E C O N O M I C S   B U L L E T I N

Monte−carlo evidence suggesting a no moment problem of the continuous updating estimator

Patrik Guggenberger
UCLA, Department of Economics

Abstract

Monte Carlo evidence is provided that suggests that the continuous updating estimator might have a moment problem. In the linear simultaneous equation model, its performance in terms of sample median and standard deviation is virtually identical to the one of the limited information maximum likelihood estimator.

Acknowledgments: I would like to thank Jinyong Hahn and Ron Mittelhammer for very helpful discussion and suggestions.
Citation: Guggenberger, Patrik, (2005) "Monte−carlo evidence suggesting a no moment problem of the continuous updating estimator." *Economics Bulletin*, Vol. 3, No. 13 pp. 1−6
Submitted: December 22, 2004. **Accepted:** March 10, 2005.
URL: http://www.economicsbulletin.com/2005/volume3/EB−04C10027A.pdf

1 Introduction

Recently, Newey and Smith (2004) have derived some desirable theoretical properties about the higher order bias of the generalized empirical likelihood (GEL) estimator. The class of GEL estimators includes as particular examples the continuous updating estimator (CUE) of Hansen, Heaton, and Yaron (1996) and the empirical likelihood (EL) estimator, see e.g. Imbens (1997). However, very little is known about the finite sample properties of these estimators. While Mariano and Sawa (1972), Fuller (1977), and Kinal (1980) show that limited information maximum likelihood (LIML) does not have any moments, two-stage least squares (2SLS) has as many moments as there are over-identifying restrictions, and Fuller's estimator has at least two moments finite, it seems to be extremely difficult to derive the corresponding results for GEL estimators in the general case. GEL estimators are not given in closed form which makes theoretical analysis very complicated.[1]

The definition of GEL estimators as the solution of a (potentially high dimensional) saddle point problem ($\arg\min - \sup$ problem) also makes finite sample investigation of GEL estimators via Monte Carlo study extremely difficult and, consequently, very few such studies can be found in the literature. Some recent exceptions include Mittelhammer, Judge, and Schoenberg (2002) and Guggenberger and Hahn (2004). Both studies are done for the linear instrumental variables (IV) model. While the former paper calculates GEL estimators directly using a combination of minimization routines including Nelder-Meade, the latter paper takes a shortcut by investigating the finite sample properties of EL indirectly by studying the finite sample properties of its two-step version that has a closed form solution and the same higher order properties. The latter paper finds strong evidence for a moment problem of the EL estimator. The former paper considers only very strongly identified scenarios (with $R^2 > .25$) and while evidence for higher variance of GEL estimators compared to 2SLS is reported, based on the strongly identified models in their Monte Carlo setup, no "no moment problem" of the GEL estimators, in particular of the CUE, is suspected. If R^2 is very small, the criterion function for the GEL estimator is typically very flat and any of the standard minimization routines is known to run into problems. Either the routine does not converge or if it does, suspicion remains whether the true minimizer has been found.

The goal of this note is to shed light on the finite sample properties of one particular GEL estimator, namely the CUE, through a comprehensive Monte Carlo study in a linear IV model with structural parameter vector of dimension one. The CUE can be computed much more easily than other GEL estimators since the $\arg\min - \sup$ saddle point problem simplifies to a $\arg\min$ problem in this case. Instead of relying on a minimization routine as done in Mittelhammer, Judge, and Schoenberg (2002), I evaluate the criterion function of the CUE for all values in the parameter space using a very fine grid.[2] This approach guarantees numerical stability and works successfully even in very weakly identified models. Corroborating the findings in Guggenberger and Hahn (2004), the Monte Carlo evidence presented here

[1] In a linear instrumental variables model, it is easily shown that the CUE coincides with LIML and 2SLS in the just-identified case (under weak conditions that guarantee that the weighting matrix in the CUE criterion function is positive definite) and therefore does not have moments in this case. However, even for the simplest over-identified case, where the parameter of interest is of dimension one and there are two instruments, a theoretical investigation into the moment question becomes very hard.

[2] This has also been done in some Monte Carlo scenarios in Hansen, Heaton, and Yaron (1996) and Imbens, Spady, and Johnson (1998).

suggests that the CUE may have a moment problem. The sample median and standard deviation (Std) properties of CUE are virtually identical to the ones of LIML which is known to not have any moments finite. While (in the over-identified case) these two estimators typically have better median bias properties than 2SLS and Fuller's estimator, their Std can exceed the ones of 2SLS and Fuller, oftentimes by several magnitudes. Donald and Newey (2000) provide theoretical intuition as to why the bias of CUE should be relatively small. They show that the CUE has an interpretation as a jackknife estimator. Consequently, as they point out, confidence intervals based on the CUE, as recently suggested by Guggenberger and Smith (2002), should be well centered.

