Statistical Inference of a Bivariate Proportional Hazard Model with Grouped Data [1]

MARK YUYING AN

Department of Economics, Duke University, Durham, NC 27708-0097.
Telephone: (919) 660-1809;   Fax: (919) 684-8974
Email: man@econ.duke.edu.

November 1996

Abstract

This paper studies the estimation of a semiparametric bivariate proportional hazard model from event time data under interval censoring. As a direct generalization of the bivariate exponential distribution of Marshall and Olkin, the model, on the one hand, controls for the effects of observed covariates, and on the other, achieves great flexibility through nonparametrically specified baseline hazards. The model is most relevant in analyzing the joint distribution of two event times arising from "systems of two components". Examples include the two infection times of the left and the right kidneys of patients and the two retirement times of married couples. To estimate this semiparametric model from grouped data, we propose a maximum likelihood estimator and a minimum chi-square estimator. Both estimation methods exploit the fact that the most flexible model structure that can be identified with grouped data is finite-dimensional. Compared with the maximum likelihood estimation, the minimum chi-square procedure is computationally more attractive but applies only to "many observations per cell" cases where the covariates are either categorical or amendable to sensible grouping. Specification tests for different model assumptions are also discussed.

Key Words: Proportional Hazard, Bivariate Durations, Grouped Data, Minimum Chi-Square Methods.

[1]Prepared for presentation at the Econometric Society Winter Meetings at New Orleans, January 4-7, 1997. I would like to thank the Duke University Arts and Sciences Research Council for financial support. An electronic copy of the paper is available as 'pub/man/papers/bivph.ps' on the ftp site 'econ.duke.edu'.

1 Introduction

In this paper we investigate the estimation of a bivariate proportional hazard model, referred to as the *generalized Marshall-Olkin model under proportional hazards* (GMOPH), from event time data under interval censoring or grouping. As a direct generalization of the bivariate exponential distribution of Marshall and Olkin (1967), the GMOPH model, on the one hand, controls for the effects of observed covariates, and on the other, achieves great flexibility through nonparametrically specified baseline hazards. The model is most relevant in analyzing the joint distribution of two event times arising from systems of two components.

Bivariate duration models for two-component systems pervade many fields of scientific research. For example, economists interested in modeling the joint retirement decisions of married couples have collected data on times to retirement for both the wife and the husband of each surveyed household (Christensen and Gupta, 1996). Medical researchers have obtained data on the infection times of the left and right kidneys of patients, and times to HIV infection of same-sex partners. Industrial analysts have accumulated data on failure times of machinery components. In analyzing these data, there are three common features. First, since the two duration times from each system are not independent, it is important to model the dependence structure. Second, in specification of the joint distribution of the two duration times, it is important to control for observed covariates including both component-specific characteristics as well as system-wide characteristics. Third, since many data report substantial ties between the two duration times, the joint distribution should allow for possible probability mass on the 45^{0} line.

The GMOPH model achieves the aforementioned goals. The model is constructed by introducing three hidden counting processes that generate fatal shocks. For k=1,2, a shock from the k-th process "kills" the k-th component of the system and a shock from the third process "kills" both components. The existence of system-wide shocks induces the dependence structure among the otherwise independent component-specific durations. By adopting the proportional hazard specification for the intensity rates of the three latent counting processes, we obtain the GMOPH model which specifies the joint distribution of the two component-specific duration times.

Statistical inference of the semiparametric GMOPH model without parametric specification of the baseline hazards has been an ardent challenge. The first difficulty stems from the fact that the joint distribution is not absolutely continuous due to the probability mass on the 45^0 line. The second difficulty stems from the infinite dimensionality of the parameter space due to the nonparametric specification of the baseline hazard. For the univariate proportional hazard model, the partial likelihood approach of Cox (1972) provides consistent and asymptotically normal estimates for the coefficients of the covariates. A similar trick does not generalize to the bivariate case. Previous estimation methods apply only to special cases of the general model (see section 2 below).

In contrast to exactly observed duration data, duration variables are often intervally censored. For example, in a controlled experiment, due to high monitoring cost, scientists often opt to follow the subjects in discrete time intervals and record the *current status* at each measurement. In medical studies, patients are often scheduled to visit the clinic in predetermined intervals for check up. In retrospective longitudinal surveys, due to limited memory, human subjects can only recall the time interval (a month, for instance) in which an event has happened. In all these situations, the duration time is known to fall into some time interval. This paper discusses the statistical inference of the GMOPH model with grouped data (case 2 interval censoring a lá Groeneboom and Wellner, 1992).

Relative to exactly observed duration data, grouped data are less informative. Conventional wisdom suggests that, if the semiparametric GMOPH model is intractable with exact observed durations, it should be more difficult with grouped data. However, in the current setting, the maximum likelihood estimation of the GMOPH model with grouped data is obtainable while remaining difficult with the exactly observed durations. The reason is that, with grouped duration date, the most flexible GMOPH model structure that can be identified is intrinsically finite dimensional akin to the piecewise constant baseline hazards.

The above observation does not imply that the grouped data are better than exactly observed data. On the contrary, this suggests a second-best solution to the original problem. If, as it is often the case, the primary purpose is to estimate the coefficients associated with the covariates, that is, with the base line hazard rates treated as nuisance parameters, then one can artificially group the exactly observed duration data and proceed as if the data were grouped. The advantage with the exactly observed durations is that the researcher is free to specify the number of intervals. As the number of groupings and the arrangement of intervals changes, the researcher obtains valuable diagnostic information regarding model specification as well as consistent estimates for the parameters of primary interest.

To exploit the key insight that under data grouping the identifiable model structure is finite dimetional, we propose both a maximum likelihood estimation procedure and a minimum chi-square procedure. Both procedures identify and consistently estimate the same set of parameters.

The minimum chi-square methods were first introduced by Berkson (1944) for logit models. Other applications include the probit model (Amemiya, 1985 and references therein), the univariate proportional hazard model under interval censoring (Ryu, 1994), and the willingness to pay model under across-interval censoring (An, 1996a). The name, minimum chi-square, comes from the fact that the objective function evaluated at the estimator is asymptotically distributed as χ^2. A minimum chi-square estimation procedure consists of two easy steps. In the first, some nonparametric method is used to consistently estimate the cell-specific survivor functions. In the second, to exploit the fact that some nonlinear transformation of the survivor function is linear in the parameters, feasible generalized least squares are used to obtain the parameter estimates. Compared with the maximum likelihood estimation, the minimum chi-square procedure is computationally more attractive but applies only to "many observations per cell" cases where the covariates are either categorical or amendable to sensible grouping.

