

Reviewing of Different Methods for Handling Longitudinal Count data

Amany. M, Mousa¹, Ahmed. A, El sheikh², Ahmed M. Gad³ and

Fatma El Zahraa S.Salama^{4*}

¹*Faculty of Graduate Studies for Statistical Research, Cairo University, Cairo, Egypt*

²*Faculty of Graduate Studies for Statistical Research, Cairo University, Cairo, Egypt*

³*Faculty of Economics and Political Science, Cairo University, Cairo, Egypt*

⁴*Faculty of Graduate Studies for Statistical Research, Cairo University, Cairo, Egypt*

¹dramanymousa04@gmail.com, ²aham103@yahoo.com, ³ahmed.gad@feps.edu.eg,

^{4*}fatmafeps90@yahoo.com

Abstract:

In this paper, we will review the methods that used to handle longitudinal data in the case of marginal models when inferences about the population average are the primary focus [1] or when future applications of the results require the expectation of the response as a function of the current covariates [7]. We will review the generalized estimating equations method (GEE), quadratic inference functions (QIF), generalized quasi likelihood (GQL) and the generalized method of moments (GMM). These methods will be reviewed by discussing its advantages and disadvantages in more details.

Keywords: Generalized estimating equations- Generalized Method of Moments- Generalized quasi likelihood-longitudinal count data- Quadratic inference functions.

1.Introduction:

In a longitudinal study, researchers conduct several observations for the same subjects over period of time. Also, the benefit of a longitudinal study is that researchers can detect developments or changes in the characteristics of the target population at both the group and the individual level. The key here is that longitudinal studies extend beyond a single moment in time. As a result, they can establish sequences of events. Another major strength of the longitudinal design is that it avoids cohort effects, because the researcher examines one group of people over time, rather than comparing several different groups that represent different ages and generations. Longitudinal data also allow researchers to discuss how a single individual's behavior changes with age.

Sometimes referred to longitudinal data as panel data, the descriptor panel data comes from surveys of individuals. In this context “panel” is a group of individuals surveyed repeatedly over time. Historically, panel data methodology within economics had been largely developed through labor economics applications. Now, economic applications of panel data methods are not confined to survey or labor economics problems and the interpretation of the descriptor “panel analysis” is much broader. This type of data has several advantages over other types of data. There is more than single outcome for every subject, this allows investigator to separate between change over time within individuals, i.e. (aging effect) and difference between subjects at baseline, i.e. (cohort effect). Each subject serves his own control, so variability becomes less because the intra-subject variability is substantially less than inter-subject variability. Longitudinal studies involve using and collecting data for long periods of time, they can be helpful in determining patterns. It is possible to learn more about cause-and-effect relationships through these types of research studies and connections can be made more clearly. The observations are not independent, for longitudinal data since there are repeated measurements from the same individual over time. So, there is needed to model the variability in the data. Specifically, there are two likely sources of variation unique to longitudinal data [2].

- Random effects: Individuals sampled at random from a population may exhibit random variation between themselves. This variability would reflect natural heterogeneity among individuals because of underlying unmeasured or difficult to measure factors (such as genetic, environmental, or personality traits that are difficult to quantify).
- Serial correlation: Repeated measurements from the same individual over time can be correlated. Further, measurements from a person at two adjacent time points (say, month 1 and month 2) are expected to be more correlated (i.e., more similar) than those from more distant time points (say, month 1 and month 5).

Longitudinal data have several types; binary–categorical–count–continues. In this paper, we will concentrate on longitudinal count data that is a type of data in which the observations can take only the non-negative integer values $\{0, 1, 2, \dots\}$ where these integers arise from counting rather than ranking. Count data are quite common in epidemiological studies. Statistical treatment of count data is distinct from binary data that take only two values usually represented by 0 and 1. It is differ from ordinal data that consist of integer, where the individual values fall on a specific scale and only the relative ranking is important. Count data have become popular dependent variables in studies in various areas, especially due to the growing of this data on human and social behavior. Examples include the number of crimes in each neighborhood, number of accidents at an intersection, number of Facebook comments, rider ship in bikes haring programs[9]. Several methods are used to find estimate of longitudinal count data, so we will review these methods in more details.

