Communications for Statistical Applications and Methods
2013, Vol. 20, No. 3, 199–205

DOI: http://dx.doi.org/10.5351/CSAM.2013.20.3.199

Size Refinement of Empirical Likelihood Tests in Time Series Models using Sieve Bootstraps

Jin Lee[1,a]

[a]Department of Economics, Ewha Womans University

Abstract

We employ sieve bootstraps for empirical likelihood tests in time series models because their null distributions are often vulnerable to the presence of serial dependence. We found a significant size refinement of the bootstrapped versions of a Lagrangian Multiplier type test statistic regardless of the bandwidth choice required by long-run variance estimations.

Keywords: Time series, empirical likelihood, size of the test, sieve bootstrap.

1. Introduction

Empirical likelihood(EL) methods (known as nonparametric likelihoods) are commonly used for robust hypothesis testing under weak identification. In time series models with dependent processes, valid testing procedures have been developed that include a generalized method of mo-ments(GMM)-based test statistic by Kleibergen (2005), Lagrangian Multiplier(LM) tests by Guggenberger and Smith(GS) (2008) and by Otsu (2006). Tests based on quadratic forms of first-order conditions in EL estimation can achieve pivotal properties because the asymptotic null distributions do not depend on the strength of identifications. They have standard Chi-squared limit distributions with degrees of freedom equal to the number of parameters instead of the number of moments; subsequently, the EL tests prove practically useful in many aspects.

EL tests typically generate size distortions when underlying processes carry serial dependence. Instrumental variables or innovation processes tend to be often serially correlated in time series models that consist of (for example) macroeconomic series. Subsequently, empirical null distributions of EL tests become sensitive to the choice of bandwidth, which is required for the estimation of long-run variance. Empirical null rejection probabilities of the tests significantly depart from the nominal level, which negatively affects accurate statistical inferences. This issue of size distortion is closely related with the well-known size problem of heteroskedasticity and autocorrelation consistent(HAC) covariance matrix estimation in time series literature (*cf*: Andrews, 1991; Newey and West, 1994). Therefore, we try to find a way to improve size performance of the EL tests.

In this work, we make use of the sieve bootstraps by Bühlman (1997) as an appropriate bootstrap method. In particular, the sieve method is a natural way to handle serial dependence in a time series. It approximates a dependent time series process through autoregressive models under the null hypothesis and generates bootstrapped data based on re-sampled residuals and fitted autoregressive models. It is expected that the bootstrapped EL tests can achieve a more accurate size performance than the

This work was supported by the National Research Foundation of Korean Grant funded by the Korean Government (NRF-2011-327-B00074).

[1] Associate Professor, Department of Economics, Ewha Womans University, Daehyun-dong, Seodaemun-gu, Seoul 120-750, Korea. E-mail: leejin@ewha.ac.kr

test statistics based on asymptotic inferences under the null hypothesis. We investigate how the sieve bootstraps work when combined to EL tests through Monte Carlo simulations. As seen in the simulation studies, bootstrap version of the EL test is shown to be less sensitive to the bandwidth choice than original EL tests. We also note that recently, sieve methods have been widely applied to reduce size distortions of test statistics in a time series context. Among others, we note Chang and Park (2004), Park (2002) and Palm *et al.* (2010) in the context of testing for unit roots and cointegration.

In Section 2, we summarize the setup of EL tests. In Section 3, sieve bootstrap algorithms are given. Simulation results are provided in the Section 4. Section 5 concludes.

2. Model

Consider R^m-valued stationary mixing process $\{z_t\}_{t=1}^T$. Let θ_0 be the true parameter for $\theta \in \Theta \subset R^p$, where p is the number of parameters, and $g : R^m \times \Theta \to R^q$ be a vector of unknown function. Moment conditions are typically written as

$$E\left[g_t(z_t, \theta_0)\right] = E\left[g_t(\theta_0)\right] = 0. \tag{2.1}$$

Therefore, the following hypothesis is of interest,

$$H_0 : \theta_0 = \theta^*. \tag{2.2}$$

In order to allow weak identifications, we decompose the parameter space as $\Theta = A \times B$, where $p = p_A + p_B$. Put $\theta = (\alpha', \beta')'$. As in Stock and Wright (2000), certain conditions are imposed such that α is weakly identified, while β is strongly identified (*cf*: GS, 2008).