The rest of the paper is structured as follows. In section 2 we introduce the GMOPH modeling strategy, briefly mention the previous estimation methods and formally define the data grouping mechanism known as case 2 interval censoring. The maximum likelihood estimator and the minimum chi-square estimation estimator are discussed in sections 3 and 4, respectively. Section 5 lists a few possible specification tests. The most important one is the test for proportional hazard assumption itself. We conclude in section 6 with some discussions on possible extensions.

2 The Model

2.1 The Counting Processes Approach

The GMOPH model is a direct generalization of the bivariate exponential (BVE) distribution due to Marshall and Olkin (1967). For a physical model that motivates the BVE distribution consider a system with two components. Suppose there are three independent Poisson processes, with intensity rates λ_1, λ_2, and λ_3, respectively, that generate random shocks. A shock released from the first process "kills" the first component. A shock from the second process "kills" the second component. A shock from the third process "kills" both. For k=1,2,3, let T_k^* be the waiting time until a shock from the k-th process. Let (T_1, T_2) be the two duration times associated with the two components. Obviously $T_1 = \min\{T_1^*, T_3^*\}$ and $T_2 = \min\{T_2^*, T_3^*\}$. It then follows that (T_1, T_2) has the joint survivor function,

$$S(t_1, t_2;) \equiv P(T_1 > t_1, T_2 > t_2) = \exp\{-\lambda_1 t_1 - \lambda_2 t_2 - \lambda_3 \max\{t_1, t_2\}\}. \qquad (1)$$

As one of the simplest and the most parsimonious models for bivariate survival times, BVE has the following noticed properties.

Remark 1

*(1) For $k = 1, 2$, the marginal distribution of T_k is exponential with parameter $\lambda_k + \lambda_3$ (*marginal lack of memory*).*

(2) $S(x_1+x_2, y_1+y_2) = S(x_1, y_1)S(x_2, y_2)$ *for all positive numbers* x_1, x_2, y_1, y_2 (joint lack of memory).

(3) $P(T_1 = T_2) = \lambda_3/(\lambda_1 + \lambda_2 + \lambda_3) > 0$ *unless* $\lambda_3 = 0$ (probability mass on the 45^0 line).

Property (3) makes the distribution not absolutely continuous. Since Marshall and Olkin (1967), several authors have tried to generalize the BVE distribution to restore absolute continuity. It turns out that it is impossible to have an absolute continuous distribution which satisfies both marginal and joint lack of memory. [2]

BVE distribution and its generalizations have rarely been used in empirical applications. To be empirically appealing, a model should be able to accommodate both component-specific and system-level covariates. Also, the lack-of-memory property should be an exception rather than the norm. We now introduce a generalization of BVE that achieves these two goals. [3]

Consider a two-component system and three independent non-homogeneous Poisson processes with intensity function $\lambda^1(t), \lambda^2(t), \lambda^2(t)$, respectively. As before, let T_k^* denote

[2] Block and Basu (1974), Sarkar (1987) and Ryu (1993) have suggested absolute continuous models by "sacrificing" marginal lack of memory, joint lack of memory, and both marginal and joint lack of memory, respectively.

[3] This generalization has been adopted by, among others, Aalen *et al* (1980), Leurgans *et al* (1982), Klein *et al* (1989), and more recently Ghosh and Gelfand (1996).

the latent duration time until a shock is released from the k-th process. The observed durations are $\mathbf{T} = (T_1, T_2)$ where $T_k = \min\{T_k^*, T_3^*\}$. Therefore,

$$S(\mathbf{t}) \equiv P(\mathbf{T} > \mathbf{t}) = P\left(T_1^* > t_1, T_2^* > t_2, T_3^* > \max\{t_1, t_2\}\right). \tag{2}$$

Since, by assumption, the latent T_k^*'s are independent,

$$S(\mathbf{t}) = \exp\left\{-\Lambda^1(t_1) - \Lambda^2(t_2) - \Lambda^3(\max\{t_1, t_2\})\right\}, \tag{3}$$

where for $k = 1, 2, 3$, $\Lambda^k(t) = \int_0^t \lambda^k(s)ds$ is the integrated hazard function associated with the k-th process.

To accommodate covariates, suppose the analyst is equipped with an independent sample of size N with the following information,

$$\left\{\mathbf{x}_n^1, \mathbf{x}_n^2, \mathbf{z}_n\right\}, \qquad n = 1, 2, ..., N, \tag{4}$$

where the row vector $\mathbf{z}_n$ measures the observed characteristics of system n, and $\mathbf{x}_n^k$ measures the observed characteristics of the k-th component of n. Denote $\mathbf{w}_n = (\mathbf{x}_n^1, \mathbf{x}_n^2, \mathbf{z}_n)$. For these covariates, assume,

A. 1 *The covariate vectors, $\{\mathbf{w}_n\}$, are uniformly bounded in n and* $\lim_{N\to\infty} N^{-1}\sum_n \mathbf{w}_n'\mathbf{w}_n = Q$, *a finite non-singular $q \times q$ matrix.*

We adopt the following convenient parametrization in the hazard function.

A. 2 *The hazard rates for the latent waiting times T_k^*'s are of the proportional hazard type, that is,*

$$\lambda^k(t|\mathbf{w}_n) = h^k(t)e^{\mathbf{w}_n\delta_k}, \qquad k = 1, 2, 3. \tag{5}$$

In A.2, the baseline hazards $h^k(.)$'s are not parametrically specified. The primary parameters of interest are the covariate coefficients $\delta = (\delta_1', \delta_2', \delta_3')'$. In practice however restrictions on δ are usually imposed. For example, it is not unreasonable to assume that $\mathbf{x}_n^1$ do not affect $\lambda^2(t)$. This restriction is accommodated by imposing the corresponding components of δ_2 to be zeros. We will discuss more on this in section 5.1. For now, we only require,

A. 3 *The parameter vector $\delta \in \Delta$ where Δ is an open and bounded subset of $\mathbf{R}^r$ for some integer $r \leq 3q$.*

Equation (3) and the parametrization in A.2 constitute the GMOPH model. Compared with Remark 1, the GMOPH model has the following properties.

Remark 2

(1) For $k = 1, 2$, the marginal distribution of T_k has hazard function $\lambda^k(t) + \lambda^3(t)$, which is clearly not of the proportional hazard type.

