Empirical likelihood methods with application to econometrics

Francesco Bravo
Lecture 5

- Non differentiable moment indicators
- Infinite dimensional moment conditions models
1 Non differentiable moment indicators

Standard moment conditions model but the moment indicator $g$ is not differentiable.

Example 1  Consider a quantile regression model with endogenous regressors. Suppose that there exists a vector of instruments $w_i$ such that

$$E [g_i (\theta_0)] = E \{w_i [I \{y_i \leq x_i' \theta_0\} - q]\} = 0 \ a.s.$$
Theorem 1 (Parente and Smith (2005)) Assume that (I) the parameter space $\Theta$ is a compact set, (II) $g(\theta)$ is continuous at each $\theta$ a.s., (III) $E\left[\sup_{\theta \in \Theta} |g(\theta)|^{\alpha}\right] < \infty$ for some $\alpha > 2$, (IV) $\rho(v)$ is twice continuously differentiable in a neighbourhood of 0, (V) $\Omega_0$ is p.d., (VI) $\theta_0 \in \text{int}\{\Theta\}$, (VII) $E[g(\theta)]$ is differentiable at $\theta_0$ with derivative matrix $G_0 = \partial E[g(\theta)]/\partial \theta'$ of full column rank, (VIII) for any $\delta_n \to 0$ $\sup_{\|\theta - \theta_0\| \leq \delta_n} \|\hat{g}(\theta) - E[g(\theta)] - \hat{g}(\theta_0)\| = o_p(n^{-1/2})$, (IX) (a) $\sup_{\theta \in N_0} \|G(\theta)\| < \infty$ and (b) $G_0'\Omega_0^{-1}G_0$ is nonsingular. Then

\[
n_{1/2} \left[ \begin{array}{c} \hat{\lambda} \\ \hat{\theta} - \theta_0 \end{array} \right] \xrightarrow{d} N \left( \left[ \begin{array}{c} 0 \\ 0 \end{array} \right], \left[ \begin{array}{cc} \Psi_0 & 0 \\ 0 & (G_0'\Omega_0^{-1}G_0)^{-1} \end{array} \right] \right),
\]

\[
2n \left( \hat{P}_\rho(\hat{\theta}, \hat{\lambda}) - \hat{P}_\rho(0) \right) \xrightarrow{d} \chi^2_{l-k}
\]

Proof. Consistency can be proved in the usual way. The asymptotic normality can be shown by replacing the (non differentiable) objective function $\hat{P}_\rho(\theta, \lambda)$ with a smooth well behaved (albeit infeasible) one, given by

\[
\hat{P}_\rho^s(\theta, \lambda) = [-G_0(\theta - \theta_0)]'\lambda - \hat{g}(\theta_0)'\lambda - \lambda'\Omega_0\lambda/2
\]

and define $[\tilde{\lambda}', \tilde{\theta}']'$ as $\min_{\theta} \sup_{\lambda} \hat{P}_\rho^s(\theta, \lambda)$. The distribution of these estimators follows easily from the
first order conditions using CLT and CMT. Following (Pakes & Pollard 1989) we show that

\begin{align*}
(1) \quad n^{1/2} \left( \hat{\theta} - \theta_0 \right) &= O_p(1), \\
(2)(a) \quad \hat{P}_\rho (\theta, \lambda) &= \hat{P}_\rho^s (\theta, \lambda) + o_p(1) \quad \text{uniformly}, \\
(2)(b) \quad \left\| \left[ \hat{\lambda}', \hat{\theta}' \right]' - \left[ \lambda', \theta' \right]' \right\| &= o_p \left( n^{-1/2} \right).
\end{align*}

These conditions are sufficient to establish the asymptotic normality of $\hat{\theta} - \theta_0$ given the asymptotic normality of $\tilde{\theta} - \theta_0$. To show (1) note that (VIII) implies that $E \left[ g \left( \tilde{\theta} \right) \right] = O_p \left( n^{-1/2} \right)$ while (VII) implies that $\| E \left[ g \left( \theta \right) \right] \| \geq C \| \theta - \theta_0 \|$ from which (1) follows evaluating the latter expression at $\hat{\theta}$. To
show (2)(a) note that by Taylor expansion about 0, T, (VII) and (VIII)

\[
\begin{align*}
| \hat{P}_\rho (\hat{\theta}, \hat{\lambda}) - [-G_0 (\hat{\theta} - \theta_0)]' \hat{\lambda} - \hat{g} (\theta_0)' \hat{\lambda} - \hat{\lambda}' \Omega_0 \hat{\lambda}/2 | & \leq \\
\left| \left( \hat{g} (\hat{\theta}) - \hat{g} (\theta_0) + [G_0 (\hat{\theta} - \theta_0)]' \right)' \hat{\lambda} \right| + \\
\hat{\lambda}' \left( \sum_{i=1}^n \rho_2 \left( \hat{\lambda}' g_i (\hat{\theta}) \right) g_i (\hat{\theta}) g_i (\hat{\theta})' / n + \Omega_0 \right) \hat{\lambda} \right| & \leq \\
\| \hat{\lambda} \| \left( \| \hat{g} (\hat{\theta}) - \hat{g} (\theta_0) \| + \| G_0 \| \| (\hat{\theta} - \theta_0) \| \right) + \| \hat{\lambda} \|^2 \times \\
\sum_{i=1}^n \rho_2 \left( \hat{\lambda}' g_i (\hat{\theta}) \right) g_i (\hat{\theta}) g_i (\hat{\theta})' / n + \Omega_0 \right| & = o_p (n^{-1}).
\end{align*}
\]

By CLT \( n^{1/2} (\hat{\theta} - \theta_0) = O_p (1) \); moreover by (VII) and (VIII)

\[
\| E \left[ g (\hat{\theta}) \right] \| \leq \| G_0 \| \| (\hat{\theta} - \theta_0) \| + o \left( \| (\hat{\theta} - \theta_0) \| \right)
\]

and

\[
\| \hat{g} (\hat{\theta}) \| \leq o_p \left( n^{-1/2} \right) + \| E \left[ g (\hat{\theta}) \right] \| = O_p \left( n^{-1/2} \right).
\]
Thus as in (1) \( |\hat{P}_\rho(\tilde{\theta}, \tilde{\lambda}) - \hat{P}_\rho^s(\tilde{\theta}, \tilde{\lambda})| = o_p(n^{-1}) \). Finally (2)(b) follows noting that by definition of \( \tilde{\theta} \) and \( \tilde{\theta} \)
\[
\hat{P}_\rho^s(\tilde{\theta}, \tilde{\lambda}) - o_p(n^{-1}) \leq \hat{P}_\rho(\tilde{\theta}, \tilde{\lambda}) \leq \hat{P}_\rho^s(\tilde{\theta}, \tilde{\lambda}) + o_p(n^{-1})
\]
which implies that \( \hat{P}_\rho^s(\tilde{\theta}, \tilde{\lambda}) = \hat{P}_\rho^s(\tilde{\theta}, \tilde{\lambda}) + o_p(n^{-1}) \). Then as in Pakes & Pollard (1989) an expansion about \( \tilde{\theta} \) using (IX) (a) gives
\[
\hat{P}_\rho^s(\tilde{\theta}, \tilde{\lambda}) - \hat{P}_\rho^s(\tilde{\theta}, \tilde{\lambda}) = [G_0(\tilde{\theta} - \tilde{\theta})]\tilde{\lambda} = o_p(n^{-1})
\]
which implies that \( ||\tilde{\theta} - \tilde{\theta}|| = o_p(n^{-1/2}) \) by VII. A similar arguments shows that \( ||\tilde{\lambda} - \tilde{\lambda}|| = o_p(n^{-1/2}) \)
The proof of the second result is standard.
2 Conditional moment conditions models

