Appendix A: Deriving the Efficient Score

Let $O_i = (\mathbf{Y}_i, A_i, \mathbf{X}_i)$, where $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, ... Y_{in_i})^{\mathrm{T}}$ is the n_i-dimensional response vector for the i^{th} independent unit, $i = 1, ..., m$, A_i is a scalar treatment assignment, and $\mathbf{X}_i$ is a matrix of auxiliary covariates. The optimal index $h_{opt}(A,t)$ is determined by solving the generalized information equality

$$-E\left[\left.\frac{\partial \psi\{\mathbf{Y}, A, \mathbf{X}, t; \beta, \gamma, h(\cdot)\}}{\partial \beta^{\mathrm{T}}}\right|_{\beta=\beta_0}\right] = E\left[\left.\psi\{\mathbf{Y}, A, \mathbf{X}, t; \beta, \gamma, h(\cdot)\}\psi^{\mathrm{T}}\{\mathbf{Y}, A, \mathbf{X}, t; \beta, \gamma, h_{opt}(\cdot)\}\right|_{\beta=\beta_0}\right], \tag{1}$$

for h_{opt}, where $h(\cdot)$ is any $p \times n_i$ function such that $E[\psi^{\mathrm{T}}\psi] < \infty$.

Conditioning on t, $h(A,t)$ takes up to K different matrix values, $h_0(t), h_1(t), ..., h_{K-1}(t)$, which may be denoted by K $p \times n_i$ constant matrices $\mathbf{h}_0, \mathbf{h}_1, ..., \mathbf{h}_{K-1}$. Similarly, we define $\Delta_k(\mathbf{X}) = E(\mathbf{Y}|A = k, \mathbf{X}, t) - \mathbf{g}(k, t; \beta)$, the n_i-dimensional vector of the difference in the conditional and marginal mean outcomes under treatment k, where $k = 0, 1, ..., K-1$. Using this construction, let $\mathbf{h} = [\mathbf{h}_0, \mathbf{h}_1, ..., \mathbf{h}_{K-1}]$ and $\Delta_K(\mathbf{X}) = \{\Delta_0^{\mathrm{T}}(\mathbf{X}), ..., \Delta_{K-1}^{\mathrm{T}}(\mathbf{X})\}^{\mathrm{T}}$. The complete index matrix $\mathbf{h}$ is therefore of dimension $p \times Kn_i$, while Δ_K is a Kn_i-dimensional vector. Estimating functions are then expressed concisely through defining two auxiliary matrix functions of A. Let $\mathbf{A}$ be the $Kn_i \times n_i$ matrix $\mathbf{A} = [I(A = 1)\mathbf{I}_n, ..., I(A = K)\mathbf{I}_n]^{\mathrm{T}}$ and $\mathbf{A}_\pi$ be the $Kn_i \times Kn_i$ block diagonal matrix composed of the diagonal matrices $\{I(A = k) - \pi_k\}\mathbf{I}_n$, where $\mathbf{I}_n$ denotes the $n_i \times n_i$ identity matrix.

Rewriting the augmented estimating equations using this notation, we obtain

$$\sum_{i=1}^{m} \mathbf{h}_i \mathbf{A}_i\{\mathbf{Y_i} - \mathbf{g}(A_i, t; \beta)\} - \mathbf{h}\mathbf{A}_{\pi_{\mathbf{i}}i}\Delta_i(\mathbf{X}_i) = \mathbf{0}. \tag{2}$$

Substituting this expression into Newey's equations we have

$$E\left[\mathbf{hA}\frac{\partial \mathbf{g}(A,t;\beta)}{\partial \beta^{\mathrm{T}}}\right] = E\big[\{\mathbf{hA}(\mathbf{Y} - \mathbf{g}(A,t;\beta)) - \mathbf{hA}_\pi\Delta_K(\mathbf{X})\} \times \\ \{(\mathbf{Y} - \mathbf{g}(A,t;\beta))^{\mathrm{T}}\mathbf{A}\mathbf{h}_{\mathbf{opt}}^{\mathrm{T}} - \Delta_K^{\mathrm{T}}(\mathbf{X})\mathbf{A}_\pi\mathbf{h}_{\mathbf{opt}}^{\mathrm{T}}\}\big]$$

