Applications of Empirical Process Methods

Myung Hwan Seo

EC518, Lent term, 2012[1]

---

[1] Cf. Office Hour: 13:30-14:30 Wednesday; S580.

Contents

1 Applications of Empirical Process Methods

1.1 Arrays

In estimation problems, the asymptotic distribution of the estimator is characterized by the convergence of empirical processes with class of functions changing with n due to the reparametrization. That is, we provide conditions for the stochastic processes

$$\{\mathbb{G}_n f_{n,\tau} : \tau \in \mathcal{T}\}$$

to converge weakly to a Gaussian process in the space of $\ell^\infty(\mathcal{T})$. These are empirical processes indexed by classes of functions $\mathcal{F}_n$ changing with n. We need

1. Given envelope functions F_n,

$$\begin{aligned} \mathbf{P}\left(F_n^2\right) &= O(1), \\ \mathbf{P}\left(F_n^2 1\left\{F_n > \eta\sqrt{n}\right\}\right) &\to 0, \quad \text{for every } \eta > 0 \\ \sup_{\rho(s,t)<\delta_n} \mathbf{P}\left(f_{n,s} - f_{n,t}\right)^2 &\to 0, \quad \text{for every } \delta_n \to 0. \end{aligned}$$

2. For every $\delta_n \to 0$, either

$$\sup_Q \int_0^{\delta_n} \sqrt{\log N \left(\varepsilon \left\| F_n \right\|_{Q,2}, \mathcal{F}_n, L_2\left(Q\right) \right)} d\varepsilon \to 0,$$

   or

$$\int_0^{\delta_n} \sqrt{\log N_{[]} \left(\varepsilon \left\| F_n \right\|_2, \mathcal{F}_n, L_2\left(\mathbf{P}\right) \right)} d\varepsilon \to 0.$$

3. The sequence of covariance functions $\mathbf{P}\left(f_{n,s} f_{n,t}\right) - \mathbf{P} f_{n,s} \mathbf{P} f_{n,t}$ converges pointwise on $\mathcal{T} \times \mathcal{T}$.

See VW section 2.11.3. for the proof.

The uniform-entropy condition in condition 2 is satisfied if $\mathcal{F}_n$ is a VC-class with VC-index bounded by some constant independent of n.

1.2 QLR test

Suppose that the null hypothesis is

$$\mathcal{H}_0 : \mu(\theta) = 0$$

and its negation is the alternative hypothesis. Also assume that $\{l_\xi, \xi \in \Xi\}$, stands for the null model, that is, there is one-to-one correspondence between m_θ and l_ξ under the null hypothesis.

The likelihood ratio test or more generally test based on the comparison of the objective function values evaluated under the null hypothesis and under the alternative hypothesis has some nice properties such as invariance to scaling. Let

$$\begin{aligned} QLR_n &= r_n^2 \left[\max_{\theta \in \Theta} \mathbb{M}_n(\theta) - \max_{\theta \in \Theta_0} \mathbb{M}_n(\theta) \right], \\ &= r_n^2 \left[\max_{\theta \in \Theta} \mathbb{M}_n(\theta) - \max_{\xi \in \Xi} \mathbb{L}_n(\xi) \right] \end{aligned}$$

where r_n is the convergence rate of the estimators and $\mathbb{L}_n(\xi) = \mathbb{P}_n l_\xi$. Standard approach based on the linearization involves somewhat messy algebra and limited to the case where the linearization is plausible. Instead, we note

1. with probability arbitrarily close to one, for any $K < \infty$,

$$\max_{\theta \in \Theta} \mathbb{M}_n(\theta) = \max_{|h_m| \leq K} \mathbb{M}_n\left(\theta_0 + h_m r_n^{-1}\right),$$
$$\max_{\xi \in \Xi} \mathbb{L}_n(\xi) = \max_{|h_l| \leq K} \mathbb{L}_n\left(\xi_0 + h_l r_n^{-1}\right),$$

2. The transformation

$$\phi(m_\theta, l_\xi) = m_\theta - l_\xi$$

is Lipschitz, and $\phi(m_{\theta_0}, l_{\xi_0}) = 0$. This gives tightness of the empirical process part using the permanence of uniform entropy condition or bracketing integral.[1]

3.
$$\max_{\theta \in \Theta} \mathbb{M}_n(\theta) - \max_{\xi \in \Xi} \mathbb{L}_n(\xi) = \max_{\theta, \xi = \xi_0} \mathbb{P}_n \phi - \left(\max_{\xi, \theta = \theta_0} -\mathbb{P}_n \phi\right),$$

is a continuous functional on ϕ.

---

[1] This holds uniformly over n if it did for the original classes.

Therefore, we the asymptotic distribution of QLR_n can be derived from the weak convergence of $r_n^2 \mathbb{M}_n \left(\theta_0 + h_m r_n^{-1}\right)$ and $r_n^2 \mathbb{L}_n \left(\xi_0 + h_l r_n^{-1}\right)$ due to the permanence property of P-Donsker. In particular,

$$\begin{aligned} r_n^2 \mathbb{M}_n \left(\theta_0 + h_m r_n^{-1}\right) &= \frac{r_n^2}{\sqrt{n}} \mathbb{G}_n m_{\theta_0 + h_m r_n^{-1}} + r_n^2 \mathbf{P} m_{\theta_0 + h_m r_n^{-1}} \\ &= \frac{r_n^2}{\sqrt{n}} \mathbb{G}_n m_{\theta_0 + h_m r_n^{-1}} + \frac{1}{2} h_m' \left[\frac{\partial^2}{\partial \theta \partial \theta'} \mathbf{P} m_{\theta_0} \right] h_m + o\left(1\right), \end{aligned}$$

where the second inequality is due to the identification condition. The weak convergence of the first term can be deriven as above, that is, as tightness follows for the same reason for the original estimation, it only remains to calculate the covariance kernel of the gaussian part (jointly with m_θ and l_ξ) and the bias term. If the gaussian part is linear and the bias is quadratic, we can show that the test statistic is asymptotically (Noncentral) chi-squared distributed.

