Nonparametric estimation of the continuous treatment effect with measurement error

Wei Huang and Zheng Zhang

10 May 2023

Outline

Background and focus of the paper

- Identifying and estimating the causal effect of a treatment or policy from observational studies is of great interest to economics, social science, and public health researchers.
- Early studies focused on whether an individual receives the treatment or not
- More recently, as opposed to such binary treatments, researchers have been investigating the causal effect of a continuously valued treatment, where the effect depends not only on the introduction of the treatment but also on the intensity
- However, all these methods require the treatment data to be measured without errors. This paper propose a broad class of novel and robust nonparametric estimators for the average dose–response function (ADRF) as in Ai et al. (2021)

Basic Setup

- This paper consider a continuously valued treatment in which the observed treatment variable is denoted by T with the probability density function $f_T(t)$ and support $\mathcal{T} \subset \mathbb{R}$.
- $Y^*(t)$ denote the potential outcome if one was treated at level t for $t \in \mathcal{T}$. In practice, each individual can only receive one treatment level T and we only observe the corresponding outcome $Y := Y^*(T)$
- $X \in \mathbb{R}^r$ be a vector of given covariates related to both T and $Y^*(t)$ for $t \in \mathcal{T}$
- Instead of observing T we observe S so that

$$S = T + U \tag{1}$$

assuming error U independent of T, X and $\{Y^*(t)\}_{t \in \mathcal{T}}$ and c.f ϕ_U is known

Basic set up

- Now we have some basic assumptions relevant in most of treatment effect literature.

  (i) *(Unconfoundedness)* $T \perp \{Y^*(t)\}_{t \in \mathcal{T}} | X$

  (ii) *(No Intereference)* for $i = 1(1)N$ outcome of individual i is not affected by treatment assigned to any other individual i.e $Y_i^*(T_i, \boldsymbol{T}_{(-i)}) = Y_i^*(T_i, \boldsymbol{T}'_{(-i)})$ for any $\boldsymbol{T}_{(-i)}, \boldsymbol{T}'_{(-i)}$, where $Y_i^*(T_i, T_{(-i)})$ is the potential outcome of individual i given the treatment T_i and others with $\boldsymbol{T}_{(-i)} := (T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_N)$, respectively.

  (iii) *(Consistency)* $Y = Y^*(t)$ a.s. if $T = t$

  (iv) *(Positivity)* $f_{T|X}(t|X) > 0$ a.s. for all $t \in \mathcal{T}$.

Unconditional ADRF

- As mentioned earlier the goal of paper is to estimate unconditional ADRF i.e $\mu(t) = \mathbb{E}\{Y^*(t)\}$ for fixed $t \in \mathbb{T}$ based on i.i.d $\{S_i, X_i, Y_i\}_{i=1}^{N}$
- Under following assumptions one can write $\mu(t)$ as

$$
\begin{aligned}
\mu(t) =& \mathbb{E}[\mathbb{E}\{Y^*(t)|X\}] \\
=& \mathbb{E}\{Y^*(t)|X, T = t\}] \\
=& \mathbb{E}\{Y|X, T = t\}] \\
=& \int_{\mathcal{X}} \int_{\mathcal{Y}} \frac{f_T(t)}{f_{T|X}(t|x)} y f_{Y|X,T}(y|x,t) f_{X|T}(x|t) dy dx \\
=& \mathbb{E}\{\pi_0(t, X) Y | T = t\}
\end{aligned}
\tag{2}
$$

where,

$$
\pi_0(t, X) = \frac{f_T(t)}{f_{T|X}(t|x)} \tag{3}
$$

Unconditional ADRF

- **GOOD NEWS** If T is fully observable and $\pi_0(t, x)$ is known a well known consistent estimator for $\mu(t)$ is Nadaraya-Watson estimator i.e

$$\mu_{NW} = \frac{\sum_{i=1}^{N} \pi_0(t, X_i) Y_i L(\{h^{-1}(t - T_i)\})}{\sum_{i=1}^{N} L(\{h^{-1}(t - T_i)\})} \quad t \in \mathcal{T} \tag{4}$$

  where L is univariate Kernel function $\int_{-\infty}^{\infty} L(x)dx = 1$,and h is bandwidth

- **BAD NEWS** Unfortunately we do not observe T but observe S and also $\pi_0(t, x)$ is unknown in practice.

