} all perform extremely well. Thus, the essential misspecification arising from negative duration dependence is diagnosed by the separate and joint tests. However, LM_h is unable to distinguish between duration dependence and unobserved heterogeneity. The implication of this result is that the "heterogeneity tests" proposed in the literature are not tests of heterogeneity alone when other sources of misspecification are also present. Conditional score tests are intended to overcome this problem. Therefore, ideally, their rejection proportion should be close to the nominal significance level of five percent (it is actually 39.6 percent for $LM_h^C(\tilde{\theta})$ and 79.4% for $LM_h^C(\theta^n)$). For $LM_d^C(\tilde{\theta})$ rejection should be close to 100%, but it is actually 79.4%.

However, as expected, the performance of all tests considered other than the $C(\alpha)$ test is model dependent, as the results of Models 3 and 4 illustrate. Here all misspecification tests have high power and all conditional score tests perform extremely well, correctly indicating the nature of misspecification in a high proportion of cases. Of course, the $C(\alpha)$ test based on the root-N consistent estimates of the parameters including the nuisance parameter should perform better than the tests based on the restricted maximum likelihood estimator, and it does.

The results are less favorable to the adjusted score tests, based on the restricted maximum likelihood estimates, when Models 5 and 9 are considered. Here there is some unobserved heterogeneity but no duration dependence. $LM_h^C(\tilde{\theta})$ still has high power, but $LM_d^C(\tilde{\theta})$, unfortunately, incorrectly identifies duration dependence in 42.2% and 76.2% of the cases, reflecting correlation between the two tests.

In Models 7, 8, 11 and 12, there is unobserved heterogeneity and positive duration dependence. It is seen that the separate heterogeneity test, LM_h has rather low power.[10] The joint test, LM_{hd} has relatively higher power which increases

[10] see Jensen (1987) for a similar result.

with the magnitude of α. The conditional tests based on the restricted maximum likelihood estimates also suffer a reduction in power for Models 5 and 8 as compared with Models 9 and 12. By contrast, the $C(\alpha)$ test of heterogeneity is very robust and has higher power in all cases. For all values of α, the power of the tests increases when the value of σ^2 is raised from 0.6 to 0.8. Nevertheless, it is somewhat problematic that there are parameter configurations in which all score tests based on the restricted maximum likelihood estimator have rather low power.

An easily implementable variant of $LM_h^c(\theta^n)$ which uses outer product of the sample scores, rather than the theoretical (expected) information matrix, to implement the test was also investigated.[11] The results obtained were rather disappointing. The experiments were repeated by increasing the sample size to 1000. This change produced some improvement in performance, but the test still had low power, as compared to the one that is based on the theoretical information matrix.

[11] Wooldridge (1989) discusses the easily implementable $C(\alpha)$ version of the conditional moment restriction tests.

4.7 Conclusion

In this chapter, conditional score tests are motivated and developed in the context of duration models. These conditional tests are shown to be useful alternatives to separate and joint tests of heterogeneity and duration dependence. A known limitation of a joint test is that if the null hypothesis is rejected, one still needs information from separate tests to indicate the nature of the required respecification. The standard separate tests may be of little help when they are correlated. Bera and Jarque (1982) derive a joint test of different restrictions for a classical linear regression model. They also suggest the Multiple Comparison Procedure in order to identify different sources of errors if the joint test results in the rejection of the null hypothesis. Their procedure, however, is aided by the additivity of their separate tests, implying that all tests they derive are asymptotically independent under the joint null. So, even when multiple misspecifications exist, each separate test is somewhat informative in carrying out a search for an appropriate model.

In the given context, the tests of heterogeneity and duration dependence are shown to be correlated with each other within a heterogeneous Weibull model. However, even though the conditional score tests, suggested in Section 4.4, are based on the restricted joint null, each score is purged of the correlation with the other relevant score. Therefore, the conditional score tests derived this way may contain more information regarding the specific source of misspecification than the standard separate tests. These conditional tests can be used to locate the exact source of misspecification using the Multiple Comparison Procedure.[12]

The second variant of the conditional score tests ($C(\alpha)$) is especially useful when it is hard to obtain maximum likelihood estimates of some nuisance parameters. Tests can easily be implemented using any inefficient but consistent estimates of the parameters rather than some hypothesised values of the nuisance parameters which may be incorrect. Based on the preliminary outcome of these tests, applied researchers can determine if it is worthwhile to continue and estimate the model with maximum likelihood methods which ensure efficiency. This procedure can be extended to test for heterogeneity in functional forms more general than the Weibull distribution. Moreover, an analogous technique can be used to test for duration dependence by using a preliminary estimate of the heterogeneity parameter from a least squares regression. Note that this can be achieved without parameterising the heterogeneity distribution.

---

[12] see Savin (1980,1984) for a survey of Multiple Comparison Procedures.

TABLE 4.1

Parameter Combinations Used in Sampling Experiments

Model	σ^2	α	Model	σ^2	α
1	0	1.0	7	0.6	1.30
2	0	0.75	8	0.6	1.45
3	0	1.30	9	0.8	1.0
4	0	1.45	10	0.8	0.75
5	0.6	1.0	11	0.8	1.30
6	0.6	0.75	12	0.8	1.45

TABLE 4.2

Percentage Rejections at α Significance Level for Model 1

	N = 50	N = 100	N = 200	N = 500
100α	10, 5, 1	10, 5, 1	10, 5, 1	10, 5, 1
LM_h	6.0,2.2,1.4	6.8,2.4,1.4	7.2,3.8,1.0	9.0,3.0,1.4
LM_d	12.2,5.6,1.2	9.0,4.6,1.2	9.4,5.0,2.6	9.6,4.4,1.6
LM_{hd}	6.8,3.4,1.0	7.2,4.8,1.0	8.2,4.4,1.2	8.0,3.4,0.8
$LM_h^C(\tilde{\theta})$	5.0,3.2,1.0	5.8,3.4,1.0	7.2,3.0,1.2	8.2,2.6,1.2
$LM_d^C(\tilde{\theta})$	8.8,4.2,2.0	9.6,5.0,2.0	9.2,4.2,1.8	9.4,5.2,1.2
$LM_h^C(\theta^n)$	11.2,8.4,6.6	11.6,7.8,4.2	9.0,7.0,4.0	8.8,5.7,3.4

TABLE 3

Percentage Rejections of H_0 at 5% Significance Level

For Models 2-4, N=200

(σ^2 = 0)

	Model #	2	3	4
	α =	0.75	1.30	1.45
LM_h		97.4	95.0	100
LM_d		99.8	99.6	100
LM_{hd}		99.8	98.2	100
$LM_h^C(\tilde{\theta})$		39.6	0.0	0.0
$LM_d^C(\tilde{\theta})$		79.4	88.8	100
$LM_h^C(\theta^n)$		8.6	8.8	6.8

TABLE 4.4

Percentage Rejections of H_o at 5% Significance Level

For Models 5-8, N=200

(σ^2 = 0.6)

	Model #	5	6	7	8
	α =	1.0	0.75	1.30	1.45
LM_h		99.8	100.0	31.0	12.0
LM_d		99.8	100.0	5.4	62.2
LM_{hd}		99.6	100.0	45.6	77.6
$LM_h^C(\tilde{\theta})$		81.6	96.2	52.2	36.2
$LM_d^C(\tilde{\theta})$		42.2	81.0	46.6	87.8
$LM_h^C(\theta^n)$		74.8	73.8	78.4	78.8

TABLE 4.5

Percentage Rejections of H_o at 5% Significance Level

For Models 9-12, N=200

(σ^2 = 0.8)

	Model #	9	10	11	12
	α =	1.0	0.75	1.30	1.45
LM_h		100.0	100.0	81.4	47.4
LM_d		100.0	100.0	36.6	15.2
LM_{hd}		100.0	100.0	84.8	83.6
$LM_h^C(\tilde{\theta})$		97.8	99.6	87.6	77.8
$LM_d^C(\tilde{\theta})$		76.2	93.0	74.2	87.2
$LM_h^C(\theta^n)$		93.6	94.2	92.8	93.6

CHAPTER 5

TESTS FOR UNOBSERVED HETEROGENEITY IN THE PRESENCE OF CENSORED OBSERVATIONS

5.1 Introduction

A peculiar feature of duration models is that data on durations are seldom complete. It is common for some observations to be censored, typically right censored. Information from the censored observations can be extracted for estimation using standard maximum likelihood techniques.[1] However, the effect of censored data on the performance of diagnostic tests is not very clear. Even though regression models with censored data may be especially sensitive to specification errors,[2] little is known about the effectiveness of the testing procedures in the presence of various types and degrees of censoring. In this chapter, the effect of various types of censoring on different versions of the score test for neglected heterogeneity is examined. A Monte Carlo analysis is made of the consequences of various types of censoring on different versions of the test for the Weibull and exponential models.

---

[1] see, for example, Lawless (1982).

[2] Horowitz and Neumann (1989), using Kennan's strike data, show that the standard diagnostics may lead to erroneous conclusions, when data are censored.

