Discussion Paper No. 552

MAXIMUM LIKELIHOOD ESTIMATION AND SPECIFICATION TEST OF THE PROPORTIONAL HAZARD MODEL USING GROUPED DURATIONS

Keunkwan Ryu

August 2001

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Maximum Likelihood Estimation and Specification Test of the Proportional Hazard Model Using Grouped Durations

Keunkwan Ryu*

August 1997
Revised Aug. 2001

Abstract

This article develops parametric and semi-parametric, maximum likelihood estimation methods of the proportional hazard model for the case when durations are grouped and heterogeneity is not fully measured. This article also extends Ryu's (1994b) specification tests for the proportional hazard model. The proposed test statistic is easy to implement, takes into account the form of alternative hypothesis to increase power, and identifies the source and direction of non-proportionality. This article identifies the nature of the bias resulting from neglected heterogeneity. This bias is essentially a sample selection bias, well-known in the literature. Gamma and discrete distributions are used to model unobserved heterogeneity. By updating the heterogeneity distribution at each stage of duration process, this article addresses the sample selection bias. The results obtained in this article are applied to tackle left-censoring problem.

KEY WORDS: Binary choice model; Grouped duration; Hausman's specification test; Invariance property; Maximum likelihood estimation; Proportional hazard model; Unobserved heterogeneity; Left-censoring.

* Keunkwan Ryu is Associate Professor, Department of Economics, Seoul National University, Seoul 151-742, KOREA, and a Visiting Foreign Scholar, ISER, Osaka University. The author is grateful to Sunil Sharma for constructive suggestions and comments. This research was supported by a Korea Research Foundation grant.

1. INTRODUCTION

In duration analysis, we often face a situation with continuous models and discrete data. It is natural to think of real time as continuous and that events can occur at any moment in time. However, duration is often measured in intervals, not the exact time elapsed, by the nature of the observation scheme. For example, many economic duration variables constructed from longitudinal surveys are at most known only up to weekly intervals.

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 the PHM with unobserved heterogeneity, the hazard rate is specified as a product of three separate terms: a baseline hazard function describing the overall shape of the hazard rate over time, a proportionality factor capturing the covariate (regression) effects across different individuals, and a random variable representing unobserved heterogeneity (hereafter, heterogeneity).

Maximum likelihood estimation (MLE) of the grouped duration models has been suggested by Thompson (1977), Prentice and Gloeckler (1978), Kiefer (1988), and Sueyoshi (1991). The first paper views a grouped duration as a sequence of binary survival indicators that follow an independent Logit probability model. As discussed in Ryu (1994b), Thompson's parameterization is quite different from the conventional proportional hazard model. The other three papers consider MLE of the PHM using grouped duration data. Additionally, Kiefer and Sueyoshi develop likelihood ratio and Lagrange multiplier tests for the proportionality assumption.

In this article, we develop a maximum likelihood estimation method of the PHM for the case in which durations are grouped and unmeasured heterogeneity exists. The motivation is that for much of the available survey data, duration variables are often interval-censored, while many of the covariates are unobserved. This article also extends Ryu's (1994b) specification tests for proportionality. The suggested tests are easy to use, take alternative hypotheses into account to increase their power, and identify the source and direction of non-proportionality without imposing a priori restrictions.

In addition, this article suggests a flexible, and tractable parameterization of the baseline hazard function. In the grouped duration context, the suggested parameterization can nest the non-parametric baseline hazard function as a special case. This article also investigates the nature of the bias resulting from neglected heterogeneity. It is essentially a sample selection bias, well-known in the literature. Gamma and discrete distributions are used to capture heterogeneity.

Let us briefly sketch the estimation and specification test ideas. Grouped duration can be viewed as a sequence of binary variables indicating whether the duration survives each interval or not. By constructing a synthetic binary data set treating each combination (individual, interval) as a new unit of indexation, we can reduce a grouped duration analysis to a sequential binary choice analysis.

If there exists heterogeneity, one needs to integrate the random variable representing heterogeneity out of each interval survival probability. Here, one has to take into account selection over time of the underlying heterogeneity. It is because the sixth year graduate students cannot be the same as the entering graduate students in terms of their underlying type distribution. This selection should be accounted for to avoid sample selection bias, a bias due to neglected heterogeneity.

To test the proportionality assumption, we can further aggregate the already grouped data. If proportionality holds, the two estimators, one from the further grouped data and the other from the original grouped data, will converge to the same quantity; however, if proportionality is violated, they will diverge from each other. Therefore, a test of the equality of these two sets of estimators will offer a test for the proportionality assumption.

Finally, to illustrate the detailed aspects of the new suggested estimation and test procedures, this article explicitly considers the following five combinations regarding (i) whether to use parametric or non-parametric baseline hazard specification and (ii) whether and how to allow for unobserved heterogeneity: (parametric baseline, Gamma heterogeneity), (parametric baseline, discrete heterogeneity), (non-parametric baseline, no heterogeneity), (non-parametric baseline, Gamma heterogeneity), and (non-parametric baseline, discrete heterogeneity).

The rest of the article is organized as follows. In Section 2, after setting up a framework for discussion, the relationship between group duration analysis and sequential binary choice analysis is shown. Also, the basics of MLE are provided in the grouped duration context. In Section 3, heterogeneity is introduced, and operationalized within the sequential binary choice representation of the grouped duration. Proportionality tests are developed in Section 4. With flexible specification of the baseline hazard function, estimation and test procedures are illustrated in detail in Section 5. In Section 6, left-censoring will be addressed using the suggested framework. Concluding remarks follow in Section 7.

2. FRAMEWORK

Let $T \in R^+$ represent a duration variable of interest. Let

$$h(t|x, v) = h_0(t) \exp(x'\beta)v, \ \ v \sim g(v), \ \ v > 0, \tag{1}$$

be the hazard rate of duration T, where $h_0(t)$ is a baseline hazard function, $\exp(x'\beta)$ is a proportionality factor, x and β are $k \times 1$ vectors of observed covariates and regression coefficients, and v captures unmeasured heterogeneity through density $g(v)$. This model can be termed as a PHM with unobserved heterogeneity.

To complete the model, one has to specify the baseline hazard function and the heterogeneity distribution. A simple way of treating the baseline hazard function is to make a parametric assumption. In the literature, exponential and Weibull functional forms have been most popular. In this article, we offer an alternative flexible parameterization together with the so called non-parametric treatment. Within the current grouped duration setting, our suggested parameterization can nest the non-parametric baseline as a special case.

The proportionality of the specification (1) refers to the constancy of β. That is, the covariate x increases or decreases the hazard rate by the same proportion throughout duration. This assumption may be too strong in some situations. Ryu (1994b) examines two empirical studies where the proportionality assumption is skeptical, and proposes a new proportionality test within a minimum χ^2 estimation context. The test in this article extends Ryu's (1994b) by generalizing the observation and aggregation schemes, by introducing heterogeneity, and by considering general covariate types.

If v degenerates to a constant, there is no unobserved heterogeneity. It is well known in the literature that neglected or mis-specified heterogeneity leads to biased estimation. To capture heterogeneity, Gamma and discrete distributions have been most widely used (see Heckman and Singer 1984 and Nickel 1979). These distributions are easy to use within the proportional hazard specification in (1). In particular, the discrete distribution can approximate any unknown distribution as the number of mass points increases with sample size.

For the identification of level in the hazard specification in (1), we need to fix level for two of three terms in the specification, $h_0(t)$, $\exp(x'\beta)$, and v. We will leave $h_0(t)$ free, and fix a level for the other two terms by (i) excluding a constant term from $x'\beta$, and by (ii) imposing one restriction on $g(v)$ (For details, see Section 4.)

A discrete observation scheme can often be represented as an equi-spaced partition Q of the support R^+: $Q = \{0, l, 2l, \cdots, rl, \infty\}$. Under Q, the researcher keeps a record of individuals' status at every l time units, until time rl elapses. Without loss of generality, let us assume $l = 1$. This assumption corresponds to taking l as the measurement unit for duration. Let $I_j = [j-1, j)$, $j = 1, \cdots, r$, and $I_{r+1} = [r, \infty)$. For each observation falling within one of r non-right-censored intervals $I_1, \cdots, I_r$, we know its duration up to a unit interval; for each observation falling within the right-censored interval I_{r+1}, we only knows its lower bound.

