

Upgraded quadratic inference functions for longitudinal data with type II time-dependent covariates

Gyo-Young Cho¹ · Oyunchimeg Dashnyam²

¹Department of Statistics, Kyungpook National University

²Department of Mathematics and Statistics, Mongolian State University of Education

Received 15 October 2013, revised 19 November 2013, accepted 12 December 2013

Abstract

Qu *et al.* (2000) proposed the quadratic inference functions (QIF) method to marginal model analysis of longitudinal data to improve the generalized estimating equations (GEE). It yields a substantial improvement in efficiency for the estimators of regression parameters when the working correlation is misspecified. But for the longitudinal data with time-dependent covariates, when the implicit full covariates conditional mean (FCCM) assumption is violated, the QIF can not provide more consistent and efficient estimator than GEE (Cho and Dashnyam, 2013). Lai and Small (2007) divided time-dependent covariates into three types and proposed generalized method of moment (GMM) for longitudinal data with time-dependent covariates. They showed that their GMM type II and GMM moment selection methods can be more efficient than GEE with independence working correlation (GEE-ind) in the case of type II time-dependent covariates. We develop upgraded QIF method for type II time-dependent covariates. We show that this upgraded QIF method can provide substantial gains in efficiency over QIF and GEE-ind in the case of type II time-dependent covariates.

Keywords: FCCM assumption, GEE, GMM, longitudinal data, marginal model, QIF, time-dependent covariates.

1. Introduction

Marginal models for longitudinal data primarily aim to estimate the population-average effects of covariates on the response of interest. The celebrated method of statistical inference among marginal models is the GEE approach proposed by Liang and Zeger (1986).

When there are time-dependent covariates, Pepe and Anderson (1994) have pointed out that the consistency of GEE estimator is not assured with arbitrary working correlation structures unless an implicit FCCM assumption is satisfied. Thus, they suggested that when we use the GEE to do marginal regression, either a diagonal working correlation matrix should be used, or FCCM assumption needs to be validated. Using working independence guarantees consistency but entails a serious loss of efficiency in many cases (Fitzmaurice *et al.*, 2004).

¹ Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea.

² Corresponding author: Lecturer, Department of Mathematics and Statistics, Mongolian State University of Education, Ulaanbaatar 210648, Mongolia. E-mail: d_louynaa@yahoo.com

To resolve this problem, Pan and Connett (2002) proposed to use resampling-based methods to choose the estimator which best predicts Y_{it} based on the covariates X_{it} among GEE estimators based on different working correlation structures.

Qu *et. al.* (2000) proposed the QIF approach which is important and powerful alternative to the GEE method. The QIF is derived by approximating the inverse of the working correlation matrix by a linear combination of several basis matrices composed of 0s and 1s:

$$R^{-1} = \sum_{i=1}^m a_i M_i \quad (1.1)$$

where M_1, \dots, M_m are known matrices and a_1, \dots, a_m are unknown constants.

Substituting (1.1) into GEE and consider the following class of estimating functions:

$$S_\beta(\beta, W) = \sum_{i=1}^N \left(\frac{\partial \mu_i}{\partial \beta} \right)' A_i^{-\frac{1}{2}} (a_1 M_1 + \dots + a_m M_m) A_i^{-\frac{1}{2}} (Y_i - \mu_i) \quad (1.2)$$

From here, the 'extended score' g_N is defined as

$$g_N(\beta) = \frac{1}{N} \sum_{i=1}^N g_i(\beta) = \frac{1}{N} \begin{pmatrix} \sum_{i=1}^N \left(\frac{\partial \mu_i}{\partial \beta} \right)' A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} (Y_i - \mu_i) \\ \vdots \\ \sum_{i=1}^N \left(\frac{\partial \mu_i}{\partial \beta} \right)' A_i^{-\frac{1}{2}} M_m A_i^{-\frac{1}{2}} (Y_i - \mu_i) \end{pmatrix}. \quad (1.3)$$

Since the vector g_N contains more estimating equations than unknown parameters, GMM is used to create the QIF, given as

$$Q_N(\beta) = g_N' C_N^{-1} g_N \quad (1.4)$$

where

$$C_N = \frac{1}{N^2} \sum_{i=1}^N g_i(\beta) g_i'(\beta). \quad (1.5)$$

The quadratic inference function estimator $\hat{\beta}$ is then defined to be

$$\hat{\beta} = \arg \min_{\beta} Q_N(\beta). \quad (1.6)$$

The QIF estimator is obtained with no need to estimate the nuisance correlation parameter. Hence, the QIF method does not rely on whether an appropriate estimation of the correlation parameter is available or not.

