“SEMIPARAMETRIC” APPROACHES TO INFERENCE IN JOINT MODELS FOR LONGITUDINAL AND TIME-TO-EVENT DATA

Marie Davidian and Anastasios A. Tsiatis

http://www.stat.ncsu.edu/~davidian/
http://www.stat.ncsu.edu/~tsiatis/

In part joint work with Xiao Song, Department of Biostatistics, University of Washington
OUTLINE

1. Introduction and example
2. Joint models
3. Conditional score approach
4. Semiparametric likelihood approach
5. Simulation evidence
6. Example, revisited
7. Discussion

Semiparametric Inference in Joint Models
1. INTRODUCTION AND EXAMPLE

Longitudinal studies in medical research:

- *Time-to-event*, e.g. death, progression to AIDS, etc. (possibly censored)
- *Intermittent longitudinal* measurements (time-dependent)
- *Time-independent covariates*, e.g. treatment group, age at baseline, gender, etc

Questions of interest: *Interrelationships* between these variables
AIDS Clinical Trials Group (ACTG) 175: Compare 4 antiretroviral regimens in asymptomatic HIV-infected patients

- Primary objective: Compare on basis of time to AIDS or death
- CD4 counts approximately every 12 weeks
- Subsequent objective 1: Characterize within-subject patterns of CD4 change
- Subsequent objective 2: Characterize relationship between features of CD4 profiles and survival
CD4 profiles: 10 randomly chosen subjects
Ideally: To address these objectives, would like to have for all subjects $i$

- $T_i = \text{time to event}$
- $X_i(u) = \text{longitudinal trajectory for all times } u \geq 0$
- $Z_i = \text{vector of time-independent covariates}$
Objective 1: Characterize within-subject patterns of CD4 change

- Estimate aspects of average longitudinal trajectories; e.g. $E\{X_i(u)\}$, $\text{var}\{X_i(u)\}$

- Trajectory of the average person

Difficulties/complications:

- Data are collected *intermittently*

- Possibly measured with error

- Informative censoring: no measurement available after death! [E.g., Wu and Carroll (1988), Hogan and Laird (1997)]

- Observe only $W_i(t_{ij})$, $j = 1, \ldots, m_i$, $t_{ij} \leq T_i$

$$W_i(t_{ij}) = X_i(t_{ij}) + e(t_{ij})$$
Objective 2: Characterize relationship between features of CD4 profiles and survival

- Establish relationships between $T_i$ and $X_i(u), u \geq 0$, and $Z_i$
- History $X_i^H(t) = \{X_i(u), u \leq t\}$
- Proportional hazards model:
  \[
  \lambda_i(u) = \lim_{du \to 0} du^{-1} \Pr\{u \leq T_i < u + du | T_i \geq u, X_i^H(u), Z_i\}
  = \lambda_0(u) \exp\{\gamma X_i(u) + \eta^T Z_i\}
  \]
- Interest in estimating $\gamma, \eta, (q \times 1)$.
- Semiparametric model (baseline hazard $\lambda_0(u)$ is unspecified)

Focus here: Objective 2
Popular focus: “Surrogate marker” problem

- Can we replace the time-to-event endpoint by CD4 in assessing treatment efficacy?

- Roughly speaking (Prentice conditions)
  - Treatment is effective on time-to-event
  - Treatment has effect on marker; e.g. CD4 ↑ more on treatment than placebo
  - Effect of treatment should manifest through the marker; i.e., risk of event given specific marker trajectory should be \parallel of treatment

- For example, \( \lambda_0(u) \exp\{\gamma X_i(u) + \eta^T Z_i\} \),
  If \( \gamma < 0 \) and \( \eta = 0 \) \( \Rightarrow \) \( X_i(u) \) is a surrogate marker
Complication 1: Time-to-event may be *censored*

