Marginal Non- and Semi-parametric Regression For Longitudinal Data

Naisyin Wang
Texas A&M University

Part of the work are done jointly with Ray Carroll, Zonghui Hu and Xihong Lin.
Longitudinal/Clustered Data

- Longitudinal outcomes or correlated measurements collected from the same subjects.
Longitudinal/Clustered Data

- Longitudinal outcomes or correlated measurements collected from the same subjects.

- Example: CD4 Count Data of HIV seroconverters (Zeger & Diggle, 1994):
  - $n = 369$ subjects.
  - Response $Y$: CD4 counts
  - An overall time trend, $\theta(T)$, $T$ : years since seroconversion.
  - Other covariates, $X$: age, packs of cigarettes, drug use, number of sex partners, and depression score.
Outline

- Data structure and model.
- Interesting issues in the existing approaches: – global or local?
- A new estimation approach.
- Theoretical properties.
- Numerical investigations.
Model and Basic Data Structure

• \( Y_i = (Y_{i1}, \ldots, Y_{im_i})^T \): responses.

• \( W_i = (W_{i1}, \ldots, W_{im_i}) \): covariates.

• \( \mathbb{E}(Y_{ij}|W_i) = \mu(W_{ij}) = \mu_{ij}. \quad \text{Var}(Y|W_i) = \Sigma_i. \)

   ♠ Parametric: \( W = X \)
   \( \star \mu_{ij} = \mu(X_{ij}^t \beta). \)

   ♠ Nonparametric: \( W = T \)
   \( \star \mu_{ij} = \mu(\theta(T_{ij})). \)

   ♠ Semiparametric: \( W = (X, T) \)
   \( \star \mu_{ij} = \mu(\{X_{ij}^t \beta + \theta(T_{ij})\}). \)

• Assume \( m_i \) being finite and \( \mu \) being a known link.
GEE Marginal Estimator

  
  ♠ Assume $V_i$ (working covariance matrix) on $\Sigma_i$.
  
  ★ $V_i = S_i^{1/2} R_i(\tau) S_i^{1/2}$,
  
  ★ $S_i$: diagonal matrix with marginal variances of $Y_{ij}$’s,
  
  ★ $R_i$: invertible working correlation matrix.
  
  ♠ $\hat{\beta}$ is consistent even though $V_i \neq \Sigma_i$; e.g. working independence (WI) estimator with $R_i = I_{m_i \times m_i}$. 
GEE Marginal Estimator


  - Assume $V_i$ (working covariance matrix) on $\Sigma_i$.
    - $V_i = S_i^{1/2} R_i(\tau) S_i^{1/2}$,
    - $S_i$: diagonal matrix with marginal variances of $Y_{ij}$’s,
    - $R_i$: invertible working correlation matrix.

  - $\hat{\beta}$ is consistent even though $V_i \neq \Sigma_i$; e.g. working independence (WI) estimator with $R_i = I_{m_i \times m_i}$.

- **Parametric:** Let $\Delta_i = \text{diag}\{\mu_{i,j}^{(1)}\}$,

  $$\sum_{i=1}^{n} \frac{\partial \mu(X_i\beta)^T}{\partial \beta} V_i^{-1}(Y_i - \mu_i) = \sum_{i=1}^{n} \{X_i^T \Delta_i\} V_i^{-1}(Y_i - \mu_i) = 0,$$

  - Most efficient estimator obtained when $V_i = \Sigma_i$. 

Non/Semiparametric Marginal Estimator

- A non-exhausting reference list:
  - Severini & Staniswalis (1994)
  - Zeger & Diggle (1994)
  - Wild & Yee (1996)
  - Hoover, et al. (1998)
  - Fan & Zhang (2000)
  - Lin & Yin (2001), with discussion.
  - Lin & Carroll (2000, 2001)
Nonparametric

- **Severini & Staniswalis (SS):**

\[
\sum_{i=1}^{n} \left\{ T_i(t)^T \Delta_i(t) \right\} V_i^{-1}(t) K_{ih}(t) \{ Y_i - \mu_i(\alpha, t) \} = 0, \quad \text{where}
\]

\[
K_{ih}(t) = \text{diag}\{ K_h(T_{ij} - t) \}, \quad \mu_{ij}(\alpha, t) = \mu\{ \alpha_0 + \alpha_1(T_{ij} - t)/h \}, \quad \text{and} \quad \hat{\theta}(t) = \hat{\alpha}_0(t).
\]

