Semiparametric Estimators for Conditional Moment Restrictions Containing Nonparametric Functions: Comparison of GMM and Empirical Likelihood Procedures

[1]Nishiyama, Y. , [2]Q.F. Liu, and [3]N. Sueishi

[1]Institute of Economic Research, Kyoto University, [2]Graduate School of Economics, Kyoto University, [3]Graduate School of Economics, Kyoto University, E-Mail: nishiyama@kier.kyoto-u.ac.jp

Keywords: *Moment restriction; Semiparametric estimation; Bias; Efficiency; GMM; Empirical likelihood.*

EXTENDED ABSTRACT

Econometric models described by conditional moment restrictions cover a wide range of interesting cases in applied econometrics. They are characterized by $E[m(W,\theta_0)\,|\,Z]=0$ where m is a k dimensional known function, W is a $d_W \times 1$ vector of economic random variables, Z is a $d_Z \times 1$ and θ_0 is a $p \times 1$ vector of unknown parameters. Given a sample $\{W_i, Z_i\}_{i=1,\Lambda,n}$, the pioneering work by Hansen (1982) proposes the generalized method of moment (GMM) estimator which is defined by the solution to

$$\arg\min_{\theta} \sum_{i=1}^{n} m(W_i,\theta)' V^{-1} m(W_i,\theta)$$

where V is a positive definite matrix. It is shown that $V = Var[m(W,\theta_0)\,|\,Z]$ minimizes the variance of th estimator. It is no doubt one of the core tools for practitioners, but it is known that the original GMM estimator possibly has a large bias especially when the number of moment conditions is large. To circumvent this problem, a modified method has been developed called continuous updating estimator (CUE) by Hansen, Heaton and Yaron (1996). Principally different alternative methods are also proposed for the same problem. One is the exponential tilting estimator (ETE) which minimizes the nonparametrically constructed Kulback-Leibler information. The other is the maximum empirical likelihood (MEL) estimator proposed by Owen (1988) and Quin and Lawless (1994). There, the likelihood, constructed nonparametrically imposing the moment restrictions, is maximized with respect to the unknown parameters.

Ai and Chen (2003) consider a semiparametric moment condition described by

$$E[m(W,\theta_0,g_0)\,|\,Z]=0$$

where $g_0 : R^p \rightarrow R^q$ is a set of unknown (nuisance) functions. It is different from the model above in that it explicitly includes an unknown function g_0 in the moment restriction. They propose a sieve minimum distance estimator for θ_0, where they principally take the same approach as GMM replacing the unknown functions by nonparametric sieve estimates. They proved its consistency and asymptotic normality, and showed further that the estimator attains the semiparametric efficiency bound for this model. The purpose of this paper is twofold. Firstly, we briefly review estimation methods for moment restriction-based models. Inference for models described by conditional moment restrictions containing nonparametric functions is a relatively less developed topic in the context of moment condition based methods. To the best of our knowledge, Ai and Chen's (2003) procedure above is the only published result for this setup. But it is obvious that MEL type approach to this problem is straightforwardly applicable. We also compare the GMM by Ai and Chen with suitably adjusted MEL estimator by Monte Carlo simulation. Our interests are the followings. It is now widely recognized that GMM estimator tends to have larger bias than the MEL in small samples. We wonder if it holds for the current model. The nonparametric functions must be estimated to obtain the estimates of parametric components, so that we would like to check out how much it affects the properties of estimator of the parameter of interests. Theoretical comparison of statistical properties of these estimators is an interesting and important issue which we shall study later.

