R Estimation for Linear Models for Cluster Correlated Data

John Kloke

28 March 2012
Outline

- Model & Assumptions
- Geometry
- Asymptotic Distribution
- Inference
- Estimates of standard error
R estimation was introduced by Hodges & Lehmann (1963) for estimating shift in the two sample location problem.


Inference & diagnostics developed by many. Summarized in the monograph by Hettmansperger & McKean (2011).

Available in the R package Rfit (Kloke & McKean 2011).
Background

R estimators are

- Robust to outliers in $Y$-space
  (HBR estimators exist which are robust in both $X$ and $Y$ space)
- Nonparametric in that strong distributional assumptions are not needed
- Highly efficient relative to LS estimators at the normal
  ($\text{ARE}(W, LS) = 0.955)$.
Example: Multicenter Clinical Trial

- Randomized, placebo controlled, multicenter clinical trial
- Outcome variable: Change in Triglyceride Level at Week 4.
- Treatments: 4 Active + Placebo
- Two Centers (Clusters): $n_1 = 47$ and $n_2 = 50$. 
Consider an experiment done over \( m \) clusters. For cluster \( k \) we have \( n_k \) observations \( y_{k1}, \ldots, y_{kn_k} \) which we model as

\[
y_k = \alpha \mathbf{1}_{n_k} + X_k \beta + e_k
\]

where

\( \alpha \) is an intercept parameter
\( \beta \) is a \( p \times 1 \) vector of regression coefficients
\( \mathbf{1}_{n_k} \) is an \( n_k \times 1 \) vector of ones
\( X_k = [x_{k1} \cdots x_{kn_k}]^T \) is a \( n_k \times p \) design matrix
\( e_k \) is a \( n_k \times 1 \) random vector of errors
Assume \( e_1, \ldots, e_m \) are independent random vectors.
Let

\[
\begin{bmatrix}
y_1 \\
\vdots \\
y_m
\end{bmatrix}, \quad
\begin{bmatrix}
X_1 \\
\vdots \\
X_m
\end{bmatrix}, \quad \text{and} \quad
\begin{bmatrix}
e_1 \\
\vdots \\
e_m
\end{bmatrix},
\]

then we have

\[
y = \alpha \mathbf{1}_n + X\beta + e.
\]

Where \( n = \sum_{k=1}^{m} n_k \) denotes the total sample size.
Assumptions

Design Assumptions:
- Assume $X$ has full column rank
- WLOG Assume $X$ is centered
- Design conditions similar to LS (e.g. Huber’s condition)

Assumptions on the Errors:
- Equal marginals: $e_{kj} \sim F, f$
- $f$ is continuous and as finite Fisher information.
- True median of $e_{kj}$ is 0
The objective or dispersion function in R estimation is defined as

\[ D(\beta) = \| y - X\beta \|_\varphi \]

where

\[ \| v \|_\varphi = \sum_{t=1}^{n} a(R(v_t))v_t = \sum_{t=1}^{n} \varphi \left( \frac{R(v_t)}{n+1} \right) v_t \]

\( R \) denotes Rank and \( a(R(v_t)) \) for \( t = 1, \ldots, n \) are the scores. The scores are nondecreasing and defined as

\[ a(t) = \varphi \left( \frac{t}{n+1} \right) \]

where \( \varphi \) is nondecreasing on \((0, 1)\) and standardized so that

\[ \int_0^1 \varphi(u) \, du = 0 \quad \text{and} \quad \int_0^1 \varphi^2(u) \, du = 1. \]
Example Score Functions

Wilcoxon (linear) scores: \( \varphi(u) = \sqrt{12} (u - \frac{1}{2}) \)

Sign scores (L1): \( \varphi(u) = \text{sign} \left( u - \frac{1}{2} \right) \)

Normal scores: \( \varphi(u) = \Phi^{-1}(u) \)

Optimal scores:

\[
\varphi(u) = -\frac{f'(F^{-1}(u))}{f(F^{-1}(u))}
\]

Normal scores dominates LS.
Note: $\| \cdot \|_\varphi$ is a pseudo-norm ($\|u\| = 0$ iff $u = u_1 1$)

The dispersion function is

$$D(\beta) = \|y - X\beta\| = a(R(y - X\beta))(y - X\beta).$$

$D(\beta)$ is a continuous, convex function of $\beta$.

