Group Duration Analysis of the Proportional Hazard Model: Minimum Chi-square Estimators and Specification Tests

by

Keunkwan Ryu *

Working Paper Number 669
Department of Economics
University of California, Los Angeles
September 1992

* The author is grateful to Sunil Sharma for helpful discussions.

ABSTRACT

Due to a discrete observation scheme, many duration variables are available only up to an interval, and not up to an exact point. The proportional hazard model is the most widely used continuous time duration models. Interval observations cause several difficulties in applying Cox's partial maximum likelihood estimation method to the proportional hazard model. First, two different death times reported to fall into the same interval are observationally equivalent. In Cox's procedure, there is no natural way of handling ties. Second, the nuisance part of the baseline hazard rate cannot be conditioned out of the estimation process. This paper develops semi-parametric minimum Chi-square estimators of the proportional hazard model for the case when durations are grouped and covariates are categorical. This paper also suggests simple specification tests. Monte Carlo simulations demonstrate the performance of the proposed estimators and the size and power properties of the proposed specification tests.

Unemployment duration, for example, is at most known only up to an interval of a week. To identify factors affecting the unemployment duration, the proportional hazard model has often been used. Hence, in an unemployment duration analysis, we face a situation with discrete data and a continuous model. If a worker's observed characteristics can be captured as a set of categorical covariates, an application of Berkson's minimum Chi-square estimation to the grouped unemployment data yields a computationally simple estimator which is asymptotically as efficient as the maximum likelihood estimator. By comparing the two sets of minimum Chi-square estimators, one from the original grouped data set and the other from a further grouped data set, we can test whether the proportionality assumption holds. Under proportionality, the two estimators should converge to the same quantity; however, if proportionality is violated, they should converge to different quantities.

KEY WORDS: Minimum Chi-square estimation; Semi-parametric estimation; Seemingly Unrelated Regressions; Hausman's specification test; Binary choice model; Partial maximum likelihood estimation.

1. INTRODUCTION

Many duration variables are often reported to fall within some intervals, and not up to exact points, due to a discrete observation scheme. Unemployment duration, for example, is at most known only up to an interval of length l=one week. However, the underlying model often assumes a continuous time framework, in which the underlying duration variable can take any continuous non-negative real value. Hence, in duration analysis, we face a situation with discrete data and a continuous model.

The proportional hazard model (PHM) is one of the most widely used continuous time duration models (Cox 1972, 1975, Cox and Oakes 1984, Kalbfleisch and Prentice 1980). Under PHM, the hazard rate of a duration is specified as a product of two separate terms: the baseline hazard describing the overall shape of the hazard rate over time, and the proportionality factor capturing the covariate (regression) effects on hazard rates across different individuals. Cox developed a semi-parametric estimation technique which can identify the covariate effects efficiently without specifying the functional form of the baseline hazard. This nuisance part of the baseline hazard is conditioned out of the estimation process in his suggested estimation procedure. One can condition the baseline hazard out of the estimation process if one knows who and how many were at risk at each moment in time. This estimation is called partial maximum likelihood estimation (PMLE), since the maximand of PMLE is equivalent to the likelihood contribution of only the relative ranks of durations on the regression coefficients.

Interval observations render it impossible to apply Cox's PMLE to his PHM. First, too many ties arise. Two death times reported to fall into the same interval are observationally equivalent. In the PMLE procedure, there is no natural way of handling ties. This problem becomes more severe as the interval length l becomes wider. Second, since we do not know the exact death time, we cannot condition the nuisance part of the baseline hazard rate out of the estimation process unless it is constant across each interval in the observation scheme. In addition, we are often interested in the changing pattern of the baseline hazard over time as well as in the regression coefficients. In this regard, positive and negative duration dependence mean increasing and decreasing baseline hazard function, respectively.

We call PHM parametric or semi-parametric, respectively, depending on whether the baseline hazard is parametrically specified or not. Of course, we need a parametric specification regarding the proportionality factor. Grouped duration can be viewed as a sequence of independent binary dependent variables indicating whether the duration survives each interval or not. By constructing a set of synthetic binary data treating each combination (individual, interval) as a new unit of indexing, we can reduce a grouped duration analysis to a binary choice analysis. Here, PHM implies a certain functional form for the binary choice probabilities. If we have many observations for each value of the covariate vector (often called, many observations per cell), application of Berkson's minimum Chi-square estimation yields a computationally simple estimator which is asymptotically as efficient as the maximum likelihood estimator. In this paper, we fully develop the minimum Chi-square estimation of PHM when the duration data are grouped and the covariates are categorical. By comparing the two sets of minimum chi-square estimators, one from the original grouped data

set and the other from a further grouped data set, we also develop simple specification tests of PHM. Under PHM, the two estimators converge to the same quantity; however, if proportionality is violated, they converge to different quantities. The suggested tests are easy to use and can take alternative hypotheses into account. Using Monte-carlo simulations, we illustrate the performance of the semi-parametric minimum Chi-square estimators and the size and power properties of the proposed specification tests.

The rest of the paper is organized as follows. In section 2, after setting up a framework for discussion, a relationship between group duration analysis and binary choice analysis is shown. Estimation and identification issues are discussed in section 3. Section 4 develops the minimum Chi-square estimation of PHM when durations are grouped and covariates are categorical. Simple specification tests of PHM are developed in sections 5 and 6. Through Monte-carlo simulations in section 7, the performance of the proposed estimators and the size and power

properties of the proposed tests are reported. Concluding remarks follow in section 8.

2. FRAMEWORK, RELATION BETWEEN GROUP DURATION AND BINARY CHOICE

First, let us introduce a framework for further discussion. Let $T \in R^+$ represent a duration variable of interest. Let $h(t,x) = h_0(t)\exp(x\beta)$ be the hazard rate of the duration T, where x is an $1 \times k$ vector of covariates, and β a $k \times 1$ vector of regression coefficients.

Quite often, the discrete observation scheme can be represented as an equi-spaced partition Q of the support R^+: $Q = \{0, l, 2l, \cdots, rl, \infty\}$. The underlying observation process can be regarded as follows: the experiment is conducted starting at $t = 0$ until $t = rl$, where each sample is recorded at each time point jl, $j = 1, \cdots, r$. Under this observation scheme, the researcher keeps a record of the objects' status at every l time units, until time rl elapses. Let $I_1 = [0, l), \cdots, I_r = [(r-1)l, rl), I_{r+1} = [rl, \infty)$. Under the observation scheme Q, we have r non-right-censored intervals $I_1, \cdots, I_r$, and a single right-censored interval I_{r+1}. For each observation falling within a non-right-censored interval, we know its duration up to an interval of length l; for each observation falling within the right-censored interval, we only know that its duration exceeds a certain lower bound (here, rl).

In the sequel, we will assume, for expositional simplicity, that $l = 1$ and $r = 2$. The resulting observation scheme is $Q = \{0, 1, 2, \infty\}$. Here, $I_1 = [0, 1)$ and $I_2 = [1, 2)$ are two non-right-censored intervals. $I_3 = [2, \infty)$ is the single right-censored interval. Durations are available only up intervals I_1, I_2, and I_3. Let α_1 be the probability that the duration T survives interval I_1, and let α_2 be the conditional probability that T survives I_2 conditional on that it has already survived I_1. Then by the above proportional hazard assumption, we have

$$\begin{aligned}\alpha_1 &= \exp[-\int_0^1 h(t,x)\,dt] \\ &= \exp[-\exp(x\beta + \gamma_1)]\end{aligned}$$

and

$$\begin{aligned}\alpha_2 &= \exp[-\int_1^2 h(t,x)\,dt] \\ &= \exp[-\exp(x\beta + \gamma_2)]\end{aligned}$$

where

$$\gamma_1 = \log[\int_0^1 h_0(t)\,dt]$$

and

$$\gamma_2 = \log[\int_1^2 h_0(t)\,dt].$$

These formulas are originally available in Prentice and Gloeckler (1978).

To take into account observed heterogeneity among individual duration variables, let i index each different observation. Assume that observations are independent. When we have n observations, i takes on an integer value from 1 through n. Accordingly, define T_i as the ith duration variable, x_i as the $1 \times k$ covariate vector of individual i, α_{1i} and α_{2i} as α_1, α_2, respectively, evaluated at $x = x_i$, $d_{1i} = 1$ if T_i survives I_1, $d_{1i} = 0$ otherwise, and $d_{2i} = 1$ if T_i survives I_2 conditional on $d_{1i} = 1$, $d_{2i} = 0$ otherwise. Under the observation scheme Q, the ith observation can be summarized as a triple (x_i, d_{1i}, d_{2i}), where d_{1i} and d_{2i} are two (r=number of non-right-censored intervals in Q) binary dependent variables indicating whether the individual survives the first and the second interval, respectively.

A grouped duration can be considered as a sequence of binary dependent variables. The effective number of terms in the sequence varies depending on whether the individual dies during the first interval or not. Note that d_{2i}'s are meaningfully defined only for those which have survived the first interval I_1. By constructing a synthetic binary data set treating each combination (individual, interval) as a new unit of indexing, we can reduce a grouped duration analysis to a binary choice analysis (Kiefer 1988, Prentice and Gloeckler 1978, Sueyoshi 1991, Thompson 1977). For each combination (interval, individual), a survivor of the jth interval independently receives the probability $\alpha_j = F(x\beta + \gamma_j)$ if he or she has covariate vector x, where $F(\cdot)$ is defined by $F(y) = \exp(-\exp(y))$. The range of $F(\cdot)$ is between zero and one, satisfying a necessary condition to be a probability. Recall that in Logit binary choice model, $F(\cdot)$ takes the form $F(y) = 1/(1 + \exp(y))$. In fact, Thompson (1977) adopted this parametrization to make group duration analysis directly comparable to Logit analysis.