In order to implement a score test of heterogeneity, the information matrix, under the null hypothesis, has to be evaluated. The theoretical information matrix can be computed when the data consist of complete observations. With censored observations, such a matrix cannot be found without additional information regarding the censoring mechanism. It is common to have censored observations due to finite observation periods. Individuals are observed over fixed time periods and some individual durations may not have ended when the finite observation period ends. Any indidvidual duration will be known exactly only if it is less than some pre-specified value. This observation process results in what is known as Type-1 censored data. In this chapter, a test for heterogeneity that is based on the theoretical information matrix is derived for Type-1 censored data. For the Weibull model, the above test involves evaluating some expressions numerically.

Sometimes data are censored for reasons other than finite observation periods. One may just lose an observation after observing it to a particular point resulting in data that are not Type-1 censored. The information matrix with these censored data cannot be determined without making further papametric restrictions on the censoring mechanism. When observations are censored randomly, involving no length biased censoring, a method is suggested where only the uncensored observations are used to implement the test of heterogeneity. This method consists of estimating the parameters of the model using all observations. These consistent and efficient estimates are then used to evaluate scores consisting only of complete observations.

Alternatively, heterogeneity tests can be based on the observed information matrix. Efron and Hinkley (1978), generally, recommend the use of the observed information matrix to implement specification tests as it is closer to the data than the corresponding, expected (theoretical), information matrix. Two obvious candidates for this matrix are the sample hessian of the log-likelihood function and the outer product of the sample scores. In this chapter it is found that in finite samples, the performance of tests based on the observed information matrix is case sensitive. The information matrix based on the sample hessian is not always positive definite, implying that the test is sometimes meaningless owing to sampling fluctuations. The following Monte Carlo analyses reflect this problem for some experiments. Tests based on the outer product of the sample scores are easy to implement but the nominal size of these tests is different from the actual size. For some data generating processes, the number of times the test is rejected under the null is found to be more than the chosen significance level.[3] Finally, Monte Carlo analyses are also used to study the performance of the test based on Kiefer's method for approximating the information matrix.

The rest of the chapter is organised as follows. In Section 5.2, different approximations of the density functions that can be used to implement the score test of heterogeneity are specified. In Section 5.3, the test is derived that is based on the uncensored data. Different versions of the test are given, depending on the choice of estimate for the information matrix. The test statistics, based on the theoretical information matrix, for all approximations of the density function given in Section 5.2, are shown to be algebraically equivalent. Section 5.4 contains a discussion of different ways of implementing the test when some observations are censored. A distinction is made between Type-1 censored data and totally uninformative and randomly censored data. For each type of data, tests are derived that are based on the theoretical information matrix. Monte Carlo analyses of all versions of the test, including the ones based on the observed information matrix, are presented in Section 5.5. Section 5.6 contains the conclusion.

---

[3] The size problem of tests based on the outer product of the sample scores is also reported by Davidson and MacKinnon (1983) in their Monte-Carlo analysis with linear regression models.

5.2 Background

Typically, in duration models, the data available to an econometrician are on the variables t_i, X_i, and C_i, where $i = 1.....N$. C_i is an indicator variable, such that:

$$C_i = \begin{cases} 1 \text{ if a spell is complete} \\ 0 \text{ if a spell is right censored.} \end{cases}$$

$$t_i = \min(T_i,\ L_i) \text{ and } C_i = 1 \text{ if } T_i \le L_i. \tag{5.2.1}$$

The T_i's are the actual durations and the L_i's are the censoring times that may not be available for all individuals in the sample. The joint probability density function of (t,C), given X, is:

$$f(t,C) = f(t)^C\ S(L)^{1-C} \tag{5.2.2}$$

where f(t) is the density function and S(t) is the survivor function of T. Given the assumption that the observations on the pairs (t_i,C_i) are independent, the log likelihood function is:

$$L = \sum_{i=1}^{N} [C_i \log f(t_i;X_i) + (1 - C_i) \log S(t_i;X_i)]. \tag{5.2.3}$$

Once the density and survivor functions of any parametric model are specified, the likelihood function can easily be maximised using standard techniques. The density function of a Weibull model, conditional on regressors, is given by

$f(t|X) = \mu\alpha t^{\alpha-1}\exp(-\mu t^{\alpha})$ where $\mu = \exp(X\beta)$. If neglected heterogeneity is suspected in the sample, it can be represented by writing $\mu = \exp(X\beta + U) = V\exp(X\beta)$ where $V = \exp(U)$ is the heterogeneity term. As V (or U) is not observable, the density of a duration conditional only on X is taken as: $f(t|X) = E_V[f(t|X,V)]$ where the expectation is taken with respect to the distribution of V (or U).

Following Lancaster (1985), a Taylor series expansion of the conditional duration distribution, $f(t|X,V)$, around the unit mean of the heterogeneity term, V, can be used to derive the following:[4]

$$f(t|X) = [\mu\alpha t^{\alpha-1}\exp(-\epsilon)]\ [1 + \frac{\sigma^2}{2}(\epsilon^2 - 2\epsilon)] \qquad (5.2.4)$$

where σ^2 is the variance of V and $\epsilon = \mu t^{\alpha}$ is the integrated hazard function that is a generalised error in the sense of Cox and Snell (1968).

Kiefer (1984) and Burdett et al (1985) derive a similar expression using a Taylor Series expansion around the zero mean of the heterogeneity term U.[5] The corresponding density function is approximated as:

[4] see Chapter 2 for details

[5] Note that assuming $E(U) = 0$ is not equivalent to assuming $E(V) = 1$.

$$f(t|X) = [\mu\alpha t^{\alpha-1}\exp(-\epsilon)]\ [1 + \frac{\sigma^2}{2}\ (1 - 3\epsilon + \epsilon^2)] \qquad (5.2.5)$$

where σ^2 is the variance of U.

Another approximation of the density function, given by Kiefer (1985) and Sharma (1989), is based on the Laguerre alternatives. Here the density function is specified as:

$$f(t|X) = [\mu\alpha t^{\alpha-1}\exp(-\epsilon)]\ [1 + \Sigma\ \delta_j L_j(\epsilon)]. \qquad (5.2.6)$$

$L_j(\epsilon)$ is the j'th Laguerre polynomial in $\epsilon(t)$ where, for all j, k = 0, 1, 2 ....etc.,

$$\int L_j(\epsilon(t))\ L_k(\epsilon(t))\ f(t|X)\ dt = \begin{bmatrix} 0 \text{ if } j \neq k \\ 1 \text{ if } j = k \end{bmatrix}. \qquad (5.2.7)$$

The family of alternatives considered here is sufficiently flexible and has power against many alternative models.[6] Testing for a Weibull specification reduces to testing for $\delta_j = 0$, for all j. Sharma (1989) points out that testing for $\delta_2 = 0$ can be interpreted as a test for heterogeneity.

For the sake of exposition, the likelihood function using (5.2.3), with the density function approximation given by (5.2.4), can be written as:

---

[6] see Kiefer (1985) or Sharma (1989) for details on Laguerre approximations for the exponential and Weibull models respectively.

$$L = \Sigma \Big[C[\ln \alpha + (\alpha-1)\ln(t) + \ln \mu + \ln (1 + \frac{\sigma^2}{2} (\epsilon^2 - 2\epsilon))]$$

$$- \epsilon + (1-C) \ln(1 + \frac{\sigma^2 \epsilon^2}{2}) \Big] \qquad (5.2.8)$$

where C is an indicator variable (defined above) that equals 1 if a duration is complete and zero if it is right censored. The test of heterogeneity comprises the following null hypothesis:

$$H_o: \sigma^2 = 0. \qquad (5.2.9)$$

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

$\theta_1' = \sigma^2$ and $\theta_2' = (\beta_o \ \ \beta_1' \ \alpha)$.

Under the standard regularity conditions,[7]

$$N^{-1/2} \begin{bmatrix} s_1(\theta) \\ s_2(\theta) \end{bmatrix} \sim n[0, \frac{1}{N} I(\theta)] \qquad (5.2.10)$$

where $I(\theta) = -E[d^2 L / d\theta d\theta']$ denote the theoretical information matrix, partitioned conformably as:

$$I(\theta) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}. \qquad (5.2.11)$$

---

[7] see Breusch and Pagan (1980), Engle (1982, 1984), Bera and McKenzie (1986) and Godfrey (1988) etc.

