Journal of Economic Theory and Econometrics, Vol. 30, No. 3, Sep. 2019, 88–99

Monte Carlo Evidences on Finite Sample Performances of the Simulated Integrated Conditional Moment Estimator for the Binary Choice Model

Hosin Song*

Abstract In this paper, I propose a simulated integrated conditional moment (SICM) estimator for the binary choice model. The asymptotic property of the proposed SICM estimator is explored via Monte Carlo experiment since its asymptotic theory has not been fully developed. In particular, the SICM estimator is compared with method of simulated moment (MSM) and ML estimators by adopting a simple parametric distributional setup in the experiment. The experiment results show that the proposed SICM estimator is valid in the sense that it is consistent and its Monte Carlo variance decreases by $1/n$ times as the sample size increases. In particular, it is found that the variance of the SICM estimator is approximately twice that of the MSM estimator with one simulator.

Keywords simulated integrated conditional moment, method of simulated moment, binary choice model.

JEL Classification C15, C18, C25

*Associate Professor, Department of Economics, Ewha Womans University, Seoul, Korea. E-mail: hsong@ewha.ac.kr. I am grateful to the editor and two anonymous referees. I also would like to thank Yookyung Koh for her research assistance.

Received March 5, 2019, Revised June 14, 2019, Accepted August 1, 2019

1. INTRODUCTION

Recently many simulation based estimation methods have been proposed. McFadden (1989), Gourieroux, Monfort and Renault (1993), McFadden and Ruud (1994), Hajivassiliou, McFadden and Ruud (1996) and Bierens and Song (2012) are related literature. This study proposes another simulation based estimator for the binary choice model. The proposed estimator is also another application to exploit the idea of Bierens and Ploberger (1997) in line with Bierens and Ginther (2001), Bierens and Wang (2012), and Bierens and Song (2012).

The motivation of this study is twofold. One is to propose a simulated integrated conditional moment (SICM) estimator for the binary choice model. SICM estimation method has been exploited for a special model such as first-price auction models by Bierens and Song (2012, 2018). In this study, SICM estimation method is extended to the binary choice model where the dependent variable is zero or one.[1] The other is to explore the asymptotic property of the SICM estimator via Monte Carlo experiments since its asymptotic theory is not fully established yet. One main reason is that the objective function is not explicitly differentiable in parameters. For the simple and clear understanding of the distribution of the SICM estimator in the finite sample, the SICM estimator is compared with other well-known estimators: method of simulated moment (MSM) and ML estimators. For their comparison, a fixed parametric distribution is used for the experiment so that the distribution itself can not be a parameter to estimate. It makes the estimation much simpler and the SICM and MSM estimators can be comparable.[2] The mean squared error (MSE) of the SICM estimator is examined in the experiment to provide a snapshot of the distribution of the SICM estimator. In particular, the mean squared errors (MSEs) of the SICM estimator, the method of simulated moment (MSM) estimator and the probit estimator are presented to compare their performance. Unlike the SICM estimator, the asymptotic property of the MSM estimator has been well established by McFadden (1989), and McFadden and Ruud (1994). The MSM estimator has been used when the computation is very cumbersome even though the true distribu-

[1]For this extension, one may consider either parametric approach or semi-nonparametric approach. In principle, either is possible. However, this study adopts parametric approach so that SICM estimator can be compared with MSM estimator which requires the parametric distributional assumption.

[2]The asymptotic theory of SICM estimator becomes more complicated when the distribution becomes a parameter to estimate as in Bierens and Song (2012). Furthermore, the MSM can not be applied to the setup in Bierens and Song (2012). Hence, the SICM and MSM estimators can not be compared in the setup of Bierens and Song (2012).

tion of the error is known. However, SICM estimation can be applied to the case where the true distribution of the error is unknown if a semi-nonparametric density function is used as in Bierens (2008, 2014), and Bierens and Song (2012, 2018).[3]

The remainder of this paper is organized as follows. In section 2, we show the identification of the binary choice model when the distribution of the error belongs to a family of parametric distributions, and propose the SICM estimator. In section 3, Monte Carlo experiments are conducted to explore the asymptotic behavior of the SICM estimator, and to compare the SICM estimator with the MSM estimator. Section 4 has some concluding remarks.

