Bayesian Analysis (2009) 4, Number 2, pp. 191–208

Bayesian Generalized Method of Moments

Guosheng Yin*

Abstract. We propose the Bayesian generalized method of moments (GMM), which is particularly useful when likelihood-based methods are difficult. By deriving the moments and concatenating them together, we build up a weighted quadratic objective function in the GMM framework. As in a normal density function, we take the negative GMM quadratic function divided by two and exponentiate it to substitute for the usual likelihood. After specifying the prior distributions, we apply the Markov chain Monte Carlo procedure to sample from the posterior distribution. We carry out simulation studies to examine the proposed Bayesian GMM procedure, and illustrate it with a real data example.

Keywords: Bayesian inference; Correlated data; Estimation efficiency; Generalized estimating equation; Generalized linear model; Gibbs sampling; Posterior distribution

1 Introduction

Bayesian methods follow Bayes' theorem and the likelihood principle. Given the specified prior distribution, statistical inferences are based on the posterior distribution of the model parameters. In biomedical applications, multivariate data often arise due to clustering or longitudinal measurements. For example, in studies with paired eyes or multiple teeth, observations from the same subject are naturally clustered, and thus do not satisfy the independence assumption. In longitudinal studies, although different subjects are assumed to be independent, the common procedure of repeatedly measuring each subject over time induces correlations among the observations on the same subject. It is typically difficult to obtain the likelihood function of such correlated data, because the underlying correlation structure is unknown. Often a random effects model can be undertaken, although the Bayesian posterior estimates might be sensitive to the usual parametric distributional assumptions for the unobservable random effects. Random effects models have been extensively studied by introducing subject-specific random effects to account for the correlation; for example, see Laird and Ware (1982) and Ware (1985). Conditional on the random effects, the observations are assumed to be independent. Likelihood-based inferences can be easily derived if a parametric distribution is assumed for the random effects, which, however, might be sensitive to the specified distribution. On the other hand, the underlying correlation can be treated as a nuisance if the population-average covariate effects are of primary interest. The generalized estimating equation (GEE) represents a robust method that produces consistent and asymptotic normal estimators even with a misspecified working correlation matrix. If the correlation structure is correctly specified, the GEE estimator is efficient

*Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX, mailto:gsyin@mdanderson.org

 DOI:10.1214/09-BA407

within the linear estimating function family (Liang and Zeger 1986; Zeger and Liang 1986; Zeger et al. 1988).

The generalized method of moments (GMM) has been extensively studied in econometrics (Hansen 1982; Newey and West 1987; Pakes and Pollard 1989; Lee 1996; Hansen et al. 1996; Newey 2004; Hall 2005). The GMM is particularly attractive and useful for improving the estimation efficiency when the likelihood formulation is difficult but the moment conditions are relatively easy to obtain. Hansen (1982) first established a comprehensive framework for the GMM and provided a rigorous justification and asymptotic theories for the estimator. Advances in computing technology have facilitated a growing interest in simulation methods for parameter estimation based on the population moment conditions (McFadden 1989). Pakes and Pollard (1989) derived asymptotic theories for the simulated method of moments when the function in the moment condition might be discontinuous. Qu et al. (2000) and Lai and Small (2007) proposed a GMM marginal regression to analyze longitudinal data, and showed its advantages over the GEE. One of the major uses of the GMM is to make inferences in semiparametric models where there are more moment conditions than unknown parameters. By combining the moments, the GMM is parsimonious and useful for constructing efficient estimators, particularly when the efficiency bound is complicated and moment conditions are available. Under some regularity conditions in Chamberlain (1987), the GMM estimator achieves the semiparametric efficiency bound in the sense of Bickel et al. (1993).

In the Bayesian paradigm, posterior inference follows the likelihood principle. Given a prior distribution $\pi(\boldsymbol{\beta})$ for an unknown parameter $\boldsymbol{\beta}$, we can derive its posterior distribution as

$$\pi(\boldsymbol{\beta}|\mathbf{y}) \propto L(\mathbf{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta}),$$

where $L(\mathbf{y}|\boldsymbol{\beta})$ is the likelihood function. However, if there is not enough information for the likelihood function, the Bayesian posterior estimation and inferences can be challenging (Zellner et al. 1997; Zellner 1997; Kim 2002; Chernozhukov and Hong 2003). Zellner (1997) proposed the Bayesian method of moments by computing the maximum entropy densities consistent with the moment conditions. Kim (2002) derived the limited information likelihood by minimizing the Kullback-Leibler information criterion distance. Chernozhukov and Hong (2003) studied a Laplace-type estimator obtained by a Markov chain Monte Carlo (MCMC) approach. For most clustered or longitudinal data, the likelihood is typically difficult to obtain without knowing the underlying correlation structure. We propose the Bayesian GMM such as to circumvent the difficulty of constructing the likelihood function. The moments typically converge to zero-mean normal distributions, based on which we can construct a substitute for the true likelihood in the asymptotic sense. We take the negative GMM quadratic function divided by two and then exponentiate it to substitute for the likelihood. Moreover, without the need to specify the correlation structure, we take a linear expansion of the inverse of the correlation matrix over a set of commonly used basis matrices, so that the Bayesian GMM estimators are properly adjusted for the correlation. The Bayesian GMM still produces valid estimates and inferences even if the underlying correlation is misspecified (for example, the working independence model). If certain information on the correla-

tion structure is available, we can incorporate the additional moment conditions into the GMM quadratic function to improve the estimation efficiency. Through the MCMC procedure, we can easily obtain the posterior estimates.

The rest of the article is organized as follows. In Section 2, we introduce the notation, and propose the Bayesian GMM in the generalized linear model framework, as well as with correlated data. In Section 3, we examine the properties of the Bayesian GMM using simulation studies, in particular, we justify its use based on the posterior probability coverage sets. We illustrate the proposed methods with a real data example in Section 4, and give concluding remarks in Section 5.