We consider how EL methods can be used in the context of (efficient) estimation and inference in conditional moment conditions. Let \((z_i, x_i)_{i=1}^n\) be i.i.d. on a data vector \((z, x)\) and assume that there exists a function \(g(z, \theta) : \mathbb{R}^l \rightarrow \mathbb{R}^k (l \geq k)\)

\[
E[g(z, \theta_0) | x] = 0.
\]  

(2)

By the LIE of conditional expectation (2) implies

\[
E[A(x, \theta_0) g(z, \theta_0)] = 0
\]  

(3)

where \(A(\cdot)\) is a matrix of instrumental variables. An interesting question is to find an \(A(\cdot)\) that yields an asymptotically efficient estimator of \(\theta_0\). By the results of Chamberlain (1987) it is well-known that

\[
A(x, \theta_0) = E[G_0' | x] \left( E \left[ g(z, \theta_0) g(z, \theta_0)' | x \right] \right)^{-1}
\]

achieves the semiparametric efficiency bound.
2.1 Unconditional approach

Newey (1993) showed that the asymptotic variance of the optimal GMM estimator based on $g(z, \theta_0) \otimes q(x)$, where $q(\cdot)$ is an approximating function corresponds to a minimum mean squared error of $A(x, \theta_0) g(z, \theta_0)$ by linear combinations of $g(z, \theta_0) \otimes q(x)$.

Thus as the dimension of $q(\cdot)$ grows with the sample size (at an appropriate rate) the variance of GMM approaches the semiparametric bound.

There are many possible choices for the approximating function $q(\cdot)$ including splines, power and Fourier series.

Thus we can approximate (3) with

$$E [g(z, \theta_0) \otimes q(x)] = 0$$

and base EL estimation on this (infinite dimensional) set of moment conditions, that is

$$\hat{\theta} = \max_{\theta \in \Theta} \sup_{\lambda} - \sum_{i=1}^{n} \log \left(1 + \lambda' g_i(\theta) \otimes q(x) \right)$$

More generally we can defined GEL estimator as

$$\hat{\theta} = \min_{\theta \in \Theta} \sup_{\lambda} \hat{P}_\rho (g_i(\theta) \otimes q(x), \lambda).$$
Theorem 2 (Donald, Imbens and Newey (2003)) Assume that (I) $E \left[ q(x)q(x) \right]$ is finite and for any $A(x)$ such that $E \|A(x)\|^2 < \infty$, $E \left[ (A(x) - q(x)' \gamma)^2 \right] \to 0$, (II) there is a $\zeta(m)$ such that for each $m = m(n)$ there is a nonsingular matrix $B$ such that $\bar{q}(x) = Bp(x)$ and $\sup_{x \in X} \|q(x)\| \leq \zeta(m)$, $\sigma_{\min} \left( E \left[ q(x)q(x)' \right] \right) > 0$ and $m^{1/2} \leq \zeta(m)$, (III) (i) $E [g(z, \theta_0) | x] = 0$ for a unique $\theta_0$, (ii) $\Theta$ compact, (iii) $E \left( \sup_{\theta \in \Theta} \|g(\theta) \|^{2} \right) \leq \zeta(m)$, (iv) for all $\theta, \theta' \in \Theta$ $\|g(\theta) - g(\theta')\| \leq \delta(z) \|\theta - \theta'\|^\alpha$ for $\alpha > 0$ and $E [\delta(z)]^2 < \infty$, (v) $\zeta(m)^2 m^2/n \to 0$. Then
\[ n^{1/2} \left( \hat{\theta} - \theta_0 \right) \sim N \left( 0, \left( G_0' \Omega_0^{-1} G_0 \right)^{-1} \right), \]
where $G_0 = E \left[ \partial g(\theta_0)/\partial \theta' | x \right]$ and $\Omega_0 = E \left[ g(\theta_0)g(\theta_0)' | x \right]$.

Proof. Let $\delta_n = o \left( n^{-1/\zeta(m)} \right)$ and $\Lambda_n = \{ \lambda : \|\lambda\| \leq \delta_n \}$. Then
\[ \sup_{\theta \in \Theta, \lambda \in \Lambda_n} \max_i \|X g_i(\theta)\| \leq \delta_n \sup_{\theta \in \Theta} \|g_i(\theta)\| \zeta(m) = o_p(1) \]
and $\Lambda_n \subseteq \hat{\Lambda}_n(\theta)$ w.p.a.1. Then as in the proof of Theorem 15 for any $\bar{\theta} = \theta_0 + o_p(m)$ and $\|\hat{g}(\bar{\theta})\| = O_p \left( (m/n)^{1/2} \right)$ Taylor expansion $\sup_{\lambda \in \hat{\Lambda}_n(\bar{\theta})} \hat{P}_p(\bar{\theta}, \lambda) \leq \rho_0 + O_p(m/n)$ and $\|\hat{\lambda}\| \leq O_p \left( m/n^{1/2} \right)$. We next show that $\|\hat{g}(\bar{\theta})\| = O_p \left( m/n^{1/2} \right)$ where $\hat{\theta} = \arg \min_{\theta \in \Theta} \hat{P}(\lambda, \theta)$ for $\lambda \in \Lambda_n$. Let $\tilde{\lambda} = -\delta_n \hat{g}(\bar{\theta}) / \|\hat{g}(\bar{\theta})\|$.
and note that
\[ \hat{P}_p \left( \hat{\theta}, \hat{\lambda} \right) \geq \rho_0 + \delta_n \left\| \hat{g} \left( \hat{\theta} \right) \right\| - C\delta_n^2 \]
and since \( m / (n \delta_n) = o \left( (m/n)^{1/2} \right) \left\| \hat{g} \left( \hat{\theta} \right) \right\| \leq O_p \left( (m/n)^{1/2} \right) \). The consistency of \( \hat{\theta} \) follows by noting that w.p.a. 1
\[ \hat{P}_p \left( \hat{\theta}, \hat{\lambda} \right) = \hat{g} \left( \hat{\theta} \right)' \hat{\Omega} \left( \hat{\theta} \right)^{-1} \hat{g} \left( \hat{\theta} \right) \to 0, \]
continuity of \( E \left[ g \left( \theta \right)' \right] \left( \hat{\theta} \right)^{-1} E \left[ g \left( \theta \right)' \right] \) and ULLN for \( \hat{g} \left( \theta \right)' \hat{\Omega} \left( \theta \right)^{-1} \hat{g} \left( \theta \right) \). The latter two follows noting that
\[ \left| E \left[ g \left( \hat{\theta} \right)' \right] \left( \hat{\theta} \right)^{-1} E \left[ g \left( \hat{\theta} \right) \right] - E \left[ g \left( \theta \right)' \right] \left( \hat{\theta} \right)^{-1} E \left[ g \left( \theta \right) \right] \right| \leq E [\delta \left( z \right)]^2 \left\| \hat{\theta} - \theta \right\|^{2\alpha}, \]
\( \hat{g} \left( \theta \right)' \hat{\Omega} \left( \theta \right)^{-1} \hat{g} \left( \theta \right) \) satisfies a LLN and
\[ \left| \hat{g} \left( \hat{\theta} \right)' \hat{\Omega} \left( \hat{\theta} \right)^{-1} \hat{g} \left( \hat{\theta} \right) - \hat{g} \left( \theta \right)' \hat{\Omega} \left( \theta \right)^{-1} \hat{g} \left( \theta \right) \right| \leq \max_{l,k} \left| \hat{\Omega} \left( \theta \right)^{-1} \right|^2 \left\| \hat{\theta} - \theta \right\|^{2\alpha} = O_p \left( 1 \right) \]
asymptotic normality follows by mean value expansion and CLT, noting that
\[
\left\| \sum_{i=1}^{n} \rho_2 \left( \hat{\lambda}' g_i \left( \hat{\theta} \right) \right) g_i \left( \hat{\theta} \right) g_i \left( \hat{\theta} \right)' / n - \Omega_0 \right\| = O_p \left( \zeta (m) \left( m^2 / n \right)^{1/2} \right),
\]
\[
\left\| \sum_{i=1}^{n} \rho_1 \left( \hat{\lambda}' g_i \left( \hat{\theta} \right) \right) G_i \left( \hat{\theta} \right) / n - G_0 \right\| = O_p \left( \left( m^2 / n \right)^{1/2} \right).
\]

