Journal of Mathematics and Statistics 10 (2): 111-116, 2014
ISSN: 1549-3644

doi:10.3844/jmssp.2014.111.116 Published Online 10 (2) 2014 (http://www.thescipub.com/jmss.toc)

EMPIRICAL LIKELIHOOD ESTIMATION BASED ON SIMULATED MOMENT CONDITIONS

Xing Wang

Department of Economics and Finance, Durham University Business School, Mill Hill Lane, Durham DH1 3LB, UK

Received 2014-01-22; Revised 2014-02-03; Accepted 2014-02-17

ABSTRACT

In this study we discuss the optimization of the Empirical Likelihood (EL) criterion function when the moment condition is nonstandard. We deal with this issue following the Method of Simulated Moment (MSM) introduced and we use importance sampling method to smooth discrete moment conditions. We have demonstrated the convergence and asymptotic normality of the empirical likelihood estimator from the simulated moment conditions.

Keywords: Empirical Likelihood, Simulated Moments, Importance Sampling

1. INTRODUCTION

Recently the Empirical Likelihood (EL) method has been increasingly popular in statistics and econometrics as an alternative to GMM, due to its desirable higher order properties, Owen and Zhou (2000) for a comprehensive introduction, Newey and Smith (2004) for higher order asymptotics and Wang (2013) for large deviation efficiency of the EL test with weakly dependent data, among others. In this study we contribute to the literature by addressing how EL deals with non standard moment conditions as:

$$E[g(x,\theta_0)] = 0$$

where, x is the observed data, θ_0 is the parameter to be estimated and g is a nonstandard function in the sense that g is diff cult to compute or can even benon-smooth. In this case both the Generalised Method Of Moments (GMM) and EL will be difficult to apply because they require explicit calculation of the sample analogue of the moment condition and existence of the derivative of g (x, θ) with respect to θ.

Pakes and Pollard (1989) (hereafter PP) and McFadden and Ruud (1994) (here-after MR) deal with this problem in GMM by simulating the moment condition where it is hard to compute. In our paper we show that the simulation method can also be used in EL and furthermore, we use importance sampling to evaluate the moment condition through observations from a different probability distribution which is relatively easy to handle. Also, we notice that as McFadden (1989) points out, importance sampling can be used to smooth discrete moment conditions, therefore we extend our estimation method to more general case where the moment conditions can even be discrete. Similar to the results of PP, the proof of consistency of our EL estimator based on simulation requires only the continuity of the simulated moment condition, but not that of the original one.

2. EMPIRICAL LIKELIHOOD WITH NON STANDARD MOMENT CONDITION

Consider the following moment condition model Equation (1):

$$E[g(x,\theta_0)] = 0 \quad (1)$$

where, x is the observed data, $\theta_0 \in \Theta \subset \mathbb{R}$ is the parameter to be estimated and g is a real function. Following the well established procedures, (e.g., Qin and Lawless (1994) and Newey and Smith (2004), the EL estimator based on (1) is defined as:

$$\hat{\theta} \equiv \arg\min_{\theta\in\Theta} \sup_{\lambda\in\mathbb{R}} R(\theta,\lambda) \quad (2)$$

where, Equation (3):

$$R(\theta,\lambda) \equiv \sum_{n=1}^{N} \log\left(1+\lambda' g(x,\theta)\right) \quad (3)$$

And λ is a vector of Lagrangian multipliers.

A problem in empirical likelihood estimation of θ by minimizing (3) is that g(.) is sometimes intractable, i.e., not in an explicit form, so that we cannot calculate its sample analogue, nor we can get its derivative. Another situation is that sometimes g(.) is not continuous in θ, but usually empirical likelihood estimation assumes that g(.) should be continuous and differentiable in the parameter of interest, so that we can demonstrate the consistency of EL estimator (see, e.g., assumption 1 of Newey and Smith (2004). To summarize these situations we list the following cases.

Case 1. g (.) is discontinuous in θ.

Example 1

McFadden (1989) considered estimation of discrete response model. Suppose we have obtained the model like Equation (4):

$$y_i = I(\beta x_i + \varepsilon_i > 0) \quad (4)$$

where, I(.) is the indicator function and ε_i is i.i.d with density p(ε). So we have the moment conditions $E[g(x,\beta)] \equiv E[y_i - I(\beta x_i + \varepsilon_i > 0)]$ and the GMM estimator $\hat{\beta}$ is based on the following sample analogue:

$$\hat{g}(x,\beta) = \frac{1}{N}\sum_{I=1}^{N}\left[y_i - I\left(\beta x_i + \varepsilon_i > 0\right)\right]$$

Problems arises because $\hat{g}(x, \beta)$ is not continuous in β.