(2) $P(T_1 = T_2) = \int_0^\infty \lambda^3(t) \exp\{-\Lambda^1(t) - \Lambda^2(t) - \Lambda^3(t)\}dt > 0$ unless $\lambda^3(t) = 0$ for all t. Hence the distribution is not absolutely continuous.

(3) The contribution, $l(\mathbf{t})$, to the likelihood of an exactly observed pair of duration times, $\mathbf{t} = (t_1, t_2)$, would be

$$l(\mathbf{t}) = \begin{cases} \lambda^1(t_1)[\lambda^2(t_2) + \lambda^3(t_2)] \exp\{-\Lambda^1(t_1) - \Lambda^2(t_2) - \Lambda^3(t_2)\}, & t_1 < t_2 \\ \lambda^2(t_2)[\lambda^1(t_1) + \lambda^3(t_1)] \exp\{-\Lambda^1(t_1) - \Lambda^2(t_2) + \Lambda^3(t_1)\}, & t_1 > t_2 \\ \lambda^3(t_1) \exp\{-\Lambda^1(t_1) - \Lambda^2(t_1) - \Lambda^3(t_1)\}, & t_1 = t_2, \end{cases} \tag{6}$$

Before we move on to study the estimation of the GMOPH model, it is worth pointing out two extra points regarding the counting processes approach. First, in contrast to the multiple-component systems studied here, there are at least two other situations involving bivariate or multivariate durations. In a *competing-risk* setting, there are multiple causes or risks for the termination of a *single duration*. Each risk is viewed as a latent duration time and the analyst observes only the minimum of the *risk-specific* durations. In a *multiple-episode* setting, there are multiple episodes of the same transitions. Each episode induces a duration time and the analyst observes a sequence of *episode-specific* durations for each subject. It is obvious that the counting-process approach and the GMOPH model apply to neither a competing-risk setting nor a multiple-episode setting.

Second, in contrast to the counting processes approach adopted here, there are other approaches to multivariate duration models. The most common approach is to generate joint distribution and the dependence structure using common unobserved heterogeneity or frailty. [4]

[4] See, for example, Hougaard (1987), Amemiya (1991) and An (1996b) for surveys and comments.

2.2 Previous Work on Estimation

Equation (6) reveals the major difficulty in the maximum likelihood-type estimation of the GMOPH model without parametric specification of the baseline hazards $h^k(t)$, $k = 1, 2, 3$. This is in contrast to the univariate case, where under proportional hazard specification, the maximum partial likelihood procedure of Cox (1972) results in consistent and asymptotically normal estimates for the covariate effects, with which the baseline hazard can be nonparametrically estimated as a byproduct. Only for several special cases of the GMOPH model, have statistical methods been developed.

- If there are no regressors and $\lambda^k(t) = \lambda_k$ for $k = 1, 2, 3$, the GMOPH model specializes to the BVE model. For the BVE model, Arnold (1968) proposes the methods of moment estimator,
$$\hat{\lambda}_k = \frac{N^{-1} B_k}{(N-1)^{-1} U}, \qquad k = 1, 2, 3, \tag{7}$$
where $U = \sum_n \min\{t_{1n}, t_{2n}\}$, $B_1 = \sum_n 1_{t_{1n} < t_{2n}}$, $B_2 = \sum_n 1_{t_{1n} > t_{2n}}$, and $B_3 = \sum_n 1_{t_{1n} = t_{2n}}$. This estimator is consistent and simple but not as efficient compared with the maximum likelihood estimation proposed by Proschan and Sullo (1974). [5]

- Assume that the three unspecified baseline hazard rates are proportional to one another, that is, $h^k(t) = \lambda_k h^3(t)$ for $k = 1, 2$ in A.2. Under this *proportional baseline hazard* model specification, a close examination of (6) shows that one can estimate the parameter $(\delta', \lambda_1, \lambda_2)$ using the partial likelihood strategy without the specification of $h^3(t)$. This is the trick employed by Klein *et al* (1989).

- Recently, Ghosh and Gelfand (1996) specify the hazard functions using the parametric Weibull distribution, $h^k(t) = t^{\lambda_k}$ and propose a Bayesian estimation procedure using Tanner and Wong's (1987) data augmentation algorithm.

In next two sections we introduce the maximum likelihood estimator and the minimum chi-square estimator for the GMOPH model under data grouping. To prepare, we will formally define the data grouping mechanism.

[5] The consistency of the Arnold's estimator follows from two simple facts. First, the random variable $\min\{T_1, T_2\} = \min\{T_1^*, T_2^*, T_3^*\}$ is exponential with parameter $\lambda = \lambda_1 + \lambda_2 + \lambda_3$. Therefore the sample mean, $(N-1)^{-1}U$, converges almost surely to its population mean λ^{-1}. Second, for $k = 1, 2, 3$, B_k is the sum of N *iid* Bernoulli random variable with the parameter λ_k/λ. Therefore the sample mean, $N^{-1}B_k$, converges almost surely to the population mean λ_k/λ.

2.3 Data Grouping Mechanism

To ease exposition, we consider only *regular interval censoring* defined as follows. Let $0 < a_1 < a_2 < ... < a_m < \infty$ be fixed measurement times. Let $a_0 = 0$ and $a_{m+1} = \infty$. Partition the whole sample space $\mathbf{R}_+^2$ for the bivariate durations into $(m+1)^2$ rectangular regions,

$$A_{ij} = \left\{(t_1, t_2) \in \mathbf{R}_+^2 : a_i \le t_1 < a_{i+1},\ a_j \le t_2 < a_{j+1}\right\}, \qquad 0 \le i, j \le m \tag{8}$$

Instead of the exact $\mathbf{t}_n = (t_{1n}, t_{2n}) \in \mathbf{R}_+^2$, the analyst observes the region that $\mathbf{t}_n$ belongs, equivalently,

A. 4 *m is fixed and For each system, the analyst observes $(m+1)^2$ binary variables,*

$$y_{nij} = 1_{(\mathbf{t}_n \in A_{ij})}, \qquad 0 \le i, j \le m, \tag{9}$$

with $\sum_{i=0}^m \sum_{j=0}^m y_{nij} = 1$ for all n.

The regular interval censoring defined here is admittedly a simple one.[6] However, many observation schemes fit the regular interval censoring described here. In some follow-up medical surveys, patients visit the clinics according to pre-fixed time intervals and at each visit only the current status is recorded. In labor economics, retrospective surveys only provide an interval (a month, for instance) in which an event occurred.