Let $\alpha_j(x, v)$ be the probability that T survives I_j conditional on that it has already survived all previous intervals. Then we have

$$\alpha_j(x, v) = \exp[-\int_{j-1}^{j} h(t|x, v)\, dt] = \exp[-\exp(x'\beta + \gamma_j)v], \tag{2}$$

where

$$\gamma_j = \log[\int_{j-1}^{j} h_0(t)\, dt],\ j = 1, \cdots, r. \tag{3}$$

These formulas were originally provided by Prentice and Gloeckler (1978) for the case $v = 1$ with probability one.

Under a parametric baseline hazard specification, the γ_j's will be functionally related. On the other hand, under a non-parametric baseline, the γ_j's will be treated as r free parameters. Sometimes, grouped duration data involve intervals of unequal length, typically with wider intervals at longer durations. Our method are easily extended to this case by redefining the γ_j's in equation (3).

Let us assume that there are n independent observations. Let i index each different observation: $i = 1, \cdots, n$. Define T_i as the ith duration variable, x_i as the covariate of individual i, and $d_{ji} = 1$ if T_i survives I_j conditional on $T_i > j - 1$, $d_{ji} = 0$ otherwise, $j = 1, \cdots, r$. Then, a grouped duration can be considered as a sequence of binary indicator variables, that is, $T_i \iff (d_{1i}, \cdots, d_{ri})$. The effective number of terms in the sequence varies depending on at which interval the individual dies. Note that d_{2i}'s are meaningfully defined only for those who have survived I_1. By constructing a synthetic binary data set treating each combination (individual, interval) as a new unit of indexation, we can reduce a grouped duration analysis to a sequential binary choice analysis (Kiefer 1988; Prentice and Gloeckler 1978; Ryu 1994b; Sueyoshi 1991; Thompson 1977). For each combination (individual, interval), a survivor of the jth interval receives the probability

$$\alpha_j(x) = P(T > j | T > j - 1, x) = E_j \alpha_j(x, v) \tag{4}$$

if he or she has covariate x, where E_j denotes taking expectation over v using the jth stage heterogeneity density, say, $g_j(v)$. Since $\alpha_j(x)$ is only defined for those who satisfy $T > j - 1$, the jth stage density $g_j(v)$ should take this selection into account.

The log-likelihood contribution of the ith individual is:

$$l_i = \sum_{j=1}^{r} s_{ji}[d_{ji} \log \alpha_{ji} + (1 - d_{ji}) \log(1 - \alpha_{ji})], \tag{5}$$

where s_{ji} is defined by $s_{ji} = d_{1i} \times \cdots \times d_{j-1i}$ with the convention $s_{1i} = 1$, and α_{ji} is a short-hand notation for $\alpha_j(x_i)$. This representation shows a similarity between grouped duration analyses and sequential binary choice analyses. Passing I_1, one of two outcomes occurs: to survive or not; Conditional on that the individual has survived I_1 (conditional on $d_{1i} = s_{2i} = 1$), again one of the same binary outcomes happens, and so forth. Here, the cumulative survival indicator s_{ji} controls whether or not the jth interval is effective for the ith individual: effective if $s_{ji} = 1$ and not effective otherwise.

By adding l_i over i, we derive the log-likelihood function for the whole sample:

$$L(\theta) = \sum_{i=1}^{n} l_i = \sum_{i=1}^{n} \sum_{j=1}^{r} s_{ji}[d_{ji} \log \alpha_{ji} + (1 - d_{ji}) \log(1 - \alpha_{ji})], \tag{6}$$

where θ denotes the collection of all model parameters such as β, and parameters in $h_0(t)$ and $g(v)$. The maximum likelihood estimator of θ, say $\hat{\theta}$, is defined as the argument maximizing (6), or alternatively as the argument solving the first-order condition: $\partial L(\hat{\theta})/\partial\theta = 0$. By expanding the first-order condition through Taylor series and rearranging terms, we derive

$$\hat{\theta} - \theta \stackrel{\text{D}}{=} -[\frac{\partial^2 L}{\partial\theta\partial\theta'}]^{-1} \frac{\partial L(\theta)}{\partial\theta}, \tag{7}$$

where '$=^{\text{D}}$' means that both hand sides of $=^{\text{D}}$ have the same asymptotic distribution as $n \to \infty$ up to $\sqrt{n}$ order (for details, see Amemiya 1985).

By taking the derivative of (5) and adding up, we obtain the score function as

$$\frac{\partial L}{\partial\theta}(\theta) = \sum_{i=1}^{n} \frac{\partial l_i}{\partial\theta} = \sum_{i=1}^{n} \sum_{j=1}^{r} s_{ji} \frac{d_{ji} - \alpha_{ji}}{\alpha_{ji}(1 - \alpha_{ji})} \frac{\partial\alpha_{ji}}{\partial\theta}. \tag{8}$$

By exploiting (i) information matrix equality, (ii) law of large numbers, (iii) $E(d_{ji} - \alpha_{ji})^2 = \alpha_{ji}(1 - \alpha_{ji})$, and (iv) independence of d_{ji} across both j and i, we can derive

$$-\frac{\partial^2 L}{\partial\theta\partial\theta'} \stackrel{\text{P}}{=} \sum_{i=1}^{n} \frac{\partial l_i}{\partial\theta} \frac{\partial l_i}{\partial\theta'} \stackrel{\text{P}}{=} \sum_{i=1}^{n} \sum_{j=1}^{r} s_{ji} \frac{1}{\alpha_{ji}(1 - \alpha_{ji})} \frac{\partial\alpha_{ji}}{\partial\theta} \frac{\partial\alpha_{ji}}{\partial\theta'}, \tag{9}$$

where '$=^p$'means that both sides of '$=^p$' have the same probability limit as $n \to \infty$ after being divided by n.

By plugging in (8) and (9) into (7), we can represent the maximum likelihood estimator $\hat{\theta}$ in an asymptotically equivalent form. As is clear from the above calculation, all we have to know is the interval survival probability α_{ji} and its derivative $\partial\alpha_{ji}/\partial\theta$. For this purpose, we need to complete the model by specifying the heterogeneity density $g(v)$ and the baseline hazard function $h(t)$.

The next section introduces heterogeneity within the sequential binary choice framework. This offers an operationally convenient way of dealing with heterogeneity. Also, this treatment will make clear the nature of the bias resulting from neglected heterogeneity. The bias is a sample selection bias, well-known in the literature.

3. UNOBSERVED HETEROGENEITY

We keep the sequential binary choice representation of the grouped duration. Accordingly, we are interested in updating the heterogeneity density $g_j(v)$ as time passes, and in computing each interval survival probability $\alpha_j(x)$ using the jth stage (interval I_j) density $g_j(v)$, $j = 1, \cdots, r$.

Given a density specification for the heterogeneity $v \sim g(v)$, we can use this density to integrate out the heterogeneity v from those *unconditional* quantities such as survival function, distribution function, and density function. However, we cannot use the same density $g(v)$ to integrate out the heterogeneity term v from those *conditional* quantities such as hazard rate, and interval survival probability. It is because the information contained in the conditioning statement has an implication on the heterogeneity. For example, those who are staying in the graduate program longer will be different from the entering class in terms of their latent type distribution, that is heterogeneity. Over time, "diligent" students will finish the program, whereas "lazy" students will still hang around, so called weeding out effect. This effect reflects selection over time.

Defining $c_j = \exp(x'\beta + \gamma_j)$, we have from (2)

$$\alpha_j(x,v) = P(T > j | T > j-1, x, v) = \exp[-\exp(x\beta + \gamma_j)v] = \exp(-c_j v). \tag{10}$$

Due to the selection over time, the jth stage heterogeneity density $g_j(v)$ will be different from the initial density, $g(v)$. The conditional density $g_j(v)$ is given by

$$g_j(v) = g(v|T > j-1, x) = \frac{P(T > j-1|x,v)g(v)}{\int_0^\infty P(T > j-1|x,v)g(v)\,dv} = \frac{e^{-(c_1+\cdots+c_{j-1})v}g(v)}{M_v(c_1+\cdots+c_{j-1})}, \tag{11}$$

where $M_v(t) = E_v(e^{-tv})$ is the moment generating function of $g(v)$. Note that $g_1(v) = g(v)$. Since, each c_j is positive, we can easily see that $g_j(v)$ is first-order stochastically decreasing in j. Let v_j be a random variable having density function $g_j(v_j)$. Then, v_j captures the heterogeneity in the jth interval, and v_j first-order stochastically dominates $v_{j'}$ for all $j < j'$.