2. BINARY CHOICE MODEL

2.1. IDENTIFICATION

Consider a binary choice model

$$Y = I(Y^* > 0) \text{ where } Y^* = X'\beta_0 + V \tag{1}$$

where the dependent variable Y is binary, $\beta_0 \in \mathbb{R}^K$ is the true parameter vector, Y^* is a latent variable and $I(\cdot)$ is the indicator function.[4]

Now I adopt the following standard assumption for the binary choice model.

Assumption 1. *(i) Equation (1) holds. (ii) V is independent of X.*

The probability of the event $Y = 1$ given X is

$$\Pr[Y = 1|X] = \Pr[V \geq -X'\beta_0|X] = 1 - \Pr[V \leq -X'\beta_0] \tag{2}$$

and the probability of its complementary event given X is

$$\Pr[Y = 0|X] = \Pr[V \leq -X'\beta_0|X] = \Pr[V \leq -X'\beta_0]. \tag{3}$$

Note the two last equations in (2)-(3) follow from the independence of V and X.

[3] Strictly speaking, Bierens (2008, 2014) exploit the semi-nonparametric(SNP) ML approach while Bierens and Song (2012, 2018) exploit SICM approach.

[4] $I(A) = 1$ if A is true, and $I(A) = 0$ otherwise.

Moreover, the following assumption is needed for the parametric identification.

Assumption 2. *(i) The distribution of V is absolutely continuous with respect to the Lebesgue measure. Specifically, for a known function G,* $\Pr[V \leq v_0] = G(v_0;\theta_0)$ *for all $v_0 \in \mathbb{R}$. (ii) The support of $X'\beta_0$ is the whole real number $\mathbb{R}$. (iii) The variance of X, $V(X)$, is positive definite.*

We restrict our interest to the parametric model where the distribution of the error V belongs to a well-known parametric distribution G. Hence, the true distribution is characterized by a parameter vector θ_0 as in Assumption 2 (i).

Now, we can easily show that the identification is achieved under Assumptions 1-2. Suppose that $(\underline{\beta},\underline{\theta})$ is observationally equivalent to $(\tilde{\beta},\tilde{\theta})$. Then,

$$\Pr[Y=1|X] = 1-G(-X'\underline{\beta};\underline{\theta}) = 1-G(-X'\tilde{\beta};\tilde{\theta}), \text{ and}$$

$$\Pr[Y=0|X] = G(-X'\underline{\beta};\underline{\theta}) = G(-X'\tilde{\beta};\tilde{\theta}).$$

Under Assumption 2 (i)-(ii), the same support implies

$$X'\underline{\beta} = X'\tilde{\beta} = t \text{ for all } t \in \mathbb{R}.$$

Then, $X'(\underline{\beta}-\tilde{\beta}) = 0$ and thus $(X-E(X))'(\underline{\beta}-\tilde{\beta}) = 0$. Therefore,

$$(\underline{\beta}-\tilde{\beta})'V(X)(\underline{\beta}-\tilde{\beta}) = 0.$$

Assumption 2 (iii) implies $\underline{\beta} = \tilde{\beta}$. Now we have the following relationship

$$G(s;\underline{\theta}) = G(s;\tilde{\theta}) \text{ for all } s \in \mathbb{R}$$

which implies

$$\underline{\theta} = \tilde{\theta}.$$

Assumption 2 (i) directly leads to

$$\underline{\theta} = \tilde{\theta} = \theta_0.$$

The parametric identification approach may look too strong from the point of nonparametric identification.[5] You may consider nonparametric identification

[5] One anonymous referee pointed out, this identification is essentially similar to one in Newey and Mcfadden (1994). Their identification condition is $E[XX']$ is finite and nonsingular. Moreover, they assume the distribution is standard normal, which is not necessary under Assumption 2.

for this problem. However, we do not handle nonparametric identification here since it is beyond the scope of this study.