Let $s(\theta_o)$ and $I(\theta_o)$ denote the score vector and the information matrix respectively, evaluated under the null hypothesis. The score, $s_1(\theta_o)$, computed for the above hypothesis is:

$$\left.\frac{\delta L}{\delta \sigma^2}\right|_{H_o} = \frac{1}{2} \Sigma(\epsilon^2 - 2C\epsilon) = s_1(\theta_o). \qquad (5.2.12)$$

Similarly, the score derived from the density given by (5.2.5) is:

$$\left.\frac{\delta L}{\delta \sigma^2}\right|_{H_o} = \frac{1}{2} \Sigma(\epsilon^2 - 2C\epsilon + C - \epsilon). \qquad (5.2.13)$$

Also, the score w.r.t δ_2, using the Laguerre polynomial approximation given by (5.2.6), can be shown to be:

$$\left.\frac{\delta L}{\delta \delta_2}\right|_{H_o} = \frac{1}{2} \Sigma(\epsilon^2 - 2C\epsilon + 2C - 2\epsilon). \qquad (5.2.14)$$

The last two expressions, (5.2.13) and (5.2.14), are identical to the corresponding expressions derived by Kiefer (1985) and Sharma (1989) but are written in a much simpler form. The form derived here avoids unnecessary expressions involving incomplete gamma functions. Moreover, in order to maximise the likelihood function, $\delta L/\delta \beta_o$ is set equal to zero. This implies that at the maximum likelihood estimates, $\Sigma(C - \tilde{\epsilon}) = 0$. Therefore, the sample scores, (5.2.12) - (5.2.14), are identical even though they are derived under three different approximations.

The score test of heterogeneity is given by:

$$LM = s_1'(\tilde{\theta}_o) I^{11}(\tilde{\theta}_o) s_1(\tilde{\theta}_o) \qquad (5.2.15)$$

where (~) denotes that the quantities have been evaluated at the restricted maximum likelihood estimates of the parameter vector, θ, and $I^{11} = [I_{11} - I_{12}(I_{22})^{-1}I_{21}]^{-1}$ is the partitioned inverse of I(θ). In order to implement the score test, I^{11} has to be evaluated. This partitioned inverse can easily be obtained once the information matrix, $I(\theta_o)$, is estimated.[8]

---

[8] see Bera and McKenzie (1986) for a discussion of alternative ways of estimating the information matrix.

5.3 Test of Heterogeneity With Uncensored Observations

When all the observations in the sample are complete, the theoretical information matrix can be used to implement the test. For a Weibull model (see Chapter 3),

$$N(I_{22})^{-1} = \begin{bmatrix} 1+m^2 q & -m\beta_1' q & -\alpha m q \\ * & \Omega^{-1}+\beta_1\beta_1' q & \alpha\beta_1 q \\ * & * & \alpha^2 q \end{bmatrix} \qquad (5.3.1)$$

where * are deduced by symmetry, $m = \phi(2)-\beta_o$, $q = 1/\phi'(1)$ and N is the sample size. $\phi(.)$ and $\phi'(.)$ are the digamma and trigamma functions respectively.

The other elements of the information matrix depend on the choice of the approximation used to specify the likelihood function. For instance, when the likelihood function is based on (5.2.4), the elements can be derived as:

$$\frac{I_{11}}{N} = 2 \quad \text{and} \quad \frac{I_{12}}{N} = [-1 \quad 0 \quad -(1/\alpha)(m+1)]. \qquad (5.3.2)$$

With (5.2.5), they are:

$$\frac{I_{11}}{N} = 5/4 \text{ and } \frac{I_{12}}{N} = [-1/2 \quad 0 \quad -(1/2\alpha)(m+2)]. \qquad (5.3.3)$$

If (5.2.6) is used, they are (see Sharma (1989)):

$$\frac{I_{11}}{N} = 1 \text{ and } \frac{I_{12}}{N} = [0 \quad 0 \quad -1/\alpha]. \qquad (5.3.4)$$

As shown in Section 2, even though the three tests are based on different approximations, $s_1(\theta_o)$ derived from the three methods is identical. Also, even though the components of the three information matrices are different, the partitioned inverse, I^{11}, derived from all three approximations is identical and is equal to 1/N(1-q).[9] For the exponential model I^{11} equals 1/N. Lancaster (1985) uses this expression to derive a test of heterogeneity. Such a test can easily be implemented when there are no censored observations in the given sample.

An interesting observation regarding the heterogeneity test for the exponential and Weibull specifications concerns the non-centrality parameter of the test. Given H_o: $\sigma^2=0$, let the sequence of alternative hypotheses be given by H_a: $\sigma^2 = \tau/N$. Then, asymptotically, the test for heterogeneity is distributed as $X^2(1,n)$[10], where the non-centrality parameter $n = (1/N)\tau^2/I^{11}$. Therefore, the non-centrality parameter for an exponential specification is τ^2 and for a Weibull specification it is $(1-q)\tau^2 \approx .39\tau^2$. Thus, for given sample size and departure from the null, the power of the heterogeneity test for an exponential specification is greater than that of a Weibull.

---

[9] Godfrey (1988) refers to the models based on such alternative approximations as locally equivalent alternatives (LEAs) with respect to the null hypothesis.

[10] see Cox and Hinkley (1974) and Chapter 4 of this thesis.

Burdett et al. (1985), following the test proposed by Kiefer (1984,1985), use sample information to calculate the variance of the $s_1(\theta)$. For example, let the mean score $(1/N)(dL/d\sigma^2)$, evaluated at $\sigma^2=0$, be $S = (1/N)\ \Sigma s_i$. The suggested variance of the mean score is $(1/N^2)\ \Sigma(s_i-s_m)^2$ where s_m is the sample mean of s. As the information matrix, using a Weibull specification, is not block diagonal, the covariance between the elements of the score vector, $dL/d\sigma^2$ and $dL/d\theta$ is ignored where $\theta = (\alpha\ \beta')'$. Therefore, the proposed variance of the test statistic is over-estimated and results in the under-rejection of the null hypothesis of no heterogeneity. Kiefer's (1985) claim that such a testing procedure is conservative, in the sense that it leads to too many rejections, cannot be true. Moreover, Kiefer recommends that a one-tailed test be used as it is based on testing for $\sigma^2=0$ against $\sigma^2>0$. This customized one-tailed test of heterogeneity can be justified when the sole possible source of misspecification is neglected heterogeneity, which in any given data may not be true.[11] The test should be considered a general misspecification test rather than being directed at testing for heterogeneity per se, hence, a two-tailed test should be

[11] see Jaggia and Trivedi (1989) and Jaggia (1990).

used.[12]

Furthermore, the performance of the Kiefer-type test is affected by the choice of the approximation of the likelihood function. As the correlation between the scores is ignored in his estimation of the variance, the test based on some approximations will be affected less than the others depending on the correlation of $s_1(\theta)$ with $s_2(\theta)$. In particular, from (5.3.4), it is seen that there is no correlation between s_1 and s_2 when the test for an exponential specification is based on the Laguerre polynomial approximation, and thus Kiefer's test is justified.[13] This, however, is not true when the same approach is used for the Kiefer's Taylor series expansion, as seen from (5.3.3). Thus, Kiefer's test, (Kiefer (1984), cannot be justified. Moreover, even if the information matrix is block diagonal, the result is based on the asymptotic distribution of the scores, which may not hold in finite samples. For a Weibull model, information matrix is not block-diagonal, even with the Laguerre polynomial approximation of the likelihood function.

For a more satisfactory test, one has to allow for the

[12] Jensen (1987) makes a similar comment regarding the use of the one-tailed test in an exponential model.

[13] Pagan and Vella (1989) and Sharma (1989) justify the use of Kiefer's procedure to test for an exponential specification.

possible correlation of $s_1(\theta)$ with $s_2(\theta)$ even though the sample information is used to derive the observed information matrix. Using the Information Matrix equality, one can use either the sample hessian of the log-likelihood function, $\Sigma(\delta^2 L / \delta\theta\delta\theta')$, or simply the outer product of the sample scores, D'D, as an estimate of $I(\theta)$. Here D is a (Nxk) matrix whose typical (i,j)th element is $dL_i/d\theta_j$, where (i=1,2...N) and $\theta_1 \ldots . \theta_k$ represent the k parameters of the model estimated under the null. When D'D is used, the score test can be easily implemented by running an artificial regression of a vector of ones on D. The test statistic is calculated as NR^2 where R^2 is the uncentred coefficient of determination derived from this artificial regression. This implementable score test is identical to the White's information matrix test as suggested by Lancaster and Chesher (1985). Even though this procedure has the merit of being easy to implement, tests based on the outer product of the sample scores approximation of the information matrix are reported to have poor small sample properties (see Davidson and MacKinnon (1983)).

5.4 Test of Heterogeneity With Censored Observations

When the data consist of complete as well as censored observations, the information matrix given above cannot be used to implement the score test of heterogeneity. Let $H_{\theta\theta} = \delta^2 L/\delta\theta\delta\theta'$ represent the sample hessian of the log-likelihood function. Consequently, $I(\theta)$ is minus the expected value of $H_{\theta\theta}$ where the expectation is taken with respect to the joint distribution of t and C conditional on the set of regressors X. Moreover, as C is a discrete random variable that assumes only 2 values, 1 and 0, the theoretical information matrix can be derived as:

$$-I(\theta) = E(H_{\theta\theta}|C=1)P(C=1) + E(H_{\theta\theta}|C=0)P(C=0) \qquad (5.4.1)$$

where $E(H_{\theta\theta}|C=1) = \int H_{\theta\theta}\, f(t|C=1)dt$ and:

$E(H_{\theta\theta}|C=0)$ is simply $H_{\theta\theta}$ evaluated at C=0 and t=L.