Alternative way of deriving $g_j(v)$ is to rely on a pure mathematical identity. Let $S(j|x,v)$ and $S(j|x)$ be the survival probability that T exceeds j, conditional on (x,v) and x, respectively. One has (i) $S(j|x,v) = \alpha_1(x,v)\times\cdots\times\alpha_j(x,v)$, (ii) $S(j|x) = E_v S(j|x,v) = \int_0^\infty S(j|x,v)g(v)\,dv$, and (iii) $S(j|x) = E_1\alpha_1(x,v)\times\cdots\times E_j\alpha_j(x,v)$, where E_j denotes the expectation taken with respect to $g_j(v)$. Using (i) and (ii) and arranging terms, we derive

$$\begin{aligned} S(j|x) &= \int_0^\infty [\alpha_1(x,v)\times\cdots\times\alpha_j(x,v)]g(v)\,dv \\ &= \int_0^\infty \alpha_1(x,v)g(v)\,dv \times\cdots\times \int_0^\infty \alpha_j(x,v)\frac{\alpha_1(x,v)\times\cdots\times\alpha_{j-1}(x,v)g(v)}{\int_0^\infty \alpha_1(x,v)\times\cdots\times\alpha_{j-1}(x,v)g(v)\,dv}\,dv. \end{aligned} \tag{12}$$

Now, by matching (iii) and (12) term by term, we can easily see that

$$\begin{aligned} g_j(v) &= \alpha_1(x,v)\times\cdots\times\alpha_{j-1}(x,v)g(v)\Big/\int_0^\infty \alpha_1(x,v)\times\cdots\times\alpha_{j-1}(x,v)g(v)\,dv \\ &= \exp[-(c_1+\cdots+c_{j-1})v]g(v)\Big/\int_0^\infty \exp[-(c_1+\cdots+c_{j-1})v]g(v)\,dv \\ &= \exp[-(c_1+\cdots+c_{j-1})v]g(v)/M_v(c_1+\cdots+c_{j-1}). \end{aligned} \tag{13}$$

Using the updated heterogeneity density in (13), we can compute the marginal (with respect to v) interval survival probability $\alpha_j(x)$:

$$\begin{aligned}\alpha_j(x) = E_j\alpha_j(x,v) &= \int_0^\infty \alpha_j(x,v)g_j(v)\,dv \\ &= \int_0^\infty \exp[-(c_1+\cdots+c_{j-1}+c_j)v]g(v)\,dv/M_v(c_1+\cdots+c_{j-1}) \\ &= M_v(c_1+\cdots+c_{j-1}+c_j)/M_v(c_1+\cdots+c_{j-1})\end{aligned} \tag{14}$$

with $\alpha_1(x) = M_v(c_1)$. The interval survival probability in (14) is represented as a ratio of the moment generating functions of $g(v)$, evaluated at two different points. Therefore, we readily notice that any heterogeneity density $g(v)$ will yield an analytical solution for the interval survival probability $\alpha_j(x)$ insofar as $g(v)$ admits an analytical moment generating function.

Now, we would like to specialize the above formulas (13) and (14) for the Gamma and discrete distributions. Let us consider the Gamma heterogeneity first. Assume that v follows a Gamma distribution with parameters a and b, denoted Gamma(a,b): $v \sim g(v) = b^a v^{a-1}e^{-bv}/\Gamma(a)$, where $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}\,dt$. It is easy to show that the moment generating function is $M_v(t) = (\frac{b}{b+t})^a$. In particular, the mean and variance are $Ev = a/b$ and $Var(v) = a/b^2$. To fix the level of v, let us assume that $Ev = 1$. This identification condition imposes $a = b$. That is, the heterogeneity is modeled through a Gamma distribution with mean one and variance $1/b$.

From (13),

$$g_j(v) = \frac{e^{-(c_1+\cdots+c_{j-1})v}g(v)}{M_v(c_1+\cdots+c_{j-1})} = \frac{b^b e^{-(b+c_1+\cdots+c_{j-1})v}v^{b-1}}{\Gamma(b)M_v(c_1+\cdots+c_{j-1})} = \frac{b_j^b v^{b-1}e^{-b_j v}}{\Gamma(b)}, \tag{15}$$

where $b_j = b+c_1+\cdots+c_{j-1}$ with $b_1 = b$. Note that $g_j(v)$ is the density corresponding to a Gamma distribution with parameters b and b_j, denoted Gamma(b,b_j). Since $E_j(v) = b/b_j$ and $Var_j(v) = b/b_j^2$, we observe that $g_j(v)$ exhibits decreasing mean and variance as j increases. That is, adverse but homogenizing selection is going on over time in terms of heterogeneity distribution. Now, from (14),

$$\alpha_j(x) = \frac{M_v(c_1+\cdots+c_{j-1}+c_j)}{M_v(c_1+\cdots+c_{j-1})} = \left(\frac{b+c_1+\cdots+c_{j-1}}{b+c_1+\cdots+c_{j-1}+c_j}\right)^b. \tag{16}$$

Next, let us consider a discrete heterogeneity distribution following Heckman and Singer (1984). Assume that v takes M finite values $v_1, \cdots, v_M$ with probabilities $p_1, \cdots, p_M$. This distribution describes that there are M unobserved types. For level identification, let us fix $v_M = 1$. For v_m, the following parameterization is convenient:

$$v_M = 1 \text{ (level normalization)},\ v_m = e^{w_m},\ m = 1, \cdots, M-1, \tag{17}$$

where w_m's are unrestricted, $-\infty < w_m < \infty$. For p_m, the following parameterization is again useful:

$$p_M = \frac{1}{1 + \sum_{m=1}^{M-1} e^{\pi_m}},\ p_m = \frac{e^{\pi_m}}{1 + \sum_{m=1}^{M-1} e^{\pi_m}},\ m = 1, \cdots, M-1, \tag{18}$$

where π_m's are unrestricted too, $-\infty < \pi_m < \infty$. The moment generating function of this discrete distribution is $M_v(t) = E_v e^{-tv} = \sum_{m=1}^{M} e^{-tv_m} p_m$.

Again, from (13), we derive

$$g_j(v_m) = \frac{e^{-(c_1+\cdots+c_{j-1})v_m} p_m}{\sum_{m=1}^{M} e^{-(c_1+\cdots+c_{j-1})v_m} p_m},\ m = 1, \cdots, M. \tag{19}$$

Also, from (14), we obtain

$$\alpha_j(x) = \frac{M_v(c_1 + \cdots + c_{j-1} + c_j)}{M_v(c_1 + \cdots + c_{j-1})} = \frac{\sum_{m=1}^{M} e^{-(c_1+\cdots+c_{j-1}+c_j)v_m} p_m}{\sum_{m=1}^{M} e^{-(c_1+\cdots+c_{j-1})v_m} p_m},\ j = 1, \cdots, r. \tag{20}$$

We can carry out the maximum likelihood estimation of the grouped PHM with heterogeneity by maximizing the log-likelihood function (6) after plugging in $\alpha_{ji} = \alpha_j(x_i)$. Under a set of quite general regularity conditions (see Amemiya 1985, ch. 4), the maximum likelihood estimator will converge to a normal distribution with mean equal to true parameters and variance matrix equal to the inverse of the information matrix. The information matrix can be consistently estimated using (9) by replacing the unknown parameters with their estimates.

4. TEST FOR PROPORTIONALITY

The popularity of the PHM has made the issue of model checking extremely important. This section extends Ryu's (1994b) test for proportionality to the general grouped duration framework with discrete/continuous covariates and observed/unobserved heterogeneity. Ryu's (1994b) framework is rather limited in the sense that it only considers observed categorical covariates, neglecting unmeasured heterogeneity and continuous covariates.

