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 sero-conversion.
  - Other covariates, X: age, packs of cigarettes, drug use, number of sex partners, and depression score.

Outline

- Data structure and model.
- Generalized Estimating Equation –parametric nonparametric, and semiparametric.
- 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.
- $\mathrm{E}(Y_{ij}|W_i) = \mu\,(W_{ij}) = \mu_{ij}$. $\mathrm{Var}(Y_i|W_i) = \Sigma_i$.
  - ♠ Parametric: $W = X$
    - ⋆ $\mu_{ij} = \mu\left(X_{ij}^t \beta\right)$.
  - ♠ Nonparametric: $W = T$
    - ⋆ $\mu_{ij} = \mu\,\{\theta\,(T_{ij})\}$.
  - ♠ Semiparametric: $W = (X, T)$
    - ⋆ $\mu_{ij} = \mu\left\{X_{ij}^t \beta + \theta\,(T_{ij})\right\}$.
- Assume m_i being finite and μ being a known link.

GEE Marginal Estimator

- Liang & Zeger (1986), Zeger & Liang (1986)
  - Assume V_i (working covariance matrix) on Σ_i.
    - $\star$ $V_i = S_i^{1/2} R_i(\tau) S_i^{1/2}$,
    - $\star$ S_i: diagonal matrix with marginal variances of Y_{ij}'s,
    - $\star$ 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

- Liang & Zeger (1986), Zeger & Liang (1986)
  - Assume V_i (working covariance matrix) on Σ_i.
    - $\star$ $V_i = S_i^{1/2} R_i(\tau) S_i^{1/2}$,
    - $\star$ S_i: diagonal matrix with marginal variances of Y_{ij}'s,
    - $\star$ R_i: invertible working correlation matrix.
  - $\hat{\beta}$ is consistent even though $V_i \neq \Sigma_i$; e.g. working independence (WI) estimator with $R_i = I_{m_i \times m_i}$.
- **<u>Parametric:</u>** Let Δ_i=diag$\{\mu_{ij}^{(1)}\}$,

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

  - 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):

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

$K_{ih}(t) = \mathrm{diag}\{K_h(T_{ij} - t)\}$, $\mu_{ij}(\alpha, t) = \mu\{\alpha_0 + \alpha_1(T_{ij} - t)/h\}$, and $\widehat{\theta}(t) = \widehat{\alpha}_0(t)$.

- Lin & Carroll (LC):

$$\sum_{i=1}^{n} \{T_i(t)^T \Delta_i(t)\} K_{ih}^{1/2}(t) V_i^{-1}(t) K_{ih}^{1/2}(t) \{Y_i - \mu_i(\alpha, t)\} = 0,$$

Semiparametric: LC:

- Estimating θ: for a given β,

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

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

- Estimating β: profile estimating equations (SS).

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

- 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 θ requiring $R_i = I$ – quite different from the parametric GEE!.
- In the semiparametric setting, $\sqrt{n}$ consistency of $\widehat{\beta}$ 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, $\widehat{\beta}$ cannot be semiparametric efficient, not even under MVN.
- Numerical results show that $\widehat{\beta}$ has smaller variances if accounting for correlation in $\hat{\theta}$ (J-L Wang).

- Numerical results show that $\widehat{\beta}$(profile) and $\widehat{\beta}$(backfitting) have different variation for correlated data.
- The semiparametric efficient score under MVN implies that accounting for correlation in $\widehat{\theta}$ is required!
- All results seem to imply that to obtain an efficient $\widehat{\beta}$, the $\widehat{\theta}$ 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 $\widehat{\theta}=\widehat{\alpha}_0$, and

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

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

- In the semiparametric setting, estimate β 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, ℓ) 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 $\widehat{\theta}$ when $V_i = \Sigma_i$ (non- & semiparametric) – this is consistent with the findings in parametric scenario.

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

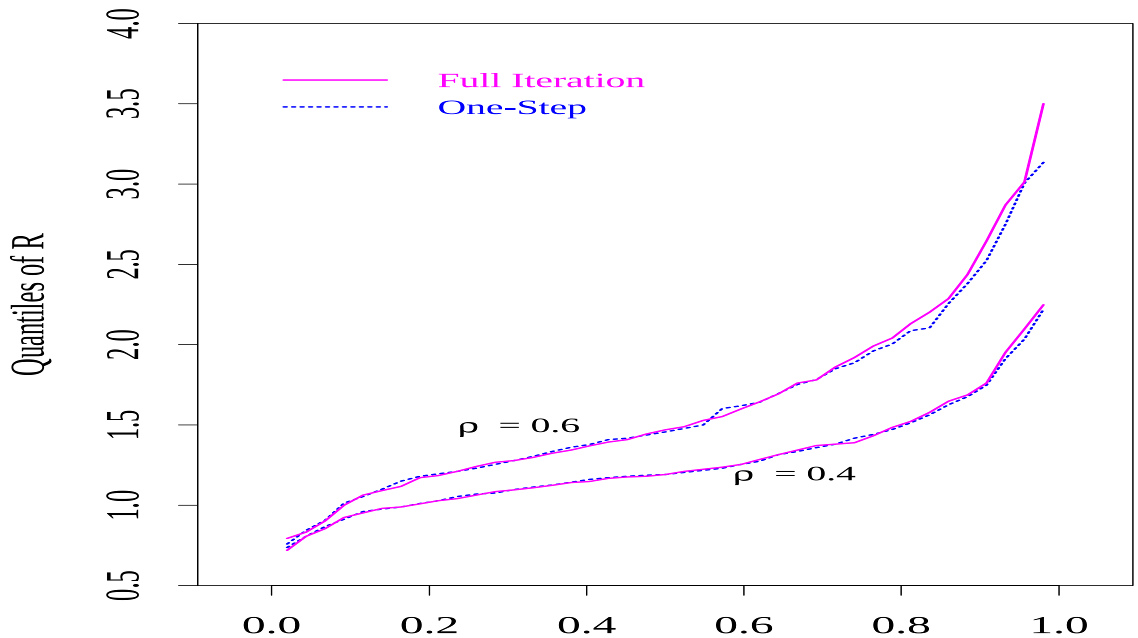
- Variance of the proposed $\widehat{\theta}$ is uniformly smaller than or equal to that of the WI estimator.