We first note that since $\mathbf{h}$ and $\mathbf{h_{opt}}$ are constant, we can extract them from the expectation, leaving

$$\mathbf{h}^{\mathrm{T}}E\left[\mathbf{A}\frac{\partial \mathbf{g}(A,t;\beta)}{\partial \beta^{\mathrm{T}}}\right] = \mathbf{h}^{\mathrm{T}}E\big[\{\mathbf{A}(\mathbf{Y} - \mathbf{g}(A,t;\beta)) - \mathbf{A}_\pi\Delta_K(\mathbf{X})\} \times \\ \{(\mathbf{Y} - \mathbf{g}(A,t;\beta))^{\mathrm{T}}\mathbf{A} - \Delta_K^{\mathrm{T}}(\mathbf{X})\mathbf{A}_\pi\}\big]\mathbf{h}_{\mathbf{opt}}^{\mathrm{T}}$$

Since $\mathbf{h}$ is nonzero, it must hold that

$$E\left[\mathbf{A}\frac{\partial \mathbf{g}(A,t;\beta)}{\partial \beta^{\mathrm{T}}}\right] = E\left[\{\mathbf{A}(\mathbf{Y}-\mathbf{g}(A,t;\beta)) - \mathbf{A}_{\pi}\Delta_K(\mathbf{X})\}\{(\mathbf{Y}-\mathbf{g}(A,t;\beta))^{\mathrm{T}}\mathbf{A} - \Delta_K^{\mathrm{T}}(\mathbf{X})\mathbf{A}_{\pi}\}\right]\mathbf{h}_{\mathbf{opt}}^{\mathrm{T}} \tag{3}$$

Evaluating the left hand side of the equation, we have

$$E\left\{\begin{bmatrix} A_0\mathbf{I}_n \\ A_2\mathbf{I}_n \\ \vdots \\ A_{K-1}\mathbf{I}_n \end{bmatrix} \frac{\partial \mathbf{g}(A,t;\beta)}{\partial \beta^{\mathrm{T}}}\right\} = \begin{bmatrix} \pi_0 \frac{\partial \mathbf{g}(0,t;\beta)}{\partial \beta^{\mathrm{T}}} \\ \pi_1 \frac{\partial \mathbf{g}(1,t;\beta)}{\partial \beta^{\mathrm{T}}} \\ \vdots \\ \pi_{K-1} \frac{\partial \mathbf{g}(K-1,t;\beta)}{\partial \beta^{\mathrm{T}}} \end{bmatrix} \tag{$\mathbf{D}^*$}$$

Evaluating the right hand side, we note that we have an expression of the form $E[A-B][A_{opt}-B_{opt}]^{\mathrm{T}}$. Interpreting the augmented estimating function as a residual, we note that $A-B \perp B_{opt}$. We can therefore evaluate $E[A-B][A_{opt}-B_{opt}]^{\mathrm{T}} = E[A-B][A_{opt}]^{\mathrm{T}}$. In (3), this becomes

$$E[\mathbf{A}\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}\{\mathbf{Y}-g(A,t;\beta)\}^{\mathrm{T}}\mathbf{A}] - E[\mathbf{A}_{\pi}\Delta(\mathbf{X})\{\mathbf{Y}-g(A,t;\beta)\}^{\mathrm{T}}\mathbf{A}] \tag{4}$$

Regarding the first term in (4), we have

$$E[\mathbf{A}\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}\{\mathbf{Y}-g(A,t;\beta)^{\mathrm{T}}\}\mathbf{A}] =$$
$$E\begin{bmatrix} A_0A_0\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2} & A_0A_1\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2} & \cdots & A_0A_{K-1}\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2} \\ A_1A_0\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2} & A_1A_1\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2} & \cdots & A_1A_{K-1}\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{K-1}A_0\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2} & A_{K-1}A_1\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2} & \cdots & A_{K-1}A_{K-1}\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2} \end{bmatrix}, \tag{5}$$