By further aggregating the already grouped duration data, we can artificially generate another coarser set of grouped duration data. The aggregation can be represented as a contiguous grouping of integers, $\{1, \cdots, r\}$. Let $Q^* = \{G_1, \cdots, G_{r^*}\}$, where $G_k = \{g_1 + \cdots + g_{k-1} + 1, \cdots, g_1 + \cdots + g_{k-1} + g_k\}$ with g_k being the number of elements in the subset G_k, $g_k \geq 1$, $k = 1, \cdots, r^*$. Then, Q^* corresponds to the following aggregation scheme: aggregate the first g_1 intervals into a big interval, say I_1^*; aggregate the next g_2 intervals into a big interval, say I_2^*; continue in the same way until we obtain the last big interval $I_{r^*}^*$. At least one of the g_k's must be greater than one. Otherwise, no aggregation occurs. Under Q^*, the duration data will take the form $d_{ki}^* = \Pi_{j \in G_k} d_{ji}$. Here, d_{ki}^* takes value one if individual i survives the big interval $I_k^* = \cup_{j \in G_k} I_j$ and zero otherwise. Moreover, $\alpha_{ki}^* = \Pi_{j \in G_k} \alpha_{ji} = \exp[-\exp(x\beta + \gamma_k^*)]$ is the survival probability of the big interval I_k^*, where $\gamma_k^* = \log \int_{t \in I_k^*} h_0(t)\, dt = \log(\sum_{j \in G_k} e^{\gamma_j})$, $k = 1, \cdots, r^*$. Obviously, the new coarser data set Q^* contains less information than the original finer data set Q. Let $\hat{\beta}$ be the estimate of β obtained by using the original data Q, and $\hat{\beta}^*$ the estimate obtained by using the further aggregated data Q^*. By comparing these two estimates, we can design a new proportionality test statistic.

If the PHM holds, both $\hat{\beta}$ and $\hat{\beta}^*$ will converge to the same β. However, if the PHM fails to hold, $\hat{\beta}$ and $\hat{\beta}^*$ will converge to different quantities. To see this, let us consider a simple case where $r = 2, r^* = 1, Q = \{0, 1, 2, \infty\}, Q^* = \{G_1\}$, and $G_1 = \{1, 2\}$. Assume that all covariates have decreasing impacts on the hazard rates:

$$h(t, x) = \begin{cases} h_0(t) \exp(x\beta^1), & \text{for } t \in I_1; \\ h_0(t) \exp(x\beta^2), & \text{for } t \in I_2, \end{cases}$$

with $\beta^1 > \beta^2 > 0$. Here, heterogeneity is assumed away by taking $v = 1$ with probability one. The aggregate estimator $\hat{\beta}^*$, obtained from the coarser data $I_1^* = I_1 \cup I_2$, is symmetrically affected by β^1 and β^2 since I_1^* is the symmetric aggregation of I_1 and I_2. On the other hand, the disaggregate estimator $\hat{\beta}$, obtained from the finer data set I_1 and I_2, is asymmetrically affected by β^1 and β^2 since there are more observations in I_1 than in I_2 (note that some individuals have died in I_1). Therefore, the influence of β^1 relative to β^2 is stronger on $\hat{\beta}$ than on $\hat{\beta}^*$. Since β^1 is larger than β^2, $\hat{\beta}$ should be stochastically larger than $\hat{\beta}^*$. As a result, the difference between $\hat{\beta}$ and $\hat{\beta}^*$ converges to a zero vector under the PHM, but to a non-zero vector under non-proportionality. The proposed test statistic uses this disparate convergence pattern: if the difference $\hat{\beta}^* - \hat{\beta}$ is significantly different from zero, reject the PHM; otherwise, do not.

The test statistic

$$R = (\hat{\beta} - \hat{\beta}^*)'[Var(\hat{\beta} - \hat{\beta}^*)]^{-1}(\hat{\beta} - \hat{\beta}^*) \tag{21}$$

follows a chi-square distribution with degrees of freedom equal to the number of parameters in β, say k. The variance inside the bracket can be expanded as

$$Var(\hat{\beta} - \hat{\beta}^*) = Var(\hat{\beta}) + Var(\hat{\beta}^*) - Cov(\hat{\beta}, \hat{\beta}^*) - Cov(\hat{\beta}, \hat{\beta}^*)'. \tag{22}$$

In calculating the above test statistic, the difficulty usually lies in computing the covariance matrix between the two estimators. It is so because variances can be easily estimated through the inverse of the observed information matrices.

Then, how to compute the covariance matrix? It depends on whether we adopt a parametric baseline specification or not. Obviously, the original finer data set contains bigger amount of information than the aggregated coarser data set. However, this ranking in data information contents does not necessarily yield an efficiency ranking between $\hat{\beta}$ and $\hat{\beta}^*$. If we make parametric baseline hazard assumption and thus estimate the same fixed number of baseline hazard parameters in both data set-ups, we obtain a result that $\hat{\beta}$ is more efficient than $\hat{\beta}^*$ (Ryu 1993a) and that their covariance reduces to $Var(\hat{\beta})$ (Hausman 1978). Obviously, the number of free parameters in the parametric specification should not exceed the number of non-right-censored intervals r^* in the coarser data set (see sub-section 5.5.). On the other hand, if we do not make a parametric functional form assumption on the baseline hazard, the number of baseline hazard parameters being estimated under the two data set-ups are different: when we are estimating the PHM using the finer data set, we are estimating more steps regarding the baseline hazard function ($r > r^*$). In this non-parametric case, the ranking in data information contents does not imply an efficiency ranking between $\hat{\beta}$ and $\hat{\beta}^*$. This lack of efficiency ordering prohibits one from using Hausman's (1978) result to simplify the covariance between $\hat{\beta}^*$ and $\hat{\beta}$ in the non-parametric baseline case. In the next section, we will explain how to compute the covariance matrix in the non-parametric baseline case.

Besides the overall chi-square test, we can also conduct individual t-tests. Under the proportionality assumption,

$$t_j = (\hat{\beta}_j - \hat{\beta}_j^*)/\sigma_{(\hat{\beta}-\hat{\beta}^*)j}, \ j = 1, \cdots, k \tag{23}$$

will have an asymptotic standard normal distribution, where $\hat{\beta}_j$ and $\hat{\beta}_j^*$ are the jth elements of $\hat{\beta}$ and $\hat{\beta}^*$, and $\sigma_{(\hat{\beta}-\hat{\beta}^*)j}$ is the square root of the jth diagonal element of $Var(\hat{\beta} - \hat{\beta}^*)$. The advantage of individual t-tests is to separately identify those covariates which exhibit non-proportional effects, and to inform the direction of those non-proportional effects. For instance, if t_j is significantly positive (negative), then we can conclude that the jth component of x has a non-proportional effect on the hazard rate and that its coefficient is larger (smaller) in the early intervals than in the later intervals. If the individual t-tests detect non-proportionality, then it would be a better idea to estimate a non-proportional hazard model by allowing different β's for those non-proportional covariates across different intervals.

If we have more than two non-right-censored intervals in the original data set ($r > 2$), we have a lot more flexibility in choosing a further aggregation. 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 in the original data set ($r = 3$) and if one suspects that the covariate impact is weaker in the third interval (if there is any difference), one may put the first two intervals together, leaving the third interval alone. That is, take $r^* = 2, Q^* = \{G_1, G_2\}, G_1 = \{1, 2\}, G_2 = \{3\}, I_1^* = I_1 \cup I_2$, and $I_2^* = I_3$. This aggregation will yield a higher power against the suspected alternative than any other aggregation. Of course, without a clear alternative in mind, we cannot design an optimal aggregation.

The suggested test is very easy to implement, and allows one to take into account the form of suspected alternative hypothesis to increase power of the test. Most of all, the test is much more convenient to use compared with other existing tests. To apply the likelihood ratio test, one has to estimate the model under a non-proportional alternative. If there are ten covariates ($k = 10$) and ten non-right-censored intervals ($r = 10$) in the original data, the most general form of non-proportional hazard model will include $100 (= 10 \times 10)$ β's, too many parameters! To reduce the number of parameters, one has to introduce a priori restrictions by assuming either that some covariates have proportional effects or that the non-proportional effects satisfy certain parametric restrictions. The situation is practically no better for the Lagrange multiplier test. Even though we only need to estimate the model under proportionality, we have to consider a very long vector of score functions corresponding to 100 β's. Again, we are forced to adopt the same a priori assumptions as in the likelihood ratio test to solve the dimensionality problem. However, our test identifies the source and nature of non-proportionality without imposing any a priori restrictions, a useful property not shared by the existing tests.

5. BASELINE HAZARD FUNCTION

This section specifies the baseline hazard function, and details the estimation and test procedures. Both parametric and non-parametric specifications are introduced and compared with each other. We first introduces the general framework in sub-section 5.1. In sub-sections 5.2 through 5.4, we specialize the general framework to the cases of (non-parametric baseline, no unmeasured heterogeneity), (non-parametric baseline, Gamma heterogeneity), and (non-parametric baseline, discrete heterogeneity). Finally, parametric specifications are studied in sub-section 5.5.