Case 2: Computation of g(.) is infeasible.

To overcome these problems in GMM, Pakes and Pollard (1989) considered simulating a good estimate $\tilde{g}(.)$ instead of using $\hat{g}(.)$ directly. Specifically, if we let $G_n(\theta)$ be a simulation of E [g(x, θ)] and $\tilde{\theta}$ be the GMM estimator based on $G_n(\theta)$, then the conditions under which $\tilde{\theta}$ converges to θ_0 are described in the following theorem.

Theorem 1

(Pakes and Pollard, 1989) $\tilde{\theta}$ converges in probability to θ_0 if:

- $\left\|G_n(\tilde{\theta})\right\| \le \inf_{\theta\in\Theta}\left\|G_n(\theta)\right\| + o_p(1)$
- $G_n(\theta_0) = o_p(1)$
- $\sup_{\|\theta-\theta_0\|>\delta}\left\|G_n(\theta)\right\|^{-1} = O_p(1), \forall\delta > 0$ where, $\|\cdot\|$ is some norm depending on θ

Remarks

The intuitions of these conditions is to require the simulation $G_n(.)$ be as closeto E[g(x, θ)] as possible. Specifically:

- $G_n(.)$ evaluated at the estimator $\tilde{\theta}$ cannot be much bigger than the smallest value of $G_n(\theta)$ in Θ
- $G_n(.)$ evaluated at the true parameter θ_0 cannot be much bigger than zero
- $G_n(.)$ evaluated outside some neighborhood of θ_0 should be large

To use the results of this theorem in EL, we consider a specific simulation method Importance Sampling which is introduced in the next section.

3. IMPORTANCE SAMPLING

Importance sampling is a simulation method which is useful to estimate an integral about a probability distribution from a different distribution. Suppose we want to evaluate the integral:

$$E_p[g(x)] = \int_D g(x)p(x)dx$$

where, g(x) is a function of x and p(x) is the density of x. If it is difficult to sample from p(x), we can choose another probability distribution Q(x) with density q(x), which is called the importance function and has the same support as p(x) and transform E_p [g(x)] as Equation (5):

$$E_p[g(x)] = \int g(x)\frac{p(x)}{q(x)}q(x)dx = E_q[g(x)w(x)] \quad (5)$$

where, w(x) = p(x)/q(x) is called the importance weight (also inverse likelihood ratio). Note that w(x) is always positive, E_q [w(x)] = 1 and this weight function reflects the important regions of the sampling space. A special case is that q(x) = p(x), when w(x) = 1.

About (5) motivates an unbiased estimator for E_p [g(x)] by sampling S independent values from Q(x) and calculating Equation (6):

$$\frac{1}{S}\sum_{s=1}^{S} g(x_{ns})w(x_{ns}) \quad (6)$$

As simulated value of g(x)w(x): Hence E_p [g(x)] can be estimated by Equation (7):

$$\tilde{E}_p[g(x)] = \frac{1}{NS}\sum_{n=1}^{N}\sum_{s=1}^{S} g(x_{ns})w(x_{ns}) \quad (7)$$

Note that g(x)w(x) is an unbiased estimator of E_p [g(x)] by construction, with expectation taken with respect to q(x). It is interesting to check the expectation of g(x)w(x) with respect to p(x). Generally it will depend on the choice of q(x), but in some circumstances this expectation can be bounded by a function that does not depend on the choice of q(x). The following result will be useful later.

Proposition 1

Assume that g(x) is nonnegative and the importance weight w(x)=p(x)/q(x) is infinitely integrable, i.e., $E_p[w(x)^{\infty}]<M$, where M is finite, then E_p [g(x)w(x)] is also bounded, in particular Equation (8):

$$E_p[g(x)w(x)] \le E_p[g(x)]M \quad (8)$$

Proof

The result is directly from the Hölder inequality:

$$E_p[g(x)w(x)] = \int g(x)\frac{p(x)}{q(x)}p(x)dx \le \left(\int g(x)p(x)dx\right)$$
$$\|w(x)\|_{\infty} \le E_p[g(x)]M$$

where, $\|.\|_{\infty}$ denotes the norm in L^{∞} space.