2 Bayesian Generalized Method of Moments

2.1 Generalized Linear Model

The generalized linear models (GLM) provide a unified framework for various discrete and continuous outcomes (McCullagh and Nelder 1989). For the ith subject $(i = 1, \ldots, n)$, we observe y_i as the outcome of interest and $\mathbf{Z}_i$ as the corresponding covariate vector. To characterize the relationship between y_i and $\mathbf{Z}_i$, we assume that the observed values y_i are from a distribution belonging to the exponential family. The density function of y_i given $\mathbf{Z}_i$ takes the form of

$$f(y_i|\mathbf{Z}_i) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right\},$$

where θ_i is a location parameter, ϕ is a scalar dispersion parameter, and $a_i(\cdot)$, $b(\cdot)$ and $c(\cdot)$ are known functions. Typically, $a_i(\phi) = \phi/w_i$ with known weights w_i, and the linear predictor $\eta_i = \boldsymbol{\beta}^T\mathbf{Z}_i$ can be linked with θ_i through a monotone differentiable function $h(\cdot)$, i.e., $\theta_i = h(\eta_i)$. This is a standard formulation of the GLM, with $\mu_i = E(y_i|\mathbf{Z}_i) = b'(\theta_i)$ and $v_i = \text{var}(y_i|\mathbf{Z}_i) = b''(\theta_i)a_i(\phi)$, where $b'(\cdot)$ and $b''(\cdot)$ are the first and second derivatives, respectively. The quasi-likelihood estimator can be obtained by solving the score-type equation

$$\sum_{i=1}^{n} \mathbf{D}_i v_i^{-1}(y_i - \mu_i) = 0,$$

where $\mathbf{D}_i = \partial\mu_i/\partial\boldsymbol{\beta}$. See Wedderburn (1974) and McCullagh (1983) for details.

In the GMM framework, we define

$$\mathbf{u}_i(\boldsymbol{\beta}) = \mathbf{D}_i v_i^{-1}(y_i - \mu_i), \quad i = 1, \ldots, n.$$

Thus, we have the population moment condition

$$E\{\mathbf{u}_i(\boldsymbol{\beta})\} = 0,$$

with the corresponding sample moment condition

$$\mathbf{U}_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{u}_i(\boldsymbol{\beta}).$$

The GMM estimator $\widehat{\boldsymbol{\beta}}$ is obtained by minimizing the following quadratic objective function

$$Q_n(\boldsymbol{\beta}) = \mathbf{U}_n^T(\boldsymbol{\beta})\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\beta})\mathbf{U}_n(\boldsymbol{\beta}),$$

where $\boldsymbol{\Sigma}_n(\boldsymbol{\beta})$ is the empirical variance-covariance matrix given by

$$\boldsymbol{\Sigma}_n(\boldsymbol{\beta}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathbf{u}_i(\boldsymbol{\beta})\mathbf{u}_i^T(\boldsymbol{\beta}) - \frac{1}{n}\mathbf{U}_n(\boldsymbol{\beta})\mathbf{U}_n^T(\boldsymbol{\beta}).$$

In general, $\widehat{\boldsymbol{\beta}}$ is computed via a two-stage iterative procedure:

(1) Insert an initial value $\boldsymbol{\beta}^{(0)}$ into $\boldsymbol{\Sigma}_n(\boldsymbol{\beta})$.

(2) At the kth iteration, obtain the estimator $\widehat{\boldsymbol{\beta}}^{(k)}$ by minimizing

$$Q_n^{(k)}(\boldsymbol{\beta}) = \mathbf{U}_n^T(\boldsymbol{\beta})\boldsymbol{\Sigma}_n^{-1}(\widehat{\boldsymbol{\beta}}^{(k-1)})\mathbf{U}_n(\boldsymbol{\beta})$$

with respect to $\boldsymbol{\beta}$ while fixing $\boldsymbol{\Sigma}_n(\widehat{\boldsymbol{\beta}}^{(k-1)})$ as known.

(3) Plug the estimator $\widehat{\boldsymbol{\beta}}^{(k)}$ back into $\boldsymbol{\Sigma}_n(\widehat{\boldsymbol{\beta}}^{(k)})$, and move to the $(k+1)$th iteration.

(4) Continue this procedure until some prespecified convergence criteria are met.

Under certain regularity conditions in Hansen (1982), the GMM estimator $\widehat{\boldsymbol{\beta}}$ exists and converges in probability to the true parameter $\boldsymbol{\beta}_0$, and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges in distribution to a multivariate normal distribution. For cross-sectional data, the usual GMM estimator of $\boldsymbol{\beta}$ is equivalent to the quasi-likelihood estimator of $\boldsymbol{\beta}$, as the dimension of $\mathbf{U}_n(\boldsymbol{\beta})$ is equal to the dimension of $\boldsymbol{\beta}$.

The objective function $Q_n(\boldsymbol{\beta})$ follows a chi-squared distribution when evaluated at $\boldsymbol{\beta}_0$ or $\widehat{\boldsymbol{\beta}}$. Intuitively, the corresponding chi-squared tests are closely related to the usual likelihood ratio tests, and $Q_n(\boldsymbol{\beta})$ behaves like $-2\log\{L(\mathbf{y}|\boldsymbol{\beta})\}$, where $L(\mathbf{y}|\boldsymbol{\beta})$ is the likelihood function. Note that the sample moment typically converges to a multivariate normal distribution, as $n \to \infty$, $\mathbf{U}_n(\boldsymbol{\beta}_0) \sim N(0, \boldsymbol{\Sigma}(\boldsymbol{\beta}_0))$, where $\boldsymbol{\beta}_0$ is the true parameter and $\boldsymbol{\Sigma}(\boldsymbol{\beta}_0)$ is the limit of $\boldsymbol{\Sigma}_n(\boldsymbol{\beta}_0)$. Therefore, we can construct a pseudo-likelihood function $\widetilde{L}(\mathbf{y}|\boldsymbol{\beta})$ to replace the original likelihood function $L(\mathbf{y}|\boldsymbol{\beta})$ which may be difficult to derive, where

$$\widetilde{L}(\mathbf{y}|\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2}Q_n(\boldsymbol{\beta})\right\} = \exp\left\{-\frac{1}{2}\mathbf{U}_n^T(\boldsymbol{\beta})\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\beta})\mathbf{U}_n(\boldsymbol{\beta})\right\}.$$

As in the usual MCMC procedure, we can derive the posterior distribution based on $\widetilde{L}(\mathbf{y}|\boldsymbol{\beta})$. Given the prior distribution $\pi(\boldsymbol{\beta})$, the posterior distribution of $\boldsymbol{\beta}$ is

$$\widetilde{\pi}(\boldsymbol{\beta}|\mathbf{y}) \propto \widetilde{L}(\mathbf{y}|\boldsymbol{\beta})\pi(\boldsymbol{\beta}).$$

The MCMC algorithm can then be used to sample from $\widetilde{\pi}(\boldsymbol{\beta}|\mathbf{y})$ to obtain the posterior inference for $\boldsymbol{\beta}$. Note that $\widetilde{L}(\mathbf{y}|\boldsymbol{\beta})$ is constructed from a multivariate normal distribution for $\mathbf{U}_n(\boldsymbol{\beta})$ asymptotically, except that the normalizing term $(2\pi)^{-p/2}|\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\beta})|^{1/2}$ is not involved, where p is the vector length of $\mathbf{U}_n(\boldsymbol{\beta})$.