Define the gradient as

$$S(\beta) = -\nabla_\beta D(\beta) = X^T a(R(y - X\beta))$$

Note that $S(\hat{\beta}) \neq 0$. 
WLOG assume $\beta_0 = 0$ (Estimator is location & scale equivariant)

$$\frac{1}{\sqrt{n}}S(0) \overset{D}{\to} N_p(0, \Sigma_\varphi)$$

Where

- $\Sigma_\varphi = \lim_{m \to \infty} \frac{1}{n} \sum_{k=1}^{m} X_k^T \Sigma_{\varphi,k} X_k$
- $\Sigma_{\varphi,k} = \text{var}(\varphi(F(e_k)))$
Asymptotics
Kloke, et. al. (2009)

Linearity

\[
\frac{1}{\sqrt{n}} S(\beta) = \frac{1}{\sqrt{n}} S(0) - \frac{1}{\tau \varphi} \frac{1}{n} X^T X \sqrt{n} \beta + o_p(\sqrt{n})
\]

where \( \tau \varphi = \int_0^1 \varphi(u) \left\{ -\frac{f'[F^{-1}(u)]}{f[F^{-1}(u)]} \right\} du \) is a scale parameter

Asymptotic Representation

\[
\sqrt{n} \hat{\beta} \varphi = \tau \varphi (X^T X)^{-1} X^T \varphi(F(e)) + o_p(1)
\]
Quadraticity

Let

\[ Q(\beta) = \frac{1}{2\tau_\phi} \beta^T (X^T X)^{-1} \beta - \beta^T S(0) + D(0) \]

denote a quadratic approximation to \( D(\beta) \).

Under regularity and design conditions

\[
\lim_{m \to \infty} P \left( \sup_{\|\beta\| \leq c/\sqrt{n}} |D(\beta) - Q(\beta)| \geq \epsilon \right) = 0
\]
The distribution of $\hat{\beta}_\varphi$ is normal with mean $\beta_0 = \beta$ and covariance matrix

$$V_\varphi = \tau_\varphi^2 (X^T X)^{-1} \Sigma_\varphi (X^T X)^{-1}$$

where

- $\Sigma_\varphi = \sum_{k=1}^{m} X_k^T \Sigma_{\varphi,k} X_k$
- $\Sigma_{\varphi,k} = \text{var}(\varphi(F(e_k)))$
- $F(e_k) = [F(e_{k1}), \ldots, F(e_{kn_k})]^T$
- $F$ is the cdf of $e_{kj}$ for $k = 1, \ldots, m$, $j = 1, \ldots, n_k$

$\tau_\varphi$ is estimated with the Koul, et. al. (1987) estimator.
Use a robust measure of location (median)

\[ \hat{\alpha}_s = \text{med}\{ y_{ki} - x_{kj}^T \hat{\beta}_\varphi \} \]

**Joint Distribution**

\[
\begin{bmatrix} \hat{\alpha}_s \\ \hat{\beta}_\varphi \end{bmatrix} \sim N_{p+1} \left( \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \begin{bmatrix} \sigma_s^2 & 0^T \\ 0 & V_\varphi \end{bmatrix} \right)
\]

where

\[ \sigma_s^2 = \frac{\tau_s^2}{n} \sum_{k=1}^m \left[ \sum_{j=1}^{n_k} \text{var}(\text{sgn}(e_{kj})) + \sum_{j \neq j'} \text{cov}(\text{sgn}(e_{kj}), \text{sgn}(e_{kj'})) \right] \]

and \( \tau_s = \frac{1}{2f(0)} \).
Assume $\mathbf{e}_k = \mathbf{1}_{n_k} b_k + \epsilon_k$ where $\epsilon_k = [\epsilon_{k1}, \ldots, \epsilon_{kn_k}]^T$ is a vector of iid errors and $\epsilon_{k1}, \ldots, \epsilon_{kn_k}, b_k$ are independent (as with ML).

