Fit Statistics for Nested Models in Which Parameter Estimates are Obtained Using Generalized Method of Moments in the Presence of Time-Dependent Covariates

Maryann N. Shane *

Abstract

The Generalized Estimating Equations (GEE) approach proposed by Liang and Zeger (1986) has become the most popular method in estimating parameters for longitudinal data. This technique accounts for correlation inherent among repeated observations, allowing the researcher to specify the nature of this working correlation. However, GEE presents major disadvantages when time-dependent covariates (TDCs), special types of predictors involving a feedback loop, are present (Pepe & Anderson, 1994; Fitzmaurice, 1995). Recently, Generalized Method of Moments (GMM) has been proposed as an alternative to GEE when estimating parameters of longitudinal data with time-dependent covariates (Lai & Small, 2007).

There has been a lack of attention paid to GMM fit statistics in the literature. This paper presents two statistics to assess the goodness-of-fit of models estimated using GMM in the presence of TDCs. Comparisons of identifying poor fit in nested models is compared to similar capabilities of the quasi-likelihood information criterion (QIC) using GEE.

Key Words: Generalized Method of Moments, time dependent covariates, longitudinal data, goodness-of-fit, nested models, information criterion

1 Introduction

This paper presents two measures of assessment of overall goodness-of-fit of nested models in which parameter estimates are obtained using Generalized Method of Moments (GMM) in the presence of time-dependent covariates (TDCs), special types of predictors that create feedback loops in the data structure.

For classical maximum likelihood estimation of independently observed data, some common statistics used in assessing model fit include the model deviance (Pregibon, 1981; Agresti, 1990; Dobson & Benett, 2008), Akaike's Information Criterion, or AIC (Akaike, 1973; Akaike, 1974), and Schwarz's Bayesian Information Criterion, or BIC (Schwarz, 1978). When using Generalized Estimating Equations (GEE) to estimate

*University of Northern Colorado, 501 20th Street, Greeley, CO 80639

model parameters for correlated data, a modification of the AIC, known as the quasi-likelihood information criterion, or QIC (Pan, 2001), is used. The literature is sparse in the discussion of methods used to assess the goodness-of-fit for models in which parameters are estimated using the Generalized Method of Moments (GMM) approach. The goal of this study is to present the need for such a statistic, as well as its establishment, whose details will be outlined in Section 4.

2 Longitudinal Data Analysis in the Presence of Time-Dependent Covariates

2.1 Time-Dependent Covariates

Time-dependent covariates (TDCs) are special types of predictors that oftentimes appear in longitudinal designs. Any covariate that changes over time and directly impacts the outcome of a study is considered a time-dependent covariate. The age of a participant in a longitudinal study to assess hypertension is an example of a time-dependent covariate because the participant's age will change (increase) over time. Annual glacial coverage in an ongoing study to assess causation and impact of global climate change, as well as the amount of chemical substance present in the half-life of a radioactive material, are also examples of time-dependent covariates. These special types of covariates introduce correlation among variables over time, and this correlation must be accounted for when constructing models for longitudinal data.

There are four types of time-dependent covariates, and distinctions are made based on the nature of their feedback. From a mathematical standpoint, TDCs are covariates that satisfy the expression:

$$E\left[\frac{\partial \mu_{is}(\beta_0)}{\partial \beta_j}\left\{y_{it} - \mu_{it}(\beta_0)\right\}\right] = 0, \tag{1}$$

where μ_{is} is the mean response for subject i at time s, β_j is the j^{th} covariate, y_{it} is the response for subject i at time t, μ_{it} is the mean response for subject i at time t, β_0 is the vector of true parameters, and s and t are different observation times, where $s = (1, \ldots, T)$ and $t = (1, \ldots, T)$. The four types of TDCs are defined by the combinations of s and t that maintain the equality in Equation (1).

A type I time-dependent covariate satisfies the equality in Equation (1) for all s and t, $s = (1, \ldots, T)$ and $t = (1, \ldots, T)$. This type of time-dependent covariate involves no feedback; the current covariate affects only the current response.

When the current covariate affects both the current and future responses, it is a type II time-dependent covariate. Type II TDCs satisfy the equality in Equation (1) for all combinations of s and t such that $s \geq t$.