In order to evaluate these expressions some additional information regarding the censoring mechanism may be needed. For instance, the conditional duration distribution, $f(t|C=1)$, can be specified only if some extra information on the censoring mechanism is available.

5.4.1 When Data are Type-1 Censored

In practice, censored data commonly occur because of finite observation periods. Some events of interest may not be complete when the data acquisition period ends. In such situations, data are said to be Type-1 censored. An individual duration is known only if it is less than some pre-determined value. As $t_i = \min(T_i, L_i)$, the actual duration is observed only if $T_i < L_i$ where the L_i's are pre-determined. Notice that L_i does not have to be the same for all individuals as all individual durations may not start on the same date, even though they are all observed until a common pre-determined date. In such a situation, L_i will be different for various individuals and hence, is random. If the L_i's are known in advance for all individuals, the information matrix, (5.4.1), needed for the heterogeneity test, can be easily computed. The following result is used to derive this information matrix:

$$E(H_{\theta\theta}|C=1) = \int H_{\theta\theta}\ f(t|C=1)dt = \int H_{\theta\theta}\ f(t|T<L)dt$$

$$= \frac{1}{P(T<L)} \int_0^L H_{\theta\theta}\ f(t)dt.$$

Also, $P(C=0) = P(T>L) = 1-P(C=1) = \exp(-\mu L)$ for the exponential model and $\exp(-\mu L^{\alpha})$ for the Weibull model. Notice that such a simplification may not be of much use if censoring times, L's, are not available for all individuals. This point has special relevance when data are not Type-1 censored and the L's corresponding to complete durations are not available.

For an exponential model, the partitioned inverse component needed to compute the test is (see Appendix 5a):

$$I^{11} = [\delta_{11} - (\delta_{12})^2/\delta_{22}]^{-1}. \qquad (5.4.2)$$

where, using $s = \mu L$, the following are derived as:

$$\delta_{11} = \Sigma[2 - \exp(-s)(s^2+2s+2)],$$

$$\delta_{12} = \Sigma[(\exp(-s)(1+s) - 1] \text{ and}$$

$$\delta_{22} = \Sigma[1-\exp(-s)].$$

Note that if the censoring times, L's, go to infinity, implying no censoring, exp(-s) = 0. Hence, $I^{11} = 1/N$, as is derived for the case of all complete observations. Also, the information matrix is a function of the unknown parameters along with the L's. Thus, the performance of the test will also depend on the parameters of the model. This is unlike the case in which the test is based on all complete observations and the partioned inverse, calculated from the information matrix, is free of unknown parameters.

Similarly, the information matrix for the Weibull model can be derived in order to implement the score test for neglected heterogeneity. However, some components of the information matrix for the Weibull model involve incomplete digamma and trigamma functions and have to be evaluated numerically (see Appendix 5a).

5.4.2 When Data are not Type-1 Censored

If censoring is not of Type-1, the components of the information matrix cannot be evaluated as above. In such a situation, the observed information matrix, in principle, can be taken to implement the test of heterogeneity. Alternatively, a test is suggested that uses the theoretical information matrix based on uncensored observations only. All observations are used to find consistent and efficient estimates of the parameters but these estimates are used to evaluate scores comprising only complete observations. As shown below, it is feasible to obtain a good estimate of the theoretical information matrix corresponding to complete observations only.

Furthermore, since parameter estimates are derived from censored data, the following correction has to be made because sample scores from purely complete observations may not be zero. For notational convenience, let L represent the log-likelihood function derived from complete as well as censored observations and L^c indicate the function based on complete observations only. Also, let the notation (~) denote maximum-likelihood estimates derived from all observations that are obtained by maximising L. Therefore, for the likelihood function based only on complete observations, $s_2(\theta) = \delta L^c/\delta\theta_2|\theta_2=\tilde{\theta}_2$ may not equal zero. The standard normality results of $s_1(\theta_o) = \delta L/\delta\theta_1|\theta=\theta_o$, given by

(5.2.10), hold even when $dL/d\theta_1$ is evaluated at $\tilde{\theta}$, where $s_2(\theta) = \delta L/\delta\theta_2$ is set equal to zero to get $\tilde{\theta}$. Once $s_1(\tilde{\theta})$ is evaluated, with $s_2(\tilde{\theta}) = 0$, the results of (5.2.10) can be used to compute the score test. However, as $\delta L^c/\delta\theta_2|\theta=\tilde{\theta}$ may not equal zero, the standard normality result of $\delta L^c/\delta\theta_1|\theta=\theta_0$ cannot be used directly. One way to deal with the problem is to use L^c for estimation as well as testing. The relevant score, $\delta L^c/\delta\theta_1$, can be evaluated at θ^c where $\delta L^c/\theta_2$ is set equal to zero to get θ^c. One can discard all censored observations and treat the set of remaining observations as the relevant sample. This approach will not only result in inefficient estimates with which to evaluate the relevant scores, it will also introduce a sample selection bias that may lead to inconsistent estimates.

If all observations are to be used for estimation, $\delta L^c/\delta\theta_1|\theta=\theta^c$ can be approximated, that corrects for the sample value of $s_2(\theta)$ being non-zero at the estimated θ. Expanding $\delta L^c/\delta\theta_1|\theta=\tilde{\theta}$ around $dL^c/\theta_1|\theta=\theta^c$, using the Taylor Series expansion,[14] one gets:

$$\delta L^C/\delta\theta|\tilde{\theta} = \delta L^C/\delta\theta|\theta=\theta^C + \delta^2 L/\delta\theta\delta\theta'(\tilde{\theta} - \theta^C).$$

[14] Engle (1984) uses a Taylor's series expansion to exposit Neyman's (1959) C-alpha test. See also Chapter 4 of this thesis for a general discussion of the C-alpha type tests.

Using the fact that $\theta_1^C = \tilde{\theta}_1 = \theta_{1o}$ and $\delta L^c/\delta\theta_2|\theta_2{=}\theta_2^C = 0$, the following can be derived:

$$s_1(\theta^C) = s_1(\tilde{\theta}) - I_{12}(\tilde{\theta}) [I_{22}(\tilde{\theta})]^{-1} s_2(\tilde{\theta}). \qquad (5.4.3)$$

Using the above, the suggested test statistic is computed as:

$$LM_a = \left[s_1(\theta^C)\right]' I^{11}(\tilde{\theta}) \left[s_1(\theta^C)\right]. \qquad (5.4.4)$$

The test specified above is algebraically identical to Neyman's (1959) C-alpha test even though the motivation for this test is quite different. The C-alpha test is used when it is computationally difficult to estimate the model using maximum likelihood methods. Here, the use of maximum likelihood estimates is stressed in order to obtain consistent and efficient estimates for the evaluation of the scores derived from complete observations. In both cases, however, the correction is made because $s_2(\theta)$ is not constrained to be equal to zero at the estimated θ.

In order to implement the above mentioned adjusted test, the components of the information matrix have to evaluated. If the censored observations are such that they could have been censored randomly at any point during their potential duration, the information matrix can be found by ignoring the fact that only complete observations are being considered. In order to derive the theoretical information matrix one has to compute the expectation of $H_{\theta\theta}$ where the expectation is taken with respect to the joint distribution of T and C. Here, it is assumed that the random variable C contains no information regarding the parameters of the model and thus can be thought of as being independent of T. Therefore, the information matrix with complete observations can be evaluated as $E_t(H_{\theta\theta})\ P(C=1)$. The first component is evaluated with respect to the marginal distribution of T and P(C=1) is estimated using the proportion of complete observations in the sample. If longer durations have a higher probability of being censored than shorter ones, then such an approach cannot be justified. For example, with Type-1 censored data, the fact that an observation is complete implies that it is smaller than the length of the observation period. In such a situation one cannot assume that the distribution of T is independent of C.

To implement the above adjusted score test of heterogeneity, the information matrix for an exponential model is:

$I_{11} = 2Np, \quad I_{12} = [-Np \quad 0]$ and:

$$I_{22} = \begin{bmatrix} Np & 0 \\ 0 & Np\Omega \end{bmatrix}$$

where p is the sample proportion of complete observations.

Using (5.4.4), the adjusted score test can be computed as:

$$LM_a = \frac{1}{Np}\left[s_1(\tilde{\theta}) + s_{21}(\tilde{\theta})\right]^2 \qquad (5.4.5)$$

where $s_{21} = \Sigma[1 - \mu t]$ is the sample score w.r.t. the intercept term, β_0.