Under Thompson's specification, we have

$$\log \frac{1 - \alpha_j}{\alpha_j} = x\beta + \gamma_j;$$

while here under PHM, we have

$$\log(-\log \alpha_j) = x\beta + \gamma_j.$$

The two specifications require quite different transformations of α_j on the left hand side to obtain linear regression type equations. However, if α_j is near one, both transformations are approximately equal

$$\frac{1-\alpha_j}{\alpha_j} \approx -\log \alpha_j.$$

If the interval length is small and thus if the corresponding interval survival probability α_j is near one, group duration analysis of PHM can be carried out through Logit analysis simply by defining the afore-mentioned synthetic binary data set. On the other hand, if the group interval is rather wide and the interval survival probability is quite different from one, we need to use $\alpha_j = F(x\beta + \gamma_j)$ with $F(y) = \exp(-\exp(y))$. Instead, if we use the Logistic counterpart for group duration analysis of PHM, interpretation of the coefficient β becomes ambiguous, bearing no relation

with the underlying PHM.

3. ESTIMATION AND IDENTIFICATION ISSUES

Without making a parametric functional form assumption on the baseline hazard $h_0(\cdot)$, we can still estimate the integrated baseline hazard over each of these r non-right-censored intervals (here, $r = 2$). In our simplified setting, we are able to estimate γ_1 and γ_2, or equivalently $\exp(\gamma_1) = \int_0^1 h_0(t)\,dt$ and $\exp(\gamma_2) = \int_1^2 h_0(t)\,dt$. This means that we can approximate the unknown functional form of $h_0(\cdot)$ up to a step function with two different values (in general, a step function with r different values, insofar as r is either finite or increasing at a rate slower than the number of observations.). This is called a semi-parametric estimation of PHM.

Once we parameterize $h_0(\cdot)$ using m free parameters, we have a fully parametric duration model. One natural selection is

$$h_0(t) = \exp(\alpha_0 + \alpha_1 t + \cdots + \alpha_{m-1} t^{m-1}).$$

The exponential distribution is an example with $m = 1$. If the number m of free parameters is equal to the number r of non-right-censored intervals (in our example, $r = 2$), then both the non-parametric and parametric baseline specifications would yield the same results. This is because of the invariance property of the extremum estimators (estimators defined as the arguments either maximizing or minimizing a certain objective function: for instance, maximum likelihood, least squares, or minimum Chi-square estimators). If m is greater than r, we are not able to identify all these m parameters from grouped duration data with only r non-right-censored intervals. We can only identify r restrictions on m parameters, since the baseline parameters enter only the likelihood function through $\gamma_1, \cdots, \gamma_r$. On the other hand, if m is smaller than r, then we are virtually imposing $r - m$ parametric restrictions on these r integrated baseline hazard rates. For identification of the parametric baseline hazard function, we readily note that the number m of free parameters in the baseline hazard function should not exceed the number r of non-right-censored intervals in the data set ($m \leq r$). When the two numbers are equal ($m = r$), both parametric and non-parametric baseline hazard specifications result in the same duration model.

Many authors have stated that a grouped duration analysis without a parametric baseline hazard specification can be thought of as a fully non-parametric procedure with respect to the baseline hazard estimation (Han and Hausman 1990, Moffitt 1985, Sueyoshi 1991). However, this type of procedure is conceptually quite different from the type implied by the conventional usage of the term non − parametrics. Note that in grouped duration analyses, the primary concern is with estimating the integrated baseline hazard rates over each of r non-right-censored intervals, and not the baseline hazard function itself. Of course, if r increases at a certain rate with a compensating decrease in the interval length l as the sample size n increases, then we can apply a standard non-parametric argument to show the consistency of even the baseline hazard function itself.

However, in most grouped duration analyses, the number r of intervals and the interval size l are often fixed by an exogenous observation scheme. In this case, we do not have freedom in choosing finer and finer intervals as the sample size becomes larger. In conventional non-parametrics, we can reduce the band-width (which is equivalent to our interval length l) at a certain rate as the sample size increases.

When the nature of grouped duration data is given by an exogenous observation scheme, we should admit that we can only identify the baseline hazard up to integration over each of the non-right-censored intervals in the observation scheme. In this case, we have virtually the same duration model whether we model the baseline hazard function parametrically or not, insofar as the number m of free parameters in the parametric baseline hazard function is equal to the number r of non-right-censored intervals.

4. MINIMUM CHI-SQUARE ESTIMATION OF GROUPED DURATION MODEL

The minimum Chi-square method in the context of the binary choice model was first proposed by Berkson (1944) for Logit analysis, but can be used for any binary choice model. Since a grouped duration analysis can be reduced to a binary choice analysis as we have seen, we may consider applying the minimum Chi-square estimation to grouped duration analyses. However, this estimator can be defined only when there are many observations for each value of the covariate vector. Often it is described as many observations per cell. A cell is defined as a distinct vector value of covariates. This situation will naturally occur if the nature of covariates is either categorical or otherwise somehow aggregated.

Let me continue to use our simplified set-up to illustrate this estimation method. Suppose that the covariate vector x_i takes on g distinct vector values $x_{(1)}, \cdots, x_{(g)}$. Let us classify the integers (indices of observations) $N = \{1, \cdots, n\}$ into g disjoint sets $N_1, \cdots, N_g$ such that $i \in N_j$ if $x_i = x_{(j)}$, $i = 1, \cdots, n$, $j = 1, \cdots, g$. Note that N_j is a set of indices whose covariates are equal to $x_{(j)}$. Let us divide the integers in N_j into those in N_{1j}, and those not in N_{1j} such that $i \in N_{1j}$ if $i \in N_j$ and if $d_{1i} = 1$. That is, those in N_{1j} are those which have covariates equal to $x_{(j)}$ and which have survived the first interval I_1. Further, divide the integers in N_{1j} into those which have survived the second interval as well and those which have not: $i \in N_{2j}$ if $i \in N_{1j}$ and if $d_{2i} = 1$. Let n_j, n_{1j}, and n_{2j} be the sizes of N_j, N_{1j}, and N_{2j}, respectively. For each group $j = 1, \cdots, g$, we can estimate each interval survival probability, α_{1j} or α_{2j}, through the relative frequency of those which have survived the corresponding interval. Let $\hat{\alpha}_{1j}$ and $\hat{\alpha}_{2j}$ be those estimators: $\hat{\alpha}_{1j} = n_{1j}/n_j$.

and let $\hat{\alpha}_{2j} = n_{2j}/n_{1j}$ (assuming $n_{1j} \neq 0$). Note that $\{\hat{\alpha}_{1j}, \hat{\alpha}_{2j}\}_{j=1}^{g}$ constitute a set of sufficient statistics for the model. In the following discussion, we shall write $x_{(j)}$ as x_j for notational convenience.

Under PHM, interval survival probabilities α_{1j} and α_{2j} can be written as

$$\alpha_{1j} = F(x_j\beta + \gamma_1)$$

and

$$\alpha_{2j} = F(x_j\beta + \gamma_2)$$

where $F(\cdot)$ is defined as $F(x) = e^{e^{-x}}$; $\gamma_1 = \log[\int_0^1 h_0(t)\,dt]$; $\gamma_2 = \log[\int_1^2 h_0(t)\,dt]$. Note that the proportional hazard assumption results in the same β in both α_{1j} and α_{2j}. If the proportionality assumption is violated, the β in α_{1j} will be in general different from the β in α_{2j}. Here, we are not going to assume any functional form for the baseline hazard $h_0(\cdot)$, leaving γ_1 and γ_2 as two free parameters. Since $F(\cdot)$ is one-to-one, we can invert the relationship to obtain

$$F^{-1}(\alpha_{1j}) = x_j\beta + \gamma_1 \tag{4.1}$$

and

$$F^{-1}(\alpha_{2j}) = x_j\beta + \gamma_2 \tag{4.2}$$

where $F^{-1}(x) = \log[-\log(x)]$.

By expanding $F^{-1}(\hat{\alpha}_{1j}) = \log[-\log(\hat{\alpha}_{1j})]$ and $F^{-1}(\hat{\alpha}_{2j}) = \log[-\log(\hat{\alpha}_{2j})]$ in a Taylor series around the true α_{1j} and α_{2j}, we obtain

$$\log[-\log(\hat{\alpha}_{1j})] = x_j\beta + \gamma_1 + u_{1j} \tag{4.3}$$

and

$$\log[-\log(\hat{\alpha}_{2j})] = x_j\beta + \gamma_2 + u_{2j} \tag{4.4}$$

where u_{1j} and u_{2j} are defined by

$$\begin{aligned} u_{1j} &= \frac{\partial F^{-1}(\alpha_{1j})}{\partial \alpha_{1j}}\Big|_{\alpha_{1j}=\alpha_{1j}^*}(\hat{\alpha}_{1j} - \alpha_{1j}) \\ &= \frac{1}{\alpha_{1j}^* \log(\alpha_{1j}^*)}(\hat{\alpha}_{1j} - \alpha_{1j}) \end{aligned}$$

and

$$u_{2j} = \frac{\partial F^{-1}(\alpha_{2j})}{\partial \alpha_{2j}}\Big|_{\alpha_{2j}=\alpha_{2j}^*}(\hat{\alpha}_{2j} - \alpha_{2j})$$
$$= \frac{1}{\alpha_{2j}^* \log(\alpha_{2j}^*)}(\hat{\alpha}_{2j} - \alpha_{2j})$$

with α_{1j}^* and α_{2j}^* lying between $\hat{\alpha}_{1j}$ and α_{1j}, between $\hat{\alpha}_{2j}$ and α_{2j}, respectively.

Under the condition that $\hat{\alpha}_{1j}$ is consistent, we can easily show that $u_{1j} - \frac{1}{\alpha_{1j}\log(\alpha_{1j})}(\hat{\alpha}_{1j} - \alpha_{1j})$ is $O(n_j^{-1})$ and hence u_{1j} can be treated as $\frac{1}{\alpha_{1j}\log(\alpha_{1j})}(\hat{\alpha}_{1j} - \alpha_{1j})$ for large n_j. Similarly, under the condition that $\hat{\alpha}_{2j}$ is consistent, u_{2j} can be treated as $\frac{1}{\alpha_{2j}\log(\alpha_{2j})}(\hat{\alpha}_{2j} - \alpha_{2j})$ for large n_{1j}. For a detailed discussion on this kind of asymptotic equivalence, see Amemiya (1985, sec. 9.2.5).

