Communications of the Korean Statistical Society
2009, Vol. 16, No. 1, 137–141

Choice of the Kernel Function in Smoothing Moment Restrictions for Dependent Processes

Jin Lee[1,a]

[a]Dept.of Economics, Konkuk Univ.

Abstract

We study on selecting the kernel weighting function in smoothing moment conditions for dependent processes. For hypothesis testing in Generalized Method of Moments or Generalized Empirical Likelihood context, we find that smoothing moment conditions by Bartlett kernel delivers smallest size distortions based on empirical Edgeworth expansions of the long-run variance estimator

Keywords: Moment conditions, Edgeworh expansions, kernel function.

1. Introduction

In moment condition models with dependent processes, there have been many procedures for testing hypotheses under moment restrictions. They include, for example, GMM-based statistic by Kleibergen (2005) and generalized empirical likelihood (GEL)-based tests by Guggenberger and Smith (2006) and Otsu's (2006). They provide useful inferences particularly in the presence of weak identifications in linear models. These tests, however, are not equipped with optimal choice of smoothing parameters and of kernel function, which are crucial in smoothing moment restrictions. Owing to Smith (2005), smoothed moment conditions lead to heteroskedasticity and autocorrelation consistent (HAC) estimators. Thus, one can find the way to select optimal smoothing parameter based on empirical expansions of HAC estimators using Velasco and Robinson (2001).

The optimal bandwidth and the kernel in terms of mean squared error criteria have been well developed in the HAC literature (*e.g.*, Andrews, 1991). For hypothesis testing in moment condition models, such optimal rate is not necessarily optimal in terms of different criteria such as coverage probabilities or type I and type II errors. Sun *et al.* (2005) consider optimal bandwidth selections for HAC estimators which minimize the weighted sum of type I and type II errors, through Edgeworth expansions of nonstandard limit of the Wald type test statistics. On the other hand, GMM- or GEL-based statistic generate standard chi-square asymptotic limit under the null, thus we can make use of Edgeworth expansions with rather direct applications of Velasco and Robinson (2001), but without imposing MSE-optimal growth rate of the bandwidth.

The optimal kernel for smoothing moment restrictions can be also obtained within a suitable class of kernels, whose convolution yields the optimal weight for HAC estimators. We find that smoothing by Bartlett kernel is optimal.

2. Main Results

Consider $\{z_t\}_{t=1}^{T}$, be a R^m-valued stationary mixing process with $\sum_{j=1}^{\infty} j^2\alpha(j)^{(\nu-1)/\nu} < \infty$ for $\nu > 1$. Let θ_0 be the true parameter for θ and $g : R^m \times \Theta \to R^q$ be a vector of unknown function. Moment

[1] Associate Professor, Department of Economics, Konkuk University, Seoul 143-701, Korea. E-mail: leejin@konkuk.ac.kr

conditions are given as

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

For dependent processes, it is natural to use smoothed moment restrictions using a kernel weighting (*e.g.*, Smith, 2005; Kitamura and Stutzer, 1997; Tripathi and Kitamura, 2001),

$$g_{tT}(\theta) = M^{-1} \sum_{j=t-T}^{t-1} w\left(\frac{j}{M}\right) g_{t-j}(\theta), \tag{2.2}$$

where M is the bandwidth and $w(\cdot)$ is a kernel function. Many testing procedures including Guggenberger and Smith (2006) employ truncated kernel for $w(\cdot)$, though not justified.

Meanwhile, in order to allow weak identifications from Stock and Wright (2000), let $\theta_0 = (\alpha', \beta')'$ be in the interior of the compact set $\Theta = A \times B \subset R^p$, where $p = p_\alpha + p_\beta$. Following Guggenberger and Smith (2006) and Otsu (2006), we impose certain conditions below such that α is weakly identified, while β is strongly identified.

Assumption 1.

(1) $E[T^{-1} \sum_{t=1}^{T} g_t(\theta)] = T^{-1/2} m_{1T}(\theta) + m_2(\beta)$, *where* $m_1 : \Theta \rightarrow R^q$, *such that* $m_{1T} \rightarrow m_1$ *uniformly on* Θ *and* $m_1(\theta_0) = 0$,

(2) $m_2 : B \rightarrow R^q$ *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 long-run variance of $\{g_t\}$ is defined as

$$\Omega(\theta) = \lim_{T\rightarrow\infty} \text{Var}\left[T^{-\frac{1}{2}} \sum_{t=1}^{T} g_t(\theta)\right]. \tag{2.3}$$

The long-run variance $\Omega(\theta)$ equals to 2π times the spectral density $f(\theta; \lambda)$ of the g_t process at zero frequency, defined as $f(\theta; \lambda) \equiv f_\theta(\lambda) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} R(j; \theta) \exp^{-ijt}$, where $R(j; \theta) = E[g_{t+j}(\theta) g_t(\theta)']$. Some standard conditions are given on $f_\theta(\lambda)$.