4. LARGE SAMPLE THEORY

Now we replace E [g (x, θ)] in the original model (1) by its simulated version computed by (7) through importance sampling and define Equation (9):

$$\tilde{g}(x_n,\theta) \equiv E[g(x,\theta)] - \tilde{E}_p[g(x,\theta)] \quad (9)$$

As mentioned above, $\tilde{E}_p[g(x,\theta_0)]$ is unbiased, so $\tilde{E}_p[g(x,\theta_0)]$ and therefore can be used as a new moment condition to estimate θ_0. We further define Equation (10 to 12):

$$\tilde{g}(\theta) \equiv \frac{1}{N}\sum_{n=1}^{N}\tilde{g}(x_n,\theta) \quad (10)$$

$$\tilde{G} \equiv E\left[\frac{\partial\tilde{g}(x_n,\theta_0)}{\partial\theta}\right] \quad (11)$$

And:

$$\tilde{\Omega} \equiv E[\tilde{g}(x_n,\theta_0)'\tilde{g}(x_n,\theta_0)] \quad (12)$$

And let their counterparts from g (x, θ) be defined analogously and denoted without accent above, e.g. $g,(\theta) \equiv \frac{1}{N}\sum_{n=1}^{N} g(x_n,\theta)$. To apply the results of theorem 1, we define the empirical likelihood estimator $\tilde{\theta}$ as the solution to the following problem Equation (13):

$$\tilde{R}(\tilde{\theta},\tilde{\gamma}) \le \min_{\theta}\sup_{\gamma\in\mathbb{R}^p}\tilde{R}(\theta,\gamma) + o_p^{(N-1)} \quad (13)$$

Where:

$$\tilde{R}(\theta,\gamma) = \frac{1}{N}\sum_{n=1}^{N}\log(1+\gamma'\tilde{g}(x_n,\theta))$$

And γ is a vector of Lagrangian multipliers which is a function of θ implicitly defined through:

$$\frac{1}{N}\sum_{n=1}^{N}\frac{\tilde{g}(x_n,\theta)}{1+\gamma'\tilde{g}(x_n,\theta)} = 0$$

e.g., Qin and Lawless (1994).

For the general asymptotic properties of empirical likelihood estimator, we make the following regularity assumption.

Assumption 1

$\theta_0 \in$ int (Θ) and Θ is a compact subset of $\mathbb{R}^p$:

- $E[\sup_{\theta\in\Theta}\|g(x,\theta)\|^{\in}] < \infty, \forall \in > 2$
- $\Omega \equiv E[g(x,\theta_0)'g(x,\theta_0)]$ is nonsingular

Assumption 2

For any δ>0, $\sup_{\|\theta-\theta_0\|>\delta}\|g(\theta)\|^{-1} = O_p\left(N^{-1}\right)$.

Furthermore, we need a smoothing condition for uniform convergence. Let the simulation residual process defined as Equation (14):

$$\omega(\theta) = \sqrt{N}(\tilde{g}(\theta) - E_p[\tilde{g}(x,\theta)]) \quad (14)$$

Assumption 3

The process $\omega(\theta)$ is stochastically equicontinuous at θ_0, i.e., for any $\epsilon > 0$, there exists a neighborhood U of θ_0, which satisfies:

$$\sup_{\theta \in U} \left|\omega(\theta) - \omega(\theta_0)\right| \le \varepsilon \quad \text{a.s}$$

The following theorem demonstrates the consistency of $\tilde{\theta}$, by checking similar conditions given in theorem 1.

Theorem 2

Given assumption 6-8, we have the following results:

- $\sup_{\|\theta-\theta_0\|>\delta} \|\tilde{g}(\theta)\|^{-1} = O_p^{(N-1)}$
- $\tilde{g}(\theta_0) = o_p(1)$
- $\tilde{g}(\tilde{\theta}) = o_p(1)$
- $\tilde{R}(\theta_0, \bar{\gamma}) = O_p\left(N^{-1/2}\right)$

where, $\bar{\gamma} = \arg\sup_\gamma \tilde{R}(\theta_0, \gamma)$

And then $\tilde{\theta}$ converges in probability to θ_0.