The $\hat{\alpha}_{1j}$'s are uncorrelated across j since they are computed from different sets of observations. The first interval survival probability $\hat{\alpha}_{1j}$ has variance $\alpha_{1j}(1-\alpha_{1j})/n_j$, $j = 1, \cdots, g$. The $\hat{\alpha}_{2j}$'s are also uncorrelated across j. And $\hat{\alpha}_{2j}$ has variance $\alpha_{2j}(1-\alpha_{2j})/n_{1j}$, conditional on n_{1j}, $j = 1, \cdots, g$. These variances can be consistently estimated by replacing the true unknown quantities α_{1j} and α_{2j} by their corresponding estimates. Also, we can show that the $\hat{\alpha}_{1j}$ and the $\hat{\alpha}_{2j}$ are uncorrelated.

Therefore, the above two equations (4.3) and (4.4) comprise a system of seemingly unrelated regressions (SUR) with heteroscedastic variances. Note that the same regression coefficient β appears in both equations. The minimum Chi-square estimator of $(\beta, \gamma_1, \gamma_2)$ is the weighted least squares estimator applied simultaneously to the above two equations. This estimator has the same asymptotic distribution as the maximum likelihood estimator, provided that each group size goes to infinity as the sample size increases, and that the probability of dying in each interval (each I_1 and I_2) and in each cell (each $j = 1, \cdots, g$) is bounded away from both zero and one in a neighborhood of the true parameter values (Amemiya 1985). Note that without the bounds on the death probability, the inverse probability transformation in (4.1), (4.2) is not well-defined.

Using matrix notation, these two equations can be written as

$$y_1 = Z\theta_1 + u_1 \tag{4.5}$$

and

$$y_2 = Z\theta_2 + u_2 \tag{4.6}$$

where $y_1 = (\log[-\log(\hat{\alpha}_{11}]), \cdots, \log[-\log(\hat{\alpha}_{1g})])'$, $\theta_1 = (\beta' : \gamma_1)'$, $u_1 = (u_{11}, \cdots, u_{1g})'$, y_2, θ_2, u_2 are similarly defined, and $Z = (X : l)$ with $X = (x_1', \cdots, x_g')'$ and $l = (1, \cdots, 1)'$. Let Ω_1 be the variance of u_1 evaluated at $\hat{\alpha}_{1j}$. Similar definitions also apply to Ω_2, therefore

$$\Omega_1 = \text{diag}\left[\frac{1 - \hat{\alpha}_{1j}}{n_j \hat{\alpha}_{1j} (\log \hat{\alpha}_{1j})^2}\right]_{j=1}^{g}$$

and

$$\Omega_2 = \text{diag}\left[\frac{1 - \hat{\alpha}_{2j}}{n_{1j} \hat{\alpha}_{2j} (\log \hat{\alpha}_{2j})^2}\right]_{j=1}^{g},$$

where $\text{diag}[q_j]_{j=1}^{g}$ means a $g \times g$ diagonal matrix having q_j as the $(j,j)th$ element.

Further, we can combine these two equations (4.5) and (4.6) into a single equation system

$$\overline{y} = \overline{Z}\theta + \overline{u} \tag{4.7}$$

where $\overline{y} = (y_1' : y_2')'$; and $\overline{Z}$ is

$$\overline{Z} = \begin{pmatrix} X: & l: & 0 \\ X: & 0: & l \end{pmatrix}$$

with 0 a $g \times 1$ vector of zeros, $\overline{\theta} = (\beta' : \gamma_1 : \gamma_2)'$, and $\overline{u} = (u_1' : u_2')'$. Note that the error term $\overline{u}$ is heteroscedastic. The variance matrix of $\overline{u}$ can be consistently estimated by

$$\overline{\Omega} = \text{diag}[\Omega_1 : \Omega_2].$$

The minimum Chi-square estimator is obtained from the weighted least squares applied to the above equation system

$$\hat{\overline{\theta}} = (\overline{Z}'\overline{\Omega}^{-1}\overline{Z})^{-1}\overline{Z}'\overline{\Omega}^{-1}\overline{y} = \overline{\theta} + (\overline{Z}'\overline{\Omega}^{-1}\overline{Z})^{-1}\overline{Z}'\overline{\Omega}^{-1}\overline{u}. \tag{4.8}$$

And its variance can be consistently estimated by

$$Var(\hat{\overline{\theta}}) = (\overline{Z}'\overline{\Omega}^{-1}\overline{Z})^{-1} = \begin{pmatrix} X'\Omega_1^{-1}X + X'\Omega_2^{-1}X & X'\Omega_1^{-1}l & X'\Omega_2^{-1}l \\ l'\Omega_1^{-1}X & l'\Omega_1^{-1}l & 0 \\ l'\Omega_2^{-1}X & 0 & l'\Omega_2^{-1}l \end{pmatrix}^{-1}. \tag{4.9}$$

By using an inverse formula for partitioned matrices (see for example, Amemiya 1985, p. 460), we can separate the regression coefficient estimator $\hat{\beta}$ out of $\hat{\overline{\theta}}$

$$\hat{\beta} = [(C_1 + C_2)X]^{-1}(C_1 y_1 + C_2 y_2) \tag{4.10}$$

with C_j $(j = 1, 2)$ defined as

$$C_j = X'\Omega_j^{-1}[I_g - l(l'\Omega_j^{-1}l)^{-1}l'\Omega_j^{-1}]. \tag{4.11}$$

Note that $C_j l = 0,\ \ j = 1, 2$. Therefore, $\hat{\beta}$ can be written as

$$\hat{\beta} = \beta + [(C_1 + C_2)X]^{-1}(C_1 u_1 + C_2 u_2). \tag{4.12}$$

Also note that $C_1\Omega_1 C_1' = C_1 X$ and $C_2\Omega_2 C_2' = C_2 X$. Using this, the variance estimator of $\hat{\beta}$ can be easily derived as

$$Var(\hat{\beta}) = [(C_1 + C_2)X]^{-1}, \tag{4.13}$$

which is the $k \times k$ upper-left corner of the $Var(\hat{\bar{\theta}})$ matrix.

Once we sort $\hat{\beta}$ out of $\hat{\bar{\theta}}$, it is straightforward to fill in the rest: deriving the formulas for $\hat{\gamma}_1$ and $\hat{\gamma}_2$, their variances, and the covariances among $\hat{\gamma}_1$, $\hat{\gamma}_2$, and $\hat{\beta}$. They are:

$$\begin{aligned}
\hat{\gamma}_j =& (l'\Omega_j^{-1}l)^{-1}l'\Omega_j^{-1}(y_j - X\hat{\beta}),\ \ j = 1, 2; \\
Var(\hat{\gamma}_j) =& (l'\Omega_j^{-1}l)^{-1} + (l'\Omega_j^{-1}l)^{-2}l'\Omega_j^{-1}XVar(\hat{\beta})X'\Omega_j^{-1}l,\ \ j = 1, 2; \\
Cov(\hat{\gamma}_1, \hat{\gamma}_2) =& (l'\Omega_1^{-1}l)^{-1}(l'\Omega_2^{-1}l)^{-1}l'\Omega_1^{-1}XVar(\hat{\beta})X'\Omega_2^{-1}l; \\
Cov(\hat{\gamma}_j, \hat{\beta}) =& -(l'\Omega_j^{-1}l)^{-1}l'\Omega_j^{-1}XVar(\hat{\beta}),\ \ j = 1, 2.
\end{aligned}$$

Note that

$$Cov(\hat{\gamma}_1, \hat{\gamma}_2) = Cov(\hat{\gamma}_1, \hat{\beta})Var(\hat{\beta})^{-1}Cov(\hat{\beta}, \hat{\gamma}_2),$$

which implies that the correlation between $\hat{\gamma}_1$ and $\hat{\gamma}_2$ comes only and indirectly through their correlations with $\hat{\beta}$. In other words, given $\hat{\beta}$ fixed, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are uncorrelated.

5. SPECIFICATION TESTS OF THE PROPORTIONAL HAZARD ASSUMPTION: OVERVIEW, THEORETICAL BACKGROUND

The immense popularity of PHM has made the issue of model checking extremely important. Many ways have been developed for testing for the proportionality assumption (see among others, Lin 1991, Sueyoshi 1991).

Lin (1991) devised a test for the proportional hazard assumption by comparing weighted and non-weighted score estimators in a continuous time/continuous data context. His non-weighted score function is simply the first derivative of the Cox partial log-likelihood function with respect to the regression parameters. If the proportional hazard assumption holds, then the non-weighted score estimator which is obtained by setting the score function equal to zero is the most efficient, consistent estimator. This estimator is in fact Cox's partial maximum likelihood estimator. On the other hand, any weighted score estimator which is obtained by setting a weighted score function (where individual scores are asymmetrically weighted depending on the time of deaths) equal to zero is still consistent but not as efficient as the unweighted estimator. However, if the proportional hazard assumption fails to hold, both estimators converge to different quantities in general. This observation gives rise to Lin's (1991) test statistic, in fact, this can be viewed as an application of Hausman's (1978) specification test idea.

Sueyoshi (1991) developed another test for the proportionality assumption by applying a Lagrange multiplier test idea to grouped duration analyses of PHM (in a continuous time/discrete data context). As we have discussed, the likelihood function of grouped duration data takes the form of a binary choice model. PHM imposes the restriction that we have the same regression coefficients β appearing in each interval survival probability α_j, $j = 1, \cdots, r$. If the proportionality assumption is violated, we will in general

have different β's across the different intervals. Sueyoshi's test is a Lagrange multiplier test to check whether the coefficients are constant across these intervals or not. This test statistic can be easily computed from a certain auxiliary least squares. The auxiliary regression takes one of two different forms depending on whether we use the BHHH information matrix or its expected value (see Berndt et. al. 1974).