The nature of the feedback involved in the presence of type III time-dependent covariates creates a complex feedback loop in which the current response affects the covariate at some future time, and the current covariate affects the response at some future time. Type III TDCs satisfy Equation (1) for some $s > t$, and it is recommended that the independent working correlation structure be used when this type of time-dependent covariate is present in analyses (Lai & Small, 2007).

The last type of time-dependent covariate, type IV, can be thought of as the "opposite" of type II TDCs. When type IV TDCs are present, the current response is associated with the current covariate, and the current response can also be associated with the covariate at some future time. By definition, a type IV TDC satisfies Equation (1) for all combinations of s and t such that $s \leq t$.

Due to the nature of the feedback imposed by time-dependent covariates on the data structure, the analysis of longitudinal data in the presence of TDCs can become challenging very quickly (Neuhaus & Kalbfleisch, 1998; Diggle et al, 2002). Moreover, depending on the method of analysis chosen, algorithm non-convergence of parameter estimation may be an additional hurdle in the process of obtaining model parameter estimates (Shane, 2013).

2.2 Generalized Estimating Equations and Time-Dependent Covariates

One of the most popular techniques in the analysis of correlated data is the Generalized Estimating Equations (GEE) approach (Liang & Zeger, 1986; Zeger & Liang, 1986). The premise behind GEE uses the assumption of a working correlation structure for the data under investigation. The researcher proposes a working correlation structure (e.g., compound symmetry, order-1 auto-regressive, exponential, etc.) that likely characterizes the nature of the correlation prevalent among repeated response measurements.

GEE uses estimating equations of the form:

$$S(\beta) = \sum_{i=1}^{N} X_i^T D_i R_i^{-1}(Y_i - \mu_i) = 0,$$

where X_i is the design matrix, D_i is a diagonal matrix, R_i is the working correlation structure, Y_i is the vector of responses, μ_i is the vector of mean responses, N is the total number of subjects, and $V_i = D_i R_i^{-1}$ is the working covariance matrix (Liang & Zeger, 1986; Zeger & Liang, 1986).

One advantage of using Generalized Estimating Equations in the analysis of longitudinal data is that regardless of whether or not the "correct" working correlation structure is selected, GEE parameter estimates are always consistent (Liang & Zeger, 1986; Fitzmaurice, 1995; Neuhaus & Kalbfleisch, 1998). Additionally, the GEE approach does not require full specification of the response distribution. It only requires information involving the mean-variance relationship of the response, hence only assuming a quasi-likelihood instead of the full likelihood (Liang & Zeger, 1986; Zeger & Liang, 1986).

However, using GEE to obtain parameter estimates in the presence of time-dependent covariates can be very difficult. Pepe and Anderson (1994) and Fitzmaurice (1995) advise that the independent working correlation structure be used or that a critical assumption behind the Generalized Estimating Equations process be checked when using it for parameter estimation when time-dependent covariates are present in the data.

GEE relies on the equality of the marginal expectation:

$$E\left[Y_{it}|X_{it}\right] = E\left[Y_{it}|X_{is}, s = 1, \ldots, T\right], \tag{2}$$

where Y_{it} is the response for subject i at time t ($t = 1, \ldots, T$), X_{it} is the covariate value for subject i at time t, and X_{is} is the covariate value for subject i at time s ($s = 1, \ldots, T$). This assumption states that the expected response at time t, given the covariate value at that same time, should be equal to the expected response at time t, given the covariate value at *any* time.

Oftentimes when time-dependent covariates are present, the equality in Equation (2) does not hold (Pepe & Anderson, 1994; Fitzmaurice, 1995); moreover, using GEE in the presence of TDCs may result in the loss of efficiency of parameter estimates associated with those covariates (Fitzmaurice, 1995).

2.3 Generalized Method of Moments

An alternate approach that can be taken in estimating parameters for correlated data is to implement Generalized Method of Moments (GMM) estimation. In comparison to using the Independent GEE approach, the use of GMM estimation improves efficiency when time-dependent covariates are present. Results from Lai and Small's (2007) simulation study show that GMM estimators are more efficient than Independent GEE estimators when time-dependent covariates of types I or II are involved, and they are equally as efficient as Independent GEE estimators when a TDC of type III is present. Moreover, in general, GMM estimators are equally as efficient as GEE estimators when the working correlation structure is correctly specified, and they are asymptotically more efficient than GEE estimators when the working correlation structure is misspecified (Lai & Small, 2007).