Similarly, the heterogeneity test for the Weibull model is:

$$LM_a = \frac{1}{(1-q)Np}\left[s_1(\tilde{\theta}) - I_{12}(\tilde{\theta})\ [I_{22}(\tilde{\theta})]^{-1} s_2(\tilde{\theta})\right]^2 \qquad (5.4.6)$$

where $I_{12}(I_{22})^{-1} = [mq-1 \quad -\beta_1 q \quad -\alpha q]$ and:

$$s_2(\theta^c) = \begin{bmatrix} \delta L/\delta\beta_0 \\ \delta L/\delta\beta_1 \\ \delta L/\delta\alpha \end{bmatrix} = \begin{bmatrix} \Sigma[1 - \mu t^{\alpha}] \\ \Sigma[(1 - \mu t^{\alpha})X] \\ \Sigma[1/\alpha + (1 - \mu t^{\alpha})\log(t)] \end{bmatrix}.$$

5.5 Monte Carlo Experiments

The Monte Carlo experiments presented below are used to compare the performance of the various versions of a test for heterogeneity described above for different types of censoring. These tests are compared using exponential and Weibull models. Two sets of experiments are conducted. Firstly, random and uninformative censoring, where C is independent of T, is considered. The performance of five test statistics is compared. The adjusted test, LM_a, is based on complete observations when all observations are used for estimation. The second and third tests are based on the observed information matrix. LM_s is based on the outer product of the sample scores and LM_h is based on the sample hessian of the log-likelihood function. The fourth test statistic, LM_{k1}, is Kiefer's test, based on the density approximation given in (5.2.5), and is one-tailed as suggested by Kiefer. The fifth statistic, LM_{k2}, is the same as LM_{k1} differing only in that a two tailed test is implemented.

In the second set of experiments, Type-1 censored data are considered. Both fixed and random censoring times are studied. Random censoring refers to a situation where each individual observation is associated with a different but pre-determined censoring time, L_i. For the exponential model, LM_t is a test based on the theoretical information matrix. This test is not implemented for the Weibull model where it involves evaluating incomplete digamma and trigamma functions numerically. Further, performance of tests based on the observed information matrix, including Kiefer's test, are examined for both models.

5.5.1 Design of Sampling Experiments

The size and power properties of the test statistics, mentioned above, are evaluated on the basis of different data generating processes. The heterogeneity term is multiplicative, and is log-normal. A random draw from the $N(0,\sigma^2)$ distribution is made for the 'U' representation of the heterogeneity term. Once a draw corresponding to a particular value of σ^2 is made, the vector of the heterogeneity term is held fixed for all replications and types of censoring. This practice reduces sample variability between replications and between experiments. Throughout the simulation experiments, the sample size taken is 200. The parameters (β_0, β_1) are set at (-5.0, 1.0). The variable X_1 is taken as a random draw from a uniform [0,1] distribution and is held fixed for all experiments.

5.5.2 Experiment 1

In the first set of experiments, the censoring mechanism is assumed to be completely non-informative. To incorporate heterogeneity, the value of σ^2 is taken as 0.6 for the exponential model and 0.8 for the Weibull model. To generate censored data in this set of experiments, the pre-selected observations on durations are multiplied by a vector drawn from a uniformly distributed random variable between [0,1]. This process of generating censored observations implies that censoring could have occurred at any point during the duration of a spell. For example, if the value of a uniformly distributed variable is 1/2, it implies that duration is censored at exactly half of its actual potential value. The tests are performed at 0%, 20% and 40% censoring denoting no, moderate and heavy censoring respectively. For both sets of experiments, tests are replicated three hundred times. To evaluate the size of the tests, allowance has to be made for the Monte Carlo error. For the 300 replications, the standard error of the 5% level of significance is 0.0126. Therefore, a 95 percent confidence interval around a 5% level of significance would range between 0.025 and 0.075.[15]

Results of Experiment 1

The percentage rejections of the null hypothesis at a 5% nominal level of significance for the exponential model are presented in Tables 5.1 and 5.2. Table 5.1 contains results of the tests when the data generating process conforms to

[15] The confidence interval is constructed using the normal approximation of the binomial distribution.

the null hypothesis. Table 5.2 compares the power when there is heterogeneity in the sample. Analogously, Tables 5.3 and 5.4 are used to compare the size and power of the tests respectively, for the Weibull model.

From Table 5.1 and Table 5.3, it is seen that the adjusted score test, LM_a, has good size properties. The observed number of rejections fall in the above mentioned confidence interval for all degrees of censoring for both exponential and Weibull models. For LM_s, the proportion of rejections under the null are more than the nominal level. The performance of this test gets worse as the proportion of censored observations increases. The performance of the test based on the sample hessian, LM_h, cannot be evaluated accurately as it is found that the estimated information matrix is not always non-negative definite. Kiefer's test under-rejects using a one-tailed test, and over-rejects when a two-tailed test is used for the exponential model. When the underlying model is Weibull, Kiefer's test grossly under-rejects. As mentioned earlier, because the correlation of $s_1(\theta)$ with $s_2(\theta)$ is totally ignored, the variance of $s_1(\theta)$ is overestimated resulting in under-rejection, especially in Weibull models.

From Tables 5.2 and 5.4 we see that the tests based on the observed information matrix are not as powerful as the LM_a.

Even though we drop some observations to implement this test, the effect of reducing the non-centrality parameter of the test, due to reduction in sample size, is less severe than the effect of inaccurately estimating the information matrix. The number of rejections of LM_s may seem comparable to that of LM_θ, but a comparison of the power of the two tests cannot be made as the size of the tests is different. On a size-corrected basis, the power of LM_s would be less than that of LM_θ.

5.5.3 Experiment 2

In the second set of experiments, Type-1 censored data are generated. To artificially create Type-1 censored data, $t = \min(T,L)$ is taken, where t is the observed duration. Different values for the predetermined L's are used to allow for various degrees of censoring. For the exponential model, the values taken for the fixed censoring times are 100 and 150 respectively. With the given parameter values in this Monte Carlo set up, about 20 and 35 percent censoring is generated under the null. To allow for random censoring times, $L = 120 + 20R$ is used where R is uniformly distributed on [0,1]. This method results in about 25 percent censored observations under the null. When heterogeneity is introduced to the sample, the number of censored observations increases by about 5 percent. The value of σ^2 is still set at 0.6 for the exponential model.

The performance of the tests LM_s, LM_h, LM_{k1} and LM_{k2} and LM_t, which is based on the theoretical information matrix, is compared. For the Weibull model the two fixed values used for L are 12 and 10 which generate about 25 and 35 censored observations respectively. For random censoring times, L = 10 + 2R is implemented. As with the exponential model, the number of censored observations increase by about 5 percent when heterogeneity is introduced. The value for the heterogeneity variance, σ^2, is 2.0.

Results of Experiment 2

The tests with Type-1 censored data are also replicated three hundred times. For the exponential model, the number of rejections of all test statistics falls within the confidence interval, constructed above, for all degrees of Type-1 censored data (Table 5.5). From Table 5.6, it is found that all tests seem to have comparable power though the tests based on the observed information matrix outperform the ones based on the theoretical information matrix for many experiments. However, the number of rejections, for a given value of σ^2 = .6, are a great deal less than the case when there are no censored observations (Table 5.1). This decrease in power can be explained by the non-centrality parameter of the test statistic. When all observations are complete, I_{11} = N and is independent of any nuisance parameter. For local departures from the null hypothesis, $\sigma^2 = \tau/N$, the non-centrality parameter of the test is equal to τ^2. This, however, is not the case when there is Type-1 censored data. From (5.4.2), $I_{11} = [\delta_{11} - (\delta_{12})^2/\delta_{22}]$ which depends on the parameters and censoring times, L. For the given parameter value of $\beta = [-5,1]'$, evaluated at the mean X, along with L = 100, the non-centrality parameter is $.06\tau^2$. This value is much smaller than the one derived when all observations are complete for a given sample size and departure from the null. Therefore, the power of the tests is affected by the length of the data acquisition period and becomes smaller as the period is reduced. Moreover, since the power of the tests also depends on the choice of parameter values, a more comprehensive monte-carlo analysis is desired.

For the Weibull model, LM_t is not computed as it requires numerical evaluation of incomplete digamma and trigamma functions. From Table 5.7 and 5.8, it is seen that the size of LM_s and LM_h is within the confidence limits mentioned above. However, the test is almost never rejected with LM_{k1} and LM_{k2} as in the case with random censoring seen above. When there is heterogeneity in the data, the sample hessian used to estimate the information matrix is not constrained to be positive definite. Thus, the resulting test statistic becomes meaningless sometimes owing to sampling fluctuations.[16] In this Monte Carlo simulation, the I^{11} element is negative for many replications where there is heterogeneity in the data making LM_h negative. This phenomenon results in a spuriously low number of rejections as LM_h has to be significantly high to result in the rejection of the test. The power of LM_s is reasonable considering the fact that there is Type-1 censored data. As in the exponential model, the non-centrality parameter of the test is reduced when data are Type-1 censored thus resulting in a decrease of power of the test.

[16] A classic example of such a phenomenon is Durbin's h test (see Godfrey (1988)).