Assumption 2. $f_\theta(\lambda)$ *is* q *times differentiable at* $\lambda = 0$, *where the* q^{th} *derivative is given by* $f_\theta^{(q)}(0) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} |j|^q R(j; \theta)$, *for positive integer* $q \in [0, \infty)$, $R(j; \theta) = E[g_{t+j}(\theta) g_t(\theta)']$ *and* $\|f_\theta^{(q)}(0)\| < \infty$.

The Assumption 2 is standard in the spectral density estimation, (*e.g.*, Hannan, 1970). The q^{th} derivative $f_\theta^{(q)}(0)$ characterize smoothness of $f_\theta(\cdot)$ at the zero frequency. If q equals to 2, for example, $f_\theta^{(2)}(0) = (-1)[d^2 f(\lambda)/d\lambda^2|_{\lambda=0}]$.

Also, one can impose standard conditions on kernel weights. Smoothed moment conditions lead to HAC type estimators of sample moments with the weight equals to the convolution of $w(\cdot)$, *i.e.*, $k(y) = \int_{-\infty}^{\infty} w(x-y) w(x) dx$.

Assumption 3.

(1) $k(x) : R \rightarrow [-1, 1]$ *is symmetric and continuous at zero with* $k(0) = 1$, $|k(x)| \leq C|x|^{-b}$ *as* $x \rightarrow \infty$ *for* $b > 2$,

(2) $K(\lambda) \geq 0$, *for all* $\lambda \in [-\pi, \pi]$, *where* $K(\lambda) = (2\pi)^{-1} \int_{-\infty}^{\infty} k(x) e^{-i\lambda x} dx$.

Assumption 3 states a regularity condition for kernel functions standard in the nonparametric context. 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 $k(x)$. The smoothness of the kernel function is characterized by

$$k_q = \lim_{x\to 0} \frac{1-k(x)}{|x|^q}, \quad \text{for } q \in [0,\infty).$$

In this analysis, we restrict our attentions to kernels with $q \le 2$, which are popularly used in practice. Higher-order kernels are often ruled out in econometrics context since it does not guarantee positive definiteness for long-run variance estimators. For example, $k_1 = 1$ for Bartlett, $k_2 = 6, 1.6449$ and 1.4212 for Parzen, Daniell and QS kernels, respectively (Priestley, 1981).

Now, in order to find a clue for optimal choice of the bandwidth and the kernel weight, we consider Edgeworth expansions of the normalized sample moments $T^{1/2}\bar{g}_T(\theta_0)$, where $\bar{g}_T = T^{-1}\sum_{t=1}^{T} g_{tT}(\theta)$ under the null, $\theta = \theta_0$. Here, $\{g_t\}$ is assumed to be Gaussian for simple Edgeworth expansions (*e.g.*, SPJ). Also, the asymptotic distribution of $T^{1/2}(\widehat{\theta}-\theta_0)$, for GMM or GEL type estimator $\widehat{\theta}$ is linearly related to $T^{1/2}\bar{g}_T(\theta_0)$, but the latter is easier to handle in a sense that Edgeworth expansions in VR can be more directly utilized. Asymptotic normality of the sample moment can be given in Smith (2005),

$$S_T = \frac{c_k^{-1}T^{\frac{1}{2}}[\bar{g}_T(\theta_0) - E\bar{g}_T(\theta_0)]}{\Omega^{\frac{1}{2}}(\theta_0)} \to N(0,1), \tag{2.4}$$

where $c_k = \int_{-\infty}^{\infty} k(z)dz$.

For valid third order Edgeworth expansions (*cf:* Velasco and Robinson, 2001, Section 6), we impose the conditions on the bandwidth, $M = N^{\delta}$, for $(1+q)^{-1} \le \delta < 1$. Let e be the size distortion at the $\alpha\%$ significance level, *i.e.*, $e = 1 - P(S_T \le z_\alpha) - \alpha$, for associated critical value z_α. One can get, after some simplifications,

$$\begin{aligned} e = {} & \frac{1}{4}\phi(z_\alpha)z_\alpha\left[\int_{-\infty}^{\infty} k^2(u)du(z_\alpha^2+1) + 2\int_{-\infty}^{\infty} k(u)du\right]\left(\frac{M}{T}\right) \\ & + \frac{1}{2}\phi(z_\alpha)z_\alpha\left[\frac{f_\theta^{(q)}(0)k_q}{q!f_\theta(0)}\right]M^{-q} + o\left(\frac{M}{T}\right), \end{aligned} \tag{2.5}$$

where $\phi(\cdot)$ is *pdf* of normal random variable.

Note that typical shape of the spectral density of most economic time series, that $f_\theta^{(q)}(0) > 0$. This indeed rules out the Bartlett kernel with $q = 1$, which makes it impossible to obtain the first order condition with respect to M.

Given this, one can directly obtain the optimal rate by the first order necessary condition for minimizations. The following lemma is basically equivalent to Velasco and Robinson (2001, Theorem 4).