Then $\Sigma_{\varphi,k} = \text{var}(\varphi(F(\mathbf{e}_k)))$ becomes

$$
\Sigma_{\varphi,k} = (1 - \rho_\varphi)\mathbf{I}_{n_k} + \rho_\varphi \mathbf{J}_{n_k} \quad (\text{CS})
$$

where $\rho_\varphi = \text{cor}(\varphi(F(e_{11})), \varphi(F(e_{12})))$. Estimate with

$$
\hat{\rho}_\varphi = M^{-1} \sum_{k=1}^m \sum_{i>j\, a(R(\hat{e}_{ki})) a(R(\hat{e}_{kj}))}
$$

where $M = \sum_{k=1}^m \binom{n_k}{2} - p$. 
Tests $H_0 : \beta = 0$

**Score Test**

$$S(0)^T \Sigma^{-1}_\varphi S(0) \xrightarrow{D} \chi^2_p$$

**Wald Test**

$$\hat{\beta}_\varphi^T \hat{V}_\varphi^{-1} \hat{\beta}_\varphi \xrightarrow{D} \chi_p$$

**Reduction in Dispersion Test (designed experiments)**

$$\frac{\|y\|_\varphi - \|y - \hat{y}\|_\varphi}{(1 - \hat{\rho}_\varphi)\hat{\tau}_\varphi/2} \xrightarrow{D} \chi^2_p$$
Denote the residuals of the JR fit by

\[ \hat{e}_{JR} = y - \hat{\alpha}_s \mathbf{1} - X \hat{\beta}_\varphi \]

Using the asymptotic representations for \( \hat{\alpha}_s \) and \( \hat{\beta}_\varphi \) the first order expression for \( \hat{e}_{JR} \) is

\[ \hat{e}_{JR} = e - \frac{\tau_s}{N} J_N \text{sgn}(e) - \tau_\varphi H x C \varphi [F(e)] \]

So that the

\[ \text{var}[\hat{e}_{JR}] = C_{JR} \]

and the Studentized residuals are

\[ r_{JR,t} = \frac{\hat{e}_{JR,t}}{\sqrt{c_{tt}}} \]

where \( c_{tt} \) is the \( t \)th diagonal element of \( C_{JR} \).
Variance Components

Let

\[ \hat{b}_k = \text{med}_{1 \leq j \leq n_k} \{ \hat{e}_{kj} \}, \quad k = 1, \ldots, m, \]

A robust estimator of \( \sigma^2_b \) is

\[ \hat{\sigma}^2_b = (\text{MAD}_{1 \leq k \leq m} \{ \hat{b}_k \})^2. \]

The natural residuals for the errors \( \varepsilon_{kj} \) are then

\[ \hat{\varepsilon}_{kj} = \hat{e}_{kj} - \hat{b}_k, \quad j = 1, \ldots, n_k, k = 1, \ldots, m. \]

Thus a robust estimate of \( \sigma^2_\varepsilon \) is

\[ \hat{\sigma}^2_\varepsilon = (\text{MAD}_{1 \leq j \leq n, 1 \leq k \leq m} \{ \hat{\varepsilon}_{kj} \})^2. \]

To complete the robust estimates of the variance components, let

\[ \hat{\sigma}^2 = \hat{\sigma}^2_\varepsilon + \hat{\sigma}^2_b \quad \text{and} \quad \hat{\rho} = \hat{\sigma}^2_b / \hat{\sigma}^2. \]
Example: Multicenter Clinical Trial

- Randomized, placebo controlled, multicenter clinical trial
- Outcome variable: Change in Triglyceride Level at Week 4.
- Treatments: 4 Active + Placebo
- Two Centers (Clusters): \( n_1 = 47 \) and \( n_2 = 50 \).
Comparison Boxplots

in Triglyceride (Baseline – Week 4)

Treatment

R Estimation for Linear Models for Cluster Correlated Data
## Results With Outlier

<table>
<thead>
<tr>
<th></th>
<th>Wilcoxon</th>
<th>ML</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>SE</td>
</tr>
<tr>
<td>trt2</td>
<td>0.29</td>
<td>8.11</td>
</tr>
<tr>
<td>trt3</td>
<td>-2.41</td>
<td>8.11</td>
</tr>
<tr>
<td>trt4</td>
<td>33.04</td>
<td>7.90</td>
</tr>
<tr>
<td>trt5</td>
<td>26.11</td>
<td>8.11</td>
</tr>
</tbody>
</table>
Studentized Residuals