- Underlying *censoring variable* $C_i$
- $(T_i, C_i) = (\text{survival}, \text{censoring})$ time for $i$
- Observe $V_i = \min(T_i, C_i)$, $\Delta_i = I(T_i \leq C_i)$
- Cox (1972, 1975) *partial likelihood*: maximize

$$\prod_{i=1}^{n} \left[ \frac{\exp\{\gamma X_i(V_i) + \eta^T Z_i\}}{\sum_{j=1}^{n} \exp\{\gamma X_j(V_i) + \eta^T Z_j\} I(V_j \geq V_i)} \right]^{\Delta_i}$$
Complication 2:

- Presumes $X_i(u), i = 1, \ldots, n$ is available at all failure times
- $X_i(u)$ only available intermittently
- Subject to error and intra-subject variation

Naive approach: Extrapolate $X_i(u)$ at failure times $u$ using available longitudinal data, substitute in partial likelihood

- E.g. use most recent observed, error-prone covariate value (in place of unobserved, “true” covariate)
- “Last Value Carried Forward” (LVCF)
- Such substitution $\Rightarrow$ biased estimation (Prentice, 1982)
2. JOINT MODELS

Popular approach: *Joint longitudinal data-survival model*

- *True* longitudinal data process: *mixed effects model* (+ stochastic process)
- *Normal measurement error*
- *Proportional hazards model* depending on true longitudinal data process
Longitudinal data process: Popular models

\[ X_i(u) = \alpha_0i + \alpha_1i u, \quad \alpha_i = (\alpha_{0i}, \alpha_{1i})^T \]

\[ X_i(u) = \alpha_0i + \alpha_1i u + \alpha_2i u^2 + \cdots + \alpha_pi u^p \]

\[ X_i(u) = \alpha_0i + \alpha_1i u + U_i(u) \]

- \( U_i(u) \) stationary *Gaussian process* (Henderson, Diggle, Zeger, et al.)

- \( U_i(u) \) *integrated Ornstein-Uhlenbeck (IOU) process* (Taylor, DeGruttola, et al.)

- Introduce *within-subject autocorrelation*, allow *time-varying trend*
Ingredients of joint model: Two *linked* model components

**Longitudinal data:** At *times* \( t_i = \{t_{ij}, j = 1, \ldots, m_i\} \)

\[
W_i(t_{ij}) = X_i(t_{ij}) + e(t_{ij})
\]

- \( e(t_{ij}) \sim \mathcal{N}(0, \sigma^2) \) **of** \( \alpha_i, U_i(u) \)
- \( \alpha_i \) usually assumed *normal* (more later...)

**Time-to-event data:** *Proportional hazards* model

\[
\lambda_i(u) = \lim_{du \to 0} du^{-1} \Pr\{u \leq T_i < u + du|T_i \geq u, X_i^H(u), Z_i\}
= \lambda_0(u) \exp\{\gamma X_i(u) + \eta^T Z_i\}
\]

- *Dependence on* \( \alpha_i \) through \( X_i(u) \)
Philosophical interlude:

- From *biological* point of view, form of $X_i(u)$ may be dictated by beliefs in underlying *biological mechanisms*
  - E.g., $X_i(u) = \alpha_{0i} + \alpha_{1i}u \Rightarrow$ *smooth trend* is predominant feature associated with prognosis
  - E.g., $X_i(u) = \alpha_{0i} + \alpha_{1i}u + U_i(u) \Rightarrow$ “*local*” “good” or “bad” periods predominant feature associated with prognosis

- From *empirical* point of view, represent hazard in terms of relevant features of longitudinal process
  - *Higher-order polynomials, splines* versus *stochastic process*
  - Ease of *implementation*

- From *practical* point of view – *intermittency* motivates assumption that within-subject autocorrelation *negligible*
  - $U_i(u)$ “*absorbed*” into $e_i(u)$
Observed data: \((V_i, \Delta_i, W_i, Z_i, t_i), \ i = 1, \ldots, m\)

\[ V_i = \min(T_i, C_i), \quad \Delta_i = I(T_i \leq C_i), \quad W_i = \{W(t_{i1}), \ldots, W(t_{im_i})\}^T \]

Implementation: Assuming normal \(\alpha_i\)