To be more concrete, consider

$$\mathcal{H}_0 : \theta_1 = 0,$$

where $\theta = (\theta_1', \theta_2')'$, and suppose

$$\frac{r_n^2}{\sqrt{n}} \mathbb{G}_n m_{\theta_0 + h_m r_n^{-1}} \xrightarrow{d} Z' h_m,$$

where $Z \sim \mathcal{N}(0, E)$. Partition $Z = (Z_1', Z_2')'$ and the second derivative matrix $V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}$ according to $(\theta_1', \theta_2')'$. Then,

$$\frac{r_n^2}{\sqrt{n}} \mathbb{G}_n l_{\xi_0 + h_l n r_n^{-1}} \xrightarrow{d} Z_2' h_l, \quad \text{and} \quad r_n^2 \mathbf{P} l_{\theta_0 + h_l r_n^{-1}} \to \frac{1}{2} h_l' V_{22} h_l.$$

and the maximizers of the limits are given by

$$\hat{h}_1 = -V^{-1} Z \text{ and } \hat{h}_2 = -V_{22}^{-1} Z_2.$$

If in addition, $E = -V$ as in the MLE, then

$$2QLR_n \xrightarrow{d} -Z' V^{-1} Z + Z_2' V_{22}^{-1} Z_2 = \mathcal{X}_{k_1}^2,$$

where k_1 is the dimension of θ_1.

1.3 Z-estimation

An estimator satisfying an approximate first order condition:

$$\Psi_n\left(\hat{\theta}\right)=\frac{1}{n}\sum_{i=1}^{n}\psi\left(w_i,\hat{\theta}\right)=o_p\left(n^{-1/2}\right). \tag{1}$$

However, it does not mean that $\hat{\theta}$ is defined by this equation, as in LAD example.

The parameter can be infinite-dimensional but let's focus on finite-dimensional θ.

Let $\Psi\left(\theta\right)=\mathrm{E}\left(\psi\left(w_i,\theta\right)\right)$

- Maintained assumptions:

    1. the consistency: $\hat{\theta}\xrightarrow{p}\theta_0$ such that $\Psi\left(\theta_0\right)=0$.
    2. Ψ is continuously differentiable at θ_0 with derivative H, which is invertible at θ_0.

- Then, for the first derivative H of Ψ

$$\Psi\left(\hat{\theta}\right)=H\left(\tilde{\theta}\right)\left(\hat{\theta}-\theta_0\right).$$

Furthermore, it follows from (1) that

$$\begin{aligned}
-\sqrt{n}\Psi\left(\hat{\theta}\right) &= \sqrt{n}\left(\Psi_n\left(\hat{\theta}\right)-\Psi\left(\hat{\theta}\right)\right)+o_p(1)\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\psi\left(w_i,\hat{\theta}\right)-\Psi\left(\hat{\theta}\right)\right)+o_p(1),\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi\left(w_i,\theta_0\right)+o_p(1)
\end{aligned}$$

where the first component on rhs is an empirical process evalutated at $\theta=\hat{\theta}$. If it is Donsker and $\hat{\theta}\xrightarrow{p}\theta_0$, then the stochastic-equicontinuity implies that its limit distribution is equivalent to that of $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi\left(w_i,\theta_0\right)$.

- LAD Example

  1. (1) is shown by e.g. Ruppert and Carroll (1980 JASA). Basic idea is $\hat{\theta}$ is the minmizer of $M_n(\theta)=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-x_i'\beta\right|$. However, there exists at most one i such that $\left|y_i-x_i'\hat{\beta}\right|$ is not-differentiable at $\beta=\hat{\beta}$ with

probability one due to continuity of the distribution and $|u|$ is differentiable almost every u except $u = 0$. Let $M_n(\theta)_{-i} = \frac{1}{n}\sum_{j \neq i}^{n} \left| y_j - x_j' \beta \right|$. Then, for any θ in a neighborhood of $\hat{\theta}$

$$\begin{aligned} 0 &\geq M_n\left(\hat{\theta}\right) - M_n(\theta) \\ &\geq M_n\left(\hat{\theta}\right)_{-i} - M_n(\theta)_{-i} - \frac{1}{n}\left| \left| y_i - x_i'\hat{\beta} \right| - \left| y_i - x_i'\beta \right| \right| \end{aligned}$$

Thus,

$$\frac{1}{n}|x_i| \geq \frac{\partial}{\partial \theta'} M_n\left(\hat{\theta}\right)_{-i} \frac{\left(\hat{\theta} - \theta\right)}{\left|\hat{\theta} - \theta\right|}.$$

By construction, $M_n(\theta)_{-i}$ is differentiable at $\theta = \hat{\theta}$ and its second derivative is zero. As θ is arbitrary and $\frac{\left(\hat{\theta}-\theta\right)}{\left|\hat{\theta}-\theta\right|}$ has norm 1, $\frac{\partial}{\partial \theta'} M_n\left(\hat{\theta}\right)_{-i}$ should shrink at the rate of n^{-1}.