2. Marginal Model

The defining feature of marginal models is a regression model relating the mean response on each occasion, via a suitable link function, to the covariates. With a marginal model, the focus is on making inferences about population means. This model type of linear model that accounts for repeated response measures on the same subject. They extend the general linear model by allowing and accounting for non-independence among the observations from the

same subject by estimating one or more parameters that capture the covariance among the residuals. So rather than assuming a single constant variance and zero covariance for all residuals, observations from the same subject are allowed to have different variances and nonzero covariances. The pattern of variances and covariances is known as the covariance structure of the R matrix.

2.1 The three main parts for the marginal model

- I. The conditional expectation of each response, $E(Y_{ij}|X_{ij}) = \mu_{ij}$, is assumed to depend on the covariates through a known link function $h^{-1}(\cdot)$, e.g., $\text{logit}(\mu_{ij})$ or $\text{log}(\mu_{ij})$,

$$h^{-1}(\mu_{ij}) = \eta_{ij} = X'_{ij}\beta.$$

- II. The conditional variance of each Y_{ij} , given X_{ij} , is assumed to depend on the mean according to

$$\text{Var}(Y_{ij}) = \phi v(\mu_{ij})$$

where $v(\mu_{ij})$ is a known “variance function” (i.e., a known function of the mean, μ_{ij} and ϕ is a scale parameter that may be fixed and known or may need to be estimated.

- III. The conditional within-subject association among the vector of repeated responses, given the covariates, is assumed to be a function of an additional vector of association parameters, say α (and also depends upon the means, μ_{ij} . For example, the components of α might represent the pairwise correlations or log odds ratios among the repeated responses.

Marginal models are a very natural way to extend generalized linear models to longitudinal responses. Marginal models specify a generalized linear model for the longitudinal responses at each occasion but also include a model for the within-subject association among the responses. A crucial aspect of marginal models is that the mean response and within-subject association are modeled separately. This separation of the modeling of the mean response and the association among responses has important implications for interpretation of the regression parameters in the model for the mean response. In particular, the regression parameters β in the marginal model have so-called population averaged interpretations. That described how the mean response in the population is related to the covariates.

3. Different method for finding the estimate for longitudinal data

We will review the methods that used to handle longitudinal data in the case of marginal models when inferences about the population average are the primary or when future

applications of the results require the expectation of the response as a function of the current covariates. Section (3.1) will review the generalized estimating equations method (GEE), section (3.2) will review the quadratic inference functions (QIF), section (3.3) will review generalized quasi likelihood (GQL) and section (3.4) will review the generalized method of moments (GMM).

3.1 Generalized estimating equations method (GEE)

The generalized estimating equation was developed [6] from generalized linear models and quasi likelihood to deal with non-normal correlated longitudinal data. This method extended generalized linear models, which include simple linear regression, for two important cases. First, given a data set consisting of repeated measures, a GEE model allows the correlation of outcomes within an individual to be estimated and taken into appropriate account in the formulae which generate the regression coefficients and their standard errors. Secondly, GEE models permit the calculation of robust estimates for the standard errors of the regression coefficients. Provided the basic linear regression relationship is correct and there is no correlation in the measured responses between individuals, robust standard errors ensure consistent inferences from a GEE model even if the chosen correlation structure is incorrect or if the strength of the correlation between repeated observations varies somewhat from individual to individual. This method provided the population-averaged estimates of the parameters.

The basic GEE strategy is to simply generalize the quasi-likelihood equations to the multivariate setting by replacing Y_i and μ_i by their vector counterparts and using a weight matrix W_i . This yield

$$\sum_{i=1}^N \frac{\partial \mu_i}{\partial \beta} W_i (Y_i - \hat{\mu}_i) = 0 \quad (1)$$

$\frac{\partial \mu_i}{\partial \beta}$ is defined as an $n_i \times p$ matrix, whose j^{th} row is $\frac{\partial \mu_{ij}}{\partial \beta^T}$. Although optimally it would take $W_i = V_i^{-1}$, in fact any positive definite and symmetric matrix can be used for W_i . If $W_i = V_i^{-1}$, it depends upon β . It provides a semi-parametric approach to longitudinal analysis of non-normal response. In practice, the working correlation matrix is usually unknown and must be estimated according to the following table.