- Two-stage ("Regression calibration"): (i) EBLUPs from normal mixed model fit at each survival time, (ii) insert in Cox partial likelihood ⇒ reduces but doesn’t eliminate bias (Pawitan & Self, 1993; Tsiatis et al., 1995; Dafni & Tsiatis, 1998; Lavalley and DeGruttola, 1996; Bycott & Taylor, 1998)


Issues:

- Sensitivity to assumptions
- Computational complexity
**ACTG 175:** Assuming $X_i(u) = \log \text{CD4}$, $X_i(u) = \alpha_0 + \alpha_1 u$

*Raw individual OLS estimates*
3. CONDITIONAL SCORE APPROACH

Our objective: \( X_i(u) = f(u)^T \alpha_i \); e.g., \( X_i(u) = \alpha_0 + \alpha_1 u \)

- Simple method (computationally and conceptually straightforward) that yields consistent, asymptotically normal estimator for \( \gamma, \eta \)

- ... And that furthermore makes no distributional assumption about the underlying random effects \( \alpha_i \)

- \( \Rightarrow \) Semiparametric model/method (\( \lambda_0(u) \) and distribution of \( \alpha_i \) unspecified)

Our approach: Exploit the conditional score idea of Stefanski & Carroll (1987) for GLIMs

- "Condition away" the random effects \( \alpha_i \)
Assume: $\sigma^2$ known for now

Notation:

- $\hat{X}_i(u)$ OLS estimator for $X_i(u)$ based on $t_i(u) = (t_{ij} < u)$ (requires $\geq 2$ measurements prior to $u$)

- $dN_i(u) = I(u \leq V_i < u + du, \Delta_i = 1, t_{i2} \leq u)$ puts point mass at $u$ for observed death time if after 2nd measurement

- $Y_i(u) = I(V_i \geq u, t_{i2} \leq u) = at \ risk$ with $\geq 2$ measurements at $u$
Assume: Hazard relationship satisfies

\[
\lambda_i(u) = \lim_{du \to 0} P\{u \leq T_i < u + du | T_i \geq u, C_i \geq u, t_i(u), \tilde{W}_i(u), \alpha_i, Z_i\} = \lim_{du \to 0} P\{u \leq T_i < u + du | T_i \geq u, C_i \geq u, t_i(u), \tilde{e}_i(u), \alpha_i, Z_i\} = \lim_{du \to 0} P\{u \leq T_i < u + du | T_i \geq u, \alpha_i, Z_i\} = \lambda_0(u) \exp\{\gamma X_i(u) + \eta^T Z_i\}.
\]

- \(\tilde{W}_i(u) = \{W_i(t_{ij}) : t_{ij} < u\}, \tilde{e}_i(u) = \{e_i(t_{ij}) : t_{ij} < u\}\)

Also assume: Distribution of \(e_i(t_{ij})\) given measurement at \(t_{ij}\), at risk at \(t_{ij}\), measurement history prior to \(t_{ij}\), \(\alpha_i, Z_i\) is \(N(0, \sigma^2)\)

- May be shown with additional assumptions on censoring, timing of measurements \(\Rightarrow\)

\[
\hat{X}_i(u)|Y_i(u) = 1, t_i(u), \alpha_i, Z_i \sim N\{X_i(u), \sigma^2 \theta_i(u)\}
\]

- \(\sigma^2 \theta_i(u) = \text{prediction variance} [\text{depends on } t_i(u)]\)
At any time $u$: Conditional on being at risk at $u$

$$\text{pr}\{dN_i(u) = r, \hat{X}_i(u) = x|Y_i(u) = 1, \alpha_i, Z_i, t_i(u)\}$$

$$= \text{pr}\{dN_i(u) = r|Y_i(u) = 1, \hat{X}_i(u) = x, \alpha_i, Z_i, t_i(u)\} \times \text{pr}\{\hat{X}_i(u) = x|Y_i(u) = 1, \alpha_i, Z_i, t_i(u)\}$$