1. INTRODUCTION

Econometric models described by conditional moment restrictions cover a wide range of interesting cases in applied econometrics, such as standard regression models, simultaneous equation models, regression models with error-in-variables, and ARCH type models. Inference based on these restrictions is one of the most widely used methods and thus statistical properties of the estimators and tests for this class of models have been intensively studied in econometric theory. Pioneering work by Hansen (1982) proposes the celebrated generalized method of moment (GMM) estimator, and tests for parameters and the applied moment restrictions. It is no doubt one of the core tools for practitioners, but it is known that the original GMM estimator possibly has a large bias especially when the number of moment conditions is large. To circumvent this problem, a modified method has been developed called continuous updating estimator (CUE) by Hansen, Heaton and Yaron (1996). Principally different alternative methods are also proposed for the same problem. One is the exponential tilting estimator (ETE) which minimizes the nonparametrically constructed Kulback-Leibler information. The other is the maximum empirical likelihood (MEL) estimator proposed by Owen (1988) and Quin and Lawless (1994). There, the likelihood, constructed nonparametrically imposing the moment restrictions, is maximized with respect to the unknown parameters.

The purpose of this paper is twofold. Firstly, we briefly review estimation methods for moment restriction-based models. Inference for models described by conditional moment restrictions containing nonparametric functions is a relatively less developed topic in the context of moment condition based methods. To the best of our knowledge, GMM type estimation by Ai and Chen (2003) is the only published result for this setup. It is obvious that MEL type approach to this problem is straightforwardly applicable. Thus, secondly, we compare the GMM by Ai and Chen with suitably adjusted MEL estimator by Monte Carlo simulation. Our interests are the followings. It is now widely recognized that GMM estimator tends to have larger bias than the MEL in small samples. We wonder if it holds for the current model. Also, as seen later, the nonparametric functions need to be estimated to obtain the estimates of parametric components, so that we would like to check out how much it affects the properties of estimator of the parameter of interests. Theoretical comparison of statistical properties of these estimators is an interesting and important issue which we shall study later.

In the following section, we quickly state some econometric models which are described by conditional or unconditional moment restrictions. Section 3 provides estimators for these models such as GMM estimator, CUE, ETE and MEL estimator. We explain econometric models expressed by conditional moment restriction containing unknown functions in Section 4 and introduce a GMM type estimator for this model. We also state an alternative MEL type estimator for this model. Section 5 compares these estimators in bias and variance by Monte Carlo experiments. Section 6 concludes.

2. MOMENT RESTRICTION BASED ECONOMETRIC MODELS

A number of econometric models can be written in terms of moment restrictions such as

$$E[m(W,\theta_0)]=0 \qquad (1)$$

or

$$E[m(W,\theta_0)\,|\,Z]=0 \qquad (2)$$

where m is a k dimensional known function, W is a $d_W\times 1$ vector of economic random variables, Z is a $d_Z\times 1$ and θ_0 is a $p\times 1$ vector of unknown parameters. (2) is especially an important class of model which includes many interesting special cases. A simple case is when

$$m(W,\theta_0)=y-x'\theta_0$$

where $W=(y,x)$, y is a scalar dependent variable, x is $d_X\times 1$ independent variables and Z is the corresponding instruments. This model includes such cases as one equation model of a system of simultaneous equation models, regression model when the regressors has measurement errors. In this case, we can apply the standard instrumental variable estimation or two stage least squares method to consistently estimate θ_0 .

In time series setup, ARCH type models have been widely studied mainly to analyze financial time series data. ARCH model originally introduced by Engle (1982) is

$$y_t=\varepsilon_t h_t^{1/2},$$
$$h_t=\alpha_0+\alpha_1 y_{t-1}^2.$$

Letting $W=(y_t, y_{t-1})$, $Z=y_{t-1}$ and $\theta_0=(\alpha_0,\alpha_1)$, this model can be regarded as a special case of (2) with

$$m(W,\theta_0)=y_t^2-(\alpha_0+\alpha_1 y_{t-1}^2).$$

A variety of extension of ARCH model is proposed to better explain the behavior of financial time series, and some of them can be rewritten in the form of (2) similarly.