Proof

The first result is to say that $\tilde{g}(\theta)$ is big outside some neighborhood of θ_0, which is from the identification of θ_0. To see this, note that from triangle inequality we have:

$$\begin{aligned}
&\sup_{\|\theta-\theta_0\|>\delta} \|\tilde{g}(\theta)\| = \sup_{\|\theta-\theta_0\|>\delta} \|g(\theta)\| - \|\tilde{g}(\theta)\| - g\|(\theta)\| \\
&\ge \sup_{\|\theta-\theta_0\|>\delta} \|g(\theta)\| - \sup_{\|\theta-\theta_0\|>\delta} \|\tilde{g}(\theta)\| - g\|(\theta)\| \\
&\ge \sup_{\|\theta-\theta_0\|>\delta} \|g(\theta)\| - \sup_{\theta} \|\tilde{g}(\theta) - g(\theta)\|
\end{aligned}$$

Given the assumption 3 of stochastic equicontinuity, $\sup_\theta \|\tilde{g}(\theta) - g(\theta)\| = o_p(1)$ and with assumption 2 we have $\mathrm{Sub}_{\|\theta-\theta_0\|>\delta} \|\tilde{g}(\theta)\|^{-1} = O_p(N-1)$.

Secondly we follow the way of McFadden (1989), McFadden and Ruud (1994), where $\sqrt{N}\tilde{g}(\theta)$ is decomposed as Equation (15):

$$\sqrt{N}\tilde{g}(\theta) = A_N + [\omega(\theta)] - \omega(\theta_0)] + B_N(\theta) + C_N(\theta) \quad (15)$$

Where:

$$A_N \equiv g(z,\theta_0) + \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\tilde{g}(x_n,\theta_0) - E_q[\tilde{g}(x_n,\theta_0)])$$

$$C_N(\theta) \equiv \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (\tilde{g}(x_n,\theta_0) - g(x_n,\theta_0)$$

$$B_N(\theta) \equiv \frac{1}{\sqrt{N}} \sum_{n=1}^{N} (E_q[\tilde{g}(x_n,\theta)] - g(x_n,\theta))$$

McFadden and Ruud (1994) have shown that $A_N = o_p^{(N1/2)}$, with i.i.d assumption on the observations and simulations. Also note that $C_N(\theta_0) = o_p(N^{1/2})$ and $B_N(\theta_0) = 0$, so we have $\sqrt{N}\tilde{g}(\theta_0) = o_p\left(N^{1/2}\right) + o_p(N^{1/2})$ and hence $\tilde{g}(\theta_0) = o_p(1)$.

To see the third results, a first order Taylor expansion of $\tilde{R}(\theta,\gamma)$ $\gamma = 0$ gives:

$$\tilde{R}(\theta,\gamma) = \gamma'(\tilde{g}(\theta)) - \frac{1}{2}\gamma' \left[\frac{1}{N}\sum_{n=1}^{N} \frac{\tilde{g}(x_n,\theta)\tilde{g}(x_n,\theta)'}{1+\dot{\gamma}'\tilde{g}(x_n,\theta)^2}\right]\gamma \quad (16)$$

where, $\dot{\gamma}$ lies between 0 and γ. According to Lemma A1 and A2 of Newey and Smith (2004) we have $\tilde{\gamma} = O_p(N^{-1/2})$ and $\frac{1}{(1+\dot{\gamma}'\tilde{g}(x_n,\theta))^2} \le -1/2$. Thus from (16) and result 1 we have:

$$\begin{aligned}
&\tilde{R}(\theta_0,\bar{\gamma}) \le O_p(N^{-1/2}) o_p(1) \\
&+O_p(N^{-1})\left(\frac{1}{N}\sum_{n=1}^{N} \tilde{g}(x_n,\theta)'\tilde{g}(x_n,\theta)\right) \\
&= O_p(N^{-1/2}) + O_p(N^{-1}) \\
&= O_p(N^{-1/2})
\end{aligned}$$

Now from the definition of $\tilde{\theta}$ we have Equation (17):

$$\begin{aligned}
&\tilde{R}(\tilde{\theta},\tilde{\gamma}) = O_p(N^{-1/2})\tilde{g}(\tilde{\theta}) + O_p(N^{-1}) \\
&\le \min_\theta \sup_{\gamma \in \mathbb{R}^p} \tilde{R}(\theta,\gamma) + o_p(N^{-1}) \\
&\le \tilde{R}(\theta_0,\tilde{\gamma}) + o_p(N^{-1}) \\
&= O_p(N^{-1/2})
\end{aligned} \quad (17)$$