3 Maximum Likelihood Estimation

Under the GMOPH model specification, the survivor function, $S_{ij}(n) \equiv S(a_i, a_j | \mathbf{w}_n)$ can be expressed as,

$$S_{ij}(n) = \begin{cases} \exp\left\{-\phi_{1n}\kappa_i^1 - \phi_{2n}\kappa_j^2 - \phi_{3n}\kappa_j^3\right\}, & \text{if } 0 \le i < j \le m \\ \exp\left\{-\phi_{1n}\kappa_i^1 - \phi_{2n}\kappa_j^2 - \phi_{3n}\kappa_i^3\right\}, & \text{if } 0 \le j < i \le m \\ \exp\left\{-\phi_{1n}\kappa_i^1 - \phi_{2n}\kappa_i^2 - \phi_{3n}\kappa_i^3\right\}, & \text{if } 0 \le i = j \le m, \end{cases} \tag{10}$$

where for $k = 1, 2, 3$, $\phi_{kn} = \exp\{\mathbf{w}_n \delta_k\}$, and

$$\kappa_l^k = \int_0^{a_l} h^k(s) ds \tag{11}$$

[6] For example, it does not allow right-censoring at times other than a_m. More complicated interval censoring and data groupings will be discussed in section 6.

is the integrated k-th baseline hazard evaluated at a_l.

For a random sample of size N with data $\{\mathbf{w}_n, y_{nij} : i, j = 0, ..., m, n = 1, ..., N\}$, the sample log likelihood is

$$l = \sum_{n=1}^{N} \sum_{i=0}^{m} \sum_{j=0}^{m} y_{nij} \log \left[S_{ij}(n) + S_{i+1,j+1}(n) - S_{i+1,j}(n) - S_{i,j+1}(n) \right]. \quad (12)$$

It follows from the definition of $S_{ij}(n)$ that the likelihood function (12) depends on the baseline hazards only through a set of $3m$ discrete points $\{\kappa_l^k : k = 1, 2, 3; l = 1, 2, ..., m\}$. The function (12) can be viewed as the log likelihood function in terms of the (r+3m) extended parameters, $\theta = (\delta', \kappa_1^1, \kappa_2^1, \cdots, \kappa_m^1, \kappa_1^2, \cdots, \kappa_m^3)'$. The maximum likelihood estimator $\hat{\theta}_N$ of θ is defined as the one that maximizes (12) subject to

$$0 \leq \kappa_1^k \leq \cdots \leq \kappa_m^k, \qquad k = 1, 2, 3. \quad (13)$$

In practice, to get around the inequality constraints (13), one usually adopts an alternative parametrization, $\gamma_1^k = \log(\kappa_1^k)$ and $\gamma_j^k = \log(\kappa_j^k - \kappa_{j-1}^k)$ for $j = 2, 3, ..., m$. The likelihood function can be rewritten in terms of $\beta = (\delta', \gamma_1^1, ..., \gamma_m^3)'$ as $l(\beta)$. The maximum likelihood estimator for β is

$$\hat{\beta}_N = \arg\max l(\beta). \quad (14)$$

To obtain $\hat{\beta}_N$, one adopts a hill-climbing algorithm. For statistical inference, the standard asymptotic theory for maximum likelihood estimation can be invoked. We state the following result without devising a formal proof, which is standard. Notice that conditional on the covariate vector $\mathbf{w}_n$, the $(m+1)^2$ random variables Y_{nij}'s are iid multinomial.

Proposition 1 *Under conditions A.1 to A.4, the mle $\hat{\beta}_N$ defined in (14) is root-N consistent and asymptotically normal with asymptotic variance-covariance matrix,*

$$V = \left[E \frac{\partial l(\beta)}{\partial \beta} \frac{\partial l(\beta)}{\partial \beta'} \right]^{-1},$$

which can be consistently estimated using its sample analog and with $\hat{\beta}_N$ replacing β.

It is interesting to point out that we have come across a recurrent feature associated with data grouping. Compared with exact observations, grouped data convey less information. This loss of information results in dramatic a change in the nature of the statistical inference.

Remark 3 *In fact the semiparametric structure which can be identified with grouped data is intrinsically* finite dimensional.

With grouped data, one can start by parametrizing the baseline hazard $h^k(t)$ as piecewise constant functions of the form,

$$h^k(t) = \sum_{j=1}^{m} c_j^k 1_{(a_{j-1} \le t < a_j)} \qquad k = 1, 2, 3. \tag{15}$$

It is easy to see that the $3m$ number of c's in (15) and the $3m$ number of γ's defined earlier are exactly one-to-one. In fact,

$$\exp\{\gamma_j^k\} = c_j^k (a_j - a_{j-1}), \qquad 1 \le j \le m; k = 1, 2, 3.$$

These $3m$ number of free parameters are the most flexible functional forms for the baseline hazard functions that can be identified by the grouped data. In the next section we will see that the minimum chi-square estimation procedure of this paper identifies and estimates exactly the same sets of parameters, β.

4 Minimum Chi-Square Estimation

As mentioned in the introduction, minimum chi-square methods apply only to "many observations per cell" cases where the observed covariates are either categorical or amenable to sensible grouping. In the current setting, assume,

A. 5 *The covariate vector* $\mathbf{w}_n$ *takes* D *distinct values,* $\{\mathbf{w}_{(1)}, \mathbf{w}_{(2)}, ..., \mathbf{w}_{(D)}\}$ *with* $D << N$.

Partition the set $\{1, 2, ..., N\}$ into D subsets (cells) $I_1, I_2, \dots I_D$ according to the rule that $n \in I_d$ if and only if $\mathbf{w}_n = \mathbf{w}_{(d)}$. Let N_d be the cardinality of the "cell" I_d. Define

$$n_{ij}(d) = \sum_{n \in I_d} \sum_{l=i}^{m} \sum_{\tau=j}^{m} y_{nl\tau}, \qquad 0 \le i, j \le m, \tag{16}$$

which is the number of observations in cell I_d whose bivariate durations (T_1, T_2) jointly survive (a_i, a_j). Clearly $n_{00}(d) = N_d$ for all d.