By further aggregating the already grouped duration data, we can artificially generate another coarser set of grouped duration data. Concretely, in our simplified setting, we may consider aggregating two intervals I_1 and I_2 into a single large interval $I \equiv I_1 \cup I_2 = [0, 2)$. In this case, the duration data will take the following form: $d_i = d_{1i}d_{2i} = 1$ if individual i survives both interval I_1 and I_2, $d_i = 0$ otherwise. And $\alpha_i = \alpha_{1i}\alpha_{2i}$ is the true survival probability of the interval $I \equiv I_1 \cup I_2$. Obviously, the new coarser data set contains a lesser

amount of information.

If we estimate the same fixed number of parameters from both the new (coarser) and the original (finer) data sets, the estimator, say $\hat{\beta}$, from the original data set will be more efficient than the one, say $\hat{\beta}^*$, from the new set.

This is the case when we make a parametric baseline hazard assumption. On the other hand, if we do not make a parametric functional form

assumption on the baseline hazard as in the previous section, the number of parameters being estimated under the two data set-ups are different. When we are estimating the model using the

finer data set (original data set), the information content of the data is richer, but in this case we are estimating more parameters (estimating more steps regarding the baseline hazard). Therefore, we cannot say that one estimator is more efficient than the other, lacking an efficiency ranking between the two estimators. Of course, whether we specify a parametric baseline hazard or not, both estimators will be consistent under the proportionality assumption.

If the proportional hazard assumption is violated, both estimators will converge to different quantities, basically by the same reason as in Lin (1991). To see this, let us assume that all covariates have decreasing impacts on the hazard rates: the β applied to the first interval I_1 (say, β^1) is bigger in a vector sense than the β applied to the second interval I_2 (say, β^2). Then $\hat{\beta}$ tends to be stochastically larger than $\hat{\beta}^*$, and $\hat{\beta}$ converges to a quantity which is bigger in a vector sense than the probability limit of $\hat{\beta}^*$. The intuitive reason is as follows: $\hat{\beta}^*$, obtained from the coarser data, is symmetrically affected by both β^1 and β^2 since the larger interval I is the symmetric aggregation of the two sub-intervals I_1 and I_2. However, $\hat{\beta}$, obtained from the finer data set, is asymmetrically affected by both β^1 and β^2 since we tend to have more observations in the first interval than in the second. In other words, the influence of β^1 relative to that of β^2 tends to be stronger in the case of using the finer data set than in the case of using the coarser data set. Since β^1 is assumed to be larger than β^2, $\hat{\beta}$ tends to be stochastically larger than $\hat{\beta}^*$.

Therefore, the difference between the two estimators converges to a zero vector under the proportional hazard assumption, but to a non-zero vector under a non-proportionality

Our specification test uses this disparate convergence pattern of the difference in the estimators: if the difference is significantly different from a zero vector, reject the proportional hazard assumption, otherwise, do not.

The test statistic

$$R = (\hat{\beta}^* - \hat{\beta})'[Var(\hat{\beta}^* - \hat{\beta})]^{-1}(\hat{\beta}^* - \hat{\beta}) \tag{5.1}$$

follows a Chi-square distribution with k (the number of parameters in β) degrees of freedom. The variance inside the bracket can be consistently estimated by

$$Var(\hat{\beta}^* - \hat{\beta}) = Var(\hat{\beta}^*) + Var(\hat{\beta}) - Cov(\hat{\beta}^*, \hat{\beta}) - Cov(\hat{\beta}, \hat{\beta}^*). \tag{5.2}$$

In computing the above test statistic, the problem usually lies with the computational burden of computing the covariance matrix between the two estimators. It is so because when the two estimators do not satisfy an informational ranking relation, their covariance matrix does not reduce to the variance of the more efficient estimator, a property enjoyed by a pair of estimators satisfying an informational nesting relation (see Hausman 1978). Here, how to compute the covariance between these two regression estimators is at stake. When there are many observations per cell, the task can be simplified through the minimum Chi-square method.

The next section obtains $\hat{\beta}^*$, its variance, and its covariance with $\hat{\beta}$.

If we have more than two non-right-censored intervals ($r > 2$) in the original data set, we have a lot more flexibility in this kind of intentional aggregation process. For testing purposes, the selection of an optimal aggregation may be guided by whatever alternative hypothesis one has in mind. For example, if there are three non-right-censored intervals ($r = 3$) in the original data set and if one suspects that the covariate impact is stronger in the third interval (if there is any difference), one may aggregate intervals I_1 and I_2 together, leaving the third interval I_3 alone. This aggregation will yield a higher power against the suspected alternative than any other kind of aggregation. Of course, without a clear prior alternative, we cannot design an optimal aggregation.

Besides the overall Chi-square test, we can conduct individual t-tests to separately see which covariate has a non-proportional impact on the hazard rate. Also, the signs of

the individual t-test statistics give us some idea in which direction the proportionality is violated.

The essence of this testing idea is to secure a testable implication by throwing away a certain amount of information (aggregation of the original data entails a certain loss of informational content), which shares Goldfeld and Quandt's (1965) heteroscedasticity test idea.

6. SPECIFICATION TESTS OF THE PROPORTIONAL HAZARD MODEL: ACTUAL DERIVATION

Now let us aggregate two sub-intervals I_1 and I_2 from section 4 into a single large interval I ($\equiv I_1 \cup I_2$). The estimator $\hat{\alpha}_j$ of the interval survival probability α_j is ($n_{1j} \neq 0$)

$$\hat{\alpha}_j = \frac{n_{2j}}{n_j} = \frac{n_{1j}}{n_j}\frac{n_{2j}}{n_{1j}} = \hat{\alpha}_{1j}\hat{\alpha}_{2j}. \tag{6.1}$$

By the same procedure as in the finer data case from section 4, we have

$$y = Z\theta + u, \tag{6.2}$$

where $y = (\log[-\log(\hat{\alpha}_1)], \cdots, \log[-\log(\hat{\alpha}_g)])'$; $Z = (X : l)$ with $X = (x_1', \cdots, x_g')'$; $\theta = (\beta' : \gamma)'$ with $\gamma = \log[\int_0^2 h_0(t)\,dt] = \log(e^{\gamma_1} + e^{\gamma_2})$; and $u = (u_1, \cdots, u_g)'$ with

$$\begin{aligned} u_j \approx & \frac{1}{\hat{\alpha}_j \log(\hat{\alpha}_j)}(\hat{\alpha}_j - \alpha_j) \\ \approx & \frac{\hat{\alpha}_{2j}}{\hat{\alpha}_j \log(\hat{\alpha}_j)}(\hat{\alpha}_{1j} - \alpha_{1j}) + \frac{\alpha_{1j}}{\hat{\alpha}_j \log(\hat{\alpha}_j)}(\hat{\alpha}_{2j} - \alpha_{2j}) \\ \approx & \frac{\log(\hat{\alpha}_{1j})}{\log(\hat{\alpha}_{1j}) + \log(\hat{\alpha}_{2j})}u_{1j} + \frac{\log(\hat{\alpha}_{2j})}{\log(\hat{\alpha}_{1j}) + \log(\hat{\alpha}_{2j})}u_{2j}, \end{aligned} \tag{6.3}$$

where $X \approx Y$ means that X and Y are asymptotically equivalent. This equivalence can be easily seen by noting that

$$\begin{aligned} \hat{\alpha}_j - \alpha_j = & \hat{\alpha}_{1j}\hat{\alpha}_{2j} - \alpha_{1j}\alpha_{2j} \\ = & \hat{\alpha}_{1j}(\hat{\alpha}_{2j} - \alpha_{2j}) + \alpha_{2j}(\hat{\alpha}_{1j} - \alpha_{1j}) \\ \approx & \alpha_{1j}(\hat{\alpha}_{2j} - \alpha_{2j}) + \alpha_{2j}(\hat{\alpha}_{1j} - \alpha_{1j}) \\ \approx & \hat{\alpha}_{1j}(\hat{\alpha}_{2j} - \alpha_{2j}) + \hat{\alpha}_{2j}(\hat{\alpha}_{1j} - \alpha_{1j}). \end{aligned} \tag{6.4}$$

Of course, this asymptotic equivalence holds only after a suitable normalization.

Using a matrix notation, u can be written as a weighted sum of two uncorrelated error vectors u_1 and u_2

$$u = Wu_1 + (I_g - W)u_2, \tag{6.5}$$

with

$$W = \text{diag}\left[\frac{\log(\hat{\alpha}_{1j})}{\log(\hat{\alpha}_{1j}) + \log(\hat{\alpha}_{2j})}\right]_{j=1}^{g}. \tag{6.6}$$

Let Ω be the variance of u evaluated at $\hat{\alpha}_j$'s

$$\begin{aligned}\Omega \approx & \text{diag}\left[\frac{1-\hat{\alpha}_{1j}}{n_j\hat{\alpha}_{1j}(\log(\hat{\alpha}_{1j}) + \log(\hat{\alpha}_{2j}))^2} + \frac{1-\hat{\alpha}_{2j}}{n_{1j}\hat{\alpha}_{2j}(\log(\hat{\alpha}_{1j}) + \log(\hat{\alpha}_{2j}))^2}\right]_{j=1}^{g} \\ = & W\Omega_1 W + (I_g - W)\Omega_2(I_g - W).\end{aligned} \tag{6.7}$$

The pooled minimum Chi-square estimator $\hat{\theta}^*$ of $\theta = (\beta' : \gamma)'$ is obtained from the weighted least squares applied to the above equation using the aggregate data $I = I_1 \cup I_2$

$$\hat{\theta}^* = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}y = \theta + (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}u. \tag{6.8}$$

And the variance of $\hat{\theta}^*$ can be consistently estimated by

$$Var(\hat{\theta}^*) = (Z'\Omega^{-1}Z)^{-1} = \begin{pmatrix} X'\Omega^{-1}X & X'\Omega^{-1}l \\ l'\Omega^{-1}X & l'\Omega^{-1}l \end{pmatrix}^{-1}. \tag{6.9}$$

Again, we can separate the regression coefficient estimator $\hat{\beta}^*$ out of $\hat{\theta}^*$

$$\hat{\beta}^* = [CX]^{-1}Cy, \tag{6.10}$$

with C defined as

$$C = X'\Omega^{-1}[I_g - l(l'\Omega^{-1}l)^{-1}l'\Omega^{-1}]. \tag{6.11}$$

Note that $Cl = 0$. Therefore,

$$\hat{\beta}^* = \beta + [CX]^{-1}Cu. \tag{6.12}$$

By noting that $C\Omega C' = CX$, we can easily estimate the variance of $\hat{\beta}^*$ as

$$Var(\hat{\beta}^*) = [CX]^{-1}, \tag{6.13}$$

which is the $k \times k$ upper-left corner of $Var(\hat{\theta}^*)$.