Remark 1: The rate conditions correspond to \( m^3 / n \to 0 \) for splines and \( m^4 / n \to 0 \) for power series. For GMM estimator the rate conditions are slightly weaker (i.e. \( m^2 / n \to 0 \) for splines and \( m^3 / n \to 0 \) for power series).

We now briefly consider consistent conditional moment tests. The GEL statistic has the usual form
\[
2n \left( \hat{P}_\rho \left( \hat{\theta}, \hat{\lambda} \right) - \hat{P}_\rho (0) \right).
\]

If \( m \) was fixed then the asymptotic distribution would be \( \chi^2_{m-k} \). However \( m \to \infty \) so the statistic we consider is a standardised version, as suggested for example by DeJong & Bierens (1994).
Theorem 3 Under the same assumptions of Theorem 54 with (v) strengthened to \( \zeta(m)^2 m^3/n \to 0 \) then

\[
\left[ 2n \left( \hat{P}_\rho \left( \hat{\theta}, \hat{\lambda} \right) - \hat{P}_\rho (0) \right) - (lm - k) \right] / (2(lm - k)) \xrightarrow{d} N(0, 1).
\]

Proof. By the usual arguments

\[
2n \left( \hat{P}_\rho \left( \hat{\theta}, \hat{\lambda} \right) - \hat{P}_\rho (0) \right) = n\hat{g} \left( \hat{\theta} \right) ' \hat{\Omega} \left( \hat{\theta} \right)^{-1} \hat{g} \left( \hat{\theta} \right) + o_p(1)
\]

and the conclusion follows by noting that the following conditions (I) \( \left\| \hat{\Omega} \left( \theta \right) - \Omega_0 \right\| = o_p \left( (lm)^{-1/2} \right) \), (II) \( \sigma_{\min} (\Omega) > 0 \), (III) \( \left\| \hat{G} \left( \hat{\theta} \right) - G_0 \right\| \xrightarrow{p} 0 \), (IV) \( G_0' \Omega_0^{-1} G_0 \) is bounded, (V) \( E \left[ g \left( \theta_0 \right)' \Omega_0 g \left( \theta_0 \right) \right] ^2 / (lmn^{1/2}) \to 0 \) are satisfied. These conditions imply that

\[
n \left( \hat{g} \left( \hat{\theta} \right) ' \hat{\Omega} \left( \hat{\theta} \right)^{-1} \hat{g} \left( \hat{\theta} \right) - n\hat{g} \left( \theta_0 \right)' \Omega_0^{-1} \hat{g} \left( \theta_0 \right) \right) / (2(lm - k)) \xrightarrow{p} 0
\]

and that

\[
\left( n\hat{g} \left( \theta_0 \right)' \Omega_0^{-1} \hat{g} \left( \theta_0 \right) - (lm - k) \right) / (2(lm - k)) \xrightarrow{d} N(0, 1)
\]

by Theorem 1 of DeJong & Bierens (1994).
2.2 Conditional (local) approach

An alternative approach is to use the so-called conditional or local EL proposed by Kitamura, Tripathi & Ahn (2004). This method avoids the explicit estimation of the matrix $A(x, \theta_0)$ of optimal instruments by using a local (i.e. kernel based) EL to estimate the conditional moment restriction itself.

Let

$$w_{ij} = K_h(x_i - x_j) / \sum_{j=1}^{n} K_h(x_i - x_j)$$

denote a set of positive weights used to carry out the localisation, where $K_h(\cdot)$ is a kernel function, and let $\pi_{ij} = \Pr(z = z_j|x = x_i)$ denote the conditional probability supported on the observed sample $(x_i)_{i=1}^{n} \times (z_j)_{j=1}^{n}$. Then by (2)

$$\max_{\pi_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \log \pi_{ij} \quad \text{s.t.} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} \pi_{ij} = 1 \quad \text{and} \quad \sum_{j=1}^{n} \pi_{ij} g(z_j, \theta_0) = 0,$$

whose solution is

$$\hat{\pi}_{ij} = w_{ij} / \left( 1 + \hat{\lambda}'_i g(z_j, \theta_0) \right) \quad i = 1, ..., n$$
where the Lagrange multipliers $\hat{\lambda}_i$ satisfy

$$\sum_{j=1}^{n} \pi_{ij} g(z_j, \theta_0) / \left(1 + \hat{\lambda}_i g(z_j, \theta_0)\right) = 0.$$ 

We can then define the maximum conditional EL (CEL) estimator for $\theta_0$ as

$$\hat{\theta} = \arg \max_{\theta \in \Theta} CEL(\theta) \quad (4)$$

where

$$CEL(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{n} T_{i,n} w_{ij} \log \left( w_{ij} / \left(1 + \hat{\lambda}_i g(z_j, \theta)\right) \right),$$

$$T_{i,n} = I \left\{ \sum_{j=1}^{n} K_h (x_i - x_j) / (nh) \geq h^\tau \right\} \text{ with } \tau \in (0, 1)$$

is a sequence of trimming functions included to avoid the well-known denominator problem of kernel estimators.