Assumption 1. *(1) $E[T^{-1}\sum_{t=1}^T g_t(\theta)] = T^{-1/2}m_{1T}(\theta) + m_2(\beta)$, where $m_1 : \Theta \to R^q$, such that $m_{1T}(\theta) \to m_1(\theta)$ uniformly on Θ, and $m_1(\theta_0) = 0$, (2) $m_2(\beta) : B \to R^q$ such that $m_2(\beta) = 0$, if and only if $\beta = \beta_0$, (3) $m_2(\beta)$ is continuously differentiable at β_0 such that $\partial m_2(\beta)/\partial\beta'|_{\beta=\beta_0}$ has full column rank which is equal to p_β.*

The sample average of moment indicators are defined by

$$\hat{g}(\theta) = T^{-1}\sum_{t=1}^{T} g_t(\theta). \tag{2.3}$$

In time series context, it is natural to use smoothed moment restrictions using a kernel weighting (*e.g.*, Kitamura and Stutzer, 1997),

$$g_{tT}(\theta) = M^{-1}\sum_{j=t-T}^{t-1} k\left(\frac{j}{M}\right) g_{t-j}(\theta), \tag{2.4}$$

where M is the bandwidth, and $k(\cdot)$ is a kernel function. For instance, testing procedures by GS (2008) employ truncated kernel for $k(\cdot)$. Following conditions on a kernel function are standard.

Assumption 2. *(1) $k(x) : R \to [-1, 1]$ is symmetric and continuous at zero with $k(0) = 1$, $|k(x)| \le C|x|^{-b}$ as $x \to \infty$ for $b > 2$, (2) $K(\lambda) \ge 0$, for all $\lambda \in [-\pi, \pi]$, where $K(\lambda) = (2\pi)^{-1}\int_{-\infty}^{\infty} k(x)e^{-i\lambda x}dx$.*

Popular kernels such as Bartlett, Daniell, Parzen, quadratic spectral(QS) satisfy the conditions (1) and (2). The function $K(\lambda)$ is a Fourier transforms of $k(x)$. Besides, we impose a mild condition on bandwidths.

Assumption 3. *(1) $M \to \infty$ as $T \to \infty$, (2) $M/T \to 0$ as $T \to \infty$.*

The long-run variance of moment indicators g_t is defined as

$$\Omega(\theta) = \lim_{T\to\infty} \mathrm{Var}\left[T^{-\frac{1}{2}} \sum_{t=1}^{T} g_t(\theta)\right]. \tag{2.5}$$

As counterparts of (2.4) and (2.5), sample average and the sample long-run variances are given by

$$\hat{g}_T(\theta) = T^{-1} \sum_{t=1}^{T} g_{tT}(\theta), \qquad \widehat{\Omega}(\theta) = M \sum_{t=1}^{T} \left(\frac{g_{tT}(\theta) g_{tT}(\theta)'}{T}\right). \tag{2.6}$$

The EL estimator can be obtained as the solution of the following problem (Newey and Smith, 2004),

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \sup_{\lambda\in\Lambda(\theta)} L(\theta,\lambda), \quad \text{where } L(\theta,\lambda) = 2\sum_{t=1}^{T} \frac{\rho\left(\lambda' g_{tT}(\theta)\right)}{T}, \tag{2.7}$$

where $\Lambda(\theta) = \{\lambda \in R^q; \lambda' g_{tT}(\theta) \in R\}$ and $\rho(x) = \ln(1-x)$. If we define $\rho_j(x)$ as the j^{th} derivative of ρ, then EL corresponds to $\rho_1(0) = \rho_2(0) = -1$.

