Journal of the Korean Data & Information Science Society
2013, **24**(4), 877–883

http://dx.doi.org/10.7465/jkdi.2013.24.4.877
한국데이터정보과학회지

Generalized methods of moments in marginal models for longitudinal data with time-dependent covariates[†]

Gyo-Young Cho[1] · Oyunchimeg Dashnyam[2]

[12]Department of Statistics, Kyungpook National University

Received 17 April 2013, revised 13 May 2013, accepted 18 May 2013

Abstract

The quadratic inference functions (QIF) method proposed by Qu *et al.* (2000) and the generalized method of moments (GMM) for marginal regression analysis of longitudinal data with time-dependent covariates proposed by Lai and Small (2007) both are the methods based on generalized method of moment (GMM) introduced by Hansen (1982) and both use generalized estimating equations (GEE). Lai and Small (2007) divided time-dependent covariates into three types such as: Type I, Type II and Type III. In this paper, we compared these methods in the case of Type II and Type III in which full covariates conditional mean assumption (FCCM) is violated and interested in whether they can improve the results of GEE with independence working correlation. We show that in the marginal regression model with Type II time-dependent covariates, GMM Type II of Lai and Small (2007) provides more efficient result than QIF and for the Type III time-dependent covariates, QIF with independence working correlation and GMM Type III methods provide the same results. Our simulation study showed the same results.

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

1. Introduction

In the marginal model, there is implicit FCCM assumption that the conditional mean of the k^{th} response, given $X_{i1}, \cdots, X_{in_i}$, depends only on X_{ik}.

$$E(Y_{it} \mid X_{i1}, \cdots, X_{in_i}) = E(Y_{it} \mid X_{it}) \tag{1.1}$$

With time-stationary covariates, this assumption necessarily holds since X_{it}=X_{ik} for all occasions $k \neq t$. Also, with time-dependent covariates that are fixed by design of the study, the assumption also holds since values of the covariates at any occasion are determined a priori by study design and in a manner completely unrelated to the longitudinal response (Fitzmaurice *et al.*, 2004).

† This research was supported by Kyungpook National University Research Fund, 2012.
1 Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea.
2 Corresponding author: Ph. D student, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea. E-mail: d_ouynaa@yahoo.com

However, when a covariate is time-dependent and stochastic, this assumption may not necessarily hold. If the assumption is violated and a nondiagonal working correlation matrix is used in GEE, biased estimates of regression coefficients may result (Pan *et al.*, 2000). Thus, it is important to check the FCCM assumption made in (1.1), namely, that the conditional mean of the Y_{it}, given the entire time-dependent covariate profile $X_{i1}, \cdots, X_{in_i}$, depends only on the covariate value at the t^{th} occasion, X_{it}.

Pepe and Anderson (1994) pointed out 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. It means that they suggest using the independent working correlation matrix as a "safe" analysis choice. However the using working independence correlation matrix in GEE guarantees consistency, but entails a serious loss of efficiency in many cases (Fitzmaurice, 1995).

The QIF method proposed by Qu *et al.* (2000) is an important and powerful alternative to the GEE and it has several useful properties such as robustness, a goodness-of-fit test and model selection (Song *et al.*, 2009). The QIF does not require more assumptions than does the GEE method, but yields a substantial improvement in efficiency for the estimator of β when the working correlation is misspecified, and equal efficiency to the GEE when the working correlation is correct.

Lai and Small (2007) proposed another alternative GMM for the marginal regression analysis of longitudinal data with time-dependent covariates. They classified the time-dependent covariates into three types: Type I, Type II, Type III and showed that their GMM has advantages over the GEE with independence working correlation for Type II time-dependent covariate.

These two methods are the methods based on GMM introduced by Hansen (1982) and both of them use the GEEs.

In this paper, we compare these methods and interest in whether can improve the results of GEE with independence working correlation for the time-dependent covariates. We show that in the marginal regression model with Type II time-dependent covariates, GMM Type II of Lai and Small (2007) provides better result than QIF and for the Type III time-dependent covariates, QIF with working correlation matrix and GMM Type III methods provide the same results.

This paper was organized as following: Section 2 describes classification of time-dependent covariates by Lai and Small (2007), Section 3 shows a review of QIF by Qu *et al.* (2000), Section 4 introduces the GMM by Lai and Small (2007), Section 5 illustrates the result of simulation studies and final section is summary.

2. Classification of time-dependent covariates

The GEE assumes that the marginal mean $\mu_{it} = E(Y_{it}|X_{it})$ is a function of the covariates through a link function g with $g(\mu_{ij}) = X'_{ij}\beta$, and the variance of Y_{ij} is a function of the mean $\text{var}(Y_{ij}) = \phi V(\mu_{ij})$, where ϕ is the dispersion parameter.