**Theorem 4 (Kitamura, Tripathi and Ahn (2004))** Assume that (I) $E[g(z, \theta) | x] \neq 0$ for each $\theta \neq \theta_0$ for every $x \in \chi_\theta$ such that $\Pr(x \in \chi_\theta)$, (II) $E[\sup_{\theta \in \Theta} \|g(z, \theta)\|^{\alpha}] < \infty$, $\alpha \geq 8$, (III) $K$ is a symmetric continuously differentiable pdf with support $[-1, 1]$ and bounded away from 0 on $[-a, a]$ for some $a \in (0, 1)$, (IV) $0 < l(x) \leq \sup_{x \in \mathbb{R}} l(x) < \infty$, $l(x)$ is twice continuously differentiable with $\sup_x \|\partial^2 l(x) / \partial x \partial x'\| < \infty$, where $l(x)$ is the Lebesgue density of $x$, (V) $E \|x\|^{1+\beta} < \infty$ for some $\beta > 0$, (VI) $E \left[\sup_{\theta \in \Theta} \|\partial g(z, \theta) / \partial \theta'\|\right] < \infty$.
Empirical likelihood methods with application to econometrics

\[ \sup_{\theta \in \Theta} \sup_{x \in \mathbb{R}^r} \left| \frac{\partial^2 E \left[ g^j (z, \theta) \mid x \right]}{\partial x \partial x'} \right| < \infty \text{ for } j = 1, \ldots, l, \]

\[ E \sup_{\theta \in \Theta} \left[ \frac{G (\theta)}{\partial \theta'} \mid x \right] \text{ are continuous,} \]

\[ \inf_{\xi, x, \theta \in \Theta} \xi' E \left[ \Omega (\theta) \mid x \right] \xi > 0, \sup_{\xi, x, \theta \in \Theta} \xi' E \left[ \Omega (\theta) \mid x \right] \xi < \infty, \]

\[ E \left[ \sup_{\theta \in \Theta} \left| G (\theta) \right|^\eta \right] < \infty \eta \geq 4, E \left[ \sup_{\theta \in \Theta} \left| \frac{\partial G (\theta)}{\partial \theta^j} \right| \right] < \infty \ j = 1, \ldots, k, \]

\[ \sup_{x, \theta \in \Theta} \left| \frac{\partial^2 E \left[ \Omega (\theta)^{jk} \mid x \right]}{\partial x \partial x'} \right| < \infty, \sup_{x, \theta \in \Theta} \left| \frac{\partial^2 E \left[ G (\theta)^{jm} \mid x \right]}{\partial x \partial x'} \right| < \infty \ j = 1, \ldots, l, m = 1, \ldots, k, \]

\[ \text{such that } n^{1-\beta} h \left( \frac{n+2}{n} \right)^{s/2} \rightarrow \infty, n^{1-2\beta-2/\alpha} h^{2s+4\tau} \rightarrow \infty, \]

\[ n^{\delta} h^{2\tau} \rightarrow \infty, n^{\delta-1/\delta} h^{\tau} \rightarrow \infty, n^{\delta-2/\alpha} h \rightarrow \infty. \]

Then

\[ n^{1/2} \left( \hat{\theta} - \theta_0 \right) \converges_d N \left( 0, \left( G'_{0x} - \Omega_{0x}^{-1} G_{0x} \right)^{-1} \right), \]

where \( G_{0x} = E \left[ \frac{\partial g (\theta_0)}{\partial \theta'} \mid x \right] \) and \( \Omega_{0x} = E \left[ g (\theta_0) \frac{g (\theta_0)'}{\mid x} \right]. \)

Proof. See Kitamura et al. (2004)

We now briefly consider testing restrictions on \( \theta_0 \). We want to test \( H_0 : h (\theta_0) = 0. \) Let

\[ CELR = 2 \left( CEL \left( \hat{\theta} \right) - CEL \left( \hat{\theta}^c \right) \right) \]

denote the conditional empirical likelihood ratio where \( \hat{\theta}^c \) is the constrained estimator.

\textbf{Theorem 5} Under the same assumptions of Theorem 4, and \( H_0 \)

\[ CELR \converges_d \chi_q^2. \]
Proof. Note that the results of Theorem 4 applies to $\Theta_0 = \{\Theta : h(\theta_0) = 0\}$ which is still compact. Thus $\hat{\theta}^c$ is also consistent. Then by a second order Taylor expansion about $\hat{\theta}$ gives

$$CEL_R = n \left( \hat{\theta}^c - \hat{\theta} \right)' \left[ \partial^2 CEL \left( \hat{\theta} \right) / \partial \theta \partial \theta' / n \right] \left( \hat{\theta}^c - \hat{\theta} \right). \quad (5)$$

By results of Kitamura et al. (2004) $\partial^2 CEL \left( \hat{\theta} \right) / \partial \theta \partial \theta' / n \overset{p}{\to} G'_{0x} \Omega^{-1}_{0x} G_0$ and

$$n^{1/2} \left( \hat{\theta} - \theta_0 \right) = \left( G'_{0x} \Omega^{-1}_{0x} G_{0x} \right)^{-1} G'_{0x} \Omega^{-1}_{0x} n^{1/2} \tilde{g} (\theta_0) + o_p \left( 1 \right).$$

Moreover by standard manipulations where

$$n^{1/2} \left( \hat{\theta}^c - \theta_0 \right) = \Upsilon \left( \theta_0 \right) G'_{0x} \Omega^{-1}_{0x} n^{1/2} \tilde{g} (\theta_0) + o_p \left( 1 \right)$$

$$\Upsilon \left( \theta_0 \right) = \left( G'_{0x} \Omega^{-1}_{0x} G_{0x} \right)^{-1} \left[ I - H_0'^{-1} \left( H_0 \left( G'_{0x} \Omega^{-1}_{0x} G_{0x} \right)^{-1} H_0' \right)^{-1} H_0 \left( G'_{0x} \Omega^{-1}_{0x} G_{0x} \right)^{-1} \right]$$

so that

$$n^{1/2} \left( \hat{\theta} - \hat{\theta}^c \right) = \left( G'_{0x} \Omega^{-1}_{0x} G_{0x} \right)^{-1} H_0' \left( H_0 \left( G'_{0x} \Omega^{-1}_{0x} G_{0x} \right)^{-1} H_0' \right)^{-1} H_0 \left( G'_{0x} \Omega^{-1}_{0x} G_{0x} \right)^{-1} \times \left( G'_{0x} \Omega^{-1}_{0x} n^{1/2} \tilde{g} (\theta_0) \right).$$
Let $\xi(\theta_0) = G'_0 x_0^{-1} n^{1/2} \hat{g}(\theta_0)$ and note that $\xi(\theta_0) \overset{d}{\rightarrow} N(0, G'_0 x_0^{-1} G_0 x)$. Then (5) is

$$CELR = \xi(\theta_0)' (G'_0 x_0^{-1} G_0 x)^{-1} H_0' \left( H_0 (G'_0 x_0^{-1} G_0 x)^{-1} H_0' \right)^{-1} \times$$

$$H_0 (G'_0 x_0^{-1} G_0 x)^{-1} \xi(\theta_0)$$

and since $(G'_0 x_0^{-1} G_0 x)^{-1} H_0' \left( H_0 (G'_0 x_0^{-1} G_0 x)^{-1} H_0' \right)^{-1} H_0$ is idempotent with rank $q$ the conclusion follows by standard results. ■

Smith (2007) extends the CEL to ECR statistics, that is

$$\max_{\pi_{ij}} \sum_{i=1}^{n} \sum_{j=1}^{n} T_{i,n}w_{ij} \left[ (\pi_{ij}/w_{ij})^{-\gamma} - 1 \right] / [\gamma (\gamma + 1)] \text{ s.t.}$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} = 1 \text{ and } \sum_{j=1}^{n} \pi_{ij} g(z_j, \theta_0) = 0,$$

and shows that all conditional ECR estimators defined as

$$\hat{\theta} = \arg \max_{\theta} CECR(\theta)$$
where

\[ CER(\theta) = \frac{2}{\gamma(\gamma+1)} \sum_{i=1}^{n} \sum_{j=1}^{n} T_{i,n} \left( \left( 1 + \tilde{\zeta}_i + \tilde{\lambda}_i g(z_j, \theta) \right)^{\gamma/(\gamma+1)} - 1 \right) \]

are asymptotically equivalent to the conditional EL estimator.