Robust test statistics have been developed in the presence of weak identification as in Assumption 1, which make use of quadratic forms of the first-order conditions. Below, we focus on two testing methods by GS (2008) and Kleibergen (2005). Let

$$D(\theta) = \sum_{t=1}^{T} \frac{\rho_1\left(\lambda' dg_{tT}(\theta)\right)}{T} \in R^{q\times p}, \quad \text{for } dg_{tT}(\theta) = \frac{\partial g_{tT}(\theta)}{\partial\theta} \in R^{q\times p}. \tag{2.8}$$

Further, define

$$\begin{aligned} G(\theta) &= T^{\frac{1}{2}} \hat{g}_T(\theta)' \widehat{\Omega}(\theta)^{-1} D(\theta), \\ V(\theta) &= D(\theta)' \widehat{\Omega}(\theta)^{-1} D(\theta). \end{aligned} \tag{2.9}$$

Then, the LM test is given by (GS, 2008; Otsu, 2006)

$$\mathrm{LM} = \frac{G(\theta) V(\theta)^{-1} G(\theta)'}{2}. \tag{2.10}$$

Suppose assumptions 1–3 hold. Then, under the null hypothesis, the test converges to the Chi-squared distribution with the degree of freedom equal to the number of parameters p,

$$\mathrm{LM} \xrightarrow{d} \chi^2_p. \tag{2.11}$$

The K-test by Kleibergen (2005) is summarized as follows. Define unsmoothed version of moment indicators,

$$\tilde{g}_T(\theta) = T^{-1} \sum_{t=1}^{T} g_t(\theta), \qquad s_T(\theta) = T^{\frac{1}{2}} \sum_{t=1}^{T} \left(\tilde{g}_T(\theta) - E\tilde{g}_T(\theta)\right), \tag{2.12}$$

and the long-run variance can be written as

$$\Sigma(\theta) = \lim_{T\to\infty} E s_T(\theta) s_T(\theta)'. \tag{2.13}$$

As is widely known, the long-run variance is estimable by conventional HAC estimator, denoted as $\widetilde{\Sigma}(\theta)$. One popular choice is Bartlett kernel-based Newey and West estimator. Further define

$$\begin{aligned} H(\theta) &= T^{\frac{1}{2}} \tilde{g}_T(\theta)' \widetilde{\Sigma}(\theta)^{-1} D(\theta), \\ W(\theta) &= D(\theta)' \widetilde{\Sigma}(\theta)^{-1} D(\theta). \end{aligned} \tag{2.14}$$

The K-test is given as

$$K = 2H(\theta) W(\theta)^{-1} H(\theta)'. \tag{2.15}$$

As in the LM test above, the K-test also achieves limiting Chi-squared distribution with the p degrees of freedom,

$$K \xrightarrow{d} \chi^2_p, \tag{2.16}$$

then LM and K are asymptotically equivalent. It is known that both LM and K-tests have advantages in terms of power as the degree of freedom is equal to p, which is often less than the number of moments q (Kitamura, 1997). However, empirical null distributions of the tests may be vulnerable in the presence of serial dependence. Some simulation results that include GS (2008) and Otsu (2006) show unstable size properties of the tests. In the next section, we employ sieve bootstraps to improve the size performances.

3. Sieve Bootstrap Tests

We make use of sieve bootstrap methods developed by Bühlman (1997). We suitably employ the method in this EL context. The algorithms are stated as a sequence.

(a) Consider autoregressive(AR) model for $\{g_t\}$ process,

$$g_t(\theta) = \sum_{j=1}^{p} \alpha_j g_{t-j}(\theta) + e_t, \tag{3.1}$$

where the lag p is set to increase with the sample size. Practical choice for p includes AIC or BIC. As is clear, dependent moment indicators are approximated by an AR(∞) model.