- **Lin & Carroll (LC):**

\[
\sum_{i=1}^{n} \left\{ T_i(t)^T \Delta_i(t) \right\} K_{ih}^{1/2}(t) V_i^{-1}(t) K_{ih}^{1/2}(t) \{ Y_i - \mu_i(\alpha, t) \} = 0,
\]
Semiparametric: LC:

- Estimating $\theta$: for a given $\beta$,

$$
\mu_{ij}(t, \alpha, \beta) = \mu\{X_{ij}^T \beta + \alpha_0 + \alpha_1(T_{ij} - t)/h\},
$$

and $\hat{\theta}(t, \beta) = \hat{\alpha}_0(t, \beta)$. Also let $V_i = V_{1i}$.

- Estimating $\beta$: profile estimating equations (SS).

$$
\sum_{i=1}^{n} \frac{\partial \mu\{X_i \beta + \hat{\theta}(T_i; \beta)\}^T}{\partial \beta} V_{2i}^{-1} \left[ Y_i - \mu\{X_i \beta + \hat{\theta}(T_i; \beta)\} \right] = 0,
$$

- Taking $R_{1i} = I$ and replacing profile by backfitting result the estimate of Zeger & Diggle (1994).

- Note: for independent data, the two have equivalent asymptotic variance (Opsomer & Ruppert, 1999).
Several Interesting Issues

Under the estimation framework described:

• Best estimated $\theta$ requiring $R_i = I$ – quite different from the parametric GEE!

• In the semiparametric setting, $\sqrt{n}$ consistency of $\hat{\beta}$ requires either $R_i = I$ or under-smoothing.

• LC still recommended $R_{1i} = R_{2i} = I$ under semiparametric setting.

• Regardless what $R_i$ to be used, $\hat{\beta}$ cannot be semiparametric efficient, not even under MVN.

• Numerical results show that $\hat{\beta}$ has smaller variances if accounting for correlation in $\hat{\theta}$ (J-L Wang).
Numerical results show that $\hat{\beta}$ (profile) and $\hat{\beta}$ (backfitting) have different variation for correlated data.

- The semiparametric efficient score under MVN implies that accounting for correlation in $\hat{\theta}$ is required!

- All results seem to imply that to obtain an efficient $\hat{\beta}$, the $\hat{\theta}$ needs to be

  ♠ “local”—to eliminate biases.

  ♠ “global”—to reduce variation.
A New Estimation Approach

• Estimating $\theta(t)$ by

$$
\sum_{i=1}^{n} \sum_{j=1}^{m_i} K_h(t - T_{ij}) \left\{ \mu_{ij}^{(1)} G_{ij}^t(t) \right\} V_i^{-1}(Y_i - \mu^*) = 0,
$$

where $\hat{\theta} = \hat{\alpha}_0$, and

$$
\mu^* = \mu \left[ I(\ell = j) \{ \alpha_0 + \alpha_1(t - T_{ij})/h \} + I(\ell \neq j) \tilde{\theta}(T_{i\ell}) \right],
$$

$
\mu_{i,j}^{(1)}G_{i,j}^t(t)$ is again the derivative term, and $\tilde{\theta}$ is a consistent estimate of $\theta$; Wang (2003)

• In the semiparametric setting, estimate $\beta$ using profile method as before; Wang, Carroll & Lin (2003).
Consider a linear case where $Y_{ij} = \theta(T_{ij}) + \epsilon_{ij}$,

$$
\hat{\theta}(t) \simeq \frac{\sum_i \sum_j K_h(T_{ij} - t) \left[ (v^i)^{jj} Y_{ij} + \sum_{\ell \neq j} (v^i)^{j\ell} \left\{ Y_{il} - \tilde{\theta}_{il} \right\} \right]}{\sum_i \sum_j K_h(T_{ij} - t) (v^i)^{jj}},
$$

where $(v^i)^{j\ell}$ denotes the $(j, \ell)$ entry of $(V^i)^{-1}$.

- Once point $j$ in cluster $i$ is used, all points within cluster $i$ are used – **global**.