Zhang & Gijbels (2003) also consider CEL\(^1\), however their interest is explicitly in parametric and non-parametric (conditional) regressions models as opposed to the conditional moment restrictions themselves as in Kitamura et al. (2004) and Smith (2007). Here we consider some of their results for nonparametric regression models. We first give an example to illustrate and motivate their approach.

**Example 2** Consider a nonparametric regression model \( y_i = \theta_0(x_i) + \varepsilon_i \) where the unobservable errors \( \varepsilon_i \) satisfy \( E[\varepsilon_i^2 | x_i] = \theta(x_i)^2 + 1 \ a.s. \) and \( E[\varepsilon_i^{2(k-1)+1} | x_i] = 0 \ a.s. \) for \( 1 \leq k \leq k_0 - 1 \). Then the \( \mathbb{R}^{k_0} \)-valued vector

\[ g(z_i, \theta) = \left( (y_i - \theta(x_i))^2 - \left( \theta(x_i)^2 + 1 \right) , (y_i - \theta(x_i))^{2(k-1)+1} \right) \]

satisfies the conditional moment condition \( E[g(z_i, \theta_0) | x_i] = 0 \ a.s. \).

Zhang & Gijbels (2003) suggest to use CEL to simultaneously estimate the unknown regression func-

\(^1\) Note that they call their procedure sieve empirical likelihood (SEL).
tion \( \theta(\cdot) \) and the unknown conditional distribution of the observations. We assume that \( \theta \) belongs to a smooth class \( \Theta \) and is equipped with norm \( \|\theta\|^2 = E \left[ \|\theta(x_i)\|^2 | x_i \right] \). Let \( z_i = [y_i, x_i'] \in \mathbb{R}^{k+1} \),

\[
F_g = \{ g(z_i, \theta) \mid \|\theta - \theta_0\| \leq \delta_n \},
F_{gg'} = \{ g(z_i, \theta) g(z_i, \theta') \mid \|\theta - \theta_0\| \leq \delta_n \}
\]

for any positive \( \delta_n \to 0 \), and \( \theta(\theta, \theta^*, s) \in \Theta(0 \leq s \leq 1) \) denote a path\(^2\) from \( \theta \) to \( \theta^* \). Also define

\[
\partial g(\theta(\cdot)) / \partial \theta = \lim_{s \to 0} \frac{g(\theta(\theta, \theta^*, s)) - g(\theta))}{s}
\]

\[
\|\partial g(\theta(\cdot)) / \partial \theta\| = \sup_{\theta^* \in \Theta, \theta^* \neq \theta} \left| \partial g(\theta(\cdot)) / \partial \theta [\theta^*(\cdot) - \theta(\cdot)] \right| / \|\theta^* - \theta\|
\]

and for a given class of functions \( \mathcal{F} \) let

\[
H^B(\epsilon, L_2(P), \mathcal{F}) = \log N^B(\epsilon, L_2(P), \mathcal{F})
\]

denote the \( \epsilon \) entropy with bracketing, where \( N^B(\cdot) \) is the bracketing covering number\(^3\).

\(^2\) For example \( \theta(\theta, \theta^*, s) = (1-s)\theta + s\theta^* \).

\(^3\) The bracketing covering number is the smallest number \( m \) such that for any \( f \in \mathcal{F} \) there exist \( f^l \leq f^u \) with

\[
\max_{1 \leq k \leq m} \int |f^u_k - f^l_k|^2 dP \leq \epsilon
\]

such that \( f^u_k \leq f \leq f^l_k \).
The following theorem establishes consistency and the convergence rate of CEL defined as

\[
\hat{\theta} = \arg \max_{\theta \in \Theta} CEL (\theta) = \arg \max_{\theta \in \Theta} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \log \left( w_{ij} / \left( 1 + \lambda_i g (z_j, \theta) \right) \right).
\]

**Theorem 6 (Zhang and Gijbels (2003))** Assume that (I) \( \sup_{\| \theta - \theta_0 \| \geq \delta} \| E [g (z_i, \theta)] \| > 0 \) a.s, (II) \( \int K (||s||) ds = 1, \int K (||s||) (1+\delta)/\delta ds < \infty \) for any \( \delta > 0, |K (s_1) - K (s_2)| \leq C |s_1 - s_2| \) for any \( s_j \in \mathbb{R} \) \( (j = 1, 2) \), the bandwidth \( h \) satisfies \( c_0 \leq h^k n^n \leq c_1 \) (III) \( x_i \) has compact and convex support \( X_0 \), (IV) \( E \sup_{\| \theta - \theta_0 \| \leq \delta, \theta \in \Theta} E [g_j^\alpha (z_i, \theta)] < \infty \) \( (j = 1, \ldots, k) \alpha > 4 \), (V) \( \sup_{\| \theta - \theta_0 \| \leq \delta, \theta \in \Theta} \sup_{x_i \in X_0} E [g_j^4 (z_i, \theta) | x_i] < \infty \) \( (j = 1, \ldots, k) \), (VI)

\[
\sup_{\| \theta - \theta_0 \| \leq \delta, \theta \in \Theta} \sup_{x_i \in X_0} \| \partial E [g_j (z_i, \theta) | x_i] / \partial \theta \| < \infty \) \( (j = 1, \ldots, k) \),
\]

(VII)

\[
\sup_{\| \theta - \theta_0 \| \leq \delta, \theta \in \Theta} \sup_{x_i \in X_0} | E [g_j (z_i, \theta (x_i + u) | x_i)] - E [g_j (z_i, \theta (x_i) | x_i)] | \rightarrow 0
\]

\[
\sup_{\| \theta - \theta_0 \| \leq \delta, \theta \in \Theta} \sup_{x_i \in X_0} | E [g_{jl} (z_i, \theta (x_i + u) | x_i)] - E [g_{jl} (z_i, \theta (x_i) | x_i)] | \rightarrow 0
\]