4.1 The Estimator

Under regular interval censoring defined earlier, for all (i, j, d) there is a natural non-parametric estimator for $S_{ij}(d) = P((T_1, T_2) > (a_i, a_j) | \mathbf{w}_{(d)})$,

$$\hat{S}_{ij}(d) = \frac{n_{ij}(d)}{N_d}. \tag{17}$$

For an observation in cell d, consider the conditional probabilities of the form $P(T_1 > a_i, T_2 > a_j | T_1 > a_l, T_2 > a_\tau, \mathbf{w}_{(d)}) = S_{ij}(d)/S_{l\tau}(d)$ for some $(l, \tau) < (i, j)$. Specifically for $1 \leq i \leq j \leq m$ consider,

$$\frac{S_{ij}(d)}{S_{i-1,j}(d)} = \exp\left\{-\phi_{1d}(\kappa_i^1 - \kappa_{i-1}^1)\right\}; \tag{18}$$

for $m \geq i \geq j \geq 1$ consider,

$$\frac{S_{ij}(d)}{S_{i,j-1}(d)} = \exp\left\{-\phi_{2d}(\kappa_j^2 - \kappa_{j-1}^2)\right\}; \tag{19}$$

and for $1 \leq i \leq m$ consider,

$$\frac{S_{i-1,i}(d)S_{i,i-1}(d)}{S_{ii}(d)S_{i-1,i-1}(d)} = \exp\left\{-\phi_{3d}(\kappa_i^3 - \kappa_{i-1}^3)\right\}. \tag{20}$$

In the above expressions, $\phi_{kd} = \exp\{\mathbf{w}_{(d)}\delta_k\}$, for $k = 1, 2, 3$. For each $1 \leq i \leq j \leq m$ and d, define

$$u_{ij}^1(d) = \log\left(-\log\left[S_{ij}(d)/S_{i-1,j}(d)\right]\right). \tag{21}$$

Let

$$\hat{u}_{ij}^1(d) = \log\left(-\log\left[\frac{\hat{S}_{ij}(d)}{\hat{S}_{i-1,j}(d)}\right]\right) = \log\left(\log\left[\frac{n_{i-1,j}(d)}{n_{ij}(d)}\right]\right), \tag{22}$$

be the estimate of $u_{ij}^1(d)$. Equation (18) can be rewritten as

$$\hat{u}_{ij}^1(d) = \mathbf{w}_{(d)}\delta_1 + \gamma_i^1 + v_{ij}^1(d), \qquad 1 \leq i \leq j \leq m \tag{23}$$

where $\gamma_i^1 = \log(\kappa_i^1 - \kappa_{i-1}^1)$ and $v_{ij}^1 = \hat{u}_{ij}^1 - u_{ij}^1$. Similarly (19) and (20) can be rewritten as

$$\hat{u}_{ij}^2(d) = \mathbf{w}_{(d)}\delta_2 + \gamma_j^2 + v_{ij}^2(d), \qquad 1 \leq j \leq i \leq m, \tag{24}$$

$$\hat{u}_{ii}^3(d) = \mathbf{w}_{(d)}\delta_3 + \gamma_i^3 + v_{ii}^3(d), \qquad 1 \leq i \leq m. \tag{25}$$

Putting together all the equations of the forms (23), (24) and (25), we get a system of $G = m * (m + 2)$ linear regression equations for each cell d. Obvious cross-equation restrictions must be imposed. When $m = 3$, $G = 15$, these restrictions are shown in table 1.

(TABLE 1 HERE)

Putting in matrix form, we can express the system of equations as

$$\tilde{\mathbf{U}} = \mathbf{X}\beta + \tilde{\mathbf{V}}, \tag{26}$$

where $\tilde{\mathbf{U}} = (\mathbf{u}'_1, \mathbf{u}'_2 ..., \mathbf{u}'_D)'$, with $\mathbf{u}_d = (\hat{u}^1_{11}(1), \hat{u}^1_{12}(1), ..., \hat{u}^3_{mm}(1))'$ is the $G \times 1$ vector associated with cell d. The other terms in (26) are defined accordingly.

Since the covariates are weakly exogenous, there is no correlation across cells. Let $\mathbf{v}_d = (v^1_{11}(1), v^1_{12}(1), ..., v^3_{mm}(1))'$ be the $G \times 1$ vector of error terms associated with cell d. Then $Cov[\mathbf{v}_d, \mathbf{v}_{d'}] = 0$ for $d' \neq d$. This implies that the variance-covariance matrix of the entire system is of the following block diagonal form

$$Var[\tilde{\mathbf{V}}] = \Omega = diag(\Sigma_1, ..., \Sigma_D), \tag{27}$$

where Σ_d is the $G \times G$ matrix $V[\mathbf{v}_d]$.

The minimum chi-square estimator of β is defined as the feasible generalized least squares estimator from (26),

$$\hat{\beta}_c = \left(\mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X}\right)^{-1} \mathbf{X}'\hat{\boldsymbol{\Omega}}^{-1}\tilde{\mathbf{Y}}, \tag{28}$$

where $\hat{\Omega}^{-1} = diag(\hat{\Sigma}^{-1}_1, ..., \hat{\Sigma}^{-1}_D)$ is a consistent estimate of Ω^{-1}. As for any other feasible GLS estimations, the validity of $\hat{\beta}_c$ depends crucially on the ability to estimate Ω. In the current context, we need to be able to consistently estimate Σ's. This, in turn, depends on the characterization of Σ's. We will discuss this issue in conjunction with the study of the large-sample properties of the minimum chi-square estimator.

4.2 Asymptotics

We will now show that the minimum chi-square estimator has the same first-order asymptotic distributions as the maximum likelihood estimator, provided that the group size N_d goes to infinity as the sample size increases. We need the following assumption:

A. 6 *The number D of cells is fixed and for every $d = 1, 2, ..., D$,* $\lim_{N\to\infty} N_d/N = c_d > 0$.

A.6 is here only to simplify the analysis. If $c_d = 0$ for some d, then we can act as if cell d does not exist.

We first state a lemma about the consistency of the left-hand-side variables in (26).

