Journal of Mathematics and Statistics 10 (2): 125-129, 2014
ISSN: 1549-3644

doi:10.3844/jmssp.2014.125.129 Published Online 10 (2) 2014 (http://www.thescipub.com/jmss.toc)

THE EFFICIENCY OF EMPIRICAL LIKELIHOOD WITH NUISANCE PARAMETERS

Xing Wang

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

Received 2014-02-03; Revised 2014-02-14; Accepted 2014-02-17

ABSTRACT

In this study we check the asymptotic efficiency of empirical likelihood in the presence of nuisance parameters combined with augmented moment conditions. We show that in the presence of nuisance parameters, the asymptotic efficiency of the empirical likelihood estimator of the parameter of interest will increase by adding more moment conditions, in the sense of the positive semidefiniteness of the difference of information matrices. As a by product, we point out a necessary condition for the asymptotic efficiency to be increased when more moment condition are added.

Keywords: Empirical Likelihood, Nuisance Parameters

1. INTRODUCTION

Likelihood inference may have some drawbacks when estimating a parameter of interest in the presence of nuisance parameters. For example, (Neyman and Scott, 1948) considered this problem and found that the maximum likelihood estimation could be either inconsistent or inefficient with many nuisance parameters. This study deal with Empirical Likelihood (EL), which is a nonparametric analogue of maximum likelihood, in the presence of nuisance parameters combined with selection of moment conditions. We show that with the existence of nuisance parameters, the asymptotic efficiency of the empirical likelihood estimator of the parameter of interest can be increase by adding more moment conditions, in the sense of the positive semidefiniteness of the difference of information matrices. Particularly, we focus on a special case, where nuisance parameters only appear in some of the moment conditions. This case leads to an important result that the asymptotic efficiency can increase with added moment condition only if it is not orthogonal with the original moment conditions.

2. MOMENT CONDITION WITH NUISANCE PARAMETERS

Consider a sequence of i.i.d. realizations $\{x_i\}_{i=1}^n$ of a random variable x from an unknown distribution F, with n being the sample size. Let θ be a p-dimensional vector of parameters in a compact parameter space $\Theta \subset \mathbb{R}^p$ associated with F. Suppose that for a true value of θ which is denoted as $\theta_0, \{x_i\}_{i=1}^n$ satisfies the following moment condition Equation (1):

$$E\left[g(x_i;\theta)\right] = 0 \quad (1)$$

where, g is a m×1 vector of real functions and the expectation is taken with respect to F. We consider the over-identified case where m≥p. Unlike (Qin and Lawless, 1994), we don't assume that the m functions of g are independent, since correlation between these functions plays an important role in the aspect of asymptotic efficiency, which we will discuss in the following section.

Now suppose the parameter θ can be decomposed as $\theta = (\beta', \phi')'$ with corresponding $\theta_0 = (\beta'_0, \phi'_0)'$ where $\beta \in B \subset \mathbb{R}^q, \phi \in \Phi \subset \mathbb{R}^{p-q}$ and $\Theta = B \times \Phi$. If we are only interest in β but not in ϕ, then ϕ is a nuisance parameter in the model and we write the corresponding moment condition as Equation (2):

$$E\left[g(x_i;\beta,\phi)\right] = 0 \quad (2)$$

For the true value β_0 of β. The empirical likelihood ratio statistic for this model is Equation (3):

$$\Re(\beta,\phi) = 2\sum_{i=1}^{n}\log\left(1+\lambda' g(x_i;\beta,\phi)\right) \quad (3)$$

where, λ is an m×1 vector of Lagrangian multipliers, which is a continuous differentiable function of $(\beta',\phi')'$ (see, e.g., (Qin and Lawless, 1994) and is determined by Equation (4):

$$\frac{1}{n}\sum_{i=1}^{n}\frac{g\left(x_i;\beta,\phi\right)}{1+\lambda' g\left(x_i;\beta,\phi\right)} = 0 \quad (4)$$