De-Convolution Kernel Approach

- Clearly from Fourier's inversion theorem we have

$$f_T(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iwt}\phi_T(w)dw = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iwt}\frac{\phi_S(w)}{\phi_U(w)}dw \tag{5}$$

  This inspired Stefinski and Caroll[1990] to estimate $f_T(t)$ by $\hat{f}_{T,h}(t) = (Nh)^{-1}\sum_{i=1}^{N} L_U\{(t-S_i)/h\}$ where,

$$L_U(v) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iwv}\frac{\phi_L(w)}{\phi_U(w/h)} \tag{6}$$

- Based on this idea Fan and Troung [1993] proposed a consistent errors-in-variable regression estimator replacing the $L\{(t-T_i)/h\}$ with $L_U\{(t-S_i)/h\}$ i,e

$$\tilde{\mu}(t) = \frac{\sum_{i=1}^{N} \pi_0(t, X_i)Y_i L_U\{h^{-1}(t-S_i)\}}{\sum_{i=1}^{N} L_U\{h^{-1}(t-S_i)\}} \tag{7}$$

- In fact we have,

$$\mathbb{E}[L_U(\{h^{-1}(t-S)\}|T, X, Y)] = L\{h^{-1}(t-T)/h\} \tag{8}$$

Estimation of $\pi_0(t, X)$

- **Bad News?** The straightforward estimate of π_0 by plug in is subject to sensitivity on low values of $f_{T|X}$ since small errors in estimating $f_{T|X}$ lead to large error. So we will treat π_0 as a whole
- if T_i are fully observable from Ai et al[2021]

$$\mathbb{E}\{\pi_0(t, X)u(X)v(T)\} = \mathbb{E}\{v(T)\}\mathbb{E}\{u(X)\} \tag{9}$$

  holds for any integrable function $u(X)$ and $v(T)$
- **Bad News?** non parametric estimation of $\mathbb{E}\{v(T)\}$ from contaminated data $\{S_i\}_{i=1}^N$ is difficult if not impossible.
- **Solution** We will estimate the projection $\pi_0(t, .) : \mathcal{X} \to \mathbb{R}$ for fixed $t \in \mathcal{T}$. We see

$$\mathbb{E}\{\pi_0(t, X)u(X)|T = t\} = \int_{\mathcal{X}} \frac{f_T(t)}{f_{T|X}(t|x)} u(x) f_{X|T}(x|t)dx = \mathbb{E}\{u(X)\} \tag{1}$$

  holds for any integrable $u(X)$

Estimation of $\pi_0(t, X)$

Theorem 3.1 for fixed $t \in \mathcal{T}$ and any integrable function $u(X)$,

$$\lim_{h_0 \to 0} \frac{\mathbb{E}[\pi(t, X)u(X)L_U\{h_0^{-1}(t - S)\}]}{\mathbb{E}[L_U\{h_0^{-1}(t - S)\}]} = \mathbb{E}[u(X)] \tag{11}$$

if and only if $\pi(t, X) = \pi_0(t, X)$ a.s

- We can solve a sample analogue of (11) for any integrable function $u(X)$ where $h_0 \to 0$ as $N \to \infty$ but this need solving infinite number of equations which is impractical
- **Solution** We can approximate infinite dimensional space of $u(x)$ using a finite known basis function with dimension K(power series, B-splines etc) i.e $u_K(X) = (u_{K,1}(X), \ldots, u_{K,K}(X))^T$ which approximates any $u(X)$ as $K \to \infty$. Now,

$$\lim_{h_0 \to 0} \frac{\mathbb{E}[\pi(t, X)u_K(X)L_U\{h_0^{-1}(t - S)\}]}{\mathbb{E}[L_U\{h_0^{-1}(t - S)\}]} = \mathbb{E}[u_K(X)] \tag{12}$$