2.2. SICM ESTIMATION

Following the idea of Bierens and Song (2012), we propose an SICM estimator for the binary choice model. Suppose two errors V and $\widetilde{V}$ have their distribution functions G and $\widetilde{G}$ respectively. Specifically, $\Pr[V \leq v] = G(v) \equiv G(v;\theta_0)$ and $\Pr[\widetilde{V} \leq v] = \widetilde{G}(v) \equiv G(v;\widetilde{\theta})$. Let $\widetilde{Y}_i = I(\widetilde{Y}_i^* \geq 0)$ where $\widetilde{Y}_i^* = X_i'\beta + \widetilde{V}_i$, and $\widetilde{V}_i$ is independent of X_i. Recall Y is defined in (1). The idea of SICM estimation is that the distribution Y given X is equivalent to that of $\widetilde{Y}$ given X provided that $\beta_0 = \beta$ and $G(v) = \widetilde{G}(v)$ for all $v \in \mathbb{R}$.[6] Recall Y is the actual dependent variable which is associated with β_0 and $G(v)$, while $\widetilde{Y}$ is the simulated dependent variable that is associated with β and $\widetilde{G}(v)$.

If two conditional distribution $Y|X$ and $\widetilde{Y}|X$ are same, then their conditional characteristic functions are the same and vice versa. Then,

$$E[\exp(\mathbf{i}t_0Y) - \exp(\mathbf{i}t_0\widetilde{Y})|X] = 0 \ \ \forall t_0 \in \mathbb{R} \tag{4}$$

which implies the following moment condition to estimate the parameter β_0.

$$E[(\exp(\mathbf{i}t_0Y) - \exp(\mathbf{i}t_0\widetilde{Y}))\exp(\mathbf{i}t_1'X)] = 0 \ \forall (t_0,t_1')' \in \Xi, \tag{5}$$

where $X \in \mathbb{R}^K$, $\Xi \equiv [-\overline{\tau},\overline{\tau}]^{1+K}$, $\overline{\tau} > 0$ is an arbitrary positive real number, and $\mathbf{i}$ is the complex number $\mathbf{i} = \sqrt{-1}$.

Note that (5) is equivalent to

$$\left|E\left((\exp(\mathbf{i}t_0Y) - \exp(\mathbf{i}t_0\widetilde{Y}))\exp(\mathbf{i}t_1'X)\right)\right|^2 = 0 \ \ \forall (t_0,t_1')' \in \Xi. \tag{6}$$

Therefore, the population objective function can be defined as

$$Q(\beta) = \int_{\Xi} |E[(\exp(\mathbf{i}t_0Y) - \exp(\mathbf{i}t_0\widetilde{Y}))\exp(\mathbf{i}t_1'X)]|^2 dt_0 dt_1. \tag{7}$$

Accordingly, the sample counterpart of (7) can be defined as

$$\widehat{Q}_n(\beta) = \int_{\Xi} \left| \left(\frac{1}{n}\sum_{j=1}^{n} \exp(\mathbf{i}t_0Y_j) - \frac{1}{n}\sum_{j=1}^{n} \exp(\mathbf{i}t_0\widetilde{Y}_j) \right) \exp(\mathbf{i}t_1'X_j) \right|^2 dt_0 dt_1 \tag{8}$$

[6] Note Assumption 2 enforces $G(v) = \widetilde{G}(v)$ for all $v \in \mathbb{R}$.

when X is bounded. Moreover, it is reduced to the following closed form.

$$
\begin{aligned}
\widehat{Q}_n(\beta) = & \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\sin(\overline{\tau}(Y_i - Y_j))}{\overline{\tau}(Y_i - Y_j)}\right)\times\left(\prod_{m=1}^{K}\frac{\sin(\overline{\tau}(X_{i,m} - X_{j,m}))}{\overline{\tau}(X_{i,m} - X_{j,m})}\right)\\
& - 2\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\sin(\overline{\tau}(\widetilde{Y}_i - Y_j))}{\overline{\tau}(\widetilde{Y}_i - Y_j)}\right)\times\left(\prod_{m=1}^{K}\frac{\sin(\overline{\tau}(X_{i,m} - X_{j,m}))}{\overline{\tau}(X_{i,m} - X_{j,m})}\right)\\
& + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\sin(\overline{\tau}(\widetilde{Y}_i - \widetilde{Y}_j))}{\overline{\tau}(\widetilde{Y}_i - \widetilde{Y}_j)}\right)\times\left(\prod_{m=1}^{K}\frac{\sin(\overline{\tau}(X_{i,m} - X_{j,m}))}{\overline{\tau}(X_{i,m} - X_{j,m})}\right). \qquad (9)
\end{aligned}
$$