Generalized Method of Moments, like the Generalized Estimating Equations, is a method that accounts for correlation inherent in the data due to repeated measurements taken on the same subjects. However, it relies on the use of moment conditions rather than on the derivation of the likelihood or quasi-likelihood functions. Moment conditions are products of derivative and residual terms at different times:

$$\frac{\partial \mu_{is}}{\partial \beta_j}\left(y_{it} - \mu_{it}\right),$$

where μ_{is} is the mean response for subject i at time s ($s = 1, ..., T$), y_{it} is the response value for subject i at time t ($t = 1, ..., T$), μ_{it} is the mean response for subject i at time t, β_j is the j^{th} covariate, and $i = 1, ..., N$ denotes the subject.

Moment conditions that are considered “valid” are placed into a vector, g_i, of moment conditions, where “valid” is defined by Lai & Small (2007) as:

$$E\left[g(\mathbf{Y_i}, \mathbf{X_i}, \beta_\mathbf{0})\right] = \mathbf{0},$$

where $\mathbf{Y_i}$ is the vector of responses for subject i, $\mathbf{X_i}$ is the vector of covariates for subject i, and $\beta_\mathbf{0}$ is the vector of true parameters.

Once the vectors of “valid” moment conditions are obtained, they are averaged over all N subjects to obtain G_N, a vector of averaged moment conditions:

$$G_N = \frac{1}{N}\sum_{i=1}^{N} g(\mathbf{Y_i}, \mathbf{X_i}, \beta_\mathbf{0})$$

Then, a quadratic form, Q_N, is constructed using G_N:

$$Q_N = {G_N}^T W_N G_N,$$

where W_N is a weight matrix, with optimal choice being the inverse of the covariance matrix of the moment conditions, $\hat{V}_N^{-1}$ (Hansen, 1982). Hansen (2007) additionally suggested the use of an iterative procedure in which an initial consistent GEE estimator $\tilde{\beta}$ is used to obtain $cov\left\{g(y_i, x_i, \beta_0)\right\}^{-1}$, then estimating β_{GMM} using $\hat{V}_N^{-1}$, yielding an estimator that is as asymptotically efficient as the traditional 2-Step GMM estimator and has consistent asymptotic variance (Hansen, 2007; Lai & Small, 2007), given by:

$$\left\{\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial g(y_i, x_i, \beta)}{\partial \beta}\right)^T \left\{\frac{1}{N}\sum_{i=1}^{N} g(y_i, x_i, \tilde{\beta})\, g(y_i, x_i, \tilde{\beta})^T\right\}^{-1} \left(\frac{1}{N}\sum_{i=1}^{N}\frac{\partial g(y_i, x_i, \beta)}{\partial \beta}\right)\right\}^{-1},$$

where $\frac{\partial g(y_i,x_i,\beta)}{\partial \beta}$ is evaluated at $\beta = \hat{\beta}_{GMM}$ (Hansen, 2007).

In order to obtain GMM parameter estimates, this quadratic form is minimized over β (Lai & Small, 2007).

3 Measures to Assess Model Goodness-of-Fit

The natural step that follows statistical modeling involves the assessment of model goodness-of-fit. Section 3 outlines several exisiting statistics used to measure overall model fit. Several statistics are listed for assessing the goodness-of-fit of models in which observations were taken independently, as well as for assessing the fit of models constructed from correlated data. Lastly, the need for a new fit statistic is presented at the end of Section 3.

3.1 Existing Fit Statistics For Uncorrelated Data

3.1.1 Model Deviance

To assess the fit of any model based on independently observed data, the model deviance is commonly used. Deviance is a very useful tool in assessing the goodness-of-fit of a model because it can be used to: 1) measure the deviation of a specific model from the data, and 2) compare two nested models.