The GEE solves the equation

$$S_\beta(\beta, W) = \sum_{i=1}^{m} \left(\frac{\partial \mu_i}{\partial \beta}\right)^T V_i^{-1} (Y_i - \mu_i) = 0 \tag{2.1}$$

where $V_i = A_i^{1/2} R_i(\alpha) A_i^{1/2}$ with A_i being the diagonal matrix of the marginal variances, $\mathrm{var}(Y_{ij})$ and $R_i(\alpha)$ being the working correlation matrix.

Using (2.1), Lai and Small (2007) classified time-dependent covariates into three types as following:

Type I : A time-dependent covariate X^j as being of Type I if it satisfies

$$E_{\beta_0}\left[\frac{\partial \mu_{is}(\beta_0)}{\partial \beta_j}\left(Y_{it}-\mu_{it}(\beta_0)\right)\right]=0,\ \text{for all } s,t,\ s\geq t,\ s=1,\cdots T,\ t=1,\cdots T \tag{2.2}$$

A sufficient condition for all covariates to be of Type I is that FCCM assumption holds. That is, values X_{itk} of this covariate are independent of each Y_{it} conditional upon the associated X_{it}.

Type II : A time-dependent covariate X^j as being of Type II if it satisfies

$$E_{\beta_0}\left[\frac{\partial \mu_{is}(\beta_0)}{\partial \beta_j}\left(Y_{it}-\mu_{it}(\beta_0)\right)\right]=0,\ \text{for all s,t},\ t=1,\cdots T \tag{2.3}$$

A sufficient condition for all covariates to be of Type II is that

$$f\left((X_{i,t+1},\cdots,X_{i,T})\mid Y_{it},X_{it}\right)=f\left((X_{i,t+1},\cdots,X_{i,T})\mid X_{it}\right)$$

For the Type II covariate, it satisfies that the response variable is independent of future values of the covariate, conditionally upon current values. That is, for Type II covariates, there may be delayed influences of past covariates upon future responses, but not vice versa.

Type III : A time-dependent covariate X^j as being of Type III if it satisfies

$$E_{\beta_0}\left[\frac{\partial \mu_{is}(\beta_0)}{\partial \beta_j}\left(Y_{it}-\mu_{it}(\beta_0)\right)\right]\neq 0,\ \text{for some } s>t. \tag{2.4}$$

Type III covariates may have feedback loops in which the response influences future values of the covariate, so only the moment conditions associated with working independence hold as following:

$$E_{\beta_0}\left[\frac{\partial \mu_{it}(\beta_0)}{\partial \beta_j}\left(Y_{it}-\mu_{it}(\beta_0)\right)\right]=0,\ t=1,\cdots T,\ j=1,\cdots,p. \tag{2.5}$$

3. Quadratic inference function

The QIF is derived by approximating the inverse of the working correlation matrix by a linear combination of several basis matrices which composed 0s and 1s:

$$R^{-1}=\sum_{i=1}^{m} a_i M_i \tag{3.1}$$

where $M_1,\cdots,M_m$ are known matrices and $a_1,\cdots,a_m$ are unknown constant.

Substituting (3.1) into (2.1), consider the following class of estimating functions:

$$S_\beta(\beta,W)=\sum_{i=1}^{N}\left(\frac{\partial \mu_i}{\partial \beta}\right)^T A_i^{-\frac{1}{2}}(a_1M_1+\cdots+a_mM_m)A_i^{-\frac{1}{2}}(Y_i-\mu_i). \tag{3.2}$$

Define the 'extended score' g_N to be

$$g_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} g_i(\beta) = \frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N}\left(\frac{\partial\mu_i}{\partial\beta}\right)^T A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} (Y_i - \mu_i) \\ \vdots \\ \sum_{i=1}^{N}\left(\frac{\partial\mu_i}{\partial\beta}\right)^T A_i^{-\frac{1}{2}} M_m A_i^{-\frac{1}{2}} (Y_i - \mu_i) \end{pmatrix}. \tag{3.3}$$

The vector g_N contains more estimating equations than unknown parameters, the GMM can be applied and define the quadratic inference function to be

$$Q_N(\beta) = g_N^{'} C_N^{-1} g_N \tag{3.4}$$

where $C_N = (1/N^2)\sum_{i=1}^{N} g_i(\beta) g_i^{'}(\beta)$.