Recall the simulated dependent variable $\widetilde{Y}$ depends on β and $\widetilde{G}(v)$. If $|X_i|$ is unbounded, a bounded one-to-one transformation of X_i, x_i^*, can be used. The dependent variables Y and $\widetilde{Y}$ are bounded, hence the objective function for the SICM estimation is

$$
\begin{aligned}
\widehat{Q}_n(\beta) = & \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\sin(\overline{\tau}(Y_i - Y_j))}{\overline{\tau}(Y_i - Y_j)}\right)\times\left(\prod_{m=1}^{K}\frac{\sin(\overline{\tau}(x^*_{i,m} - x^*_{j,m}))}{\overline{\tau}(x^*_{i,m} - x^*_{j,m})}\right)\\
& - 2\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\sin(\overline{\tau}(\widetilde{Y}_i - Y_j))}{\overline{\tau}(\widetilde{Y}_i - Y_j)}\right)\times\left(\prod_{m=1}^{K}\frac{\sin(\overline{\tau}(x^*_{i,m} - x^*_{j,m}))}{\overline{\tau}(x^*_{i,m} - x_{j,m})}\right)\\
& + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\sin(\overline{\tau}(\widetilde{Y}_i - \widetilde{Y}_j))}{\overline{\tau}(\widetilde{Y}_i - \widetilde{Y}_j)}\right)\times\left(\prod_{m=1}^{K}\frac{\sin(\overline{\tau}(x^*_{i,m} - x^*_{j,m}))}{\overline{\tau}(x^*_{i,m} - x^*_{j,m})}\right) \qquad (10)
\end{aligned}
$$

where $x^*_{i,m} = \arctan\big((X_{i,m} - \overline{X}_m)/S_m\big)$ is used as a bounded one to one transformation when the m-th component in X_i, $X_{i,m}$, is not bounded. S_m is the sample standard deviation of X_m. $\overline{\tau} = 1$ is used in the estimation by following Bierens and Song (2012).[7]

Therefore,

$$\hat{\beta}_{SICM} = \arg\min_{\beta}\widehat{Q}_n(\beta) \qquad (11)$$

where $\widehat{Q}_n(\beta)$ is defined in (10). Moreover,

$$\beta_0 = \arg\min_{\beta} Q(\beta) \qquad (12)$$

where

$$Q(\beta) = \int_{\Xi}\left|E\left[\left(\exp(\mathbf{i}t_0 Y_j) - \exp(\mathbf{i}t_0\widetilde{Y}_j)\right)\exp(\mathbf{i}t_1' x^*_j)\right]\right|^2 dt_0 dt_1. \qquad (13)$$

[7]Note that $|\arctan(t)| < \pi/2$ for any $t \in \mathbb{R}$.

The consistency of $\hat{\beta}_{SICM}$ can be achieved by applying the uniform law of large numbers and definitions of (11) and (12).[8] Also note $\widehat{Q}_n(\beta) \xrightarrow{p} Q(\beta)$ follows from the fact

$$n^{-1}\sum_{j=1}^{n}\exp(\mathbf{i}t_0 Y_j)\exp(\mathbf{i}t_1' x_j^*) \xrightarrow{p} E\left[\exp(\mathbf{i}t_0 Y_1)\exp(\mathbf{i}t_1' x_1^*)\right], \text{ and}$$

$$n^{-1}\sum_{j=1}^{n}\exp(\mathbf{i}t_0 \widetilde{Y}_j)\exp(\mathbf{i}t_1' x_j^*) \xrightarrow{p} E\left[\exp(\mathbf{i}t_0 \widetilde{Y}_1)\exp(\mathbf{i}t_1' x_1^*)\right]$$

where $x_1^* = (x_{1,1}^*, \ldots, x_{1,K}^*)'$. However, the asymptotic distribution theory of $\hat{\beta}_{SICM}$ is not trivially established since the objective function is not explicitly differentiable in parameters. Therefore, this study aims to provide a snapshot about its distribution via Monte Carlo experiments.