2 Monte Carlo Experiment

2.1 Model and Experimental Design

I consider the properties of the CUE applied to a simple linear IV model as in Hahn and Hausman (2002)

$$\begin{aligned} y_i &= x_i\beta_0 + \varepsilon_i, \\ x_i &= z_i'\pi + u_i \qquad i = 1, \ldots, n, \end{aligned} \tag{1}$$

where y_i, x_i, and z_i denote the dependent variable, an (endogenous) regressor, and the vector of instrumental variables. The value of β_0 is set equal to 0. Let z_i and $(\varepsilon_i, u_i)'$ be i.i.d. $\mathcal{N}(0, I_K)$ and $\mathcal{N}(0, \Omega)$, respectively, where I_K is the K-dimensional identity matrix and Ω is a 2×2-matrix with diagonal and off diagonal elements 1 and ρ, respectively. Let $\mathbb{R}^2 \equiv E\left[(\pi' z_i)^2\right] / (E\left[(\pi' z_i)^2\right] + E\left[v_i^2\right]) = \pi'\pi/(\pi'\pi + 1)$ denote the theoretical R^2 of the first stage regression. Assume $\pi = (\eta, \eta, \ldots, \eta)'$ and thus

$$\mathbb{R}^2 = \frac{K \cdot \eta^2}{K \cdot \eta^2 + 1}. \tag{2}$$

The sample size n is set equal to 100. I simulate data for all the possible 45 parameter combinations of

$$\begin{aligned} K &= (1, 5, 20), \\ \mathbb{R}^2 &= (.005, .01, .02, .05, .2), \\ \rho &= (0, .3, .5). \end{aligned}$$

Different choices of K allow us to investigate how the degree of over-identification affects the performance of estimators. The parameter ρ determines the degree of endogeneity that ranges from none ($\rho = 0$) to strong ($\rho = .5$). Finally, equation (2) implies $\eta = \left(\mathbb{R}^2 / \left(K\left(1 - \mathbb{R}^2\right)\right)\right)^{1/2}$, and therefore $\mathbb{R}^2$ (together with K) pins down the strength of the instruments.

The CUE exploits the moment condition $E\varphi_i(\beta_0) = 0$, where $\varphi_i(\beta) \equiv (y_i - x_i\beta)z_i$. It is defined as[3]

$$\begin{aligned} \widehat{\beta}_C &\equiv \arg\min_{\beta \in B} Q_n(\beta), \text{ where} \\ Q_n(\beta) &\equiv [\sum_{i=1}^n \varphi_i(\beta)]'[\sum_{i=1}^n \varphi_i(\beta)\varphi_i(\beta)']^{-1}[\sum_{i=1}^n \varphi_i(\beta)], \end{aligned}$$

[3] Newey and Smith (2004, p.222) show that this definition of the CUE $\widehat{\beta}_C$ coincides with the one given in Hansen, Heaton, and Yaron (1996), where in $Q_n(\beta)$ the term $\sum_{i=1}^n \varphi_i(\beta)\varphi_i(\beta)'$ is replaced by $\sum_{i=1}^n \varphi_i(\beta)\varphi_i(\beta)' - \sum_{i=1}^n \varphi_i(\beta) \sum_{i=1}^n \varphi_i(\beta)'$.

for a subset B of the real numbers. I consider three choices for B, namely $[-c, c]$ for $c = 1, 5$, and 25. Instead of relying on a minimization routine, $\widehat{\beta}_C$ is calculated by evaluating $Q_n(\beta)$ for all $\beta \in B$ on a grid with stepsize .01 and I pick $\widehat{\beta}_C$ as the minimizer of $Q_n(\beta)$ on this grid. If the criterion function $Q_n(\beta)$ is very flat and/or has multiple local minima, minimization routines can be unstable and may lead to wrong results. The heavy computational burden and unreliability of minimization routines for GEL estimators has been widely recognized, see Hansen, Heaton, and Yaron (1996) and Imbens, Spady, and Johnson (1998) who also suggest the grid search method for scenarios where numerical optimization routines fail to converge. In my experience, especially in scenarios where instruments are only weakly correlated with the endogenous regressor, the saddle point problem of the GEL estimator becomes particularly intractable. The grid search approach used here however, finds the correct minimizer within the specified compact set.[4] I do not, of course, recommend such a time consuming algorithm in practice. It is implemented here, because with only one-dimensional β it can be done in reasonable time.