The key feature of the Bayesian GMM is that it is a moment-based approach, so that we can circumvent the difficulties of deriving the likelihood when moments are relatively easier to obtain. In addition, the Bayesian GMM is more robust as the likelihood may be vulnerable to certain parametric assumptions. However, the posterior distribution of the model parameters is complicated, and typically is not log-concave, thus we use the adaptive rejection Metropolis sampling algorithm within the Gibbs sampler proposed by Gilks et al. (1995).

2.2 Correlated Data

In the presence of correlation, the statistical analysis can be quite challenging as the underlying correlations need to be properly adjusted for valid inferences. Methodological development for analyzing correlated data has been greatly advanced. Let $(\mathbf{y}_1, \ldots, \mathbf{y}_n)$ be independent vectors of the response variables with means $(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n)$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iK_i})^T$ for the ith cluster of size K_i, $i = 1, \ldots, n$. Let the mean $\boldsymbol{\mu}_i = E(\mathbf{y}_i|\mathbf{Z}_i)$, and the variance of $\mathbf{y}_i$ be given by $\mathbf{V}_i = \zeta h(\boldsymbol{\mu}_i)$, where $h(\cdot)$ is the variance function and ζ is the scale parameter. The mean μ_{ik} is linked with the covariate vectors $\mathbf{Z}_{ik}$ through

$$\eta(\mu_{ik}) = \boldsymbol{\beta}^T \mathbf{Z}_{ik}, \quad k = 1, \ldots, K_i; \ i = 1, \ldots, n.$$

Let $\mathbf{A}_i$ be the diagonal matrix of the marginal variance of $\mathbf{y}_i$, let $\mathbf{R}_i$ be the true correlation matrix, and let $\mathbf{C}_i$ be the working correlation matrix which may not be identical to $\mathbf{R}_i$. The generalized estimating equation (GEE) is given by

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{D}_i{}^T \mathbf{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i) = 0, \tag{1}$$

where $\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{C}_i\mathbf{A}_i^{1/2}$ and $\mathbf{A}_i^{1/2} = \text{diag}\{h^{1/2}(\boldsymbol{\mu}_i)\}$. Under certain regularity conditions given in Liang and Zeger (1986), the estimator solved from (1), denoted by $\widehat{\boldsymbol{\beta}}$, is consistent and asymptotically follows a normal distribution. That is, as $n \to \infty$,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Gamma}),$$

where $\boldsymbol{\beta}_0$ is the true parameter value, $\boldsymbol{\Gamma}$ is the limit of $\mathbf{H}_1^{-1}\mathbf{H}_2\mathbf{H}_1^{-1}$, and

$$\begin{aligned}
\mathbf{H}_1 &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{D}_i{}^T \mathbf{V}_i^{-1}\mathbf{D}_i, \\
\mathbf{H}_2 &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{D}_i{}^T \mathbf{V}_i^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_i)(\mathbf{y}_i - \boldsymbol{\mu}_i)^T \mathbf{V}_i^{-1}\mathbf{D}_i.
\end{aligned}$$

The consistent estimator of $\boldsymbol{\Gamma}$, referred to as the sandwich or robust variance estimator, is obtained by evaluating the matrices $\mathbf{H}_1$ and $\mathbf{H}_2$ at their empirical estimates.

Maximum likelihood methods usually depend on the assumption of the underlying distribution, which may cause bias or efficiency loss if the distribution is misspecified. When observations in the same cluster are correlated, the estimator based on the working independence model, although still consistent, may not be efficient, as it completely ignores the correlation information. A natural way to enhance the estimation efficiency is to incorporate a weight matrix to account for the within-cluster correlation. However, the true correlation matrix $\mathbf{R}$ often has a complicated structure and is typically unknown in real applications. To circumvent the direct estimation of $\mathbf{R}$, Qu et al. (2000) proposed a GMM approach for the marginal regression model. Following the same route, we can linearly expand the inverse of $\mathbf{R}$, $\mathbf{R}^{-1}(\boldsymbol{\alpha})$, as

$$\mathbf{R}^{-1}(\boldsymbol{\alpha}) = \sum_{j=1}^{J} \alpha_j \mathbf{C}_{(j)},$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)$ are unknown constants and $(\mathbf{C}_{(1)}, \ldots, \mathbf{C}_{(J)})$ are a set of known basis matrices. For example, if the cluster size $K = 4$, then $\mathbf{C}_{(j)}$ can be the identity matrix $\mathbf{I}$, or

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \ldots.$$

The linear span of $\mathbf{C}_{(j)}$ should accommodate or adequately approximate the true correlation structure. The estimating equation then becomes

$$\frac{1}{n} \sum_{i=1}^{n} \mathbf{D}_i{}^T \mathbf{A}_i^{-1/2} (\alpha_1 \mathbf{C}_{(1)} + \ldots + \alpha_J \mathbf{C}_{(J)}) \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i) = 0.$$

Regardless of the complexity of $\mathbf{R}^{-1}(\boldsymbol{\alpha})$, it can be represented by a linear combination of some commonly used basis matrices. For example, if $\mathbf{R}(\boldsymbol{\alpha})$ is an exchangeable matrix, then $\mathbf{R}^{-1}(\boldsymbol{\alpha}) = \alpha_1 \mathbf{C}_{(1)} + \alpha_2 \mathbf{C}_{(2)}$, where $\mathbf{C}_{(1)} = \mathbf{I}$ and $\mathbf{C}_{(2)}$ is of the form with 0 on the diagonal and 1 elsewhere; and if $\mathbf{R}(\boldsymbol{\alpha})$ is a first-order autoregressive AR(1) correlation matrix, $\mathbf{R}^{-1}(\boldsymbol{\alpha}) = \alpha_1 \mathbf{C}_{(1)} + \alpha_2 \mathbf{C}_{(2)} + \alpha_3 \mathbf{C}_{(3)}$, where $\mathbf{C}_{(1)} = \mathbf{I}$, $\mathbf{C}_{(2)}$ is of the sandwich form with two main off-diagonals of 1 and 0 elsewhere, and $\mathbf{C}_{(3)}$ is a matrix with 1 at the corners of $(1, 1)$ and (K, K), and 0 elsewhere. The coefficient α_j takes an implicit role as a weighting scheme that can automatically adjust for the importance and contribution of each $\mathbf{C}_{(j)}$ to the entire correlation structure. The unknown coefficients, the α_j's, can take any real values; but they do not need to be sampled in the Bayesian GMM procedure.

