



Supplementary materials for this article are available online.  
Please click the JCGS link at <http://pubs.amstat.org>.

# Stochastic Generalized Method of Moments

Guosheng YIN, Yanyuan MA, Faming LIANG, and Ying YUAN

The generalized method of moments (GMM) is a very popular estimation and inference procedure based on moment conditions. When likelihood-based methods are difficult to implement, one can often derive various moment conditions and construct the GMM objective function. However, minimization of the objective function in the GMM may be challenging, especially over a large parameter space. Due to the special structure of the GMM, we propose a new sampling-based algorithm, the stochastic GMM sampler, which replaces the multivariate minimization problem by a series of conditional sampling procedures. We develop the theoretical properties of the proposed iterative Monte Carlo method, and demonstrate its superior performance over other GMM estimation procedures in simulation studies. As an illustration, we apply the stochastic GMM sampler to a Medfly life longevity study. Supplemental materials for the article are available online.

**Key Words:** Generalized linear model; Gibbs sampling; Iterative Monte Carlo; Markov chain Monte Carlo; Metropolis algorithm; Moment condition.

## 1. INTRODUCTION

The generalized method of moments (GMM) plays an important role in statistical estimation and inference, particularly in econometrics (Hansen 1982; Newey and West 1987; Pakes and Pollard 1989; Lee 1996; Hansen, Heaton, and Yaron 1996; Newey 2004; and Hall 2005; among others). When the likelihood formulation is difficult, the GMM provides an attractive alternative for estimation and inference. Hansen (1982) developed a comprehensive framework and asymptotic theory for the GMM estimator. One of the main advantages of the GMM is that it can be used to make inference when there are more moment conditions than the number of unknown parameters. By combining the moments, the GMM is parsimonious and useful for obtaining efficient estimators (Chamberlain 1987).

---

Guosheng Yin is Associate Professor, Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong (E-mail: [gyin@hku.hk](mailto:gyin@hku.hk)). Yanyuan Ma is Associate Professor (E-mail: [ma@stat.tamu.edu](mailto:ma@stat.tamu.edu)) and Faming Liang is Professor (E-mail: [fliang@stat.tamu.edu](mailto:fliang@stat.tamu.edu)), Department of Statistics, Texas A&M University, College Station, TX 77843. Ying Yuan is Associate Professor, Department of Biostatistics—Unit 1411, M. D. Anderson Cancer Center, The University of Texas, Houston, TX 77230 (E-mail: [yyuan@mdanderson.org](mailto:yyuan@mdanderson.org)).

© 2011 American Statistical Association, Institute of Mathematical Statistics,  
and Interface Foundation of North America

*Journal of Computational and Graphical Statistics*, Volume 20, Number 3, Pages 714–727  
DOI: 10.1198/jcgs.2011.09210

In practice, if there is not enough information to construct the likelihood function, likelihood-based methods can be very difficult (Kim 2002; Chernozhukov and Hong 2003). Under such situations, Kim (2002) derived the limited information likelihood by minimizing the Kullback–Leibler distance. Chernozhukov and Hong (2003) developed a Laplace-type estimator obtained by a Markov chain Monte Carlo (MCMC) approach with a focus on non-smooth sample moments. Similarly, for non-smooth estimating functions, MCMC sampling-based approaches have been studied because such estimating equations are difficult to solve directly, and also the variance of the parameter estimate is hard to obtain (Tian et al. 2004; Tian, Liu, and Wei 2007). In the Bayesian framework, Yin (2009) proposed to sample the GMM using the usual Metropolis algorithm, while the convergence of the Markov chain appears to be slow due to the complexity of the pseudo-likelihood function. In this article, we use the special quadratic structure of the GMM, and propose sampling the GMM based on a novel iterative Monte Carlo procedure. The samples of the model parameters, although not exactly from an MCMC procedure, can be used for statistical inference purposes. The stochastic GMM sampler reduces the classical GMM minimization problem to a series of conditional sampling steps. In addition, the variances of the parameter estimates can be easily obtained using the empirical variance of a large number of parameter samples.

More specifically, we take the negative GMM quadratic function divided by two and then exponentiate it to construct the pseudo-likelihood function. In the proposed stochastic GMM sampler, the parameters in the variance–covariance matrix of the quadratic form are fixed as the sampled values at the previous iteration, such that the only parameters to sample are in the moments. This is the key difference from the usual Metropolis algorithm used by Chernozhukov and Hong (2003) and Yin (2009). The new sampling procedure converges faster and the samples mix better, because the sampling objective function is greatly simplified. The stochastic GMM sampler improves the estimation stability by inheriting the numerical advantages of the usual Gibbs sampler. The estimates of model parameters can be obtained easily, even for a high-dimensional case in which minimization over a large parameter space is often quite challenging.

The rest of this article is organized as follows. In Section 2 we introduce the notation, and review the classical frequentist GMM estimator and the Bayesian GMM estimator. In Section 3 we propose the stochastic GMM sampler, and also derive the asymptotic convergence property for the new stochastic GMM estimator. We examine the finite-sample properties using simulation studies, and illustrate the proposed stochastic GMM sampling scheme with a real data example in Section 4. Finally, we give concluding remarks in Section 5.