The QIF has several useful properties over the GEE. It does not require more assumptions than does the GEE method, but it yields a substantial improvement in efficiency for the estimators of regression parameters when the working correlation is misspecified, and same efficiency to the GEE when the working correlation is correct. However for the marginal

regression model for time-varying stochastic covariates, specifically when the FCCM assumption is violated, the QIF does not provide consistent, asymptotically normal and efficient estimators over GEE-ind estimator. So when we use the QIF in marginal regression for longitudinal data with time-varying stochastic covariates, either FCCM assumption needs to be validated or QIF with independence working correlation matrix (QIF-ind) should be used (Cho and Dashnyam, 2013).

Lai and Small (2007) showed that there are available valid estimating equations that are not exploited by the usual GEE-ind estimator. They proposed the GMM for the marginal regression analysis of longitudinal data with time-dependent covariates which use these additional estimating equations. Lai and Small (2007) classified time-dependent covariates into three types based on GEE such as type I, II and III and showed that their methods provide more efficient estimator than GEE-ind for type II time-dependent covariates and same efficiency for type III time-dependent covariates.

2. Upgraded quadratic inference functions for type II time-dependent covariates

Cho and Dashnyam (2013) concluded that the GMM type II provides better result than QIF/GEE-ind. This is because the GMM type II method uses more valid moment conditions. Using more valid moment conditions leads to more efficient estimator. On the other hand, the QIF with non independence working correlation cannot be consistent and asymptotically normal because they use some invalid moment conditions. The QIF-ind guarantees unbiasedness and consistency, but it cannot be efficient because it removed a large number of estimating functions. So we sought the way to improve efficiency of QIF estimators.

According to Lai and Small (2007), for a type II time-dependent covariates, there are $T(T+1)/2$ valid moment conditions, so $T(T+1)/2 - T$ valid moment conditions are not used in GEE-ind. To use these valid estimating functions, we proposed the upgraded QIF-AR(1) and upgraded QIF-exch methods for type II time-dependent covariates as changing second basis matrix related to QIF-AR(1) and QIF-exch. The idea is borrowed from GMM type II method of Lai and Small (2007).

This is made as 0s when corresponding estimating function is invalid and 1s when corresponding estimating function is valid in the M_1 matrix related to AR(1) and exchangeable working correlation. Then invalid estimating functions can drop from moment vectors in QIF. By doing so, M_1 becomes linear combination of the valid estimating functions. This means that expectation of the extended score vector can be zero so that upgraded QIF is unbiased, consistent and asymptotically normal. Also, the QIF can use more correct estimating functions than GEE/QIF-ind so that using the upgraded QIF-AR(1) and QIF-exch methods can improve the efficiency of GEE/QIF-ind in the case of type II time-dependent covariates.

While type I covariates are intuitively easier to identify, the difference between a type II and III covariate may be harder to determine without a formal test. If we use upgraded QIF method for type II in the case of type III, the estimator will be inconsistent because it uses invalid estimating functions. Thus, it is important to determine the type of covariates correctly. So if we do not know the type of covariates, we will use the test provided by Lai and Small (2007) for whether a time-dependent covariate is of a certain type.