- 1st piece: Bernoulli$[\lambda_0(u)du \exp\{\gamma X_i(u) + \eta^T Z_i\}]$

- 2nd piece: $\mathcal{N}\{X_i(u), \sigma^2 \theta_i(u)\}$

At time $u$, likelihood $\{dN_i(u), \hat{X}_i(u)\}|Y_i(u) = 1, \alpha_i, Z_i, t_i(u)$: To order $du$

$$= \exp \left[ X_i(u) \left\{ \frac{\gamma \sigma^2 \theta_i(u)dN_i(u) + \hat{X}_i(u)}{\sigma^2 \theta_i(u)} \right\} \right]$$

$$\times \left\{ \frac{\lambda_0(u) \exp(\eta^T Z_i)du}{\{2\pi \sigma^2 \theta_i(u)\}^{1/2}} \right\} \exp \left\{ -\frac{\hat{X}_i^2(u) + X_i^2(u)}{2\sigma^2 \theta_i(u)} \right\}$$
Thus: Sufficient statistic for $\alpha_i$

$$S_i(u, \gamma, \sigma^2) = \gamma \sigma^2 \theta_i(u) dN_i(u) + \hat{X}_i(u)$$

(conditional on $Y_i(u) = 1$)

Suggestion: Conditioning on $S_i(u, \gamma, \sigma^2)$ would remove the dependence of the conditional distribution on (the “nuisance parameter”) $\alpha_i$

- Form estimating equations in same spirit as the partial likelihood score equations
Usual partial likelihood equations: $X_i(u)$ known

- **Intensity** $\lim_{du \to 0} du^{-1} \Pr\{dN_i(u) = 1|X_i(u), Z_i, Y_i(u)\} = \lambda_0(u) \exp\{\gamma X_i(u) + \eta^T Z_i\} Y_i(u) = \lambda_0(u) E_{0i}(u, \gamma, \eta)$

- $\Rightarrow$ **Estimating equations**

$$\sum_{i=1}^n \int \{X_i(u), Z_i^T\}^T \left\{dN_i(u) - E_{0i}(u, \gamma, \eta, \sigma^2) \lambda_0(u) du \right\} = 0$$

$$\sum_{i=1}^n \{dN_i(u) - E_{0i}(u, \gamma, \eta, \sigma^2) \lambda_0(u) du\} = 0$$

$$E_0(u, \gamma, \eta) = \sum_{j=1}^n E_{0j}(u, \gamma, \eta), \quad \hat{\lambda}_0(u) du = \sum_{j=1}^n dN_j(u)/E_0(u, \gamma, \eta)$$

- **Substitute** $\hat{\lambda}_0(u) du$, rearrange, **solve** in $(\gamma, \eta)$

$$\sum_{i=1}^n \int \left[\{X_i(u), Z_i^T\}^T - \frac{E_1(u, \gamma, \eta)}{E_0(u, \gamma, \eta)}\right] dN_i(u) = 0$$

$$E_{1j}(u, \gamma, \eta) = \{X_i(u), Z_i^T\}^T E_{0j}(u, \gamma, \eta),$$

$$E_1(u, \gamma, \eta) = \sum_{j=1}^n E_{1j}(u, \gamma, \eta)$$
Conditional score estimating equations: By analogy

- **Conditional intensity** – can show

\[
\lim_{du \to 0} du^{-1} \Pr\{dN_i(u) = 1|S_i(u, \gamma, \sigma^2), Z_i, t_i(u), Y_i(u)\} = \lambda_0(u) \exp\{\gamma S_i(u, \gamma, \sigma^2) - \gamma^2 \sigma^2 \theta_i(u)/2 + \eta^T Z_i\} Y_i(u) = \lambda_0(u) E^{*}_0(u, \gamma, \eta, \sigma^2)
\]

\[
\Rightarrow \text{Estimating equations}
\]

\[
\sum_{i=1}^n \int \{S_i(u, \gamma, \sigma^2), Z_i^T\}^T \{dN_i(u) - E^{*}_0(u, \gamma, \eta, \sigma^2)\lambda_0(u) du\} = 0
\]

\[
\sum_{i=1}^n \{dN_i(u) - E^{*}_0(u, \gamma, \eta, \sigma^2)\lambda_0(u) du\} = 0
\]