Normal Q–Q Plot

Theoretical Quantiles

Sample Quantiles
### Results

Outlier Removed

<table>
<thead>
<tr>
<th></th>
<th>Wilcoxon</th>
<th></th>
<th></th>
<th></th>
<th>ML</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>SE</td>
<td>t value</td>
<td></td>
<td>Est</td>
<td>SE</td>
<td>t value</td>
</tr>
<tr>
<td>trt2</td>
<td>0.34</td>
<td>7.57</td>
<td>0.05</td>
<td></td>
<td>3.20</td>
<td>13.39</td>
<td>0.24</td>
</tr>
<tr>
<td>trt3</td>
<td>-2.39</td>
<td>7.57</td>
<td>-0.32</td>
<td></td>
<td>-12.07</td>
<td>13.39</td>
<td>-0.90</td>
</tr>
<tr>
<td>trt4</td>
<td>36.28</td>
<td>7.57</td>
<td>4.79</td>
<td></td>
<td>35.04</td>
<td>13.39</td>
<td>2.62</td>
</tr>
<tr>
<td>trt5</td>
<td>25.18</td>
<td>7.57</td>
<td>3.32</td>
<td></td>
<td>29.28</td>
<td>13.39</td>
<td>2.19</td>
</tr>
</tbody>
</table>

---

John Kloke

R Estimation for Linear Models for Cluster Correlated Data
p-values for tests of $H_0: \Delta_1 = \Delta_2 = \Delta_3 = \Delta_4 = 0$

<table>
<thead>
<tr>
<th></th>
<th>Wald (Wilcoxon)</th>
<th>LRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>With Outlier</td>
<td>1e-05</td>
<td>0.5887</td>
</tr>
<tr>
<td>Outlier Removed</td>
<td>0+</td>
<td>0.002652</td>
</tr>
</tbody>
</table>
Let $\hat{\tau}_\varphi$ denote a consistent estimator of the scale parameter $\tau_\varphi$ and $\hat{\rho}$ denote a consistent estimator of the correlation parameter $\rho$.

The reduction in dispersion test is defined as

$$F_\varphi = \frac{(\|y - \hat{y}_R\|_\varphi - \|y - \hat{y}_F\|_\varphi)/q}{(1 - \hat{\rho})\hat{\tau}_\varphi/2}$$

where $\hat{y}_R$ is the reduced model fit, $\hat{y}_F$ is the full model fit and $q$ is the difference in the number of parameters.

Compare to $F_{q, n - p - 1}$ critical values.
**Simulation**

**Setup:**
3 Treatments (Control + 2 Active)
5 Centers
30 Subjects per Center (10 per treatment)
Errors: Normal, Logistic, Laplace, CN(ε = 0.25, σ_c = 10)
Effect Sizes: Δ_1 = Δ_2 = θ.
Rank-Based and ML Test of

\[ H_0 : \Delta_1 = \Delta_2 = 0 \text{ versus } H_A : \Delta_1 \neq 0 \text{ or } \Delta_2 \neq 0 \]

\[ \text{power} = \frac{\#\text{Reject}}{1000} \ (\alpha = 0.05 \ & \ \text{sim size} = 1000) \]
Results

Normal

Logistic

Laplace

Contaminated Normal

Power vs. θ
Estimates of $\Sigma_\varphi$

**Compound Symmetric**

\[ \hat{\rho}_\varphi = \frac{1}{M - p} \sum_{k=1}^{m} \sum_{i>j} a(R(\hat{e}_{ki}))a(R(\hat{e}_{kj})) \]

where $M = \sum_{k=1}^{m} \binom{n_k}{2}$

**Empirical** To simplify notation, let $a_{ki} = a(R(\hat{e}_{ki}))$. Estimate $\sigma_{ij}$ with

\[ \hat{\sigma}_{ij} = \sum_{k=1}^{m} (a_{ki} - \bar{a}.i)(a_{kj} - \bar{a}.j) \]

where $\bar{a}.i = \sum_{k=1}^{m} a_{ki}$.