## 2. FREQUENTIST AND BAYESIAN GMM ESTIMATORS

In the classical GMM framework (Hansen 1982), we denote the population moment condition by  $E\{\mathbf{u}_i(\boldsymbol{\beta})\} = 0$ , for  $i = 1, \dots, n$ . The corresponding sample moment is

$$\mathbf{U}_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{u}_i(\boldsymbol{\beta}), \quad (2.1)$$

in which the dimension of  $\mathbf{U}_n(\boldsymbol{\beta})$  may be higher than that of  $\boldsymbol{\beta}$ . The GMM estimator  $\widehat{\boldsymbol{\beta}}$  is obtained by minimizing the 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 version of the variance–covariance matrix of  $\mathbf{U}_n(\boldsymbol{\beta})$ :

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

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

- (1) Insert the initial value  $\boldsymbol{\beta}^{(0)}$  into  $\boldsymbol{\Sigma}_n(\boldsymbol{\beta})$ .
- (2) At the  $k$ th 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  $\boldsymbol{\beta}$ , in which we fix  $\boldsymbol{\Sigma}_n(\widehat{\boldsymbol{\beta}}^{(k-1)})$  as known by plugging in  $\widehat{\boldsymbol{\beta}}^{(k-1)}$  from the  $(k-1)$ th iteration.

- (3) Plug the estimator  $\widehat{\boldsymbol{\beta}}^{(k)}$  back into  $\boldsymbol{\Sigma}_n(\widehat{\boldsymbol{\beta}}^{(k)})$ , and go back to step (2).
- (4) Continue this procedure until some prespecified convergence criteria are met.

Hence, we first specify an initial value for  $\boldsymbol{\beta}$  and continue updating the estimator until some prespecified convergence criteria are met. In step (2), the minimization is typically achieved by the usual Newton–Raphson algorithm. Under suitable regularity conditions, the classical 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 (Hansen 1982). The minimization can also be taken directly in a single step instead of using such a two-stage procedure, which produces an asymptotically equivalent estimator (Hansen, Heaton, and Yaron 1996). However, in certain cases, the minimization procedure needs to minimize a possibly complex function. If  $\mathbf{U}_n(\boldsymbol{\beta})$  is a highly nonlinear function of  $\boldsymbol{\beta}$ , the iterative procedure may converge to a local optimal GMM estimator.

In contrast to the minimization procedure, Chernozhukov and Hong (2004) and Yin (2009) proposed a Metropolis sampling approach to the GMM, in which they drew samples of  $\boldsymbol{\beta}$  from

$$\exp\left\{-\frac{1}{2}\mathbf{U}_n^T(\boldsymbol{\beta})\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\beta})\mathbf{U}_n(\boldsymbol{\beta})\right\}, \quad (2.2)$$

directly. Based on the mean and empirical standard deviation of the samples, they were able to make the same statistical inference as the classical GMM. By using the Metropolis algorithm, the multivariate minimization problem is reduced to an MCMC sampling problem. In particular, Chernozhukov and Hong (2003) provided regularity conditions and theoretical justification for the convergence in total variation of moments norm. However, the model parameter  $\boldsymbol{\beta}$  stays in both  $\mathbf{U}_n(\boldsymbol{\beta})$  and  $\boldsymbol{\Sigma}_n^{-1}(\boldsymbol{\beta})$  as shown in (2.2). Such a complicated structure often causes slow convergence of the Metropolis algorithm. From a computational perspective, we propose using the Bayesian mechanism and a new iterative Monte Carlo algorithm as a means of computing an estimator that is effectively equivalent to the frequentist GMM estimator.

### 3. STOCHASTIC GMM SAMPLER

Following the same route of Chernozhukov and Hong (2003) and Yin (2009), we propose a new iterative Monte Carlo sampling procedure, the stochastic GMM sampler, to alleviate the slow convergence of the usual Metropolis algorithm. We note that the main difficulty in drawing samples from (2.2) is that  $\beta$  appears in both  $U_n(\beta)$  and  $\Sigma_n^{-1}(\beta)$ , and thus it is difficult to find a sensible candidate distribution in the Metropolis procedure. Our Monte Carlo method takes iterative steps with partial substitution of the parameter samples in the covariance matrix. The detailed description of the new procedure is given as follows.

- (1) Insert an initial value  $\beta^{(0)}$  into  $\Sigma_n(\beta)$ .
- (2) At the  $k$ th iteration, plug in the sample of the  $(k - 1)$ th iteration,  $\beta^{(k-1)}$ , such that the covariance matrix  $\Sigma_n(\beta^{(k-1)})$  is fixed and known. Then, draw a sample  $\beta^{(k)}$  from

$$\exp\left\{-\frac{1}{2}U_n^T(\beta)\Sigma_n^{-1}(\beta^{(k-1)})U_n(\beta)\right\},$$

where  $\Sigma_n^{-1}(\beta^{(k-1)})$  does not involve any parameter to be sampled.

- (3) Continue such sampling until the chain appears to be converging and stable.

More specifically, we give two illustrative examples. If the GMM only involves a single model parameter ( $\beta$  is a scalar), then at the  $k$ th iteration, we sample  $\beta^{(k)}$  from

$$[\beta|\beta^{(k-1)}] \sim \exp\left\{-\frac{U_n^2(\beta)}{2\Sigma_n(\beta^{(k-1)})}\right\}.$$

In a more general situation, we take a three-dimensional parameter vector  $\beta = (\beta_0, \beta_1, \beta_2)^T$  as an example. At the  $k$ th iteration, we sample from the following conditional distributions:

$$\begin{aligned} & [\beta_0|\beta_0^{(k-1)}, \beta_1^{(k-1)}, \beta_2^{(k-1)}] \\ & \sim \exp\left\{-\frac{1}{2}U_n^T(\beta_0, \beta_1^{(k-1)}, \beta_2^{(k-1)})\Sigma_n^{-1}(\beta_0^{(k-1)}, \beta_1^{(k-1)}, \beta_2^{(k-1)})U_n(\beta_0, \beta_1^{(k-1)}, \beta_2^{(k-1)})\right\}, \\ & [\beta_1|\beta_0^{(k)}, \beta_1^{(k-1)}, \beta_2^{(k-1)}] \\ & \sim \exp\left\{-\frac{1}{2}U_n^T(\beta_0^{(k)}, \beta_1, \beta_2^{(k-1)})\Sigma_n^{-1}(\beta_0^{(k)}, \beta_1^{(k-1)}, \beta_2^{(k-1)})U_n(\beta_0^{(k)}, \beta_1, \beta_2^{(k-1)})\right\}, \\ & [\beta_2|\beta_0^{(k)}, \beta_1^{(k)}, \beta_2^{(k-1)}] \\ & \sim \exp\left\{-\frac{1}{2}U_n^T(\beta_0^{(k)}, \beta_1^{(k)}, \beta_2)\Sigma_n^{-1}(\beta_0^{(k)}, \beta_1^{(k)}, \beta_2^{(k-1)})U_n(\beta_0^{(k)}, \beta_1^{(k)}, \beta_2)\right\}. \end{aligned}$$