Table (1) Working Correlation Structures and Estimators

Structure	Working Correlation	Estimator
Independent	$\text{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ 0 & j \neq k \end{cases}$	The working correlation is not estimated in this case
Exchangeable	$\text{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha & j \neq k \end{cases}$	$\hat{\alpha} = \frac{1}{(N^* - p)\phi} \sum_{i=1}^N \sum_{j < k} e_{ij} e_{ik}$, $N^* = 0.5 \sum_{i=1}^N n_i(n_i - 1)$

Unstructured	$\text{Corr}(Y_{ij}, Y_{ik}) = \begin{cases} 1 & j = k \\ \alpha_{jk} & j \neq k \end{cases}$	$\hat{\alpha}_{jk} = \frac{1}{(k-p)\phi} \sum_{i=1}^N e_{ij} e_{ik}$
Autoregressive AR(1)	$\text{Corr}(Y_{ij}, Y_{ik}) = \alpha^t, \quad t=0,1,2,\dots,n_i-j$	$\hat{\alpha} = \frac{1}{(K_1-p)\phi} \sum_{i=1}^N \sum_{j < n_i-1} e_{ij} e_{i,j+1},$ $K_1 = \sum_{i=1}^N (n_i - 1)$

Source: Salama et al (2016).

where: e_{ij} is the standardized Pearson residual:

$$e_{ij} = \frac{y_{ij} - u_{ij}}{\sqrt{v(u_{ij})(1-h_{ij})}} \quad (2)$$

h_{ij} : The elements of the "hat" matrix for all i, j and ϕ is the dispersion parameter is estimated by:

$$\hat{\phi} = \frac{1}{N-p} \sum_{i=1}^N \sum_{j=1}^{n_i} e_{ij}^2 \quad (3)$$

where $N = \sum_{i=1}^N n_i$ is the total number of measurements and p is the number of regression parameters. Then to find the covariance matrix for $\hat{\beta}$ the sandwich estimator will be used to get it as follow.

$$\left[\sum_{i=1}^N D_i' V_i^{-1} D_i \right]^{-1} \hat{M} \left[\sum_{i=1}^N D_i' V_i^{-1} D_i \right]^{-1} \quad (4)$$

with

$$\hat{M} = \sum_{i=1}^N D_i' V_i^{-1} \text{Cov}(Y_i) V_i^{-1} D_i$$

3.1.1 Advantages of generalized estimating equations (GEE):

- Computationally more simple than MLE for categorical data, since estimation in GEEs is carried out with quasi-likelihood methods which is computationally easier than full-likelihood methods.
- Only requires the correct specification of marginal mean and variance as well as the link function which connects the covariates of interest and marginal means.
- The parameter estimates are consistent and asymptotically normally distributed even when the "working" correlation structure of responses is mis-specified [11].

3.1.2 Limitations for generalized estimating equations (GEE):

- There is no likelihood function since the GEE does not specify completely the joint distribution.
- Empirical based standard errors underestimate the true ones, unless very large sample size.

- Misspecification of the correlation structure may lead to a great loss of efficiency although consistency remains holds.
- This method is very sensitive to outliers or contaminated data which is worrisome in practice.

Under mis-specification of the correlation coefficient, it was pointed [11] estimates that obtained by the method of moments, the asymptotic estimates of the parameter (α) which specifies the working correlation matrix may not be well defined or exist so the correlation coefficient estimates did not converge to the true values. Furthermore, for discrete random vectors, the correlation matrix was usually complicated, and it was difficult to attain multivariate distributions with specified correlation structures. These limitations lead researchers to work on this area to develop new methodologies. Several alternative approaches for estimating the correlation coefficients have been proposed. Correctly specifying “working” correlation structure can enhance the efficiency of the parameter estimates when the sample size is not large enough. Therefore, how to select intra-subject correlation matrix plays a vital role in GEE with improved finite-sample performance; also, the variance function is another potential factor affecting the goodness-of-fit of GEE. Correctly specified variance function can assist in the selection of covariates and an appropriate correlation structure. Different criteria might be needed due to the goal of model selection. Although the GEE provides a consistent estimator regardless of whether the working correlation is correctly specified or not, the estimator can be inefficient under mis-specified correlation structures.

