STAT/BIOST 571: Homework 3

Philip Pham

February 10, 2019

Problem 1: Quasilikelihood and semiparametric methods for the general linear model (14 points)

This question examines the effect of different correlation structures, designs, and sample sizes in fitting a general linear model in a quasi-likelihood and semiparametric framework. It is also an exercise in writing code systematically; please take care to break the required programming into small tasks, and write individual functions to do each of these tasks. Please write all code "by hand", using matrix algebra and simple moment-based estimators. You may find the `mvtnorm` *package helpful.*

For the marginal model

$$E(Y_{ij}|x_{ij}) = \beta_0 + \beta_1 x_{ij},$$

consider estimation by weighted least squares, where the cluster weights are the inverse of the estimated cluster covariance matrix. Calculate quasi-likelihood standard errors as if your assumed form of the covariance matrix is known to be correct (even if, in actuality, you have assumed the wrong form of the covariance) and semi-parametric standard errors using the sandwich estimator that accounts for clustering. All of the notation follows the lecture notes.

Throughout, the following are true in the data-generating mechanism

- $\beta_0 = 0$, $\beta_1 = 0.5$
- $\mathbf{Y}_i|\boldsymbol{X}_i \sim N(\boldsymbol{X}_i\boldsymbol{\beta}, \sigma^2\boldsymbol{R}_i)$ *with* $\sigma^2 = 1$.

The factors that will vary are

- *The number of clusters is 15, 30, or 60*
- *The design:*
  - *Design I has* $m_i = 4$, *for all clusters. In each cluster, we see* $\{x_{i1}, x_{i2}, x_{i3}, x_{i4}\} = \{7, 10, 13, 16\}$
  - *Design II has* $m_i = 3$ *for all clusters. We see equal numbers of clusters with* $\{x_{i1}, x_{i2}, x_{i3}\} = \{7, 10, 13\}, \{7, 10, 16\}, \{7, 13, 16\},$ *or* $\{10, 13, 16\}$
- *The true covariance and the assumed covariance matrices are of the form* $\sigma^2\boldsymbol{R}_i$*:*