where $U^{\otimes 2}$=UU^{T}. Since each individual is only assigned to one treatment, only one of A_0, A_1,...,A_{K-1} is nonzero. The non diagonal blocks of (5) are identically 0. The diagonal blocks contain terms of the form $E[A_kA_k\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2}] = E[A_k\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\otimes 2}] = \pi_k V(\mathbf{Y}|A=k)$. Matrix (5) is written as

$$\begin{bmatrix} \pi_0 V(\mathbf{Y}|A=0) & 0 & \cdots & 0 \\ 0 & \pi_1 V(\mathbf{Y}|A=1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi_{K-1}V(\mathbf{Y}|A=K-1) \end{bmatrix} \tag{$\mathbf{C}_1$}$$

Evaluating the second term of (4), we have

$$E[\mathbf{A}_\pi \Delta_K(\mathbf{X})\{\mathbf{Y}-g(A,t,\beta)\}^{\mathrm{T}}\mathbf{A}] =$$

$$E\left\{ \begin{bmatrix} (A_0-\pi_0)\mathbf{I}_n & \dots & 0 \\ \vdots & (A_1-\pi_1)\mathbf{I}_n & \vdots \\ 0 & \cdots & (A_{K-1}-\pi_{K-1})\mathbf{I}_n \end{bmatrix} \begin{bmatrix} \Delta_0(\mathbf{X}) \\ \Delta_1(\mathbf{X}) \\ \vdots \\ \Delta_{K-1}(\mathbf{X}) \end{bmatrix} \times \right.$$

$$\left. \{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\mathrm{T}} \begin{bmatrix} A_0\mathbf{I}_n & A_1\mathbf{I}_n & \cdots & A_{K-1}\mathbf{I}_n \end{bmatrix} \right\}$$

$$=E\left\{ \begin{bmatrix} (A_0-\pi_0)A_0\Delta_0(\mathbf{X})\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\mathrm{T}} & \cdots & (A_0-\pi_0)A_{K-1}\Delta_0(\mathbf{X})\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\mathrm{T}} \\ (A_1-\pi_1)A_0\Delta_1(\mathbf{X})\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\mathrm{T}} & \ddots & (A_1-\pi_1)A_{K-1}\Delta_1(\mathbf{X})\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ (A_{K-1}-\pi_{K-1})A_0\Delta_{K-1}(\mathbf{X})\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\mathrm{T}} & \cdots & (A_{K-1}-\pi_{K-1})A_{K-1}\{\mathbf{Y}-\mathbf{g}(A,t;\beta)\}^{\mathrm{T}} \end{bmatrix} \right\}$$

$$=E\left\{ \begin{bmatrix} (A_0-\pi_0)A_0\Delta_0(\mathbf{X})\Delta_A^{\mathrm{T}}(\mathbf{X}) & \cdots & (A_0-\pi_0)A_{K-1}\Delta_0(\mathbf{X})\Delta_A^{\mathrm{T}}(\mathbf{X}) \\ (A_1-\pi_1)A_0\Delta_1(\mathbf{X})\Delta_A^{\mathrm{T}}(\mathbf{X}) & \ddots & (A_1-\pi_1)A_{K-1}\Delta_1(\mathbf{X})\Delta_A^{\mathrm{T}}(\mathbf{X}) \\ \vdots & \ddots & \vdots \\ (A_{K-1}-\pi_{K-1})A_0\Delta_{K-1}(\mathbf{X})\Delta_A^{\mathrm{T}}(\mathbf{X}) & \cdots & (A_{K-1}-\pi_{K-1})A_{K-1}\Delta_A^{\mathrm{T}}(\mathbf{X}) \end{bmatrix} \right\}$$

$$= \begin{bmatrix} \pi_0(1-\pi_0)\Delta_0(\mathbf{X})\Delta_0^{\mathrm{T}}(\mathbf{X}) & \cdots & -\pi_0\pi_{K-1}\Delta_0(\mathbf{X})\Delta_{K-1}^{\mathrm{T}}(\mathbf{X}) \\ -\pi_1\pi_0\Delta_1(\mathbf{X})\Delta_0^{\mathrm{T}}(\mathbf{X}) & \ddots & -\pi_1\pi_{K-1}\Delta_1(\mathbf{X})\Delta_{K-1}^{\mathrm{T}}(\mathbf{X}) \\ \vdots & \ddots & \vdots \\ -\pi_{K-1}\pi_0\Delta_{K-1}(\mathbf{X})\Delta_0^{\mathrm{T}}(\mathbf{X}) & \cdots & \pi_{K-1}(1-\pi_{K-1})\Delta_{K-1}^{\mathrm{T}}(\mathbf{X}) \end{bmatrix} \qquad (\mathbf{C_2})$$