Estimation of $\pi_0(t, X)$

Let $\rho(.)$ be a globally concave and increasing function. Define a strictly concave function

$$G_t^*(\lambda) = \lim_{h_0 \to 0} \frac{\mathbb{E}[\rho\{\lambda^T u_K(X)\} L_U\{h_0^{-1}(t-S)\}]}{\mathbb{E}[L_U\{h_0^{-1}(t-S)\}]} - \lambda^T \mathbb{E}[u_K(X)]$$

and $\lambda_t^* = \text{argmax}_{\lambda \in \mathbb{R}^k} G_t^*(\lambda)$ Hence $\nabla G_t^*(\lambda_t^*) = 0$ which implies (12) holds if

$$\pi^*(t, X) = \rho'\{\lambda_t^{*T} u_K(X)\} \tag{13}$$

Hence this paper proposes

$$\hat{\pi}(t, X) = \rho'\{\hat{\lambda}_t^T u_K(X)\} \tag{14}$$

with $\hat{\lambda}_t = \text{argmax}_{\lambda \in \mathbb{R}^k} \hat{G}_t(\lambda)$

$$\hat{G}_t(\lambda) = \lim_{h_0 \to 0} \frac{\sum_{i=1}^N \rho\{\lambda^T u_K(X)\} L_U\{h_0^{-1}(t-S_i)\}}{\sum_{i=1}^N L_U\{h_0^{-1}(t-S_i)\}} - \lambda^T\{N^{-1} \sum_{i=1}^N u_K(X)\} \tag{15}$$

Equivalent Duel Solution interpretation

The proposed estimator $\hat{\pi}(t, X)$ is also a duel solution to a local generalised empirical likelihood maximization problem: for fixed $t \in \mathcal{T}$,

$$\begin{cases} \max_{\{\pi_i\}_{i=1}^N} -\frac{\sum_{i=1}^N D(\pi_i) L_U(\{t-S_i\}/h_0)}{\sum_{i=1}^N L_U(\{t-S_i\}/h_0)} \\ \text{subject to } \frac{\sum_{i=1}^N \pi_i u_K(X_i) L_U(\{t-S_i\}/h_0)}{\sum_{i=1}^N L_U(\{t-S_i\}/h_0)} = \frac{1}{N}\sum_{i=1}^N u_K(X_i), \end{cases}$$

where $D(v)$ is a distance measure from v to 1 for $v \in \mathbb{R}$, which is continuously differentiable and satisfies that $D(1) = 0$ and

$$\rho(-v) = D\{(D')^{-1}(v)\} - v \cdot (D')^{-1}(v).$$

Now once we get $\hat{\pi}(t, X)$ we obtain estimator of $\mu(t)$:

$$\hat{\mu}(t) = \frac{\sum_{i=1}^N \hat{\pi}(t, X_i) Y_i L_U\{(t - S_i)/h_0\}}{\sum_{i=1}^N L_U\{(t - S_i)/h\}} = \frac{1}{N}\sum_{i=1}^N u_K(X_i) \tag{17}$$

Proof of Theorem 3.1

We first prove that for every fixed $t \in \mathcal{T}$ and any integrable function $u(X)$, $\mathbb{E}\{\pi(t, X)u(X) \mid T = t\} = \mathbb{E}\{u(X)\}$ holds if and only if $\pi(t, X) = \pi_0(t, X)$ a.s.. The sufficient part is obvious and we here show the necessary part. Since for all $t \in \mathcal{T}$ and any integrable function $u(\boldsymbol{X})$, we have $\mathbb{E}\{\pi(t, X)u(X) \mid T = t\} = \mathbb{E}\{u(X)\}$, comparing to (11), we see that

$$\mathbb{E}\left[\{\pi(t, X) - \pi_0(t, X)\}\, u(X) \mid T = t\right] = 0$$

for all $t \in \mathcal{T}$ and any integrable function $u(X)$. Taking $u(X) = \exp\left(a^T X\right)$ for $a \in \mathbb{R}^r$, we have