Lemma 1 *For every (i, j, d), the estimators $\hat{u}^k_{ij}(d)$ in (23), (24) and (25) are root-N-consistent for $u^k_{ij}(d)$, that is, plim $_{N\to\infty} v^k_{ij}(d) = 0$.*

Proof. Under A.6, $N_d \to \infty$ as $N \to \infty$. For every (i, j, d), $\hat{S}_{ij}(d)$ defined in (17) is the sample mean of N_d number of *iid* Bernoulli trials with success probability $S_{ij}(d)$, and hence is root-N-consistent for $S_{ij}(d)$. Each $u^1_{ij}(d)$ defined as in (21) is a continuous function of two $S_{l\tau}(d)$'s. This proves the consistency of $\hat{u}^1_{ij}(d)$ defined in (22). The consistency of $\hat{u}^2_{ij}(d)$ and $\hat{u}^3_{ii}(d)$ follows from similar lines of argument. Notice that every $u^3_{ii}(d)$ is a continuous function of *four* $S_{l\tau}(d)$'s. □

The following result characterizes the error structure

Lemma 2 *Let $\sigma_d(i, j, k; i', j'k') = Cov\left[v^k_{ij}(d), v^{k'}_{i'j'}(d)\right]$ be a typical element of Σ_d, then*

(1) There exist functions, V^1_{ij}, V^2_{ij} and V^3_{ii}, of $S_{ij}(d)$'s such that

$$\sigma_d(i, j, k; i', j'k') \approx \begin{cases} 0, & \textit{if } (i, j, k) \neq (i', j', k') \\ V^1_{ij}(d), & \textit{if } (i, j) = (i', j') \textit{ and } k = k' = 1 \\ V^2_{ij}(d), & \textit{if } (i, j) = (i', j') \textit{ and } k = k' = 2 \\ V^3_{ii}(d), & \textit{if } i = j = i' = j' \textit{ and } k = k' = 3 \end{cases} \tag{29}$$

(2) Replacing $S_{ij}(d)$ with $\hat{S}_{ij}(d)$ in the right-hand-side of (29) provides consistent estimates for the $\sigma_d(i, j, k; i', j'k')$'s.

Proof. For the expressions of the V^1_{ij}, V^2_{ij} and V^3_{ii} and a proof of the lemma see the Appendix. □.

With this, we have

Proposition 2 *The minimum chi-square estimator defined earlier is consistent, asymptotically normally distributed and is equivalent to the maximum likelihood estimator. The variance-covariance matrix can be estimated using,*

$$\hat{V} = \left(\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X}\right)^{-1}. \tag{30}$$

Proof. The proof for the consistency and asymptotic normality of the minimum chi-square estimator as a application of the feasible GLS estimator is now standard. See for example, Amemiya (1985) for a general treatment and Ryu (1994) for the application to univariate proportional hazard models. The only condition one needs is the ability to consistently estimate the variance matrix of the error terms. In the current setting it is achieved as shown in Lemma 2. □

5 Specification Tests

Different parts of the semiparametric GMOPH model can be tested. In this section we mention a few of them.

5.1 Testing Coefficient Variability

So far in the discussion, we have basically ignored the fact that the covariate vector $\mathbf{w}_n$ consists of component-specific characteristics $\mathbf{x}_n^k$ as well as system-wide characteristics $\mathbf{z}_n$. In practice, the model is rarely taken in its saturated form so that the total number r of covariate coefficients δ is equal to $3q$. Restrictions are usually imposed. For example, it is quite natural to assume that $\lambda^1(t)$ is independent of $\mathbf{x}_n^2$, $\lambda^2(t)$ is independent of $\mathbf{x}_n^1$, $\lambda^3(t)$ is independent of $\mathbf{x}_n^k$. Also, when $\mathbf{x}_n^1$ and $\mathbf{x}_n^2$ contain the same variables, it is usually of interest to find out whether the corresponding coefficients in $\lambda^1(t)$ and $\lambda^2(t)$ are equal. All these can be stated as linear restrictions on δ, and can be tested using likelihood ratio tests.

5.2 Testing Independence

The key feature of the GMOPH model is that the dependence between the two observed duration variables can be attributed to the existence of the latent system-wide shocks. In fact T_1 and T_2 are independent if and only if $\lambda^3(t) = 0$ for all $t \geq 0$. Testing this hypothesis in the current setting amounts to testing that $(\gamma_1^3, \gamma_2^3, ..., \gamma_m^3)$ are jointly 0.

5.3 Testing Restrictions on Baseline Hazards

Under data grouping defined in this paper, the maximum identifiable model structure is $3m$ discrete points on the integrated baseline hazards. Any further restriction the analyst puts on the baseline hazards can be tested.

For example, if one is willing to adopt the parametric Weibull hazard specification, that is, $h^k(t) = \alpha_k t^{\alpha_k - 1}$, then $\gamma_j^k = \log[(a_j)^{\alpha_k} - (a_{j-1})^{\alpha_k}]$. The Weibull baseline hazard model can be easily estimated and tested using a likelihood ratio test. Testing other parametric forms of the baseline hazard can be done in a similar fashion.

Alternatively, one might want to test the "proportional baseline hazard" assumption used in Klein *et al* (1989). Under these assumption, for $k = 1, 2$, $\gamma_j^k = \log(\lambda_k) + \gamma_j^3$. These restrictions can also be tested.

In labor economics, researchers are interested in whether the baseline hazard rates are monotonic increasing (positive duration dependence) or monotonic decreasing (negative duration dependence). If, for the data grouping, the intervals are of equal length, then the monotonic increasing baseline hazards can be formulated as $\gamma_1^k \leq \gamma_2^k \leq ... \leq \gamma_m^k$.

5.4 Testing Proportionality

Even though the specification A.2 achieves maximum flexibility within the proportional hazard framework, the proportionality is itself an assumption which should be tested. For this a test procedure proposed by Ryu (1994) for univariate proportional hazard models is most appropriate.

Ryu proposes to estimate the model with artificial further aggregation of the already grouped data. Let the original estimator for the covariate coefficients δ be $\hat{\delta}$ and the new estimator be $\tilde{\delta}$. Ryu's test exploits the fact that, under the null hypothesis that the proportionality is correct, both estimators will be consistent. That is, if proportionality holds, both $\tilde{\delta}$ and $\hat{\delta}$ will converge to the same δ. Otherwise they will converge to different quantities. The test statistic is then based on the distance of the two estimates,

$$R = (\tilde{\delta} - \hat{\delta})'[V(\tilde{\delta} - \hat{\delta})]^{-1}(\tilde{\delta} - \hat{\delta}). \tag{31}$$

This is a little different from a Hausman (1978)'s test since the two estimators do not necessarily satisfy an informational nesting relation which was the key for a Hausman test. When the original "finer" data are used in the estimation, we are estimating possibly more points of the baseline hazard function. This will not necessarily make $\hat{\delta}$ more efficient than $\tilde{\delta}$. So the estimation of the variance of the difference between the two estimators is a little involved. Ryu (1994) proposes a way to facilitate this calculation.

6 Conclusions

In this paper we have investigated issues surrounding a bivariate duration model refered to as the generalized Marshall-Olkin model. The main points of the paper are follows.