2.3. MSM ESTIMATION

One well-known simulation based estimator is the MSM estimator by McFadden (1989). The distribution of V is known as a distribution function $G(v;\theta_0)$.[9] Then we can define the MSM estimator for the binary choice model as follows.

$$\hat{\beta}_{MSM} = \arg\min_{\beta}\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widetilde{P}_i(\beta))W_i\right)'\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widetilde{P}_i(\beta))W_i\right)$$

where $\widetilde{P}_i(\beta) = \widetilde{\Pr}[Y_i = 1|X_i] = \frac{1}{S}\sum_{s=1}^{S} I(\widetilde{Y}_{s,i}^* \geq 0) = \frac{1}{S}\sum_{s=1}^{S} I(\widetilde{V}_{s,i}^* \geq -X_i'\beta)$ where $\widetilde{Y}_{s,i}^* = X_i'\beta + \widetilde{V}_{s,i}$, and $\Pr[\widetilde{V}_{s,i} \leq t] = G(t;\theta_0)$. Moreover, $\widetilde{V}_{s,i}$ is independent of X_i. S is the number of simulated errors $\widetilde{V}_{s,i}$ from a known distribution $G(v;\theta_0)$, which are needed to obtain a simulator $\widetilde{\Pr}[Y_i = 1|X_i]$. The asymptotic distribution theory was well established by McFadden (1989).

The performance of the MSM estimator may depend on the choice of instruments W_i. To have an idea of optimal instruments, think of the following

---

[8] Of course, the continuity assumption of $Q(\beta)$ and the uniqueness of β_0 are needed. In particular, the consistency is trivial when the distribution $G(\cdot;\theta_0)$ is known. This case is considered in this study.

[9] In general, θ_0 consists of a location parameter and scale parameter. For example, if $G(v;\theta_0) = \int_{-\infty}^{v}\frac{1}{\sqrt{2\pi}}\exp(-0.5t^2)dt$, then $\theta_0 = (\mu_0, \sigma_0^2) = (0, 1)$.

maximum likelihood estimation. Let $P_i(\beta) = \Pr[Y_i = 1|X_i] = 1 - G(-X'\beta; \theta_0)$. Then, the log likelihood function is defined as

$$\log L(\beta) = \sum_{i=1}^{n} [Y_i \log P_i(\beta) + (1 - Y_i)\log(1 - P_i(\beta))].$$

The first order condition can be obtained from the log likelihood function, and it can be used as the moment condition.

$$\frac{\partial \log L(\beta)}{\partial \beta} = \sum_{i=1}^{n} u_i \frac{\frac{\partial P_i(\beta)}{\partial \beta}}{P_i(\beta)(1 - P_i(\beta))} = 0 \tag{14}$$

where $u_i(\beta) \equiv Y_i - P_i(\beta)$ is treated as an error. It follows from (14) that $W_{o,i}(\beta) = (\partial P_i(\beta)/\partial \beta)/(P_i(\beta)(1 - P_i(\beta)))$ are optimal instruments in GMM since those instruments asymptotically deliver the ML estimator.[10] In the Monte Carlo experiment, we will consider two kinds of instruments: one is usual instruments W_c delivering a consistent estimator and the other is the optimal instruments W_o.

3. MONTE CARLO EXPERIMENTS

Monte Carlo experiments are conducted to see the validity of the SICM estimator for the binary choice model. The true data generating process for the experiment is

$$Y_i^* = \beta_0 X_i + V_i$$

where $V_i \sim \mathscr{N}(0,1)$, $X_i \sim \mathscr{E}(1)$ where $\mathscr{E}(1)$ is the exponential distribution with mean 1, and V_i is independent of X_i. The true value of $\beta_0 = -1$. There is no constant regressor in this setup so that Assumption 2 (iii) can be satisfied.