			$\mathbf{SD}(\hat{\beta}_1)$			$\mathbf{E}(\widehat{\mathbf{SE}}_{\mathbf{1,QL}})$			$\mathbf{E}(\widehat{\mathbf{SE}}_{\mathbf{1,sand}})$		
			Assumed Corr			Assumed Corr			Assumed Corr		
n	Design	True Corr	Uncor	Exch	Expon	Uncor	Exch	Expon	Uncor	Exch	Expon
15	I	Exchangeable $\rho = 0.5$	0.03849	0.02853	0.03791	0.03768	0.02768	0.03702	0.02571	0.02571	0.02616
15	I	Exchangeable $\rho = 0.9$	0.03849	0.01469	0.02464	0.03701	0.0138	0.02324	0.01149	0.01149	0.01207
15	I	Exponential $\rho = 0.5$	0.03849	0.03177	0.03804	0.03776	0.03101	0.03726	0.03627	0.03627	0.03586
15	I	Exponential $\rho = 0.9$	0.03849	0.0174	0.02468	0.03707	0.01639	0.02334	0.02026	0.02026	0.02001
15	II	Exchangeable $\rho = 0.5$	0.04464	0.03459	0.04015	0.04401	0.03372	0.03934	0.03493	0.03174	0.03213
15	II	Exchangeable $\rho = 0.9$	0.04464	0.0191	0.02488	0.04334	0.01811	0.02364	0.02609	0.01432	0.01483
15	II	Exponential $\rho = 0.5$	0.04464	0.03664	0.04026	0.04412	0.03592	0.03957	0.04029	0.03847	0.03836
15	II	Exponential $\rho = 0.9$	0.04464	0.02075	0.02501	0.04339	0.01969	0.02379	0.02877	0.01975	0.01953
30	I	Exchangeable $\rho = 0.5$	0.02722	0.01968	0.02682	0.0269	0.01936	0.02647	0.01857	0.01857	0.01897
30	I	Exchangeable $\rho = 0.9$	0.02722	0.00949	0.01624	0.02669	0.0092	0.01577	0.00831	0.00831	0.00876
30	I	Exponential $\rho = 0.5$	0.02722	0.02214	0.02684	0.02692	0.02184	0.02651	0.02621	0.02621	0.02591
30	I	Exponential $\rho = 0.9$	0.02722	0.01153	0.01618	0.02671	0.0112	0.01571	0.01464	0.01464	0.01446
30	II	Exchangeable $\rho = 0.5$	0.03148	0.02376	0.02816	0.03134	0.0235	0.02793	0.02525	0.02283	0.02316
30	II	Exchangeable $\rho = 0.9$	0.03148	0.01201	0.01599	0.03115	0.01172	0.01562	0.01905	0.01023	0.01057
30	II	Exponential $\rho = 0.5$	0.03148	0.02534	0.02813	0.03138	0.02514	0.02795	0.02906	0.02776	0.02766
30	II	Exponential $\rho = 0.9$	0.03148	0.01331	0.01602	0.03117	0.013	0.01565	0.02107	0.01422	0.01404
60	I	Exchangeable $\rho = 0.5$	0.01925	0.01374	0.01894	0.01914	0.01363	0.01881	0.01336	0.01336	0.01369
60	I	Exchangeable $\rho = 0.9$	0.01925	0.00639	0.01102	0.01908	0.00629	0.01086	0.00597	0.00597	0.00629
60	I	Exponential $\rho = 0.5$	0.01925	0.01554	0.01893	0.01913	0.01543	0.01881	0.0188	0.0188	0.0186
60	I	Exponential $\rho = 0.9$	0.01925	0.00789	0.01096	0.01908	0.00778	0.0108	0.0105	0.0105	0.01038
60	II	Exchangeable $\rho = 0.5$	0.02225	0.01663	0.01986	0.02222	0.01656	0.0198	0.01801	0.01636	0.01657
60	II	Exchangeable $\rho = 0.9$	0.02225	0.00799	0.01073	0.02214	0.0079	0.01061	0.01376	0.00736	0.0076
60	II	Exponential $\rho = 0.5$	0.02225	0.01782	0.01985	0.02223	0.01776	0.0198	0.02063	0.01981	0.01972
60	II	Exponential $\rho = 0.9$	0.02225	0.00897	0.01072	0.02215	0.00887	0.0106	0.01504	0.01015	0.01002

Table 1: Table of standard deviations of the estimated slopes and average of model-based and sandwich-based standard error estimates

- *For the true covariance, consider exchangeable and exponential correlation structures, with $\rho = 0.5$ or $\rho = 0.9$ (distances between observations in the exponential model based on x_{ij}).*
- *For the assumed covariance, consider these and additionally the uncorrelated homoscedastic covariance. Any covariance parameters should be estimated using moment-based methods.*

Solution: See the results in Table 1. Correct standard errors are found when the assumed correlation structure agrees with the true correlation structure when using GLS (columns 2 and 3). Ignoring the correlation within the clusters overestimates the standard error (first column). When incorrectly assuming exponential correlation, the standard error is overestimated. When incorrectly assuming exchangeable correlation, the standard error is underestimated.

Quasi-likelihood doesn't help very much when the correlation structure is misspecified. The standard error estimates are almost identical to just using GLS. The standard errors are slightly closer to those of the sandwich estimator, so there is some insignificant gain.

Comparing the sandwich estimation with GLS when the assumed correlation is correct, one sees that the standard error estimates are underestimated for smaller n. For $n = 60$, the estimates are very good. When the correlation structure is misspecified, sandwich estimation comes closest to the actual the standard error, especially when n is large (verified numerically, results not shown). It follows that misspecifying the correlation structure as exchangeable or exponential doesn't affect the standard error of $\hat{\beta}_1$ very much.

This is also the case when assuming no correlation with design I. However, when using design II, where not all the clusters have identical covariates, assuming no correlation increases the standard error of our estimator significantly. This effect is even more pronounced when there is more correlation ($\rho = 0.9$ versus $\rho = 0.5$).

Details of the calculation are described below, and code is in the appendix.

Let X_i be the covariates for each cluster with a column of 1s prepended. Let y_i be the cluster responses.