Regardless of the comparison – whether a candidate model is compared to the full data or two candidate models are compared – deviance is always calculated as the different in the log-likelihoods of the two models, the likelihood ratio (Agresti, 1990). It is the information not explained by a specified model. Moreover, the model deviance always follows a Chi-square distribution with degrees of freedom equal to the difference in the number of parameters of the two models being compared:

$$D(y;\hat{\mu_2}) - D(y;\hat{\mu_1}) = 2\left[L(\hat{\mu_2};y) - L(\hat{\mu_1};y)\right] \sim \chi^2_p \,,$$

where p denotes the difference in the number of parameters in the two models, and $L(\hat{\mu};y)$ is the maximum value of the log-likelihood under the given models (Agresti, 1990). In the case where a specific model is compared to the data, $p = N - k$, where N denotes the number of observations in the data, and k denotes the number of parameters in the specified model (Agresti, 1990).

3.1.2 Akaike's Information Criterion

To assess the fit of models derived using maximum likelihood estimation, researchers oftentimes rely on Akaike's Information Criterion (AIC), an information-based fit statistic proposed by Akaike (1973). AIC assumes no distribution; rather, it is a single value that is used descriptively to represent the amount of information lost from fitting a specific model to the data. It does not require the comparison of two nested models but rather compares a model to the actual data (Akaike, 1973). AIC is:

$$AIC = -2\,L(\beta;y) + 2k \,,$$

where $L(\beta;y)$ is the maximum of the log-likelihood of the specified model, and k is the number of parameters in the model, not including interaction terms.

AIC is an information-based theoretical criterion that can be used to compare nested models (Akaike, 1974). It can be used to compare models with different numbers of parameters. The process involves maximizing the likelihood function individually for all potential nested models being compared and choosing the model that gives the largest discrepancy between its maximized likelihood and the dimension of the model (Akaike, 1974; Schwarz, 1978). Better fitting models are associated with having smaller AIC values.

This statistic can be used to compare models with varying numbers of parameters because models with greater numbers of parameters are penalized more (by the term $2k$). Thus, AIC can be thought of as a fit statistic that almost always selects parsimonious models. It takes into account the value of the maximum of the log-likelihood, then adds the penalty to it; thus, the larger the value of the AIC, the farther the model deviates from the observed data. In the sense of multiple regression models, the value of AIC for models with different combinations of predictors can be used to select which model is most "ideal" (Akaike, 1974).

3.1.3 BIC

Another commonly used information criterion that measures the goodness-of-fit of a model is the Bayesian Information Criterion (BIC), proposed by Schwarz (1978):

$$BIC = -2\ L(\beta; y) + k \cdot ln(n)\ ,$$

where $L(\beta; y)$ is the likelihood derived from the data, k is the number of predictors in the model, and n is the number of observations in the data.

BIC is very similar to AIC in that it uses the maximum value of the log-likelihood of the model, and it penalizes for the number of parameters included in the model. The penalty for BIC (i.e., the term $k \cdot ln(n)$) is larger than that of the AIC because BIC penalizes a model for the number of parameters it includes, as well as for the number of observations in the data. The larger the number of observations in the data, the larger the value $ln(n)$ grows. For this reason, BIC tends to favor more parsimonious models than AIC (Schwarz, 1978).

3.2 Existing Fit Statistics For Correlated Data

3.2.1 QIC

When assessing the fit of models with parameters estimated using maximum likelihood estimation procedures, either the model deviance or information-based goodness-of-fit statistics, such as AIC and BIC, are used. These statistics required the use of the full likelihood of the response. When using the Generalized Estimating Equations approach, we use a quasi-likelihood instead of the full likelihood. Analogously, in assessing the fit of models with parameters estimated using GEE, the quasi-likelihood information criterion (QIC), proposed by Pan (2001), is commonly used:

$$QIC = -2Q(\hat{\beta}(R); I, Y) + 2 \cdot trace(\hat{\Omega}_I \hat{V}_r)\ ,$$

where Q is the quasi-likelihood of the response under the independence model (in this case), $\hat{\beta}$ is the vector of parameter estimates obtained using GEE, I is the independent correlation structure, Y is the vector of observed responses, and $\hat{\Omega}_I \hat{V}_r$ denotes the working covariance structure. The independent working correlation structure has been specified here because it has been shown in simulation studies that the working independence model shows best performance (Pan, 2001).