For each parameter combination 5,000 simulation repetitions are used and each time I calculate the CUE and, for comparison, the LIML, 2SLS, and Fuller's[5] estimator for each choice of B. I restrict the latter estimators to B as well, that is, if, for example, the LIML estimator is bigger (smaller) than c $(-c)$ for a given data sample, it is set equal to c $(-c)$. Results regarding the sample mean, median, and Std of the estimators over the 5,000 samples are reported. To gain insight into how binding the restriction onto the interval $[-c, c]$ is, I also report the percentage of times that the estimators fall into the "10%-tail" $[-c, -.9c] \cup [.9c, c]$ of the interval $[-c, c]$.

2.2 Simulation Results

Table I below reports results on the subset of all 8 parameter combinations of $K = (5, 20)$, $\mathbb{R}^2 = (.005, .05)$, and $\rho = (0, .5)$ for the sample mean, median, Std, and the 10%-tail-probabilities for the estimators CUE, LIML, 2SLS, and Fuller when $B = [-c, c]$ for $c = 1$, 5, and 25 for sample size $n = 100$ using 5000 simulation repetitions. First, the two main findings from Table I are discussed and then some additional insight from the simulation results not reported here (but available upon request) is added.

First, while all estimators can have significant median bias when ρ is large, overall, this bias is smaller for CUE and LIML than for 2SLS and Fuller, e.g. for $K = 5$, $\mathbb{R}^2 = .05$, $\rho = .5$ the bias for 2SLS is four times the bias of CUE. One interesting feature of the median bias of the CUE is that it increases heavily when c increases from 1 to 5, while for all other estimators c does not seem to impact the median at all. While for $c = 1$ the median bias of CUE is typically smaller than the one of LIML, when $c \geq 5$ the biases of the two estimators are virtually the same.

Second, the Std of CUE and LIML is virtually identical across all parameter combinations and values of c and is considerably larger than the corresponding Std of 2SLS and Fuller, both of which have two moments finite in the over-identified cases discussed here. Also, while the Std of CUE and LIML dramatically increases with c, it remains virtually unchanged for 2SLS and Fuller when c increases. The large Std of CUE and its similarity to LIML suggest a no moment problem of the CUE. The "10% tail-probabilities" provide additional evidence for the "no moments" conjecture. Both CUE and LIML

[4] Sensitivity analysis with respect to the stepsize shows that there is virtually no change in the results by using even smaller stepsizes than .01.

[5] I pick $\alpha = 1$ in the definition of the Fuller estimator, see Fuller (1977, p.942).

take on large values with much higher probability then both 2SLS and Fuller. For example, in the case $K = 20$, $\mathbb{R}^2 = .005$, $\rho = 0$, about 50% of the observations of CUE and LIML have absolute value $> .9$ and 14% of the observations have absolute value > 4.5. Still 3% of the observations exceed 22.5 in absolute value! The corresponding numbers for 2SLS are 0%, 0%, and 0% and for Fuller 34%, 0%, and 0%.

The Monte Carlo evidence on the parameter combinations not reported in Table I supports the findings from above. 1) Overall, the median bias of CUE and LIML is very similar and smaller than the one of 2SLS and Fuller. 2) Overall, the Std of CUE and LIML is very similar and considerably larger than the one of 2SLS and Fuller. In the just-identified cases, $K = 1$, CUE, LIML, and 2SLS coincide and, in this case, the latter also has very large sample Std while the one of Fuller remains small. For larger $\mathbb{R}^2$, the sample Std of CUE and LIML becomes smaller, e.g. when $K = 20$ the Std is about 2.5 times smaller when $\mathbb{R}^2 = .2$ compared to the case $\mathbb{R}^2 = .005$.

