Empirical likelihood methods with application to econometrics

Francesco Bravo
Lecture 4

- GEL and nonparametric curve estimation
- GEL with dependent data
1 GEL and smoothing

We consider GEL in the context of curve estimation problems. Let $\bar{K}_h(z) := K(z/h)$ denote a kernel function.

1.1 Density estimation

Let $(z_i)_{i=1}^n$ be i.i.d. observations from an unknown distribution $F$ with density $f$. The kernel density estimator for $f$ at an arbitrary fixed point $z$ is

$$\hat{f}(z) = \sum_{i=1}^n \frac{K_h(z_i - z)}{(nh)}.$$

Often confidence intervals for $f(z)$ are required (for example to test a hypothesis about $f(z)$).

One possibility is to use the bootstrap. However because of the bias in the kernel estimate $h$ should be of order $n^{-1/(r+1)}$ (where $r$ is the order of the kernel) to obtain valid confidence intervals with a two-sided percentile $t$ method. This results typically in poor empirical coverage, because the required undersmoothing affects greatly the precision of the variance estimation needed for the percentile $t$ method.

Alternatively Chen (1996) propose to use EL. The rationale behind this proposal is that EL does not require explicit variance estimation and therefore this should result is better confidence intervals.
Let \( g_{ih}(f(z)) = K_{ih}(z_i - z)/h - f(z) \), and note that assuming \( f \) to be sufficiently smooth
\[
E[K_{ih}(z_i - z)/h] = f(z) + k d^r f(z)/dz^r h^r/r! + o(h^r).
\]
Thus the “moment indicator” \( g_{ih}(\cdot) \) is biased, and therefore confidence intervals do not have the right coverage. One way to solve the problem is to correct explicitly the confidence interval by shifting it by \( k d^r f(z)/dz^r h^r/r! \). Another way is to undersmooth.

The following theorem show that GEL with undersmoothing can be used to obtain confidence intervals for \( f(z) \) with asymptotic correct coverage. The theorem generalises some of Chen’s (1996) results. Let
\[
I_\alpha = \{ f(z) | \left( \widehat{P}_\rho \left( f(z), \lambda \right) - \widehat{P}_\rho(0) \right) \leq c_\alpha \}
\]
where \( \Pr \left( \chi^2_1 \leq c_\alpha \right) = \alpha \) denote a GEL confidence interval with nominal coverage \( \alpha \).

**Theorem 1** Assume that (I) the \( r \)th \((r \geq 2)\) order kernel \( K(\cdot) \) is bounded and has a compact support, (II) \( f \) has continuous derivatives up to the \( r \)th order in a neighbourhood of \( z \) and \( f(z) > 0 \), (III) \( h \to 0 \), \( nh \to \infty \), and \( n^{1/2}h^{r+1/2} \to 0 \). Then
\[
\Pr (f(z) \in I_\alpha) = \alpha + o(1).
\]
Proof. Let \( g_{ih}(f(z)) := g_{ih}, \mu_j = E\left(g_{ih}^j\right) \ (j = 1, 2, \ldots) \) and let \( \delta_n = n^{-1/2} \max_i |g_{ih}|. \) Define \( \Lambda_n = \left\{ |\lambda| \leq n^{-1/2} \delta_n^{-1/2} \right\} \) and note that \( \max_i |g_{ih}| = o_p(\sqrt{n}) \) provided \( n^{1/2}h \to \infty. \) Then \( \max_i \sup_{\lambda \in \Lambda_n} |\lambda g_{ih}| \leq \delta_n^{1/2} = o_p(1) \) and by the usual arguments

\[
0 = \hat{P}_n \left( f(z), \lambda \right) \leq 2 \left| \lambda \right| |\hat{g}_h| - \hat{\lambda}^2 \hat{g}_h^2.
\]

Note that by LLN \( \hat{g}_h^2 \xrightarrow{p} \mu_2 = O(h^{-1}) \) and that \( \hat{g}_h = O_p(n^{-1/2} + h^r). \) Therefore \( \left| \hat{\lambda} \right| \leq |\hat{g}_h| / \mu_2 = O_p(n^{-1/2}h + h^{r+1}), \) and note that \( \hat{\lambda} \in \Lambda_n. \) Then by the usual expansion for the FOC for \( \hat{\lambda} \) we get \( \hat{\lambda} = -\hat{g}_h / \mu_2 + o_p(1). \) Thus expanding \( \hat{P}_n \left( f(z), \hat{\lambda} \right) \) we get

\[
\hat{P}_n \left( f(z), \hat{\lambda} \right) - \hat{P}_n (0) = -2\hat{\lambda}\hat{g}_h + \hat{\lambda}^2 \hat{g}_h^2 + o_p(1) = n\hat{g}_h^2 / \mu_2 + o_p(1).
\]

Let \( Z = n^{1/2} (\hat{g}_h - \mu_1) / \mu_2^{1/2} \) and note that \( Z \xrightarrow{d} N(0, 1) \) so that

\[
\hat{P}_n \left( f(z), \hat{\lambda} \right) - \hat{P}_n (0) \xrightarrow{a} (Z + \mu_1 / \mu_2^{1/2})^2 \xrightarrow{d} \chi_1^2
\]

iff \( n^{1/2} \mu_1 / \mu_2^{1/2} \to 0, \) which is implied by (III). \( \blacksquare \)
1.2 Local linear regression

Let \( z_i = (y_i, x_i) \) \((i = 1, \ldots, n)\) be i.i.d. \( \mathbb{R}^2 \)-valued random vectors, and consider the following nonparametric regression

\[
y_i = m(x_i) + \varepsilon_i
\]

where \( m(\cdot) \) is an unknown regression function, and \( E(\varepsilon_i | x_i) = 0 \), and \( V(\varepsilon_i | x_i) = \sigma^2_\varepsilon < \infty \). One possible way to estimate \( m(\cdot) \) is to use the Nadaraya-Watson or the Gassler-Müller estimators. Here we consider the local linear smoothing estimators - see for example the monograph of Fan & Gijbels (1996).

Let

\[
\hat{m}(x) = \sum_{i=1}^{n} W_i y_i / \sum_{i=1}^{n} W_i
\]

denote the local linear estimator for \( m(x) \) where

\[
W_i = K_h (x_i - x) \left[ \frac{s_{h2} - (x_i - x) s_{h1}}{h} \right],
\]

\[
s_{hl} = \sum_{i=1}^{n} \left[ (x_i - x) / h \right]^l K_h (x_i - x) / (nh).
\]
Standard results of Fan & Gijbels (1996) show that

\[(nh/\beta(x/h))^{1/2} (\hat{m}(x) - m(x)) / (V(x) f(x))^{1/2} \overset{d}{\to} N(B_n, 1)\]

where

\[B_n = \alpha(x/h) (\beta(x/h) V(x) / f(x))^{1/2} m''(x) (nh)^{1/2} h^2\]

with \(V(x) = V(y_i|x_i = x)\), and the functions \(\alpha(\cdot)\) and \(\beta(\cdot)\) involve \(\int_{-1}^{x/h} u^j K(u) du\) \((j = 1, 2, 3)\). By letting \(nh^5 \to 0\) \(B_n \to 0\) we obtain a confidence interval with asymptotic correct coverage using the normal approximation with estimated standard error \(\left[\beta(x/h) \hat{V}(x) / \left( nh \hat{f}(x) \right) \right]^{1/2}\) and \(\hat{V}(\cdot), \hat{f}(\cdot)\) are kernel estimates.