1. The GMOPH model is introduced to represent the component-specific duration times of multiple-component systems. The dependence among the observed durations are due to the fact that all the components of a system are subject to a common system-wide

shock process. The main feature of the model is that it allows ties among the observed durations due to the existence of probability mass on the 45^0 line. As a consequence of this probability mass, the joint distribution is not absolutely continuous. This counting-process approach and the GMOPH model apply to neither a competing setting nor a multiple-episode setting.

2 The full power of the GMOPH model lies in its semiparametric version in which the baseline hazard rates are not parametrically specified. However, estimation of the full-blown semiparametric GMOPH model even with exactly observed duration data remains a very difficult, if not impossible, task.

3. With grouped data, the maximum identifiable model structure is intrinsically finite dimensional. We show that both the maximum likelihood estimation and minimum chi-square estimation can be used to consistently estimate the primary parameter vector together with a set of parameters which push the semiparametric model to its maximum flexibility supported by the data. Statistical inference follows from the standard asymptotic theory. The main disadvantage of the maximum likelihood estimator is its great computing burden, whereas the minimum chi-square estimator applies only to the "many observations per cell" cases.

4. If, as it is often the case, the main purpose is to estimate the coefficients δ associated with the covariates, that is, with the baseline hazard rates treated as nuisance parameters, the methods proposed in this paper suggest a "second-best" solution to the original problem. With exactly observed duration data, one can artificially group the data and proceed as if the data were grouped. The advantage with the exactly observed durations is that the researcher is free to specify the number of intervals. As the number of groupings and the arrangement of intervals changes, the researcher obtains valuable diagnostic information regarding model specification as well as consistent estimates for the δ parameters of primary interest.

This paper has focused on the bivariate GMOPH model which applies to two-component systems. Extension to multivariate cases are conceptually straightforward, though the notational complexity will be formidable. In general, for K-component systems, there will be $K+1$ latent counting processes. There will be $K+1$ identifiable baseline hazards with Km γ parameters.

For simplicity we have assumed regular interval censoring throughout the paper. More complicated data grouping mechanisms are found in use in field surveys. An immediate extension seems to allow right-censoring to occur at every observation time a_j, instead of just at a_m. This more liberal right censoring is a special case of the *across-interval censoring* as discussed in An and Ayala (1996). Across-interval censoring occurs due to missed clinic visits or invalid or lost survey records so that the duration times are known to fall into more than one interval. For these more complicated data groupings, the maximum likelihood estimator proposed in this paper can still be used without conceptual difficulty. The minimum chi-square estimation depends on the availability of nonparametrical estimates of the cell-specific survivor functions. An and Ayala (1996), Yu and Wong (1996) have recently developed the simple self-consistent algorithm to consistently estimate survivor functions from arbitrarily grouped bivariate data.

Among the many possible research projects this paper did not touch, two are of great importance. The first is to study the small-sample properties of the two estimators proposed in this paper and the second is to pursue the "second-best" solution to the original problem mentioned earlier.

Appendix: Proof of Lemma 2

Consider a term, $v_{ij}^1(d)$, say, in (23). Define $\mu_{ij}^1(d) \equiv S_{ij}(d)/S_{i-1,j}(d)$ and $\hat{\mu}_{ij}^1 = \hat{S}_{ij}(d)/\hat{S}_{i-1,j}(d)$. Taking Taylor series expansion of $\log[-\log(\hat{\mu}_{ij}^l(d))]$ around $\mu_{ij}^1(d)$ and using Lemma 1, we have

$$
\begin{aligned}
v_{ij}^1(d) &= \log[-\log(\hat{\mu}_{ij}^1(d)] - \log[-\log(\mu_{ij}^1(d))] \\
&= \frac{\hat{\mu}_{ij}^1(d) - \mu_{ij}^1(d)}{\mu_{ij}^1(d)\log\left(\mu_{ij}^1(d)\right)} + O\left(\frac{1}{N_d}\right).
\end{aligned}
$$

For large N (therefore large N_d), the last term $O(.)$ can be ignored. Hence the variance of $v_{ij}^1(d)$, conditional on $n_{i-1,j}(d)$, can be approximated by the variance of $\left[\hat{\mu}_{ij}^1(d) - \mu_{ij}^1(d)\right]\left[\mu_{ij}^1(d)\log\left(\mu_{ij}^1(d)\right)\right]^{-1}$. Conditional on $n_{i-1,j}(d)$, the random variable $\hat{\mu}_{ij}^1(d) = n_{ij}(d)/n_{i-1,j}(d)$ is the sample mean of $n_{i-1,j}$ number of *iid* Bernoulli trials with success probability μ_{ij}^1. Therefore,

$$
Var\left[\hat{\mu}_{ij}^1(d)\right] = \mu_{ij}^1(d)[1-\mu_{ij}^1(d)](n_{i-1,j}(d))^{-1}.
$$

Putting these together, we get

$$\sigma_d(i,j,1;i,j,1) \approx \left[1 - \frac{S_{ij}(d)}{S_{i,j-1}(d)}\right] \left[\log\left(\frac{S_{ij}(d)}{S_{i,j-1}(d)}\right)\right]^{-2} [n_{i-1,j}(d)]^{-1} .$$

Similar line of argument shows

$$\sigma_d(i,j,1;i,j,2) \approx \left[1 - \frac{S_{ij}(d)}{S_{i,j-1}(d)}\right] \left[\log\left(\frac{S_{ij}(d)}{S_{i,j-1}(d)}\right)\right]^{-2} [n_{i,j-1}(d)]^{-1} ,$$

and

$$\sigma_d(i,i,3;i,i,3) = Var[\hat{\mu}_i^3(d)] \left[\mu_i^3 \log(\mu_i^3)\right]^{-2} ,$$

where $\mu_i^3(d) = S_{i,i-1}S_{i-1,i}/S_{i-1,i-1}/S_{ii}$ and $\hat{\mu}_i^3(d)$ is its estimator. It is easy to show that for $(ijk) \neq (i'j'k')$ the two error terms v_{ij}^{k} and $v_{i'j'}^{k'}$, provided both are defined in (23), (24) or (25), are asymptotically uncorrelated (Ryu, 1994). This completes the proof. $\square$.

References

Aalen, O., O. Borgan, N. Keiding, and J. Thormann (1980): "Interaction Between Life History Events, Nonparametric Analysis for Prospective and Retrospective Data in the Presence of Censoring," *Scandinavian Journal of Statistics*, **7**, 161-171.