5.1. A general framework

In the nonparametric baseline case, γ_j's (γ_k^*'s) can be treated as r (r^*) free parameters in the finer (coarser, respectively) data set. It is because $h_0(t)$ is left unspecified. On the other hand, in the parametric case, γ_j's (γ_k^*'s) are functionally related due to the parametric restriction imposed on $h_0(t)$. To be able to use the same parametric baseline hazard specification across the finer and the coarser data sets, we should restrict the number of free parameters in the parametric specification not to exceed r^*, the number of non-right-censored intervals in the coarser data set.

Let θ denote the vector of all model parameters in the finer data set, which include β, parameters in unmeasured heterogeneity if any, and parameters in the baseline hazard function. Let θ^* denote the vector of all parameters in the model applicable to the coarser data set. In the non-parametric baseline case, the number of parameters are reduced due to aggregation in the coarser data set, resulting in $\dim(\theta^*) < \dim(\theta)$. But, in the parametric baseline case, we have $\theta = \theta^*$.

Under general regularity conditions (see Amemiya 1985, ch. 4), we have from (7)-(9) in Section 2.

$$
\begin{aligned}
\hat{\theta} - \theta &\overset{\mathrm{D}}{=} [\sum_{i=1}^{n}\sum_{j=1}^{r} s_{ji}\frac{1}{\alpha_{ji}(1-\alpha_{ji})}\frac{\partial \alpha_{ji}}{\partial \theta}\frac{\partial \alpha_{ji}}{\partial \theta'}]^{-1}[\sum_{i=1}^{n}\sum_{j=1}^{r} s_{ji}\frac{d_{ji}-\alpha_{ji}}{\alpha_{ji}(1-\alpha_{ji})}\frac{\partial \alpha_{ji}}{\partial \theta}] \\
&\sim N(0, [\sum_{i=1}^{n}\sum_{j=1}^{r} s_{ji}\frac{1}{\alpha_{ji}(1-\alpha_{ji})}\frac{\partial \alpha_{ji}}{\partial \theta}\frac{\partial \alpha_{ji}}{\partial \theta'}]^{-1}).
\end{aligned} \tag{24}
$$

where the variance-covariance matrix can be evaluated at the estimated parameter values for practical use.

Next, let us consider the coarser data set $Q^* = \{G_1, \cdots, G_{r^*}\}$ introduced in Section 4. Note that $d_{ki}^* = \Pi_{j\in G_k} d_{ji}$ denotes the outcome whether the ith individual survives

I_k^* or not. Accordingly, $\alpha_{ki}^* = \Pi_{j\in G_k}\alpha_{ji}$ denotes the probability of surviving the interval I_k^*. Now θ^* is of a smaller dimension than θ due to aggregation, with the difference being $r - r^*$. By a similar procedure, we obtain

$$\begin{aligned}\hat{\theta}^* - \theta &\stackrel{\text{D}}{=} [\sum_{i=1}^{n}\sum_{k=1}^{r^*} s_{ki}^* \frac{1}{\alpha_{ki}^*(1-\alpha_{ki}^*)}\frac{\partial\alpha_{ki}^*}{\partial\theta^*}\frac{\partial\alpha_{ki}^*}{\partial\theta^{*\prime}}]^{-1}[\sum_{i=1}^{n}\sum_{k=1}^{r^*} s_{ki}^* \frac{d_{ki}^*-\alpha_{ki}^*}{\alpha_{ki}^*(1-\alpha_{ki}^*)}\frac{\partial\alpha_{ki}^*}{\partial\theta^*}] \\ &\sim N(0,[\sum_{i=1}^{n}\sum_{k=1}^{r^*} s_{ki}^* \frac{1}{\alpha_{ki}^*(1-\alpha_{ki}^*)}\frac{\partial\alpha_{ki}^*}{\partial\theta^*}\frac{\partial\alpha_{ki}^*}{\partial\theta^{*\prime}}]^{-1}).\end{aligned} \tag{25}$$

Finally, let us compute the covariance between $\hat{\theta}$ and $\hat{\theta}^*$. For this purpose, let us rewrite the original estimator $\hat{\theta}$ as follows:

$$\hat{\theta} - \theta \stackrel{\text{D}}{=} [\sum_{i=1}^{n}\sum_{j=1}^{r} s_{ji}\frac{1}{\alpha_{ji}(1-\alpha_{ji})}\frac{\partial\alpha_{ji}}{\partial\theta}\frac{\partial\alpha_{ji}}{\partial\theta'}]^{-1}\sum_{i=1}^{n}\sum_{k=1}^{r^*}\sum_{j\in G_k} s_{ji}\frac{d_{ji}-\alpha_{ji}}{\alpha_{ji}(1-\alpha_{ji})}\frac{\partial\alpha_{ji}}{\partial\theta}]. \tag{26}$$

Now note that d_{ji} and $d_{ki'}^*$ are independent either $i \neq i'$ or $j \notin G_k$. This is due to the independence of survival indicators either for different individuals or across non-overlapping intervals. So, let us compute the covariance between d_{ji} and $d_{ki'}^*$ for the case $i = i'$ and $j \in G_k$. Observing that $d_{ji}d_{ki}^* = d_{ki}^*$ for all $j \in G_k$, we can easily obtain $cov(d_{ji}, d_{ki}^*) = \alpha_{ki}^*(1-\alpha_{ji})$.

Therefore,

$$\begin{aligned} cov(\hat{\theta},\hat{\theta}^*) &= [\sum_{i=1}^{n}\sum_{j=1}^{r} s_{ji}\frac{1}{\alpha_{ji}(1-\alpha_{ji})}\frac{\partial\alpha_{ji}}{\partial\theta}\frac{\partial\alpha_{ji}}{\partial\theta'}]^{-1} \\ \times[\sum_{i=1}^{n}\sum_{k=1}^{r^*}(\sum_{j\in G_k} s_{ji}\frac{1}{\alpha_{ji}}\frac{\partial\alpha_{ji}}{\partial\theta})&\frac{1}{1-\alpha_{ki}^*}\frac{\partial\alpha_{ki}^*}{\partial\theta^{*\prime}}][\sum_{i=1}^{n}\sum_{k=1}^{r^*} s_{ki}^*\frac{1}{\alpha_{ki}^*(1-\alpha_{ki}^*)}\frac{\partial\alpha_{ki}^*}{\partial\theta^*}\frac{\partial\alpha_{ki}^*}{\partial\theta^{*\prime}}]^{-1}.\end{aligned} \tag{27}$$

Using (24), (25), and (27), we can obtain the variance matrices of $\hat{\beta}$, $\hat{\beta}^*$, and their covariance matrix, by figuring out the relevant blocks corresponding to β's out of θ's. Then, these results can be used to compute the specification test statistics R and t_j's in Section 4.

5.2. Non-parametric baseline, no unmeasured heterogeneity

This sub-section analyzes the case of non-parametric baseline hazard function without unmeasured heterogeneity. Without imposing a parametric assumption on the baseline hazard function $h_0(t)$, we can still estimate the integrated baseline hazard over each of r non-right-censored intervals $I_1, \cdots, I_r$. In the current setting, we are able to consistently estimate $\gamma_1, \cdots, \gamma_r$ or, equivalently, $\exp(\gamma_1) = \int_0^1 h_0(t)\, dt, \cdots, \exp(\gamma_r) = \int_{r-1}^r h_0(t)\, dt$. This means that we can approximate the unknown function $h_0(t)$ up to a step function with r different steps, insofar as r is either finite or increasing at a rate slower than the sample size n. The model parameters can be represented as a $(k+r) \times 1$ vector, $\theta = (\beta', \gamma_1, \cdots, \gamma_r)'$.