At each step of sampling, we may use the adaptive rejection Metropolis sampling algorithm (Gilks, Best, and Tan 1995). The proposed iterative Monte Carlo procedure is different from the usual Gibbs sampling, in that at each iteration we fix the covariance matrix,  $\Sigma_n(\beta)$ , by plugging in the sample of  $\beta$  from the previous iteration. This augmentation simplifies the sampling procedure tremendously. Because in the original form of the asymptotic distribution, the same  $\beta$  is contained in both the variance term  $\Sigma_n(\beta)$  and the moment term  $U_n(\beta)$ , it is difficult to approximate the shape of the conditional density.

We now demonstrate the equivalence of the stochastic GMM estimator and the classical frequentist GMM estimator. Let  $\beta^*$  denote the unique classical GMM estimator which minimizes  $\mathbf{U}_n^T(\beta)\Sigma_n^{-1}(\beta)\mathbf{U}_n(\beta)$  in the article by Hansen (1982). Let  $\{\beta^{(1)}, \dots, \beta^{(K)}\}$  be a sequence of the stochastic GMM samples generated according to the two-stage iterative Monte Carlo procedure described previously. Let  $\widehat{\beta}$  and  $\widehat{\Gamma}$  be the sample mean and sample variance of  $\{\beta^{(1)}, \dots, \beta^{(K)}\}$ ,

$$\widehat{\beta} = \frac{1}{K} \sum_{k=1}^K \beta^{(k)},$$

$$\widehat{\Gamma} = \frac{1}{K-1} \sum_{k=1}^K (\beta^{(k)} - \widehat{\beta})(\beta^{(k)} - \widehat{\beta})^T.$$

Then, we have the following theoretical convergence properties for the stochastic GMM sampler.

**Theorem 1.** *Assume that the true variance function  $\text{var}\{\mathbf{u}_1(\beta)\}$  evaluated at the classical GMM estimator  $\beta^*$  is finite and has a finite first derivative with respect to  $\beta$ . As the data sample size  $n \rightarrow \infty$  and the number of the stochastic GMM samples  $K \rightarrow \infty$ , both  $\widehat{\beta} - \beta^*$  and  $\widehat{\Gamma} - \text{var}(\beta^*)$  go to zero in probability.*

The theorem lays out the theoretical foundation for the validity of the new iterative Monte Carlo sampling method, for which the proof is outlined in the Appendix.

## 4. NUMERICAL STUDIES

### 4.1 SIMULATIONS

We conducted simulation studies to examine the performance of the proposed stochastic GMM estimator. For ease of exposition, we first considered the generalized linear models (GLMs), while the GMM is applicable in a much broader range of problems. In particular, we examined the linear, logistic, and Poisson regression models. For the  $i$ th subject,  $i = 1, \dots, n$ , let  $y_i$  be the outcome of interest, and let  $\mathbf{Z}_i$  be the corresponding covariate vector. Under the linear regression model,

$$y_i = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i} + \epsilon_i,$$

we took the true parameter values  $\beta_0 = 0.2$ ,  $\beta_1 = 0.5$ ,  $\beta_2 = -0.5$ , and  $\epsilon_i \sim N(0, \sigma^2)$  a zero-mean normal distribution with variance  $\sigma^2 = 0.25$ . The covariate  $Z_{1i}$  was generated from the standard normal distribution  $N(0, 1)$ , and  $Z_{2i}$  was a binary variable taking a value of 0 or 1 with probability 0.5. With  $\beta = (\beta_0, \beta_1, \beta_2)^T$  and  $\mathbf{Z}_i = (1, Z_{1i}, Z_{2i})^T$ , the sample moment is given by

$$\mathbf{U}_n(\beta) = \sum_{i=1}^n \mathbf{Z}_i (y_i - \beta^T \mathbf{Z}_i).$$

We also considered the logistic regression model, which takes the form of

$$\text{logit}(p_i) = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i},$$

under which the binary outcome  $y_i$  equals 1 with probability  $p_i$ , or 0 with probability  $1 - p_i$ . Under the logistic model, the sample moment is given by

$$\mathbf{U}_n(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{Z}_i \left\{ y_i - \frac{\exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}{1 + \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)} \right\}.$$

For the Poisson log-linear model,

$$\log(\mu_i) = \beta_0 + \beta_1 Z_{1i} + \beta_2 Z_{2i},$$

we simulated  $y_i$  as a Poisson variable with mean  $\mu_i$ , and correspondingly, the sample moment is given by

$$\mathbf{U}_n(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{Z}_i \{y_i - \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)\}.$$

In the logistic and Poisson models, the true values of the regression parameters were taken to be the same as those in the linear regression model, and the covariates were generated similarly as well, that is,  $Z_{1i} \sim N(0, 1)$  and  $Z_{2i} \sim \text{Bernoulli}(0.5)$ .