TABLE I

	Mean			Median			Std			10%-tail-prob		
c	1	5	25	1	5	25	1	5	25	1	5	25
					$K = 5$, $\mathbb{R}^2 = .005$, $\rho = 0$							
CUE	-.01	-.04	-.06	-.01	-.02	-.01	.77	2.22	5.56	.48	.13	.03
LIML	-.01	-.03	-.03	-.01	-.01	-.01	.78	2.21	5.49	.49	.12	.03
2SLS	-.01	-.01	-.01	.00	.00	.00	.48	.55	.55	.09	.00	.00
Fuller	.00	.00	.00	.00	.00	.00	.59	.64	.64	.17	.00	.00
					$K = 5$, $\mathbb{R}^2 = .005$, $\rho = 0.5$							
CUE	.08	.30	.34	.22	.39	.40	.77	2.11	4.93	.48	.11	.02
LIML	.23	.38	.47	.41	.41	.41	.75	2.09	5.02	.49	.11	.02
2SLS	.43	.47	.47	.46	.46	.46	.40	.47	.47	.15	.00	.00
Fuller	.40	.45	.45	.44	.44	.44	.49	.58	.58	.24	.00	.00
					$K = 20$, $\mathbb{R}^2 = .005$, $\rho = 0$							
CUE	-.01	-.02	.04	-.01	-.01	-.01	.74	2.31	5.65	.44	.14	.03
LIML	-.02	-.02	.00	-.03	-.03	-.03	.79	2.31	5.42	.51	.14	.03
2SLS	.00	.00	.00	.00	.00	.00	.23	.23	.23	.00	.00	.00
Fuller	-.02	-.02	-.02	-.03	-.03	-.03	.70	.90	.90	.34	.00	.00
					$K = 20$, $\mathbb{R}^2 = .005$, $\rho = .5$							
CUE	.10	.33	.47	.23	.42	.44	.74	2.14	5.08	.43	.12	.02
LIML	.24	.40	.47	.43	.43	.43	.75	2.14	4.93	.50	.12	.02
2SLS	.48	.48	.48	.49	.49	.49	.20	.20	.20	.02	.00	.00
Fuller	.35	.46	.46	.45	.45	.45	.61	.79	.79	.33	.00	.00
					$K = 5$, $\mathbb{R}^2 = .05$, $\rho = 0$							
CUE	-.01	.00	-.07	.00	.00	.00	.59	1.37	2.99	.22	.04	.01
LIML	-.01	.00	-.01	-.01	-.01	-.01	.59	1.31	2.77	.21	.04	.01
2SLS	.00	.00	.00	.00	.00	.00	.36	.38	.38	.02	.00	.00
Fuller	-.01	.00	.00	.00	.00	.00	.47	.50	.50	.08	.00	.00
					$K = 5$, $\mathbb{R}^2 = .05$, $\rho = .5$							
CUE	-.03	-.05	-.06	.03	.06	.06	.58	1.37	3.07	.22	.04	.01
LIML	.00	-.03	-.04	.05	.05	.05	.58	1.33	2.88	.21	.04	.01
2SLS	.22	.22	.22	.24	.24	.24	.33	.35	.35	.03	.00	.00
Fuller	.13	.14	.14	.13	.13	.13	.44	.47	.47	.07	.00	.00
					$K = 20$, $\mathbb{R}^2 = .05$, $\rho = 0$							
CUE	.00	-.03	-.10	.00	.00	.00	.68	1.87	4.29	.33	.08	.02
LIML	.01	.02	.04	.01	.01	.01	.68	1.76	4.00	.35	.07	.01
2SLS	.00	.00	.00	.00	.00	.00	.21	.21	.21	.00	.00	.00
Fuller	.01	.01	.01	.01	.01	.01	.61	.75	.75	.23	.00	.00
					$K = 20$, $\mathbb{R}^2 = .05$, $\rho = .5$							
CUE	.04	.11	.10	.11	.17	.18	.66	1.77	4.21	.31	.07	.02
LIML	.07	.08	.03	.13	.13	.13	.66	1.71	3.96	.32	.07	.02
2SLS	.39	.39	.39	.39	.39	.39	.19	.19	.19	.01	.00	.00
Fuller	.17	.21	.21	.18	.18	.18	.56	.68	.68	.19	.00	.00

References

Donald, S.G. and W.K. Newey (2000): "A jackknife interpretation of the continuous updating estimator", *Economics Letters* 67, 239-243.

Fuller, W.A. (1977): "Some Properties of a Modification of the Limited Information Estimator", *Econometrica* 45(4), 939-953.

Guggenberger, P. and J. Hahn (2004): "Finite Sample Properties of the 2-step Empirical Likelihood Estimator", *working paper*.

Guggenberger, P. and R.J. Smith (2002): "Generalized empirical likelihood estimators and tests under partial, weak and strong identification", forthcoming in *Econometric Theory*.

Hahn, J. and J. Hausman (2002): "A New Specification Test for the Validity of Instrumental Variables", *Econometrica* 70, 163-189.

Hansen, L.P., J.Heaton, and A.Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators", *Journal of Business and Economic Statistics* 14(3), 262-280.

Imbens, G. (1997): "One-Step Estimators for Over-Identified Generalized Method of Moments Models", *Review of Economic Studies* 64, 359 - 383.

Imbens, G., R.H. Spady and P. Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models", *Econometrica* 66(2), 333-357.

Kinal, T.W. (1980): "The existence of moments of k-class estimators", *Econometrica* 48(1), 241-249.

Mariano, R.S. and T. Sawa (1972): "The Exact Finite-Sample Distribution of the Limited Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables", *Journal of the American Statistical Association* 67, 159-163.

Mittelhammer, R.C., G.G. Judge, and R. Schoenberg (2002): "Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods", *Festschrift in Honor of Thomas Rothenberg,* In Press.

Newey, W. K. and R. J. Smith (2004): "Higher order properties of GMM and Generalized Empirical Likelihood estimators", *Econometrica* 72(1), 219-255.