Alternatively we can use GEL with the same undersmoothing. Let \(g_{ih}(m(x)) = W_i (y_i - m(x_i))\) denote the moment indicator, and note that

\[E[g_{ih}(m(x))] = \gamma(x/h) d^2 m(x) / dx^2 h^2 / 2 + O(h^3)\]

(1)

where \(\gamma(\cdot)\) is a function involving \(\int_{-1}^{x/h} u^j K(u) du\) \((j = 1, 2, 3)\).
The following theorem shows that GEL with undersmoothing can be used to obtain confidence intervals for \( m(x) \) at a fixed \( x \) with asymptotic correct coverage. Let

\[
I_\alpha = \left\{ m(x) \mid \left( \hat{P}_p \left( m(x), \hat{\lambda} \right) - \hat{P}_p (0) \right) \leq c_\alpha \right\}
\]

where \( \Pr \left( \chi^2_1 \leq c_\alpha \right) = \alpha \) denote a GEL confidence interval with nominal coverage \( \alpha \).

**Theorem 2** Assume that (1) the kernel \( K(\cdot) \) is bounded with compact support \([-1, 1]\), (II) \( f(\cdot) \) (the marginal density of \( x_i \)), \( V(y_i|x_i = \cdot) \) and \( m(\cdot) \) have continuous derivatives up to the second order in a neighbourhood of \( x \), and both \( f(x) > 0 \) and \( V(y_i|x_i = x) > 0 \), (III) \( h \to 0 \), \( nh \to \infty \), \( nh^5 \to 0 \), (IV) \( E|y_i|^\alpha < \infty \) for some \( \alpha > 4 \). Then

\[
\Pr ( m(x) \in I_\alpha ) = \alpha + o(1)
\]

**Proof.** Let \( g_{ih}(m(x)) := g_{ih}, \mu_j = E \left( \hat{g}_h^j \right) (j = 1, 2, \ldots) \) and let \( \delta_n = n^{-1/2} \max_i |g_{ih}|. \) Define \( \Lambda_n = \left\{ \lambda \mid |\lambda| \leq n^{-1/2} \delta_n^{-1/2} \right\} \) and note that \( \max_i |g_{ih}| = o_p(n^{1/2}) \) provided \( \alpha > 2 \). Then \( \max_i \sup_{\lambda \in \Lambda_n} |\lambda g_{ih}| \leq \delta_n^{1/2} = o_p(1) \) and by the usual arguments

\[
0 = \hat{P}_p \left( m(x), \hat{\lambda} \right) \leq 2 \left| \lambda \right| |\hat{g}_h| - \lambda^2 \hat{g}_h^2.
\]

Note that \( V(\hat{g}_h) = \mu_2/nh + o_p(1) \) so that using (1) by \( \hat{g}_h = O_p \left( (nh)^{-1/2} + h^2 \right) \), while by \( \hat{g}_h^2 \overset{p}{\to} \mu_2/h \).
Therefore \( |\hat{\lambda}| \leq h |\hat{g}_h| / \mu_2 = O_p \left( (nh)^{-1/2} + h^2 \right) \), and note that \( \hat{\lambda} \in \Lambda_n \). Then by the usual expansion for the FOC for \( \hat{\lambda} \) we get \( \hat{\lambda} \overset{a}{=} -nh\hat{g}_h / \mu_2 \). Thus expanding \( \hat{P}_\rho \left( m(x) , \hat{\lambda} \right) \) we get

\[
\hat{P}_\rho \left( m(x) , \hat{\lambda} \right) - \hat{P}_\rho \left( 0 \right) = nh\hat{g}_h^2 / \mu_2 + o_p(1).
\]

Let \( Z = (nh)^{1/2} (\hat{g}_h - \mu_1) / \mu_2^{1/2} \) and note that \( Z \overset{d}{\to} N(0,1) \) so that

\[
\hat{P}_\rho \left( f(z) , \hat{\lambda} \right) = \left( Z + (nh)^{1/2} \mu_1 / \mu_2^{1/2} \right)^2 + o_p(1) \overset{d}{\to} \chi^2_1
\]

iff \( (nh)^{1/2} \mu_1 / \mu_2^{1/2} \to 0 \), which is implied by (III). \( \blacksquare \)
1.3 Other applications

Similar arguments could be used to construct confidence intervals for the Nadaraya-Watson estimator of \( m(x) \), or for the derivative of \( m(x) \).

Another interesting application is with quantile regressions. Here we follow Whang (2006). Let

\[
g_i(\theta) = x_i [I \{y_i \leq x_i^t \theta\} - q]
\]

denote the moment indicator and note that \( E[g_i(\theta_0)] = 0 \) under the conditional quantile restriction \( \text{Pr}(y_i - x_i^t \theta_0 \leq 0|x_i) = q \) a.s.

Because the indicator function is nondifferentiable, Whang (2006) proposes to use the following smoothed version of \( g_i(\cdot) \)

\[
g_{ih}(\theta) = x_i [K_h(x_i^t \theta - y_i) - q].
\]

He shows that under regularity conditions similar to those used in the above theorems

\[
2 \sum_{i=1}^{n} \log \left(1 + \hat{\lambda}' g_{ih}(\theta_0)\right) \overset{d}{\rightarrow} \chi^2_k.
\]

Using (essentially) the same arguments as those used in Theorem 38 it is possible to generalise his
Theorem 2\textsuperscript{1} to GEL. To be specific let $R_{\alpha} = \left\{ \beta \left| \left( \hat{P}_{\rho} (\theta, \lambda) - \hat{P}_{\rho} (0) \right) \leq c_{\alpha} \right. \right\}$ where $\Pr (\chi_{k}^{2} \leq c_{\alpha}) = \alpha$ denote a GEL confidence region with nominal coverage $\alpha$. Then
\[ \Pr (\theta \in R_{\alpha}) = \alpha + o (1). \]

\textsuperscript{1} Theorem 1, which states that the smoothed EL estimator for $\theta$ defined as $\arg \min_{\theta \in \Theta} \sum \log \left( 1 - \lambda' g_{ih} (\theta) \right)$, is also valid for the GEL estimator. It can also be shown that Theorem 3 holds for GEL, under further differentiability assumptions about $\rho (\cdot)$, similar to those described in Section 3.2.
2 Generalised empirical likelihood and dependent data

We consider two types of dependency structures: martingales and $\alpha$-mixing.

2.1 (Generalised) Dual likelihood for martingales

Mykland (1995) introduced the concept of dual likelihood (DL henceforth) as device for using likelihood methods in the context of martingale inference.

As noted by Mykland (1995) “Martingales methods are a powerful tools for dependent variables inference....A major weakness, however, is that for small samples the quality of the approximation in the martingale CLTs can be quite poor”.