In the experiment, 1000 replications are conducted. Each replication consists of a sample of n observations, and SICM, MSM and probit ML estimations are conducted in each replication. We consider $n = 500, 1000$ and 2000 where the sample size doubles. In the SICM estimation, $\overline{\tau} = 1$ is used by following Bierens and Song (2012). In the MSM estimation, $S = 1$, $S = 10$, $S = 100$ and $S = 1000$ are considered for the number of simulations per observation. In particular, we consider two kinds of instruments. In the first case, we use an instrument $W_{c,i} = \exp(X_i)$ delivering a consistent MSM estimator. In the second case, we

[10] In practice, we can consider using $W_{o,i}(\overline{\beta})$ where $\overline{\beta}$ is the estimated value of any consistent estimator of β_0.

use the optimal instruments $W_{o,i} = \phi(-\hat{\beta}_c X_i)X_i / \Big((1-\Phi(-\hat{\beta}_c X_i))\Phi(-\hat{\beta}_c X_i)\Big)$ where $\hat{\beta}_c$ is the estimate of the MSM estimator obtained by using $W_{c,i} = \exp(X_i)$.[11] The performance of the probit ML estimator is also presented as a benchmark.

Table 1 presents the performance of the SICM estimator. The MSE decreases as the sample size increases, which confirms that the SICM estimator is consistent. Moreover, notice that the variance of the SICM estimator approximately halves as the sample size doubles. It implies that the asymptotic variance of the SICM estimator is well-defined as in usual CAN (consistent and asymptotically normal) estimators.[12]

For comparison, Tables 2-3 present the performance of the MSM estimator and the probit ML estimator. The MSM estimator shows smaller variance than the SICM estimator in all three sample sizes. The experiment results in Table 2 suggest that the SICM estimator has a larger variance than the MSM estimator. It is noticeable that the MSE of the MSM estimator with one simulator ($S = 1$) is about half the MSE of the SICM estimator in all sample sizes.[13] In other words, the MSE of MSM estimator with $(n,S) = (500,1)$ is close to that of the SICM estimator with $n = 1000$, and the MSE of MSM estimator with $(n,S) = (1000,1)$ is close to that of the SICM estimator with $n = 2000$.

Table 3 displays the performance of the MSM estimator using optimal instruments. In all sample sizes, the MSEs of the MSM estimators using $S = 100$ are almost the same as those of the probit ML estimators. Table 3 is presented in the Appendix.

	$n = 500$	$n = 1000$	$n = 2000$
MSE	0.0305	0.0169	0.0090
Bias	-0.0158	-0.0080	-0.0048
Variance	0.0303	0.0168	0.0090

Table 1: EXPERIMENT RESULTS OF SICM ESTIMATORS. $\overline{\tau} = 1$ is used for the SICM estimation. In each case, 1000 replications are conducted.

[11] In this sense, the optimal GMM takes the form of two-step estimation in this experiment.

[12] In other words, $\sqrt{n}(\hat{\beta}_{SICM} - \beta_0)$ converges in distribution to a certain random variable whose variance is well defined.

[13] Note their MSEs are mainly dependent on their variance since their squared bias is very small. Moreover, you can notice that the SICM estimator exploits one simulated dependent variable $\widetilde{Y}_j$ per one observation Y_j.

	$S=1$	$S=10$	$S=100$	$S=1000$	Probit
($n=500$)					
MSE	0.0153	0.0093	0.0088	0.0088	0.0081
Bias	-0.0146	-0.0094	-0.0106	-0.0113	-0.0112
Variance	0.0151	0.0092	0.0087	0.0086	0.0079
($n=1000$)					
MSE	0.0092	0.0049	0.0044	0.0044	0.0039
Bias	-0.0112	-0.0077	-0.0079	-0.0080	-0.0076
Variance	0.0091	0.0048	0.0044	0.0043	0.0039
($n=2000$)					
MSE	0.0044	0.0024	0.0022	0.0022	0.0019
Bias	-0.0036	-0.0021	-0.0033	-0.0032	-0.0028
Variance	0.0044	0.0024	0.0022	0.0022	0.0019

Table 2: EXPERIMENT RESULTS OF MSM ESTIMATORS WITH $W_{c,i}$. S is the number of simulators for each observation for the MSM estimator. In each case, 1000 replications are conducted.

4. CONCLUDING REMARKS

In this paper, we propose the SICM estimation for the binary choice model, and provide a snapshot of the asymptotic behavior of the SICM estimator via Monte Carlo experiments since the asymptotic distribution theory of SICM estimator has not been well developed.