5.6 Conclusion

In this chapter, various versions of the test for neglected heterogeneity are developed for censored data and are analysed using Monte Carlo experiments. Different versions of the test basically depend on the choice of estimate for the information matrix. When the data are Type-1 censored, tests based on the theoretical information matrix are developed. This method is an extension of Lancaster's (1985) tests based on complete observations. Tests based on the outer product of the sample scores perform well with Type-1 censored data. These tests actually outperform the ones based on the theoretical information matrix for some experiments.

When censoring is totally random and uninformative, as with the cases considered in this chapter, the tests based on the outer product of the sample scores do not perform well. The nominal significance of these is found to be greater than the real level, especially when the number of censored observations is large. An adjusted score testing procedure that does not suffer from this limitation and performs well for all degrees of censoring is suggested. This test is only based on the complete observations in the sample even though all observations are used to estimate the parameters of the model.

The Monte Carlo experiments show that tests based on the sample hessian of the log-likelihood function are not very useful. Since the estimated information matrix is not constrained to be positive definite it is quite frequently non-positive definite, a result that renders tests meaningless. Also, the performance of the Kiefer type test is found to be especially poor for the Weibull model.

Finally, even though the focus of this chapter has been on tests for neglected heterogeneity, the results derived in this chapter can be used quite generally. These results can be used to test for other types of misspecifications, separately or jointly with the test of heterogeneity.

TABLE 5.1

Percentage Rejections at 5% Level of Significance
For Exponential Models with no Heterogeneity

	No Censoring	20% Censoring	40% Censoring
LM_a	0.063333	0.030000	0.063333
LM_s	0.143333	0.263333	0.390000
LM_h	0.130000	0.216667	0.436667
LM_{k1}	0.006667	0.000000	0.003333
LM_{k2}	0.110000	0.250000	0.383333

TABLE 5.2

Percentage Rejections at 5% Level of Significance
For Exponential Models with Heterogeneity

	No Censoring	20% Censoring	40% Censoring
LM_a	1.000000	0.986667	0.983333
LM_s	0.980000	0.933333	0.756667
LM_h	0.506667	0.420000	0.323333
LM_{k1}	0.746667	0.630000	0.486667
LM_{k2}	0.516667	0.386667	0.236667

TABLE 5.3

Percentage Rejections at 5% Level of Significance
For Weibull Models with no Heterogeneity

	No Censoring	20% Censoring	40% Censoring
LM_a	0.036667	0.073333	0.076666
LM_s	0.093333	0.143333	0.160000
LM_h	0.013333	0.026667	0.023333
LM_{k1}	0.000000	0.000000	0.000000
LM_{k2}	0.000000	0.000000	0.000000

TABLE 5.4

Percentage Rejections at 5% Level of Significance
For Weibull Models with Heterogeneity

	No Censoring	20% Censoring	40% Censoring
LM_a	0.846667	0.836667	0.750000
LM_s	0.843333	0.796667	0.733333
LM_h	0.040000	0.060000	0.050000
LM_{k1}	0.000000	0.000000	0.000000
LM_{k2}	0.000000	0.0000000	0.000000

TABLE 5.5

Percentage Rejections at 5% Level of Significance
For Exponential Model with no Heterogeneity

	L = 150	L = 100	L = 120 + 20R
LM_t	0.060000	0.040000	0.033333
LM_s	0.073333	0.050000	0.050000
LM_h	0.073333	0.080000	0.050000
LM_{k1}	0.033333	0.050000	0.023333
LM_{k2}	0.056667	0.046667	0.033333

TABLE 5.6

Percentage Rejections at 5% Level of Significance
For Exponential Model with Heterogeneity

	L = 150	L = 100	L = 120 + 20R
LM_t	0.676667	0.483333	0.670000
LM_s	0.700000	0.533333	0.703333
LM_h	0.760000	0.636667	0.793333
LM_{k1}	0.753333	0.593333	0.783333
LM_{k2}	0.663333	0.493333	0.676667

TABLE 5.7

Percentage Rejections at 5% Level of Significance
For Weibull Model with no Heterogeneity

	L = 12	L = 10	L = 10 + 2R
LM_s	0.080000	0.053333	0.046667
LM_h	0.093333	0.080000	0.070000
LM_{k1}	0.000000	0.000000	0.000000
LM_{k2}	0.003333	0.000000	0.003333

TABLE 5.8

Percentage Rejections at 5% Level of Significance
For Weibull Model with Heterogeneity

	L = 12	L = 10	L = 10 + 2R
LM_s	0.673333	0.503333	0.613333
LM_h	0.273333	0.163333	0.236667
LM_{k1}	0.190000	0.060000	0.150000
LM_{k2}	0.056667	0.016667	0.026667

Appendix 5a

In order to derive the information matrix when data are Type-1 censored, the following results regarding generalised errors, ϵ are needed. These results are based on the fact that ϵ,under the null, is distributed unit exponentially.

$$\int_0^s \epsilon \exp(-\epsilon)d\epsilon = 1 - \exp(-s)(1+s). \qquad (5A.1)$$

$$\int_0^s \epsilon^2 \exp(-\epsilon)d\epsilon = 2 - \exp(-s)(s^2+2+2s). \qquad (5A.2)$$

$$\int_0^s \epsilon^3 \exp(-\epsilon)d\epsilon = 6 - \exp(-s)(s^3+3s^2+6s+6). \qquad (5A.3)$$

$$\int_0^s \epsilon^4 \exp(-\epsilon)d\epsilon = 24 - \exp(-s)(s^4+4s^3+12s^2+24s+24). \qquad (5A.4)$$

$$\int_0^s \epsilon^q \log \epsilon \exp(-\epsilon)d\epsilon. \qquad (5A.5)$$

$$\int_0^s \epsilon^q (\log \epsilon)^2 \exp(-\epsilon)d\epsilon. \qquad (5A.6)$$

(5A.5) and (5A.6) are the incomplete digamma and trigamma functions respectively, of order q, that have to be evaluated numerically.

Exponential Model

The information matrix, using the likelihood function defined in (5.2.8) with α=1, can be derived as:

$$-E(\delta^2 L/\delta\sigma^2 \delta\sigma^2) = \Sigma[(1/4)\ E(\epsilon^4 + 4C\epsilon^2 - 4C\epsilon^3)]. \qquad (5A.7)$$

Here, using (5.4.1),

$$E(\epsilon^4+4C\epsilon^2-4C\epsilon^3) = E(\epsilon^4+4\epsilon^2-4\epsilon^3|C=1)P(C=1) + E(\epsilon^4|C=0)P(C=0)$$

where $P(C=0) = P(T>L) = \exp(-\mu L) = 1-P(C=1)$ and

$E(\epsilon^4|C=0) = (\mu L)^4$.

Therefore, (A.7)

$$= \frac{1}{4}\ \Sigma[\int_0^L (\epsilon^4+4\epsilon^2-4\epsilon^3)f(t)dt \quad + \quad (\mu L)^4\exp(-\mu L)]$$

$$= \frac{1}{4}\ \Sigma[\int_0^s (\epsilon^4+4\epsilon^2-4\epsilon^3)f(\epsilon)d\epsilon + s^4\exp(-s)],\ \text{where } s = \mu L.$$

Using (5A.1) to (5A.4), it is further simplified as:

$$= \delta_{11} = \Sigma[2 - \exp(-s)(s^2+2s+2)] = I_{11}. \qquad (5A.8)$$

Similarly, the other components of the information matrix can be found as:

$$-E(\delta^2 L/\delta\sigma^2 \delta\beta') = \Sigma[E(C\epsilon-\epsilon^2)X] = \Sigma[E(C\epsilon-\epsilon^2)E(X)]$$

$$= [\delta_{12} \quad 0] = I_{12} \qquad (5A.9)$$

where $\delta_{12} = \Sigma[(\exp(-s)(1+s) - 1]$.

$$-E(\delta^2 L/\delta\beta\delta\beta') = \Sigma[\epsilon(X'X)] = \Sigma[E(\epsilon)E(X'X)] = I_{22}$$

$$= \begin{bmatrix} \delta_{22} & 0 \\ 0 & \delta_{22}\ \Omega \end{bmatrix}. \tag{5A.10}$$

where $\delta_{22} = \Sigma[1-\exp(-s)]$.

The expressions, (5A.8) - (5A.10) can be used to compute the partitioned inverse needed to compute the heterogeneity test. As $I^{11} = [I_{11} - I_{12}(I_{22})^{-1}I_{21}]^{-1}$, it is derived as:

$$I^{11} = [\delta_{11} - (\delta_{12})^2/\delta_{22}]^{-1}. \tag{5A.11}$$

Weibull Model

The above results given for the exponential model can be used for the Weibull model with $s=\mu t^{\alpha}$. However, in addition to these the following has to be derived:

$$-E(\delta^2 L/\delta\sigma^2\,\delta\alpha) = \frac{1}{\alpha}\,\Sigma[(C\epsilon - \epsilon^2)\ (\log(\epsilon) - X\beta)]$$

$$= \frac{1}{\alpha}\,\Sigma[\beta_0\exp(-s)(1+s) - \beta_0 - s^2\log(s)\exp(-s)$$

$$+ \int_0^s (\epsilon-\epsilon^2)\log(\epsilon)\exp(-\epsilon)\ d(\epsilon)]. \tag{5A.12}$$

The last expression, from (5A.5), is an incomplete gamma function that has to be evaluated numerically. Similarly $-E(\delta^2 L/\delta\alpha^2)$ involves evaluating an incomplete trigamma function.