$$\mathbb{E}\left[\{\pi(t, X) - \pi_0(t, X)\} \exp\left(a^T X\right) \mid T = t\right] = 0$$

for all $a \in \mathbb{R}^r$. Thus, according to the uniqueness of Laplace transform, we have that $\pi(t, \cdot) = \pi_0(t, \cdot)$ a.s.. Next, we show that

$$\lim_{h_0 \to 0} \frac{\mathbb{E}\left[\pi(t, X)u(X)L_U\left\{(t - S)/h_0\right\}\right]}{\mathbb{E}\left[L_U\left\{(t - S)/h_0\right\}\right]} = \mathbb{E}\{\pi(t, X)u(X) \mid T = t\}$$

Proof of Theorem 3.1

Note,

$$
\begin{aligned}
&\lim_{h_0\to 0} \frac{\mathbb{E}\left[\pi(t,X)u(X)L_U\left\{(t-S)/h_0\right\}\right]}{\mathbb{E}\left[L_U\left\{(t-S)/h_0\right\}\right]} \\
=&\lim_{h_0\to 0} \frac{\mathbb{E}\left(\pi(t,X)u(X)\mathbb{E}\left[L_U\left\{(t-S)/h_0\right\} \mid T,X\right]\right)}{\mathbb{E}\left(\mathbb{E}\left[L_U\left\{(t-S)/h_0\right\} \mid T,X\right]\right)} \\
=&\lim_{h_0\to 0} \frac{\mathbb{E}\left(\pi(t,X)u(X)\mathbb{E}\left[L_U\left\{(t-S)/h_0\right\} \mid T\right]\right)}{\mathbb{E}\left(\mathbb{E}\left[L_U\left\{(t-S)/h_0\right\} \mid T\right]\right)} \quad (S \perp X \mid T) \\
=&\lim_{h_0\to 0} \frac{h_0^{-1}\mathbb{E}\left[\pi(t,X)u(X)L\left\{(t-T)/h_0\right\}\right]}{h_0^{-1}\mathbb{E}\left[L\left\{(t-T)/h_0\right\}\right]} \quad (\text{by } (8)).
\end{aligned}
$$

Proof of Theorem 3.1

For the numerator, we have

$$
\begin{aligned}
&\lim_{h_0\to 0} h_0^{-1}\mathbb{E}\left[\pi(t,X)u(X)L\left\{(t-T)/h_0\right\}\right] \\
=&\lim_{h_0\to 0} h_0^{-1}\iint \pi(t,x)u(x)L\left\{\left(t-t'\right)/h_0\right\} f_{T,X}\left(t',x\right)dt'dx \\
=&-\lim_{h_0\to 0}\iint \pi(t,x)u(x)L(z)f_{T,X}\left(t-zh_0,x\right)dzdx \\
=&-\iint \pi(t,x)u(x)L(z)f_{T,X}(t,x)dzdx \\
=&\mathbb{E}\{\pi(t,X)u(X)\mid T=t\}\cdot f_T(t).
\end{aligned}
$$

Similarly, we have

$$
\lim_{h_0\to 0} h_0^{-1}\mathbb{E}\left[L\left\{(t-T)/h_0\right\}\right]=f_T(t).
$$

The results then follows.