To simplify notations, let:

$$g\left(x_i;\theta\right) = g_i(\theta), \hat{g}(\theta) = n^{-1}\sum_{i=1}^{n} g_i(\theta)$$

$$G_1 = E\left[\frac{\partial g\left(x_i;\theta_0\right)}{\partial\theta}\right], \Omega_{11} = E\left[g\left(x;\theta_0\right)g\left(x;\theta_0\right)'\right]$$

Like ordinary parametric likelihood, empirical likelihood deals with nuisance parameter by profiling out ø (see, e.g., section 3.5 of (Owen, 2001)). Let $\tilde{\phi} = \tilde{\phi}(\beta)$ be the minimizer of R(β,ϕ)with respect to ϕ. The profile log-empirical likelihood ratio for β is Equation (5):

$$\Re(\beta) = \min_{\phi\in\Phi}\Re(\beta,\phi) \quad (5)$$

And EL estimator for β is Equation (6):

$$\beta = \arg\min_{\beta\in B}\Re(\beta) \quad (6)$$

Assumption 1

$\theta_0 = (\beta_0, \phi_0)$ solves E[g(x; θ)] = 0 uniquely, or equivalently, both β_0 and ϕ_0 are strongly identified.

Remark 1

This condition combined with m≥p makes the parameter well identified. In the study of (Stock and Wright, 2000), they considered the problem of weak identification of the parameter, by assuming that the subvector β of θ is completely identified, but ϕ is not, in the sense that the population moment function is steep in β around β_0 but is nearly flat in α. This idea provides us a framework to analysis problems mixed with nuisance parameters, weak identification and partial identification (Phillips, 1989).

Assumption 2

(a) $\theta_0 \in$ int (Θ). (b) Ω_{11} is positive definite and nonsingular (c). g(x, θ) is twice continuously differentiable in a neighborhood of θ_0 and G_1 is of full rank p (d). $\|\partial g(x,\theta)\|^3$, $\|\partial g(x,\theta)/\partial\theta\|$ and $\|\partial^2 g(x,\theta)/\partial\theta\partial\theta'\|$ are all bounded from above.

We derive the properties of the EL estimator of β_0 in the next theorem.

Theorem 1

Under assumption 1-2:

$$\sqrt{n}\left(\tilde{\beta}-\beta_0\right)\xrightarrow{d} N\left(0, V_\beta^1\right)$$

Where:

$$V_\beta^1 = \left\{E\left[\frac{\partial g}{\partial\beta}+\frac{\partial g}{\partial\tilde{\phi}}\frac{\partial\tilde{\phi}}{\partial\beta}\right]'\tilde{\Omega}_{11}^{-1}E\left[\frac{\partial g}{\partial\beta}+\frac{\partial g}{\partial\tilde{\phi}}\frac{\partial\tilde{\phi}}{\partial\beta}\right]\right\}^{-1},$$

$$\tilde{\Omega}_{11} = E\left[g\left(\tilde{\phi},\beta\right)g\left(\tilde{\phi},\beta\right)'\right]$$

Proof

The proof is similar to (Qin and Lawless, 1994). Differentiate R(β) with respect to β and λ respectively gives Equation (7 and 8):

$$\frac{\partial\Re(\beta)}{\partial\beta} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1+\lambda' g_i\left(\beta,\tilde{\phi}\right)}\left(\frac{\partial g_i\left(\beta,\tilde{\phi}\right)}{\partial\beta}+\frac{\partial g_i\left(\beta,\tilde{\phi}\right)}{\partial\tilde{\phi}}\frac{\partial\tilde{\phi}}{\partial\beta}\right)\lambda \quad (7)$$

$$\frac{\partial\Re(\beta)}{\partial\lambda} = \frac{1}{n}\sum_{i=1}^{n}\frac{g_i\left(\beta,\tilde{\phi}\right)}{1+\lambda' g_i\left(\beta,\tilde{\phi}\right)} \quad (8)$$