Besides the GLMs, we also considered the semiparametric proportional hazards model (Cox 1972) for the time-to-event data, which has been widely used in survival analysis. Let  $T_i$  be the failure time for subject  $i$ , and let  $C_i$  be the censoring time; then  $Y_i = \min(T_i, C_i)$  is the observed time. Let  $\Delta_i = I(T_i \leq C_i)$  be the failure time indicator, where  $I(\cdot)$  is the indicator function. Assume that  $T_i$  and  $C_i$  are conditionally independent given covariate  $\mathbf{Z}_i$ , and the observed data  $\{Y_i, \Delta_i, \mathbf{Z}_i\}$  are independent and identically distributed for  $i = 1, \dots, n$ .

Under the proportional hazards model, the hazard function for the  $i$ th subject with a covariate vector  $\mathbf{Z}_i$  is

$$\lambda(t|\mathbf{Z}_i) = \lambda_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i), \tag{4.1}$$

where  $\boldsymbol{\beta}$  is a  $p$ -dimensional vector of unknown regression coefficients and the baseline hazard  $\lambda_0(t)$  is a completely unspecified and unknown infinite-dimensional parameter. Using the counting process notation, we can define the at-risk process as  $J_i(t) = I(Y_i \geq t)$ , the counting process as  $N_i(t) = I(Y_i \leq t, \Delta_i = 1)$ , and then the martingale is given by

$$M_i(t) = N_i(t) - \int_0^t J_i(u) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i) d\Lambda_0(u),$$

where the baseline cumulative hazard function  $\Lambda_0(t) = \int_0^t \lambda_0(u) du$ .

The estimation and inference procedure of model (4.1) follows the partial likelihood function, in which  $\lambda_0(t)$  cancels out so that the only unknown parameter left is  $\boldsymbol{\beta}$ . Let  $D$  denote the observed data, and then the partial likelihood is given by

$$L_{\text{partial}}(D|\boldsymbol{\beta}) = \prod_{i=1}^n \left\{ \frac{\exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}{\sum_{k=1}^n J_k(Y_i) \exp(\boldsymbol{\beta}^T \mathbf{Z}_k)} \right\}^{\Delta_i},$$

and the corresponding score function is

$$U_n(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^{\infty} \left\{ \mathbf{Z}_i - \frac{\sum_{k=1}^n J_k(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_k) \mathbf{Z}_k}{\sum_{k=1}^n J_k(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_k)} \right\} dM_i(t).$$

In the Cox model, we took  $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)^T$  with the true parameter values  $\beta_0 = 0.2$ ,  $\beta_1 = 0.5$ , and  $\beta_2 = -0.5$ , and the corresponding covariates  $\mathbf{Z}_i = (Z_{0i}, Z_{1i}, Z_{2i})^T$  were simulated from Bernoulli(0.5),  $N(0, 1)$ , and Bernoulli(0.5), respectively. The baseline hazard was taken as a constant of 1, and we simulated censoring times from an exponential distribution,  $\text{Exp}(0.25)$ , resulting in a censoring rate of 24%.

Under each of the GLMs and the Cox model, we examined the performance of the proposed stochastic GMM estimator. For comparison, we also implemented the Metropolis approach (Chernozhukov and Hong 2003; Yin 2009), the Langevin Metropolis–Hastings (MH) method (Roberts and Tweedie 1996; Roberts and Rosenthal 1998), and the frequentist GMM method. The Metropolis approach to the GMM is to directly sample from the pseudo-likelihood function. In the Langevin-MH procedure, the discretized Langevin diffusion is used as the proposal density to adaptively mimic the target posterior distribution. We took the sample size  $n = 100, 200$ , and 500, and for each configuration, we simulated 1000 datasets. For the sampling procedures including our stochastic GMM sampler, the Metropolis, and the Langevin-MH approaches, we took 10,000 samples after a burn-in period of 100 iterations.

The simulation results are summarized in Table 1, where we present the average of the parameter estimates over 1000 simulations (Est); the empirical standard deviation (SD); the average of the standard error estimates (SE); and the 95% coverage probability (CP). We can see that under the linear regression model all of the four estimation methods, that is, the proposed stochastic GMM estimator, the Metropolis algorithm, the Langevin-MH procedure, and the frequentist GMM approach, yielded very similar estimates for the model parameters and standard errors. Furthermore, the corresponding coverage probabilities are close to the nominal level 95%. Under the logistic regression model, the stochastic GMM, Metropolis, and Langevin-MH approaches performed similarly, especially when  $n = 200$  and 500. However, all three sampling methods produced relatively larger bias and slightly wider interval estimates than the frequentist GMM method. Under the Poisson log-linear model, the Metropolis algorithm and Langevin-MH performed relatively worse than the other two estimation procedures in terms of both the point and variance estimates, especially when  $n = 100$ . There appeared to be some bias in the point estimates using the Metropolis and Langevin-MH procedures, while as the sample size increased to 500 the bias diminished. Under the Cox proportional hazards model, we used the partial likelihood method for comparison, and the simulation results are quite similar across the four methods. As a summary, when the sample size increased from  $n = 100$  to 500, the performance of the three sampling methods substantially improved. In the situations considered, the usual Metropolis algorithm and the Langevin-MH procedure performed similarly, while the proposed stochastic GMM sampling procedure generally produced better estimates than those two. The classical GMM which requires direct minimization of the objective quadratic function performed quite competitively.

Table 1. Comparison of parameter estimates using the proposed stochastic GMM, Metropolis, Langevin Metropolis–Hastings (Langevin-MH), and frequentist minimization methods.