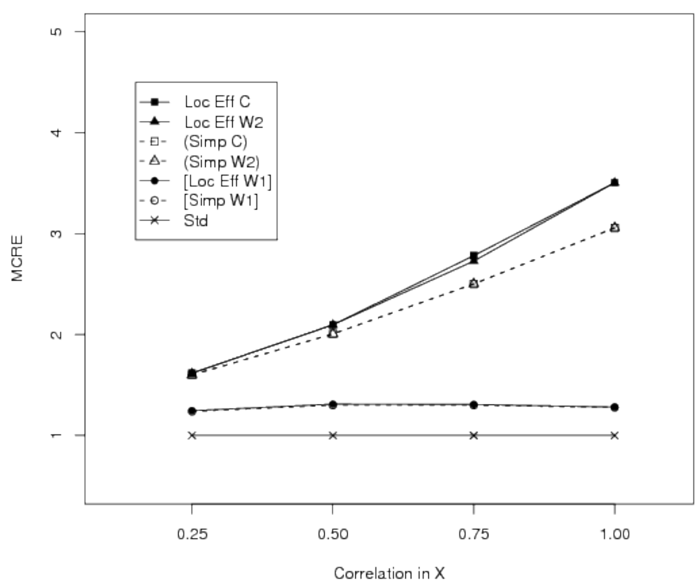
From $(\mathbf{C_2})$, we see that generally, the second term of (4) contains block diagonal terms $\pi_k(1-\pi_k)E_{\mathbf{X}}\left\{\Delta_k^{\otimes 2}(\mathbf{X})\right\}$, and block off-diagonal terms $-\pi_j\pi_k E_{\mathbf{X}}\{\Delta_j(\mathbf{X})\Delta_k^{\mathrm{T}}(\mathbf{X})\}$.

Referring back to (3), we see that $\mathbf{h_{opt}} = [\mathbf{C}_1 - \mathbf{C}_2]^{-1}\mathbf{D}^*$, as labeled above.

Appendix B: Supplementary Figures, Simulation Results

Figure 1: MCRE of Locally Efficient and Simple Augmented GEE Relative to Standard (Unaugmented) GEE. Continuous clustered outcomes. Estimators corresponding to each curve are denoted by 'Estimator-Outcome Regression' using the abbreviations: Loc Eff-Locally Efficient, Simp-Simple Augmented, Std-Standard; C-Correct, W1-Wrong 1, W2-Wrong 2. All estimators use exchangable working covariance for $V(Y|A)$ and $V\{E(Y|X,A)\}$. The order of curves in the legend follows the order of curves on the figure, with sets of superimposed curves denoted by '()' and '[]'.