CHAPTER 6

SUMMARY AND CONCLUSION

In this thesis, the consequences of multiple misspecifications, concurrently existing in the data, on tests for parametric duration models are examined. Most specification tests found in the econometrics literature have been constructed in such a way as to ascertain the validity of only one specification at a time. Such a procedure implies making auxiliary assumptions, in addition to the assumptions being tested, each time a test is carried out. It is argued that such separate tests are not generally robust in the presence of other misspecifications, and hence can lead to erroneous conclusions.

The problems encountered 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 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. For example, the problem of separate tests persists when the underlying model is generalised gamma, and heterogeneity is tested for in an exponential or a Weibull model.

As the joint presence of more than one source of misspecification cannot be a priori ruled out, omnibus test statistics that have power against several forms of misspecification are stressed. In Chapter 3, such tests are motivated and developed for exponential and Weibull models. Tests based on parametric alternatives are derived within a heterogeneous gamma model. Such tests are score tests or simply tests of moment restrictions of appropriately defined generalised errors. For example, a score test of no heterogeneity is equivalent to a test of a second moment restriction on ϵ where the integrated hazard, ϵ, is the generalised error in the sense of Cox and Snell.

The above tests, based on a specified parametric alternative, have a limitation in that they place restrictions on the alternative and hence may not have good properties when the specified alternative is also incorrect. Here, it is suggested that a joint test of all moment restrictions of ϵ should be used rather than a test of the second moment only. Such a test has the merit of being based on an unspecified alternative and thus is not restrictive in any way. Given any parametric model, the integrated hazard function has a unit exponential distribution under the null and hence, if the model is correctly specified, its moment restrictions must be satisfied. Tests of higher order moment restrictions of ϵ are derived using the Newey-Tauchen framework.
An empirical application, using Kennan's strike data, is

provided as illustration for all the above mentioned tests. It is found that when separate tests are implemented, results indicate that restrictive exponential and Weibull models are adequate. However, when joint tests for misspecification are implemented, both exponential and Weibull models are found to be inadequate. 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 be present in any given model. Further, it is shown that erroneous conclusions may be reached, as in a number of previous analyses of Kennan's strike data, if standard separate tests are implemented.

A known limitation of a joint test is that if the null hypothesis is rejected, one still needs information from separate tests to indicate the nature of the required respecification. In order to have a valid separate test for any one misspecification, all auxiliary (nuisance) parameters have to be estimated. The maximum-likelihood estimation of nuisance parameters can sometimes be computationally cumbersome. Thus, Neyman's C-alpha type tests are motivated and developed in Chapter 4. These tests may be based on any root-N consistent estimates of these parameters and are asymptotically equivalent to separate tests that are based on maximum likelihood estimates. An example of such a test for neglected heterogeneity, that is based on the consistent estimates using least squares, is provided. The Monte Carlo results indicate that such a test performs well.

A problem arises when any root-N estimates of nuisance parameters are also not available. Here, joint and separate tests have to be implemented under the joint null hypothesis. Bera and Jarque (1982) deal with a similar situation in the context of a classical linear regression model. They suggest Multiple Comparison Procedures in order to identify different sources of errors when the joint test results in the rejection of the null hypothesis. Their procedure, however, is aided by the additivity of their separate tests, implying that all tests they derive are asymptotically independent under the joint null. So, even when multiple misspecifications exist, each separate test is somewhat informative in carrying out a search for an appropriate model.

In the given context, the tests of heterogeneity and duration dependence are shown to be correlated with each other within a heterogeneous Weibull model. Each separate test, therefore, provides no useful information regarding a particular misspecification. The conditional score tests suggested in this chapter may be used to attain such information. Even though the suggested conditional score tests are based on the restricted joint null, each score is purged of any correlation with the other relevant score. Therefore, the conditional score tests derived this way may contain more information regarding the specific source of misspecification than the standard separate tests. These conditional tests can be used to locate the exact source of misspecification using the Multiple Comparison Procedure; see Savin (1980,1984) for a survey of Multiple Comparison Procedures. The Monte Carlo experiments conducted in this chapter suggest that conditional separate score tests do in fact contain more information than the standard separate tests.

All the tests suggested in Chapters 3 and 4 are based on the theoretical (expected) information matrix. Such a matrix is hard to obtain when there are censored observations in the sample. In principle, one may use the observed information matrix to implement such tests. Tests based on the observed information matrix are shown to perform poorly for many of the simulation experiments conducted in Chapter 5.

Using the test of heterogeneity as an example, alternative methods are suggested where more information from the null hypothesis is used to implement tests in the presence of censored observations. Even though the focus of this chapter is on tests for neglected heterogeneity, the results derived can be used quite generally. These results can be used to test for other types of misspecifications separately or jointly with the test of heterogeneity.

When the data are Type-1 censored, tests based on the theoretical information matrix are developed. This method is an extension of Lancaster's (1985) tests based on complete observations. With Type-1 censored data, the test is shown to depend on the chosen observation period. Using the non-null distribution of this test, it is shown that the power of tests based on the censored observations is reduced if the length of the observation period is decreased. The Monte Carlo results exhibit this reduction in power. For this type of censoring, tests based on the observed information matrix also perform well.

When data are not Type-1 censored and censoring is totally random and uninformative, the tests based on the observed information matrix do not perform well. An adjusted score testing procedure that does not suffer from this limitation and performs extremely well for all degrees of censoring is suggested. This test is only based on the complete observations in the sample even though all observations are used to estimate the parameters of the model.

REFERENCES

Bera, A.K. and C.M. Jarque (1982), "Model Specification Test: A Simultaneous Approach", Journal of Econometrics, 20, 59-82.

Bera, A.K. and C.M. Jarque (1987), "Additivity and Separability of the Lagrange Multiplier, Likelihood Ratio and Wald Tests", Journal of Quantitative Economics, 3(1), 55-63.

Bera, A.K. and C.R. McKenzie (1986), "Alternative Forms and Properties of the Score Test", Journal of Applied Statistics, 13, 13-25.

Breusch, T. S. and A. R. Pagan (1980), "Lagrange Multiplier Test and Its Applications to Model Specification In Econometrics", Review of Economic Studies, 47, 239-253.

Burdett, K., N.M. Kiefer, and S. Sharma (1985), "Layoffs and Duration Dependence in a model of Turnover", Journal of Econometrics, 28, 51-62.

Chesher, A.D. (1983), "The Information Matrix Test: Simplified Calculation Via a Score Test Interpretation", Economics Letters, 13, 45-48.

Chesher, A.D. (1984), "Testing for Neglected Heterogeneity", Econometrica, 52, 865-872.

Cox, D. R. (1961), "Tests of Separate Families of Hypotheses", in Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, 105-123, Berkeley: University of California.

Cox, D. R. (1962), "Further Results on Test of Separate Families of Hypotheses", Journal of the Royal Statistical Society, B, 24, 406-424.

Cox, D.R. (1983), "Some Remarks on Overdispersion", Biometrica, 70, 269-274.

Cox, D.R. and D.V. Hinkley (1974), Theoretical Statistics, Chapman-Hall, London.

Cox, D.R. and E.J. Snell (1968), "A general definition of residuals", Journal of the Royal Statistical Society B, 30, 248- 275.

Davidson, R and J.G. MacKinnon (1983), "Small Sample Properties of Alternative Forms of the Lagrange Multiplier Test", _Economics Letters_, 12, 269-275.

Efron, B. and D.V. Hinkley (1978), "Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information", _Biometrika_, 65, 457-482.

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.

Engle, R.F. (1982), "A General Approach to Lagrange Multiplier Model Diagnostics", _Journal of Econometrics_, 20, 83-104.

Glaser, R.E. (1980), "Bathtub and Related Failure Rate Characterizations", _Journal of the American Statistical Association_, 75, 667-672.

Godfrey, L.G. (1988), _Misspecification Tests in Econometrics: The Lagrange Multiplier Principle and Other Approaches_, Cambridge: Cambridge University Press.

Godfrey, L.G. and M.R. Wickens (1981), "Testing Linear and Log- Linear Regressions for Functional Forms", _Review of Economic Studies_, 48, 487-496.

Heckman, J. and B. Singer (1982), "The Identification Problem in Econometric Models for Duration Data", in _Advances in Econometrics_, ed. W. Hildenbrand, 39-77, Cambridge University Press.

Heckman, J. and B. Singer (1984a), "A Method for minimizing the impact of Distributional Assumptions in Econometric Models for Duration Data", _Econometrica_, 52, 271-320.

Heckman, J. and B. Singer (1984b), "The identifiability of the Proportional Hazard Model", _Review of Economic Studies_, 51, 231-241.