Denote the right hand side of (7) and (8) as $Q_{1n}(\beta,\lambda)$ and $Q_{2n}(\beta, \lambda)$ respectively. Since $\hat{\beta}$ and $\hat{\lambda}$ maximize $\Re(\beta), Q_{2n}\left(\hat{\beta},\hat{\lambda}\right) = 0$ and first order Taylor expansion around $(\beta_0, 0)$ gives:

$$0 = Q_{1n}\left(\hat{\beta},\hat{\lambda}\right)$$

$$= Q_{1n}\left(\beta_0,0\right)+\frac{\partial Q_{1n}\left(\beta_0,0\right)}{\partial\beta}\left(\tilde{\beta}-\beta_0\right)+\frac{\partial Q_{1n}\left(\beta_0,0\right)}{\partial\lambda}\hat{\lambda}+o_p(\delta)$$

$$0 = Q_{2n}\left(\hat{\beta},\hat{\lambda}\right)$$

$$= Q_{2n}\left(\beta_0,0\right)+\frac{\partial Q_{2n}\left(\beta_0,0\right)}{\partial\beta}\left(\tilde{\beta}-\beta_0\right)+\frac{\partial Q_{2n}\left(\beta_0,0\right)}{\partial\lambda}\lambda+o_p(\delta)$$

where, $\delta = \left\|\tilde{\beta}-\beta_0\right\|+\left\|\hat{\lambda}\right\|$ So $\tilde{\beta}$ and $\hat{\lambda}$ can be solved as:

$$\begin{bmatrix}\hat{\lambda}\\ \tilde{\beta}-\beta_0\end{bmatrix}=S_n^{-1}\begin{bmatrix}-Q_{1n}(\beta_0,0)+o_p(\delta)\\ op(\delta)\end{bmatrix}$$

$$\begin{bmatrix}\left(I-S_{11}^{-1}E\left(\frac{\partial g}{\partial\beta}+\frac{\partial g}{\partial\tilde{\phi}}\frac{\partial\tilde{\phi}}{\partial\beta}\right)S_{22.1}^{-1}E\left(\frac{\partial g}{\partial\beta}+\frac{\partial g}{\partial\tilde{\phi}}\frac{\partial\tilde{\phi}}{\partial\beta}\right)'\right)\\ S_{11}^{-1}Q_{1n}(\beta_0,0)+o_p(1)\\ E\left(\frac{\partial g}{\partial\beta}+\frac{\partial g}{\partial\tilde{\phi}}\frac{\partial\tilde{\phi}}{\partial\beta}\right)'S_{11}^{-1}Q_{1n}(\beta_0,0)+o_p(1)\end{bmatrix}$$

Where:

$$S_n=\begin{bmatrix}\frac{\partial Q_{1n}}{\partial u'} & \frac{\partial Q_{1n}}{\partial\beta}\\ \frac{\partial Q_{2n}}{\partial u'} & 0\end{bmatrix}_{(\beta_0,0)}\rightarrow\begin{bmatrix}S_{11} & S_{12}\\ S_{21} & 0\end{bmatrix}$$

$$=\begin{bmatrix}-E(gg') & E\left(\frac{\partial g}{\partial\beta}+\frac{\partial g}{\partial\tilde{\phi}}\le\beta\right)\\ E\left(\frac{\partial g}{\partial\beta}+\frac{\partial g}{\partial\tilde{\phi}}\le\beta\right) & 0\end{bmatrix}$$

From lemma 1 of (Qin and Lawless, 1994) under Assumption 2 we $\partial Q_{1n}(\beta_0,0)=(1/n)\sum_{i=1}^{n}g(x_i,\theta)=O_p(n^{-1/2})$ and $\delta=O_p(n^{-1/2})$. So we obtain:

$$\sqrt{n}(\tilde{\beta}-\beta_0)=S_{22.1}^{-1}E\left(\frac{\partial g}{\partial\beta}+\frac{\partial g}{\partial\tilde{\phi}}\frac{\partial\tilde{\phi}}{\partial\beta}\right)'$$
$$S_{11}^{-1}Q_{1n}(\beta_0,0)+o_p(1)\xrightarrow{d}N(0,V_\beta^1)$$

Remark 2

The structure of the asymptotic variance-covariance matrix V_β^1 is different from those in (Stock and Wright, 2000), in which they decompose $E[\hat{g}(\theta)]$ as $E[\hat{g}(\theta)]=1^{-1/2}m_1(\theta)+m_2(\beta)$ where $m_1(\theta)$ involves both of the two parameters and $m_2(\beta)$ involves β and the true value of ϕ.