3. Comparison upgraded quadratic inference functions with quadratic inference function

Extended score vectors for the QIF-exch and QIF-AR(1) can be rewritten as follows respectively

$$g_N(\beta) = \frac{1}{N} \begin{pmatrix} \sum_{i=1}^N \left(\sum_{t=1}^{T'} X_{is} \frac{\sigma_{is}}{\sigma_{it}} (Y_{it} - \mu_{it}) \right) \\ \sum_{i=1}^N \left(\underbrace{\sum_{t=1}^{T-1} \sum_{t < s} X_{is} \frac{\sigma_{is}}{\sigma_{it}} (Y_{it} - \mu_{it})}_{s.1} + \underbrace{\sum_{t=2}^T \sum_{t > s} X_{is} \frac{\sigma_{is}}{\sigma_{it}} (Y_{it} - \mu_{it})}_{s.2} \right) \end{pmatrix}$$

and

$$g_N(\beta) = \frac{1}{N} \begin{pmatrix} \sum_{i=1}^N \left(\sum_{t=1}^{T'} X_{is} \frac{\sigma_{is}}{\sigma_{it}} (Y_{it} - \mu_{it}) \right) \\ \sum_{i=1}^N \left(\underbrace{\sum_{t=1}^{T-1} \sum_{\substack{t < s \\ s-t=1}} X_{is} \frac{\sigma_{is}}{\sigma_{it}} (Y_{it} - \mu_{it})}_{s.1} + \underbrace{\sum_{t=2}^T \sum_{\substack{t > s \\ t-s=1}} X_{is} \frac{\sigma_{is}}{\sigma_{it}} (Y_{it} - \mu_{it})}_{s.2} \right) \end{pmatrix}$$

For the type II time-dependent covariate, the expectation of first elements of these vectors are zero. Also the expectation of sums (s.1) of the second elements can be zero according to the definition of type II time-dependent covariate. But the expectation of sums (s.2) can not be zero so it will cause bias in estimates.

Extended score vectors of the upgraded QIF-exch and QIF-AR(1) are as follows respectively

$$g_N(\beta) = \frac{1}{N} \begin{pmatrix} \sum_{i=1}^N \left(\sum_{t=1}^{T'} X_{is} \frac{\sigma_{is}}{\sigma_{it}} (Y_{it} - \mu_{it}) \right) \\ \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{t < s} X_{is} \frac{\sigma_{is}}{\sigma_{it}} (Y_{it} - \mu_{it}) \end{pmatrix}$$

and

$$g_N(\beta) = \frac{1}{N} \begin{pmatrix} \sum_{i=1}^N \left(\sum_{t=1}^{T'} X_{is} \frac{\sigma_{is}}{\sigma_{it}} (Y_{it} - \mu_{it}) \right) \\ \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{\substack{t < s \\ s-t=1}} X_{is} \frac{\sigma_{is}}{\sigma_{it}} (Y_{it} - \mu_{it}) \end{pmatrix}$$

In the case of the upgraded QIF methods, excluding sums (s.2) from the second elements, the expectation of these will be zero so that $E(g_N(\beta)) = 0$. So upgraded QIF-exch and QIF-AR(1) estimators can be unbiased, consistent and asymptotically normal. Using more correct estimating functions than GEE/QIF-ind, they can provide more efficient estimators in the case of type II time-dependent covariates.

4. Simulation study

To examine the performance of our proposed upgraded QIF methods for type II time-dependent covariates, we performed a simulation study using the following model

$$Y_{it} = \gamma_0 + \gamma_1 X_{it} + \gamma_2 X_{it-1} + b_i + \epsilon_{it}$$

$$X_{it} = \rho X_{it-1} + e_{it}$$

where $b_i, e_{it}, \epsilon_{it}$ are mutually independent, $b_i \sim N(0, 1)$, $\epsilon_{it} \sim N(0, 1)$ and $e_{it} \sim N(0, 1)$.

At first, let X_{it} be a stationary process as $X_{it} \sim N(0, \sigma_e^2 / (1 - \rho^2))$ for $t = 0, \dots, T$. In the second, we consider this model with standardized AR(1) covariate, specifically $X_{it} \sim N(0, 1)$ for $t = 0, \dots, T$.

This model was considered by Lai and Small (2007) and Diggle *et. al.* (2002) and has a time-dependent covariate and a response variable that depends on both current and lagged values of the covariate.