Amemiya, Takeshi (1985): *Advanced Econometrics*, Harvard University Press.

Amemiya, Takeshi (1991): "Structural Duration Models," CEPR Publication # 243, Stanford University.

An, Mark Y. (1996a): "Semiparametric Estimation of Willingness to Pay Distribution," Department of Economics Working Paper 96-20 , Duke University

An, Mark Y. (1996b): "Econometric Models for Multivariate Durations," Manuscript, Duke University.

An, Mark Y. and Roberto Ayala (1996): "Nonparametric Estimation of a Survivor Function with Across-Interval Censored Data," Department of Economics Working Paper 96-02, Duke University.

Arnold, B.C (1968): "Parameter Estimation for a Multi-variate Exponential Distribution," *Journal of the American Statistical Association*, bf 63, 848-852.

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

Block, H. W. and A. P. Basu (1974): "A continuous Bivariate Exponential Extension," *Journal of the American Statistical Association*, **69,** 1031-1037.

Christensen, B.J. and N. D. Gupta (1996): "A Dynamic Programing Model of the Retirement Behavior of Married Couples," Manuscript, University of Aarhus, Denmark.

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

Ghosh, S. and A. E. Gelfand (1996): "Latent Waiting Time Models for Bivariate Event Times with Censoring," Preprint, University of Connecticut.

Groeneboom, P. And J.A. Wellner (1992): *Information Bounds and Nonparametric Maximum Likelihood Estimation*, Birkhauser Verlag.

Hausman, J.(1978): "Specification Tests in Econometrics," *Econometrica*, November 1978, 1251-1271

Hougaard, P. (1987): "Modelling Multivariate Survival," *Scandinavian Journal of Statistics*, **14**, 291-304.

Klein, J.P., N. Keiding and C. Kamby (1989): "Semiparametric Marshall-Olkin Models Applied to the occurrence of Metastases at Multiple Sites after Breast Cancer," *Biometrics*, **45**, 1073-1086.

Leurgans, S., W.Y. Tsai, and J. Crowley (1982): "Freund's Bivariate Exponential Distribution and Censoring," in *Survival Analysis*, J. Crowley and R.A. Johnson (eds.) 230-242.

Marshall, A.W. and I. Olkin (1967): "A Multivariate Exponential Distribution," *Journal of the American Statistical Association*, **62**, 30-44.

Proschan, F and P. Sullo (1974): "Estimating the Parameters of a Bivariate Exponential Distribution in Several Sampling Situations," in *Reliability and Biometry*, F. Proschan and R.J. Serfling (eds.), 423-440.

Ryu, Keunkwan (1993): "An Extension of Marshall and Olkin's Bivariate Exponential Distribution" *Journal of the American Statistical Association*, **88**: 1458-1465.

Ryu, Keunkwan (1994) "Group Duration Analysis of the Proportional Hazard Model: Minimum chi-Squared Estimators and Specification Tests," *Journal of the American Statistical Association*, **89**: 1386-1397.

Sarkar, S.K. (1987): "A Continuous Bivariate Exponential Distribution," *Journal of the American Statistical Association*, **82**, 667-675.

Tanner, M. and W. Wong (1987): "The Calculation of Posterior Distributions by Data Augmentation (with discussion)," *Journal of American Statistical Association*, **82**, 528-550.

Yu, Q. And G.Y.C. Wong (1996):Multivariate Survival Analysis with Interval-Censored Data," Manuscript, Department of Mathematics, SUNY at Binghamton.

Table 1: Cross-equation Restrictions for the Minimum Chi-Square Methods

(The number of intervals: m=3)

Dependent Variables	δ_1	δ_2	δ_3	γ_1^1	γ_2^1	γ_3^1	γ_1^2	γ_2^2	γ_3^2	γ_1^3	γ_2^3	γ_3^3	Error
$\hat{u}_{11}^1(d)$	$\mathbf{w}_{(d)}$	0	0	1	0	0	0	0	0	0	0	0	$v_{11}^1(d)$
$\hat{u}_{12}^1(d)$	$\mathbf{w}_{(d)}$	0	0	1	0	0	0	0	0	0	0	0	$v_{12}^1(d)$
$\hat{u}_{13}^1(d)$	$\mathbf{w}_{(d)}$	0	0	1	0	0	0	0	0	0	0	0	$v_{13}^1(d)$
$\hat{u}_{22}^1(d)$	$\mathbf{w}_{(d)}$	0	0	0	1	0	0	0	0	0	0	0	$v_{22}^1(d)$
$\hat{u}_{23}^1(d)$	$\mathbf{w}_{(d)}$	0	0	0	1	0	0	0	0	0	0	0	$v_{23}^1(d)$
$\hat{u}_{33}^1(d)$	$\mathbf{w}_{(d)}$	0	0	0	0	1	0	0	0	0	0	0	$v_{33}^1(d)$
$\hat{u}_{11}^2(d)$	0	$\mathbf{w}_{(d)}$	0	0	0	0	1	0	0	0	0	0	$v_{11}^2(d)$
$\hat{u}_{21}^2(d)$	0	$\mathbf{w}_{(d)}$	0	0	0	0	1	0	0	0	0	0	$v_{21}^2(d)$
$\hat{u}_{31}^2(d)$	0	$\mathbf{w}_{(d)}$	0	0	0	0	1	0	0	0	0	0	$v_{31}^2(d)$
$\hat{u}_{22}^2(d)$	0	$\mathbf{w}_{(d)}$	0	0	0	0	0	1	0	0	0	0	$v_{22}^2(d)$
$\hat{u}_{32}^2(d)$	0	$\mathbf{w}_{(d)}$	0	0	0	0	0	1	0	0	0	0	$v_{32}^2(d)$
$\hat{u}_{33}^2(d)$	0	$\mathbf{w}_{(d)}$	0	0	0	0	0	0	1	0	0	0	$v_{33}^2(d)$
$\hat{u}_{11}^3(d)$	0	0	$\mathbf{w}_{(d)}$	0	0	0	0	0	0	1	0	0	$v_{11}^3(d)$
$\hat{u}_{22}^3(d)$	0	0	$\mathbf{w}_{(d)}$	0	0	0	0	0	0	0	1	0	$v_{22}^3(d)$
$\hat{u}_{33}^3(d)$	0	0	$\mathbf{w}_{(d)}$	0	0	0	0	0	0	0	0	1	$v_{33}^3(d)$