(b) Run a regression and obtain the residuals $\{\hat{e}_t\}$. For example, if we consider a linear model $y_t = \beta Y_t + e_t$, and $Y_t = \phi z_t + v_t$, then $g_t = z_t e_t$, where the z_t is an instrumental variable. Then, $\hat{g}_t = z_t \hat{e}_t$, and $\hat{e}_t = y_t$ under the null hypothesis of $\beta = 0$.

(c) Recenter the residuals as $\hat{e}_{c,t} = \hat{e}_t - (T-k)^{-1} \sum_{t=1}^{T} \hat{e}_t$, and resample these residuals. Denote them as $\{\hat{e}^*_t\}$.

(d) Reconstruct the process $\{\hat{g}\}$ based on estimated coefficients as

$$\hat{g}^*_t(\theta_0) = \sum_{j=1}^{p} \hat{\alpha}_j \hat{g}^*_{t-j}(\theta_0) + \hat{e}^*_t. \tag{3.2}$$

We label $\hat{g}_t^*(\theta_0)$ as bootstrapped moment indicators.

(e) Compute the EL tests. For LM and K, bootstrapped tests are denoted as LM^*, K^*.

(f) Replicate B times steps from (b)–(e) and obtain the empirical distribution of the tests. Decide bootstrap critical values at the $\alpha * 100\%$ significance level, which are labelled as $cv^*(\alpha)$. We choose $B = 500$ in our simulation.

(g) Reject the null hypothesis if

$$\mathrm{LM}\,(\text{or } K) > cv^*(\alpha). \tag{3.3}$$

In the next section, we conduct a simulation study to see the size performance of the original EL tests and their bootstrapped versions.

4. Simulation Results

We conduct a simulation study to see how effectively the bootstrap methods work. Simulation design basically follows GS (2008) and Otsu (2006). Consider a linear model with a single endogenous variable,

$$\begin{aligned} y &= Y\theta_0 + e, \qquad (4.1) \\ Y &= Z\Pi + u, \end{aligned}$$

where Y is n by 1 and instruments Z are n by k matrix. As seen above, the value of Π reflects the strength of instrumental variable(IV). We set the value of Π equal to 0.5. Note that the null distribution of the test does not depend on the value of Π, which is confirmed by trying different values in our simulation, though not reported. The number of moments k is set to 1, for simplicity. The null hypothesis is given by $\theta_0 = 0$, and the moment condition equals to $E(Z_t e_t) = 0$ for $t = 1, \ldots, n$. The sample size is n is set to 100.

Serial dependence is allowed in both innovation process of y and instruments. We simply assume that they follow AR(1) processes.

$$e_t = \rho e_{t-1} + v_t \quad \text{and} \quad Z_t = \delta Z_{t-1} + \epsilon_t, \tag{4.2}$$

where v_t and ϵ_t are independent Normal$(0, 1)$. The AR coefficients are equally set as $\rho = \delta = 0.5$ or 0.9. Large values of AR process are expected to deteriorate the size performance of the test. Further, we set $\mathrm{Cov}(u, v) = 0.5$. Different choice of this covariance structures have limited impact on the performance of the tests. We only report one case.

In computing long-run variance for LM and K test given in Section 2, we consider a rule for the bandwidth choice as $M = T^\gamma$, for $\gamma = [0.2, 0.5)$. By this rule, we include the values of M from 2 up to 10. Note that the optimal choice of bandwidths are not formally provided in the EL context. Also, for the LM test, Bartlett kernel is used for smoothing the moment indicators. To compute the null rejection probabilities, we conduct 1000 iterations. Bootstrap replications are set to 500 to save time.