Lemma 1.

$$M^* = \left(2q\frac{B}{A}\right)^{\frac{1}{q+1}} \times T^{\frac{1}{q+1}}, \quad \textit{where } A = k_q^{-1}\left[\int_{-\infty}^{\infty} k^2(u)du\left(z_\alpha^2+1\right) + 2\int_{-\infty}^{\infty} k(u)du\right] \textit{ and } B = \left|\frac{f_\theta^{(q)}(0)}{q!f_\theta(0)}\right|.$$

The M^* involves known functions of kernels in A and unknown quantities $f_\theta(0)$ and $f_\theta^{(q)}(0)$ in B, which are consistently estimable. The optimal rate of growth for the bandwidth is $T^{1/(q+1)}$, which is in fact the minimum rate of the growth validating the third order Edgeworth expansions in VR.

Table 1: Values of A in Theorem 1 with associated critical values (cv) in normal distribution.

kernel	$cv = 0$	0.5	1.28	1.64	1.96	2.58
Parzen	0.34	0.36	0.49	0.58	0.68	0.94
Daniell	1.82	1.98	2.82	3.46	4.16	5.87
QS	2.81	2.99	3.97	4.71	5.52	7.50

This rate of growth equals to the optimal rate in SPJ for consistent HAC estimators. SPJ obtain the optimal bandwidth by imposing constraints on the weights for size distortions and type II errors, which otherwise may yield corner solutions for minimizations in (2.5). In the presence of weak identifications where the test statistic is not consistent under the alternatives (see Otsu, 2006), only size distortions enter into an objective function to derive the optimal bandwidth.

Next, we turn to the minimum size distortions as well as the choice of the kernel function. The minimum size distortions with the optimal bandwidth M^* yields the following result.

Theorem 1. *Let the minimum size distortions be e^* with M^* plugged into (2.5). Then, $e^* = 1/4\,(2q)^{1/(1+q)}\phi(z_\alpha)z_\alpha[(1+q^{-1})A^{q/(q+1)}B^{1/(q+1)}] \times T^{-q/(q+1)}$ and $A_{parzen} < A_{Daniell} < A_{QS}$ for any z_α.*

The quantities A_{parzen}, $A_{Daniell}$ and A_{QS} denote the the quantities A in Lemma 1 involving Parzen, Daniell and QS kernel, respectively. The proof is simple. Among quadratic kernels with $q = 2$, it suffices to show the ranking of the kernels based on the quantity A. Simple numerical computations yield values of A for each kernel as follows.

$$\begin{aligned} A &= 0.0899z_\alpha^2 + 0.3399, && \text{for Parzen} \\ &= 0.6079z_\alpha^2 + 1.8237, && \text{for Daniell} \\ &= 0.7036z_\alpha^2 + 2.8144, && \text{for QS,} \end{aligned} \tag{2.6}$$

where $k_2 = 6, 1.6449, 1.4212$; $\int_{-\infty}^{\infty} k^2(z)dz = 0.5392, 1, 1$; and $\int_{-\infty}^{\infty} k(z)dz = 3/4, 1, 3/2$ for Parzen, Daniell and QS, respectively. Thus, we obtain $A_{parzen} < A_{Daniell} < A_{QS}$ for any values of z_α. This completes the proof.

Besides, the Theorem 1 shows that Bartlett kernel with $q = 1$ is strictly dominated by quadratic kernels with $q = 2$. This implies, in turn, that for smoothing moment restrictions, truncated kernel, though popularly used in GMM or GEL context, is not optimal in terms of null rejection probabilities. Also, it is obvious that the kernel with smaller value of A is simply preferred to that with larger values of A.

In Table 1 below, we present some simple numerical exercises to see the relative ranking of the qudratic kernels.

3. Conclusion

We consider the choice of kernel function in the context of GMM and GEL for dependent processes. Using the results of third order Edgeworth expansions for null rejection probabilities, it is found that Bartlett kernel for smoothing moment conditions, in turn, Parzen kernel for long-run variance estimation generates the smallest size distortions among quadratic kernels. This contrasts with popular use of truncated kernels in practice.

References

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

Guggenberger, P. and Smith, R. (2006). Generalized empirical likelihood tests in time series models with potential identification failure, Manuscript, UCLA.

Hannan, E. J. (1970). *Multiple Time Series*, Wiley, New York.

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

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

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

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Academic Press, New York.

Smith, R. J. (2005). Automatic positive semidefinite HAC covariance matrix and GMM estimation, *Econometric Theory*, **21**, 158–170.

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

Sun, Y. X., Phillips, P. C. B. and Jin, S. (2005). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing, *Econometrica*.

Tripathi, G. and Kitamura, Y. (2003). Testing conditional moment restrictions, *The Annals of Statistics*, **31**, 2059–2095.

Velasco, C. and Robinson, P. (2001). Edgeworth expansions for spectral density estimates and studentized sample mean, *Econometric Theory*, **17**, 497–539.

Received September 2008; Accepted October 2008