We split the moment conditions corresponding to each $\mathbf{C}_{(j)}$, $j = 1, \ldots, J$. In this case, there are more estimating equations than unknown parameters if $J > 1$. We construct a $(J \times p)$-vector $\mathbf{U}_n(\boldsymbol{\beta})$ by concatenating the J moment conditions, and building

them into a quadratic objective function. Therefore, we first split and concatenate the population moment condition to a $(J \times p)$-vector:

$$E\{\mathbf{u}_i(\boldsymbol{\beta})\} = E\begin{pmatrix} \mathbf{D}_i{}^T\mathbf{A}_i^{-1/2}\mathbf{C}_{(1)}\mathbf{A}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i) \\ \vdots \\ \mathbf{D}_i{}^T\mathbf{A}_i^{-1/2}\mathbf{C}_{(J)}\mathbf{A}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i) \end{pmatrix} = 0.$$

The corresponding sample moment condition is given by

$$\mathbf{U}_n(\boldsymbol{\beta}) = \frac{1}{n}\begin{pmatrix} \sum_{i=1}^n \mathbf{D}_i{}^T\mathbf{A}_i^{-1/2}\mathbf{C}_{(1)}\mathbf{A}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i) \\ \vdots \\ \sum_{i=1}^n \mathbf{D}_i{}^T\mathbf{A}_i^{-1/2}\mathbf{C}_{(J)}\mathbf{A}_i^{-1/2}(\mathbf{y}_i - \boldsymbol{\mu}_i) \end{pmatrix}.$$

We then follow the Bayesian GMM development to obtain the substituted likelihood

$$\widetilde{L}(\mathbf{y}|\boldsymbol{\beta}) \propto \exp\left\{-\frac{1}{2}\mathbf{U}_n(\boldsymbol{\beta})^T\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\beta})\mathbf{U}_n(\boldsymbol{\beta})\right\},$$

where

$$\boldsymbol{\Sigma}_n(\boldsymbol{\beta}) = \frac{1}{n^2}\sum_{i=1}^n \mathbf{u}_i(\boldsymbol{\beta})\mathbf{u}_i^T(\boldsymbol{\beta}) - \frac{1}{n}\mathbf{U}_n(\boldsymbol{\beta})\mathbf{U}_n^T(\boldsymbol{\beta}).$$

We can specify the prior distribution for $\boldsymbol{\beta}$ and then sample from the posterior distribution accordingly.

Table 1: Comparisons between the Bayesian GMM and the Bayesian likelihood-based estimation under the generalized linear models.

		Bayesian GMM						Bayesian Likelihood					
		$\beta_0 = 0.2$		$\beta_1 = 0.5$		$\beta_2 = -0.5$		$\beta_0 = 0.2$		$\beta_1 = 0.5$		$\beta_2 = -0.5$	
Model	n	Est	SD	Est	SD	Est	SD	Est	SD	Est	SD	Est	SD
Linear	200	.161	.050	.528	.034	−.324	.077	.161	.050	.529	.033	−.327	.071
	500	.191	.032	.511	.022	−.465	.046	.190	.032	.512	.022	−.464	.045
	1000	.199	.022	.502	.015	−.478	.033	.198	.022	.502	.016	−.477	.032
Logistic	200	.402	.216	.382	.166	−.505	.312	.401	.208	.369	.156	−.496	.297
	500	.196	.129	.406	.098	−.481	.190	.194	.130	.399	.096	−.474	.187
	1000	.229	.090	.437	.071	−.580	.132	.224	.091	.432	.069	−.571	.131
Poisson	200	.272	.094	.389	.064	−.544	.143	.284	.088	.380	.063	−.537	.131
	500	.184	.056	.452	.040	−.511	.088	.188	.057	.451	.044	−.509	.089
	1000	.216	.040	.464	.034	−.544	.063	.220	.039	.464	.031	−.546	.063

One simulated data set, Est is the posterior mean and SD is the posterior standard deviation of 10,000 posterior samples.

3 Simulation Studies

3.1 Generalized Linear Models

We carried out simulation studies to examine the performance of the proposed Bayesian GMM estimation and inference procedures. We first considered the GLM framework that includes the linear, logistic and Poisson models. Under the linear regression model,

$$y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \epsilon,$$

we took the true parameter values $\beta_0 = 0.2$, $\beta_1 = 0.5$, $\beta_2 = -0.5$, and $\epsilon \sim N(0, \sigma^2)$, a zero-mean normal distribution with variance $\sigma^2 = 0.25$. The covariate Z_1 was generated from the standard normal distribution, and Z_2 was a binary variable taking a value of 0 or 1 with probability 0.5. We took sample sizes of $n = 200$, 500 and 1,000. We examined the performance of the proposed Bayesian GMM and also implemented the usual Bayesian likelihood-based method for comparison. As the likelihood can be easily derived in these cases, it serves as a benchmark for numerical comparison. In particular, under the linear model, we assumed that σ was known in the Bayesian likelihood procedure, i.e., in the Gibbs sampling, we only took the posterior samples of the β's while fixing σ at the true value. This would make the same number of unknown parameters for the Bayesian GMM and the Bayesian likelihood-based method, as the Bayesian GMM takes σ as a nuisance parameter which does not appear in the posterior distribution. We simulated one data set for each sample size considered, and took 10,000 posterior samples after the burn-in period of 500 iterations for the posterior inference. Similarly, we considered the logistic regression model in the form of

$$\text{logit}(p) = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2,$$

under which the outcome y was simulated as a binary variable taking a value of 1 with probability p, or 0 with probability $1-p$. Moreover, under the Poisson log-linear model,

$$\log(\mu) = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2,$$

we simulated y as a Poisson variable with mean μ. In the logistic and Poisson models, the true values of the regression parameters were the same as those in the linear regression model, and the covariates were generated similarly as well, i.e., $Z_1 \sim N(0,1)$ and $Z_2 \sim$ Bernoulli(0.5). We took noninformative priors for all of the parameters such that the posterior estimation was dominated by the data. In particular, we assigned each β a prior distribution of $N(0, 10,000)$.

In Table 1, we summarize the posterior estimates, and compare the results between the Bayesian GMM and Bayesian likelihood-based estimation procedures. For each configuration, we report the posterior means and posterior standard deviations of the regression parameters. When the sample size is relatively small ($n = 200$), we can see that there are certain differences in the posterior estimates between the two methods, and that the differences diminish and become negligible as the sample size increases. The posterior means for the β's using the Bayesian GMM are quite close to the true parameter values, especially for large sample sizes, and the corresponding posterior

standard deviations decrease with increasing sample sizes. Overall, there are no notable differences in the estimation results obtained from the Bayesian GMM and Bayesian likelihood method.