CAEPR mini-course in empirical likelihood methods
as $u \to 0$ where $g_{jl}(\cdot) = g_j(\cdot)g_l(\cdot)$ ($j, l = 1, \ldots, k$), (VIII)

$$
\sup_{\|\theta - \theta_0\| \leq \delta, \theta \in \Theta} \sup_{x_i \in X_0} \sigma_{\min} \left( E \left[ g(z_i, \theta) g(z_i, \theta)' | x_i \right] \right)
$$

is positive definite (IX)

$$
H^B(\epsilon, L_2(P), F_g) \leq A(P)(\epsilon/\delta_n)^{-\beta}
$$

$$
H^B(\epsilon, L_2(P), F_{gg'}) \leq B(P)(\epsilon/\delta_n)^{-\beta},
$$

(X) there exists a measurable function $M(z)$ and a $\zeta > 0$ such that for any $\theta_j \in \Theta$ ($j = 1, 2$)

$$
|g_j(z_i, \theta_1) - g_j(z_i, \theta_2)| \leq M(z)(\theta_1 - \theta_2)
$$

$$
\sup_{x_i \in X_0} E \left[ \exp(\zeta M(z)) | x_i \right] < \infty \ a.s.,
$$

$$
|\partial E[g_j(z_i, \theta_1)]/\partial \theta - \partial E[g_j(z_i, \theta_2)]/\partial \theta| \leq C |\theta_1 - \theta_2|
$$

(XI)

$$
\sup_{x_i \in X_0} \left\| E \left[ g(z_i, \theta_0(x_i + u)) g(z_i, \theta_0(x_i + u))' \right] - E \left[ g(z_i, \theta_0(x_i)) g(z_i, \theta_0(x_i))' \right] \right\| \leq C |u|
$$

If (I)-(IX) hold and $0 < \eta < 2(\alpha - 4)/(\alpha(2 + \beta))$ then $\hat{\theta} \xrightarrow{P} \theta_0$. In addition if (X-XI) hold then for any
\[ 0 \leq \pi \leq \pi^* \text{ and } \eta_1(\pi) < \eta < \eta_2(\pi) \]

\[
\left\| \hat{\theta} - \theta_0 \right\| = O_p \left( \max \left\{ n^{-1/(2+\beta^*)}, n^{-\pi/(1-\beta^*)} \right\} \right)
\]

where \( \beta^* = 1/(q+r) \), \( \zeta_1 = 2/(2+\beta^*) \), \( \zeta_2 = 2(1-\beta^*)/(2-\beta^*) \), \( 1 \leq \xi \leq 2 \),

\[
\pi^* = \min \left\{ \frac{1}{3} \left( \frac{2}{(1-\beta^*) - \zeta_2 - 1} + 2\zeta_2 - \xi \zeta_1 \right), \frac{(6+\beta^*) \left( 2/(1-\beta^*) - \zeta_2 - 1 \right) + 4\zeta_2 - 2\beta^* - \xi \zeta_1 (2-\beta^*)}{2} \right\},
\]

\[
\eta_1(\pi) = \min \left\{ \pi \left( \zeta_1 - \zeta_2 \right), \frac{2\pi}{1-\beta^*} - \pi \left( 1 + \zeta_2 \right) \right\},
\]

\[
\eta_2(\pi) = \min \left\{ (1 - 2\zeta_2 \pi + \xi \zeta_1 \pi)/3, 2(1 - 2\zeta_2 \pi + \beta_2 \pi + \xi \zeta_1 (2-\beta^*) \pi/(6+\beta^*)) \right\}.
\]

**Proof.** Under the assumptions it is possible to show that

\[
\sup \left( 1 + \max_i \left\| g(z_i, \theta) \right\| \right) \left\| \sum_{j=1}^n w_{ij} g(z_i, \theta) \right\| = O_p(1)
\]

uniformly in \( x_i \in X_0, \theta \in \Theta \) and \( \left\| \theta - \theta_0 \right\| \leq \delta_n < n^{-1/\alpha} \), which implies by standard arguments(cf. proof CAEPR mini-course in empirical likelihood methods
of Theorem 4) that

\[
\hat{\lambda}_i = \left( \sum_{j=1}^{n} w_{ij} g(z_j, \theta) g'(z_j, \theta) \right)^{-1} \sum_{j=1}^{n} w_{ij} g(z_j, \theta) + o_p(1)
\]

\[
= \left( E \left[ g(z_i, \theta_0) g'(z_i, \theta_0) \mid x_i \right] \right)^{-1} \sum_{j=1}^{n} w_{ij} g(z_j, \theta) + o_p(1)
\]

uniformly in \( x_i \in X_0, \theta \in \Theta \) and \( \| \theta - \theta_0 \| \leq \delta_n < n^{-1/\alpha} \).

Next use ULLN show that \( \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} g(z_j, \theta) / n \) and \( \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} g(z_j, \theta) g'(z_j, \theta) / n \) converge uniformly to their expected values and proceed (similarly to the proof of Theorem 7) to show that

\[
\Pr \left( - \sup_{\| \theta - \theta_0 \| \geq \delta} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \log \left( 1 + \lambda_i' g(z_j, \theta) \right) \geq - \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \log \left( 1 + \lambda_i' g(z_j, \theta_0) \right) \right) \to 1.
\]

To obtain the convergence rate note that

\[
2CEL(\theta) = - \sum_{i=1}^{n} \left( \sum_{j=1}^{n} w_{ij} g(z_j, \theta) \right) \left( \sum_{j=1}^{n} w_{ij} g(z_j, \theta) g'(z_j, \theta) \right)^{-1} \left( \sum_{j=1}^{n} w_{ij} g(z_j, \theta) \right) + o_p(1)
\]
and that as $\left\| \hat{\theta} - \theta_0 \right\| \leq \delta_n \to 0$

$$2 \left( CEL(\theta) - CEL(\theta_0) \right) = - \sum_{i=1}^{n} \left( \sum_{j=1}^{n} w_{ij} g(z_j, \theta)' \right) \left( E \left[ g(z_i, \theta) g(z_i, \theta)' | x_i \right] \right)^{-1} \times \left( \sum_{j=1}^{n} w_{ij} g(z_j, \theta) \right) + o_p(1).$$

The convergence rate is then obtained using the same approach used by Shen & Wong (1994), that consists of improving iteratively the rate by obtaining increasingly faster uniform approximation rates for $CEL(\theta)$ -see Zhang & Gijbels (2003) for more details. ■

**Remark 2** When $\beta^* \leq 1/4 \left\| \hat{\theta} - \theta_0 \right\| = O_p \left( n^{-1/(2+\beta^*)} \right)$, which is the optimal convergence rate in ordinary nonparametric regression.
2.3 Applications of Conditional EL

2.3.1 Specification testing

Tripathi & Kitamura (2003) use CEL to test whether (2) holds over a compact set $X_0$ that is

$$E[g(z, \theta_0) | x] = 0 \text{ a.s. for } \theta_0 = \Theta \text{ and } x \in X_0.$$ 

The same Lagrangian argument of the previous section shows that the CEL test statistic is

$$2CEL(\hat{\theta}) = 2 \sum_{i=1}^{n} \sum_{j=1}^{n} I_i w_{ij} \log \left( w_{ij} / \left( 1 + \lambda_i g(z_j, \hat{\theta}) \right) \right)$$

where $I_i = I \{x_i \in X_0\}$. Tripathi & Kitamura (2003) show that

$$\left( 2CEL(\hat{\theta}) - h^{-s} q c_1(K) \right) / \left( (2h^{-s} q)^{1/2} c_2(K) \right) \xrightarrow{d} N(0, 1)$$

where $c_1(K) = \int K(u)^2 du$ and $c_2(K) = \int \left[ \int K(v) K(u-v) dv \right]^2 du$. 