*DL tries to overcome this problem by introducing a likelihood ratio statistic as an inferential tool for martingales.*

The type of martingale we consider in this section is that of a compensated sum of jumps

$$m_t(\theta) = \sum_{0 \leq s \leq t} \Delta m_s(\theta) - \Lambda_t(\theta).$$

Note that (2) assumes that $m(\theta)$ does not have infinite total variation. We now give two examples of dependent data structures that can be cast in (2)
Example 1 (Mykland (1995))

(I) Survival data. Let $H(t) = \int_0^t Y(s) \, d\Lambda_s$ denote an $n$-dimensional vector of cumulative hazards where $Y(s)$ is an $n \times k$ matrix of regressors and $\Lambda_s$ is a $k$-dimensional vector of cumulative coefficients. To estimate (and/or to make inference about) $\theta = \Lambda_t$ one can use the martingale

$$m_t(\theta) = \int_0^t (Y(s)' Y(s))^{-1} Y(s)' dN_t - \Lambda_t$$

where $N_t$ is an $n$-dimensional vector that jumps from 0 to 1 when a patient dies.

II) Time series. Let $y_i = \theta [y_{i-1} \ldots y_{i-p}]' + \varepsilon_i$. Then to estimate $\theta$ one can use the martingale

$$m_t(\theta) = [y_{t-1} \ldots y_{t-p}]' (y_t - \theta [y_t \ldots y_{t-p}]).$$

Note that in this case the compensator $\Lambda_t = 0$.

We now introduce the notion of generalised DL (GDL henceforth).

$$\hat{P}^d_p(\lambda, \theta) = -\lambda' \Lambda_t(\theta) + \sum_{1 \leq s \leq t} \rho(\lambda' \Delta m_s(\theta))$$

which specialises to (up to a sign reversal) (log) DL

$$\hat{P}^d_{EL}(\lambda, \theta) = -\lambda' \Lambda_t(\theta) + \sum_{1 \leq s \leq t} \log (1 - \lambda' \Delta m_s(\theta))$$
for $\rho(v) = \log(1 - v)$. Thus GDL is a concave transformation of a martingale.

The following theorem shows that $\hat{P}_\rho^d(\lambda, \theta)$ can be used in the usual way to obtain asymptotically valid inference for the hypothesis $H_0: \theta = \theta_0$. Let $\hat{\lambda} = \sup_{\lambda \in \Lambda(\theta_0)} \hat{P}_\rho^d(\lambda, \theta)$ where $\Lambda(\theta_0)$ is an open neighbourhood of 0, and let $m^n_t(\theta)$ $0 \leq t \leq t_n$ denote a $k$-dimensional triangular array of martingales. Let $[m^n(\theta), m^n(\theta)]_{t_n} = \sum_{0 \leq t \leq t_n} \Delta m^n_t(\theta) \Delta m^n_t(\theta)'$ denote the quadratic variation of $m^n_t(\theta)$ and let $\sigma_n = \sigma_{\min}(\sqrt{[m^n(\theta), m^n(\theta)]_{t_n}})$. The following theorem generalises Theorem 1 of Mykland (1995).

**Theorem 3** Assume that (I) $S_n = m^n_{t_n}(\theta_0)' [m^n(\theta_0), m^n(\theta_0)]_{t_n} m^n_{t_n}(\theta_0)$ is tight, (II) $\sup_{0 \leq t \leq t_n} \|\Delta m^n_t(\theta_0)\|^2 / \sigma_n = o_p(1)$. Then

$$2 \left( \hat{P}_\rho^d(\hat{\lambda}, \theta_0) - \hat{P}_\rho^d(0) \right) \xrightarrow{d} \chi^2_k.$$

**Proof.** Let $[m^n(\theta), m^n(\theta)]_{t_n} = Q_n D_n Q_n$ denote the spectral decomposition of the quadratic variation of $m^n_t(\theta)$, and $U_n(\theta_0) = D_n^{-1/2} Q_n m^n_{t_n}(\theta_0)$. Note that for $\|\lambda\| < 1 / \nu_n = \sigma_n^{1/2} / (\sup_{0 \leq t \leq t_n} \|\Delta m^n_t(\theta_0)\|)$ an expansion about 0 and T give

$$\left| \hat{P}_\rho^d(\lambda, \theta_0) - \hat{P}_\rho^d(0) + \lambda' U_n(\theta_0) + \lambda' \lambda / 2 \right| \leq (\rho_2(\|\lambda\|, \nu_n) - 1) \lambda' \lambda / 2 \leq o_p(1).$$
hence $\widehat{\lambda}$ is consistent, and by concavity of $\widehat{P}_\rho^d (\cdot)$ is tight because of (I). Then

$$2 \left( \widehat{P}_\rho^d (\lambda, \theta_0) - \rho (0) \right) = U_n (\theta_0)' U_n (\theta_0) + o_p (1)$$

and the result follows by Theorem 5.1 of Helland (1982) and CMT. ■
A situation of particular practical relevance and theoretical interest that does not fit into this framework is that of unstable autoregressive processes.

### 2.1.1 GDL for unstable autoregressive processes

Consider valid asymptotic inference in the context of autoregressive models with possibly multiple and/or complex unit roots.

Let $\theta_0 = [\theta_{10} \cdots \theta_{p0}]'$ denote the vector of unknown parameters of the autoregressive process of order $p$ ($AR(p)$)

$$z_t = \theta_0' z_{t-1} + \varepsilon_t \quad t = 1, \ldots, n$$

(3)

where $z_t = [z_t \cdots z_{t-p+1}]'$, and the innovations $\varepsilon_t$ form a martingale difference sequence with respect to the natural filtration $\mathcal{F}_t$, that is $\varepsilon_t$ is $\mathcal{F}_t$-measurable with $E(\varepsilon_t | \mathcal{F}_{t-1}) = 0 \ a.s.$ for every $t$. Inference for $\theta$ in (3) can be based on

$$m_t(\theta) = (z_t - \theta' z_{t-1}) z_{t-1},$$

which is a martingale for $\theta_0$.

It is well known that the distribution of the least squares estimator of $\theta_0$ (and related test statistics) depends on the location of the roots of the characteristic polynomial of (3), that is

$$\Phi_p(z) = 1 - \sum_{j=1}^{p} \theta_j z^j$$

(4)
The following theorem establishes the weak convergence in the Skorohod space $D[0,1]^p$ of the DGEL statistic for the hypothesis $H : \theta = \theta_0$ for unstable autoregressive processes.

**Theorem 4** Assume that (I) the initial values $y_0, \ldots, y_{1-p}$ are fixed $\mathcal{F}_0$-measurable and are $O_p(1)$, (II) $E(\varepsilon_t^2 | \mathcal{F}_{t-1}) = \sigma^2$ a.s for every $t$, $\sup_t E\left( |\varepsilon_t|^{2+\alpha} | \mathcal{F}_{t-1}\right) < \infty$ for some $\alpha > 0$. Then if all of the roots of (4) are either outside or on the unit circle, with at least one unit root on the unit circle

$$2\left( \hat{P}_\rho^d(\lambda, \theta_0) - \rho(0) \right) \xrightarrow{w} \Xi' \Delta^{-1} \Xi,$$

where

$$\Xi = \left[ \int_0^1 \xi dB \quad \int_0^1 \eta dB \quad \int_0^1 \zeta dB \quad \cdots \quad \int_0^1 \zeta dB \right]' \beta,$$

$$\Delta = \text{diag} \left[ \int_0^1 F \quad \int_0^1 \tilde{F} \quad \int_0^1 H_1 \quad \cdots \quad \int_0^1 H_l \right],$$

and the random elements in $\Xi$, $\Delta$ are as defined in Theorem 3.5.1 of Chan & Wei (1988).