QIC can be used to select a working correlation structure in GEE. The QIC is calculated for various working correlation structures, and the one with smallest QIC is selected (Pan, 2001). Using this information criterion in this way, a small value of the QIC suggests that the specified working correlation is very close to the true correlation structure present in the data.

3.2.2 QICu

An alternative statistic to the QIC is also available. Pan (2001) shows that when all modeling specifications in GEE are correct, $\hat{\Omega}_I^{-1}$ is asymptotically equivalent to $\hat{V}_r$, and $trace(\hat{\Omega}_I \hat{V}_r) \approx trace(I) = p$, then the QIC is equivalent to the AIC (Pan, 2001) and can be estimated using:

$$QIC_u \equiv -2Q(\hat{\beta}(R); I, Y) + 2p \, ,$$

where, again, p is the number of parameters in the specified model.

QIC_u can be used for variable selection, but it cannot be used to select the working correlation structure (Pan, 2001). To select a model, the QIC_u should be calculated for all candidate models (i.e., various combinations of predictors), and the model with the smallest QIC_u is selected.

3.3 GMM and Model Goodness-of-Fit

The literature is sparse in its discussions involving the assessment of fit of model parameters estimated using Generalized Method of Moments when time-dependent covariates are present. Existing measures of fit – including deviance, AIC, BIC, QIC, and QICu – are not appropriate statistics in assessing the fit of models in which parameter estimates were obtained using GMM because these statistics were derived using either the full likelihood or the quasi-likelihood of the data, whereas GMM does not employ the use of any likelihood function.

The majority of discussions involving GMM and model fit revolve around the idea of overidentification of models (Newey, 1985; Hall, 1999; Andrews, 1999; Hansen, 1982; Hansen, Heaton, & Yaron, 1996). Overidentification results when more than the necessary number of moment conditions are used in the estimation of model parameters. However, this is not a primary interest in this study.

Useful results about the distribution of the minimum of the GMM quadratic form, $Q_N(\beta)$, is presented by Hansen, Heaton, and Yaron (1996). It has been shown that the minimand of the quadratic form multiplied by the number of subjects, N, follows a Chi-squared distribution with degrees of freedom equal to the difference between the number of moment conditions in the population and the number of estimated parameters (Hansen, Heaton, & Yaron, 1996):

$$min(Q_N(\beta)) \sim \chi^2(m-k), \tag{3}$$

where m denotes the number of moment conditions used in the estimation process, and k denotes the number of parameters in the model.

Results presented in the form of Equation (3) allow researchers to test hypotheses about the quadratic form, which gives insight as to whether or not "sufficient" moment conditions were used in the estimation process. However, this assesses the appropriate inclusion – or exclusion – of a set of moment conditions rather than assessing the overall fit of the model.

The results from Newey (1985) and Hall (1999) point in a similar direction. Hall (1999) establishes hypotheses tests for the overidentifying moment conditions based on the mathematical results presented by Newey (1985), but very little is discussed about the overall goodness-of-fit of the model.

Discussions involving a GOF-like statistic presented by Lai and Small (2007) introduce hypothesis test for time-dependent covariate type. The statistic used in this hypothesis test relies on the difference in minimands of quadratic forms under the two models:

$$C_N = N\left[Q_1(\hat{\beta}_{GMM}) - Q_2(\hat{\beta}_{GMM})\right] \sim \chi^2(r-q),$$

where Q_1 and Q_2 are the minimums of quadratic forms of two candidate models, and $r-q$ is the difference in dimensions of these models.

The main goal of our research is to establish a fit statistic – an information-based criterion, much like AIC or BIC – that has no distribution but is represented by a scalar to assess overall goodness-of-fit of models that were constructed using Generalized Method of Moments. Specifically, interest is geared toward establishing a statistic for the fit of models in which parameter estimates were obtained using GMM using data with at least one time-dependent covariate.

4 Statistics for Assessing Model Fit

Section 4 outlines the process for obtaining two statistics that can be used in conjunction with each other to assess the overall fit of models constructed using Generalized Method of Moments in the presence of time-dependent covariates. These methods rely neither on the likelihood, nor the quasi-likelihood, but rather on the use of moment conditions, as is the case with the GMM process.