Lazar and Mykland (1999) consider higher order properties of $\hat{\beta}$ through Edgeworth expansion of $R(\beta,\phi)$ They find that $\hat{\beta}$ may not achieve higher order accuracy which can be obtained by ordinary likelihood in the presence of nuisance parameters, also they show that the empirical likelihood ratio statistic does not admit Bartlett correction, unlike the case without nuisance parameters.

3. MORE MOMENT CONDITIONS

Now we focus on the asymptotic efficiency of $\hat{\beta}$ when there are more moment condition being added. Suppose based on model (1), we have the following new model by adding one more moment indicator f (.) Equation (9):

$$E[h(x_i;\beta_0,\phi_0)]=E\begin{bmatrix}g(x_i;\beta_0,\phi_0)\\ f(x_i;\beta_0,\phi_0)\end{bmatrix}=\begin{bmatrix}0\\0\end{bmatrix}\quad(9)$$

For more notations we define:

$$G\equiv E\left[\frac{\partial h(x;\theta_0)}{\partial\theta}\right],G_2\equiv E\left[\frac{\partial f(x;\theta_0)}{\partial\theta}\right]$$
$$\Omega\equiv E\left[h(\beta_0,\phi_0)h(\beta_0,\phi_0)'\right]=\begin{bmatrix}\Omega_{11} & \Omega_{12}\\ \Omega_{21} & \Omega_{22}\end{bmatrix}$$

In this model, following the setup in the previous section, the parameter vector $\theta=(\beta',\phi')'$ can be identified by (1) alone and now we are interested in whether the covariance matrix V_β^1 can be improved with extra information given by f. Let the estimator of β based on both g and f denoted as $\tilde{\beta}$ and the corresponding covariance matrix as V_β^2: In general, well established results have shown that at least using f will not be harmful, i.e., it will not increase the asymptotic variance of $\tilde{\theta}$. And, nor will dropping f will decrease the asymptotic variance of the estimator, relative to that of the estimator based on both g and f. See, corollary 1 of (Qin and Lawless, 1994).

Remark 3

A similar and relevant situation may be worth mention, which is described in (Newey and Windmeijer, 2005; Han and Philips, 2006), for instance. They assume that the number of moment conditions is increased with the sample size. Thus in this case extra information are provided by both extra data and extra moment conditions, while in our case only by the latter one with fixed sample size n. They also allow the moment conditions are weak, while we assume both g and f are strong as indicated in assumption 1. Estimation under many weak moment conditions is also discussed by (Andrews and Stock, 2005).

Proposition 1

The asymptotic efficiency of EL estimator of β can be increased by adding more moment conditions.

Proof

Since we can always block the component of the vector of the moment function, for simplicity and without loss of generality, we assume that both g and f are of dimensional one.

For convenience let:

$$E\left(\frac{\partial g}{\partial \beta}+\frac{\partial g}{\partial \tilde{\phi}}\frac{\partial \tilde{\phi}}{\partial \beta}\right)\equiv E_1, E\left(\frac{\partial f}{\partial \beta}+\frac{\partial f}{\partial \tilde{\phi}}\frac{\partial \tilde{\phi}}{\partial \beta}\right)\equiv E_2$$

The inverse of V_β^2, or the information matrix of β with both g and f is Equation (10):

$$\begin{aligned} I_\beta^2 &= E\left[\frac{\partial h}{\partial \beta}\right]'\left(E[hh']\right)^{-1}E\left[\frac{\partial h}{\partial \beta}\right] \\ &= \left[E_1\ E_2\right]\begin{bmatrix}\Omega_{11}\ \Omega_{12}\\ \Omega_{21}\ \Omega_{22}\end{bmatrix}^{-1}\begin{bmatrix}E_1\\ E_2\end{bmatrix}\end{aligned} \tag{10}$$