**Proof.** As in Chan & Wei (1988) we use a suitable normalising matrix, say $N_n$, to decompose the original $AR(p)$ process $y_t$ into several separate processes according to the nature of the roots of (4),
so that for $p' = a + b + 2 \sum_{k=1}^{l} d_k \leq p$

$$N_n y_t = Y_{nt} = \left[ \frac{u_t(1)}{n} \ldots \frac{u_t(a)}{n^a} \frac{v_t(1)}{n} \ldots \frac{v_t(b)}{n^b} \frac{x_t(d_1)'}{n} \ldots \frac{x_t(d_l)'}{n^{d_l}} \frac{z_t/\sqrt{n}}{n^{1/2}} \ldots \frac{z_{t-q+1}/n^{1/2}}{n^{1/2}} \right]$$

where

$$u_t(j)/n^j = (1 - L)^{-j} \varepsilon_t (j = 1, \ldots, a),$$

$$v_t(j)/n^j = (1 + L)^{-j} \varepsilon_t (j = 1, \ldots, b),$$

$$(1 - 2 \cos \theta_k L + L^2)^{d_k} x_t(d_k)(j) = \varepsilon_t$ for $(j = 1, \ldots, d_k)$

$$x_t(d_j) = \left[ \frac{x_t(d_k)(1)}{n} \frac{x_{t-1}(d_k)(1)}{n} \ldots \frac{x_t(d_k)(d_k)}{n^{d_k}} \right]'$$

and $\Psi(B) z_t = \varepsilon_t$ with $\Psi(B)$ is a polynomial of degree $q = p - p'$ with all the roots outside the unit circle. Let $\Lambda_n = \{ \lambda : \| N_n^{-1} \lambda \| \leq C \}$ for some $C > 0$. By results of Chan & Wei (1988) $\sum Y_{nt} \varepsilon_t$ and $\sum Y_{nt} Y'_{nt}$ are bounded in probability; furthermore by Chuang & Chan (2002) $\| \sum Y_{nt} Y'_{nt} (\varepsilon_t^2 - \sigma^2) \| = o_p(1)$, and max $\| Y_{nt} \varepsilon_t \| = o_{a.s.}(1)$ and thus

$$\max_t \sup_{\lambda \in \Lambda_n} \left| \lambda' (N_n^{-1})' Y_{nt} \varepsilon_t \right| = o_p(1).$$
Then by the usual Taylor expansion about 0

\[
\sup_{\lambda \in \Lambda_n} \hat{P}_p (\theta_0, \lambda) \leq \rho_0 - \lambda' \left( N_n^{-1} \right)' \sum_t Y_{nt} \varepsilon_t + \\
\lambda' \left( N_n^{-1} \right)' \sum_{i=1}^n \rho_2 \left( \lambda' \left( N_n^{-1} \right)' Y_{nt} \varepsilon_t \right) Y_{nt} Y_{nt} \varepsilon_t^2 N_n^{-1} \hat{\lambda} / 2
\]

\[
\leq \rho_0 - \left\| N_n^{-1} \hat{\lambda} \right\| \left\| \sum_t Y_{nt} \varepsilon_t \right\| - \zeta_n \left\| N_n^{-1} \hat{\lambda} \right\|^2 / 4,
\]

where \( \zeta_n > 0 \ a.s. \). Then \( \left\| N_n^{-1} \hat{\lambda} \right\| \leq \left\| \sum_t Y_{nt} \varepsilon_t \right\| = O_p (1) \). To find a stochastic approximation of \( N_n^{-1} \hat{\lambda} \) we expand the FOCs about 0 and use CMT to obtain

\[
0 = \sum \rho'_1 \left( \left( N_n^{-1} \hat{\lambda} \right)' Y_{nt} \varepsilon_t \right) N_n^{-1} Y_{nt} \varepsilon_t = -N_n^{-1} \sum_t Y_{nt} \varepsilon_t - \sum_{i=1}^n N_n^{-1} Y_{nt} Y_{nt} \varepsilon_t^2 N_n^{-1} \hat{\lambda} + o_p (1)
\]

which yields

\[
N_n^{-1} \hat{\lambda} = \left( \sum Y_{nt} Y_{nt} \varepsilon_t^2 \right)^{-1} \sum_t Y_{nt} \varepsilon_t + o_p (1) \tag{7}
\]

since by results of Chan & Wei (1988) \( \sum Y_{nt} Y_{nt}' \) is nonsingular \( a.s. \). The conclusion of the theorem
follows by the usual second order expansion about 0 and (7) which shows that

\[
2 \left( \hat{P}^d_\rho (\lambda, \theta_0) - \rho (0) \right) = \sum Y_{nt} \varepsilon'_t \left( \sum Y_{nt} Y'_{nt} \varepsilon^2_t \right)^{-1} \sum Y_{nt} \varepsilon_t + o_p (1)
\]

and by the results of Chan & Wei (1988).

To evaluate the power properties of the DGEL statistic we consider the same type of (local) alternative hypotheses considered by Jeganathan (1991), that is \( H_n : \theta_n = \theta_0 + N_n \delta_n \), where \( N_n \) is a suitable normalising matrix, and \( \delta_n \) is a sequence of (nonrandom) vectors such that \( \delta_n \to \delta_0 \) for \( n \to \infty \) and \( 0 < \delta'_0 \delta_0 < \infty \). For example, in the case of an unstable AR (1) process \( H_n : \theta_n = 1 + n^{-1} \delta_n \).

**Proposition 1** Under the same assumptions of Theorem 42, and under \( H_n : \beta_n = \beta_0 + N_n \delta_n \), if \( \delta_n \to \delta_0 \equiv [u_0 \ v_0 \ \cdots \ u_l \ v_l \ z] \), then

\[
2 \left( \hat{P}^d_\rho (\lambda, \theta_n) - \rho (0) \right) \xrightarrow{w} \Xi'_\delta_0 \Delta^{-1}_\delta \Xi_\delta_0,
\]

where

\[
\Xi_\delta_0 = \left[ \int_0^1 \xi'_1 (u_0) dB \ \int_0^1 \xi'_1 (v_0) dB \ \int_0^1 \xi'_1 (u_1, v_1) dB \ \cdots \ \int_0^1 \xi'_1 (u_l, v_l) dB \ W' \right]',
\]

\[
\Delta_\delta_0 = \text{diag} \left[ \int_0^1 \Gamma_1 (u_0) \ \int_0^1 \Gamma_2 (v_0) \ \int_0^1 H (u_1, v_1) \ \cdots \ \int_0^1 H (u_l, v_l) \ \Gamma_3 \right],
\]

and the random elements in \( \Xi_\delta_0, \Delta_\delta_0 \) are as defined in Theorem 1 of Jeganathan (1991).
Proof. See Bravo (2006) (for the ECR statistic) ■

GDL can be used to obtain valid inferences in unstable autoregressive models that contain deterministic components. Consider

\[ z_t^d = \eta' d_t + z_t \]
\[ z_t = \theta_0' z_{t-1} + \varepsilon_t \quad t = 1, \ldots, n, \]

where \( z_t = [z_t \cdots z_{t-p+1}]' \) and \( d_t = [1 \cdots t_{q}^{-1}]' \) is a polynomial trend\(^2\).