Model	$n$	Estimation method	$\beta_0 = 0.2$				$\beta_1 = 0.5$				$\beta_2 = -0.5$			
			Est	SD	SE	CP	Est	SD	SE	CP	Est	SD	SE	CP
Linear	100	Stochastic GMM	0.199	0.071	0.072	94.2	0.500	0.051	0.051	93.9	-0.502	0.102	0.101	94.6
		Metropolis	0.199	0.071	0.074	95.7	0.500	0.051	0.056	96.9	-0.502	0.102	0.105	95.4
		Langevin-MH	0.199	0.071	0.073	95.0	0.500	0.051	0.058	97.0	-0.502	0.102	0.102	94.4
		Frequentist GMM	0.203	0.073	0.070	94.0	0.501	0.051	0.050	94.1	-0.503	0.102	0.099	94.0
	200	Stochastic GMM	0.202	0.053	0.050	93.7	0.500	0.036	0.036	94.6	-0.502	0.074	0.071	94.4
		Metropolis	0.199	0.049	0.051	96.4	0.500	0.036	0.037	95.6	-0.500	0.071	0.072	95.3
		Langevin-MH	0.202	0.053	0.050	93.2	0.500	0.036	0.038	96.5	-0.502	0.074	0.070	93.6
		Frequentist GMM	0.199	0.052	0.050	92.9	0.501	0.035	0.035	94.9	-0.498	0.073	0.070	93.8
	500	Stochastic GMM	0.200	0.032	0.032	94.6	0.500	0.023	0.022	94.6	-0.501	0.045	0.045	94.9
		Metropolis	0.200	0.032	0.032	94.8	0.500	0.023	0.023	95.0	-0.501	0.045	0.045	95.3
		Langevin-MH	0.200	0.032	0.031	94.3	0.500	0.023	0.024	95.6	-0.501	0.045	0.044	94.1
		Frequentist GMM	0.200	0.031	0.032	95.1	0.500	0.023	0.022	95.2	-0.499	0.045	0.045	94.7
Logistic	100	Stochastic GMM	0.246	0.352	0.338	93.4	0.629	0.248	0.259	92.7	-0.592	0.506	0.481	92.6
		Metropolis	0.247	0.353	0.349	94.3	0.626	0.247	0.276	91.6	-0.594	0.508	0.498	93.3
		Langevin-MH	0.241	0.354	0.334	92.8	0.602	0.242	0.270	93.0	-0.567	0.492	0.458	92.2
		Frequentist GMM	0.210	0.316	0.300	94.2	0.518	0.233	0.227	94.7	-0.516	0.441	0.427	95.2
	200	Stochastic GMM	0.208	0.233	0.220	93.4	0.547	0.172	0.173	93.1	-0.531	0.310	0.313	94.4
		Metropolis	0.211	0.226	0.224	94.2	0.552	0.174	0.179	93.9	-0.512	0.310	0.318	95.2
		Langevin-MH	0.208	0.235	0.220	93.3	0.544	0.172	0.184	94.0	-0.529	0.310	0.308	94.0
		Frequentist GMM	0.206	0.217	0.210	94.9	0.511	0.161	0.158	95.0	-0.512	0.299	0.298	95.6
	500	Stochastic GMM	0.209	0.142	0.134	93.5	0.519	0.104	0.102	94.6	-0.525	0.195	0.190	94.5
		Metropolis	0.209	0.142	0.135	93.9	0.519	0.104	0.103	94.7	-0.525	0.194	0.192	94.8
		Langevin-MH	0.208	0.142	0.132	93.4	0.517	0.104	0.107	95.4	-0.523	0.194	0.185	93.6
		Frequentist GMM	0.203	0.132	0.132	95.0	0.506	0.102	0.099	93.9	-0.506	0.188	0.187	95.0

Table 1. (Continued.)

Model	$n$	Estimation method	$\beta_0 = 0.2$				$\beta_1 = 0.5$				$\beta_2 = -0.5$			
			Est	SD	SE	CP	Est	SD	SE	CP	Est	SD	SE	CP
Poisson	100	Stochastic GMM	0.156	0.133	0.135	92.9	0.500	0.098	0.101	95.4	-0.527	0.212	0.213	94.4
		Metropolis	0.133	0.132	0.172	96.1	0.644	0.158	0.286	98.4	-0.585	0.229	0.321	97.9
		Langevin-MH	0.133	0.132	0.172	96.1	0.646	0.159	0.286	98.2	-0.585	0.229	0.321	98.1
		Frequentist GMM	0.189	0.132	0.128	93.5	0.497	0.094	0.094	93.5	-0.505	0.204	0.195	93.1
	200	Stochastic GMM	0.175	0.096	0.094	93.9	0.501	0.070	0.068	94.6	-0.513	0.145	0.144	94.2
		Metropolis	0.169	0.097	0.103	94.3	0.525	0.082	0.114	97.3	-0.520	0.148	0.164	95.2
		Langevin-MH	0.170	0.096	0.101	95.0	0.524	0.084	0.111	96.0	-0.521	0.149	0.163	95.5
		Frequentist GMM	0.193	0.094	0.091	93.9	0.503	0.068	0.066	92.9	-0.494	0.136	0.138	95.6
	500	Stochastic GMM	0.191	0.057	0.058	94.6	0.501	0.042	0.043	95.0	-0.507	0.089	0.089	95.1
		Metropolis	0.190	0.058	0.060	94.9	0.504	0.044	0.051	95.8	-0.507	0.089	0.092	95.4
		Langevin-MH	0.190	0.058	0.060	94.9	0.503	0.046	0.048	95.7	-0.508	0.090	0.091	95.4
		Frequentist GMM	0.202	0.059	0.057	95.0	0.499	0.044	0.042	93.8	-0.503	0.088	0.087	95.2
Cox	100	Stochastic GMM	0.205	0.235	0.237	95.8	0.516	0.126	0.131	95.5	-0.530	0.241	0.250	94.8
		Metropolis	0.205	0.235	0.237	96.1	0.515	0.126	0.131	95.4	-0.529	0.241	0.251	94.9
		Langevin-MH	0.205	0.235	0.238	96.4	0.516	0.126	0.132	95.3	-0.529	0.241	0.252	95.1
		Partial likelihood	0.202	0.237	0.241	95.9	0.518	0.136	0.131	94.6	-0.502	0.247	0.245	94.7
	200	Stochastic GMM	0.202	0.167	0.169	95.8	0.512	0.090	0.091	94.5	-0.511	0.175	0.172	94.0
		Metropolis	0.202	0.167	0.169	95.9	0.512	0.090	0.091	94.4	-0.511	0.175	0.172	94.2
		Langevin-MH	0.202	0.167	0.170	95.5	0.512	0.090	0.092	94.3	-0.510	0.175	0.173	93.8
		Partial likelihood	0.200	0.167	0.167	95.4	0.507	0.091	0.090	94.9	-0.506	0.165	0.169	96.2
	500	Stochastic GMM	0.198	0.106	0.105	94.2	0.501	0.056	0.056	94.9	-0.505	0.107	0.106	95.1
		Metropolis	0.198	0.106	0.105	94.0	0.501	0.056	0.056	95.0	-0.505	0.107	0.106	95.0
		Langevin-MH	0.198	0.106	0.105	94.1	0.501	0.056	0.056	95.1	-0.505	0.107	0.106	94.9
		Partial likelihood	0.200	0.102	0.104	95.4	0.505	0.057	0.056	94.9	-0.500	0.104	0.105	95.9