3.1.3 The selection of “working” correlation structure

The methods was reviewed[13] that can be used to select the correct working correlation structure. the adequacy of “working” correlation structure can be examined through

$$\Gamma = (D_i' V_i^{-1} D_i)^{-1} \hat{M}_{LZ} \quad (5)$$

where

$$\hat{M}_{LZ} = \sum_{i=1}^K D_i' V_i^{-1} Cov(Y_i) V_i^{-1} D_i$$

and

$Cov((Y_i) = \hat{\tau}_i \hat{\tau}_i'$ with $\hat{\tau}_i = Y_i - \mu_i$ is an estimator of the variance covariance matrix of Y_i . The statistic $R_{J(R)} = \sqrt{(1 - R_{J1})^2 + (1 - R_{J2})^2}$ where $R_{J1} = \text{trace}(\Gamma)/p$ and $R_{J2} = \text{trace}(\Gamma^2)/p$ respectively. If the “working” correlation structure R is correctly specified R_{J1} and R_{J2} will be thus close to 1, leading to $R_{J(R)}$ approaching 0. Thus R_{J1} , R_{J2} , and $R_{J(R)}$ can all be used for correlation structure selection. A criterion was proposed[10] for selecting “working” correlation structure based on the minimization of the generalized error sum of squares (ESS) given as follow:

$$\begin{aligned} ESS &= \sum_{i=1}^K (Y_i - \mu_i)' V_i^{-1} (Y_i - \mu_i) \\ &= \sum_{i=1}^K Z_i'(\beta) R_i^{-1}(a) Z_i(\beta) \end{aligned}$$

where

$$Z_i(\beta) = A^{1/2}(Y_i - \mu_i)$$

The criterion is defined by

$$SC = \frac{ESS}{(N-p-q)}$$

where $N = \sum_{i=1}^K n_i$ is the total number of observations, p is the number of regression parameters, and q is the number of correlation coefficients within the “working” correlation structure. Another extended criterion from SC was proposed by Carey and Wang(2011), where the Gaussian pseudolikelihood (GP) is adopted, and it is given by :

$$GP(R) = -0.5 \times \sum_{i=1}^K (Y_i - \mu_i)' V_i^{-1} (Y_i - \mu_i) + \log(|V_i|)$$

where a better “working” correlation structure yields a larger GP. In their work, they also showed that GP criterion held better performance than RJ via simulation. Besides those criteria mentioned above, there are more criteria that can be used to decide which correlation matrix (R) can be used.

3.2 Quadratic inference functions (QIF)

To overcome some of the difficulties in the use of the GEE method, many improvements have been proposed in the literature. Among these improvements, the quadratic inference function a method was introduced [8] of quadratic inference functions(QIF) that does not involve direct estimation of the correlation parameter, and remains optimal even if the working correlation structure is mis-specified. The method does not need direct estimation of the correlation parameter. The idea is to represent the inverse of the working correlation matrix by the linear combination of basis matrices, a representation that is valid for the working correlations most commonly used. To find the joint estimation of β and v , the joint estimating equations was developed by employing the first two moments of the response variable as:

$$\sum_{i=1}^n \frac{\partial \mu'_i}{\partial \beta} A_i^{-\frac{1}{2}} R(\alpha)_i^{-1} A_i^{-\frac{1}{2}} (f_i - \mu_i) = 0 \quad (6)$$

This method has received some attention in the recent years, as it improved on the GEE method in several aspects simultaneously:

- The QIF provides statistical inference under the same model assumptions as the GEE requires. Like the GEE method, a mis-specified working correlation does not affect the consistency of the regression parameter estimation. It provides a robust sandwich estimator for the variance of the regression parameter estimator. When the working correlation structure is correctly specified, both the QIF and GEE are equally efficient. However, when the working correlation structure is mis-specified, the QIF is more efficient than the GEE, especially when time-varying covariates are included in the analysis.
- The QIF provides a goodness-of-fit test for checking the validity of the assumption for the first moment mean model. This is a crucial condition for the estimation consistency that the GEE requires but is unable to validate easily. Several summary statistics was introduced[4] similar to those considered in the GLMs under the independence working correlation
- The QIF is robust against outliers or contaminated data since it has been proved to have a bounded influence function.