Let \( h = [h_1 (r) \cdots h_q (r)]' \), \( H = \int_0^1 hh' \), and for a random vector \( X \in D [0, 1]^q \), let \( X_h \) denote its projection on the space orthogonal to \( \left( \int_0^1 h' \right) H^{-1} h \). Let \( 2 \left( \hat{P}_\rho^d (\lambda, \theta_0, \hat{\eta}) - \rho (0) \right) \) denote the profile GDL statistic.

**Proposition 2** Under the same assumptions of Theorem 42, and under \( H_0 : \theta = \theta_0 \), then

\[ 2 \left( \hat{P}_\rho^d (\lambda, \theta_0, \hat{\eta}) - \rho (0) \right) \overset{w}{\to} \Xi_h' (\Delta - \Upsilon H^{-1} \Upsilon')^{-1} \Xi_h, \]

\(^2\) In fact the result holds for deterministic sequences of general form satisfying the following regularity condition: for \( j = 1, \ldots, q \) there exist \( \lambda_j \) and linearly independent bounded functions \( h_j (r) \) such that \( d_{j[n]} / n^{\lambda_j} \to h_j (r) \) as \( n \to \infty \) uniformly in \( r \in [0, 1] \).
where

\[
\Xi_h = \left[ \int_0^1 \xi'_h dB \int_0^1 \eta'_h dB \int_0^1 \zeta'_{1h} dB \cdots \int_0^1 \zeta'_{lh} dB \ N'_h (0, \Sigma) \right]',
\]

\[
\Upsilon = \sigma \int_0^1 \left[ \xi dB \ \eta dB \ \xi dB \ \cdots \ H_l dB \ \Sigma \right] h',
\]

and \( \Delta \) is as in (5).

**Proof.** See Bravo (2006) (for the ECR statistic) \( \blacksquare \)

**Example 2 (Double unit roots with intercept)** We consider an AR(2) process with an intercept

\[
z = \sum_{j=1}^{2} \theta_j z_{t-j} + \eta_0 + \varepsilon_t, \quad t = 1, \ldots, n. \tag{8}
\]

The null hypothesis of interest is that of double (positive) unit roots in the characteristic polynomial \( \Phi_2(z) = 1 - \sum_{j=1}^{2} \theta_j z^j \) of (8), that is under \( H_0 \) \( z_{1,2} = 1 \). This hypothesis can be tested using the profile GDL statistic \( \widehat{P}^d_\rho (\lambda, \theta_0, \hat{\eta}) \) with \( \theta_0 = \begin{bmatrix} 2 & -1 \end{bmatrix}' \). By Proposition 44

\[
2 \left( \widehat{P}^d_\rho (\lambda, \theta_0, \hat{\eta}) - \rho (0) \right) \xrightarrow{w} \left[ \int_0^1 \widehat{B}_m dB \int_0^1 B_m dB \right] F_m^{-1} \left[ \int_0^1 \widehat{B}_m dB \int_0^1 B_m dB \right]', \tag{9}
\]
where

\[
F_m = \begin{bmatrix}
\int_0^1 B_m^2 & \int_0^1 \overline{B}_m B_m \\
\int_0^1 \overline{B}_m B_m & \int_0^1 \overline{B}_m^2
\end{bmatrix},
\]

and \( B_m \) and \( \overline{B}_m \) are, respectively, a demeaned and an integrated demeaned Brownian motions.
2.2 Blocking and smoothing methods for $\alpha$ mixing processes

- The proof of GEL Theorem 15 shows that the GEL estimator $\hat{\theta}$ is still $n^{1/2}$-consistent and asymptotically normal even with weakly dependent observations, but is less efficient than the efficient GMM estimator. More importantly GEL test statistics are no longer asymptotically chi-squared distributed.

- One way to solve this problem is to consider blocking techniques as suggested by Kitamura (1997).

- Alternatively one can use kernel smoothing techniques as suggested by Kitamura & Stutzer (1997), Smith (1997) and Smith (2004), among others. We first briefly consider the latter. Let

$$g_{st}(\theta) = \sum_{s=t-n}^{n-1} K_h(s) g_{t-s}(\theta) / (nh)$$

denote a smoothed moment indicator where $K_h(\cdot)$ is a kernel function. The kernel weights $K_h(s)/h$ are in the same spirit of those used in the heteroskedasticity autocorrelation consistent (HAC) covariance matrix estimation; see for example Newey & West (1987) and Andrews (1991).

---

3 For the exactly identified case it is possible to show that the GEL test statistic for $H_0 : \theta = \theta_0$ has asymptotic distribution $\sum_{j=1}^{k} \omega_j \chi^2_{1,j}$ where the $\omega_j$ are the eigenvalues of $[E[g(\theta_0)g(\theta_0)^\top]]^{-1} \sum_{j=1}^{\infty} COV[g_1(\theta_0)g_j(\theta_0)]$. 

CAEPR mini-course in empirical likelihood methods
To be specific we can define a smoothed GEL criterion function as

$$\hat{P}_{\rho}^s (\theta, \lambda) := \sum_{i=1}^{n} \rho (k \lambda' g_{st}(\theta)) / n$$  \hspace{1cm} (10)

where $k$ is a normalisation constant. Smith (2004) shows that (10) can be used to obtain efficient estimates and asymptotically valid inference.

An alternative (asymptotically equivalent) approach is to consider blocking techniques, see for example Politis & Romano (1992).

The basic idea is to construct “new” observations by considering blocks of the original observations, and base estimation and inference on the resulting sequence of blocks.

This procedure preserves nonparametrically the dependent structure of the data, delivering therefore valid asymptotic inference. Let $l = l(n), m = m(n)$ such that $1 \leq l \leq m$, and $\lim_{n \to \infty} m = \infty$. For each $i \in \mathbb{N}$, let

$$b_{i,m,l} = \left[ z'_{(i-1)l+1} \hspace{0.5cm} \cdots \hspace{0.5cm} z'_{(i-1)l+m} \right]'$$

be a block of $m$ consecutive observations starting from $(i - 1)l + 1$. Note that $m$ is the block length and $l$ is the separation between block starting points. (If $l = m$ the resulting sequence of blocks is nonoverlapping, while if $l = 1$ it is fully overlapping).
Define now the blockwise moment function

\[ \psi(b_{i,m,l}, \theta) := \psi_i(\theta) = \sum_{j=1}^{m} g(z_{(i-1)l+j}, \theta) / m, \]

and note that \( E[\psi_i(\theta_0)] = 0 \ \forall i. \)

Blockwise GEL (BGEL henceforth) estimation and inference for \( \theta_0 \) is based on the BGEL criterion function

\[ \hat{P}_{\rho}^b(\theta, \lambda) := \sum_{i=1}^{q} \rho(\lambda' \psi_i(\theta)) / q \]

where \( q = \lfloor (n - m) / l + 1 \rfloor \) is the total number of blocks and \( \lfloor \cdot \rfloor \) is the integer part function.