Once we sort $\hat{\beta}^*$ out of $\hat{\theta}^*$, it is again straightforward to fill in the rest: deriving the formula for $\hat{\gamma}^*$, its variance, and the covariance between $\hat{\gamma}^*$ and $\hat{\beta}^*$. They are:

$$\begin{aligned}\hat{\gamma}^* &= (l'\Omega^{-1}l)^{-1}l'\Omega^{-1}(y - X\hat{\beta}^*);\\ Var(\hat{\gamma}^*) &= (l'\Omega^{-1}l)^{-1} + (l'\Omega^{-1}l)^{-2}l'\Omega^{-1}XVar(\hat{\beta}^*)X'\Omega^{-1}l;\\ Cov(\hat{\gamma}^*, \hat{\beta}^*) &= -(l'\Omega^{-1}l)^{-1}l'\Omega^{-1}XVar(\hat{\beta}^*).\end{aligned}$$

The covariance between $\hat{\beta}^*$ and $\hat{\beta}$ can be consistently estimated by

$$Cov(\hat{\beta}^*, \hat{\beta}) = (CX)^{-1}C[W\Omega_1 C_1' + (I_g - W)\Omega_2 C_2'][(C_1 + C_2)X]^{-1},$$

which can be easily shown by noting that (i) $\hat{\beta}^* = \beta + [CX]^{-1}Cu \approx \beta + [CX]^{-1}C[Wu\_1 + (I_g - W)u_2]$, (ii) $\hat{\beta} = \beta + [(C_1 + C_2)X]^{-1}(C_1u_1 + C_2u_2)$, and that (iii) u_1 and u_2 are uncorrelated.

By comparing the two estimators, $\hat{\beta}^*$ in this section and $\hat{\beta}$ from section 4, we can develop a test statistic for PHM. Under the null hypothesis that PHM holds, both $\hat{\beta}$ and $\hat{\beta}^*$ are consistent. If the null hypothesis does not hold, $\hat{\beta}$ and $\hat{\beta}^*$ converge to different quantities as the sample size increases. This disparate behavior of the two estimators offers a clue to the design of our test statistic. The test statistic

$$R = (\hat{\beta}^* - \hat{\beta})'[Var(\hat{\beta}^* - \hat{\beta})]^{-1}(\hat{\beta}^* - \hat{\beta}) \tag{6.14}$$

will follow a Chi-square distribution with k (number of parameters in β) degrees of freedom. The variance inside the bracket can be consistently estimated by

$$Var(\hat{\beta}^* - \hat{\beta}) = Var(\hat{\beta}^*) + Var(\hat{\beta}) - Cov(\hat{\beta}^*, \hat{\beta}) - Cov(\hat{\beta}, \hat{\beta}^*). \tag{6.15}$$

7. MONTE-CARLO SIMULATIONS

The true model is specified through the hazard rate $h(t, x)$ of an individual with covariate vector $x = (x_1, x_2, x_3)'$

$$h(t, x) = (0.36t^{0.2})\exp(0.2x_1 - 0.5x_2 + 0.8x_3).$$

Note that the baseline hazard follows a Weibull distribution with an increasing hazard rate over time. Assume that x_1, x_2 and x_3 are all discrete: $x_1 = -1$, or 1; $x_2 = -0.6$, 0, or 0.5; $x_3 = -0.9$, -0.3, 0.3, or 0.9. Therefore, we have $2 \times 3 \times 4 = 24$ different cells ($g = 24$), each of which takes on a distinct vector value $x = (x_1, x_2, x_3)'$. The observation scheme is assumed to be $Q = \{0, 1, 2, \infty\}$. It is characterized by the two non-right-censored intervals ($r = 2$), $I_1 = [0, 1)$ and $I_2 = [1, 2)$. The proportional hazard assumption is satisfied by $\beta_1 = 0.2, \beta_2 = -0.5$, and $\beta_3 = 0.8$. This same $\beta = (0.2, -0.5, 0.8)'$ affects both interval survival probabilities α_1 and α_2. We make each group size equal in the simulation. Equal group sizes imply independence among x_1, x_2 and x_3. The total sample size is the group size multiplied by the number of groups. The interval survival indicators, d_{1i}, d_{2i}, are each generated by independent coin tosses with success probabilities, α_{1i} and α_{2i}, respectively. In lieu of actual coin tosses, uniform random numbers are generated by the GAUSS mathematical language. Obviously, we need d_{2i} only when $d_{1i} = 1$. The combined indicator d_i which takes the value one if individual i survives the large interval I, zero otherwise, is obtained by multiplying the two sub-indicators: $d_i = d_{1i} d_{2i}$.

Denote by $\hat{\beta}$ the estimates of β using the finer data set, I_1 and I_2; and denote by $\hat{\beta}^*$ the estimates using the coarser data set, $I \equiv I_1 \cup I_2$. Calculating these estimates is straightforward since it involves only matrix operations. Again, we use GAUSS to perform the necessary matrix operations. Based on a thousand replications, we computed the empirical mean and the empirical standard deviation of the estimators, $\hat{\beta}^*$ and $\hat{\beta}$. We tried seven different group sizes to observe convergence patterns of the estimates as the sample size (=24× group size) increases. These estimated β's and the standard deviations are reported in Table 1. Notice that as the group size increases, accordingly, as the sample size increases, both $\hat{\beta}^*$'s and $\hat{\beta}$'s converge to the true parameter values. In general, between $\hat{\beta}^*$ and $\hat{\beta}$ we cannot choose which estimator is more efficient. The reason is that whenever we are using a finer data set, we are estimating more parameters. Even for large group sizes, the performance of both estimators are quite comparable, with the difference in the mean square errors very small.

To see how the variance formulas perform, we compared the empirically obtained variances with the corresponding theoretical variances. Theoretical variances are evaluated

at the true parameter values. Table 2 and Table 3 show the empirical and the theoretical variance estimates of $\hat{\beta}^*$ and $\hat{\beta}$, respectively, for group sizes 20, 50, and 100. As expected, for larger group sizes we have more favorable match-ups. Regarding the variance estimates, the theoretical variance formulas seem to match up quite well with the corresponding empirical variances even for small group sizes. Since components of the x vector are generated to be independent, covariances among the estimators of β are expected to be small. In fact, the off-diagonal elements (covariances) are much smaller than the diagonal elements (variances). Throughout Table 2 and Table 3, all correlations (normalized off-diagonal elements) are between 0.01 and 0.07.

Now, let us consider several alternative models to the proportional hazard model. The true regression coefficients applied to the first interval I_1 are set as before. The true regression coefficients applied to the second interval I_2 are set by adding a vector value Δ to the first interval coefficients. Under the proportionality assumption, Δ is set equal to zero. As Δ deviates from the zero vector, the model moves farther and farther away from PHM. When we are estimating a non-proportional hazard model (non-zero Δ) wrongly assuming proportionality, the estimates are in general inconsistent. But, by the assumed independence among the components in the covariate vector x, only the non-proportional coefficients are inconsistently estimated. On the other hand, all the other proportional coefficients are still consistently estimated.

Table 4 shows empirical means of the two estimates, β^* and $\hat{\beta}$ under several alternative models. The coefficients corresponding to the non-zero elements in Δ are the non-proportional coefficients. The coefficients corresponding to the zero elements in Δ are the proportional coefficients. All the proportional coefficients are close to the true values in terms of the empirical means. Note that the pooled estimates of the non-proportional coefficients (components of $\hat{\beta}^*$ corresponding to the non-zero elements of Δ) are always smaller than the corresponding non-pooled estimates (components of $\hat{\beta}$ corresponding to the non-zero elements of Δ). Since the true non-proportional coefficients are monotonically decreasing over time (note that $\beta^2 = \beta^1 + \Delta$, and $\Delta \leq 0$), their non-pooled estimators, which receive more weight from the first interval than from the second, are stochastically larger than the corresponding pooled estimators.

Table 5 shows the empirical sizes (when Δ is a zero vector) and powers (when Δ is a non-zero vector) of the proposed overall χ^2 tests in terms of the empirical percentage of rejections out of 1000 repetitions. Results in the top block under $\Delta' = (0, 0, 0)$ show empirical sizes under each of the three nominal sizes. As the group sizes increase, empirical sizes converge to the corresponding nominal sizes. However, for group sizes up to 50, the test tends to reject the true null hypothesis more frequently than supposed. Results in the rest of the blocks under non-zero Δ's show empirical powers. As the group sizes increase, the power increases toward 100 %, exhibiting consistency of the test.

To measure a "distance" between the model implied by the non-zero Δ (call it f_1) and the wrongly assumed PHM (regarding $\Delta = 0$; call it f_0), we modify one of the Kullback-Leibler (1951) information measures to be suitable to the current discrete observation scheme. The discrete Kullback-Leibler information is defined as

$$KL = E^*[\log(P^*(T_d)) - \log(P(T_d))],$$

where (i) E^* denotes the expectation taken with respect to model f^* (recall that the data are generated according to model f^* in our simulations), (ii) T_d is the interval censored observation on the underlying continuous duration T, and (iii) $P^*(T_d)$, $P(T_d)$ denotes the probability that T_d will be observed under model f^*, f, respectively. Given a particular cell x, we can easily compute $KL(x)$. To define a unique measure for each Δ, we take the expectation of $KL(x)$ with respect to the distribution of x. We denote the resulting quantity as KL in Tables 5,6,7. For details on this distance measure, see the Appendix. In Table 5, we readily notice that as the distance between the null and alternative increases, the power tends to increase. Except in some cases when $n_g = 100$, the rankings based on power are exactly equal to the rankings based on the distance measure.