3. GMM, MEL AND RELATED ESTIMATORS

Suppose we would like to estimate θ_0 of model (1) given a sample $\{W_i\}_{i=1,\Lambda,n}$. GMM estimator is the solution to

$$\tilde{\theta}_V=\arg\min_\theta \sum_{i=1}^n m(W_i,\theta)'V^{-1}m(W_i,\theta) \quad (3)$$

where V is a symmetric positive definite weight matrix. Hansen (1982) showed $\tilde{\theta}_V$ is $\sqrt{n}-$ consistent for θ_0 and asymptotically normally distributed. Its asymptotic variance is minimized when $V=V^*(\theta)\equiv Var(m(W,\theta))$. Since $V^*(\theta)$ is unknown, a natural two step procedure is proposed. In the first step, we set $V=I$ and obtain an initial estimate $\tilde{\theta}_I$. In the next step, plugging $\hat{V}=\frac{1}{n}\sum_{i=1}^n m(W_i,\tilde{\theta}_I)m(W_i,\tilde{\theta}_I)'$ in (3), we obtain a feasible efficient estimate

$$\tilde{\theta}^*=\arg\min_\theta \sum_{i=1}^n m(W_i,\theta)'\hat{V}^{-1}m(W_i,\theta).$$

However, it is well known that this estimator may have large bias especially when we have many moment conditions. To circumvent this problem, an alternative estimator

$$\tilde{\theta}_{CUE}=\arg\min_\theta \sum_{i=1}^n m(W_i,\theta)'\hat{V}^*(\theta)^{-1}m(W_i,\theta)$$

is proposed by Hansen, Heaton and Yaron (1996), where $\hat{V}^*(\theta)$ is an estimate of $V^*(\theta)$ given θ. This is called the continuous updating estimator (CUE). Owen (1988), Qin and Lawless (1994) and Imbens (1997) propose a different approach to this moment condition based estimation. Letting

$$R(\theta)=\max_{p_1,\Lambda\, p_n}\left\{\prod_{i=1}^n np_i \mid \sum_{i=1}^n p_i m(W_i,\theta)=0,\ p_i\geq 0, \sum_{i=1}^n p_i=1\right\}$$

the maximum empirical likelihood estimator (MEL) is defined by

$$\hat{\theta}_{EL}=\arg\max_\theta R(\theta).$$

It is regarded as a nonparametric MLE imposing semiparametric constraints. Another alternative called exponential tilting estimator (ETE) is proposed by Kitamura and Stutzer (1997) and Imbens, Spady and Johnson (1998). The idea is to minimize the Kullback-Leibler information criterion instead of maximizing the nonparametric likelihood in the MELE, or

$$Q(\theta)=\min_{p_1,\Lambda,p_n}\left\{\prod_{i=1}^n p_i \ln p_i \mid \sum_{i=1}^n p_i m(W_i,\theta)=0,\ p_i\geq 0, \sum_{i=1}^n p_i=1\right\}$$

and estimate the parameter by

$$\hat{\theta}_{ET}=\arg\min_\theta Q(\theta).$$

Smith (1997) and Newey and Smith (2005) show that these estimators can be written in a unifying manner in terms of “generalized empirical likelihood estimator” (GEL).

There are a number of research which compare these alternative estimators both empirically and theoretically.

4. CONDITIONAL MOMENT RESTRICTION CONTAINING UNKNOWN FUNCTIONS

The models (1) and (2) and their estimators in the previous section have been intensively studied. There are cases where the moment restriction may depend on unknown functions, when we assume the model of the following form.

$$E[m(W,\theta_0,g_0)\mid Z]=0 \quad (4)$$

where $g_0: R^p\rightarrow R^q$ is a set of unknown (nuisance) functions. A typical example of this model is a partly linear regression with an endogenous nonparametric component, or

$$E[Y_1 - \theta_0' X - g_0(Y_2) \mid Z] = 0,$$

where Y_1, Y_2 are endogenous variables, X is a vector of exogenous variables and Z is the instruments. A similar model of partly linear regression model is considered by Carroll (1982) and Robinson (1987), but their model is a regression model of the form

$$E(Y_1 \mid X_1, X_2) = \theta_0' X_1 + g_0(X_2)$$

and thus it does no allow endogenous variables on the right hand side.