NOTE: Est is the average of the parameter estimates over 1000 simulations, SD is the empirical standard deviation, SE is the average of the standard error estimates, and CP is the 95% coverage probability.

To further examine the convergence of the proposed stochastic GMM sampler, we compare the cumulative mean squared error (MSE) under the linear, logistic, Poisson, and Cox models, respectively. We simulated 1000 datasets with  $n = 100$ , and computed the averaged MSE based on the accumulating samples increased by every 10 iterations. That is, for each component of  $\beta$ , we increased  $K$  by 10 to evaluate

$$\text{MSE}_K = \frac{1}{K} \sum_{k=1}^K (\beta^{(k)} - \beta_{\text{True}})^2,$$

where  $\beta^{(k)}$  is the  $k$ th sample and  $\beta_{\text{True}}$  is the true parameter value. In Figure 1, we show the plot of the cumulative MSE for each component of  $\beta$  under different models. We can see that the convergence of the proposed stochastic GMM sampler is in general better than that of the Metropolis algorithm, especially under the linear and Poisson cases. Compared with the Langevin-MH approach, the stochastic GMM performed substantially better under the Poisson model, but slightly worse under the logistic model. Under the Cox proportional hazards model, the three sampling approaches performed comparably. The main advantage of the stochastic GMM over the Langevin-MH is its ease of implementation. The Langevin-MH algorithm involves the first partial derivatives of the objective function, and the appearance of parameters in the covariance matrix makes these derivatives particularly difficult to evaluate.

## 4.2 EXAMPLE

As an illustration, we applied the proposed stochastic GMM to data from a Medfly life longevity study with 1,203,646 observations to determine the relationship between the Medfly longevity and potentially important risk factors (Carey et al. 1992; Carey, Liedo, and Vaupel 1995; Carey and Liedo 1995; Koenker and Geling 2001). In the study, pupae were sorted into five sizes from 4 mm up to 8 mm, and the pupal size was known to be strongly associated with the adult size of the fly, which would affect the fly's survival. There were a total of 167 cages, and each cage only contained one size. The initial density of flies and the percentage of male flies in each cage varied considerably, both of which could also affect the survival.

In our analysis, we took a log transformation on the survival times of the Medflies, and included the covariates of sex (female = 1, male = 0), the size of the Medfly, the fly density, and the percentage of male Medflies in each cage in an accelerated failure time model (Kalbfleisch and Prentice 2002). In the stochastic GMM, we took 10,000 samples after 100 burn-in iterations. The samples obtained from the stochastic GMM sampler and the corresponding density functions are exhibited in Figure 2. We can see that the convergence of the samples is very fast, the samples mix well, and the densities appear to be normal. In Table 2, we present the parameter estimates and the standard errors using our stochastic GMM sampling procedure, and the classical GMM minimization method. We can see that both the point estimates and the standard errors for all of the covariate effects are very close between these two approaches due to the extremely large sample size. As a result, the proposed stochastic GMM sampler yielded a similar conclusion to that from the