It is interesting to note that compared to other commonly used procedures CEL has two advantages. Assume that $g$ and $x$ are scalar and consider a weighted sum of squared residuals of a kernel regression (Ait-Sahalia, Bickel & Stoker 2001)

$$S_a = h \sum_{i=1}^{n} \sum_{j=1}^{n} a(x_i) \left[ w_{ij} g(z_j, \hat{\theta}) \right]^2$$

where $a(\cdot) : [0, 1] \rightarrow \mathbb{R}^+$ is a function such that $\int a(x)^2 dx = 1$. An example of $S_a$ is the statistic proposed by Hardle & Mammen (1993). Let $\sigma^2(x) = V(g(z_j, \theta_0) | x)$ denote the conditional variance and note that

$$T_{s_a} = h^{-1/2} \left( S_a - c_1(K) \int \sigma a dx \right) / \left( 2c_2(K) \int \sigma^2 a^2 dx \right)^{1/2} \xrightarrow{d} N(0, 1),$$

First note that, as opposed to $CEL$, $T_{s_a}$ is not invariant to reparameterisations of $g$. Second $CEL$ has an advantage in terms of power. To be specific let

$$E \left[ g(z, \theta_n) | x \right] = n^{-1/2} h^{s/4} \delta(x) \ a.s.$$

denote a sequence of alternative hypothesis, let $f(x)$ denote the distribution of $x$ and

$$\mu(a, \delta) = \int_0^1 \delta^2(x) f(x) a^2(x) / \left( 2c_2(K) \int \sigma^2 a^2 dx \right)^{1/2}.$$
Then $T_{sa} \xrightarrow{d} N(\mu(a, \delta), 1)$ and thus the asymptotic local power of the test with critical value $c_\alpha$ is

$$\pi(a, \delta) = 1 - \Phi(c_\alpha - \mu(a, \delta)).$$

Given that $\delta$ is unknown one way to evaluate $\pi(a, \delta)$ is to use the notion of average local power, i.e. $\bar{\pi}(a, \delta) = \int \pi(a, \delta) \, dP_\delta$ where $P_\delta$ is the distribution under the alternative. In case of parametric likelihood functions the latter is typically known, however in the present context of moment conditions models this is clearly not the case. Tripathi & Kitamura (2003) solve this problem by using the non-parametrically denote a sequence of alternative hypothesis, let $f(x)$ denote the distribution of the $x_i$ that is consistent with the conditional moment condition

$$\hat{\delta}(x) = \sum_{i=1}^{n} K_h(x_i - x) g(x_i, \theta) / \sum_{j=1}^{n} K_h(x_i - x)$$

They then redefine the average local power as

$$\bar{\pi}(a, \hat{\delta}(x)) = \int_{C[0,1]} \pi(a, \hat{\delta}(x)) \, dP_\delta(x),$$

where $C[0,1]$ is the set of continuous function over $[0,1]$ and show using some calculus of variations.
that \( \bar{\pi} \left( a, \hat{\delta} (x) \right) \) is maximised at

\[
a = 1 / \left( \sigma (x) \int_0^1 \sigma^{-2} (x) \, dx \right)
\]

which happens to be the value of \( a \) that the statistic \( T_{s_a} \) takes if we are considering the CEL. Thus CEL has maximal average local power.

Smith (2007) shows that the CCR statistic

\[
CCR \left( \hat{\theta} \right) = \frac{2}{\gamma (\gamma + 1)} \sum_{i=1}^{n} \sum_{j=1}^{n} I_i \left[ \left( 1 + \hat{\zeta}_i + \hat{\lambda}_i g \left( z_j, \hat{\theta} \right) \right)^{\gamma/(\gamma+1)} - 1 \right]
\]

can be used to test (2). To be specific he shows that under \( H_0 \)

\[
\left( CCR \left( \hat{\theta} \right) - h^{-s} q c_1 (K) \right) / \left( (2h^{-s} q)^{1/2} c_2 (K) \right) \overset{d}{\to} N \left( 0, 1 \right).
\]

By the asymptotic equivalence between CEL and CECR it is expected that all of the members of the latter enjoy the optimality property of the former in terms of average local power.
2.3.2 Non nested conditional moment restrictions

Otsu & Whang (2007) use CEL to develop three non nested tests: moment encompassing, Cox-type and efficient score encompassing test.

Consider two competing conditional moment restriction models

\[
H^g_0 : E[g(z_i, \theta) | x_i] = 0 \text{ if } \theta = \theta_0 \\
H^h_0 : E[h(z_i, \beta) | x_i] = 0 \text{ if } \beta = \beta_0
\]

Let CEL\(_g(\theta)\) denote the conditional EL as defined in Theorem 4, let \(\hat{\theta}\) denote the Conditional MELE and similarly for model \(h\). If the model is misspecified \(\hat{\theta}\) and \(\hat{\beta}\) converge to their pseudo-true value defined as

\[
\hat{\theta}^* = \min_{\theta \in \Theta} E \left[ E_{\lambda} \left( \log \left( 1 + \lambda' g(z_i, \theta) \right) \right) | x_i \right], \\
\hat{\beta}^* = \min_{\beta \in \beta} E \left[ E_{\lambda} \left( \log \left( 1 + \lambda' h(z_i, \beta) \right) \right) | x_i \right]
\]

To compute the three non nested statistics Otsu & Whang (2007) use the implied probabilities

\[
\hat{\pi}^g_{ij} = w_{ij} / \left( 1 + \lambda'_{ij} g(z_j, \hat{\theta}) \right) \\
\hat{\pi}^h_{ij} = w_{ij} / \left( 1 + \lambda'_{ij} h(z_j, \hat{\beta}) \right)
\]
where $w_{ij}$ are the kernel weights $\hat{\theta}$ and $\hat{\beta}$ are any $n^{1/2}$ consistent estimators. Let $m_{ij}(\theta, \beta) = m_{ij}(z_j, x_i, \theta, \beta)$ denote a moment indicator that can be expressed as $M(x_i, \theta, \beta) \tilde{m}(z_j, \theta, \beta)$ for an appropriate matrix $M(\cdot)$ and vector $\tilde{m}(\cdot)$. Typically $M(\cdot) = I$ and $\tilde{m}(\cdot) = h(z_j, \beta)$ (i.e. it is the alternative model).

Consider now the difference between the moment indicators evaluated at the implied probabilities (6) and those evaluated at the unconstrained one $w_{ij}$

$$T_m(\hat{\theta}, \hat{\beta}) = \sum_{i=1}^{n} \sum_{j=1}^{n} I_i \sum_{j=1}^{n} \pi_{ij} m_{ij}(\hat{\theta}, \hat{\beta}) / n - \sum_{i=1}^{n} I_i \sum_{j=1}^{n} w_{ij} m_{ij}(\hat{\theta}, \hat{\beta}) / n$$

where as before $I_i = I\{x_i \in X_0\}$ is trimming function for a subset $X_0$ of $X$.