Obviously another model of interest of this class includes index models with endogenous variables,

$$E[Y_1 - g_0(\theta_{01}' X, \theta_{02}' Y_2) \mid Z] = 0.$$

To estimate the model formulated by (4), Ai and Chen (2003, AC hereafter) propose a sieve minimum distance (SMD) estimator interpreted as a natural extension of the GMM estimator. Putting $\alpha_0 = (\theta_0, g_0)$ and

$$f(z, \alpha_0) = \int m(w, \alpha_0) dF_{W|Z}(w \mid z)$$

where $F_{W|Z}(w \mid z)$ is the distribution of W conditional on Z, we can rewrite (4) into

$$f(z, \alpha_0) = 0.$$

Thus α_0 should satisfy

$$\alpha_0 = \arg\min_{\alpha} E[f(Z, \alpha)' \Omega(Z)^{-1} f(Z, \alpha)]$$

for any positive definite matrix $\Omega(Z)$. Though f is unknown, for a given α, we can estimate it nonparametrically. Then a possible estimator is

$$\hat{\alpha} = \arg\min_{\alpha} \frac{1}{n} \sum_{i=1}^{n} \hat{f}(Z_i, \alpha)' \hat{\Omega}(Z_i)^{-1} \hat{f}(Z_i, \alpha)$$

where $\hat{f}$ and $\hat{\Omega}$ are estimates of f and Ω respectively. They propose to use a sieve estimator for $\hat{f}$, and also for estimator (or approximate) of g_0, which yields a feasible estimator for α_0. Rewriting the estimator as in AC, we immediately know that it corresponds to the unconditional moment conditions

$$E[h_i(Z) m(W, \alpha_0)] = 0, \ i = 1, \Lambda, k_n \qquad (5)$$

where $\{h_i(z)\}_{i=1,\Lambda k_n}$ is the sieve basis. In view of (5), it is easy to apply MEL procedure instead of GMM, where we also use sieve estimate for g_0 to make it feasible. The next section compare these two estimators by simulation. There is not yet any justification for the MEL method, it would be likely that these two estimators have similar first order properties (see e.g. Newey and Smith (2004)).

5. MONTE CARLO COMPARISON OF THE ESTIMATORS

We implement a Monte Carlo experiment to investigate the following two points. Firstly, we would like to check out if GMM procedure by AC also has the bias problem when the number of instruments increases. In standard GMM settings, many authors have pointed out the bias of GMM estimator increases as the number of moment conditions grows. But this result is observed when new (and different) instrumental variables are added. Here, we consider the case when different functions of the same instrumental variables are included. Secondly, we would like to see if the SMDand MEL are mostly comparable. Lastly, we also investigate the influence of sieve estimate of g_0 on the estimator of parametric components.

5.1. Monte Carlo settings

We compare the two estimators by AC and its EL counterpart by a Monte Carlo study. We employ the same model as AC as the DGP. The original DGP used there was

$$\text{DGP1:} \quad \begin{cases} Y_1 = X_1 \theta_0 + g_0(Y_2) + u \\ Y_2 = X_1 + X_2 + \sigma u + \varepsilon, \end{cases}$$

or a partially linear simultaneous equation model, where

$$\theta_0 = 1,$$

$$g_0(Y_2) = \frac{\exp(Y_2)}{1 + \exp(Y_2)}$$

$$\sigma = 0.9.$$

$\varepsilon \sim N(0,1)$, $X_1, X_2 \sim U(0,1)$ and $u \sim N(0, X_1^2 + X_2^2)$. ε, X_1, X_2 are mutually independent. The parameter of interest is θ_0 and

g_0 is an infinite dimensional nuisance parameter. They generated an iid sample of size 1,000 and computed the bias, variance, quantiles of the estimator. The replication is 1000. As the approximation of g_0 function, they took

$$g_0(y) \approx \Phi(y)(\beta_0 + \beta_1 y + \beta_2 y^2 + \beta_3 y^3)$$

where Φ is the standard normal distribution function.

To see the asymptotic performance, we take n=100, 300, 1,000.