To obtain the estimators of β and ν , specification of the working correlation structure and estimation of α in $R(\alpha)$ are required prior to solving equation. To handle this problem, A class of basis matrices was employed [8] to represent the inverse of R^{-1} . The QIF is derived by observing that the inverse of the working correlation matrix can be approximated by a linear combination of several basis matrices. The QIF is derived by observing that the inverse of the working correlation matrix can be approximated by a linear combination of several basis matrices:

$$R^{-1} \approx \sum_{l=0}^k \alpha_l M_l$$

Where M_0 is the identity matrix, M_1, \dots, M_k are known basis matrices with 0 or 1 as components and $\alpha_0, \dots, \alpha_k$ are unknown coefficients. The above expression holds exactly for some common working correlation structures. In practical implementation, one must choose the basis for the inverse of the correlation matrix, determine the magnitude of λ , and calculate the standard error and confidence interval of the resulting estimator.

3.2.1 Choice of the basis for the inverse of the correlation matrix

Qu and Li (2006) discussed the choice of basis matrices, if the working correlation is exchangeable, then $R^{-1} = \alpha_0 M_0 + \alpha_1 M_1$ where M_1 is 0 on the diagonal and 1 elsewhere. If it is an Auto Regressive1 (AR1) working correlation, $R^{-1} = \alpha_0 I + \alpha_1 M_1^* + \alpha_2 M_2^*$ where M_1^* has 1 on the sub-diagonal and 0 elsewhere and M_2^* has 1 on the two corner components of the diagonal. The advantage of this approach is that it does not require estimation of linear coefficients a_j 's which can be viewed as nuisance parameters, since the generalized estimating equation is a linear combination of elements of the estimating functions. Depending on the above linear combination of the elements of the following extended score vector: to find the estimate of β and ν that depend on minimize $Q_n(\beta, \nu)$ that take the following formula:

$$Q_n(\beta, \nu) = n \bar{g}_n(\beta, \nu) C_n^{-1}(\beta, \nu) \bar{g}_n(\beta, \nu)$$

where

$$\bar{g}_n(\beta, \nu) = \frac{1}{n} \sum_{i=1}^n g_i(\beta, \nu) = \frac{1}{n} \begin{pmatrix} \sum_{i=1}^n \frac{\partial \mu'_i}{\partial(\beta, \nu)} A_i^{-\frac{1}{2}} M_1 A_i^{-\frac{1}{2}} (Y_i - \hat{\mu}_i) \\ \vdots \\ \sum_{i=1}^n \frac{\partial \mu'_i}{\partial(\beta, \nu)} A_i^{-\frac{1}{2}} M_m A_i^{-\frac{1}{2}} (Y_i - \hat{\mu}_i) \end{pmatrix} \quad (7)$$

and

$$C_n^{-1}(\beta, \nu) = \frac{1}{n} \sum_{i=1}^n g_i(\beta, \nu) g'_i(\beta, \nu)$$

The QIF estimator $(\hat{\beta}, \hat{\nu}) = \text{argmin } Q_n(\beta, \nu)$ and the asymptotic covariance matrix for the statistical inference on (β, ν) can be estimated using the following equation:

$$\hat{\Sigma} = \left[\left\{ \frac{1}{n} \sum_{i=1}^n \frac{\partial g_i(\beta, \nu)}{\partial(\beta, \nu)} \right\}' \left\{ \frac{1}{n} \sum_{i=1}^n g_i(\hat{\beta}, \hat{\nu}) g'_i(\hat{\beta}, \hat{\nu}) \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\partial g_i(\beta, \nu)}{\partial(\beta, \nu)} \right\} \right]^{-1} \quad (8)$$

3.2.2 This method has important limitations in application

- It needs to specify that the inverse of the working correlation is a linear combination of several basis matrices.
- There is no idea which working correlation structure is the most appropriate, then it must be chosen in an ad hoc manner.
- No matter how chosen, the dimension of the extended score vector is $m\alpha$; where m is the number of basis elements and α is the number of parameters.