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

$$\widehat{\beta} = \arg\min_{\beta} Q_N(\beta)\ . \tag{3.5}$$

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 estimators are consistent, asymptotically normal, and equal or more efficient than GEE estimators depending on working correlation matrix.

4. Generalized method of moments

The use of GMM to estimate the p-dimensional parameter β requires an $r \geq p$ vector $g(Y_i, X_i, \beta)$ of "valid" moment conditions means of which is zero

$$E_{\beta_0}[g(Y_i, X_i, \beta)] = 0. \tag{4.1}$$

Lai and Small (2007) proposed GMM method as incorporating all possible valid equations, but no invalid ones.

For a Type I time-dependent covariate, there are T^2 valid moment conditions in (2.2), for a Type II time-dependent covariate, there are $T(T+1)/2$ valid moment conditions in (2.3) and for a Type III time-dependent covariate there are T valid moment conditions in (2.5). The valid moment conditions for each of the p covariates are combined into the vector g. The sample version of (4.1) is

$$G_N(\beta) = \frac{1}{N}\sum_{i=1}^{N} g(Y_i, X_i, \beta) \tag{4.2}$$

The GMM estimator is $\widehat{\beta} = \arg\min_{\beta} Q_N(\beta)$, where $Q_N(\beta) = G_N^{'}(\beta) W_N G_N(\beta)$ and W_N is the positive definite weight matrix. The GMM estimator has the same large sample properties as QIF estimator.

5. QIF and GMM for longitudinal data with Type II and Type III time-dependent covariates

The FCCM assumption is violated in the case of Type II and Type III time-dependent covariates. In our previous paper (Cho and Dashnyam, 2013), we showed that

- The QIF with independence working correlation is the same as GEE with independence working correlation, so it provides an unbiased and consistent estimator in the case of time-dependent stochastic covariates.
- The estimator of QIF with exchangeable or AR(1) working correlation cannot be consistent and asymptotically normal because some of the extended score vector cannot be valid moment conditions.

In summary, we say that the QIF with independence working correlation is the best choice in marginal regression for longitudinal data with time-dependent stochastic covariates.

Lai and Small (2007) showed that their method can be more efficient than GEE with independence correlation in the case of Type II and same efficient in the case of Type III time-dependent covariates. The GMM method of Lai and Small (2007) uses only valid moment conditions, so their method can be consistent and asymptotically normal even in Type II and Type III. The consistency and asymptotic normality of the GMM estimator rest on having valid moment conditions

6. Simulation study

Case of the Type II covariate:

Lai and Small (2007) and Diggle *et al.* (2002) considered the model of time-dependent covariate vector X_i which is standardized AR(1) Gaussian process with autocorrelation parameter ρ.

$$Y_{it} = \gamma_o + \gamma_1 X_{it} + \gamma_2 X_{it-1} + b_i + e_{it} \tag{6.1}$$

$$X_{it} = \rho X_{it-1} + \epsilon_{it} \tag{6.2}$$

where $b_i, e_{it}, \epsilon_{it}$ are mutually independent, $b_i \sim N(0,1)$, $e_{it} \sim N(0,1)$ and $\epsilon_{it} \sim N(0, 1-\rho^2)$ $X_{i0} \sim N(0,1)$ and $X_{it} \sim N(0,1)$. This model implies

$$E(Y_{it}|X_{i1}, \cdots X_{in}) = \gamma_o + \gamma_1 X_{it} + \gamma_2 X_{it-1} \tag{6.3}$$

but yields the marginal mean

$$E(Y_{it}|X_{it}) = \beta_o + \beta_1 X_{it} \tag{6.4}$$

where $\beta_o = \gamma_0$ and $\beta_1 = \gamma_1 + \gamma_2 * \rho$. So FCCM assumption is violated. Furthermore,

$$E\left[X_{is}(Y_{is} - E(Y_{it}|X_{it}))\right] = \begin{cases} 0 & \text{for } s \geq t \\ \left(\gamma_2 \rho^{t-s-1} - \gamma_2 \rho^{t-s+1}\right) \dfrac{1}{1-\rho^2} & \text{for } s < t \end{cases}$$

So the covariate is Type II time dependent covariate. In this case, by simulation study of Lai and Small (2007), their GMM Type II method is unbiased and is substantially more efficient than GEE with working independence matrix (almost twice as efficient).

We did simulation study using this model with the same coefficients as Lai and Small (2007). Simulation number is 500 and each data set contains N=100 subjects with t=6 time measurements. We consider three estimators: 1) GEE or QIF with independence working correlation (GEE-ind/QIF-ind), 2) QIF using AR(1) working correlation (QIF-AR(1)), 3) QIF using exchangeable working correlation (QIF-exch). Although current version of QIF can handle four kinds of working correlation structure, QIF with unstructured working correlation use adaptive QIF which used variance matrix of responses instead of basic matrices (Qu *et al.*, 2003). So unstructured working correlation is not compared.