Note that

$$\partial\alpha_{ji}/\partial\theta = -\alpha_{ji} e^{x_i'\beta+\gamma_j} z_{ji}, \tag{28}$$

where $z_{ji} = (x_i', e_j')'$ is a $(k+r) \times 1$ column vector and $e_j = (0, \cdots, 0, 1, 0, \cdots, 0)'$ is an $r \times 1$ unit column vector with one occupying the jth location. Thus,

$$\frac{\partial L}{\partial\theta}(\theta) = -\sum_{i=1}^{n}\sum_{j=1}^{r} s_{ji}\frac{d_{ji}-\alpha_{ji}}{1-\alpha_{ji}} e^{x_i'\beta+\gamma_j} z_{ji} = -\sum_{j=1}^{r} Z_j' S_j \Omega_j^{-1} u_j, \tag{29}$$

where $Z_j = (z_{j1}, \cdots, z_{jn})'$ is an $n \times (k+r)$ matrix, Ω_j is an $n \times n$ diagonal matrix of $(1-\alpha_{ji})/(\alpha_{ji}(e^{x_i'\beta+\gamma_j})^2)$, S_j is an $n \times n$ diagonal matrix of s_{ji}, and u_j is an $n \times 1$ vector of $(d_{ji}-\alpha_{ji})/(\alpha_{ji}e^{x_i'\beta+\gamma_j})$. Using $Var(u_j) = \Omega_j$, $j = 1, \cdots, r$, the information matrix can be approximated as

$$-\frac{\partial^2 L}{\partial\theta\partial\theta'} \overset{\text{P}}{=} \frac{\partial L}{\partial\theta}\frac{\partial L}{\partial\theta'} \overset{\text{P}}{=} [\sum_{j=1}^{r} Z_j' S_j \Omega_j^{-1} u_j][\sum_{j=1}^{r} Z_j' S_j \Omega_j^{-1} u_j]' \overset{\text{P}}{=} \sum_{j=1}^{r} Z_j' S_j \Omega_j^{-1} Z_j. \tag{30}$$

Using these results, we have

$$\begin{aligned}\hat{\theta} - \theta &\overset{\text{D}}{=} -[\sum_{j=1}^{r} Z_j' S_j \Omega_j^{-1} Z_j]^{-1} \sum_{j=1}^{r} Z_j' S_j \Omega_j^{-1} u_j \\ &\sim N(0, [\sum_{j=1}^{r} Z_j' S_j \Omega_j^{-1} Z_j]^{-1}).\end{aligned} \tag{31}$$

Similarly for the coarser data set $Q^* = \{G_1, \cdots, G_{r^*}\}$, we obtain

$$\hat{\theta}^* - \theta \stackrel{\mathrm{D}}{=} -[\sum_{k=1}^{r^*} Z_k^{*'} S_k^* \Omega_k^{*-1} Z_k^*]^{-1} \sum_{k=1}^{r^*} Z_k^{*'} S_k^* \Omega_k^{*-1} u_k^* \sim N(0, [\sum_{k=1}^{r^*} Z_k^{*'} S_k^* \Omega_k^{*-1} Z_k^*]^{-1}), \quad (32)$$

where $Z_k^* = (z_{k1}^*, \cdots, z_{kn}^*)'$ is an $n \times (k+r^*)$ matrix with $z_{ki}^* = (x_i', e_k^{*'})'$ being a $(k+r^*) \times 1$ vector (e_k^* is an $r^* \times 1$ unit vector with one in the kth location), Ω_k^* is an $n \times n$ diagonal matrix of $(1 - \alpha_{ki}^*)/(\alpha_{ki}^*(e^{x_i'\beta + \gamma_k^*})^2)$, S_k^* is an $n \times n$ diagonal matrix of s_{ki}^*, and u_k^* is an $n \times 1$ vector of $(d_{ki}^* - \alpha_{ki}^*)/(\alpha_{ki}^* e^{x_i'\beta + \gamma_k^*})$. Again, note that S_1^* is an identity matrix of order n, and that $Var(u_k^*) = \Omega_k^*$, $k = 1, \cdots, r^*$.

Finally, the covariance between $\hat{\theta}$ and $\hat{\theta}^*$ is

$$\begin{aligned} &cov(\hat{\theta}, \hat{\theta}^*) \\ =&[\sum_{j=1}^{r} Z_j' S_j \Omega_j^{-1} Z_j]^{-1} [\sum_{k=1}^{r^*} (\sum_{j \in G_k} f_j Z_j' S_j) \Omega_k^{*-1} Z_k^*][\sum_{k=1}^{r^*} Z_k^{*'} S_k^* \Omega_k^{*-1} Z_k^*]^{-1}, \end{aligned} \quad (33)$$

where $S_j \Omega_k^{*-1} S_k^* = S_j \Omega_k^{*-1}$ (for all $j \in G_k$) has been used, a property resulting from $s_{ji} s_{ki}^* = s_{ji}$ for all $j \in G_k$.

5.3. Non-parametric baseline, Gamma heterogeneity

This sub-section analyzes the case of non-parametric baseline with unmeasured heterogeneity modeled according to a Gamma distribution. Assume that v_i follows a Gamma distribution with parameters a and b, denoted Gamma(a, b): $v_i \sim$ i.i.d. $g(v) = b^a v^{a-1} e^{-bv}/\Gamma(a)$, where $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t}\, dt$. It is easy to show that the moment generating function is $M_v(t) = (\frac{b}{b+t})^a$. The mean and variance are $Ev = a/b$ and $Var(v) = a/b^2$. To fix the level of v, let us assume that $Ev = 1$, rendering $a = b$. That is, the unmeasured heterogeneity is modeled through a Gamma distribution with mean one and variance $1/b$.

The model parameters can be represented as a $(k+1+r) \times 1$ vector, $\theta = (\beta' : b : \gamma_1, \cdots, \gamma_r)'$. Using the moment generating function, $M_v(t) = (\frac{b}{b+t})^b$, we obtain

$$\alpha_{ji} = (\frac{b + c_{1i} + \cdots + c_{j-1i}}{b + c_{1i} + \cdots + c_{j-1i} + c_{ji}})^b, j = 1, \cdots, r, \quad (34)$$

where $c_{ji} = \exp(x_i'\beta + \gamma_j)$. Now, it is straightforward to compute $\partial\alpha_{ji}/\partial\theta = (\partial\alpha_{ji}/\partial\beta' : \partial\alpha_{ji}/\partial b : \partial\alpha_{ji}/\partial\gamma_1, \cdots, \partial\alpha_{ji}/\partial\gamma_r)'$. We have

$$\begin{aligned}
\frac{\partial\alpha_{ji}}{\partial\beta} &= -b^2[\alpha_{ji}]^{\frac{b-1}{b}} \frac{c_{ji}}{(b + c_{1i} + \cdots + c_{j-1i} + c_{ji})^2} x_i, \\
\frac{\partial\alpha_{ji}}{\partial b} &= \frac{\alpha_{ji}}{b} \log \alpha_{ji} + b[\alpha_{ji}]^{\frac{b-1}{b}} \frac{c_{ji}}{(b + c_{1i} + \cdots + c_{j-1i} + c_{ji})^2}, \\
\frac{\partial\alpha_{ji}}{\partial\gamma_k} &= b[\alpha_{ji}]^{\frac{b-1}{b}} \frac{1}{(b + c_{1i} + \cdots + c_{j-1i} + c_{ji})^2} \\
&\quad \times [c_{ji}(c_{1i}\delta_{1k} + \cdots + c_{j-1i}\delta_{j-1k}) - c_{ji}\delta_{jk}(b + c_{1i} + \cdots + c_{j-1i})],
\end{aligned} \tag{35}$$

where δ_{jk} takes value one if $j = k$, and zero otherwise.

Under the coarser grouping, the model parameters becomes a $(k + 1 + r^*) \times 1$ vector, $\theta^* = (\beta' : b : \gamma_1^*, \cdots, \gamma_{r^*}^*)'$. Similarly, we have

$$\alpha_{ki}^* = (\frac{b + c_{1i}^* + \cdots + c_{k-1i}^*}{b + c_{1i}^* + \cdots + c_{k-1i}^* + c_{ki}^*})^a, k = 1, \cdots, r^*, \tag{36}$$

where $c_{ki}^* = \exp(x_i'\beta + \gamma_k^*)$. From this, we obtain

$$\begin{aligned}
\frac{\partial\alpha_{ki}^*}{\partial\beta} &= -b^2[\alpha_{ki}^*]^{\frac{b-1}{b}} \frac{c_{ki}^*}{(b + c_{1i}^* + \cdots + c_{k-1i}^* + c_{ki}^*)^2} x_i, \\
\frac{\partial\alpha_{ki}^*}{\partial b} &= \frac{\alpha_{ki}^*}{b} \log \alpha_{ki}^* + b[\alpha_{ki}^*]^{\frac{b-1}{b}} \frac{c_{ki}^*}{(b + c_{1i}^* + \cdots + c_{k-1i}^* + c_{ki}^*)^2}, \\
\frac{\partial\alpha_{ki}^*}{\partial\gamma_j} &= b[\alpha_{ki}^*]^{\frac{b-1}{b}} \frac{1}{(b + c_{1i}^* + \cdots + c_{k-1i}^* + c_{ki})^2} \\
&\quad \times [c_{ki}^*(c_{1i}^*\delta_{1j} + \cdots + c_{k-1i}^*\delta_{k-1j}) - c_{ki}^*\delta_{kj}(b + c_{1i}^* + \cdots + c_{k-1i}^*)],
\end{aligned} \tag{37}$$

where δ_{kj} takes value one if $k = j$ and zero otherwise.