For the simple and clear comparison, we restrict our attention to the binary choice model where the distribution of the error term is known. The experiment results show that the SICM estimator is consistent, and its variance decreases by $1/n$ times as the sample size increases. In particular, it is found that the variance of the SICM estimator is approximately twice that of the MSM estimator with one simulator.

Admittedly, there are some drawbacks with the SICM estimation. As seen in the experiment results, the dispersion of the SICM estimator is larger than the MSM estimator. In addition, SICM estimation requires much more computation time than MSM estimation.

In spite of those limitations, there is still much room for the application of the SICM estimation. First, the SICM estimator can be used even when the true distribution of the error is not known. Related to this, semi-nonparametric SICM estimation allowing for nonparametric specification for the error in the binary choice model can be a good complementary alternative to a semi-nonparametric (SNP) ML estimator as Bierens (2014).[14] Second, the SICM estimation can be used to test the correctness of the functional form by applying the idea of the integrated conditional moment (ICM) test in Bierens and Ploberger (1997). Therefore, the correctness of the estimated parameter and specified distribution can be tested. Even the true distribution is unknown, the correctness of the estimated parameter and distribution still can be tested by considering a semi-nonparametric (SNP) distribution in Bierens (2008, 2014).

APPENDIX

	$S = 1$	$S = 10$	$S = 100$	$S = 1000$	Probit
$(n = 500)$					
MSE	0.0145	0.0085	0.0081	0.0081	0.0081
Bias	-0.0120	-0.0099	-0.0106	-0.0110	-0.0112
Variance	0.0143	0.0084	0.0080	0.0079	0.0079
$(n = 1000)$					
MSE	0.0085	0.0044	0.0040	0.0039	0.0039
Bias	-0.0105	-0.0076	-0.0076	-0.0076	-0.0076
Variance	0.0084	0.0043	0.0039	0.0039	0.0039
$(n = 2000)$					
MSE	0.0041	0.0022	0.0019	0.0019	0.0019
Bias	-0.0034	-0.0020	-0.0028	-0.0027	-0.0028
Variance	0.0041	0.0022	0.0019	0.0019	0.0019

Table 3: EXPERIMENT RESULTS OF MSM ESTIMATORS WITH $W_{o,i}$. S is the number of simulators for each observation for the MSM estimator. In each case, 1000 replications are conducted.

[14]Bierens and Song (2018) is one example.

REFERENCES

Bierens, H. J., and W. Ploberger (1997). Asymptotic Theory of Integrated Conditional Moment Tests, Econometrica, 65, 1129-1151.

Bierens, H. J., and D. Ginther (2001). Integrated Conditional Moment Testing of Quantile Regression Models, Empirical Economics, 26, 307-324.

Bierens, H. J. (2008). Semi-nonparametric Interval-Censored Mixed Proportional Hazard Models: Identification and Consistency Results,Econometric Theory, 24, 749-794.

Bierens, H. J., and L. Wang (2012). Integrated Conditional Moment Tests for Parametric Conditional Distributions,Econometric Theory, 28, 328-362.

Bierens, H. and H. Song (2012). Semi-Nonparametric Estimation of Independently and Identically Repeated First-Price Auctions via an Integrated Simulated Moments Method,Journal of Econometrics, 168, 108-119.

Bierens, H. J. (2014). Consistency and Asymptotic Normality of Sieve ML Estimators Under Low-Level Conditions, Econometric Theory, 30, 1021-1076.

Bierens, H. J., and H. Song (2018). Semi-Nonparametric Modeling and Estimation of First-Price Auctions Models with Auction-Specific Heterogeneity, Working paper.

Gourieroux, C., A. Monfort and E. Renault (1993). Indirect Inference, Journal of Applied Econometrics, 8, 85-118.

Hajivassiliou, V., D. Mcfadden and P. Ruud (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results, Journal of Econometrics, 72, 85-134.

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

McFadden, D., and P. A. Ruud (1994). Estimation by Simulation, Review of economics and statistics, 76, 591-608.

Newey, W. K., and D. McFadden (1994). Large Sample Estimation and Hypothesis Testing, Handbook of Econometrics, 4, 2111-2130.