2. Set $\psi(w_i, \theta) = -x_i \operatorname{sgn}(y_i - x_i'\beta)$, which constitutes a VC class. Thus, with a proper moment condition, the stochastic equicontinuity follows. Calculate H etc to characterize the asymptotic variance matrix.

1.4 Semi-parametric 2-step Estimation

The approach here is similar to section 4.1. Suppose that a nonparametric estimate $\hat{\tau}$ is given and the estimator $\hat{\theta}$ of the finite-dimensional parameter θ satisfies

$$\sum_{i=1}^{n} \psi\left(w_i, \hat{\theta}, \hat{\tau}\right) = 0.$$

This is typically the FOC of a minimization problem. Suppose $\hat{\theta} \xrightarrow{p} \theta_0$. Then,

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi\left(w_i, \hat{\theta}, \hat{\tau}\right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi\left(w_i, \theta_0, \hat{\tau}\right) + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \psi\left(w_i, \tilde{\theta}, \hat{\tau}\right) \sqrt{n}\left(\hat{\theta} - \theta_0\right). \end{aligned}$$

Note that the first term equals

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi\left(w_i, \theta_0, \hat{\tau}\right) = \left[\nu_n\left(\hat{\tau}\right) - \nu_n\left(\tau_0\right)\right] + \nu_n\left(\tau_0\right) + \sqrt{n}\mathbf{P}\left(\psi_{\theta_0, \hat{\tau}}\right),$$

where $\nu_n\left(\tau\right) = \mathbb{G}_n \psi_{\theta_0, \tau}$. And

1. For $\nu_n(\tau_0)$, an ordinary CLT applies.
2. Assume $\sqrt{n}\mathbf{P}(\psi_{\theta_0,\hat{\tau}}) = o_p(1)$.
3. And

$$\nu_n(\hat{\tau}) - \nu_n(\tau_0) = o_p(1)$$

   provided that

   (a) the stochastic equicontinuity of ν_n with respect to some $\mathcal{T}$ and ρ,
   (b) $\rho(\hat{\tau}, \tau_0) \xrightarrow{p} 0$
   (c) $\mathbf{P}(\hat{\tau} \in \mathcal{T}) \to 1$.

The stochastic equicontinuity is verified by checking one of the conditions[2] we studied previously. With usual nonparametric density or regression estimates, uniform consistency has been established (see e.g. Andrews 1995 ET), yielding (*b*) and (*c*) in condition 3. Issues associated with the nonparametric kernel estimation such as trimming will be discussed in another topic.

In the GLS example,

$$\psi\left(w_i, \theta_0, \tau\right) = \frac{x_i \left(y_i - x_i' \beta_0\right)}{\sigma^2 \left(x_i\right)},$$

Condition 2 needs to be verified in practice. See Andrews (1994, Econometrica). Sometimes,

$$\mathbf{P}\left(\psi_{\theta_0, \tau}\right) = 0,$$

---

[2] Smooth function class on a **bounded** $\bar{\mathcal{W}}$: the class $\mathcal{F}$ of all bounded functions on a bounded, convex, nonempty set $\bar{\mathcal{W}} \subset \mathbb{R}^s$ that possess uniformly bounded partial derivatives up to order $[\alpha]$ for some $\alpha > 0$ and whose highest derivatives are Lipschitz of order $\alpha - [\alpha]$. The entropy with bracketing is $O\left(\left(\frac{1}{\varepsilon}\right)^{s/\alpha}\right)$, where the constant depends on the smoothness α, size of $\bar{\mathcal{W}}$ in terms of $\mathbf{P}$, and the dimension s of w. Note that higher dimension requires smoother functions as we need $s/\alpha < 2$.

for all τ, which is the case for the GLS due to the conditional moment restriction. It may also demand

$$n^{1/4} \sup_x \left| \hat{\tau}(x) - \tau_0(x) \right| = o_p(1),$$

as in the partially linear regression. We do not discuss details here. In some other occasions,

$$\sqrt{n}\mathbf{P}\left(\psi_{\theta_0,\hat{\tau}}\right) \xrightarrow{d} \mathcal{N}(0, A),$$

often when τ is finite-dimensional.

Example 1 *Partially Linear Regression by Robinson (Econometrica, 1988)*

$$\begin{aligned} y_i &= x_i'\beta + g(z_i) + \varepsilon_i \\ \mathrm{E}(\varepsilon_i | x_i, z_i) &= 0 \end{aligned}$$

The $\hat{\beta}$ *solves the FOC*

$$\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \mathrm{E}\widehat{(x_i|z_i)}\right)\left(y_i - \mathrm{E}\widehat{(y_i|z_i)} - \hat{\beta}'\left(x_i - \mathrm{E}\widehat{(x_i|z_i)}\right)\right) = 0.$$