We consider three estimators: 1) GEE or QIF with independence working correlation (GEE-ind/QIF-ind), 2) upgraded QIF with AR(1) working correlation (upgraded QIF-AR(1)), 3) upgraded QIF with exchangeable working correlation (QIF-exch).

We investigated several design features for this model and simulated 1000 data sets each of which contained $n=50, 100$ subject measures, 5, 10 time occasions and $\rho = 0.3, 0.5, 0.8$. We calculate bias of parameters $\hat{\beta}_1$ and mean squared error (MSE) of the estimators which is defined as

$$MSE = E \left(\hat{\beta}_1 - \beta_1 \right)^2 .$$

The simulated relative efficiency (SRE) of QIF with respect to GEE-ind is defined as the ratio of the MSE of GEE-ind/QIF-ind to that of the upgraded QIF estimators:

$$SRE = \frac{MSE_{GEE-ind}}{MSE_{QIF}}$$

The simulation result is shown in the following tables.

Table 4.1 Bias, MSE and SRE of the upgraded QIF estimators for type II stationary AR(1) time-dependent covariate ($\rho = 0.3$)

| Working correlation | | $n=50$ | | $n=100$ | |
|---------------------|-------------------------|---------|---------|---------|---------|
| | | $t=5$ | $t=10$ | $t=5$ | $t=10$ |
| GEE-ind/QIF-ind | Bias($\hat{\beta}_1$) | -0.0059 | -0.0027 | -0.0042 | -0.0061 |
| | MSE | 0.0304 | 0.0156 | 0.0154 | 0.0082 |
| | SRE | 1.00 | 1.00 | 1.00 | 1.00 |
| Upgraded QIF-AR(1) | Bias($\hat{\beta}_1$) | -0.0063 | -0.0028 | -0.0024 | -0.0047 |
| | MSE | 0.0216 | 0.0103 | 0.0103 | 0.0053 |
| | SRE | 1.4077 | 1.5082 | 1.4891 | 1.5475 |
| Upgraded QIF-exch | Bias($\hat{\beta}_1$) | -0.0072 | -0.0022 | -0.0045 | -0.0058 |
| | MSE | 0.0212 | 0.0103 | 0.0105 | 0.0053 |
| | SRE | 1.4332 | 1.5142 | 1.4679 | 1.5329 |

Table 4.2 Bias, MSE and SRE of the upgraded QIF estimators for type II stationary AR(1) time-dependent covariate ($\rho = 0.5$)

| Working correlation | | $n=50$ | | $n=100$ | |
|---------------------|-------------------------|---------|---------|---------|---------|
| | | $t=5$ | $t=10$ | $t=5$ | $t=10$ |
| GEE-ind/QIF-ind | Bias($\hat{\beta}_1$) | -0.0128 | -0.0054 | 0.0069 | -0.0066 |
| | MSE | 0.0318 | 0.0187 | 0.0158 | 0.0094 |
| | SRE | 1.00 | 1.00 | 1.00 | 1.00 |
| Upgraded QIF-AR(1) | Bias($\hat{\beta}_1$) | -0.0093 | -0.0055 | 0.0068 | -0.0052 |
| | MSE | 0.0224 | 0.0121 | 0.0112 | 0.0062 |
| | SRE | 1.4248 | 1.5581 | 1.4182 | 1.5225 |
| Upgraded QIF-exch | Bias($\hat{\beta}_1$) | -0.0113 | -0.0059 | 0.0059 | -0.0051 |
| | MSE | 0.0223 | 0.0122 | 0.0111 | 0.0061 |
| | SRE | 1.4315 | 1.5349 | 1.4323 | 1.5459 |