- Only the contribution of point $j$ to the estimate is through its response, the rest points are through residuals – **local**.
Theoretical Properties

- Obtain the smallest variation in $\hat{\theta}$ when $V_i = \Sigma_i$ (non- & semiparametric) – this is consistent with the findings in parametric scenario.

- For the 1st order properties of $\hat{\theta}$, only one-step update from the $WI$ estimate of $\theta$ is needed to get the minimum asymptotic variance.

- Variance of the proposed $\hat{\theta}$ is uniformly smaller than or equal to that of the $WI$ estimator.

- No under-smoothing is needed to obtain $\sqrt{n}$ consistency for $\hat{\beta}$. 
• $\hat{\beta}$ is asymptotically normal.

• Under MVN, $\hat{\beta}$ is \textbf{semiparametric efficient}.

• In general cases, $\hat{\beta}$ is more efficient than the \textbf{WI estimator}.

• $\hat{\theta}$ (profile) is at least as efficient as $\hat{\theta}$ (backfitting) for a wide selections of $\hat{\theta}$ under mild conditions (Hu, Wang & Carroll, 2003).

• Under linear link, kernel and spline are “equivalent”, an extension of Silverman (1984); see Lin, Wang, Welsh & Carroll (2003).
Numerical Studies

• Simulation study for the nonparametric estimator.
• Semiparametric efficiency evaluation.
• CD4 data example.

• In the simulation studies:
  ♠ $\theta(t) = \sin(2t)$.
  ♠ All correlation structures considered are compound symmetry.
  ♠ For nonparametric setting, we consider WI, one-step update and fully iterated estimators.
  ♠ For semiparametric setting, we consider WI, Zeger-Diggle–profile (ZD) and the proposed estimators.
Fig. 1. Quantile plot of $R$ for the one-step and fully iterated estimators vs the WI estimator among 500 simulated datasets.
Model: \( Y_{ij} = X_{ij} \beta + \theta(T_{ij}) + \epsilon_{ij} \). \( X, T, \epsilon \): zero mean Gaussian process with correlation parameters, \( \rho_X, \rho_T \), and \( \rho \), respectively.

\[
\text{cor}(X_{ij}, T_{ik}) = \delta_{jk} \rho_{xt}; \quad \delta_{jj} = 1, \quad \delta_{jk} = .6, \quad \rho_T = 0.3,
\]

\[
\rho_X = \rho_{xt} = 0.3 \text{ or } 0.6.
\]
CD4 Data Example

- CD4 Count Data of HIV seroconverters (Zeger & Diggle, 1994)
  - $n = 369$ subjects; $Y$: CD4 counts.
  - An overall time trend, $\theta(T)$, $T$: years since seroconversion.
  - Other covariates, $X$: age, packs of cigarettes, drug use, number of sex partners, and depression score.
  - Working covariance structure—“random intercept plus serial correlation and measurement error” of ZD.
    - A random intercept and an exponential decay serial correlation by specifying the covariance structure as $\tau^2 I + \nu^2 J + \omega^2 H$, where $J$ is a matrix of 1’s and $H(j, k) = \exp(-\alpha |T_{ij} - T_{ik}|)$. 
Regression Coefficients in the CD4 cell counts study in HIV seroconverters using the Semiparametric Efficient and the Working Independence Estimate. For the semiparametric efficient estimates, the working covariance parameter, 
\[ \hat{\xi} = (11.32, 3.26, 22.15, 0.23) \] for Scenario I, and 
\[ \hat{\xi} = (14.1, 6.9, 16.1, 0.22), \] for Scenario II.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>SE</td>
<td>Estimate</td>
</tr>
<tr>
<td>Age</td>
<td>.014</td>
<td>.035</td>
<td>.010</td>
</tr>
<tr>
<td>Smoking</td>
<td>.984</td>
<td>.182</td>
<td>.549</td>
</tr>
<tr>
<td>Drug</td>
<td>1.049</td>
<td>.526</td>
<td>.584</td>
</tr>
<tr>
<td>Sex Partners</td>
<td>-.054</td>
<td>.059</td>
<td>.080</td>
</tr>
<tr>
<td>Depression</td>
<td>-.033</td>
<td>.021</td>
<td>-.045</td>
</tr>
</tbody>
</table>
Estimated $\theta$ for CD4 Data

![Graph showing the estimated $\theta$ over years since seroconversion.](image)