\[
E^{*}_0(u, \gamma, \eta, \sigma^2) = \sum_{j=1}^n E^{*}_{0j}(u, \gamma, \eta, \sigma^2),
\]

\[
\hat{\lambda}_0(u) du = \sum_{j=1}^n dN_j(u)/E^{*}_0(u, \gamma, \eta, \sigma^2)
\]
• Substitute $\hat{\lambda}_0(u)du$, rearrange – solve in $(\gamma, \eta)$

$$\sum_{i=1}^{n} \int \left[ \{S_i(u, \gamma, \sigma^2), Z_i^T\}^T \frac{E_1^*(u, \gamma, \eta, \sigma^2)}{E_0^*(u, \gamma, \eta, \sigma^2)} \right] dN_i(u) = 0$$

$$E_{1j}^*(u, \gamma, \eta, \sigma^2) = \{S_j(u, \gamma, \sigma^2), Z_j^T\}^T E_{0j}^*(u, \gamma, \eta, \sigma^2),$$

$$E_1^*(u, \gamma, \eta, \sigma^2) = \sum_{j=1}^{n} E_{1j}^*(u, \gamma, \eta, \sigma^2)$$

**Remarks:**

• *Easy* to compute

• Can substitute $\hat{\sigma}^2$ based on OLS residuals

• *Consistent, asymptotically normal*, SEs via *sandwich*

• Reduces to *usual partial likelihood* equations when $\sigma^2 = 0$ [so $\hat{X}_i(u) = X_i(u)$]

4. SEMIPARAMETRIC LIKELIHOOD APPROACH

Potential drawbacks of conditional score approach:

- Possible *inefficiency*
- Includes *any* distribution for $\alpha_i$

**Alternative approach:** *Likelihood formulation*

- Take $X_i(u) = f^T(u)\alpha_i$ as before
- Make *realistic* but not *overly restrictive* assumption on $\alpha_i$

**Assume:** $\alpha_i$ have *density* $p(\alpha_i)$ in a class of “*smooth*” densities $\mathcal{H}$
  (Gallant and Nychka, 1987)
  - Densities in $\mathcal{H}$ *do not have* jumps, kinks or oscillations but may be *skewed*, *multi-modal*, *fat-* or *thin-tailed* relative to the normal (and the *normal* is $\in \mathcal{H}$)
  - $\alpha_i = g(\mu, Z_i) + Rz_i$, $R$ lower triangular, $z_i$ has density $h \in \mathcal{H}$
Representation of $h \in \mathcal{H}$:

$$h(z) = P^2_\infty(z)\varphi_q(z) + \text{small lower bound for tail behavior}$$

- $P_\infty(z)$ is an infinite-dimensional polynomial
- $\varphi_q(z)$ is $q$-variate standard normal density

Practical approximation: Truncate

$$h_K(z) = P^2_K(z)\varphi_q(z)$$

- $P_K(z)$ is $K$th order polynomial; e.g., for $K = 2$
  $$P_K(z) = a_{00} + a_{10}z_1 + a_{01}z_2 + a_{20}z_1^2 + a_{02}z_2^2 + a_{11}z_1z_2$$
- Vector of coefficients $a$ must satisfy $\int h_K(z)\,dz = 1$
- $K = 0$ is standard normal $\Rightarrow \alpha_i \sim \mathcal{N}\{g(\mu, Z_i), RR^T\}$
- “SemiNonParametric” (SNP)
**As before:** Proportional hazard model assumption, normal errors same as for conditional score approach

**Implementation issues:**

- **Polar coordinate reparameterization** in terms of vector of angles \( \phi \) to enforce \( \int h_K(z) \, dz = 1 \), numerical stability
- **Parameters** of interest: \( \Omega = \{ \lambda_0(\cdot), \gamma, \eta, \sigma^2, \mu, R, \phi \} \)
- \( K \) controls degree of **flexibility**, must be chosen **somehow**
- Need a **likelihood** function