Table 1 and Table 2 show the results and are summarized as follows. Firstly, we observe that the original LM and K-tests generate size distortions that show unstable rejection probabilities over the different values of the bandwidths. The type I errors of the LM test tends to decrease with the bandwidth, where the under-rejection problem becomes particularly severe in the case of large value of

Table 1: Rejection probabilities under the Null at the 5% significance level

AR coefficient = 0.5	Bandwidth				
Test	2	4	6	8	10
LM	0.084	0.053	0.031	0.015	0.005
K	0.091	0.075	0.074	0.079	0.084
LM*	0.042	0.048	0.055	0.050	0.049
K*	0.041	0.042	0.045	0.046	0.045

[note] sample size = 100, strength of IV = 0.5, the number of iteration equals to 1000. Bootstrap replication = 500.

Table 2: Rejection probabilities under the Null at the 5% significance level

AR coefficient = 0.9	Bandwidth				
Test	2	4	6	8	10
LM	0.272	0.123	0.058	0.028	0.002
K	0.416	0.291	0.220	0.195	0.180
LM*	0.029	0.027	0.029	0.030	0.034
K*	0.027	0.026	0.029	0.028	0.031

[note] Refer to the note in the Table 1.

the bandwidth. The K-test over-rejects regardless of the bandwidth choice. Unlike the i.i.d. setup, the tests behave very differently as the magnitude of dependence increases. However, sieve bootstrapped version of tests achieve sizes close to the nominal level of 5%, which are minimally affected by the bandwidth selection. Though minor in magnitude, LM test works better than K-test when bootstraps are combined. Table 2 lists the results when the AR coefficient is equal to 0.9, which implies a strong correlation of the data. Sizes of the LM and K-test show a sharply decreasing pattern with the bandwidth. As expected, the unstable pattern becomes more severe than in the previous case. However, bootstrapped versions (though slightly under-rejected) yield a reasonable size performance. Their performance could be further improved if large values of AR lags are allowed at the cost of computing time. From the above results, we confirm the advantages of applying sieve bootstraps to the EL tests. Gauss programs used in the simulations are available upon request.

5. Conclusion

We consider sieve bootstraps for empirical likelihood tests in time series moment condition models. Test statistics often generate size distortions due to serial dependence. We found a significant size refinement of the bootstrapped versions of the LM type test statistic that was nearly irrespective of the bandwidth choice in long-run variance estimation.

Acknowledgement

The author is grateful to Hosin Song and two anonymous referees.

References

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica*, **59**, 817–858.

Bühlman, P. (1997). Sieve bootstrap for time series, *Bernoulli*, **3**, 123–148.

Chang, Y. and Park, J. (2004). A sieve bootstrap for the test of a unit root, *Journal of Time Series Analysis*, **24**, 379–400.

Guggenberger, P. and Smith, R. (2008). Generalized empirical likelihood tests in time series models with potential identification failure, *Journal of Econometrics*, **142**, 134–161.

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes, *Annals of Statistics*, **25**, 2084–2102.

Kitamura, Y. and Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimations, *Econometrica*, **65**, 861–874.

Kleibergen, F. (2005). Testing parameters in GMM without assuming that they are identified, *Econometrica*, **73**, 1103–1123.

Newey, W. and Smith, R. (2004). Higher order properties of GMM and empirical likelihood estimators, *Econometrica*, **72**, 219–255.

Newey, W. and West, K. (1994). Automatic lag selection in covariance matrix estimation, *Review of Economic Studies*, **61**, 631–653.

Otsu, T. (2006). Generalized empirical likelihood inference for nonlinear and time series models under weak identification, *Econometric Theory*, **22**, 513–527.

Palm, F. C., Smeekes, S. and Urbain, J.-P. (2010). A sieve bootstrap test for cointegration in a conditional error correction model, *Econometric Theory*, **26**, 647–681.

Park, J. (2002). An Invariance Principle for sieve bootstrap in time series, *Econometric Theory*, **18**, 469–490.

Stock, J. and Wright, J. (2000). GMM with weak identification, *Econometrica*, **68**, 1055–1096.

Received March 22, 2013; Revised May 1, 2013; Accepted May 21, 2013