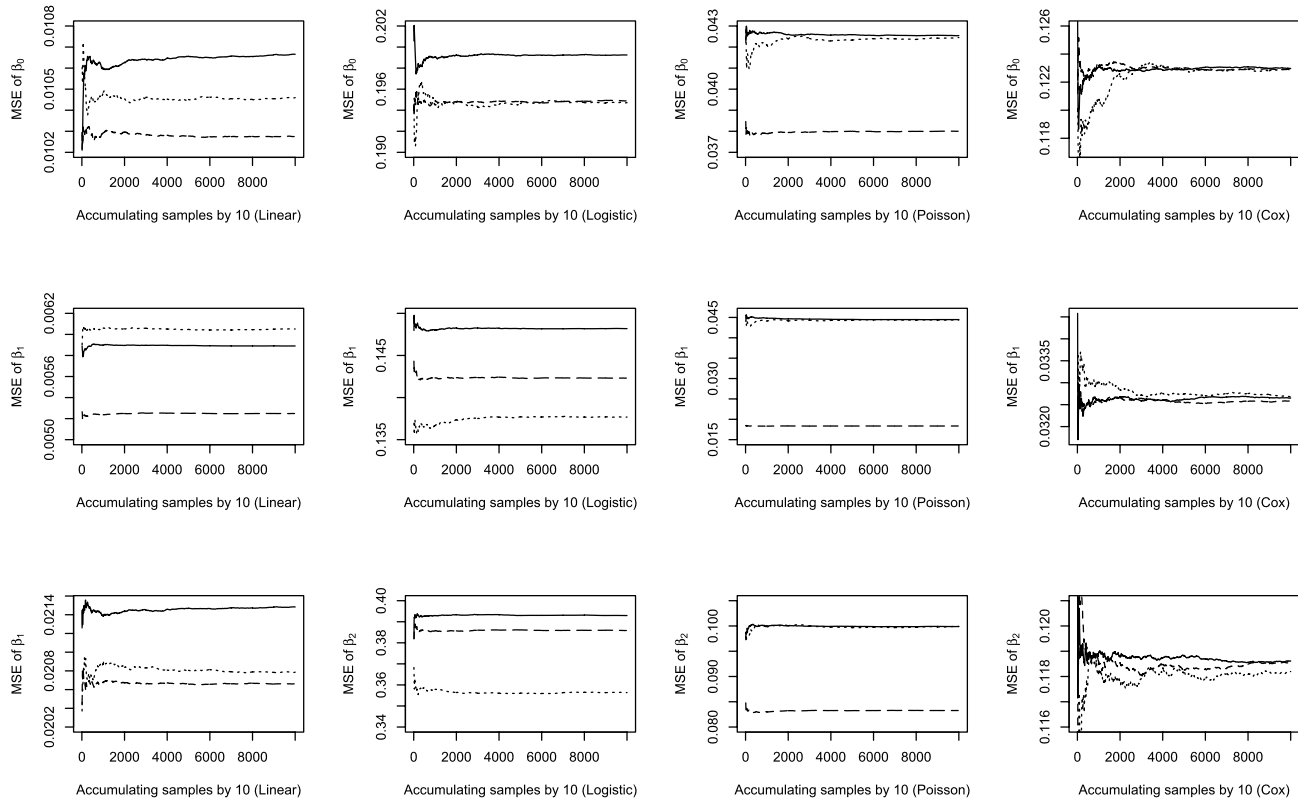


Figure 1. Mean squared errors by accumulating every 10 samples averaged over 1000 simulations. Under the linear, logistic, Poisson, and Cox models, respectively, the solid line is obtained using the Metropolis algorithm (Chernozhukov and Hong 2003), the dotted line denotes the Langevin Metropolis–Hastings method (Roberts and Tweedie 1996), and the dashed line corresponds to the stochastic GMM estimator.

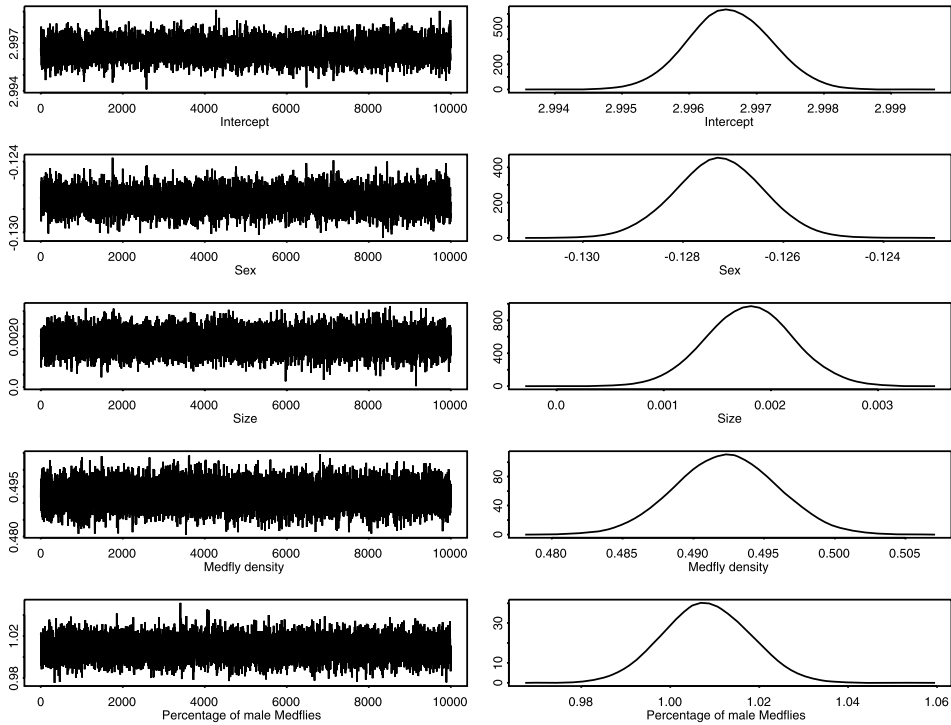


Figure 2. The trace plots for the samples using the stochastic GMM for the Medfly data analysis are given in the left panels, and the corresponding densities are exhibited in the right panels.

frequentist GMM. All of the four covariates considered strongly affected flies’ survival. Male Medflies survived significantly longer than female. The size of the Medfly was also significantly associated with survival but not as strongly as other covariates: a Medfly with a bigger size would survive longer. Furthermore, the higher was the density of the flies in a cage, the longer was the Medfly’s survival; and more male flies in a cage also would lead to better survival of the flies. Intuitively, our findings are consistent with the literature of Medfly life longevity that bigger sizes, more male flies, and stronger cohorts would lead to better survival.

Table 2. Analysis of the Medfly longevity data using the proposed stochastic GMM sampling procedure and the frequentist GMM minimization method.