The simulation results are shown in Table 6.1. QIF with independence working correlation is the same efficient as GEE with working independence correlation and QIF with AR(1) and exchangeable working correlation cannot be efficient than GEE with independence working correlation.

Table 6.1 Results of simulation study for the model which has a Type II time-dependent covariate. The bias, mean squared error (MSE) and ratio of MSE of GEE independence to MSE of estimator (SRE) for the parameter β_1 in (6.4)

Estimator	Bias	MSE	SRE
GEE-ind/QIF-ind	-0.003	0.013	1
QIF-AR(1)	-0.003	0.014	0.987
QIF-exch	-0.003	0.091	0.151

If this results are compared to Lai and Small's results, the GMM Type II method can provide more efficient estimator than QIF with independence working correlation in the case of the Type II time-dependent covariates.

<u>Case of the Type III covariate:</u>

Lai and Small (2007) considered the model with Type III time-dependent covariate

$$Y_{it}=\beta X_{it}+\kappa Y_{i,t-1}+u_{it} \tag{6.5}$$

$$X_{it}=\gamma Y_{i,t-1}+v_{it} \tag{6.6}$$

where $u_{i1},\cdots,u_{iT},v_{i1},\cdots,v_{iT}$ are mutually independent mean zero normal random variables with variances σ_u^2 for $u_{i1},\cdots,u_{iT}$ and σ_v^2 for $v_{i1},\cdots,v_{iT}$ and (Y_{it},X_{it}) is stationary.

The marginal mean model is

$$E(Y_{it}|X_{it})=\left[\beta+(\kappa\gamma)\left(\frac{\beta^2\sigma_u^2+\sigma_v^2}{\gamma^2\sigma_u^2+\sigma_v^2-2\sigma_v^2\beta\kappa\gamma-\sigma_v^2\kappa^2}\right)\right]X_{it}\equiv\theta X_{it} \tag{6.7}$$

The response Y_{it} has a feedback effect on the covariate process X_{it} and consequently X_{it} is Type III time-dependent covariates. In this case GMM Type III method by Lai and Small (2007) is the almost same with GEE independence.

By our simulation results using this model, QIF with independence working correlation showed better result than QIF with AR(1) and exchangeable working correlation and the same result as GEE with independence working correlation.

Table 6.2 Results of simulation study for the model which has a Type III time-dependent covariate. The bias, mean squared error (MSE) and ratio of MSE of GEE independence to MSE of estimator (SRE) for the parameter θ in (6.7)

Estimator	Bias	MSE	SRE
GEE-ind/QIF-ind	0.001	0.002	1
QIF-AR(1)	0.002	0.002	0.984
QIF-exch	-0.005	1.308	0.001

Comparing our simulation results with simulation results of Lai and Small (2007), we can say that for the Type III time-dependent covariate, QIF with working independence and GMM Type III methods have the same results.

7. Summary

In this paper, we compared QIF and GMM methods in the case of Type II and Type III in which FCCM is violated.

For the marginal regression model with Type II time-dependent covariates, GMM Type II of Lai and Small (2007) provides better result than QIF with independence working correlation and for the Type III time-dependent covariates, QIF with independence working correlation and GMM Type III methods provide the same results.

We did simulation study using the models which are discussed by Lai and Small (2007). The results of our simulation study were compared to the simulation results of Lai and Small (2007) and the simulation study verified our conclusion.

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.

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. (1995). A caveat concerning independence estimating equations with multivariate binary data. *Biometrics*, **51**, 309-317.

Fitzmaurice, G. M, Liard, N. M. and Ware, J. H. (2004). *Applied longitudinal analysis.* Wiley, New York.

Hansen, L. (1982). Large sample properties of generalized methods of moments estimators. *Econometrica,* **50**, 1029–1055.

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., Thomas, A. L. and John, E. C. (2000). Note on marginal linear regression with correlated response data. *The American Statistician,* **54**, 191-195.

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.

Qu, A. and Lindsay, B. G. (2003). Building adaptive estimating equation when inverse of covariance estimation is difficult. *Journal of the Royal Statistical Society B*, **65,** 127–142.

Song, P. X.-K., Jiang, Z., Park, E. J. and Qu, A. (2009) Quadratic inference functions in marginal models for longitudinal data. *Statistical Medicine*, **28**, 3683–3696.