Solving $\tilde{g}(\tilde{\theta})$ out of (17) gives Equation (18):

$$\left\|\tilde{g}(\tilde{\theta})\right\| = o_p(1) \tag{18}$$

Then the following argument is similar to Pakes and Pollard (1989). By result 1we have just proved, for arbitrary δ>0, there exists a bounded, positive constant M such that $\sup_{\|\theta-\theta_0\|>\delta}\left\|\tilde{g}(\tilde{\theta})\right\|^{-1} < M$. On the other hand, since $\left\|\tilde{g}(\tilde{\theta})\right\|$ is $o_p(1)$, for N large enough $\left\|\tilde{g}(\tilde{\theta})\right\|^{-1} > M$ with probability approaching one. Hence:

$$\sup_{\|\theta-\theta_0\|>\delta}\left\|\tilde{g}(\theta)\right\|^{-1} < M\left\|\tilde{g}(\tilde{\theta})\right\|^{-1}$$

Which implies $\tilde{\theta}$ must be within the neighborhood of θ_0 of radius δ, by noting that $\tilde{g}(\theta)$ is continuous. The convergence follows since δ can be arbitrary small.

Remarks

The consistency of $\tilde{\theta}$ does not depend on the choice of number of simulations S, although S does affect the asymptotic efficiency of $\tilde{\theta}$.

Assumption 4

g(x, θ) is differentiable at θ_0 and G = E[∂g(x,θ_0)/∂θ] is of full rank.

Theorem 3

Given assumption 1-3, $\sqrt{n}(\tilde{\theta}-\theta_0) \xrightarrow{d} N(0,V)$, where:

$$V = (G'\tilde{\Sigma}^{-1}G)^{-1}$$

Proof

Firstly we show that $\sqrt{n}(\tilde{\theta}-\theta_0)$ is stochastically bounded. Since $\hat{g}(\tilde{\theta}) = o_p(1)$ hence $C_N(\hat{\theta}) = O_p(1)$ and by expanding $C_N(\tilde{\theta})$ we have:

$$C_N(\tilde{\theta}) = \sqrt{n}(\tilde{\theta}-\theta_0)\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial m(x_n,\theta_0)}{\partial\theta} + O\left(\tilde{\theta}-\theta_0\right)\right)\right)$$
$$= O_p(1)$$

With the consistency $\tilde{\theta} \xrightarrow{p} \theta_0$ we have $\sqrt{n}(\tilde{\theta}-\theta_0) = O_p(1)$. Based on theorem 1, the following proof is similar to Parente and Smith (2008). Now we define Equation (19):

$$(\dot{\theta},\dot{\gamma}) = \arg\min_{\theta}\sup_{\gamma\in\mathbb{R}^p}\tilde{R}(\theta,\gamma) \tag{19}$$

Let $G_n(\theta) = \partial\tilde{m}_m(\theta)/\partial\theta, G(\theta_0) = \frac{1}{N}\sum_{n=1}^{N}G_n(\theta_0), \tilde{\Omega}_n$
$= \frac{1}{N}\sum_{n=1}^{N}\tilde{m}_n(\theta_0), \tilde{m}_n(\theta_0)'$.

Expand the first order condition for the saddlepoint problem of (??) around θ_0 and γ_0 = 0 Equation (20 and 21):

$$\frac{\partial\tilde{R}(\theta,\gamma)}{\partial\theta} = 0 = -\sum_{n=1}^{N}\frac{G_n(\dot{\theta})'\gamma}{1+\dot{\gamma}'(b_n-\tilde{m}_n(\dot{\theta}))}$$
$$\simeq \frac{1}{N}\sum_{n=1}^{N}G_n(\theta_0)'\dot{\gamma} \tag{20}$$

$$\frac{\partial\tilde{R}(\theta,\gamma)}{\partial\gamma} = 0 = -\sum_{n=1}^{N}\frac{\tilde{m}_n(\dot{\theta})'}{1+\dot{\gamma}'(b_n-\tilde{m}_n(\dot{\theta}))}$$
$$\simeq -\tilde{g}(\theta_0) - \frac{1}{N}\sum_{n=1}^{N}G_n(\theta_0)(\theta-\dot{\theta}_0) + \tilde{\Omega}_n\dot{\gamma} \tag{21}$$