4.1 GMM Model Fit Statistic

Currently, there is no universal method in assessing the goodness-of-fit of models in which parameters are estimated using Generalized Method of Moments in the presence of time-dependent covariates. One approach is to use the minimum of the quadratic form to assess the overall "goodness" of the model; the smaller the minimum of the quadratic form, the better the model.

Distributional results from Hansen, Heaton, and Yaron (1996), as well as Hall's (1999) discussion on non-nested models (§4.6), can be used to form hypothesis tests for overidentifying restrictions of moment conditions. However, these methods test whether or not an estimate β formed from a set of moment conditions deviates from β_0; it does not test overall model fit.

Additionally, there is no published work in the body of literature that suggests any of these methods were employed in the presence of time-dependent covariates. There is the need to establish a statistic that is based on moment conditions rather than on the specification of the full likelihood or quasi-likelihood functions. Further, there is a need for a more information-based fit statistic to assess the overall goodness-of-fit of models in which parameters are estimated using GMM in the presence of TDCs. Without such a statistic, it would be difficult to assess whether or not a model and its parameter estimates are reasonable or even meaningful.

4.2 Moment-Based Goodness-of-Fit Statistics

Two measures may be used in conjunction with each other to assess the goodness-of-fit of a model in which parameter estimates are obtained using Generalized Method of Moments in the presence of time-dependent covariates. This method relies on neither the likelihood function, nor the quasi-likelihood function; instead, it relies on the use of moment conditions and an established measure of distance. The first statistic utilizes the minimum of the quadratic form used in the GMM estimation process, and the second statistic is the Kullback-Leibler information statistic (Kullback & Leibler, 1951).

4.2.1 Minimum of the GMM Quadratic Form

Based on Hansen, Heaton, and Yaron's (1996) findings, the goodness-of-fit of two nested models, M_1 and M_2, can be compared using the statistic:

$$C = (minimum_{QF_1} - minimum_{QF_2}) \sim \chi^2_{(m)},$$

where m denotes the difference in the number of parameters between the two candidate models. In other words, the difference between the minimum of the GMM quadratic forms of two competing models follows a Chi-squared distribution with degrees of freedom equal to the difference between the number of parameters in the two models (Hansen, Heaton, & Yaron, 1996).

This statistic, C, can be used to test the hypothesis:

H_0: The candidate model with less parameters is sufficient in explaining the variation in the response.

H_A: The candidate model with less parameters is *not* sufficient in explaining the variation in the response.

If the true model, M_0, is known and fully specified using the full data, the statistic C can be used to assess whether or not an alternate model with less parameters is adequate in explaining the variability present in the response, due to trivial deviation from the full model (Hansen, Heaton, & Yaron, 1996; Hall, 1999).

4.2.2 Kullback-Leibler Information Criterion

The Kullback-Leibler information based measure of distance can be used as a measure to assess the discrepancy of a candidate model from the "null" model – i.e, the data (Kullback & Leibler, 1951; Csiszár, 1975; White, 1982).

Using an estimation process such as the 2-Step Generalized Method of Moments (Lai & Small, 2007), a set of parameter estimates, $\hat{\beta}$, is obtained. This $\hat{\beta}$ is used to calculate an estimate for the Kullback-Leibler Information Criterion (KLIC):

$$\widehat{KLIC} = \min_{\gamma} \left(\frac{1}{N} \sum_{i=1}^{N} e^{\gamma^T \, g_i(y_i, \hat{\beta})} \right),$$

where γ is a vector of unknown parameters.

The Kullback-Leibler information statistic is similar to other measures of goodness-of-fit, such as AIC, BIC, QIC, and QICu, in that it is represented by a single number and follows no distribution; therefore, no hypothesis test can be formed using the Kullback-Leibler information based measure of distance as a basis for inferential conclusions.

Rather, the Kullback-Leibler information statistic can be used to compare nested models. The candidate model with the smallest value of the Kullback-Leibler information criterion is selected.

5 Conclusions

This paper highlighted the need for an information-based criterion, similar to Akaike's Information Criterion or the Quasi-likelihood Information Criterion, that can be used to assess the overall goodness-of-fit of models in which parameter estimates are obtained using Generalized Method of Moments in the presence of at least one time-dependent covariate. Because the algorithm behind GMM employs the use of moment conditions rather than the likelihood or quasi-likelihood functions, there is a need for a measure of fit that is derived from moment conditions.