Thus,

$$\psi(w_i, \theta_0, \tau) = (x_i - \tau_1(z_i))(y_i - \tau_2(z_i) - \beta'(x_i - \tau_1(z_i))),$$

where $\tau_1 = \mathrm{E}(x_i|z_i)$ *and* $\tau_2 = \mathrm{E}(y_i|z_i)$. *And*

$$\begin{aligned} &\sqrt{n}\mathrm{E}\psi(w_i, \theta_0, \tau) \\ &= \sqrt{n}\mathrm{E}(x_i - \tau_1(z_i))(g(z_i) - \tau_2(z_i) + \beta'\tau_1(z_i)) \\ &= \mathrm{E}\left[n^{1/4}(\mathrm{E}(x_i|z_i) - \tau_1(z_i))\right]\left[n^{1/4}(g(z_i) - \tau_2(z_i) + \beta'\tau_1(z_i))\right]. \end{aligned}$$

2 Extensions

2.1 Partial-Sum Processes

Two famous convergence results by Donsker concern (i) the weak convergence of empirical distribution functions and (ii) that of the partial-sum process, i.e., for $s \in [0,1]$,

$$S_n(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[ns]} \xi_i \Rightarrow W(s),$$

where ξ_i is a sequence of independent mean zero random variables with finite variance. This is also called *functional central limit theorem (FCLT).* The FCLT is a building block for the unit root type nonstationary time series analysis. Note that if $y_t = y_{t-1} + x_t$, then $n^{-1/2}\left(y_{[ns]} - y_0\right)$ is a partial-sum process. Billingsley (1968) provides an elegant treatment of this result. Several different approaches are available. We rephrase his Theorem 15.6 here.

Theorem 1 *Let* X_n *be a sequence of stochastic processes whose sample paths are on* $[0,1]$, *right-continuous and have left-limits. And let* X *be a stochstic process on* $[0,1]$ *that is continuous at* 1 *a.s. Suppose (i) fidi holds and (ii)*

$$\mathrm{E}\left(\left|X_n(t)-X_n(t_1)\right|^{\gamma}\left|X_n(t)-X_n(t_2)\right|^{\gamma}\right)\leq\left[g(t_2)-g(t_1)\right]^{2\alpha},$$

for any $0\leq t_1\leq t\leq t_2\leq 1$, *and for some* $\gamma\geq 0,\alpha>\frac{1}{2}$, *and a continuous and nondecreasing function* g. *Then,* $X_n\Rightarrow X$.

In case of *iid* X_i, this is easy to check the FCLT with $\gamma=2$ and $\alpha=1$. The FCLT has been extended to incorporate various dependent sequences $\{\xi_t\}$ as well, see e.g. Davidson (1994).

Remark 1 *The partial-sum process should not be confused with*

$$\tilde{S}_n(s)=\frac{1}{\sqrt{sn}}\sum_{i=1}^{[ns]}\xi_i,$$

which appears in studying various testing problems such as structural break tests. This process has completely different asymptotic property than the partial-sum *process.. For instance, it does not converge weakly and* $\max_s \tilde{S}_n(s)$ *diverges. However, we can still find a sequence of independent mean zero normal variables* Z_i *such that*

$$\Pr\left\{\sup_{1\leq k\leq n}\left|\sum_{i=1}^{k}\xi_i-\sum_{i=1}^{k}Z_i\right|>C\right\}$$

is controlled at a certain order. This is called strong approximation *and has proven useful in some interesting testing problems.*

Next, we consider a generalization, that is, the weak convergence of the *sequential empirical process*

$$\mathbb{Z}_n(s,f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[ns]} \left(f(\xi_i) - \mathbf{P}f \right) = \sqrt{\frac{[ns]}{n}} \mathbb{G}_{[ns]} f,$$

where the index (s,f) ranges over $[0,1] \times \mathcal{F}$.

Theorem 2 *If $\mathcal{F}$ is Donsker,*

$$\mathbb{Z}_n \Rightarrow \mathbb{Z},$$

which is a Gaussian process with the covariance kernel

$$\operatorname{cov}\left(\mathbb{Z}(s,f), \mathbb{Z}(t,g)\right) = (s \wedge t)\left(\mathbf{P}fg - \mathbf{P}f\mathbf{P}g\right).$$

Proof. See Theorem 2.12.1 in VW. ■

This result is for the *iid* ξ_is. Replacing Ottaviani's inequality in the proof, one may extend this to the partial-sum processes with dependent data (e.g. Ch.10.2 in Lin & Bai 2011 Springer).

The following is the maximal inequality for the partial sum process (compare it with the previous term's maximal inequality for the empirical processes).

Lemma 3 *[Kolmogorov Inequality]Let* $\{x_i\}$ *be a sequence of* independent *r.v.'s with* $\mathrm{E}x_i = 0$ *and variance* σ_i^2 *Then,*

$$\Pr\left\{\max_{1\leq k\leq n}\left|\sum_{i=1}^{k} x_i\right| > \varepsilon\right\} \leq \frac{\sum_{i=1}^{n}\sigma_i^2}{\varepsilon^2}. \tag{2}$$

This has been extended to various dependent sequences.

2.2 Testing for the Parameter Constancy

Examples include Davies (1977), Andrews (1993), Hansen (1996), Andrews (2001), etc.

Revisit to Example **??**. Given γ, the null hypothesis is a simple exclusion test for δ after rewriting the model as

$$y_i = x_i'\beta + x_i'\delta \cdot 1\left\{q_i > \gamma\right\} + \varepsilon_i.$$

If $q_i = i$, then the test becomes the structural break test.