Our main concerns are:

(i) Does the number of sieve basis affects the bias of the SMD estimator?

(ii) Does MEL have a similar properties as SMD? Especially, is there any difference in bias properties?

(iii) How large is the efficiency loss due to the estimation of unknown function?

5.2. Results

It is well known that GMM estimator tends to have larger bias when the number of moment conditions is large. In view of (5), the number of moment conditions is determined by how many sieve basis is used to form the moment conditions. We first investigate this effect. Table 1 shows properties of the estimator for different number of moment conditions. SMD1 uses $\{1, X_1, X_2\}$ as the instruments, while SMD2 uses $\{1, X_1, X_2, X_1^2, X_1X_2, X_2^2\}$. Similarly, SMD$i$ (i=1,2,…,6) uses power functions of $\{X_1, X_2\}$ up to i-th degree. The bias clearly increases with the number of moment conditions. We may point out that the variance of SMD1 seems excessively small, but we do not know why it happens at present. Otherwise the variance tends to become smaller as the number of instruments increases. If we exclude SMD1, SMD4 seems to be the best choice in this estimation. But in the following, we mostly use SMD3 as the reference following AC.

Estimator	Bias	SE	RMSE
SMD1	0.0043	0.1359	0.1360
SMD2	-0.0092	0.2907	0.2909
SMD3	-0.0345	0.1433	0.1473
SMD4	-0.0664	0.1295	0.1456
SMD5	-0.0928	0.1220	0.1533
SMD6	-0.1244	0.1179	0.1715

Table 1. Bias, standard error and RMSE of SMD estimator for increased number of sieve basis, n=1000.

Table 2 compares the properties of inefficient SMD (or when the weight matrix is set to be an identity matrix) with the two kinds of efficient estimators which differ only in the initial values. “Efficient SMD3 (AC)” indicates SMD3 with initial value determined as in AC, namely we use the inefficient SMD3 as the initial value. “Efficient SMD3 (Averaged initial value)” means that we give the fixed initial value computed from averaging the inefficient SMD3 estimates over 1000 replications. We found, surprisingly, that RMSE is smaller for inefficient SMD3 than efficient SMD3 by AC for not only smaller sample of n=100, but also n=1000. But efficient SMD3 computed from with the fixed initial value mostly outperforms the inefficient SMD3. This indicate that the initial value choice may be quite crucial in this SMD method.

Estimator	n	Bias	SE	RMSE
Inefficient SMD3	100	-0.2678	0.3700	0.4565
	300	-0.0884	0.2528	0.2676
	1000	-0.0345	0.1433	0.1473
Efficient SMD3 (AC)	100	-0.2591	0.4550	0.5231
	300	-0.1213	0.2653	0.2914
	1000	-0.0317	0.1145	0.1187
Efficient SMD3 (Averaged initial value)	100	-0.1470	0.2200	0.2640
	300	-0.1065	0.1913	0.2188
	1000	0.0010	0.1877	0.1877

Table 2. Performance of SMD and MEL under DGP1 for different choice of initial value and the sample size.

As the alternative of SMD3, MEL is implemented. Table 3 compares SMD3 and MEL3, where MEL3 indicates MEL estimator with the same moment conditions as SMD3. MEL3 seems to have smaller bias than SMD as we expected, but the standard error of MEL3 is larger by some unknown reasons.

Estimator	bias	SE	RMSE
SMD3	-0.0345	0.1433	0.1473
MEL3	0.0010	0.1877	0.1877

Table 3. Comparison of bias, standard error and RMSE of SMD and MEL under DGP1, n=1000.

Lastly, we investigate how much efficiency loss exists due to the estimation of unknown functions. We simply estimate the parameter θ_0 using the true "unknown function". The results are in Table 4. Since the SMD estimator is shown to attain the semiparametric efficiency bound of the corresponding model, we expected that the efficiency loss which comes from the nonparametric component estimation would reduce as the sample size increases, but we did not obtain such a result. Comparing Table 3 and 4, the efficiency ratio for inefficient SMD3 is around 1.64-2.42 and we do not observe efficiency improvement for larger sample size. For inefficient SMD3(AC), the ratio is quite stable, around 2.55-2.77 for all three sample sizes. Though in order for the asymptotic theory of AC to apply, the number of instruments must satisfy $k_n = O(n^{1/7})$ and thus, this comparison we made may be inadequate, but we at least know that the efficiency loss due to the nonparametric estimation can be quite large depending on k_n, and in terms of variance, it may be about 6-7 times as large as the known function case.