**Likelihood:** Under certain assumptions on nature of censoring, timing of measurements, likelihood as function of \( \Omega \propto \)

\[
\prod_{i=1}^{n} \int \left[ \lambda_0(V_i) \exp\{\gamma X_i(V_i) + \eta^T Z_i\} \right]^{\Delta_i} \exp \left[ - \int_0^{V_i} \lambda_0(u) \exp\{\gamma X_i(u) + \eta^T Z_i\} \, du \right] \\
\times \frac{1}{(2\pi \sigma^2)^{m_i/2}} \exp \left[ - \sum_{j=1}^{m_i} \frac{(W_i(t_{ij}) - X_i(t_{ij}))^2}{2\sigma^2} \right] h_K(z_i) \, dz_i
\]
Implementation: *EM algorithm*

- **E-step**: Intractable integration by quadrature
- **M-step**: Maximization in \((\mu, R, \phi)\) and \((\gamma, \eta, \sigma^2, \lambda_0)\) *separates*; one-step NR update for \((\gamma, \eta)\)
- SEs, CIs by *profile likelihood*

**Choice of \(K\): Adaptive** based on inspection of *information criteria*

- If \(\ell_K(\hat{\Omega})\) is maximized loglikelihood for fixed \(K\), \(N = \) total number of observations, \(\Omega (p \times 1)\), minimize

\[
\{-\ell_K(\hat{\Omega}) + pC(N)\}/N
\]

- AIC, \(C(N) = 1\); BIC, \(C(N) = \log N/2\);
  Hannan-Quinn (HQ), \(C(N) = \log \log N\)
- AIC prefers “larger” models, BIC “smaller,” HQ intermediate

Song, Davidian, and Tsiatis (2002), *Biometrics*
5. SIMULATION EVIDENCE

Simple situation: \( \eta = 0 \)

- \( X_i(u) = \alpha_{0i} + \alpha_{1i}u \)
- \( E(\alpha_i) = (4.173, -0.0103)^T, \gamma = -1.0, \lambda_0(u) = 1, \)
  \( C_i \sim \exp(1/110) \), additional censoring at 80 weeks, \( \sigma^2 = 0.6 \)
- Nominal \( t_{ij} = (0, 2, 4, 8, 16, 24, 32, 40, 48, 56, 64, 72, 80) \), 10% missing
- Case 1: \( \alpha_i \sim \) bivariate mixture of normals
- Case 2: \( \alpha_i \sim \) bivariate normal
Methods:

- I, Ideal, $X_i(u)$ known for all $u$
- LV, LVCF
- CS, Conditional score, $\sigma^2$ estimated
- $K = 0$, likelihood with normal $\alpha_i$
- SNP, $K$ chosen by HQ

Next slide: 200 Monte Carlo replications, $n = 200$, $\sigma^2 = 0.60$

- MC SD, Average estimated SE, CP of 95% Wald CI
### Case 1: Mixture Scenario

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>LV</th>
<th>CS</th>
<th>K = 0</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>-1.02</td>
<td>-0.82</td>
<td>-1.05</td>
<td>-1.01</td>
<td>-1.01</td>
</tr>
<tr>
<td>SD</td>
<td>0.09</td>
<td>0.07</td>
<td>0.17</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>SE</td>
<td>0.09</td>
<td>0.07</td>
<td>0.14</td>
<td>0.11</td>
<td>0.11</td>
</tr>
<tr>
<td>CP</td>
<td>0.96</td>
<td>0.31</td>
<td>0.94</td>
<td>0.97</td>
<td>0.95</td>
</tr>
</tbody>
</table>

### Case 2: Normal Scenario

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>LV</th>
<th>CS</th>
<th>K = 0</th>
<th>SNP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>-1.01</td>
<td>-0.81</td>
<td>-1.04</td>
<td>-1.00</td>
<td>-1.00</td>
</tr>
<tr>
<td>SD</td>
<td>0.08</td>
<td>0.07</td>
<td>0.15</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>SE</td>
<td>0.08</td>
<td>0.07</td>
<td>0.13</td>
<td>0.10</td>
<td>0.10</td>
</tr>
<tr>
<td>CP</td>
<td>0.96</td>
<td>0.25</td>
<td>0.93</td>
<td>0.96</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Observations:

- *Conditional score* is *inefficient* (but *much* faster...)
- Interesting *robustness* to specification of distribution of $\alpha_i$
- *Not shown*: Likelihood inference on aspects of $\alpha_i$ *compromised* by incorrect distributional assumption
- *Advantage of likelihood inference*: Insight on distribution of $\alpha_i$ $\Rightarrow$ model refinements
True and MC average density estimates:
Intercept marginal – Average and raw estimates:

![Graphs showing density estimates for intercept marginal.]