**Sandwich**

\[ \sum_{k=1}^{m} X_k^T a(R(\hat{e}_k))a(R(\hat{e}_k))^T X_k. \]
Simulation

- Score Test
- $m = 12, 25, 100$ Clusters
- $k = 2, 4$ Treatments
- $n = 2$ Replicates
- $\rho = 0.25, 0.75$
- CS & AR(1)
- Normal errors
- Simulation size = 1000
Simulation Results
CS, k=2, Empirical Level

<table>
<thead>
<tr>
<th>m</th>
<th>rho</th>
<th>empirical</th>
<th>sandwich</th>
<th>cs</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.25</td>
<td>0.07</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>12</td>
<td>0.75</td>
<td>0.11</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td>25</td>
<td>0.25</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
</tr>
<tr>
<td>25</td>
<td>0.75</td>
<td>0.07</td>
<td>0.06</td>
<td>0.06</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>100</td>
<td>0.75</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
</tr>
</tbody>
</table>
## Simulation Results

**CS, k=2, Empirical Power**

<table>
<thead>
<tr>
<th>m</th>
<th>rho</th>
<th>empirical</th>
<th>sandwich</th>
<th>cs</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.25</td>
<td>0.65</td>
<td>0.64</td>
<td>0.61</td>
</tr>
<tr>
<td>12</td>
<td>0.75</td>
<td>0.68</td>
<td>0.66</td>
<td>0.67</td>
</tr>
<tr>
<td>25</td>
<td>0.25</td>
<td>0.63</td>
<td>0.63</td>
<td>0.63</td>
</tr>
<tr>
<td>25</td>
<td>0.75</td>
<td>0.62</td>
<td>0.60</td>
<td>0.62</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.63</td>
<td>0.63</td>
<td>0.62</td>
</tr>
<tr>
<td>100</td>
<td>0.75</td>
<td>0.59</td>
<td>0.58</td>
<td>0.59</td>
</tr>
</tbody>
</table>
### Simulation Results

**AR(1), k=2, Empirical Level**

<table>
<thead>
<tr>
<th>m</th>
<th>rho</th>
<th>empirical</th>
<th>sandwich</th>
<th>cs</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.25</td>
<td>0.10</td>
<td>0.09</td>
<td>0.11</td>
</tr>
<tr>
<td>12</td>
<td>0.75</td>
<td>0.10</td>
<td>0.09</td>
<td>0.17</td>
</tr>
<tr>
<td>25</td>
<td>0.25</td>
<td>0.08</td>
<td>0.07</td>
<td>0.09</td>
</tr>
<tr>
<td>25</td>
<td>0.75</td>
<td>0.06</td>
<td>0.05</td>
<td>0.14</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.06</td>
<td>0.06</td>
<td>0.08</td>
</tr>
<tr>
<td>100</td>
<td>0.75</td>
<td>0.05</td>
<td>0.05</td>
<td>0.13</td>
</tr>
</tbody>
</table>
Simulation Results
AR(1), k=2, Empirical Power

<table>
<thead>
<tr>
<th>m</th>
<th>rho</th>
<th>empirical</th>
<th>sandwich</th>
<th>cs</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.25</td>
<td>0.56</td>
<td>0.56</td>
<td>0.63</td>
</tr>
<tr>
<td>12</td>
<td>0.75</td>
<td>0.55</td>
<td>0.51</td>
<td>0.68</td>
</tr>
<tr>
<td>25</td>
<td>0.25</td>
<td>0.56</td>
<td>0.56</td>
<td>0.64</td>
</tr>
<tr>
<td>25</td>
<td>0.75</td>
<td>0.50</td>
<td>0.48</td>
<td>0.65</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.58</td>
<td>0.58</td>
<td>0.67</td>
</tr>
<tr>
<td>100</td>
<td>0.75</td>
<td>0.49</td>
<td>0.48</td>
<td>0.66</td>
</tr>
</tbody>
</table>
## Simulation Results

CS, k=4, Empirical Level

<table>
<thead>
<tr>
<th>m</th>
<th>rho</th>
<th>empirical</th>
<th>sandwich</th>
<th>cs</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.25</td>
<td>0.16</td>
<td>0.16</td>
<td>0.05</td>
</tr>
<tr>
<td>12</td>
<td>0.75</td>
<td>0.20</td>
<td>0.16</td>
<td>0.07</td>
</tr>
<tr>
<td>25</td>
<td>0.25</td>
<td>0.09</td>
<td>0.09</td>
<td>0.05</td>
</tr>
<tr>
<td>25</td>
<td>0.75</td>
<td>0.13</td>
<td>0.11</td>
<td>0.06</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>100</td>
<td>0.75</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
</tr>
</tbody>
</table>
## Simulation Results