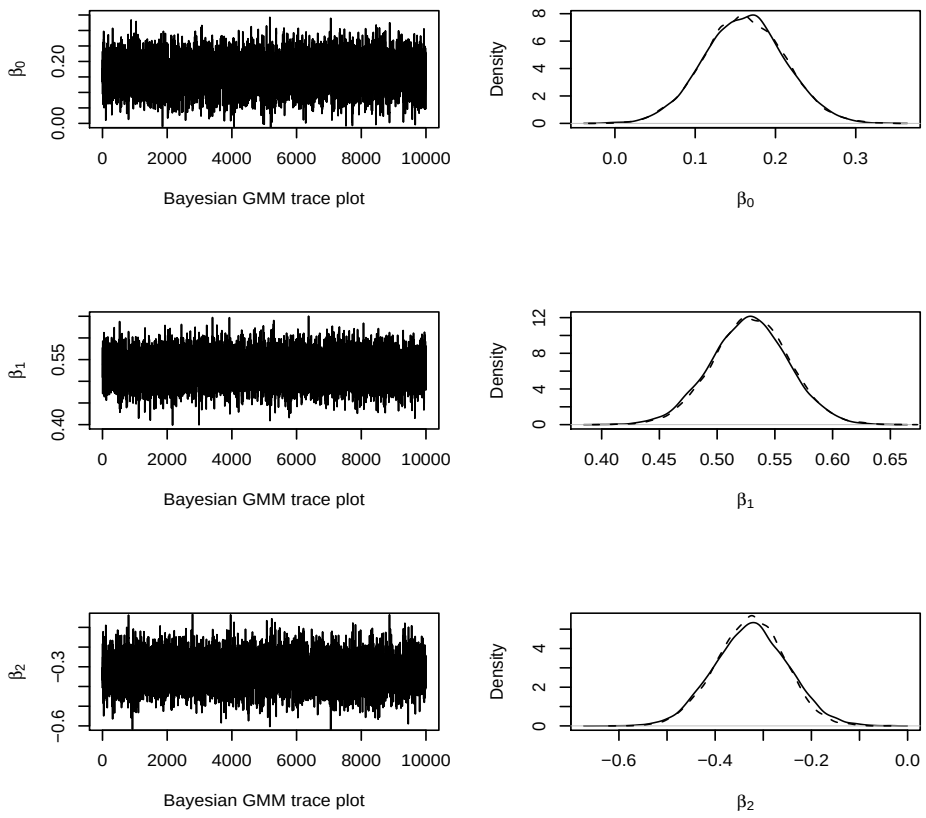

Figure 1: Posterior trace plots using the Bayesian GMM, and posterior density plots using the Bayesian GMM (——) and likelihood method (- - - -) for the linear regression model with $n = 200$.

To examine the convergence of the Markov chains obtained by the Bayesian GMM, we show the trace plots in Figure 1 for the linear model case with $n = 200$, side by side with the posterior densities of each parameter using the Bayesian GMM and the Bayesian likelihood-based method. We can see that the chains converged fast, mixed well and appeared to be stationary for each β. Furthermore, we conducted more rigorous convergence diagnostics for these Markov chains using the methods recommended by Cowles and Carlin (1996). For example, based on the Z-score by Geweke (1992), we found that all the Z-scores were not significant, which took values of 0.288, 0.884, and -0.162 for β_0, β_1 and β_2, respectively. Thus, the convergence of the Markov chains

was satisfactory. In addition, the posterior densities using the Bayesian GMM and the Bayesian likelihood method are very similar, which indicates that the moment-based approach can adequately recover the information in the likelihood.

3.2 Posterior Probability Coverage Set

We further examined the validity of the posterior distribution and inference based on the Bayesian GMM using the probability coverage of posterior sets suggested by Monahan and Boos (1992). Their method is intuitive and numerically convenient, which has been used to justify the Bayesian empirical likelihood (Lazar 2003). For ease of exposition, we consider a single parameter β. If the "likelihood" produces valid Bayesian inferences, then for any prior $\pi(\beta)$, posterior credibility sets which are supposed to contain posterior probability α should in fact contain the true β with proportion α when β is generated from $\pi(\beta)$ and the data are generated from the true density function of the data given β. Our proposed alternative likelihood $\tilde{L}(\mathbf{y}|\beta)$ would be justified in terms of probability coverage if and only if the posterior $\tilde{\pi}(\beta|\mathbf{y}) \propto \tilde{L}(\mathbf{y}|\beta)\pi(\beta)$ is valid by coverage for every continuous prior distribution on β. That is, if we define

$$H = \int_{-\infty}^{\beta} \tilde{\pi}(s|\mathbf{y})ds,$$

H should follow a uniform distribution on (0, 1); see Monahan and Boos (1992). Under the linear regression model with $n = 200$, we took the normal prior distributions for the β's, i.e., $\beta_0 \sim N(0.2, 25)$, $\beta_1 \sim N(0.5, 25)$ and $\beta_0 \sim N(-0.5, 25)$, respectively. We replicated 1,000 simulations, and for each data set, we took 10,000 posterior samples with 100 burn-in iterations, based on which we computed the statistic H. In Figure 2, we show the quantile-quantile (q-q) plots for the H statistics versus the quantiles of Uniform(0, 1), using the Bayesian GMM and Bayesian likelihood method, respectively. The q-q plot for each β is closely matching with the diagonal line. In addition, we do not see any difference between the q-q plots using the Bayesian GMM and those using the Bayesian likelihood method. This demonstrates that our Bayesian GMM can serve as a valid alternative for the Bayesian likelihood method.