(a) n_i=(2,4,6,8,10,12), $\sigma_b^2 = 0$



(b) n_i=(2,4,6,8,10,12), $\sigma_b^2 = 6$

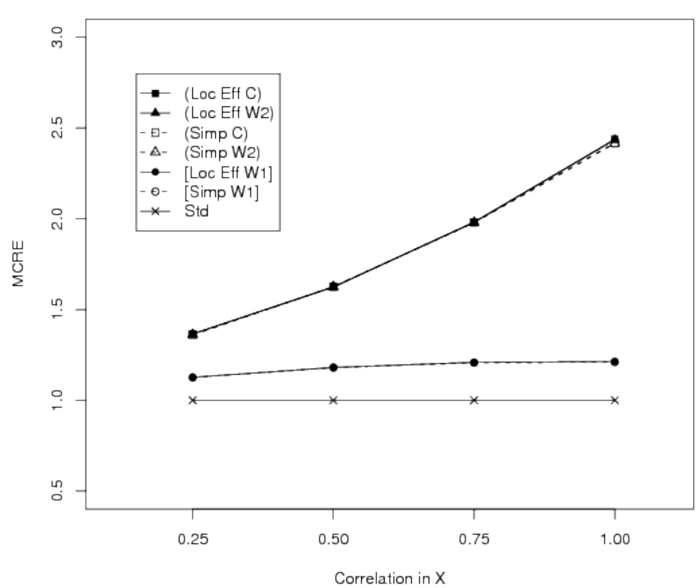
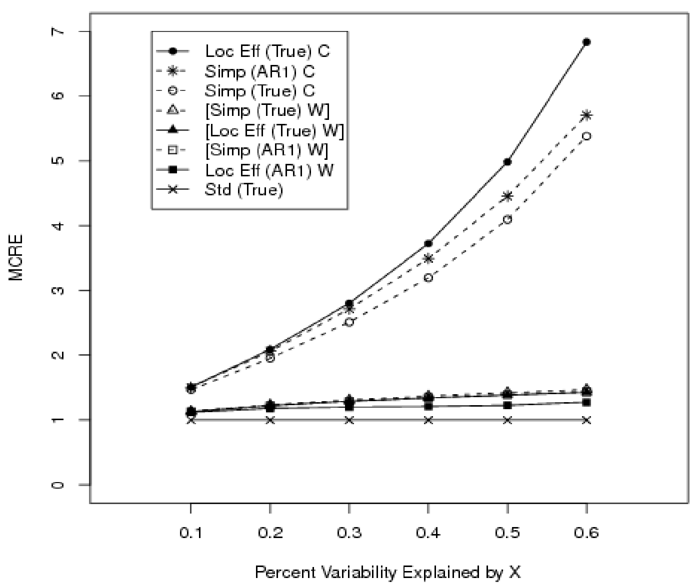


Figure 2: MCRE of Locally Efficient and Simple Augmented GEE Relative to Standard (Unaugmented) GEE. Continuous longitudinal outcomes. Estimators corresponding to each curve are denoted by 'Estimator (Marginal Working Covariance) Outcome Regression' using the abbreviations: Loc Eff-Locally Efficient, Simp-Simple Augmented, Std-Standard; AR1-Autoregressive(1) $V(Y|A)$, True-Exchangeable/AR1 for $V\{E(Y|X,A)\}$ and $V(Y|X,A)$, respectively; C-Correct, W1-Wrong 1, W2-Wrong 2;α=0.3. The order of curves in the legend follows the order of curves on the figure, with the set of superimposed curves denoted by '[]' and '{}'.