**Theorem 5** Assume that (I) \( z_t \) is a strong mixing sequence of size \(-2\alpha/(\alpha - 2)\) where \( \alpha > 2 \), (II) The parameter space \( \Theta \) is compact, (III) (a) \( E\left( \sup_t \sup_{\theta \in \Theta} \| g_t(\theta) \|^2 \right) < \infty \) for some \( \delta > 0 \), (b) \( \lim_{n \to \infty} E[\tilde{g}(\theta)] \) exists uniformly in \( \Theta \) and is continuous, (IV) \( \lim_{n \to \infty} \sum_{t=1}^{n} E(\sup_{\theta \in \Theta} \| G_t(\theta) \| / n < \infty, \]

(V) \( \Omega(\theta) := \lim_{n \to \infty} V \left[ n^{1/2} \tilde{g}(\theta) \right] \) exists and it is positive definite uniformly in \( \Theta \), (VI) \( \rho(\cdot) \) is twice continuously differentiable in an open neighbourhood of \( 0 \), (VII) (a) \( G_t(\theta) \) is twice-continuously differentiable in a convex neighbourhood \( \mathcal{N} \) of \( \theta_0 \) \( \forall t \) (b) \( E(\sup_{\theta \in \mathcal{N}} (\| G_t(\theta) \|^\alpha + \| \partial G_t(\theta) / \partial \theta_j \|^{(\alpha-1)/\alpha}) < \infty \)

CAEPR mini-course in empirical likelihood methods
\( (j = 1, \ldots, k) \) \((c)\) \(\lim_{n \to \infty} E \left[ \partial^k \hat{G}(\theta) / \partial \theta_j \right] \) and \(\lim_{n \to \infty} E \left[ \sum_{t=1}^{n} \sum_{j=1}^{l} \partial G_t(\theta) / \partial \theta_j g_t(\beta) / n \right] \) exist uniformly in \( \Theta \) and are continuous, \( (iv) \) \(\lim_{n \to \infty} \sum_{t=1}^{n} E \sup_{\beta \in \mathcal{N}} \left\| \partial^k G_t(\theta) / \partial \theta_j \right\| / n < \infty, \) \((v)\) \(\text{rank } [G_0] = k \) where \(G_0 = \lim_{n \to \infty} E \left[ \hat{G}(\theta_0) \right] \), \((vi)\) \(G'(\beta_0)^T \Omega(\beta_0)^{-1} G(\beta_0) \) is nonsingular. Then for \(m = o \left( n^{1/2} \right) \)

\[
\begin{bmatrix}
\left( n^{1/2} / m \right) \hat{\lambda} \\
n^{1/2} \left( \hat{\theta} - \theta_0 \right)
\end{bmatrix} \xrightarrow{d} N \left( \begin{bmatrix}
\Psi_0 \\
0
\end{bmatrix}, \begin{bmatrix}
0 & 0 \\
0 & \left( G_0' \Omega_0^{-1} G_0 \right)^{-1}
\end{bmatrix} \right).
\]

**Proof.** The proof is similar to that of Theorem 15 and uses ULLN as in Andrews (1987) applied to \(\psi_i(\theta)\) and its first derivative, and LLN and CLT for strongly mixing processes to show that \(\left\| \hat{\psi}(\theta) \right\| = O_p \left( n^{-1/2} \right) \) and \(\left\| \hat{\lambda} \right\| = O_p \left( m/n^{1/2} \right) \). Then by mean value expansion

\[
m \left\| \sum - (\partial^k \psi_i(\hat{\theta}) / \partial \theta_j) \psi_i(\hat{\theta})' / q + \sum (\partial^k \psi_i(\theta_0) / \partial \theta_j) \psi_i(\theta_0)' / q \right\| \leq \\
2 \left( \sup_{\theta \in \mathcal{N}} \sum \left\| \partial^k g(\theta) / \partial \theta \right\|^\alpha / n \right)^{1/\alpha} \left( \sup_{\theta \in \mathcal{N}} \sum \left\| \partial^{k+1} g(\theta) / \partial \theta' \partial \theta_j \right\|^{(\alpha-1)/\alpha} / n \right)^{\alpha/(\alpha-1)} \times \\
\left( m/n^{1/2} \right) n^{1/2} \left\| \theta - \theta_0 \right\| + O_p \left( m^2/ln \right) = o_p \left( 1 \right) \text{ for } k = 0, 1,
\]
Empirical likelihood methods with application to econometrics

\[
\left\| \sum_{j=1}^{l} \psi_{ij}(\bar{\theta}) \frac{\partial \psi_i(\bar{\theta})}{\partial \theta'}/q - \frac{1}{m} \sum_{j=1}^{l} \sum g_{ij}(\theta_0) \frac{\partial g_i(\theta_0)}{\partial \theta'}/n \right\| = o_p(1)
\]

and therefore by CMT and

\[
\| \bar{\lambda} \| = o_p(1) \left\| \sum \rho_2 \left( \bar{\lambda}' \psi_i(\bar{\theta}) \right) \bar{\lambda} \psi_i(\bar{\theta})' \frac{\partial \psi_i(\bar{\theta})}{\partial \theta'}/q \right\| = o_p(1/m) .
\]

Furthermore,

\[
\left\| \sum \rho_1 \left( \bar{\lambda}' \psi_i(\bar{\theta}) \right) \frac{\partial \psi_i(\bar{\theta})}{\partial \theta'}/q + G(\theta_0) \right\| = o_p(1) ,
\]

\[
\left\| \sum \sum_{j=1}^{l} \rho_2 \left( \bar{\lambda}' \psi_i(\bar{\theta}) \right) \bar{\lambda} \frac{\partial^2 \psi_i(\bar{\theta})}{\partial \theta' \partial \theta_j}/q \right\| = o_p(1) ,
\]

and hence

\[
n^{-1} \begin{bmatrix} m \frac{\partial^2 W(\bar{\lambda}, \bar{\theta})}{\partial \lambda \partial \lambda'} & \frac{\partial^2 W(\bar{\lambda}, \bar{\theta})}{\partial \lambda \partial \theta'} \\ \frac{\partial^2 W(\bar{\lambda}, \bar{\theta})}{\partial \theta \partial \lambda'} & \frac{\partial^2 W(\bar{\lambda}, \bar{\theta})}{\partial \theta \partial \theta'} \end{bmatrix} \overset{p}{\rightarrow} \begin{bmatrix} \Omega_0 & G_0 \\ G_0' & 0 \end{bmatrix} = M_0 .
\]

Then

\[
n^{1/2} \begin{bmatrix} \hat{\lambda}'/m, (\hat{\theta} - \theta_0)' \end{bmatrix} = \begin{bmatrix} \Psi(\beta_0), N(\beta_0)' \end{bmatrix} n^{1/2} \hat{\psi}(\beta_0) + o_p(1) ,
\]

and the conclusion follows (see Bravo (2007) for more details).
We consider specifications tests based on additional moment conditions as originally developed by Newey (1985). Let $\beta = [\alpha', \theta']'$ where $\alpha$ is an $\mathbb{R}^p$-valued vector of additional parameters, and suppose that there exists an an $\mathbb{R}^s$-valued $(s \leq p)$ vector of functions $h(z_t, \beta) := h_t(\beta)$ satisfying

$$E[h_t(\beta_0)] = 0, \quad \forall t$$

(11)

The information contained in this additional set of moment conditions can naturally be incorporated into BGEL estimation. To be specific let $l_t(\beta) = [g_t(\theta)', h_t(\beta)']'$ denote the “augmented” moment indicator, and let $\psi_i^a(\beta) = \sum_{j=1}^m l_{(i-1)+j}(\beta) / m$ denote the blockwise version of $l(\cdot)$. Finally let

$$\widehat{P}^b_\rho(\beta, \lambda, \varphi) = \sum_{i=1}^q \rho(\mu' \psi_i^a(\beta)) / q$$

where $\mu = [\lambda', \varphi']'$ and $\varphi$ is an $\mathbb{R}^s$-valued vector of unknown auxiliary parameters associated with $h_t(\beta)$.