Besides the overall χ^2 tests, we also carried out individual t-tests. In fact, we used normal approximation based on asymptotics. If we have a certain prior knowledge about the direction in which the alternative deviates from the null, we can conduct one-sided t-tests; otherwise use two-sided t-tests. These results are reported in Table 6 and 7, respectively. Those t-tests based on the non-proportional coefficients show a reasonable powers; while those t-tests based on the proportional coefficients do not show any significant

powers in excess of the assumed nominal sizes. When there is only one non-proportional coefficient, the corresponding t-test yields a higher power than the overall χ^2 test. It is because we pin point the exact source of non-proportionality. In this case, the t-test is not diluted by other proportional components, while the χ^2 test is. When there are two non-proportional coefficients, we cannot choose which one, between the t-tests and the χ^2 test, is more powerful. However, as the group size increases, the overall χ^2 test surpass either of the relevant t-tests.

8. CONCLUDING REMARKS

Often, we face grouped duration data due to a certain discrete observation mechanism. The discrete time setting in duration analyses may cause a conceptual problem in that there arises inconsistency between the two choices in time units. This kind of problem can be avoided by using a continuous time duration model. The proportional hazard model is one of the most widely used example of the continuous time duration model.

In this paper, the relationship between group duration analysis and binary choice analysis is shown. Estimation and identification issues are also discussed. When covariates are all discrete, application of Berkson's minimum Chi-square estimation yields a computationally simple estimator which is asymptotically as efficient as the maximum likelihood estimator.

We also suggest simple specification tests for the proportional hazard model by comparing two sets of minimum Chi-square estimators, one from a finer data set, and the other from a coarser data set. The suggested tests are easy to use and can take alternative hypotheses into account, thereby increasing the power of these tests.

Using Monte-carlo simulations, we illustrate the performance of the proposed minimum Chi-square estimators and the size and power properties of the proposed specification tests. Performance of the overall Chi-square tests and individual t-tests are compared.

APPENDIX: KULLBACK-LEIBLER INFORMATION MEASURE

I. GENERAL THEORY

Let $Q = \{0, a_1, \cdots, a_r, \infty\}$ be a partition of the positive real line R^+. Then, Q can represent the following discrete observation scheme: the underlying continuous duration data on T are only available up to intervals $I_1 = [0,1), \cdots, I_r = [a_{r-1}, a_r), I_{r+1} = [a_r, \infty)$ (recall that in the text $Q = \{0, 1, 2, \infty\}$, and that the data are available only up to intervals $I_1 = [0,1)$, $I_2 = [1,2)$, $I_3 = [2,\infty)$.) Let us denote the observed discrete duration data as T_d (note that T_d is an interval censored version of the underlying T).

Let f^* and f be two alternative distributions for T. Based on the observation scheme Q (equivalently, based on T_d), define the discrete version of the Kullback-Leibler distance (information) measure between those two distributions as:

$$KL_Q = E^*[\log(P^*(T_d)) - \log(P(T_d))],$$

where (i) E^* denotes the expectation taken with respect to model f^*, and (ii) $P^*(T_d), P(T_d)$ denotes the probability that T_d will be observed under model f^*, f, respectively. By Jensen's inequality, we have $-KL_Q = E^*[\log P(T_d)/P^*(T_d)] \leq \log E^*[P(T_d)/P^*(T_d)] = \log 1 = 0$ with equality if and only if $P^*(T_d) = P(T_d)$ for every realization of T_d. Therefore, the distance measure KL_Q is non-negative, and strictly positive if and only if two distributions have different observable implications within the observation scheme Q.

Let P_j^* and P_j be $P^*(T \in I_j)$, $P(T \in I_j)$, $j = 1, \cdots, r, r+1$: $P_j^* = \int_{I_j} f^*(t)\,dt$, $P_j = \int_{I_j} f(t)\,dt$. Then KL_Q can be more concretely written as

$$KL_Q = \sum_{j=1}^{r+1} P_j^*[\log P_j^* - \log P_j]. \tag{A.1}$$

Assumption 1. Two alternative distributions f^* and f satisfy $\int_0^\infty f^*(t)[\log f^*(t) - \log f(t)]\,dt < \infty$, $\int_0^\infty f(t)[\log f(t) - \log f^*(t)]\,dt < \infty$.

This assumption states that the Kullback-Leibler distance measure between f^* and f exists, whether measured based on f^* or f.

Theorem 1. Suppose $\{Q_k\}_{k=1,2,\cdots}$ is a sequence of partitions of the support R^+ such that[1] (1) $Q_k \subset Q_{k+1}$ and (2) $mesh\{Q_k\} \to 0$ as $k \to \infty$. Then under Assumption 1, $\{KL_{Q_k}\}_{k=1,2,\cdots}$ is an increasing sequence with limit equal to the Kullback-Leibler distance measure based on the exact observation.

Proof of Theorem 1.

The proof of (1) is completed by using Lemma 1. The second part (2) is proved in steps using Lemmas 2,3, and 4.

(1) By (A.1), it suffices to prove for each $j = 1, \cdots, r, r+1$ and for each $m_j \in I_j$

$$P^*_{1j}[\log P^*_{1j} - \log P_{1j}] + P^*_{2j}[\log P^*_{2j} - \log P_{2j}] \geq (P^*_{1j} + P^*_{2j})[\log(P^*_{1j} + P^*_{2j}) - \log(P_{1j} + P_{2j})],$$

where

$$P^*_{1j} = \int_{a_{j-1}}^{m_j} f^*(t)\,dt,\ P^*_{2j} = \int_{m_j}^{a_j} f^*(t)\,dt,\ P^*_j = P^*_{1j} + P^*_{2j}$$

and

$$P_{1j} = \int_{a_{j-1}}^{m_j} f(t)\,dt,\ P_{2j} = \int_{m_j}^{a_j} f(t)\,dt,\ P_j = P_{1j} + P_{2j}.$$

Equivalently, suffices to show

$$\left(\frac{P^*_{1j}}{P_{1j}}\right)^{\frac{P^*_{1j}}{P^*_{1j}+P^*_{2j}}} \cdot \left(\frac{P^*_{2j}}{P_{2j}}\right)^{\frac{P^*_{2j}}{P^*_{1j}+P^*_{2j}}} \geq \frac{P^*_{1j} + P^*_{2j}}{P_{1j} + P_{2j}}.$$

Lemma 1. $\left(\frac{P^*_{1j}}{P_{1j}}\right)^{\frac{P^*_{1j}}{P^*_{1j}+P^*_{2j}}} \cdot \left(\frac{P^*_{2j}}{P_{2j}}\right)^{\frac{P^*_{2j}}{P^*_{1j}+P^*_{2j}}} \geq \frac{P^*_{1j}}{P^*_{1j}+P^*_{2j}}\left(\frac{P^*_{1j}}{P_{1j}}\right) + \frac{P^*_{2j}}{P^*_{1j}+P^*_{2j}}\left(\frac{P^*_{2j}}{P_{2j}}\right)$ with equality if and only if $\frac{P^*_{1j}}{P^*_{2j}} = \frac{P_{1j}}{P_{2j}}$ (For proof, see Royden (1968), Lemma 1. on page 112.)

But, we can easily show that the right hand side of the inequality of Lemma 1 is greater than or equal to $\frac{P^*_{1j}+P^*_{2j}}{P_{1j}+P_{2j}}$ with equality if and only if $\frac{P_{1j}}{P^*_{1j}} = \frac{P_{2j}}{P^*_{2j}}(=\frac{P_j}{P^*_j})$. Therefore, we have proved that as the data partition gets finer, the Kullback-Leibler distance measure between two distributions increases except when $\frac{P_{1j}}{P^*_{1j}} = \frac{P_{2j}}{P^*_{2j}}$ for each sub-partitioned interval j. If $\frac{P_{1j}}{P^*_{1j}} = \frac{P_{2j}}{P^*_{2j}}$ for each sub-partitioned interval j, then the sub-partition does not add any

[1] Here *mesh* is defined as the supremum of the distances between any two contiguous partition points.

new information with regards to distinguishing between the two alternative distributions, and the Kullback-Leibler information measure stays the same.

(2) For $\{Q_k\}_{k=1,2,\cdots}$, define

$$Z_k(t) = \sum_{I_j \in Q_k} 1_{\{t \in I_j\}} [\log P_j^* - \log P_j],$$

$$Z(t) = \int_0^\infty 1_{\{t \in du\}} [\log f^*(t) - \log f(t)],$$

where I_j's are intervals in Q_k (I dropped the superscript k to simplify the notation) and P_j^* abd P_j are the corresponding interval probabilities under model f^* and f. Note that $E^* Z_k = KL_{Q_k}$ and $EZ = KL_{R^+}$, where KL_{R^+} is the Kullback-Leibler distance measure based on the exact observation.