We first estimate β with the ordinary least squares (OLS) estimator, $\hat{\beta}_{\text{OLS}} = \left(X^\intercal X\right)^{-1} X^\intercal y$, where X is the concatenation of the cluster covariates and y is the concatenation of the cluster responses.

We can get the covariances for $\hat{\beta}$ with

$$\operatorname{cov}\left(\hat{\beta}\right) = \left(\sum_{i=1}^{n} X_i^\intercal W X_i\right)^{-1} \left(\sum_{i=1}^{n} X_i^\intercal W \hat{\Sigma} W X_i\right) \left(\sum_{i=1}^{n} X_i^\intercal W X_i\right)^{-1}.$$

For the different estimation methods and assumed covariances, we vary the form of $\hat{\Sigma}$ and W, both of which are assumed to be the same for all clusters.

To obtain W, we first estimate σ^2 with

$$\hat{\sigma}^2 = \frac{1}{\sum_{i=1}^{n} m_i - 2} \sum_{i=1}^{n} \left(y_i - X_i \hat{\beta}_{\text{OLS}}\right)^\intercal \left(y_i - X_i \hat{\beta}_{\text{OLS}}\right). \tag{1}$$

Let the standardized OLS residuals for each cluster be $\tilde{\epsilon}_i = \frac{y_i - X_i \hat{\beta}_{\text{OLS}}}{\hat{\sigma}}$.

If we assume that the correlation structure is exchangeable, we estimate

$$\hat{\rho}_{\text{exch}} = \frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{m_i - 1} j} \sum_{i=1}^{n} \sum_{j=1}^{m_i - 1} \sum_{k=j+1}^{m_i} \left(\tilde{\epsilon}_i \tilde{\epsilon}_i^\intercal\right)_{jk}, \tag{2}$$

so $W_{ii}^{-1} = 1$ for all i and $W_{ij}^{-1} = \hat{\rho}_{\text{exch}}$ for all $i \neq j$.

If we assume that the correlation structure is exponential, we estimate

$$\hat{\rho}_{\text{expon}} = \frac{1}{\sum_{i=1}^{n} \sum_{j=1}^{m_i - 1} j} \sum_{i=1}^{n} \sum_{j=1}^{m_i - 1} \left(\tilde{\epsilon}_i \tilde{\epsilon}_i^\intercal\right)_{j,j+1}, \tag{3}$$

so $W_{ij}^{-1} = \hat{\rho}_{\text{expon}}^{|j-i|}$ for any i and j.

If we don't assume any correlation structure, $W = I$.

Now that we know W, we can estimate β with

$$\hat{\beta} = \frac{1}{\sum_{i=1}^{n} m_i} \sum_{i=1}^{n} m_i \left(X_i^\intercal W X_i\right)^{-1} X_i^\intercal W y_i. \tag{4}$$

By default, for the general linear model, we would have that $\hat{\Sigma} = W^{-1}$, in which case, we have that

$$\operatorname{cov}\left(\hat{\beta}\right) = \left(\sum_{i=1}^{n} X_i^\intercal W X_i\right)^{-1}. \tag{5}$$

In the quasi-likelihood model, we have an additional dispersion factor α, so $\hat{\Sigma} = \hat{\alpha} W^{-1}$, where

$$\hat{\alpha} = \frac{1}{\sum_{i=1}^{n} m_i - 2} \left(y - X\hat{\beta} \right)^{\mathsf{T}} \left(y - X\hat{\beta} \right), \tag{6}$$

which results in the covariance

$$\operatorname{cov}\left(\hat{\beta}\right) = \hat{\alpha} \left(\sum_{i=1}^{n} X_i^{\mathsf{T}} W X_i \right)^{-1}. \tag{7}$$

For the sandwich estimator, we use the empirical estimate

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - X_i\hat{\beta} \right)^{\mathsf{T}} \left(y_i - X_i\hat{\beta} \right). \tag{8}$$

Problem 2: Efficiency of OLS for linear models with correlated data (6 points)

Review the example on slides 2.34 – 2.35, which can also be found on pages 60 – 62 of the Diggle et al. textbook. We will generalize this example by considering the mean model

$$E(Y_{ij}) = \beta_0 + \beta_1 x_j$$

for arbitrary $\boldsymbol{x} = (x_1, x_2, \ldots, x_5)$ *that is the same for all subjects, but which may or may not be equal to* $\boldsymbol{t} = (-2, -1, 0, 1, 2)$ *(as is the case in the original version of the example). Note that the correlation structure is still determined based on t, as in the original example, but now the mean model contains x rather than t.*