Test statistics for the additional moment conditions (11) may be constructed by imposing the restriction
\( \varphi = 0 \) into the estimation of \( \hat{P}^b(\beta, \mu) \).

\[
D^p = 2c_n \left( \hat{P}^b_{\varphi} \left( \hat{\beta}, \hat{\mu} \right) - \hat{P}^b_{\mu} \left( \hat{\beta}^c, \hat{\mu}^c \right) \right),
\]

\[
LM^p = (n/m^2) (\varphi^c) \left[ S_{\varphi} \Delta \left( \hat{\beta}^c \right) S'_{\varphi} \right]^{-1} \varphi^c,
\]

\[
S^p = \sum_{i=1}^{q} s_i \left( \hat{\beta} \right)' / q^{1/2} S_{\varphi} \Delta \left( \hat{\beta} \right) S'_{\varphi} \sum_{i=1}^{q} s_i \left( \hat{\beta} \right) / q^{1/2}
\]

where \( c_n = (q/mn) \) is a correction factor that account for the overlap in the blocks,

\[
\Delta (\beta) = \Xi (\beta)^{-1} \left( I - L (\beta) \Gamma (\beta)^{-1} L (\beta)' \Xi (\beta)^{-1} \right)
\]

\[
s_i (\beta) = \rho_1 \left( \hat{\lambda}' \psi_i (\beta) \right) \psi_i^o (\beta),
\]

and \( S_{\varphi} = [0, I] \) is a selection matrix such that \( S_{\varphi} \mu = \varphi \). It is then possible

**Theorem 6** Assume that the assumptions of Theorem 5 hold for \( h(\beta) \) with parameter space \( B = A \times \Theta \) assumed compact. Then under (11)

\[
D^p, LM^p, MC^p, S^p \xrightarrow{d} \chi^2 (s).
\]

**Proof.** See Bravo (2007) for details. ■

CAEPR mini-course in empirical likelihood methods
Remark 1  The score statistic $S^p$ can be interpreted as the blockwise version of Newey’s (1985) specification test. Some algebra shows that

$$S^p = n \left[ \hat{l}(\theta_0)' \Delta (\theta_0) \hat{l}(\theta_0) - \hat{g}(\beta_0)' \Psi (\beta_0) \hat{g}(\beta_0) \right] + o_p(1),$$

which is in fact the asymptotic approximation of

$$n \left[ \hat{l}(\hat{\theta})' \hat{\Xi}(\hat{\theta})^{-1} \hat{l}(\hat{\theta}) - \hat{g}(\hat{\beta})' \hat{\Omega}(\hat{\beta})^{-1} \hat{g}(\hat{\beta}) \right]$$

i.e. the difference of two quadratic forms in the estimated sample moments as suggested by Newey (1985).
Example 3  We consider as in Andersen & Sorensen (1996) GMM estimation of the stochastic volatility model

\[
y_t = \sigma_t u_t \\
\ln \sigma_t^2 = \beta_1 + \beta_2 \ln \sigma_{t-1}^2 + \beta_3 \varepsilon_t
\]

where \([u, \varepsilon_t]' \sim N(0, I)\). The moment condition is

\[
E \left[ \begin{array}{c}
|y_t| - (2/\pi)^{1/2} \exp \left( \frac{\mu}{2} + \frac{\sigma^2}{8} \right) \\
y_t^2 - \exp \left( \frac{\mu + \sigma^2}{2} \right) \\
|y_t y_{t-j}| - (2/\pi) \exp \left( \frac{\mu + \sigma^2}{2} \right) \exp \left( \frac{\beta_1^2 \sigma^2}{4} \right)
\end{array} \right] = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}
\]

for \(j = 1, 3, 5\), where \(\mu = \beta_1 / (1 - \beta_2)\) and \(\sigma^2 = \beta_3^2 / (1 - \beta_2^2)\).
### Empirical size for lognormal stochastic volatility model with $n = 1000$