CS, $k=4$, Empirical Power

<table>
<thead>
<tr>
<th>m</th>
<th>rho</th>
<th>empirical</th>
<th>sandwich</th>
<th>cs</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.25</td>
<td>0.62</td>
<td>0.61</td>
<td>0.48</td>
</tr>
<tr>
<td>12</td>
<td>0.75</td>
<td>0.66</td>
<td>0.63</td>
<td>0.54</td>
</tr>
<tr>
<td>25</td>
<td>0.25</td>
<td>0.56</td>
<td>0.56</td>
<td>0.48</td>
</tr>
<tr>
<td>25</td>
<td>0.75</td>
<td>0.57</td>
<td>0.54</td>
<td>0.51</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.47</td>
<td>0.47</td>
<td>0.45</td>
</tr>
<tr>
<td>100</td>
<td>0.75</td>
<td>0.47</td>
<td>0.47</td>
<td>0.48</td>
</tr>
</tbody>
</table>

John Kloke

R Estimation for Linear Models for Cluster Correlated Data
## Simulation Results

**AR(1), k=4, Empirical Level**

<table>
<thead>
<tr>
<th>m</th>
<th>rho</th>
<th>empirical</th>
<th>sandwich</th>
<th>cs</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.25</td>
<td>0.15</td>
<td>0.16</td>
<td>0.10</td>
</tr>
<tr>
<td>12</td>
<td>0.75</td>
<td>0.17</td>
<td>0.16</td>
<td>0.23</td>
</tr>
<tr>
<td>25</td>
<td>0.25</td>
<td>0.08</td>
<td>0.08</td>
<td>0.10</td>
</tr>
<tr>
<td>25</td>
<td>0.75</td>
<td>0.11</td>
<td>0.11</td>
<td>0.20</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.08</td>
<td>0.08</td>
<td>0.13</td>
</tr>
<tr>
<td>100</td>
<td>0.75</td>
<td>0.05</td>
<td>0.06</td>
<td>0.21</td>
</tr>
</tbody>
</table>
## Simulation Results

### AR(1), k=4, Empirical Power

<table>
<thead>
<tr>
<th>m</th>
<th>rho</th>
<th>empirical</th>
<th>sandwich</th>
<th>cs</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.25</td>
<td>0.51</td>
<td>0.52</td>
<td>0.55</td>
</tr>
<tr>
<td>12</td>
<td>0.75</td>
<td>0.44</td>
<td>0.43</td>
<td>0.52</td>
</tr>
<tr>
<td>25</td>
<td>0.25</td>
<td>0.44</td>
<td>0.43</td>
<td>0.50</td>
</tr>
<tr>
<td>25</td>
<td>0.75</td>
<td>0.29</td>
<td>0.29</td>
<td>0.47</td>
</tr>
<tr>
<td>100</td>
<td>0.25</td>
<td>0.41</td>
<td>0.41</td>
<td>0.53</td>
</tr>
<tr>
<td>100</td>
<td>0.75</td>
<td>0.27</td>
<td>0.27</td>
<td>0.48</td>
</tr>
</tbody>
</table>
Generalized Estimator

Model

\[ y_k = X_k \beta + e_k. \]

Assume \( V_k = \text{var}[e_k] \).
Let \( \hat{V}_k \) be a robust, consistent estimate of \( V_k \).
Transform responses the responses to Working Independence:

\[ y_k^\dagger = \hat{V}_k^{-1/2} y_k \]

and model as

\[ y_k^\dagger = X_k^\dagger \beta + e_k^\dagger \]

estimate \( \beta \) using JR.
Example AR(1)

Use Koul & Saleh (1993) estimate of $\rho$ based on the residuals of an initial fit. Use sandwich estimator for standard errors.

**Simulation**

- Normal errors, $\rho = 0.5$
- $m = 16$ subjects, $n = 5$ measurements
- Simulation size = 1000
- One treatment effect, one baseline covariate (normal).
- Empirical levels: 0.048 0.050 ($\alpha = 0.05$)
Summary

R Estimators are

- Robust
- Applicable to a wide range of problems
- Complete inference (estimates, standard errors, diagnostics tools, CIs, tests of GLH)
- Implemented in R: Rfit
- Can be more efficient/powerful then traditional LS/ML methods.
- Sandwich estimator appears to work well for large samples.
Future Work

- Add clustered correlated inference to Rfit
- Applications of R (JR, GJR) methods in practice
- Model Selection