Now the following lemmas hold:

Lemma 2 For all $k = 1, 2, \cdots, EZ_k \leq EZ$: i.e.

$$\sum_j P_j^* [\log P_j^* - \log P_j] \leq \int_0^\infty f^*(t)[\log f^*(t) - \log f(t)].$$

Proof For each I_j, it suffices to show

$$P_j^* (\log P_j^* - \log P_j) \leq \int_{I_j} f^*(t)[\log f^*(t) - \log f(t)]\, dt.$$

By the following representation

$$\int_{I_j} f^*(t)[\log f^*(t) - \log f(t)]\, dt / P_j^* = E^*[\log f^*(T) - \log f(T) | T \in I_j],$$

it also suffices to show

$$E^*[\log f(T) - \log f^*(T) | T \in I_j] \leq \log P_j - \log P_j^*.$$

But, by Jensen's inequality,

$$\begin{aligned}
(LHS) &= E^*[\log f(T)/f^*(T) | T \in I_j] \\
&\leq \log E^*[f(T)/f^*(T) | T \in I_j] \\
&= \log[\int_{I_j} f^*(t) f(t)/f^*(t)\, dt / P_j^*] \\
&= \log P_j - \log P_j^* = (RHS)
\end{aligned}$$

with the equality holding if and only if $f(t)/f^*(t)$ =(constant) for every $t \in I_j$. When $f(t)/f^*(t)$ =(constant) for every $t \in I_j$, the interval observation $t \in I_j$ and the exact observation within I_j are equivalent with regards to distinguishing between f^* and f, and therefore their Kullback-Leibler distance measures take the same numerical value.

Lemma 3 $\{Z_k\}_{k=1,2,\cdots}$ is a submartingale relative to σ-fields $\{F_k\}_{k=1,2,\cdots}$, where F_k denotes the σ-field generated by Q_k.

Proof (i) Obviously, $F_k \subset F_{k+1}$; (ii) Z_k is F_k measurable;

$$\begin{aligned}
\text{(iii) } E|Z_k| &= \sum_j P_j^* |\log P_j^* - \log P_j| \\
&= \sum_{P_j^* \geq P_j} P_j^* [\log P_j^* - \log P_j] + \sum_{P_j^* < P_j} P_j^* [\log P_j - \log P_j^*] \\
&\leq \sum_{P_j^* \geq P_j} P_j^* [\log P_j^* - \log P_j] + \sum_{P_j^* < P_j} P_j [\log P_j - \log P_j^*] \\
&\leq \sum_{P_j^* \geq P_j} \int_{I_j} f^*(t)[\log f^*(t) - \log f(t)]\, dt + \sum_{P_j^* < P_j} \int_{I_j} f(t)[\log f(t) - \log f^*(t)]\, dt \\
&\leq \int_{f^* \geq f} f^*(t)[\log f^*(t) - \log f(t)]\, dt + \int_{f^* < f} f(t)[\log f(t) - \log f^*(t)]\, dt,
\end{aligned}$$

which is finite under Assumption 1; (iv) With probability one, $E^*[Z_{k+1}|F_k] \geq Z_k$, which is easily derived by the similar arguments as in Proof (1) above.

Lemma 4 $Z_k \to Z$ a.e. $t \in R^+$. (see Theorem 35.4, Billingsley, 1986).

Now by combining Lemma 4 with the bounded convergence theorem, we have $\lim_k EZ_k = EZ$. That is, the limit of the discrete Kullback-Leibler information is equal to the continuous Kullback-Leibler information.

When we introduce fixed right censoring at $t = c$, we can similarly prove that as the *mesh* of partitions (up to the censoring point c) goes to zero, its discrete Kullback-Leibler distance measure converges to

$$\int_0^c f^*(t)[\log f^*(t) - \log f(t)]\, dt + P_c^*[\log P_c^* - \log P_c],$$

where $P_c^* = \int_c^\infty f^*(t)\, dt$ and $P_c = \int_c^\infty f(t)\, dt$. Note that the above formula is the Kullback-Leibler distance when data on T are continuously available up to the point c and then

right-censored. Of course, setting $c = \infty$ yields the Kullback-Leibler distance based on the exact observation without any censoring.

II. APPLICATION TO GROUP DURATION ANALYSIS OF PHM

To measure a "distance" between the model implied by the non-zero Δ (call it f_1) and the wrongly assumed PHM (regarding $\Delta = 0$; call it f_0), we apply the above discrete Kullback-Leibler information measure. The discrete Kullback-Leibler information is

$$KL = E^*[\log(P^*(T_d)) - \log(P(T_d))],$$

where (i) E^* denotes the expectation taken with respect to model f^* (recall that the data are generated according to model f^* in our simulations), (ii) T_d is the interval censored observation on the underlying continuous duration T, and (iii) $P^*(T_d)$, $P(T_d)$ denotes the probability that T_d will be observed under model f^*, f, respectively. Given a particular cell x, we can easily compute $KL(x)$, the discrete Kullback-Leibler distance measure conditional on x. To define a unique measure for each Δ, we take the expectation of $KL(x)$ with respect to the distribution of x. Note that the discrete observation scheme we are discussing is $Q = \{0, 1, 2, \infty\}$. The $KL(x)$ is

$$\begin{aligned} KL(x) = &\sum_{j=1}^{3} P_j^*[\log P_j^* - \log P_j] \\ =&(1 - \alpha_{1x}^*)[\log(1 - \alpha_{1x}^*) - \log(1 - \alpha_{1x})] \\ &+\alpha_{1x}^*(1 - \alpha_{2x}^*)[\log \alpha_{1x}^*(1 - \alpha_{2x}^*) - \log \alpha_{1x}(1 - \alpha_{2x})] \\ &+\alpha_{1x}^*\alpha_{2x}^*[\log \alpha_{1x}^*\alpha_{2x}^* - \log \alpha_{1x}\alpha_{2x}]. \end{aligned}$$

The notation should be self-explanatory referring to the text. By using the formulas for α^*'s and α's, and by noting that $\alpha_{1x}^* = \alpha_{1x}$, we have

$$\begin{aligned} KL(x) =&e^{-e^{x\beta+\gamma_1}}(1 - e^{-e^{x(\beta+\Delta)+\gamma_2}})[\log(1 - e^{-e^{x(\beta+\Delta)+\gamma_2}})) - \log(1 - e^{-e^{x\beta+\gamma_2}})] \\ &+e^{-e^{x\beta+\gamma_1}}e^{-e^{x(\beta+\Delta)+\gamma_2}}[-e^{x(\beta+\Delta)+\gamma_2} + e^{x\beta+\gamma_2}]. \end{aligned}$$

REFERENCES

Amemiya, T. (1985), Advanced Econometrics, Boston, MA: Harvard University Press.

Berkson, J. (1944), "Application of the Logistic Function to Bio-Assay," Journal of the American Statistical Association, 39, 357–365.

Berndt, E. K., Hall, B. H., Hall, R. E., and Hausman, J. A. (1974), "Estimation and Inference in Nonlinear Structural Models," Annals of Economic and Social Measurement, 3, 653–665.

Billingsley, P. (1986), Probability and Measure, Wiley.

Cox, D. R. (1972), "Regression Models and Life-Tables," Journal of the Royal Statistical Society B, 34, 187–220.

Cox, D. R. (1975), "Partial Likelihood," Biometrica, 62, 269-279.

Cox, D. R., and Oakes, D. (1984), Analysis of Survival Data, London: Chapman & Hall.

Goldfeld, S., and Quandt, R. (1965), "Some Tests for Homoskedasticity," Journal of the American Statistical Association, 60, 539–547.

Han, A. K., and Hausman, J. A. (1990), "Flexible Parametric Estimation of Duration and Competing Risk Models," Journal of Applied Econometrics, 5, 1–28.

Hausman, J. A. (1978), "Specification Tests in Econometrics," Econometrica, 46, 1251-1272.

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

Kiefer, N. M. (1988), "Analysis of Grouped Duration Data," Contemporary Mathematics, 80, 107–137.

Kullback, S. and R. A. Leibler, 1951, "On Information and Sufficiency," Annals of Mathematical Statistics, 22, 79–86.

Lin, D. Y. (1991), "Goodness-of-Fit Analysis for the Cox Regression Model Based on a Class of Parameter Estimators," Journal of the American Statistical Association, 86, 725–728.

Moffitt, R. (1985), "Unemployment Insurance and the Distribution of Unemployment

Spells," Journal of Econometrics, 28, 85–101.

Mortenson, D. T. (1986), "Job Search and Labor Market Analysis," in: Handbook of Labor Economics II, eds., O. Ashenfelter, and R. Layard, Elsevier Science Publishers, BV, 849–919.

Prentice, R., and Gloeckler, L. (1978), "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data," Biometrics, 34, 57–67.

Royden, H. L. (1968), Real Analysis, NY: Macmillan.

Sueyoshi, G. T. (1991), "Evaluating Simple Alternatives to the Proportional Hazard Model: Unemployment Insurance Receipt and the Duration of Unemployment," mimeo., San Diego, CA: University of California.

Thompson, W. A. (1977), "On the Treatment of Grouped Observations in Life Studies," Biometrics, 33, 463–470.

Table 1. Min χ^2 Estimators: Empirical Mean (S.d.) Based on 1000 Repetitions

group size	$10\hat{\beta}_1^*$	$10\hat{\beta}_2^*$	$10\hat{\beta}_3^*$
(n_g)	$10\hat{\beta}_1$	$10\hat{\beta}_2$	$10\hat{\beta}_3$
20	1.86(0.64)	-4.66(1.44)	7.53(0.94)
	1.85(0.62)	-4.60(1.40)	7.38(0.93)
50	1.98(0.43)	-4.91(0.93)	7.86(0.69)
	1.96(0.41)	-4.86(0.90)	7.76(0.66)
100	1.99(0.30)	-4.98(0.66)	7.94(0.47)
	1.97(0.29)	-4.95(0.63)	7.89(0.45)
200	1.99(0.20)	-5.00(0.47)	7.97(0.32)
	1.99(0.19)	-4.99(0.45)	7.94(0.32)
500	2.00(0.13)	-5.00(0.29)	7.99(0.21)
	2.00(0.13)	-5.00(0.28)	7.98(0.21)
1000	2.00(0.09)	-5.00(0.21)	8.00(0.15)
	2.00(0.09)	-4.99(0.20)	7.99(0.14)
2000	2.00(0.07)	-5.00(0.15)	8.00(0.11)
	2.00(0.06)	-5.00(0.14)	7.99(0.10)

1. True values: $10(\beta_1, \beta_2, \beta_3) = (2, -5, 8)$.
2. The upper half are estimates based on $I = I_1 \cup I_2 = [0, 2)$.
3. The lower half are estimates based on $I_1 = [0, 1)$, $I_2 = [1, 2)$.