In Section 4, two statistics were proposed as methods to assess the overall goodness-of-fit of models constructed using Generalized Method of Moments estimation: the first approach utilizes the difference in minimums of the GMM quadratic forms, and the second method uses the Kullback-Leibler information criterion. Neither method relies on the derivation of the likelihood or the quasi-likelihood; both rely on the use of moment conditions, similar to the process behind GMM.

In theory, the two methods should yield results that agree; the candidate model with the smallest value of the estimated Kullback-Leibler information criterion should also be the most parsimonious model that does not deviate significantly from the full data, in terms of accounting for response variability.

Some alternatives may exist for assessing the fit of correlated models, but more work is necessary to establish a statistic that is universally agreed upon.

References

Andrews, D. W. K. (1999). Consistent Moment Selection Procedures for Generalized Method of Moments Estimation. *Econometrica, 67*(3), 543-564.

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. *Proceedings of the Second International Symposium on Information Theory*, 267-281.

Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control, 19,*(6), 716-723.

Agresti, A. (1990). *Categorical Data Analysis*. Barnett, V., Bradley, R. A., Fisher, N. I., Hunter, J. S., Kadane, J. B., Kendall, D. G., Smith, A. F. M., Stigler, S. M., Teugels, J., & Watson, G. S. (Eds.). Hoboken, NJ: John Wiley & Sons, Inc.

Csiszár, I (1975). I-Divergence Geometry of Probability Distributions and Minimization Problems. *The Annals of Probability, 3*(1), 146-158.

Department of Economics, University of Chicago. (2007). *Generalized Method of Moments Estimation*. Chicago, IL: Hansen, L. P.

Diggle, P. J., Heagerty, P., Liang, K-Y., & Zeger, S. L. (2002). *Analysis of Longitudinal Data*. Atkinson, A. C., Copas, J. B., Pierce, D. A., Schervish, M. J., Titterington, D. M., Carroll, R. J., Hand, D. J., & Smith, R. L. (Eds.). New York, NY: Oxford University Press Inc.

Dobson, A. J. & Barnett, A. G. (2008). *An Introduction to Generalized Linear Models*. Carlin, B. P., Faraway, J. J., Tanner, M., & Zidek, J. (Eds.) Boca Raton, FL: Chapman and Hall, CRC Press.

Fitzmaurice, G. M. (1995). A Caveat Concerning Independence Estimating Equations With Multivariate Binary Data. *Biometrics, 51*, 309-317.

Hall, A. R. (1999). Hypothesis Testing in Models Estimated by GMM. *Generalized Method of Moments Estimations*, 96-127.

Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica, 50*(4), 1029-1054.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite Sample Properties of Some Alternative GMM Estimators. *Journal of Business & Economic Statistics, 14*(3), 262-280.

Kullback, S. & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics, 22*(1), 79-86.

Liang, K-Y. & Zeger, S. L. (1986). Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika, 73*(1), 13-22.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*. Cox, D. R., Hinkley, D. V., Reid, N., Rubin, D. B., & Silverman, B. W. (Eds.). Boca Raton, FL: Chapman and Hall, CRC Press.

Neuhaus, J. M. & Kalbfleisch, J. D. (1998). Between- and Within-Cluster Covariate Effects in the Analysis of Clustered Data. *Biometrics, 54*(2), 638-645.

Newey, W. K. (1985). Generalized Method of Moments Specification Testing. *Journal of Econometrics, 29*, 229-256.

Pan, W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics,* 57, 120-125.

Pepe, M. S. & Anderson, G. L. (1994). A Cautionary Note on Inference for Marginal Regression Models with Longitudinal Data and General Correlated Response Data. *Communication in Statistics – Simulation and Computation, 23*(4), 939-951.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics, 6*(2), 461-464.

White, H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica, 50*(1), 1-25.

Zeger, S. L. & Liang, K-Y. (1992). An Overview of Methods for the Analysis of Longitudinal Data. *Statistics in Medicine, 11*, 1825-1839.

Zeger, S. L. & Liang, K-Y. (1986). Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics,* 42, 121-130.