3.3 Longitudinal Data

Repeated measurements are common in longitudinal studies, and thus we examined the frequentist properties of the Bayesian GMM under the linear regression model for such data. For ease of exposition, we took all of the clusters to be of the same size, i.e. $K_i \equiv K = 4$ for $i = 1, \ldots, n$. We considered the linear regression model with correlated errors given by

$$y_{ik} = \beta_0 + \beta_1 Z_{1ik} + \beta_2 Z_{2ik} + \epsilon_{ik}, \quad i = 1, \ldots, n; \; k = 1, \ldots, K,$$

where $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iK})^T$ was assumed to follow a multivariate normal distribution with mean zero and covariance $\sigma^2\mathbf{R}$. We assumed a typical exchangeable correlation

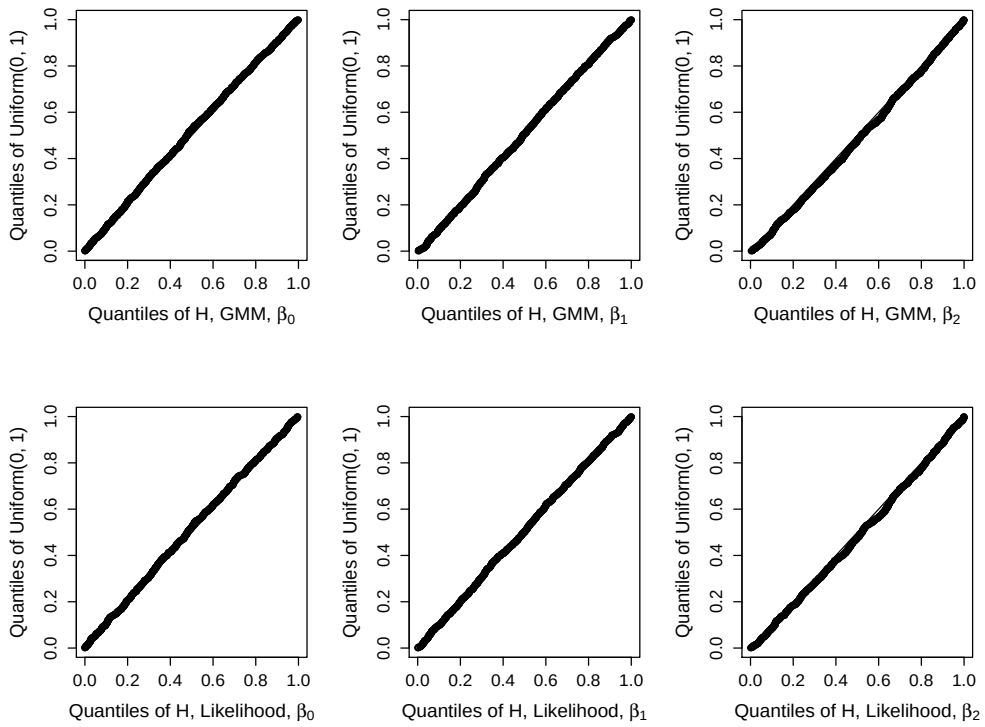

Figure 2: Quantile-quantile plots for the H statistics versus Uniform(0, 1) using the Bayesian GMM and likelihood method under the linear regression model with $n = 200$.

matrix $\mathbf{R} = (1-\rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}^T$ where ρ is the correlation coefficient, σ^2 is the marginal variance and $\mathbf{1}$ is a K-vector of 1. The true values of the parameters were $\beta_0 = 0.2$, $\beta_1 = 0.5$, $\beta_2 = -0.5$, and $\sigma = 1$, and we took $\rho = 0.25$ and 0.5. The prior distributions for the parameters were noninformative, for example, the prior distributions for the β's were $N(0, 10,000)$, and $\tau = \sigma^{-2}$ followed a Gamma(0.0001, 0.0001) distribution with mean one. We took the number of clusters as $n = 100$ and replicated 500 simulations for each setup so that we could examine the frequentist properties of the Bayesian GMM. For each data realization, after the burn-in period, we recorded 1,000 posterior samples, upon which we then based our computation of the posterior means, posterior standard deviations and the coverage probabilities of the 95% credible intervals.

We examined two frequentist GMM, two Bayesian GMM and two Bayesian likelihood-based estimation procedures. In Table 2, the first row contains simulation results obtained from the frequentist GMM with only one basis matrix $\mathbf{I}$ (known as the working independence model). The second row shows simulation results obtained from the frequentist GMM with two basis matrices $\mathbf{C}_{(1)} = \mathbf{I}$ and an exchangeable matrix with diagonal elements of 0 and off-diagonal elements of 1, denoted as $\mathbf{C}_{(2)} = \text{Exch}$. The

Table 2: Comparisons of simulation results using the frequentist GMM, the Bayesian GMM and the Bayesian likelihood-based estimation method with correlated data.

	$\beta_0 = 0.2$				$\beta_1 = 0.5$				$\beta_2 = -0.5$			
Method	Ave	ESD	ASD	CP	Ave	ESD	ASD	CP	Ave	ESD	ASD	CP
$\rho = 0.25$												
Freqn GMM (**I**)	.199	.082	.082	94.8	.498	.052	.050	94.8	−.502	.100	.099	94.2
Freqn GMM (**I**+Exch)	.192	.088	.079	91.8	.501	.048	.046	92.6	−.495	.101	.092	92.4
Bayes GMM (**I**)	.202	.082	.083	95.8	.501	.051	.050	95.0	−.502	.100	.100	92.8
Bayes GMM (**I**+Exch)	.197	.081	.082	94.4	.498	.048	.047	92.8	−.497	.093	.094	94.6
Bayes Like (Full)	.198	.080	.082	95.4	.498	.048	.047	93.6	−.498	.093	.094	94.6
Bayes Like (Fix ρ, σ)	.197	.080	.081	93.6	.497	.048	.046	93.4	−.497	.093	.094	95.4
$\rho = 0.5$												
Freqn GMM (**I**)	.205	.094	.092	94.0	.500	.051	.049	93.2	−.503	.102	.098	93.0
Freqn GMM (**I**+Exch)	.198	.091	.086	92.8	.501	.040	.039	94.2	−.501	.083	.078	92.8
Bayes GMM (**I**)	.201	.093	.094	95.0	.500	.050	.050	96.2	−.502	.100	.100	94.2
Bayes GMM (**I**+Exch)	.197	.089	.089	94.2	.498	.041	.040	93.2	−.498	.080	.079	94.2
Bayes Like (Full)	.203	.089	.090	94.1	.502	.042	.040	94.1	−.504	.078	.080	95.3
Bayes Like (Fix ρ, σ)	.200	.087	.089	94.6	.501	.041	.040	92.2	−.494	.079	.080	95.4

500 simulated data sets, Ave is the average of the parameter estimates over 500 simulations, ESD is the empirical standard deviation, ASD is the average of the standard error estimates, and CP is the 95% coverage probability.

third row shows simulation results from the Bayesian GMM with only one basis matrix **I**; and the fourth row exhibits estimation results from the Bayesian GMM with two basis matrices $\mathbf{C}_{(1)} = \mathbf{I}$ and $\mathbf{C}_{(2)} =$ Exch. When the correlation matrix of the errors is exchangeable, it can be fully characterized by these two basis matrices. Therefore, we expect the posterior estimates based on the Bayesian GMM with (**I** + Exch) to be comparable to those produced with the Bayesian likelihood method. The fifth row of Table 2 gives the results obtained from the typical Bayesian likelihood method, in which the five unknown parameters including three β's, ρ and σ were all updated in the Gibbs iterations. The results shown in the sixth row correspond to an ideal application of the Bayesian likelihood method in which ρ and σ were fixed at the true value and thus were not updated in the MCMC procedure, which would make the same number of unknown parameters for the Bayesian GMM and the Bayesian likelihood-based estimation procedures. For each simulated data set, we computed the parameter estimates, the corresponding standard errors and the 95% confidence intervals using the frequentist GMM; and computed the posterior means, the posterior standard deviations and the 95% credible intervals using the Bayesian methods.