Under the null hypothesis that $H_g$ is correct $T_m \xrightarrow{p} 0$ otherwise $T_m$ diverges. Then we can define the CEL-based moment encompassing test statistic for $H_0^g$ as

$$T_m^2 = nT_m(\hat{\theta}, \hat{\beta})' \Phi_m(\hat{\theta}, \hat{\beta})^{-g} T_m(\hat{\theta}, \hat{\beta})$$
where 

\[ g \] denotes generalised inverse, and

\[
\Phi_m(\theta, \beta) = \sum_{i=1}^{n} \psi_{im}(\theta, \beta) \psi_{im}(\theta, \beta)' / n,
\]

\[
\psi_{im}(\theta, \beta) = -I_i M (x_i, \theta, \beta) \hat{J}_i(\theta, \beta) \hat{\Omega}_i^g(\theta)^{-1} g(z_i, \theta) + \hat{H}_m(\theta, \beta) \Delta \psi_i(\theta)
\]

\[
\hat{J}_i(\theta, \beta) = \sum_{j=1}^{n} w_{ij} g(z_j, \theta) \tilde{m}_j(z_j, \theta, \beta)' / n, \quad \hat{\Omega}_i^g(\theta) = \sum_{j=1}^{n} w_{ij} g(z_j, \theta) g(z_j, \theta)' / n,
\]

\[
\hat{H}_m(\theta, \beta) = \sum_{i=1}^{n} I_i M (x_i, \theta, \beta) \hat{J}_i(\theta, \beta) \hat{\Omega}_i^g(\theta)^{-1} \hat{G}_i(\theta),
\]

\[
\hat{G}_i(\theta) = \sum_{j=1}^{n} w_{ij} \left( \partial g(z_j, \theta) / \partial \theta' \right) / n,
\]

and

\[
n^{1/2} \left( \hat{\theta} - \theta_0 \right) = -\Delta^{-1} \sum_{j=1}^{n} \psi_j(\theta_0) + o_p(1). \quad (7)
\]

To define a Cox-type conditional non-nested statistic consider the difference between the quadratic....
Empirical likelihood methods with application to econometrics

\[ T_c \left( \hat{\theta}, \hat{\beta} \right) = \sum_{i=1}^{n} I_i \hat{h}_i^g \left( \hat{\beta} \right) \hat{h}_i^h \left( \hat{\beta} \right)^{-1} \hat{h}_i^g \left( \hat{\beta} \right) / n - \sum_{i=1}^{n} I_i \hat{h}_i \left( \hat{\beta} \right) \hat{\Omega}_i^h \left( \hat{\beta} \right)^{-1} \hat{h}_i \left( \hat{\beta} \right) / n \]

where

\[ \hat{h}_i^g \left( \beta \right) = \sum_{j=1}^{n} \tilde{\pi}_{ij}^g h \left( z_j, \beta \right), \quad \hat{h}_i \left( \beta \right) = \sum_{j=1}^{n} w_{ij} h \left( z_j, \beta \right), \quad \hat{\Omega}_i^h \left( \beta \right) = \sum_{j=1}^{n} w_{ij} h \left( z_j, \beta \right) h \left( z_j, \beta \right)' / n. \]

The CEL-base Cox test statistic is

\[ C_g = n^{1/2} T_c \left( \hat{\theta}, \hat{\beta} \right) / \hat{\phi}_{c}^{1/2} \]
where

\[
\hat{\phi}_c = \sum_{i=1}^{n} \psi_{ic}(\theta, \beta)^2 / n,
\]

\[
\psi_{ic}(\theta, \beta) = -2 I_i \hat{h}_i(\beta) \hat{\Omega}_i^h(\theta)^{-1} \hat{J}_i^h(\theta, \beta) \hat{\Omega}_i^g(\theta)^{-1} g(z_i, \theta) + \hat{H}_c(\theta, \beta) \Delta \psi_i(\theta)
\]

\[
\hat{J}_i^h(\theta, \beta) = \sum_{j=1}^{n} w_{ij} g(z_j, \theta) h(z_j, \beta)' / n,
\]

\[
\hat{H}_c(\theta, \beta) = 2 \sum_{j=1}^{n} I_i \hat{h}_i(\beta) \hat{\Omega}_i^h(\beta)^{-1} \hat{J}_i^h(\theta, \beta) \hat{G}_i(\theta).
\]

Finally the CEL-based efficient score encompassing test, which focuses on the probability limit of the asymptotic linear form of the asymptotically efficient estimators for \( \beta_0 \) in \( H_0^h \), that is

\[
n^{1/2} \left( \hat{\beta} - \beta_0 \right) = -n^{-1/2} I^h(\beta_0) \sum_{i=1}^{n} G_i(\beta)' \Omega_i^h(\beta)^{-1} h(z_i, \beta)
\]

where

\[
I^h(\beta) = E \left[ G_i^h(\beta)' \Omega_i^h(\beta)^{-1} G_i^h(\beta) \right], \quad G_i^h(\beta) = E \left[ \partial h(z_i, \beta) / \partial \beta' | x_i \right],
\]

\[
\Omega_i^h(\beta) = E \left[ h(z_i, \beta) h(z_i, \beta)' | x_i \right].
\]
Let $\hat{G}^h_i(\beta) = \sum_{j=1}^n w_{ij} \partial h(z_j, \beta) / \partial \beta'$; similarly to the moment encompassing test consider the difference between two efficient scores

$$T_s(\hat{\beta}) = \sum_{i=1}^n \sum_{j=1}^n I_i \hat{G}^h_i(\hat{\beta}) \hat{\Omega}_i^h(\hat{\beta})^{-1} \hat{\pi}_{ij}^g h_{ij}(\hat{\beta}) / n - \sum_{i=1}^n \sum_{j=1}^n I_i \hat{G}^h_i(\hat{\beta}) \hat{\Omega}_i^h(\hat{\beta})^{-1} w_{ij} h_{ij}(\hat{\beta}) / n$$

can be used to test $H_0^g$. The CEL based efficient score encompassing test is

$$T_s^2 = nT_s(\hat{\beta})' \Phi_s(\hat{\beta})^{-g} T_s(\hat{\beta})$$
where

\[
\Phi_s (\beta) = \sum_{i=1}^{n} \psi_{is} (\beta) \psi_{is} (\beta)' / n,
\]
\[
\psi_{is} (\theta, \beta) = -I_i \hat{G}_i^h (\beta) \hat{J}_i^h (\theta, \beta) \hat{\Omega}_i^g (\theta)^{-1} g(z_i, \theta) + \hat{H}_s (\beta) \Delta \psi_i (\theta)
\]
\[
\hat{J}_i^h (\theta, \beta) = \sum_{j=1}^{n} w_{ij} g(z_j, \theta) h(z_j, \beta)' / n,
\]
\[
\hat{H}_c (\theta, \beta) = 2 \sum_{j=1}^{n} I_i \hat{h}_i (\beta) \hat{\Omega}_i^h (\beta)^{-1} \hat{J}_i^h (\theta, \beta) \hat{\Omega}_i^g (\theta)^{-1} \hat{G}_i (\theta) / n.
\]

Under regularity conditions that are similar to those of Theorem 4 Otsu & Whang (2007) show that

\[
T_m^2 \xrightarrow{d} \chi_{\text{rank} (\Phi_m (\beta_0))}^2
\]
\[
C_g \xrightarrow{d} N (0, 1)
\]
\[
T_s^2 \xrightarrow{d} \chi_{\text{rank} (\Phi_s (\beta_0))}^2
\]


Pakes, A. & Pollard, D. (1989), ‘Simulation and the asymptotics of optimization estimators’, *Economet-