<table>
<thead>
<tr>
<th>$m^{(a)}$</th>
<th>$\alpha$</th>
<th>$J$</th>
<th>$D^{ET}$</th>
<th>$S^{ET}$</th>
<th>$LM^{ET}$</th>
<th>$LM^{ET}_{\hat{\pi}_i}$</th>
<th>$D^{CU}$</th>
<th>$LM^{CU}$</th>
<th>$LM^{CU}_{\hat{\pi}_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>.100</td>
<td>.214</td>
<td>.146$^\dagger$</td>
<td>.168</td>
<td>.241</td>
<td>.160</td>
<td>.222</td>
<td>.240</td>
<td>.138</td>
</tr>
<tr>
<td></td>
<td>.050</td>
<td>.139</td>
<td>.099</td>
<td>.064$^\dagger$</td>
<td>.198</td>
<td>.108</td>
<td>.146</td>
<td>.157</td>
<td>.086</td>
</tr>
<tr>
<td></td>
<td>.010</td>
<td>.048</td>
<td>.046</td>
<td>.031$^\dagger$</td>
<td>.136</td>
<td>.051</td>
<td>.114</td>
<td>.122</td>
<td>.059</td>
</tr>
<tr>
<td>12</td>
<td>.100</td>
<td>.187</td>
<td>.124$^\dagger$</td>
<td>.152</td>
<td>.203</td>
<td>.149</td>
<td>.209</td>
<td>.222</td>
<td>.129</td>
</tr>
<tr>
<td></td>
<td>.050</td>
<td>.119</td>
<td>.084</td>
<td>.058$^\dagger$</td>
<td>.176</td>
<td>.100</td>
<td>.137</td>
<td>.145</td>
<td>.081</td>
</tr>
<tr>
<td></td>
<td>.010</td>
<td>.041</td>
<td>.039</td>
<td>.021$^\dagger$</td>
<td>.115</td>
<td>.048</td>
<td>.107</td>
<td>.113</td>
<td>.055</td>
</tr>
<tr>
<td>$m^{*(c)}$</td>
<td>.100</td>
<td>.181</td>
<td>.119$^\dagger$</td>
<td>.148</td>
<td>.223</td>
<td>.144</td>
<td>.192</td>
<td>.215</td>
<td>.125</td>
</tr>
<tr>
<td></td>
<td>.050</td>
<td>.115</td>
<td>.080</td>
<td>.056$^\dagger$</td>
<td>.193</td>
<td>.097</td>
<td>.126</td>
<td>.140</td>
<td>.078</td>
</tr>
<tr>
<td></td>
<td>.010</td>
<td>.039</td>
<td>.037</td>
<td>.019$^\dagger$</td>
<td>.126</td>
<td>.046</td>
<td>.098</td>
<td>.109</td>
<td>.053</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$m^{(a)}$</th>
<th>$J^f$</th>
<th>$D^{ET,f}$</th>
<th>$S^{ET,f}$</th>
<th>$LM^{ET,f}$</th>
<th>$LM^{ET,f}_{\hat{\pi}_i}$</th>
<th>$D^{CU,f}$</th>
<th>$LM^{CU,f}$</th>
<th>$LM^{CU,f}_{\hat{\pi}_i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>.100</td>
<td>.173</td>
<td>.124$^\dagger$</td>
<td>.147</td>
<td>.187</td>
<td>.137</td>
<td>.173</td>
<td>.189</td>
</tr>
<tr>
<td></td>
<td>.050</td>
<td>.112</td>
<td>.084</td>
<td>.056$^\dagger$</td>
<td>.154</td>
<td>.092</td>
<td>.113</td>
<td>.124</td>
</tr>
<tr>
<td></td>
<td>.010</td>
<td>.038</td>
<td>.039</td>
<td>.009</td>
<td>.106</td>
<td>.043</td>
<td>.088</td>
<td>.096</td>
</tr>
<tr>
<td>12</td>
<td>.100</td>
<td>.157</td>
<td>.109$^\dagger$</td>
<td>.127</td>
<td>.160</td>
<td>.125</td>
<td>.169</td>
<td>.177</td>
</tr>
<tr>
<td></td>
<td>.050</td>
<td>.099</td>
<td>.073</td>
<td>.048$^\dagger$</td>
<td>.139</td>
<td>.084</td>
<td>.110</td>
<td>.116</td>
</tr>
<tr>
<td></td>
<td>.010</td>
<td>.034</td>
<td>.034</td>
<td>.008$^\dagger$</td>
<td>.090</td>
<td>.040</td>
<td>.086</td>
<td>.090</td>
</tr>
<tr>
<td>$m^{*(c)}$</td>
<td>.100</td>
<td>.148</td>
<td>.108$^\dagger$</td>
<td>.119$^\dagger$</td>
<td>.178</td>
<td>.120</td>
<td>.153</td>
<td>.176</td>
</tr>
<tr>
<td></td>
<td>.050</td>
<td>.098</td>
<td>.074</td>
<td>.045$^\dagger$</td>
<td>.154</td>
<td>.083</td>
<td>.100</td>
<td>.114</td>
</tr>
<tr>
<td></td>
<td>.010</td>
<td>.031</td>
<td>.033</td>
<td>.007$^\dagger$</td>
<td>.100</td>
<td>.033</td>
<td>.078</td>
<td>.089</td>
</tr>
</tbody>
</table>

(a) Lag-length parameter, (b) Calculated with implied probabilities $\hat{\pi}_i$ as in (??), (c) Optimally chosen lag-length parameter using Newey & West (1994). A dagger in each row indicates closest value to the nominal level.

CAEPR mini-course on empirical likelihood methods.
The following table reports the empirical powers of the test statistics considered in the previous table with \( m \) chosen by Newey & West's (1994) nonparametric method. The power of each test is calculated at 9 points in the interval \( \beta_1 = [-.768, 0.068] \) using 1000 replications and Monte Carlo size corrected critical values.

### Empirical power for lognormal stochastic volatility model \( n = 1000 \)

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>-.768</th>
<th>-.668</th>
<th>-.568</th>
<th>-.468</th>
<th>-.368</th>
<th>-.268</th>
<th>-.168</th>
<th>-.068</th>
<th>.068</th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>.504</td>
<td>.302</td>
<td>.148</td>
<td>.104</td>
<td>0</td>
<td>.054</td>
<td>.072</td>
<td>.124</td>
<td>.154</td>
</tr>
<tr>
<td>ET</td>
<td>.574</td>
<td>.376(^\dagger)</td>
<td>.211</td>
<td>.128</td>
<td>0</td>
<td>.025</td>
<td>.048</td>
<td>.147</td>
<td>.185</td>
</tr>
<tr>
<td>( S )( ET )</td>
<td>.611(^\dagger)</td>
<td>.358</td>
<td>.220</td>
<td>.134</td>
<td>0</td>
<td>.049</td>
<td>.062</td>
<td>.094</td>
<td>.134</td>
</tr>
<tr>
<td>ET( (a) )</td>
<td>.434</td>
<td>.240</td>
<td>.130</td>
<td>.103</td>
<td>0</td>
<td>.075(^\dagger)</td>
<td>.085(^\dagger)</td>
<td>.155(^\dagger)</td>
<td>.175</td>
</tr>
<tr>
<td>LM( \pi_i )( CU )</td>
<td>.562</td>
<td>.364</td>
<td>.234(^\dagger)</td>
<td>.136(^\dagger)</td>
<td>0</td>
<td>.024</td>
<td>.038</td>
<td>.062</td>
<td>.105</td>
</tr>
<tr>
<td>LM( \pi_i )( CU (a) )</td>
<td>.449</td>
<td>.261</td>
<td>.154</td>
<td>.117</td>
<td>0</td>
<td>.039</td>
<td>.071</td>
<td>.155(^\dagger)</td>
<td>.188(^\dagger)</td>
</tr>
<tr>
<td>J( f )</td>
<td>.367</td>
<td>.214</td>
<td>.111</td>
<td>.073</td>
<td>0</td>
<td>.004</td>
<td>.054</td>
<td>.094</td>
<td>.121</td>
</tr>
<tr>
<td>ET( f, a )</td>
<td>.419</td>
<td>.266</td>
<td>.158</td>
<td>.090</td>
<td>0</td>
<td>.002</td>
<td>.063(^\dagger)</td>
<td>.121(^\dagger)</td>
<td>.149(^\dagger)</td>
</tr>
<tr>
<td>ET( s, f )</td>
<td>.446(^\dagger)</td>
<td>.254</td>
<td>.165</td>
<td>.096</td>
<td>0</td>
<td>.042(^\dagger)</td>
<td>.028</td>
<td>.071</td>
<td>.106</td>
</tr>
<tr>
<td>ET( (a), f )</td>
<td>.316</td>
<td>.170</td>
<td>.097</td>
<td>.083</td>
<td>0</td>
<td>.006</td>
<td>.053</td>
<td>.117</td>
<td>.138</td>
</tr>
<tr>
<td>LM( \pi_i )( CU, f )</td>
<td>.410</td>
<td>.258(^\dagger)</td>
<td>.175(^\dagger)</td>
<td>.123(^\dagger)</td>
<td>0</td>
<td>.008</td>
<td>.024</td>
<td>.042</td>
<td>.082</td>
</tr>
<tr>
<td>LM( \pi_i )( CU (a), f )</td>
<td>.327</td>
<td>.185</td>
<td>.115</td>
<td>.084</td>
<td>0</td>
<td>.012</td>
<td>.051</td>
<td>.118</td>
<td>.148</td>
</tr>
</tbody>
</table>

\((a)\) Calculated with implied probabilities \( \pi_i \) as in (??)

CAEPR mini-course in empirical likelihood methods


Helland, I. (1982), ‘Central limit theorems for martingales with discrete or continuous time’, *Scandina-