Assumptions

- **Assumption 2** The kernel function $L(\cdot)$ is an even function such that $\int_{-\infty}^{\infty} L(u)\mathrm{d}u = 1$ and has finite moments of order 3 .
- **Assumption 3** Assume
  - (i) the support $\mathcal{X}$ of $\boldsymbol{X}$ is a compact subset of $\mathbb{R}^r$. The support $\mathcal{T}$ of the treatment variable T is a compact subset of $\mathbb{R}$.
  - (ii) (Strict Positivity) there exist a positive constant $\eta_{\min}$ such that $f_{T|X}(t \mid x) \geq \eta_{\min} > 0$, for all $x \in \mathcal{X}$.
- **Assumption 4**
  - (i) The densities $f_T(t), f_{T|X}(t \mid \boldsymbol{X})$, and $f_{T|Y,X}(t \mid Y, X)$ are third-order continuously differentiable w.r.t. t almost surely.
  - (ii) The derivatives of $f_{T|X}(t \mid X)$ and $f_{T|Y,X}(t \mid Y, X)$, denoted by $\{\partial_t^d f_{T|X}(t \mid X), \partial_t^d f_{T|Y,X}(t \mid Y, X)$ for $d = 0, 1, 2, 3\}$, are integrable almost surely in t.
- **Assumption 5** For every $t \in \mathcal{T}$,
  - (i) the function $\pi_0(t, x)$ is s-times continuously differentiable w.r.t. $x \in \mathcal{X}$, where $s > r/2$ is an integer;
  - (ii) there exist $\lambda_t \in \mathbb{R}^K$ and a positive constant $\alpha > 0$ such that $\sup_{x\in\mathcal{X}} \left| (\rho')^{-1} \{\pi_0(t, x)\} - \lambda_t^\top u_K(x) \right| = O\left(K^{-\alpha}\right)$.

Assumptions

- **Assumption 6**
  - (i) For every K, the eigenvalues of $\mathbb{E}\left[u_K(\boldsymbol{X})u_K(\boldsymbol{X})^\top \mid T=t\right]$ are bounded away from zero and infinity, and twice differentiable w.r.t. t for $t \in \mathcal{T}$.
  - (ii) There is a sequence of constants $\zeta(K)$ satisfying $\sup_{x\in\mathcal{X}} \|u_K(\boldsymbol{x})\| \leq \zeta(K)$, such that $\zeta(K)\left\{K^{-\alpha} + h_0^2 + h^2\right\} \to 0$ as $N \to \infty$, where $\|\cdot\|$ denotes the Euclidean norm.
- **Assumption 7**
  - For every $t \in \mathcal{T}$, there exist $\gamma_t \in \mathbb{R}^K$ and a positive constant $\ell > 0$ such that $\sup_{x\in\mathcal{X}} \left|m(t,x) - \gamma_t^\top u_K(x)\right| = O\left(K^{-\ell}\right)$, where $m(t,x) = \mathbb{E}[Y \mid T=t, X=x]$.
- **Assumption 8** Following are bounded for some $\delta > 0$, for all $t \in \mathcal{T}$.
  - $R_1^{2+\delta}(t) = \mathbb{E}[|\pi_0(t,X)Y - \mu(t)|^{2+\delta} \mid T=t]$
  - $R_2^{2+\delta}(t) = \mathbb{E}[\pi_0(t,X)m(t,X) - \mu(t)|^{2+\delta} \mid T=t]$ and
  - $R_3^{2+\delta}(t) := \mathbb{E}[|\pi_0(t,X)\{Y - m(t,X)\}|^{2+\delta} \mid T=t]$

Ordinary Smooth and Super Smooth error

- An ordinary smooth error of order $\beta \geq 1$ satisfies,

$$\lim_{t\to\infty} t^{\beta}\phi_U(t) = c \text{ and } \lim_{t\to\infty} t^{\beta+1}\phi_U^{(1)}(t) = -c\beta \qquad (18)$$

  for some $c > 0$

- A super smooth error of order $\beta \geq 1$ satisfies,

$$d_0|t|^{\beta_0}e^{-|t|^{\beta}/\gamma} \leq \phi_U(t) \leq d_1|t|^{\beta_1}e^{-|t|^{\beta}/\gamma} \text{ as } |t| \to \infty \qquad (19)$$

  for positive constants d_0, d_1, γ and constants β_0 and β_1

- **Assumption O** (Ordinary Smooth Case)
  - $\|\phi_L\|_\infty < \infty$, $\int_{-\infty}^{\infty} |t|^{\beta+1} \{|\phi_L(t)| + |\partial_t \phi_L(t)|\} \mathrm{d}t < \infty$ and $\int_{-\infty}^{\infty} \left|t^{\beta}\phi_L(t)\right|^2 \mathrm{d}t < \infty$.
- **Assumption S** (Super Smooth Case)
  - $\phi_L(t)$ is support on $[-1, 1]$ and bounded.