Since without f, the information on β is:

$$I_\beta^1 = E_1\left[E(gg')\right]^{-1}E_1 = E_1' S_{11}^{-1}E_1$$

We have:

$$\begin{aligned} I_\beta^2 - I_\beta^1 &= E\left[\frac{\partial h}{\partial \beta}\right]'\left(E[hh']\right)^{-1}E\left[\frac{\partial h}{\partial \beta}\right] - E_1'\left(E[gg']\right)^{-1}E_1 \\ &= E_1'\left[S_{11}^{-1}S_{12}\Omega_{22}S_{21}S_{11}^{-1}\right]E_1 + E_2\left[-\Omega_{22}S_{21}S_{11}^{-1}\right]E_1 \\ &+E_1\left[-S_{11}^{-1}S_{12}\Omega_{22}\right]E_2 + E_2\Omega_{22}E_2 \\ &= \left(E_1' S_{11}^{-1}S_{12} - E_1\right)\Omega_{22}\left(E_1' S_{11}^{-1}S_{12} - E_2\right)'\end{aligned}$$

Which is positive semidefinite, providing E(gg') is p.d as Assumption 2 indicates.

Example 1

Suppose we have a sequence of i.i.d observations of univariate random variable $x_1,\ldots,x_n$. Let E(x) = μ and var (x) = σ^2 Thus we have the following two moment conditions Equation (11 and 12):

$$E\left[g(x;\beta)\right] = E(x-u) = 0 \tag{11}$$

$$E\left[f(x;\beta,\phi)\right] = E\left((x-u)^2-\sigma^2\right) = 0 \tag{12}$$

And now we are only interested in the estimation of μ. The empirical likelihood estimator of μ is Equation (13):

$$\hat{\mu} = \arg\min_{\mu}\sum_{i=1}^{n}\log\left(1+t'\begin{pmatrix}x_i-\mu\\ (x_i-\mu)^2-\hat{\sigma}^2\end{pmatrix}\right)$$

And:

$$\begin{aligned} nVar(\hat{\mu}) &= \left(\left[\frac{\partial g}{\partial \beta}\frac{\partial f}{\partial \beta}\right]\begin{bmatrix}E(gg)E(gf)\\ E(fg)E(ff)\end{bmatrix}^{-1}\begin{bmatrix}\frac{\partial g}{\partial \beta}\\ \frac{\partial f}{\partial \beta}\end{bmatrix}\right)^{-1} = \Omega_{11}^{-1} \\ &= \sigma^2 - \frac{\left(E(x-\mu)^3\right)^2}{E\left((x-\mu)^2-\sigma^2\right)^2} \le \sigma^2\end{aligned} \tag{13}$$

Notice that without g_2, nVar(û) equals σ^2.

In the above example, we notice that $E\left(\frac{\partial f}{\partial \beta}\right)=0$ and this feature simplifies the calculation dramatically. So we consider the following more special model, where g does not have nuisance parameter, but f has a nuisance parameter only, although it brings some information from the data Equation (14):

$$E\left[h(x;\beta_0,\phi_0)\right] = E\begin{bmatrix}g(x;\beta_0)\\ f(x,\phi_0)\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix} \tag{14}$$

The gradient vector of h in (14) is:

$$\frac{\partial h}{\partial \theta} = \begin{bmatrix}\frac{\partial g}{\partial \beta} & 0\\ 0 & \frac{\partial f}{\partial \phi}\end{bmatrix}$$