If γ were prespecified, it can be tested by the standard Wald test, say $W_n\left(\gamma\right),$ or the Likelihood Ratio test, say $LR_n\left(\gamma\right),$ under the auxiliary assumption of normality of the error. Otherwise, it is natural to consider a continuous functional of the function $W_n\left(\gamma\right)$ or $LR_n\left(\gamma\right).$ Typical is

$$\sup_{\gamma\in\Gamma} W_n\left(\gamma\right) \text{ or } \sup_{\gamma\in\Gamma} LR_n\left(\gamma\right).$$

We elaborate on the structural break test here.

$$y_t = x_t'\beta_1 1\left\{t \le \tau\right\} + x_t'\beta_2 1\left\{t > \tau\right\} + \varepsilon_t.$$

Let $\hat{\beta}_1(\tau)$ and $\hat{\beta}_2(\tau)$ denote the OLSE and $\hat{\delta}(\tau) = \hat{\beta}_2(\tau) - \hat{\beta}_1(\tau)$ with a consistent estimator $\hat{V}(\tau)$ of its asymptotic variance for a given τ. Then,

$$W_n(\tau) = n\hat{\delta}(\tau)\hat{V}(\tau)^{-1}\hat{\delta} \Rightarrow \frac{B(\tau)' B(\tau)}{\tau(1-\tau)},$$

on $\tau \in [\eta, 1-\eta]$, where B is the standard Brownian Bridge. The limit process is called the standardized tied-down Bessel process of order k, the dimension of δ, and for each fixed τ its distribution is χ_k^2. And

$$\sup_{\tau\in[\eta,1-\eta]} W_n(\tau) \Rightarrow \sup_{\tau\in[\eta,1-\eta]} \frac{B(\tau)' B(\tau)}{\tau(1-\tau)}.$$

Thus, the critical values can be tabulated as a function of η, typical choices being 0.05 or 0.15. See Andrews (1993).

To see this, note that it follows from the ULLN[3] (see Andrews 1993 Econometrica) and FCLT that

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_1(\tau)-\beta\right) &= \left(\frac{1}{n}\sum_{t=1}^{[n\tau]} x_t x_t'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{[n\tau]} x_t \varepsilon_t\right) \\ &\Rightarrow (\tau M)^{-1}\Omega^{1/2}W(\tau),\end{aligned}$$

where $M = \mathrm{E}(x_t x_t'), \Omega = \mathrm{E}(x_t \varepsilon_t)$, and W is the standard Brownian motion. Then,

$$\begin{aligned}\sqrt{n}\left(\hat{\beta}_2(\tau)-\beta\right) &\Rightarrow ((1-\tau)M)^{-1}\Omega^{1/2}(W(1)-W(\tau)) \\ \sqrt{n}\hat{\delta}(\tau) &\Rightarrow (\tau(1-\tau)M)^{-1}\Omega^{1/2}(\tau W(1)-W(\tau)).\end{aligned}$$

And

$$(\tau(1-\tau)M)^{-1}\Omega^{1/2}(\tau W(1)-W(\tau)) \sim \mathcal{N}\left(0,(\tau(1-\tau))^{-1}M^{-1}\Omega M^{-1}\right).$$

---

[3] Of course, this can be implied by the FCLT or Kolmogorov inequality at the expense of stronger moment condition on x_t

Remark

1. For the validity of the weak convergence theory, the trimming, that is $\eta > 0$, is indespensible. Otherwise, the supremum of the Wald statistics will diverge.

2. However, the trimming comes with price, that is, it loses power if the true break occurrs outside of the interval $[\eta, 1-\eta]$. Hidalgo and Seo (2011) develop a different asymptotic approximation to the supremum statistic without trimming, incorporating different normalization and then strong approximation to the process on the whole interval $[0,1]$.

3. Another issue is HAC estimation, which arises when x_t and ε_t are serially correlated.

4. When the change point is due to a stochastic variable, the limit distribution

is not distribution free anymore. That is, for $\gamma_1 < \gamma_2$,

$$\begin{aligned}
&\operatorname{cov}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t\varepsilon_t 1\left\{q_t > \gamma_1\right\}, \frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t\varepsilon_t 1\left\{q_t > \gamma_2\right\}\right) \\
&= \mathrm{E}\left(x_t x_t' \varepsilon_t^2 1\left\{\gamma_1 < q_t \leq \gamma_2\right\}\right),
\end{aligned}$$

when $\{\varepsilon_t\}$ is a mds. The critical values need to be simulated. See e.g. Hansen (1996 Econometrica).

3 Estimation of the Change-Point

We consider 3 scenarios: Correctly specified, local-to-linear setup, and misspecified case.

Reparametrize and write the objective function as

$$\mathbb{M}_n(\theta) = \frac{1}{n}\sum_{t=1}^{n}\left(y_t - x_t'\beta\right)^2 + \frac{1}{n}\sum_{t=1}^{[n\tau]}\left(x_t'\delta\right)^2 - 2x_t'\delta\left(y_t - x_t'\beta\right),$$

where $\theta \in \Theta$, which is compact and in particular $\tau \in [\eta, 1-\eta]$. Let

$$\mathbb{M}_n(\theta) \xrightarrow{p} \mathbb{M}(\theta). \tag{3}$$

- For the consistency, we need to check
    1. uniform convergence over some compact Θ of $\mathbb{M}_n$.
    2. well-separatedness of θ_0 as a unique minimizer of $\mathbb{M}(\theta)$ on Θ.