By combining the above results with the general framework introduced in sub-section 5.1, we can make inferences on θ and θ^*, and carry out the proportionality test.

5.4. Non-parametric baseline, discrete heterogeneity

This sub-section analyzes the case of non-parametric baseline hazard functionwith unmeasured heterogeneity modeled according to a discrete distribution. Assume that v_i takes M finite values $v_1, \cdots, v_M$ with probabilities $p_1, \cdots, p_M$. This distribution describes that there are M unobserved types. For level identification, let us fix $v_M = 1$. For v_m, the following parameterization is convenient:

$$v_M = 1 \text{ (level normalization)},\ v_m = e^{w_m},\ m = 1, \cdots, M-1, \tag{38}$$

where w_m's are unrestricted, $-\infty < w_m < \infty$. For p_m, the following parameterization is useful:

$$p_M = \frac{1}{1 + \sum_{m=1}^{M-1} e^{\pi_m}},\ p_m = \frac{e^{\pi_m}}{1 + \sum_{m=1}^{M-1} e^{\pi_m}},\ m = 1, \cdots, M-1, \tag{39}$$

where π_m's are unrestricted too, $-\infty < \pi_m < \infty$.

The model parameters can be represented as a $(k + 2M - 2 + r) \times 1$ vector, $\theta = (\beta' : w_1, \cdots, w_{M-1} : \pi_1, \cdots, \pi_{M-1} : \gamma_1, \cdots, \gamma_r)'$. Using the moment generating function, $M_v(t) = E_v e^{-tv} = \sum_{m=1}^{M} e^{-tv_m} p_m$, we obtain

$$\alpha_{ji} = \frac{M_v(c_{1i} + \cdots + c_{j-1i} + c_{ji})}{M_v(c_{1i} + \cdots + c_{j-1i})} = \frac{\sum_{m=1}^{M} e^{-(c_{1i}+\cdots+c_{j-1i}+c_{ji})v_m} p_m}{\sum_{m=1}^{M} e^{-(c_{1i}+\cdots+c_{j-1i})v_m} p_m},\ j = 1, \cdots, r. \tag{40}$$

From this, we can compute $\partial\alpha_{ji}/\partial\theta = (\partial\alpha_{ji}/\partial\beta' : \partial\alpha_{ji}/\partial w_1, \cdots, \partial\alpha_{ji}/\partial w_{M-1} : \partial\alpha_{ji}/\partial\pi_1, \cdots, \partial\alpha_{ji}/\partial\pi_{M-1} : \partial\alpha_{ji}/\partial\gamma_1, \cdots, \partial\alpha_{ji}/\partial\gamma_r)'$. For notational convenience, let us define $a_{ji} = c_{1i} + \cdots + c_{ji}$ and $\delta_{ji} = \sum_{m=1}^{M} e^{-a_{ji}v_m} p_m$, we have

$$\begin{aligned}
\frac{\partial\alpha_{ji}}{\partial\beta} &= \frac{1}{\delta_{ji}} \sum_{m=1}^{M} p_m v_m (\alpha_{ji} a_{j-1i} e^{-a_{j-1i}v_m} - a_{ji} e^{-a_{ji}v_m}),\\
\frac{\partial\alpha_{ji}}{\partial w_k} &= \frac{1}{\delta_{ji}} p_k v_k (\alpha_{ji} a_{j-1i} e^{-a_{j-1i}v_k} - a_{ji} e^{-a_{ji}v_k}),\\
\frac{\partial\alpha_{ji}}{\partial\pi_k} &= \frac{1}{\delta_{ji}} p_k (e^{-a_{ji}v_k} - \alpha_{ji} e^{-a_{j-1i}v_k}) + \frac{1}{\delta_{ji}} p_k \sum_{m=1}^{M} p_m (\alpha_{ji} e^{-a_{j-1i}v_m} - e^{-a_{ji}v_m})\\
\frac{\partial\alpha_{ji}}{\partial\gamma_k} &= \frac{c_{ki}}{\delta_{ji}} \sum_{m=1}^{M} p_m v_m [\alpha_{ji} e^{-a_{j-1i}v_m} 1_{(k\le j-1)} - e^{-a_{ji}v_m} 1_{(k\le j)}],
\end{aligned} \tag{41}$$

where $a_{0i} = 0$.

Under the coarser grouping, the model parameters becomes a $(k + 2M - 2 + r^*) \times 1$ vector, $\theta^* = (\beta' : w_1, \cdots, w_{M-1} : \pi_1, \cdots, \pi_{M-1} : \gamma_1^*, \cdots, \gamma_{r^*}^*)'$. Similarly, we have

$$\alpha_{ki}^* = \frac{\sum_{m=1}^{M} e^{-(c_{1i}^*+\cdots+c_{k-1i}^*+c_{ki}^*)v_m} p_m}{\sum_{m=1}^{M} e^{-(c_{1i}^*+\cdots+c_{k-1i}^*)v_m} p_m}, k = 1, \cdots, r^*, k = 1, \cdots, r^*, \tag{42}$$

where $c^*_{ki} = \exp(x_i'\beta + \gamma^*_k)$. From this, we can compute $\partial\alpha^*_{ji}/\partial\theta^* = (\partial\alpha^*_{ji}/\partial\beta' : \partial\alpha^*_{ji}/\partial w_1, \cdots, \partial\alpha^*_{ji}/\partial w_{M-1} : \partial\alpha^*_{ji}/\partial\pi_1, \cdots, \partial\alpha^*_{ji}/\partial\pi_{M-1} : \partial\alpha^*_{ji}/\partial\gamma^*_1, \cdots, \partial\alpha^*_{ji}/\partial\gamma^*_r)'$. For notational convenience, let us define $a^*_{ji} = c^*_{1i} + \cdots + c^*_{ji}$ and $\delta^*_{ji} = \sum_{m=1}^M e^{-a^*_{ji}v_m} p_m$. We have

$$
\begin{aligned}
\frac{\partial\alpha^*_{ji}}{\partial\beta} &= \frac{1}{\delta^*_{ji}} \sum_{m=1}^M p_m v_m (\alpha^*_{ji} a^*_{j-1i} e^{-a^*_{j-1i}v_m} - a^*_{ji} e^{-a^*_{ji}v_m}), \\
\frac{\partial\alpha^*_{ji}}{\partial w_k} &= \frac{1}{\delta^*_{ji}} p_k v_k (\alpha^*_{ji} a^*_{j-1i} e^{-a^*_{j-1i}v_k} - a^*_{ji} e^{-a^*_{ji}v_k}), \\
\frac{\partial\alpha^*_{ji}}{\partial\pi_k} &= \frac{1}{\delta^*_{ji}} p_k (e^{-a^*_{ji}v_k} - \alpha^*_{ji} e^{-a^*_{j-1i}v_k}) + \frac{1}{\delta^*_{ji}} p_k \sum_{m=1}^M p_m (\alpha^*_{ji} e^{-a^*_{j-1i}v_m} - e^{-a^*_{ji}v_m}) \\
\frac{\partial\alpha^*_{ji}}{\partial\gamma^*_k} &= \frac{c^*_{ki}}{\delta^*_{ji}} \sum_{m=1}^M p_m v_m [\alpha^*_{ji} e^{-a^*_{j-1i}v_m} 1_{(k\le j-1)} - e^{-a^*_{ji}v_m} 1_{(k\le j)}],
\end{aligned}
\tag{43}
$$

where $a^*_{0i} = 0$.

Again, by combining the above results with the general framework introduced in sub-section 5.1, we can make inferences on θ and θ^*, and carry out the proportionality test.