- No under-smoothing is needed to obtain $\sqrt{n}$ consistency for $\widehat{\beta}$.

- $\widehat{\beta}$ is asymptotically normal.
- Under MVN, $\widehat{\beta}$ is semiparametric efficient.
- In general cases, $\widehat{\beta}$ is more efficient than the WI estimator.
- $\widehat{\theta}$ (profile) is at least as efficient as $\widehat{\theta}$ (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

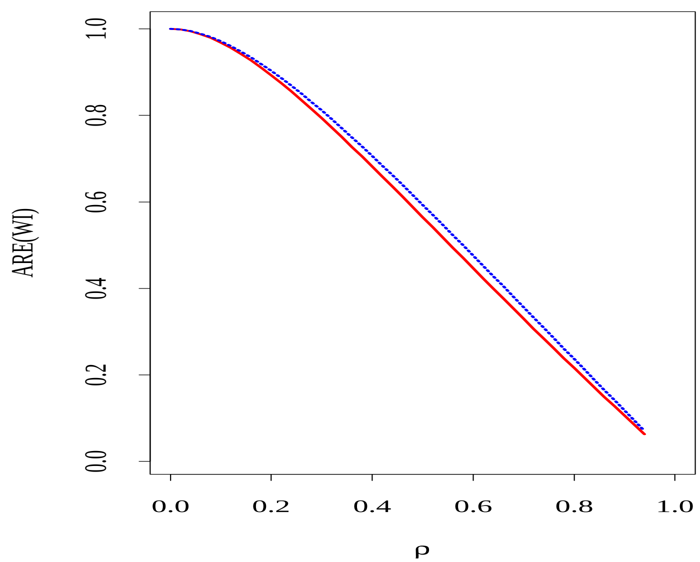
- Simulation study for the nonparametric estimator.
- Semiparametric efficiency evaluation.
- CD4 data example.
- In the simulation studies:
  - $\theta(t) = sin(2t)$.
  - All correlation structures considered are compound symmetry.
  - For nonparametric setting, we consider WI, one-step update and fully iterated estimators.
  - For semiparametric setting, we consider WI, Zeger-Diggle–profile (ZD) and the proposed estimators.

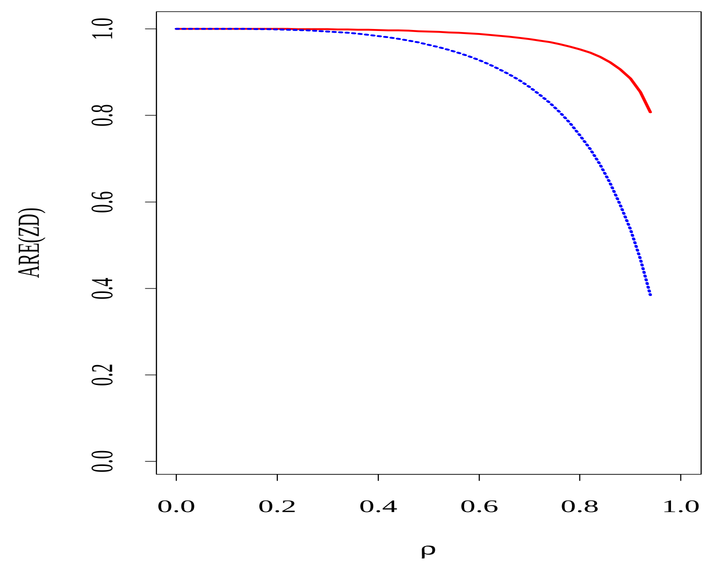


$$R = \frac{SSE(\text{WI estimator})}{SSE(\text{new estimator})}$$

Fig. 1. Quantile plot of R for the one-step and fully iterated estimators vs the WI estimator among 500 simulated datasets.

Numerical Efficiency Study for β





Model: $Y_{ij} = X_{ij}\beta + \theta(T_{ij}) + \epsilon_{ij}$. X, T, ϵ: zero mean Gaussian process with correlation parameters, ρ_X, ρ_T, and ρ, respectively.

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

$\rho_X = \rho_{xt} = 0.3$ 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, $\widehat{\xi}$ =$(11.32, 3.26, 22.15, 0.23)$ *for Scenario I, and* $\widehat{\xi}$ = *(*$14.1, 6.9, 16.1, 0.22$*), for Scenario II.*

	Working Independ		Semi. Efficient Scenario I		Semi. Efficient Scenario II	
	Estimate	SE	Estimate	SE	Estimate	SE
Age	.014	.035	.010	.033	.008	.032
Smoking	.984	.182	.549	.144	.579	.139
Drug	1.049	.526	.584	.331	.584	.335
Sex Partners	- .054	.059	.080	.038	.078	.039
Depression	- .033	.021	-.045	.013	-.046	.014

Estimated θ for CD4 Data