- Let $\sum^{\tau} = \sum_{t=1}^{[n\tau]}$ and $\sum_{\tau} = \sum_{t=[n\tau]+1}^{n}$. then,

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left(y_t - x_t'\beta\right)^2 \\
&= \frac{1}{n}\sum^{\tau_0}\left(\varepsilon_t + x_t'\left(\left(\beta_0+\delta_0\right)-\beta\right)\right)^2 + \frac{1}{n}\sum_{\tau_0}\left(\varepsilon_t + x_t'\left(\beta_0-\beta\right)\right)^2 \\
&\xrightarrow{p} \tau_0 \mathrm{E}\left(\varepsilon_t + x_t'\left(\left(\beta_0+\delta_0\right)-\beta\right)\right)^2 + \left(1-\tau_0\right)\mathrm{E}\left(\varepsilon_t + x_t'\left(\beta_0-\beta\right)\right)^2 .
\end{aligned}$$

And if $\tau \leq \tau_0$

$$\frac{1}{n}\sum_{t=1}^{[n\tau]} x_t'\delta\left(y_t - x_t'\beta\right) \xrightarrow{p} -\tau \mathrm{E} x_t'\delta x_t'\left(\beta-\beta_0-\delta_0\right)$$

and otherwise, it converges to

$$-\tau_0 \mathrm{E} x_t'\delta x_t'\left(\beta-\beta_0-\delta_0\right) - \left(\tau-\tau_0\right)\mathrm{E} x_t'\delta x_t'\left(\beta-\beta_0\right).$$

Putting them together

$$\begin{aligned} &\frac{1}{n}\sum_{t=1}^{[n\tau]}\left(x_t'\delta\right)^2-2x_t'\delta\left(y_t-x_t'\beta\right)\\ \xrightarrow{p}\ &\tau\mathrm{E}\left(x_t'\delta\right)^2+\tau\left[2\mathrm{E}x_t'\delta x_t'\left(\beta-\beta_0-\delta_0\right)\right]1\left\{\tau\leq\tau_0\right\}\\ &+2\left[-\tau_0\mathrm{E}x_t'\delta x_t'\delta_0+\tau\mathrm{E}x_t'\delta x_t'\left(\beta-\beta_0\right)\right]1\left\{\tau>\tau_0\right\}. \end{aligned}$$

Then, it follows that

$$\mathbb{M}\left(\theta_0\right)=\mathrm{E}\left(\varepsilon_t^2\right)$$

and for all $\theta\neq\theta_0$,

$$\mathbb{M}\left(\theta\right)-\mathbb{M}\left(\theta_0\right)>0.$$

Clearly $\mathbb{M}$ is continuous and thus θ_0 is well-separated on any compact set Θ.

- Let $\tau < \tau_0$. Then,

$$
\begin{aligned}
&\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \\
\\
&= \tau_0 \left(\mathrm{E}\left(x_t'\left(\beta-\beta_0\right)\right)^2 - 2\mathrm{E}x_t\left(\beta-\beta_0\right)x_t'\delta_0 + \mathrm{E}\left(x_t'\delta_0\right)^2 \right) \\
&+ \left(1-\tau_0\right)\mathrm{E}\left(x_t'\left(\beta_0-\beta\right)\right)^2 \\
&+ \tau\mathrm{E}\left(x_t'\delta\right)^2 + \tau\left[2\mathrm{E}x_t'\delta x_t'\left(\beta-\beta_0\right) - 2\mathrm{E}x_t'\delta x_t'\delta_0\right] \\
&= \left(\tau_0-\tau\right)\left(\mathrm{E}\left(x_t'\delta_0\right)^2 - 2\mathrm{E}x_t\left(\beta-\beta_0\right)x_t'\delta_0 + \mathrm{E}\left(x_t'\left(\beta_0-\beta\right)\right)^2 \right) \\
&+ \left(1-\tau_0\right)\mathrm{E}\left(x_t'\left(\beta_0-\beta\right)\right)^2 \\
&+ \tau\left(\mathrm{E}\left(x_t'\left(\delta-\delta_0\right)\right)^2 + 2\mathrm{E}x_t'\left(\delta-\delta_0\right)x_t'\left(\beta-\beta_0\right) + \mathrm{E}\left(x_t'\left(\beta_0-\beta\right)\right)^2 \right).
\end{aligned}
$$

Note here that the function is quadratic in $\beta - \beta_0$ and $\delta - \delta_0$ while it is linear in $\tau - \tau_0$

- For the convergence rate,

  1. note that $\mathbb{M}(\theta)$ is not differentiable at θ_0 wrt to τ. In particular, it is reasonable to set

  $$d(\theta_1, \theta_2) = \sqrt{|\tau_1 - \tau_2|} + |\beta_1 - \beta_2| + |\delta_1 - \delta_2| .$$

  2. The modulus of continuity of the centered empirical process can be derived from the Kolmogorov inequality[4], i.e.,

  $$\Pr\left\{ \sup_{d(\theta,\theta_0)<\delta} |\mathbb{G}_n(\theta)| > \varepsilon \right\} \leq \frac{O(\delta)}{\varepsilon}$$

     where $\mathbb{G}_n(\theta) = \sqrt{n}\left(\mathbb{M}_n(\theta) - \mathbb{M}_n(\theta_0) - \mathrm{E}\left(\mathbb{M}_n(\theta) - \mathbb{M}_n(\theta_0)\right)\right)$.