Heckman, J. J. and J. R. Walker (1987), "Using Goodness of Fit and Other Criteria to Choose Among Competing Duration Models: A Case Study of Hutterite Data", _Sociological Methodology 1987_, 247-307.

Holly, A. (1987), "Specification Tests : An Overview", Chapter 2 in T. Bewley (Ed), Advances in Econometrics - Fifth World Congress, vol. I. Cambridge: Cambridge University Press.

Horowitz, J.L. and G.R. Neumann (1989), "Specification Testing in Censored Regression Models: Parametric and Semiparametric Methods", Journal of Applied Econometrics, 4(S), S61-S86.

Hougaard, P. (1984), "Life table methods for heterogeneous populations: Distributions describing the heterogeneity", Biometrica, 71, 75-83.

Hougaard, P. (1986), "Survival models for heterogeneous populations derived from stable distributions", Biometrica, 73, 387-396.

Jaggia, S. (1990), "Specification Tests based on the Heterogeneous Generalised Gamma Model of Duration: With an Application to Kennan's Strike Data", resubmitted to The Journal of Applied Econometrics.

Jaggia, S. and P.K. Trivedi (1989), "Joint and Separate Score Tests for State Dependence and Unobserved Heterogeneity", presented at the North American Meetings of Econometric Society, Univerisity of Michigan, Ann Arbor, June 21-24, 1989.

Jaggia, S. and Satish Thosar (1989), "Contests for Corporate Control: A Survival Analysis", unpublished manuscipt.

Jensen, P. (1987), "Testing for Unobseved Heterogeneity and Duration Dependence in Econometric Duration Models", University of Aarhus, unpublished paper.

Kalbfleish, J.D. and R.L.Prentice (1980), The Statistical Analysis of Failure Time Data, New York: Wiley.

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. (1985), "Specification Diagnostics Based on Laguerre Alternatives for Econometric Models of Duration", Journal of Econometrics, 28, 135-154.

Kiefer, N.M. (1988), "Econometric Duration data and Hazard Functions", Journal of Economic Literature, 25, 646-679.

Lancaster, T. (1979), "Econometric Methods for the Duration of Unemployment", Econometrica, 47, 939-956.

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

Lancaster, T. (1984), "Notes and Comments: The Covariance Matrix of the Information Matrix Test", Econometrica, 52, 1051-1053.

Lancaster, T. (1985), "Generalized Residuals and Heterogeneous Duration Models: With Applications to the Weibull Model", Journal of Econemetrics, 28, 155-169.

Lancaster, T. and A. 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", University of Hull manuscript.

Lancaster, T. and S. Nickel (1980), "The Analysis of Re-Employment Probabilities for the Unemployed", Journal of the Royal Statistical Society A, 141-165.

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

Manton, K.G., E. Stallard, and J.W. Vaupel (1986), "Alternative Models for the Heterogeneity of Mortality Risks Among the Aged", Journal of the American Statistical Association, 81, 635-644.

Moon, C.G. (1988), "Simultaneous Specification Test in a Binary Logit Model", Communications in Statistics - Theory and Methods, 17(10), 3361-3387.

Moran, P. A. P. (1970), "On Asymptotically Optimal Tests of Composite Hypothesis", Biometrika, 57, 47-55.

Newey, W.K. (1985), "Maximum Likelihood Specifiaction Testing and Conditional Moment Tests", Econometrica, 53, 1047-1070.

Neyman, J. (1959), "Optical Asymptotic Tests of Composite Statistical Hypotheses", in Probability and Statistics, ed. U. Grenander, 213-234, Ney York: Wiley.

Pagan, A.R. and A.D. Hall (1983), "Diagnostic Tests as Residual Analysis", Econometric Reviews, 2, 159-218.

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.

Pagan, A. R. and F. Vella (1989), "Diagnostic Tests For Models Based on Individual Data: A Survey", Journal of Applied Econometrics, 4(S), S29-S59.

Pereira, B. de B. (1978), "Tests and Efficiencies of Separate Regression Models", Biometrika, 65, 319-27.

Ridder, G. and W. Verbaekel (1983), "On the Estimation of the Proportional Hazard Model in the Presence of Unobserved Heterogeneity", Faculty of Actuarial Science and Econometrics, Report AE, 22/83, University of Amsterdam.

Savin, N. E. (1980), "The Bonferroni and Scheffe Multiple Comparison Procedures", Review of Economic Studies, 67, 255- 273.

Savin, N.E. (1984), "Multiple Hypothesis Testing", in Handbook of Econometrics, Vol 2, Z. Griliches and M.D. Intriligator, eds. Elsevier Science, New York.

Salant, S.W. (1977), "Search Theory and Duration Data: A Theory of Sorts", Quarterly Journal of Economics, 91, 39-57.

Sharma, S. (1989), "Specification Diagnostics for Econometric Models of Duration", unpublished UCLA working paper.

Tauchen, G.E. (1985), "Diagnostic Testing and Evaluation of Maximum Likelihood Models", Journal of Econometrics, 30, 415-443.

Trussell, J. and T. Richards (1985), "Correcting For Unmeasured Heterogeneity in Hazard Models Using The Heckman-Singer Procedure", Sociological Methodology 1985, 242-276.

Vaupel, J.W., K.G. Manton and E. Stallard (1979), "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality", Demography, 16, 439-454.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models", Econometrica, 50, 1-25.

White, H (1987), "Specification Testing in Dynamic Models", in T. Bewley ed. Advances in Econometrics - Fifth World Congress, Vol. 1, 1-58, New York: Cambridge University Press.

Wooldridge, J.M. (1989a), "A C(α) Version of the Newey-Tauchen-White Test", unpublished MIT working paper.

Wooldridge, J.M. (1989b), "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances", unpublished working paper.

CURRICULUM VITAE

SANJIV JAGGIA

Department of Economics
901 Ballantine Hall
Indiana University
Bloomington, IN 47405
(812) 855-1021.

Home Address
Bicknell 100
Bloomington, IN 47406
(812) 857-8925.

PERSONAL

Birthdate: January 13, 1960
Marital Status: Married
Birth Place: India
Visa Status: Permanent Resident of the U.S.

AREAS OF INTEREST

Teaching: Undergraduate and Graduate Statistics and Econometrics; Undergraduate Theory of Economics and International Trade.
Research: Issues in General Specification Testing: Theory and Applications.

EDUCATION

Indiana University	Economics	Ph.D (1990)
Panjab University, India	Economics	M.A. (1982)
Panjab University, India	Economics	B.A. (1981)

DISSERTATION

Topic: "Specification Diagnostics in the Presence of Multiple Misspecifications for Parametric Duration Models."

Synopsis: The thesis is aimed primarily at identifying various sources of misspecification and testing for them in the context of duration models. It is emphasised that the conventional approach of testing each parametric restriction in isolation is inconclusive when multiple misspecifications exist concurrently. Incorrect inferences may be drawn in such situations, especially when separate tests are correlated. The need to compute an omnibus statistic is stressed and some joint tests of this form are developed. Such a statistic tests all the relevant assumptions made within a given model jointly and hence has power against several forms of misspecification. Furthermore, modified separate tests are developed that can provide the additional information needed to pin-point the exact error, once the joint null hypothesis is rejected. Finally, some adjustments to standard tests, that are valid in the presence of censored data as well, are suggested in this thesis. An empirical application and extensive Monte Carlo evidence are provided as illustration for all the above mentioned tests.

Thesis Chairman: Professor Pravin K. Trivedi

HONORS

Received the Taulman Miller Prize as the outstanding international student in economics (1987).

EXPERIENCE

Indiana University — Associate Instructor — 1984-90
Independently taught Principles of Micro Economics (E103) three times and Statistics (E270) nine times.

World Bank — Research Assistant — 1988-89
Participated in the construction of a simulation experiment to evaluate pricing and tax policies of perennial crops in developing countries.

WORKING PAPERS

"Joint and Separate Score Tests for State Dependence and Unobserved Heterogeneity", Co-authored with Pravin Trivedi, resubmitting to The Journal of Econometrics.

"Specification Tests based on the Heterogeneous Generalised Gamma Model of Duration: With an Application to Kennan's Strike Data", resubmitted to The Journal of Applied Econometrics.

"Contests for Corporate Control: A Survival Analysis", Co-authored with Satish Thosar, presenting at the FMA meetings in October, 1990.

PROFESSIONAL MEMBERSHIPS

The Econometric Society
American Economic Association

REFERENCES

Professor Pravin K. Trivedi
Department of Economics
901 Ballantine Hall
Indiana University
Bloomington, IN 47405
(812) 855-3567

Professor Esfandiar Maasoumi
Department of Economics
Southern Methodist University
Dallas, Texas 75275
(214) 692-4298

Professor R. Jeffery Green
Department of Economics
901 Ballantine Hall
Indiana University
Bloomington, In 47405
(812) 855-7835

Professor Phillip Saunders
Department of Economics
901 Ballantine Hall
Indiana University
Bloomington, In 47405
(812) 855-7808