3.3 Generalized quasi likelihood (GQL)

This approach was developed[12] to estimate β , where the generalized quasi-likelihood estimator of β is the root of the score equation:

$$\sum_{i=1}^N D_i^T \Sigma_i(\rho)^{-1} (f_i - \mu_i) \quad (9)$$

where $f_i = (\Sigma_i(\rho)^{-1})^{1/2}$ is the true covariance matrix of y_i that can be expressed as $\Sigma_i(\rho)^{-1} = A_i^2 C(\rho) A_i^2$ with $A_i = \text{diag}(\sigma_{i1}, \dots, \sigma_{ij})$ and $C(\rho)$ as the true correlation matrix of Y_i , ρ being a correlation index parameter. To overcome the difficulty of unknown $C(\rho)$ in practice a general stationary auto-correlation structure has been suggested given by

$$C(\rho) = C(\rho_1, \rho_2, \dots, \rho_{j-1}) = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{T-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{T-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \rho_{T-3} & \dots & 1 \end{pmatrix}$$

where for $l = 1, 2, \dots, T-1$, ρ_l , represents the lag l autocorrelation. The GQL estimate of β is then computed by solving the above equation. In practice ρ is unknown therefore the lag correlations can be consistently estimated by using the well known method of moments. For $l = |j - j'|$, $j \neq j'$ and $j, j' = 1, 2, \dots, T$ the autocorrelation of lag l , ρ_l , is estimated by the method of moments as:

$$\hat{\rho}_l = \frac{\sum_{i=1}^N \sum_{t=1}^{T-l} \tilde{y}_{it} \tilde{y}_{it+l} / N(T-1)}{\sum_{i=1}^N \sum_{t=1}^{T-1} \tilde{y}_{it}^2 / NT}$$

where \tilde{y}_{it} is the standardized residual, defined as $\tilde{y}_{it} = \frac{y_{it} - \mu_{it}}{\sigma_{it}^{1/2}}$. Let $\hat{\beta}_{GQL}$ denote this estimator, which is consistent for β . Under some mild conditions, it can be shown that $\hat{\beta}_{GQL}$ has the asymptotic covariance matrix V_G^* given by:

$$V_G^* = \lim_{N \rightarrow \infty} \left\{ \sum_{i=1}^N \frac{\partial \mu_i'}{\partial \beta} A_i^2 C^{-1}(\rho_1, \dots, \rho_{n-1}) A_i^2 \frac{\partial \mu_i}{\partial \beta} \right\}^{-1} \quad (10)$$

Through a simulation study it is showed[12] that GQL performs the best in estimating both the regression and the true correlation parameters, even though the longitudinal correlations are estimated separately by the method of moments.

3.4 Generalized Method of Moments (GMM)

Another approach that could be used to estimate parameters for marginal models of correlated data is to apply generalized method of moment's estimation (GMM)[3].

This method, like the Generalized Estimating Equations, accounts for correlation inherent in the data due to repeated measurements taken on the same subjects. Unlike GEE, GMM relies on the use of moment conditions, expressions with zero expectation, rather than on the derivation of the likelihood or quasi-likelihood functions. GMM depends on minimizing a quadratic form, $J_t(\beta)$, over the parameters [3, 5].

$$J_t(\beta) = \ddot{G}_N(\beta)^T W_N \ddot{G}_N(\beta) \quad (12)$$

where W_N is a weight matrix, a positive semidefinite matrix. The vector \ddot{G}_N in the above equation is the average of moment conditions for all N subjects:

$$\ddot{G}_N = \frac{1}{N} \sum_{i=1}^N g(Y_i, X_i, \beta_0),$$

where Y_i is the vector of responses for subject i , X_i is the vector of covariates for subject i , β_0 is the vector of true parameters. When time-dependent covariates are present, the vector $g_i(Y_i, X_i)$ is composed of only the moment conditions that are considered “valid” for subject i , defined [5] as satisfying the expression:

$$E[g_i(Y_i, X_i)] = 0$$

Using moment conditions that were products of derivative and residual terms at different times[5]:

$$g_i = \frac{\partial u_{is}}{\partial \beta_j} (y_{it} - u_{it})$$

$$W_N = \hat{V}_N^{-1} = \text{Cov}^{-1}(g_i)$$

In obtaining the covariance matrix used to construct the quadratic form, $J_t(\beta)$, for obtaining GMM parameter estimates, additionally suggested the use of an iterative procedure in which an initial consistent GEE estimator β_0 is used to obtain $\text{cov}(g(Y_i, x_{it}, \beta_0))^{-1}$, then estimating GMM β yielding an estimator that is as asymptotically efficient as the traditional 2-Step GMM estimator and has consistent asymptotic variance [3, 5], given by:

$$\left\{ \left(\frac{1}{N} \sum_{i=1}^N \frac{g(Y_i, X_{it}, \beta)}{\partial \beta} \right)^T \hat{V}_N^{-1} \left(\frac{1}{N} \sum_{i=1}^N \frac{g(Y_i, X_{it}, \beta)}{\partial \beta} \right) \right\}^{-1}$$

where $\frac{g(Y_i, X_{it}, \beta)}{\partial \beta}$ is evaluated at $\beta = \hat{\beta}_{GMM}$

3.4.1 GMM estimators have become widely used for the following reasons:

- GMM estimators have large sample properties that are easy to characterize in ways that facilitate comparison. A family of such estimators can be studied a priori in ways that make asymptotic efficiency comparisons easy. The method also provides a natural way to construct tests which take account of both sampling and estimation error.
- In practice, researchers find it useful that GMM estimators can be constructed without specifying the full data generating process (which would be required to write down the maximum likelihood estimator). This characteristic has been exploited in analyzing partially specified economic models, in studying potentially mis-specified dynamic

models designed to match target moments, and in constructing stochastic discount factor models that link asset pricing to sources of macroeconomic risk.

- The method uses assumptions about specific moments of the random variables instead of assumptions about the entire distribution, which makes GMM more robust than ML, at the cost of some efficiency. The assumptions are called moment conditions
- The generalized method of moments was used [5] to make optimal use of the estimating equations that are made available by the covariates for marginal regression analysis of longitudinal data with time-dependent covariate.

4. Conclusion

There are different methods to Handel marginal model for longitudinal count data when inferences about the population average are the primary focus, but every method need some condition that should be considered to handle the data for example the generalized estimating equations and the generalized methods of moments both accounts for correlation inherent in the data due to repeated measurements taken on the same subjects. But GMM relies on the use of moment conditions, expressions with zero expectation, rather than on the derivation of the likelihood or quasi-likelihood functions. The quadratic inference function was introduced this method does not involve direct estimation of the correlation parameter, it only depends on specific basis and remains optimal even if the working correlation structure is misspecified.

References

- 1) Diggle, P., Heagerty, P., Liang, K. Y., & Zeger, S. (2002). Analysis of longitudinal data. USA: Oxford University Press.
- 2) Frees, E., w. (2004) Longitudinal and Panel Data: Analysis and Applications for the Social Sciences, Cambridge University press.
- 3) Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators, *Econometrica*, 50(4), 1029-1054
- 4) Hilbe, J., M., and Hardin, J., W., (2014) modeling count data with generalized distribution, *stata Journal*, 14(3), 562–579.
- 5) Lai, T. L., & 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, Series B*, 69(1), 79-99.
- 6) Liang, k., Y. and Zeger, S., L. (1986) Longitudinal data analysis using generalized linear models, *Biometrika*, 73 (1), 13–22.
- 7) 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. *Communs Statist. Simuln Computn*, 23, 939-951.
- 8) Qu, A., B. G. Lindsay, and B. Li. (2000) Improving generalised estimating equations using quadratic inference functions, *Biometrika*, 87 (4), 823–836.
- 9) Salama, S., F., Gad, M., A., and Mousa, A. (2016) Modeling Longitudinal Count Data with Missing Values: A Comparative Study, *Academic Journal of Applied Mathematical Sciences*, Academic Research Publishing Group, vol. 2(3), 19-26.

- 10) Shults, J., Chaganty, N.(1998) Analysis of Serially Correlated Data Using Quasi-LeastSquares, Biometrics, 54, 1622-1630.
- 11) Sutradhar, B., C., and Das, K., (1999) "On the efficiency of regression estimators in generalized linear models for longitudinal data," Biometrika, 86, 2, 459–465.
- 12) Sutradhar, B.C. (2003), An overview on regression models for discrete longitudinal responses, Statistical Sci, 18, 377-393.
- 13) Wang,M.,(2014)Generalized estimating equations in longitudinal data, Advances in statistics, 2014.