Table 4.3 Bias, MSE and SRE of the upgraded QIF estimators for type II stationary AR(1) time-dependent covariate ($\rho = 0.8$)

| Working correlation | | $n=50$ | | $n=100$ | |
|---------------------|-------------------------|---------|---------|---------|---------|
| | | $t=5$ | $t=10$ | $t=5$ | $t=10$ |
| GEE-ind/QIF-ind | Bias($\hat{\beta}_1$) | -0.0023 | -0.0100 | -0.0036 | -0.0037 |
| | MSE | 0.0240 | 0.0178 | 0.0119 | 0.009 |
| | SRE | 1.00 | 1.00 | 1.00 | 1.00 |
| Upgraded QIF-AR(1) | Bias($\hat{\beta}_1$) | -0.0011 | -0.0084 | -0.0032 | -0.0027 |
| | MSE | 0.0184 | 0.0128 | 0.0094 | 0.0065 |
| | SRE | 1.2993 | 1.3856 | 1.2698 | 1.3949 |
| Upgraded QIF-exch | Bias($\hat{\beta}_1$) | -0.0019 | -0.0084 | -0.0029 | -0.0033 |
| | MSE | 0.0180 | 0.0120 | 0.0088 | 0.0060 |
| | SRE | 1.3319 | 1.4810 | 1.3547 | 1.5089 |

Table 4.4 Bias, MSE and SRE of the QIF estimators for type II standardized AR(1) time-dependent covariate ($\rho = 0.3$)

| Working correlation | | $n=50$ | | $n=100$ | |
|---------------------|-------------------------|---------|---------|---------|---------|
| | | $t=5$ | $t=10$ | $t=5$ | $t=10$ |
| GEE-ind/QIF-ind | Bias($\hat{\beta}_1$) | -0.0025 | -0.0102 | -0.0042 | -0.0070 |
| | MSE | 0.0314 | 0.0176 | 0.0165 | 0.0094 |
| | SRE | 1.00 | 1.00 | 1.00 | 1.00 |
| Upgraded QIF-AR(1) | Bias($\hat{\beta}_1$) | -0.0001 | -0.0071 | -0.0027 | -0.0051 |
| | MSE | 0.0215 | 0.0120 | 0.0114 | 0.0061 |
| | SRE | 1.4619 | 1.4638 | 1.4438 | 1.5435 |
| Upgraded QIF-exch | Bias($\hat{\beta}_1$) | -0.0015 | -0.0079 | -0.0033 | -0.0057 |
| | MSE | 0.0215 | 0.0115 | 0.0114 | 0.0059 |
| | SRE | 1.4559 | 1.5251 | 1.4470 | 1.5894 |

Table 4.5 Bias, MSE and SRE of the QIF estimators for standardized AR(1) time-dependent covariate ($\rho = 0.5$)

| Working correlation | | $n=50$ | | $n=100$ | |
|---------------------|-----------------------------|---------|---------|---------|---------|
| | | $t=5$ | $t=10$ | $t=5$ | $t=10$ |
| GEE-ind/QIF-ind | Bias($\widehat{\beta}_1$) | -0.0135 | -0.0055 | 0.0084 | -0.0072 |
| | MSE | 0.0414 | 0.0246 | 0.0206 | 0.0123 |
| | SRE | 1.00 | 1.00 | 1.00 | 1.00 |
| Upgraded QIF-AR(1) | Bias($\widehat{\beta}_1$) | -0.0097 | -0.0055 | 0.0083 | -0.0054 |
| | MSE | 0.0285 | 0.0156 | 0.0143 | 0.0079 |
| | SRE | 1.4508 | 1.5807 | 1.4458 | 1.5511 |
| Upgraded QIF-exch | Bias($\widehat{\beta}_1$) | -0.0117 | -0.0060 | 0.0073 | -0.0053 |
| | MSE | 0.0287 | 0.0160 | 0.0143 | 0.0079 |
| | SRE | 1.4440 | 1.5401 | 1.4455 | 1.5532 |

Table 4.6 Bias, MSE and SRE of the QIF estimators for standardized AR(1) time-dependent covariate ($\rho = 0.8$)