5.5. Parametric baseline, Gamma and discrete heterogeneity

Once we parameterize $h_0(t)$ using m free parameters, we have a fully parametric duration model. This article proposes one to use the following flexible specification.

$$
\gamma_j = \log \int_{j-1}^{j} h_0(t)\,dt = \delta_0 + \delta_1 j + \cdots + \delta_{m-1} j^{m-1}. \tag{44}
$$

This parameterization is polynomial in time, and it guarantees positivity of the integrated baseline hazard over each interval, I_j. The exponential distribution (constant baseline hazard) is a special case when $m = 1$. The monotonic hazard feature of the Weibull distribution is well captured by choosing $m = 2$. As m increases, the resulting parametric specification becomes more flexible.

As a practical guide, if the baseline hazard function is expected to be monotone, use $m = 2$. If U or inverted U-shaped, use $m = 3$. Without any clear idea on the shape, apply model selection criteria to the choice of m. If data are rich enough, use the non-parametric baseline specification.

If the number m of free parameters is equal to the number r of non-right-censored intervals, then both the parametric and non-parametric baseline specifications would yield the same results. This is because of the invariance property of the maximum likelihood estimation. If m is greater than r, then we are not able to identify all these m parameters from grouped duration data with just r non-right-censored intervals. We can only identify r restrictions on m parameters, because the baseline hazard 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 those r integrated baseline hazard rates. For identification of the parametric baseline hazard function (44), we readily note that the number of free parameters in $h_0(t)$ should not exceed the number of non-right-censored intervals in the data set, $m \leq r$.

In the parametric baseline cases, the maximum likelihood estimation and specification tests are straightforward: (1) obtain two sets of parametric maximum likelihood estimators, one from the original data set and the other from the new coarser data set, (2) estimate each variance as the inverse of the observed information matrix, and (3) estimate their covariance as the variance of the estimator from the original data set, a result due to Hausman (1978).

From the results in the non-parametric baseline case, we can derive the partial derivatives of α_{ji} and α^*_{ki} with respect to each model parameters. The partial derivatives with respect to β and parameters in unmeasured heterogeneity are the same as before. Only changes occur in the partial derivatives of α_{ji} and α^*_{ki} with respect to the parameters in the baseline hazard function. But, this can be handled easily using the chain rule of differentiation. For example, $\partial\alpha_{ji}/\partial\delta_k = (\partial\gamma_1/\partial\delta_k, \cdots, \partial\gamma_r/\partial\delta_k)(\partial\alpha_{ji}/\partial\gamma_1, \cdots, \partial\alpha_{ji}/\partial\gamma_r)'$. Similarly for $\partial\alpha^*_{ji}/\partial\delta_k$.

To allow maximum flexibility within the parametric baseline specification (44), one can use $m = r^*$, resulting in

$$\gamma_j = \log \int_{j-1}^{j} h_0(t)\, dt = \delta_0 + \delta_1 j + \cdots + \delta_{r^*-1} j^{r^*-1}, j = 1, \cdots, r. \tag{45}$$

Under this parameterization, the coarser estimator $\hat{\beta}^*$ is in fact equivalent to the semi-parametric estimator, whereas the original estimator $\hat{\beta}$ is an estimator obtained after imposing $r - r^*$ restrictions on the baseline hazard function.

6. APPLICATION TO LEFT-CENSORING

The suggested framework can be used to deal with left-censoring issue in duration analysis. If every individual is observed from the start of his or her episode (called, flow sampling), there is no problem of left-censoring. Often, however, individuals are observed to be already in the middle of an episode (called, stock sampling). The resulting duration variable is said to be left-censored, and is known to complicate the estimation (see, for example, Amemiya 1985, ch. 11).

Considering that a longer duration is more likely to be in progress at a random start time of observation, the fact that an episode is left-censored implies that the corresponding duration is more likely to be longer than a typical duration (described as *length biased sampling*). We would like to address this problem by assuming that there is a latent unobserved variable affecting duration, say v. The essential feature of left-censoring can be described by the changing distribution of v over the duration process. Therefore, by updating the distribution of v at each stage of the process as suggested in this article, we can account for the selectivity, thus for the left-censoring.

To be concrete, for left-censored observations, we replace $\alpha_j(x)$ in eq. (14) with the following modified interval survival probability:

$$\alpha_j(x|s) = E_{s+j}\alpha_j(x, v) = \int_0^{\infty} \alpha_j(x, v) g_{s+j}(v)\, dv, \tag{46}$$

where $g_{s+j}(v)$ is the $s+j$th stage heterogeneity density, $g_{s+j}(v) = g(v|T > s+j-1, x)$ (see eq. (11)). Note that to be able to compute $\alpha_j(x|s)$, we need information on how long the process has already been in progress, that is, elapsed duration s, by the time of first observation.

7. CONCLUDING REMARKS

Often, duration data are available in a grouped form due to a certain discrete observation mechanism, while many covariates are unobservable. This article develops a general sequential binary choice framework of grouped duration model with unobserved heterogeneity. Considering that many economic duration data are grouped and many variables are missing, the proposed methods will prove useful in many situations.

This article introduces a new flexible parameterization of the baseline hazard function. Non-parametric baseline hazard function is covered as a special case of this flexible parameterization. This article proposes a new, and operationally convenient, way of tackling unobserved heterogeneity from a sample selection perspective. Gamma and discrete distributions have been used to capture heterogeneity. By updating the heterogeneity distribution, we can account for selection of type over duration, and thus can avoid the neglected heterogeneity bias. In fact, the bias resulting from neglected heterogeneity is essentially a sample selection bias. This interpretation allows us to keep the sequential binary choice representation of the grouped duration model, and still enables us to estimate the model easily. Also, this article extends Ryu's (1994b) proportionality test to a general grouped duration setting. The test can detect the source and direction of non-proportionality without adopting a priori restrictions, not shared by existing test procedures.

The general framework suggested in this article proves useful in addressing left-censoring problem.

REFERENCES

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

Amemiya, T. (1973), "Regression Analysis when the Dependent Variable is Truncated Normal," Econometrica, 41, 997–1017.

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.

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," Biometrika, 62, 269-279.

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

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.

Hausman, J. A., and Wise (1979), Econometrica.

Heckman, J., 1979, " Sample Selection Bias as a Specification Error," Econometrica, 47, 153–161.

Heckman, J., 1976, "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models," The Annals of Economic and Social Measurement, 5, 475–492.

Heckman, J., and Singer, B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," Econometrica, 52, 271-320.

Jennrich, R., 1969, "Asymptotic Properties of Non-Linear Least Squares Estimators," The Annals of Mathematical Statistics, 40, 633–643.

Kalbfleisch, J. D., and 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.

Lee, L. F., 1982, "Some Approaches to the Correction of Selectivity Bias," Review of Economic Studies, 49, 355–372.

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.

Newey, W., and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," in R. F. Engle and D. McFadden, eds., Handbook of Econometrics, IV, Elsevier.

Moffitt, R. (1985), "Unemployment Insurance and the Distribution of Unemployment Spells," Journal of Econometrics, 28, 85–101.

Newey, W. (1984), "A Method of Moments Interpretation of Sequential Estimators," Economics Letters, 14, 201–206.

Newey, W. and K. West, 1987, "A Simple Positive Semi-Definite, Heteroscedasticity and Auto-correlation Consistent Covariance Matrix," Econometrica, 55, 703–708.

Nickell, S. (1979), "Estimating the Probability of Leaving Unemployment," Econometrica, 47, 1249-1266.

A. Pagan, 1986, "Two Stage and Related Estimators and Their Applications," Review of Economic Studies, 53, 517–538.

A. Pagan, 1984, "Econometric Issues in the Analysis of Regressions with Generated Regressors," International Economic Review, 25, 221–247.

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

Ryu, Keunkwan (1994a), "Modified Heckman's Two-step Estimation and Specification Test in Sample Selection Models," mimeo., UCLA.

Ryu, K. (1994b), "Group Duration Analysis of the Proportional Hazard Model: Minimum Chi-square Estimators and Specification Tests," Journal of the

American Statistical Association, 89, 1386–1397.

Ryu, Keunkwan (1993a), "Monotonicity of the Fisher Information and the Kullback-Leibler Divergence Measure," Economics Letters, 42, 121-128.

Ryu, K. (1993b), "An Extension of Marshall and Olkin's Bivariate Exponential Distribution," Journal of the American Statistical Association, 88, 1458–1465.

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.

White, H., 1980, "A Heteroscedasticity-Consistent Covariance Matrix Estimator and A Direct Test for Heteroscedasticity," Econometrica, 48, 817–838.