  3. The preceding implies the convergence rate $r_n = n^{1/2}$ and thus

  $$|\hat{\tau} - \tau_0| = O_p\left(n^{-1}\right), \quad \text{and} \quad \left|\hat{\beta} - \beta_0\right|, \left|\hat{\delta} - \delta_0\right| = O_p\left(n^{-1/2}\right).$$

---

[4] Careful reading reveals that we need to take as the ϕ function the square root of the right side of the Kolmogorov inequality in (2).

- To verify them, see our previous derivation of $\mathbb{M}(\theta)$ and note that

$$
\begin{aligned}
&\mathbb{M}_n(\theta) - \mathbb{M}_n(\theta_0) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(-2y_t + x_t'(\beta+\beta_0)\right)x_t'(\beta-\beta_0) \\
&+ \frac{1}{n}\sum_{t=1}^{[n\tau]}\left(x_t'(\delta+\delta_0) - 2(y_t - x_t'\beta)\right)x_t'(\delta-\delta_0) + 2x_t'\delta_0 x_t'(\beta-\beta_0) \\
&+ \frac{1}{n}\sum_{t=[n\tau_0]+1}^{[n\tau]}\left(x_t'\delta_0\right)^2 - 2(y_t - x_t'\beta_0)x_t'\delta_0
\end{aligned}
$$

  and apply the Kolmogorov inequality after the centering by expectation and scaling by $\sqrt{n}$.

- Asymptotic Distribution:

  1. that of $\hat{\tau}$ is non-regular
  2. Oracle property due to the super-consistency of $\hat{\tau}$. That is,

$$\left|\hat{\beta}\left(\hat{\tau}\right)-\hat{\beta}\left(\tau_0\right)\right|=o_p\left({}^{-1/2}\right).$$

- It follows from the FCLT and stoch. equi. that

$$\frac{1}{n}\sum_{t=[n\tau_0]+1}^{[n\hat{\tau}]}x_tx_t'=o_p\left(1\right),\quad\frac{1}{\sqrt{n}}\sum_{t=[n\tau_0]+1}^{[n\hat{\tau}]}x_t\varepsilon_t'=o_p\left(1\right)$$

  However, considering $\sum^{\tau}x_ty_t=\sum^{\tau_0}x_t\left(\varepsilon_t+x_t'\beta_1\right)+\sum_{\tau_0}^{\tau}x_t\left(\varepsilon_t+x_t'\beta_2\right),$ we need

$$\frac{1}{\sqrt{n}}\sum_{t=[n\tau_0]+1}^{[n\hat{\tau}]}x_tx_t'\leq\mathrm{E}\left|x_t\right|^2\frac{[n\hat{\tau}]-[n\tau_0]}{\sqrt{n}}+o_p\left(1\right)=o_p\left(1\right),$$

  due to first the FCLT and second the super-consistency.

3.0.1 Local-to-Linear Setup

The model becomes linear asymptotically, that is,

$$\delta_0 = \delta_{0n} = cn^{-\alpha},$$

where $0 < \alpha < \frac{1}{2}$ and $c \neq 0$. We can proceed in the similar way before but this time using the profiled criterion function, which is straightforward since for a give τ, $\left(\hat{\beta}(\tau), \hat{\delta}(\tau)\right)$ is the OLS estimator. The formal proof of the consistency and convergence rate, however, need to invoke a different version of the rate theorem than we learned reflecting the arrary nature of our data generating process. It can be found in Section 3.4 of van der Vaart and Wellner (1996). We skip the verification of the conditions of the theorem here, but it turns out that the proper distance is

$$d(\theta_1, \theta_2) = |\tau_1 - \tau_2|^{\frac{1}{2-4\alpha}} + |\beta_1 - \beta_2| + |\delta_1 - \delta_2|.$$

and $r_n = n^{1/2}$.

- The main motivation to consider this asymptotics was that this results in a slower convergence rate for $\hat{\tau}$, which in turn yields a more tractable asymptotic distribution for $\hat{\tau}$. This turns out to be the minimizer of a Gaussian process. Its density function is also known in the literature.

- For $\tau > \tau_0$, and $\tau = \tau_0 + hn^{-1+2\alpha}$ for some $0 \leq h \leq K < \infty$,

$$\begin{aligned} r_n^2 \left(\frac{1}{n} \sum_{t=[n\tau_0]+1}^{[n\tau]} \left(x_t'\delta_n\right)^2 - 2x_t'\delta_n\varepsilon_t \right) &= n^{-2\alpha} \sum_{t=[n\tau_0]+1}^{[n\tau_0]+hn^{2\alpha}} \left(x_t'c\right)^2 - 2n^{-\alpha} \sum_{t=[n\tau_0]+1}^{[n\tau_0]+hn^{2\alpha}} \left(x_t'c\right)\varepsilon_t \\ &\Rightarrow h\mathrm{E}(x_t'c)^2 - 2\sqrt{E\left(x_t'c\right)^2 \varepsilon_t^2} W\left(h\right). \end{aligned}$$

The convergence follows from the FCLT. Similarly, we can show the weak convergence when $\tau < \tau_0$, where the limit is the minus of the above and the Brownian motion is independent of each other.