| Working correlation | | $n=50$ | | $n=100$ | |
|---------------------|-----------------------------|---------|---------|---------|---------|
| | | $t=5$ | $t=10$ | $t=5$ | $t=10$ |
| GEE-ind/QIF-ind | Bias($\widehat{\beta}_1$) | -0.0006 | -0.0106 | -0.0001 | -0.0025 |
| | MSE | 0.0678 | 0.0453 | 0.0325 | 0.0217 |
| | SRE | 1.00 | 1.00 | 1.00 | 1.00 |
| Upgraded QIF-AR(1) | Bias($\widehat{\beta}_1$) | -0.0007 | -0.0082 | -0.0016 | -0.0024 |
| | MSE | 0.0505 | 0.0308 | 0.0243 | 0.0147 |
| | SRE | 1.3421 | 1.4677 | 1.3376 | 1.4781 |
| Upgraded QIF-exch | Bias($\widehat{\beta}_1$) | -0.0008 | -0.0089 | -0.0022 | -0.0022 |
| | MSE | 0.0498 | 0.0301 | 0.0236 | 0.0141 |
| | SRE | 1.3607 | 1.5055 | 1.3752 | 1.5386 |

The simulation results show that the upgraded QIF-AR(1) and upgraded QIF-exch estimators are more efficient than GEE/QIF-ind estimator. They are 40-50% more efficient than GEE-ind in average. To compare with the estimators of usual QIF method, bias of upgraded QIF methods became smaller and efficiency of upgraded QIF methods improved substantially than QIF estimators. Specifically, efficiency of upgraded QIF-exch estimator became better than upgraded QIF-AR(1) in many cases. It is because the number of valid estimating functions in upgraded QIF-exch method is more than the number of valid estimating functions in QIF-AR(1) method.

Actually, upgraded QIF-exch method can use all valid estimating functions of type II time-dependent covariates, because M1 matrix of upgraded QIF-exch consists of 1s below of diagonal. So using both I and M1 of upgraded QIF-exch matrix can be all estimating functions corresponding to type II time-dependent covariates. Thus we recommend to use upgraded QIF-exch as upgraded QIF method in the case of type II time-dependent covariates.

Thus, one may say that the number of valid estimating equations is important over true working correlation matrix for the efficiency of QIF estimators of time-dependent covariates. But this is verified only in simulation study and it is of interest to study.

5. Summary

In this paper, we proposed upgraded QIF-AR(1) and upgraded QIF-exch methods for type II time-dependent covariates. This is made as selecting out valid estimating functions from M_1 matrix related to AR(1) and exchangeable working correlation. Using only valid estimating functions, the upgraded QIF-AR(1), upgraded QIF-exch methods can provide consistent and asymptotically normal estimators in the case of type II time dependent covariates. Also the upgraded QIF-AR(1) and QIF-exch estimators can be more efficient GEE/QIF-ind estimators, because the upgraded QIF-AR(1) and upgraded QIF-exch methods use more valid estimating functions than usual QIF methods. We note that upgraded QIF-exch method works better than upgraded QIF-AR(1) in the case of type II time-dependent covariates because it can use all valid estimating functions of type II time-dependent covariates. This is verified by our simulation results.

References

- Cho, G. Y. and Dashnyam, O. (2013) Quadratic inference functions in marginal models for longitudinal data with time-varying stochastic covariates. *Journal of the Korean Data & Information Science Society*, **24**, 651-658.
- Cho, G. Y. and Dashnyam, O. (2013) Generalized methods of moments in marginal models for longitudinal data with time-dependent covariates. *Journal of the Korean Data & Information Science Society*, **24**, 1-7.
- Diggle, P. J. Heagerty, P. Liang, K-Y. and Zeger, S. L. (2002). *Analysis of longitudinal data*, Oxford University Press, New York.
- Fitzmaurice, G. M, Liard, N.M. and Ware, J.H. (2004). *Applied longitudinal analysis*, Wiley, New York.
- Lai, Tz. L. and Small, D. (2007). Marginal regression analysis of longitudinal data with time-dependent covariates: A generalized method of moments approach. *Journal of the Royal Statistical Society B*, **69**, 79-99.
- Pan, W. T and Connett, J. E. (2002). Selecting the working correlation structure in generalized estimating equations with application to the lung health study. *Statistica Sinica*, **12**, 475-490.
- Pepe, M. S. and Anderson, G. L. (1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics-Simulation*, **23**, 939-951.
- Qu, A. Lindsay, B. G. and Li, B. (2000). Improving generalized estimating equations using quadratic inference functions. *Biometrika*, **87**, 823-836.