In Table 2, we present the "Ave", the average of the parameter estimates (based on the frequentist GMM) or the average of the posterior means (each posterior mean was based on 1000 posterior samples using the Bayesian methods) over 500 replicated data sets; and the empirical standard deviation of the 500 estimates of the β's denoted as "ESD". We also show the "ASD", which was obtained by averaging the frequentist standard error estimates or the posterior standard deviations over 500 simulated data

sets; and the coverage probability "CP", which is the percentage of the frequentist 95% confidence intervals or the Bayesian 95% credible intervals that covered the true parameter value. We can see that the biases are negligible, and the parameter estimates are very close based on all of the six estimation methods. When the correlation is relatively low with $\rho = 0.25$, the Bayesian GMM with basis matrices of $(\mathbf{I} + \text{Exch})$ yielded similar posterior variance estimates compared to what were produced with only one basis matrix of $\mathbf{I}$. However, when the correlation is high with $\rho = 0.5$, adding the additional basis matrix "Exch" produced substantially smaller variances. Furthermore, the variances of the parameter estimates based on the Bayesian full likelihood approach are very close to those obtained from the Bayesian GMM with $(\mathbf{I} + \text{Exch})$. When the parameters ρ and σ were fixed at the true values, the Bayesian likelihood method yielded slightly better results. The coverage probabilities of the 95% confidence/credible intervals from all of the six estimation procedures were quite accurate. There were no notable differences in terms of the estimation bias and standard deviation between the Bayesian GMM and Bayesian likelihood approaches. We thus conclude that the Bayesian GMM can basically recover the information in the likelihood; it takes ρ and σ as nuisance parameters; and, more importantly, it does not rely on the multivariate normal assumption for the errors. In addition, comparing the frequentist GMM and the proposed Bayesian GMM, we obtained very similar results in terms of the parameter and variance estimates. This would suggest that the proposed MCMC procedure can be used to solve the frequentist GMM when the corresponding objective function is difficult to minimize numerically. Especially, when $\boldsymbol{\beta}$ is of high-dimension, minimization over a large dimensional space can be very challenging, whereas the Metropolis algorithm within the Gibbs sampler reduces the problem to sample from one-dimensional conditional densities.

4 Nursing Intervention Study

In longitudinal studies, the same response variable is measured at consecutive times, and the collection of responses on the same subject forms a cluster of outcomes. A nursing intervention study provided an interesting example of a longitudinal study for our proposed procedure. The Managing Uncertainty in Cancer research group in the School of Nursing at the University of North Carolina at Chapel Hill conducted a study to determine whether phone consultations with patients would increase their understanding of breast cancer and improve their ability to cope with the disease and condition (Mishel et al. 2005). Patients diagnosed with breast cancer face many uncertainties, and may feel hopeless in their ability to confront the disease. Such patients face uncertainty about the cause and progression of the disease, about possible treatments they will receive, about treatment side-effects and their prognosis. These uncertainties may undermine a patient's confidence in herself, in her beliefs, in her ability to determine the meaning of illness-related events and to overcome breast cancer. The major aim of this study was to implement an uncertainty-management intervention and determine its efficacy. It involved assessing the ability of the intervention to increase cancer knowledge, enhance self-care, promote a self-help response, and improve the quality of life among patients with breast cancer who were undergoing treatment.

In this study, patients diagnosed with breast cancer were randomly assigned to either the experimental/intervention arm or the control arm. One outcome of interest was cancer knowledge, which was defined as the knowledge patients gained about their cancer through the intervention. The outcome was measured at three time points for each patient: at baseline (for assessment measures), at four months post-baseline (for efficacy measures), and at seven months post-baseline (for durability measures). During the four months following recruitment, a nurse-client manager contacted each patient in the experimental group periodically by phone to relieve any concerns and answer any questions related to the patient's breast cancer, treatments, or treatment-related symptoms. Patients in the control arm did not receive the phone calls. The intervention ceased after four months. The durability measurement at month seven assessed whether the impact of the intervention endured. The same questions were repeatedly asked of patients in both arms over the three time points in order to evaluate the impact of the nurse intervention.

As the outcome of interest, we took the increment of cancer knowledge at months four and seven compared to the patient's baseline measurement. The covariates included the nursing intervention (experimental arm=1 and control arm=0) and ethnicity (white=1 and others=0). There were a total of 279 patients in our analysis. We took 50,000 posterior samples after 1,000 burn-in iterations. Based on the Bayesian GMM, the posterior means and standard deviations for the intercept, intervention effect and ethnicity were 0.611 (0.267), 1.063 (0.332) and 0.313 (0.333), respectively. For comparison, we also implemented the Bayesian likelihood method, for which we assumed a bivariate normal error distribution. Correspondingly, the posterior means and posterior standard deviations for the intercept, intervention effect and ethnicity were 0.615 (0.265), 1.062 (0.329) and 0.308 (0.340), respectively. The Bayesian likelihood-based method has two more parameters for the error distribution, i.e., the correlation coefficient ρ and the marginal variance σ^2. The posterior mean and standard deviation for ρ were 0.637 (0.036), and those for $\tau = \sigma^{-2}$ were 0.109 (0.008). We can see that the two Bayesian methods yielded very similar results. Our analysis showed that the nursing intervention clearly increased patients' cancer knowledge, and we found no difference in the increment of cancer knowledge across patient ethnicity. In addition, we implemented the frequentist GMM to analyze the nursing intervention data. The parameter estimates and standard errors for the intercept, intervention effect and ethnicity were 0.615 (0.265), 1.063 (0.328) and 0.307 (0.328), respectively. The results obtained from the Bayesian GMM and the frequentist GMM match very well, and thus the same inference and conclusion can be drawn.