Estimator	n	Bias	SE	RMSE
Inefficient SMD3	100	0.0014	0.1603	0.1607
	300	-0.0004	0.0966	0.0966
	1000	-0.0039	0.0576	0.0577
Efficient SMD3 (AC)	100	0.0115	0.1603	0.1607
	300	0.0077	0.0899	0.0902
	1000	-0.0033	0.0545	0.0546
MEL3	100	0.0027	0.1662	0.1663
	300	-0.0026	0.0850	0.0850
	1000	-0.0049	0.0454	0.0457

Table 4. Bias, standard error and RMSE when g_0 is known under DGP1.

We lastly give other findings in the following.

Continuous updating estimator is less biased compared with standard two-step GMM estimator, but it is often reported from Monte Carlo studies that it has a larger dispersion. Also standard optimization algorithms do not always seem to work well so that the estimate may not converge. We also experienced these phenomena. For numerical optimization, we used fminsearch of Matlab, which uses simplex search method, and fmincon for constrained optimization. We refer to Guggenberger (2005) for possible no moment problem of CUE, and Kunitomo and Matsushita (2003) for the same problem of MEL.

6. CONCLUSIONS

The purpose of this paper is to investigate the statistical properties of estimators of parametric components of the model described by conditional moment restrictions containing unknown functions. Two estimators, SMDand MEL are compared. We found that (1) MEL seems to have a smaller bias thanSMD, (2) the optimization seems to be quite sensitive to the initial value and thus one may need to be cautious in practical applications, (3) estimation of nonparametric component may give a significant efficiency loss to the estimator.

Research on the estimation of the model considered in this paper is still in its early stage. Further research on asymptotic theory for MEL procedure is necessary and also how to estimate or approximate the unknown functions seems to be critically important in practice. Also computational problem should be resolved.

7. REFERENCES

Ai, C. and X. Chen (2003), Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions, *Econometrica*, 71: 1795-1843.

Carroll, R. J. (1982) Adapting for Heteroscedasticity in Linear Models, Annals of Statistics, 9: 194-200.

Florens, C., E. Jondeau and H. Le Bihan (2001), Assessing GMM Estimators of the Federal Reserve Reaction Function, mimeo.

Guggenberger, P. (2005), Monte Carlo Evidence Suggesting a No Moment Problem of the Continuous Updating Estimator, *Economics Bulletin*, forthcoming.

Hansen, L. P., J. Heaton and A. Yaron (1996), Finite-Sample Properties of Some Alternative GMM Estimators, *Journal of Business and Economic Statistics*, 14: 262-280.

Imbens, G. W., R. H. Spady and P. Johnson (1998), Information Theoretic Approaches to Inference in Moment Condition Models, *Econometrica* 66: 333-357.

Kitamura, Y. and M. Stutzer, (1997), An Information Theoretic Alternative to

Generalized Method of Moments Estimation, *Econometrica*, 65: 861-874.

Kunitomo, N. and Y. Matsushita, (2003), Finite Sample Distributions of the Empirical Likelihood Estimator and GMM Estimator, Discussion paper CIRJE-F-200, Graduate School of Economics, University of Tokyo.

Robinson, P. M. (1987) Asymptotically Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form, Econometrica, 56: 875-891.

Smith R. J. (1997), Alternative Semi-Parametric Likelihood Approach to Generalised Method of Moments Estimation, *The Economic Journal*, 107: 503-519.

Stock, J. H., J. H. Wright and M. Yogo (2002),A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments, *Journal of Business and Economic Statistics* 20: 518 - 529.