*Semiparametric Inference in Joint Models*
6. EXAMPLE, REVISITED

Model:

\[ W(t_{ij}) = X_i(t_{ij}) + e(t_{ij}), \quad X_i(u) = \alpha_{0i} + \alpha_{1i}u, \quad e(t_{ij}) \sim N(0, \sigma^2) \]

\[ \alpha_i = \mu_0(1 - Z_i) + \mu_1 Z_i + Rz_i, \quad Z_i = I(\text{Ttrt=ZDV}) \]

\[ \lambda_i(u) = \lambda_0(u) \exp\{\gamma X_i(u) + \eta Z_i\} \]

- 2467 subjects

- **Conditional Score**: Estimate \((\gamma, \eta, \sigma^2)\) with **no assumptions** on \(\alpha_i\) (so no model for \(\alpha_i\))

- **Likelihood**: Assume \(h \in \mathcal{H}\) and estimate \(\{\gamma, \eta, \sigma^2, \mu, R, \lambda_0(u)\}\) with \(K = 0, 1, 2, 3, 4\); all criteria chose \(K = 3\) or 4
<table>
<thead>
<tr>
<th></th>
<th>CS</th>
<th>$K = 0$</th>
<th>$K = 2$</th>
<th>$K = 3$</th>
<th>$K = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(0.207)$</td>
<td>$(0.091)$</td>
<td>$(0.092)$</td>
<td>$(0.092)$</td>
<td>$(0.092)$</td>
</tr>
<tr>
<td>$\eta$</td>
<td>$0.145$</td>
<td>$0.003$</td>
<td>$0.002$</td>
<td>$0.001$</td>
<td>$-0.007$</td>
</tr>
<tr>
<td></td>
<td>$(0.264)$</td>
<td>$(0.132)$</td>
<td>$(0.132)$</td>
<td>$(0.132)$</td>
<td>$(0.131)$</td>
</tr>
<tr>
<td>loglike</td>
<td>—</td>
<td>8558.465</td>
<td>9018.782</td>
<td>9310.945</td>
<td>9347.163</td>
</tr>
<tr>
<td>AIC</td>
<td>—</td>
<td>$-0.412$</td>
<td>$-0.435$</td>
<td>$-0.449$</td>
<td>$-0.450$</td>
</tr>
<tr>
<td>HQ</td>
<td>—</td>
<td>$-0.397$</td>
<td>$-0.418$</td>
<td>$-0.432$</td>
<td>$-0.433$</td>
</tr>
<tr>
<td>BIC</td>
<td>—</td>
<td>$-0.364$</td>
<td>$-0.385$</td>
<td>$\mathbf{-0.397}$</td>
<td>$-0.396$</td>
</tr>
</tbody>
</table>
Estimated joint density, $K = 4$: 

![3D graph showing estimated joint density]
Estimated slope marginal, $K = 0, 2, 3, 4$:
7. DISCUSSION

- Naive methods yield *biased inferences*
- Regression calibration *may also yield bias*
- *Likelihood* or methods like *conditional score* are required, yield *unbiased inferences*
- Conditional score is *easy to compute*, readily extends to *multiple longitudinal processes*, and does not require as *restrictive assumptions on censoring and timing of measurements*
- But is *inefficient*, does not accommodate additional stochastic process, only permits inference on hazard parameters
- Full likelihood approaches are *computationally intensive*; *robustness* to violation of assumptions still *unclear*
- *Philosophical issue*: Modeling the longitudinal process $X_i(u)$