(a) *Derive a general expression for the relative efficiency of OLS compared to the optimal GLS in estimating β_0 and β_1 in this problem. Your formula should be valid for a homoscedastic exponential covariance matrix with arbitrary ρ and for arbitrary $\boldsymbol{x}$. That is, derive a general version of the expressions on the bottom of 2.44 that is valid for any choice of covariates. Note that it is acceptable for your solution to be written using matrix notation and matrix algebra.*

Solution: Let there be n subjects. Each subject i has m_i observations $x_i = \begin{pmatrix} x_{i1} & x_{i2} & \ldots & x_{im_i} \end{pmatrix}^{\mathsf{T}}$. Let X_i be the $m_i \times 2$ covariate matrix, where $X_{ij1} = 1$ and $X_{ij2} = x_{ij}$. Let X be the $\sum_{i=1}^{n} m_i \times 2$ matrix obtained by stacking $X_1, X_2, \ldots, X_n$.

The true model is $Y_{ij} = \beta_0 + \beta_1 x_{ij} + \epsilon_{ij}$. Let the obvervations between subjects be independent with each other. Let Σ_i denote the covariance structure for within-subject observations, that is $\Sigma_{ijk} = \operatorname{cov}(\epsilon_{ij}, \epsilon_{ik})$. Let Σ be the $\sum_{i=1}^{n} m_i \times \sum_{i=1}^{n} m_i$ block diagonal matrix

$$\Sigma = \begin{pmatrix} \Sigma_1 & & & \\ & \Sigma_2 & & \\ & & \ddots & \\ & & & \Sigma_n \end{pmatrix}. \tag{9}$$

Let Y_i be the vector of observations for subject i. Let Y be obtained by concatenating the $Y_1, Y_2, \ldots, Y_n$. If $\beta = \begin{pmatrix} \beta_0 & \beta_1 \end{pmatrix}^\intercal$, we can write $Y = X\beta + \epsilon$, where $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma)$.
The OLS estimate for is

$$\hat{\beta}_{\text{OLS}} = \left(X^\intercal X\right)^{-1} X^\intercal Y = \beta + \left(X^\intercal X\right)^{-1} X^\intercal \epsilon, \tag{10}$$

which has covariance

$$\begin{aligned} \operatorname{cov}\left(\hat{\beta}_{\text{OLS}}\right) &= \left(X^\intercal X\right)^{-1} X^\intercal \operatorname{cov}(\epsilon) X \left(X^\intercal X\right)^{-1} \\ &= \left(X^\intercal X\right)^{-1} X^\intercal \Sigma X \left(X^\intercal X\right)^{-1}. \end{aligned} \tag{11}$$

The optimal GLS estimate, which can be derived by maximizing likelihood is

$$\hat{\beta}_{\text{GLS}} = \left(X^\intercal \Sigma^{-1} X\right)^{-1} X^\intercal \Sigma^{-1} Y = \beta + \left(X^\intercal \Sigma^{-1} X\right)^{-1} X^\intercal \Sigma^{-1} \epsilon, \tag{12}$$

which has covariance

$$\begin{aligned} \operatorname{cov}\left(\hat{\beta}_{\text{GLS}}\right) &= \left(X^\intercal \Sigma^{-1} X\right)^{-1} X^\intercal \Sigma^{-1} \operatorname{cov}(\epsilon) \Sigma^{-1} X \left(X^\intercal \Sigma^{-1} X\right)^{-1} \\ &= \left(X^\intercal \Sigma^{-1} X\right)^{-1}. \end{aligned} \tag{13}$$