Covariate	Stochastic GMM			Frequentist GMM		
	Estimate	Std. error	Z Stat.	Estimate	Std. error	Z Stat.
Intercept	2.99660	0.00060	5001.8	2.99663	0.00061	4897.2
Sex	-0.12727	0.00086	-147.5	-0.12732	0.00086	-147.3
Size	0.00180	0.00040	4.5	0.00181	0.00039	4.6
Density	0.49234	0.00351	140.3	0.49233	0.00339	145.1
% male	1.00812	0.00954	105.6	1.00802	0.00982	102.7

## 5. DISCUSSION

We have proposed a new sampling scheme for the GMM estimation, which can be implemented in a straightforward manner. Our proposed stochastic GMM sampling algorithm reduces the minimization problem to a sampling procedure based on a series of one-dimensional conditional densities as the usual Gibbs sampler. Using the stochastic GMM sampler, the point estimate of the model parameter can be obtained as the average of the samples, and the standard error can be easily computed as the empirical standard deviation of the samples. When  $\beta$  is of high dimension, and when it is difficult to numerically minimize  $Q_n(\beta)$  over  $\beta$ , the stochastic GMM estimator with its simple iterative Monte Carlo sampling scheme can be very useful for statistical estimation and inference. Although the classical Metropolis–Hastings algorithm can also be used as a sampling tool for the GMM, both  $U_n(\beta)$  and  $\Sigma_n^{-1}(\beta)$  contain the model parameter  $\beta$  to be sampled. Such a complicated functional structure makes the Metropolis sampling difficult and the convergence of the MCMC slow. The new stochastic GMM sampling algorithm through partial parameter substitution immensely improves the convergence property of the samples.

## SUPPLEMENTARY MATERIALS

**Appendix:** The supplemental files include the Appendix which gives the proof of Theorem 1. (appendix.pdf)

**C++ Code:** The supplemental files include C++ programs which can be used to replicate the Medfly data analysis. The programs need a C++ library BiostatGeneral.lib which can be requested from the Department of Biostatistics, M. D. Anderson Cancer Center. (cpp.zip)

**Data:** The supplemental files include the Medfly dataset. The original data contain 1,203,646 observations (with duplicates), and the current data contain 19,072 observations (with a replication weight that can be used to extract to the original data). (fly.txt)

## ACKNOWLEDGMENTS

We thank the referees, associate editor, and editor for many insightful suggestions which strengthened the work immensely. Yin's research was supported by a grant from the Research Grants Council of Hong Kong, Ma's research was supported by a US NSF grant, Liang's research was supported by grants from US NSF (DMS-1007457 and CMMI-0926803) and King Abdullah University of Science and Technology (KUS-C1-016-04), and Yuan's research was supported by a U.S. National Cancer Institute R01 grant (R01CA154591-01A1).

*[Received November 2009. Revised February 2011.]*

## REFERENCES

- Carey, J. R., and Liedo, P. (1995), "Sex Mortality Differentials and Selective Survival in Large Medfly Cohorts," *The Gerontologist*, 35, 588–596. [723]

- Carey, J. R., Liedo, P., Orozco, D., and Vaupel, J. W. (1992), "Slowing of Mortality Rates at Older Ages in Large Medfly Cohorts," *Science*, 258, 457–461. [723]
- Carey, J. R., Liedo, P., and Vaupel, J. W. (1995), "Mortality Dynamics of Density in the Mediterranean Fruit Fly," *Experimental Gerontology*, 30, 605–629. [723]
- Chamberlain, G. (1987), "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334. [714]
- Chernozhukov, V., and Hong, H. (2003), "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115, 293–346. [715-717,720,724]
- Cox, D. R. (1972), "Regression Models and Life-Tables," *Journal of the Royal Statistical Society, Ser. B*, 34, 187–220. [719]
- Gilks, W. R., Best, N. G., and Tan, K. K. C. (1995), "Adaptive Rejection Metropolis Sampling," *Applied Statistics*, 44, 455–472. [717]
- Hall, A. R. (2005), *Generalized Method of Moments*, New York: Oxford University Press. [714]
- Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054. [714-716,718]
- 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. [714,716]
- Kalbfleisch, J. D., and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data* (2nd ed.), New York: Wiley. [723]
- Kim, J.-Y. (2002), "Limited Information Likelihood and Bayesian Analysis," *Journal of Econometrics*, 107, 175–193. [715]
- Koenker, R., and Geling, O. (2001), "Reappraising Medfly Longevity: A Quantile Regression Survival Analysis," *Journal of the American Statistical Association*, 96, 458–468. [723]
- Lee, M. J. (1996), *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variable Models*, New York: Springer-Verlag. [714]
- Newey, W. K. (2004), "Efficient Semiparametric Estimation via Moment Restrictions," *Econometrica*, 72, 1877–1897. [714]
- Newey, W. K., and West, K. D. (1987), "Hypothesis Testing With Efficient Method of Moments Estimation," *International Economic Review*, 28, 777–787. [714]
- Pakes, A., and Pollard, D. (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057. [714]
- Roberts, G. O., and Rosenthal, J. S. (1998), "Optimal Scaling of Discrete Approximations to Langevin Diffusions," *Journal of Royal Statistical Society, Ser. B*, 60, 255–268. [720]
- Roberts, G. O., and Tweedie, R. L. (1996), "Exponential Convergence of Langevin Distributions and Their Discrete Approximations," *Bernoulli*, 2, 341–363. [720,724]
- Tian, L., Liu, J., Zhao, Y., and Wei, L. J. (2004), "Statistical Inference Based on Non-Smooth Estimating Functions," *Biometrika*, 91, 943–954. [715,716]
- Tian, L., Liu, J. S., and Wei, L. J. (2007), "Implementation of Estimating Function-Based Inference Procedures With Markov Chain Monte Carlo Samplers" (with discussion), *Journal of the American Statistical Association*, 102, 881–888. [715]
- Yin, G. (2009), "Bayesian Generalized Method of Moments" (with discussion), *Bayesian Analysis*, 4, 191–208. [715-717,720]