Table 2. Estimates of $Var(\hat{\beta}^*)$: Emp. (Thy.)

Emp. Estimates Are Based on 1000 Repetitions

n_g		$10\hat{\beta}_1^*$	$10\hat{\beta}_2^*$	$10\hat{\beta}_3^*$
20	$10\hat{\beta}_1^*$	0.410(0.447)	-0.035(-0.039)	0.028(0.034)
	$10\hat{\beta}_2^*$		2.081(2.210)	-0.008(-0.120)
	$10\hat{\beta}_3^*$			0.889(1.073)
50	$10\hat{\beta}_1^*$	0.181(0.183)	-0.015(-0.005)	0.016(0.020)
	$10\hat{\beta}_2^*$		0.857(0.915)	-0.015(-0.038)
	$10\hat{\beta}_3^*$			0.474(0.446)
100	$10\hat{\beta}_1^*$	0.091(0.090)	-0.001(-0.008)	0.008(0.006)
	$10\hat{\beta}_2^*$		0.432(0.443)	-0.009(-0.019)
	$10\hat{\beta}_3^*$			0.220(0.216)

Table 3. Estimates of $Var(\hat{\beta})$: Emp. (Thy.)

Emp. Estimates Are Based on 1000 Repetitions

n_g		$10\hat{\beta}_1$	$10\hat{\beta}_2$	$10\hat{\beta}_3$
20	$10\hat{\beta}_1$	0.390(0.417)	-0.014(-0.025)	0.028(0.024)
	$10\hat{\beta}_2$		1.946(2.045)	0.008(-0.080)
	$10\hat{\beta}_3$			0.862(1.011)
50	$10\hat{\beta}_1$	0.166(0.169)	-0.018(-0.003)	0.013(0.013)
	$10\hat{\beta}_2$		0.803(0.835)	-0.010(-0.026)
	$10\hat{\beta}_3$			0.434(0.415)
100	$10\hat{\beta}_1$	0.085(0.084)	0.002(-0.005)	0.008(0.004)
	$10\hat{\beta}_2$		0.394(0.432)	-0.005(-0.013)
	$10\hat{\beta}_3$			0.206(0.204)

Table 4. Min χ^2 Estimators:

Emp. Mean Based on 1000 Repetitions

$10\Delta'$	n_g	$10\hat{\beta}_1^*$ $10\hat{\beta}_1$	$10\hat{\beta}_2^*$ $10\hat{\beta}_2$	$10\hat{\beta}_3^*$ $10\hat{\beta}_3$
(-0.5,0,0)	100	1.71	-4.96	7.94
		1.76	-4.94	7.90
	200	1.71	-4.99	8.00
		1.76	-4.98	7.98
	500	1.72	-5.00	7.99
		1.78	-5.00	7.99
	1000	1.71	-5.00	8.00
		1.77	-5.01	8.00
	2000	1.71	-5.00	8.00
		1.77	-5.00	8.00
(0,-1,0)	100	1.98	-5.55	7.95
		1.97	-5.38	7.90
	200	1.99	-5.54	7.97
		1.98	-5.40	7.94
	500	1.99	-5.58	7.99
		1.99	-5.44	7.96
	1000	2.00	-5.56	7.99
		1.99	-5.43	7.98
	2000	2.00	-5.57	8.00
		2.00	-5.43	7.99

Table 4. (continued)

$10\Delta'$	n_g	$10\hat{\beta}_1^*$ $10\hat{\beta}_1$	$10\hat{\beta}_2^*$ $10\hat{\beta}_2$	$10\hat{\beta}_3^*$ $10\hat{\beta}_3$
(0,0,-1)	100	1.99	-4.98	7.38
		1.98	-4.96	7.43
	200	2.00	-4.96	7.41
		2.00	-4.97	7.50
	500	1.99	-4.99	7.43
		2.00	-5.00	7.54
	1000	2.00	-4.99	7.43
		2.00	-5.01	7.53
	2000	2.00	-5.00	7.44
		2.00	-5.01	7.55
(0,-1,-1)	100	1.99	-5.50	7.40
		1.98	-5.36	7.46
	200	1.99	-5.52	7.43
		1.99	-5.39	7.51
	500	2.00	-5.57	7.42
		2.00	-5.45	7.52
	1000	2.00	-5.55	7.42
		2.00	-5.44	7.52
	2000	2.00	-5.56	7.44
		2.00	-5.45	7.54

1. Estimating misspecified models by wrongly assuming PHM.
2. The upper half are estimates based on $I = I_1 \cup I_2 = [0, 2)$.
3. The lower half are estimates based on $I_1 = [0, 1)$, $I_2 = [1, 2)$.
4. True values: $10\beta^1 = (2, -5, 8)$ on I_1; $10(\beta^1 + \Delta)$ on I_2.

Table 5. Emp. Size and Power (%):

Overall χ^2 Test

$10\Delta'$		nominal size (%)		
KL$\times 10^4$	**n_g**	**10**	**5**	**1**
(0, 0, 0)	20	21	13	4
0.0	50	18	11	4
	100	11	6	2
	200	11	5	1
	500	12	6	2
	1000	10	5	1
	2000	9	5	1
(-0.5, 0, 0)	100	17	12	3
1.2	200	22	13	4
	500	40	27	10
	1000	61	50	27
	2000	90	84	64

Table 5. (continued)

$10\Delta'$		nominal size (%)		
$KL \times 10^4$	n_g	10	5	1
(0, -1, 0)	100	21	13	5
0.99	200	19	12	3
	500	33	23	10
	1000	57	45	21
	2000	86	77	53
(0, 0, -1)	100	19	10	3
2.1	200	26	18	6
	500	52	39	20
	1000	82	72	50
	2000	99	96	88
(0, -1,-1)	100	24	16	5
3.2	200	35	25	12
	500	75	64	39
	1000	96	93	82
	2000	100	100	99

Table 6. Emp. Size and Power (%):

Individual one-sided t-test

		β_1			β_2			β_3		
$10\Delta'$		**nominal size**			**nominal size**			**nominal size**		
KL$\times 10^4$	n_g	**10**	**5**	**1**	**10**	**5**	**1**	**10**	**5**	**1**
(0, 0, 0)	20	11	5	1	16	10	3	7	4	1
0.0	50	10	5	1	16	9	1	5	3	1
	100	7	4	1	15	7	1	5	2	0
	200	10	5	1	12	6	1	6	3	1
	500	9	5	1	12	7	2	8	4	0
	1000	12	7	1	11	6	1	9	4	1
	2000	9	5	1	9	5	1	9	4	1
(-0.5, 0, 0)	100	28	16	5	13	7	1	6	3	0
1.2	200	38	24	9	13	6	1	7	4	1
	500	64	50	24	11	6	1	10	5	1
	1000	84	75	49	11	6	1	11	5	1
	2000	98	96	84	10	5	1	12	6	1

Table 6. (continued)

		β_1			β_2			β_3		
$10\Delta'$		nominal size			nominal size			nominal size		
$KL\times10^4$	n_g	10	5	1	10	5	1	10	5	1
(0, -1, 0)	100	9	4	1	34	22	8	7	4	1
0.99	200	9	4	1	40	25	8	6	3	1
	500	8	4	1	61	47	21	6	3	0
	1000	7	4	1	83	73	47	5	2	0
	2000	8	3	1	97	94	78	5	2	0
(0, 0, -1)	100	10	5	1	13	8	1	22	13	4
2.1	200	13	7	2	12	7	2	39	27	11
	500	13	8	2	9	4	1	75	64	36
	1000	14	8	2	7	3	1	96	91	72
	2000	17	9	3	6	3	0	100	100	98
(0, -1, -1)	100	11	6	2	31	21	7	21	12	4
3.2	200	10	5	1	38	28	10	36	25	9
	500	11	6	2	55	42	20	67	54	31
	1000	13	7	2	72	58	33	91	85	63
	2000	14	8	2	91	84	62	100	99	93

Table 7. Emp. Size and Power (%):

Individual two-sided t-test

		β_1			β_2			β_3		
$10\Delta'$		**nominal size**			**nominal size**			**nominal size**		
KL$\times 10^4$	n_g	**10**	**5**	**1**	**10**	**5**	**1**	**10**	**5**	**1**
(0, 0, 0)	20	14	7	2	16	9	3	19	12	4
0.0	50	13	7	3	14	8	1	20	12	4
	100	10	6	1	10	5	1	12	7	2
	200	11	6	2	10	5	1	12	6	1
	500	12	6	1	13	7	1	13	6	1
	1000	13	6	1	11	7	1	9	4	1
	2000	10	5	1	10	4	1	10	4	1
(-0.5, 0, 0)	100	17	10	3	11	5	1	13	7	2
1.2	200	25	16	6	11	5	1	12	6	2
	500	50	38	18	11	6	1	11	6	1
	1000	75	64	39	11	6	2	10	6	1
	2000	96	92	78	11	6	2	10	5	1

Table 7. (continued)

		β_1			β_2			β_3		
$10\Delta'$		nominal size			nominal size			nominal size		
$KL\times10^4$	n_g	10	5	1	10	5	1	10	5	1
(0, -1, 0)	100	12	7	2	24	14	5	15	9	3
0.99	200	10	6	2	26	15	5	12	7	2
	500	10	5	1	47	34	16	13	7	2
	1000	9	5	1	73	62	38	12	6	1
	2000	8	4	1	94	88	70	14	7	1
(0, 0, -1)	100	13	8	1	12	6	1	16	8	3
2.1	200	13	7	1	14	8	2	27	19	7
	500	14	7	2	11	7	1	64	50	27
	1000	12	6	2	12	6	2	91	83	63
	2000	13	7	3	13	8	2	100	99	94
(0, -1, -1)	100	15	8	2	22	15	5	15	8	2
3.2	200	11	5	2	28	17	7	25	17	6
	500	11	5	1	42	32	13	54	44	21
	1000	11	6	1	58	45	25	85	77	53
	2000	13	7	1	84	75	55	99	97	89