(a) α=0.1



(b) α=0.5

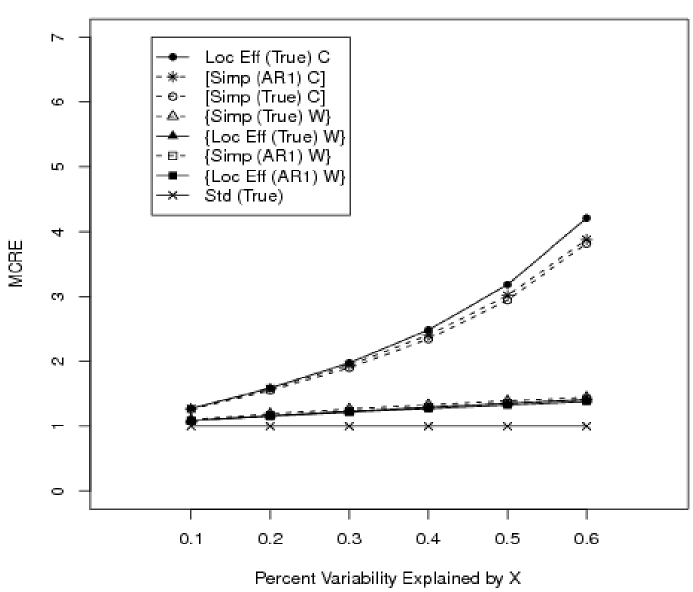
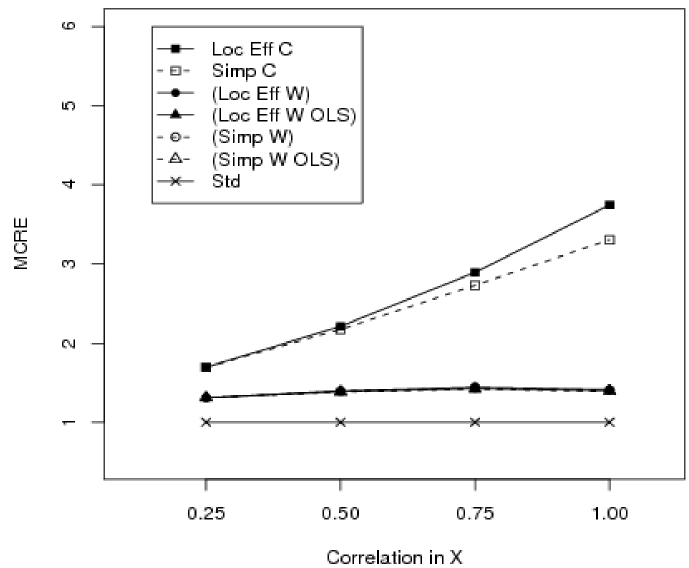
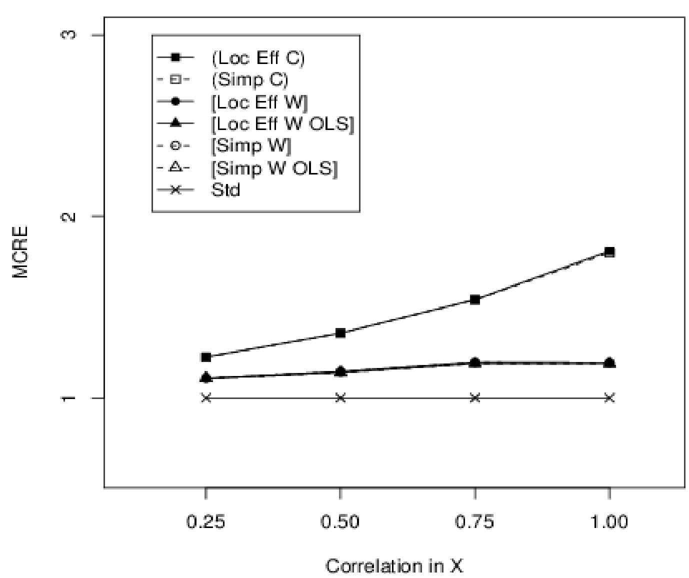


Figure 3: MCRE of Locally Efficient and Simple Augmented GEE Relative to Standard (Unaugmented) GEE. Binary clustered outcomes. Estimators corresponding to each curve are denoted by 'Estimator-Outcome Regression' using the abbreviations: Loc Eff-Locally Efficient, Simp-Simple Augmented, Std-Standard; C-Correct, W1-Wrong 1, W2-Wrong 2. All estimators use exchangable working covariance for $V(Y|A)$ and $V\{E(Y|X,A)\}$. The order of curves in the legend follows the order of curves on the figure, with sets of superimposed curves denoted by '()'.

(a) n_i=(2,4,6,8,10,12), $\theta = 1$



(b) n_i=(2,4,6,8,10,12), $\theta = 0.8$



Appendix C: QIC for selecting working covariance structures

Table 1: QIC for selecting working covariance structures. Conditional model: $E(CD4_{ij}|Trt_i, Week_{ij}, \mathbf{X}_i) = \eta_0 + \eta_1 A_i + \eta_2 Week_{ij} + \eta_3 Sex_i + \eta_4 CD4_{0_i}$. Marginal model: $E(CD4_{ij}|Trt_i, Week_{ij}) = \eta_0 + \eta_1 Trt_i + \eta_2 Week_{ij}$

Conditional Model

Working Covariance Structure	QIC
Independence	1053.44
Exchangeable	1051.9
AR1	1052.29
Unstructured	1049.72

Marginal Model

Working Covariance Structure	QIC
Independence	1047.59
Exchangeable	1047.1
AR1	1046.56
Unstructured	1049.35