5 Discussions

In practical applications, we often have limited information on the likelihood function. When the likelihood is difficult to derive, Bayesian inference is challenging and may be vulnerable to some additional assumptions. We have proposed the Bayesian GMM that can be implemented in a straightforward manner. The Bayesian GMM is most attractive when the likelihood function is difficult to derive. Numerical comparisons have shown

that the Bayesian GMM can basically recover the information in the likelihood, and thus may serve as a good substitute for the true likelihood. The proposed method can be viewed as a Bayesian analog of the frequentist GEE approach, which has gained much popularity in statistical analysis of longitudinal or clustered data. In addition, when it is difficult to numerically minimize $Q_n(\boldsymbol{\beta})$ over $\boldsymbol{\beta}$, especially when $\boldsymbol{\beta}$ is high-dimensional, the Bayesian GMM can be useful for approximating the frequentist GMM estimate and obtaining approximate frequentist inferences.

The proposed Bayesian GMM is based on the moment conditions, instead of the likelihood. In many semiparametric models, the likelihood is difficult to derive or maximize due to nuisance parameters (which can be infinite-dimensional), such as the quantile regression (Koenker and Bassett 1978), the proportional hazards model (Cox 1972) or other profile likelihoods in Murphy and van der Vaart (2000). Our procedure can be applied as long as the moments are correctly specified. In longitudinal studies, the GEE under the working independence model, still produces consistent estimators, while the corresponding variance need to be adjusted for the correlation based on the sandwich-form estimator. However, our formulation of the Bayesian GMM automatically accounts for the underlying correlation, such that the posterior samples of the parameters yield valid inference.

References

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore: Johns Hopkins University Press. 192

Chamberlain, G. (1987). "Asymptotic efficiency in estimation with conditional moment restrictions." *Journal of Econometrics*, 34: 305–334. 192

Chernozhukov, V. and Hong, H. (2003). "An MCMC approach to classical estimation." *Journal of Econometrics*, 115: 293–346. 192

Cowles, M. K. and Carlin, B. P. (1996). "Markov chain Monte Carlo convergence diagnostics: A comparative review." *Journal of the American Statistical Association*, 91: 883–904. 199

Cox, D. R. (1972). "Regression models and life tables (with discussion)." *Journal of Royal Statistical Society, Series B*, 34: 187–220. 205

Geweke, J. (1992). "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments." In Bernardo, J. M., Berger, J., Dawid, A. P., and Smith, A. F. M. (eds.), *Bayesian Statistics 4*, 169–193. Oxford: Oxford University Press. 199

Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995). "Adaptive rejection Metropolis sampling within Gibbs sampling." *Applied Statistics*, 44: 455–472. 195

Hall, A. R. (2005). *Generalized Method of Moments*. New York: Oxford University Press. 192

Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators." *Econometrica*, 50: 1029–1054. 192, 194

Hansen, L. P., Heaton, J., and Yaron, A. (1996). "Finite-sample properties of some alternative GMM estimators." *Journal of Business & Economic Statistics*, 14: 262–280. 192

Kim, J. Y. (2002). "Limited information likelihood and Bayesian analysis." *Journal of Econometrics*, 107: 175–193. 192

Koenker, R. and Bassett, G. J. (1978). "Regression quantiles." *Econometrica*, 46: 33–50. 205

Lai, T. L. and Small, D. (2007). "Marginal regression analysis of longitudinal data with time-dependent covariates: a generalised method of moments approach." *Journal of the Royal Statistical Society, Series B*, 69: 79–99. 192

Laird, N. M. and Ware, J. H. (1982). "Random-effects models for longitudinal data." *Biometrics*, 38: 963–974. 191

Lazar, N. A. (2003). "Bayesian empirical likelihood." *Biometrika*, 90: 319–326. 200

Lee, M. J. (1996). *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*. New York: Springer-Verlag. 192

Liang, K.-Y. and Zeger, S. L. (1986). "Longitudinal data analysis using generalized linear models." *Biometrika*, 73: 13–22. 192, 195

McCullagh, P. (1983). "Quasi-likelihood function." *Annals of Statistics*, 11: 59–67. 193

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. London: Chapman and Hall, 2nd edition. 193

McFadden, D. (1989). "A method of simulated moments for estimation of discrete response models without numerical integration." *Econometrica*, 57: 995–1026. 192

Mishel, M. H., Germino, B. B., Gil, K. M., Belyea, M., Laney, I. C., Stewart, J., Porter, L., and Clayton, M. (2005). "Benefits from an uncertainty management intervention for African-American and Caucasian older long-term breast cancer survivors." *Psychooncology*, 14: 962–978. 203

Monahan, J. F. and Boos, D. D. (1992). "Proper likelihoods for Bayesian analysis." *Biometrika*, 79: 271–278. 200

Murphy, S. A. and van der Vaart, A. W. (2000). "On the Profile Likelihood." *Journal of the American Statistical Association*, 95: 449–465. 205

Newey, W. K. (2004). "Efficient semiparametric estimation via moment restrictions." *Econometrica*, 72: 1877–1897. 192

Newey, W. K. and West, K. D. (1987). "Hypothesis testing with efficient method of moments estimation." *International Economic Review*, 28: 777–787. 192

Pakes, A. and Pollard, D. (1989). "Simulation and the asymptotics of optimization estimators." *Econometrica*, 57: 1027–1057. 192

Qu, A., Lindsay, B. G., and Li, B. (2000). "Improving generalised estimating equations using quadratic inference functions." *Biometrika*, 87: 823–836. 192, 196

Ware, J. H. (1985). "Linear models for the analysis of longitudinal studies." *The American Statistician*, 39: 95–101. 191

Wedderburn, R. W. M. (1974). "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method." *Biometrika*, 61: 439–447. 193

Zeger, S. L. and Liang, K. Y. (1986). "Longitudinal data analysis for discrete and continuous outcomes." *Biometrics*, 42: 121–130. 192

Zeger, S. L., Liang, K.-Y., and Albert, P. S. (1988). "Models for longitudinal data: A generalized estimating equation approach." *Biometrics*, 44: 1049–1060. 192

Zellner, A. (1997). "The Bayesian method of moments (BMOM): theory and applications." *Advances in Econometrics*, 12: 85–105. 192

Zellner, A., Tobias, J., and Ryu, H. (1997). "Bayesian method of moments (BMOM) analysis of parametric and semiparametric regression models." In *Proceedings of the Section on Bayesian Statistical Science*, 211–216. Alexandria, Virginia: American Statistical Association. 192

Acknowledgments

We would like to thank the referee, associate editor and editor for very insightful and constructive comments that substantially improved the paper.