The information on β is:

$$\begin{aligned} I_\beta^2 &= E\left[\frac{\partial h}{\partial \beta}\right]'\left[E(hh')\right]^{-1}E\left[\frac{\partial h}{\partial \beta}\right] \\ &= \left[E\left(\frac{\partial g}{\partial \beta}\right)0\right]\begin{bmatrix}\Omega_{11}\ \Omega_{12}\\ \Omega_{21}\ \Omega_{22}\end{bmatrix}^{-1}\begin{bmatrix}E\left(\frac{\partial f}{\partial \beta}\right)\\ 0\end{bmatrix} \\ &= E\left[\frac{\partial g}{\partial \beta}\right]'\left[\Omega_{11}^{-1}\left(I+\Omega_{12}\left(\Omega_{22}-\Omega_{12}\Omega_{11}^{-1}\Omega_{21}\right)^{-1}\Omega_{21}\Omega_{11}^{-1}\right)\right]E\left[\frac{\partial f}{\partial \beta}\right]\end{aligned}$$

where, I is the corresponding identity matrix. Now we have Equation (15 and 16):

$$I_{\beta}^{2} - I_{\beta}^{1} = E\left[\frac{\partial f}{\partial \beta}\right]' VE\left[\frac{\partial f}{\partial \beta}\right] \quad (15)$$

Where:

$$V = \left[\Omega_{11}^{-1}\Omega_{12}\left(\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{21}\right)^{-1}\Omega_{21}\Omega_{11}^{-1}\right] \quad (16)$$

By assumption E(gg') is positive semidefinite, so $\left(\Omega_{22} - \Omega_{12}\Omega_{11}^{-1}\Omega_{21}\right)^{-1}$ is also p.s.d and so is V. Thus we see that f provide extra information for β. However, if in (10), $E(gf) = \Omega_{12} = 0, V = 0, \text{so } I_{\beta}^{2} = I_{\beta}^{1}$. So we have the following proposition.

Proposition 2

Additional moment conditions which contains only nuisance parameters will provide extra information on the parameter of interest only if they are correlated to the original moment conditions.

Remark 4

Whether g and f are correlated is a testable condition. Since $E[g(x,\beta_0)] = E[f(x,\phi_0)] = 0$, to test the correlation of g and f it is equivalent to test the following additional moment condition Equation (17):

$$E\left[\rho\left(x;\beta_0,\phi_0\right)\right] = E\left[g\left(x,\beta_0\right)f\left(x,\phi_0\right)\right] = 0 \quad (17)$$

And this can be done by standard EL test procedure.

4. CONCLUSION

In this study we have discussed the efficiency of the EL estimator in the presence of nuisance parameters, via standard asymptotic method. We are particularly interested in whether the asymptotic efficiency of the parameter of interest can be improved by adding more moment conditions. We found that a necessary condition for augmented moment condition to be useful to improve the asymptotic efficiency is that it is correlated to the original moment condition. It is worth mentioning that here we incorporate more moment conditions with sample size being fixed, while researchers like (Newey and Windmeijer, 2005; Han and Philips, 2006) consider increasing the number of moment conditions brought by increasing sample size.

For future research, it would be worth checking the efficiency of the EL test with nuisance parameters, because it will be an extension to the results found in (Wang, 2013), where the large deviation efficiency of the EL test with weakly dependent data is established.

5. REFERENCES

Andrews, D. W. K. and J. H .Stock, 2005. Inference with Weak Instruments. NBER Research.

Han, C. and P.C.B. Philips, 2006. GMM with many moment conditions. Econometrica, 74: 147-192. DOI: 10.1111/j.1468-0262.2006.00652.x

Lazar, N.A. and P.A. Mykland, 1999. Empirical likelihood in the presence of nuisance parameters. Biometrika, 86: 203-211. doi: 10.1093/biomet/86.1.203

Newey, W.K. and F. Windmeijer, 2005. GMM with Many Weak Moment Conditions. Centre for Microdata Methods.

Neyman, J. and E.L. Scott, 1948. Consistent estimates based on partially consistent. Observ. Econ., 16: 1-32.

Owen, A.B., 2001. Empirical Likelihood. 1st Edn., CRC Press, London, ISBN-10: 1420036157, pp: 304.

Phillips, P.C.B., 1989. Partially identified econometric models. Econometr. Theory, 5: 181-240. DOI: 10.1017/S0266466600012408

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

Stock, J.H. and J.H. Wright, 2000. GMM with weak identification. Econometrica, 68: 1055.1096.

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