About (20) and (21) imply:

$$\sqrt{N}(\dot{\theta}-\theta) \simeq -(G(\theta_0)'\tilde{\Omega}_n^{-1}G(\theta_0))^{-1}G(\theta_0)\tilde{\Omega}_n^{-1}\sqrt{N}\tilde{g}(\theta_0)$$

Note that from Lemma 1 we have:

$$\sqrt{N}\tilde{g}(\theta_0) \xrightarrow{d} N(0,\tilde{\Sigma})$$

Also from i.i.d assumption and unconditional simulation:

$$\frac{1}{N}\sum_{n=1}^{N}G_n(\theta_0) \xrightarrow{p} E[G_n(\theta_0)] = G$$

So $\sqrt{n}(\dot{\theta}-\theta_0) \to N(0,V)$. Next we show $\tilde{\theta}$ and $\dot{\theta}$ are asymptotically equivalent.

The definition of $\tilde{\theta}$ implies:

$$\tilde{R}(\tilde{\theta},\tilde{\gamma}) \le \tilde{R}(\dot{\theta},\dot{\gamma}) + o_p(N^{-1}) \le \tilde{R}(\dot{\theta},\tilde{\gamma}) + o_p(N^{-1})$$

Then with the similar expansion as (17) we have:

$$O_p(N^{-1/2})\tilde{g}(\tilde{\theta}) + O_p(N^{-1})$$
$$\le O_p(N^{-1/2})\tilde{g}(\dot{\theta}) + O_p(N^{-1}) + o_p(N^{-1})$$
$$\Rightarrow \tilde{g}(\tilde{\theta}) - \tilde{g}(\dot{\theta})O_p(N^{-1/2})$$

So $\tilde{g}(\tilde{\theta}) - \tilde{g}(\dot{\theta}) = o_p(1)$. Thus according to the continuity of $\tilde{g}$ we have $\tilde{\theta} = \dot{\theta} + o_p(1)$.

5. CONCLUSION

We have presented EL estimation with moment condition which is intractable and we also mentioned that simulation by importance sampling can be used to smooth moment condition with discreteness in parameter. This is a different way from Parente and Smith (2008) approach. Rather than simulating the moment indicator, they put different assumption on it to ensure the EL estimator to have standard first order asymptotic properties.

It is important to note that these asymptotic results of our estimator rely heavily on i.i.d assumptions on observations and simulations and for time series model our EL estimator may fail since the general conditions for uniform convergence and the law of large numbers will not be satisfied. So if we want to use EL by simulating moment conditions with dependent data through importance sampling, more assumptions on stochastic convergence (e.g., Pollard (1984) and chapter 4 of Billingsley (2013) should be added and the choice of importance function should also be carefully considered, to make the simulated moments satisfy certain conditions. These are the directions of our further research.

6. REFERENCES

Billingsley, P., 2013. Convergence of Probability Measures. 2th End., John Wiley and Sons, ISBN-10: 111862596X.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica, 57: 995-1026. DOI: 10.2307/1913621

McFadden, D. and P. Ruud, 1994. Estimation by Simulation. Rev. Econom. Stat., 76: 591-608. DOI: 10.2307/2109765

Newey, W.K. and R.J. Smith, 2004. Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica, 72: 219-255. DOI: 10.1111/j.1468-0262.2004.00482.x

Owen, A. and Y. Zhou, 2000. Safe and effective importance sampling. J. Am. Stat. Assoc., 95: 449, 135-143. DOI: 10.1080/01621459.2000.10473909

Parente, P. and R.J. Smith, 2008. GEL methods for nonsmooth moment indicators. Cemmap working paper CWP19/08.

Pakes, A. and D. Pollard, 1989. Simulation and the asymptotics of optimization estimators. Econometrica, 57: 1027-1057. DOI: 10.2307/1913622

Pollard, D., 1984. Convergence of Stochastic Processes. 1st Edn., Springer-Verlag, New York, ISBN-10: 0387909907, pp: 215.

Qin, J. and J. Lawless, 1994. Empirical likelihood and general estimating equations. Ann. Stat., 22: 300-325. DOI: 10.1214/aos/1176325370

Wang, X., 2013. The rate function of s-mixing process and its application in evaluating empirical likelihood tests. J. Math. Stat., 9: 334-338. DOI: 10.3844/jmssp.2013.334.338