Much of these matrix multiplications can be written as summations:

$$\begin{aligned} X^\intercal X &= \sum_{i=1}^n X_i^\intercal X_i \\ X^\intercal \Sigma X &= \sum_{i=1}^n X_i^\intercal \Sigma_i X_i \\ X^\intercal \Sigma^{-1} X &= \sum_{i=1}^n X_i^\intercal \Sigma_i^{-1} X_{i}., \end{aligned}$$

which is probably computationally faster if n is very large.
In any case, using Equations 11 and 13, we can calculate relative efficiency

$$\begin{aligned} e\left(\hat{\beta}_0\right) &= \left[\left(X^\intercal \Sigma^{-1} X\right)^{-1}\right]_{11} \Big/ \left[\left(X^\intercal X\right)^{-1} X^\intercal \Sigma X \left(X^\intercal X\right)^{-1}\right]_{11} \\ e\left(\hat{\beta}_1\right) &= \left[\left(X^\intercal \Sigma^{-1} X\right)^{-1}\right]_{22} \Big/ \left[\left(X^\intercal X\right)^{-1} X^\intercal \Sigma X \left(X^\intercal X\right)^{-1}\right]_{22} \end{aligned}$$

by taking the corresponding entries of the covariance matrix.

(b) *Reproduce the lines in the table on 2.35 that pertain to β_1 for the following choices of covariate vectors*

$$\begin{aligned} \boldsymbol{x} &= (-2, -1, 0, 1, 2) \\ \boldsymbol{x} &= (-1, -2, 0, 2, 1) \\ \boldsymbol{x} &= (0, -1, 1, 3, 2) \\ \boldsymbol{x} &= (0, -1, 1, 5, 2) \end{aligned}$$

	ρ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.99
x	Value										
(-2, -1, 0, 1, 2)	$e(\hat{\beta}_0)$	0.9978	0.9917	0.983	0.9729	0.9631	0.9554	0.9521	0.9558	0.9701	0.9961
	$e(\hat{\beta}_1)$	0.9969	0.9893	0.9797	0.97	0.9615	0.9554	0.9522	0.9522	0.9554	0.9608
(-1, -2, 0, 2, 1)	$e(\hat{\beta}_0)$	0.9978	0.9917	0.983	0.9729	0.9631	0.9554	0.9521	0.9558	0.9701	0.9961
	$e(\hat{\beta}_1)$	0.9959	0.9818	0.9554	0.9154	0.8621	0.7974	0.7249	0.6486	0.5726	0.507
(0, -1, 1, 3, 2)	$e(\hat{\beta}_0)$	0.9972	0.9888	0.9758	0.9596	0.9432	0.9302	0.9247	0.931	0.9541	0.9942
	$e(\hat{\beta}_1)$	0.9959	0.9818	0.9554	0.9154	0.8621	0.7974	0.7249	0.6486	0.5726	0.507
(0, -1, 1, 5, 2)	$e(\hat{\beta}_0)$	0.9949	0.9817	0.9636	0.9445	0.9281	0.9178	0.9169	0.9279	0.9541	0.9943
	$e(\hat{\beta}_1)$	0.9911	0.9644	0.9206	0.8626	0.794	0.7194	0.6432	0.5689	0.4992	0.4416

Table 2: Efficiency results comparing OLS with GLS.

Solution: Results can be seen Table 2. Code is in the appendix.

(c) *Explain the key differences between the relative efficiencies you just calculated. Phrase your answers in a manner that will be understandable by a quantitavely sophisticated non-statistician (e.g., an epidemiologist collaborator).*

Solution: If there's not much correlation, i.e. small ρ, there's not much efficiency gain in using GLS over OLS.

For β_1, the efficiency gain does depend on the observed covariates, however. Recall that the estimated for β_1 can be written as a sum of weighted pairwise slopes. The weight is higher for for pairs where x_{ij_1} and x_{ij_2} are further part because the effect size becomes bigger relative to the noise. Similarly, when the errors are positively correlated, more of the variance is explained away, so we are more confident about the slope estimate β_1.

These two factors interact multiplicatively. Thus, when

$$\mathbf{x} \in \left\{(-1,-2,0,2,1),(0,-1,1,3,2),(0,-1,1,5,2)\right\},$$

there are pairs of observations where the difference in covariates is large and the correlation is high. GLS is able to leverage this to produce a more efficient estimator. In particular when $\mathbf{x} = (0,-1,1,5,2)$, there are neighboring observations where the difference in covariates can be quite large, leading to the relatively most efficient estimator when using GLS.