3.0.2 Under Misspecification

- Now assume that

  $$y_t = \beta_t + \varepsilon_t,$$

  where $\beta_t = \beta(t/n)$, which is a function on $[0,1]$ and twice continuously differentiable. Here with slight abuse of notation, $\beta_t = \beta(t)$ if t lies between 0 and 1. Then,

  $$\begin{aligned}\mathbb{M}(\theta) &= p\lim_{n\to\infty}\left(\frac{1}{n}\sum_{t=1}^{[n\tau]}\left(\varepsilon_t+\beta_t-\beta_1\right)^2+\frac{1}{n}\sum_{t=[n\tau]+1}^{n}\left(\varepsilon_t+\beta_t-\beta_2\right)^2\right)\\ &= \sigma^2+\int_0^\tau\left(\beta_t-\beta_1\right)^2+\int_\tau^1\left(\beta_t-\beta_2\right)^2dt.\end{aligned}$$

  Observe that $\mathbb{M}$ is differentiable and the FOC is

  $$\begin{aligned}\tau\beta_1 &= \int_0^\tau\beta_t, \quad (1-\tau)\beta_2=\int_\tau^1\beta_t,\\ \left(\beta_\tau-\beta_1\right)^2 &= \left(\beta_\tau-\beta_2\right)^2\end{aligned}$$

and SOC is

$$\begin{bmatrix} \tau & 0 & 2(\beta_1 - \beta_\tau) \\ 0 & (1-\tau) & 2(\beta_\tau - \beta_2) \\ 2(\beta_1 - \beta_\tau) & 2(\beta_\tau - \beta_2) & 2\beta_\tau^{(1)}\left(-\beta_1 + \beta_2\right) \end{bmatrix}.$$

Its determinant is

$$\begin{aligned} &\tau\left(\left(1-\tau\right)2\beta_\tau^{(1)}\left(-\beta_1+\beta_2\right) - 4\left(\beta_\tau - \beta_2\right)^2\right) - 4\left(1-\tau\right)\left(\beta_1 - \beta_\tau\right)^2 \\ &= 4\left[\left(1-\tau\right)\tau\beta_\tau^{(1)}\frac{\left(-\beta_1+\beta_2\right)}{2} - \left(\beta_1 - \beta_\tau\right)^2\right], \end{aligned}$$

at $\tau = \tau_0$, which is assumed to be positive and which is true for some β. This implies that a proper d for the convergence rate computation is the standard Euclidean norm.

- The modulus of continuity for $\sqrt{n}\left(\mathbb{M}_n\left(\theta\right) - \mathbb{M}_n\left(\theta_0\right) - \mathbb{M}\left(\theta\right) + \mathbb{M}\left(\theta_0\right)\right)$ does not change. However, reflecting the change in d, the bound becomes $O\left(\sqrt{\delta}\right)$. This in turn results in the convergence rate of $n^{1/3}$.

3.1 Weakly Dependent Data

- Absolutely regular empirical processes (Doukhan, Massart, & Rio 1995)
  - norm:

    $$\|f\|_{2,\beta} = \left[\int_0^1 \beta^{-1}(u)\left[Q_f(u)\right]^2 du\right]^{1/2},$$

    where β^{-1} denotes the cadlag inverse of the monotone function $\beta(t) = \beta_{[t]}$, which is a β-mixing coefficient such that $\sum_{n\geq 0}\beta_n < \infty$, and Q_f denotes the quantile function of $|f(W_0)|$. This norm is bounded below by L_2-norm and above by a constant multiple of L_2-norm (Lemma 1 & 2).
  - For the entropy with bracketing wrt $\|\cdot\|_{2,\beta}$, $H_\beta(\varepsilon, \mathcal{F})$, and a class of functions satisfying the condition that $\|f\|_{2,\beta} \leq \delta$, $\mathcal{F}_\delta$, let

    $$\rho(\delta) = \int_0^\delta \sqrt{H_\beta(u, \mathcal{F}_\delta)}du.$$

  - Then, (Theorem 3) for some constant A and for all large n

$$\mathrm{E}\left\{\sup_{f\in\mathcal{F}_\delta}\left|\mathbb{G}_n f\right|\right\}\leq A\rho\left(\delta\right).$$

- Remarks

  1. A maximal inequality for a class such as $\acute{\mathcal{F}} = \{f_\tau : |\tau-\tau_0|\leq\delta\}$ can be obtained by carefully comparing $\mathcal{F}_\delta$ with $\acute{\mathcal{F}}$.
  2. An upper bound for $\rho\left(\delta\right)$ can be computed by the entropy with bracketing wrt the L_p-norm, as it is smaller than $\left\|\cdot\right\|_{2,\beta}$.

- Dependent Heterogeneous arrays (Hansen 1996 Econometric Theory)

  - the class of Lipschitz functions
  - various mixingale arrays such as MDS, mixing, NED arrays.
  - stochastic equicontinuity of the empirical process but modulus of continuity

  - the seminorm for the index τ depends on the mixingale norm

Andrews (1993)Andrews and Ploberger (1995)

References

Andrews, D. (1993): “Tests for Parameter Instability and Structural Change with Unknown Change Point,” *Econometrica*, 61, 821–856. 2.2, 3.1

Andrews, D. W. K. (2001): “Testing When a Parameter Is on the Boundary of the Maintained Hypothesis,” *Econometrica*, 69(3), 683–734. 2.2

Andrews, D. W. K., and W. Ploberger (1995): “Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative,” *The Annals of Statistics*, 23(5), 1609–1629. 3.1

Davies, R. (1977): “Hypothesis Testing when a Nuisance parameter is Present Only under the Alternative,” *Biometrika*, 64, 247–254. 2.2

Hansen, B. E. (1996): “Inference when a nuisance parameter is not identified under the null hypothesis,” *Econometrica*, 64, 413–430. 2.2

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Process*. Springer, New York. 3.0.1