Asymptotics for estimated weight function

Theorem 4.1 Suppose that the error U is ordinary smooth of order β satisfying (18) and that Assumption O holds. Under Assumptions 2-6 and $\zeta(K)\{K/(Nb_0^{1+2\beta})\}^{1/2} \to 0$ as $N \to \infty$, for every fixed $t \in \mathcal{T}$, then

$$\sup_{x\in\mathcal{X}} |\widehat{\pi}(t,x) - \pi_0(t,x)| = O_p\left(\zeta(K)\{K^{-\alpha} + b_0^2\} + \zeta(K)\left\{\frac{K}{Nb_0^{1+2\beta}}\right\}^{1/2}\right),$$

$$\int_{\mathcal{X}} |\widehat{\pi}(t,x) - \pi_0(t,x)|^2 \, \mathrm{d}F_X(x) = O_p\left(\{K^{-2\alpha} + b_0^4\} + \frac{K}{Nb_0^{1+2\beta}}\right)$$

$$\frac{1}{N}\sum_{i=1}^{N} |\widehat{\pi}(t,X_i) - \pi_0(t,X_i)|^2 = O_p\left(\{K^{-2\alpha} + b_0^4\} + \frac{K}{Nh_0^{1+2\beta}}\right).$$

Asymptotics for estimated weight function

Theorem 4.3 Suppose that the error U is supersmooth of order β satisfying (19) and Assumption S holds. Under Assumptions $2-6$ and $\zeta^2(K)K\cdot(Nh_0)^{-1}.\ e^{(2h_0^{-\beta}/\gamma)}\to 0$ as $N\to\infty$, for every fixed $t\in\mathcal{T}$, then

$$\sup_{x\in\mathcal{X}}|\widehat{\pi}(t,x)-\pi_0(t,x)|=O_p\left(\zeta(K)\left[\{K^{-\alpha}+b_0^2\}+\frac{e^{(h_0^{-\beta}/\gamma)}}{\sqrt{h_0}}\cdot\sqrt{\frac{K}{N}}\right]\right),$$

$$\int_{\mathcal{X}}|\widehat{\pi}(t,x)-\pi_0(t,x)|^2\,\mathrm{d}F_X(x)=O_p\left(\{K^{-2\alpha}+b_0^4\}+\frac{e^{(2h_0^{-\beta}/\gamma)}}{h_0}\cdot\frac{K}{N}\right),$$

$$\frac{1}{N}\sum_{i=1}^{N}|\widehat{\pi}(t,X_i)-\pi_0(t,X_i)|^2=O_p\left(\{K^{-2\alpha}+b_0^4\}+\frac{\exp(2h_0^{-\beta}/\gamma)}{h_0}\cdot\frac{K}{N}\right)$$

Asymptotics for unconditional ADRF

Let us define for $i = 1, \ldots N$,
$\eta_{h,h_0}(S_i, X_i, Y_i; t) = \phi_h(S_i, X_i, Y_i; t) + \psi_{h_0}(S_i, X_i, Y_i; t)$ where,

- $\phi_h(S_i, X_i, Y_i; t) = [\pi_0(t, X_i)Y_i L_{U,h}(t - S_i) - \mathbb{E}\{\pi_0(t, X)Y L_{U,h}(t - S)\}] - \mu(t)[L_{U,h}(t - S_i) - \mathbb{E}\{L_{U,h}(t - S)\}]$
- $\psi_{h_0}(S_i, X_i, Y_i; t) = \mu(t)[L_{U,h_0}(t - S_i) - \mathbb{E}\{L_{U,h_0}(t - S)\}] - [m(t, X_i)\pi_0(t, X_i)L_{U,h_0}(t - S_i) - \mathbb{E}\{m(t, X_i)\pi_0(t, X)L_{U,h}(t - S)\}]$
- $L_{U,h}(v) = h^{-1}L_U(v/h)$
- $V_j = f_T^{-2}(t)(R_j^2 f_T) * f_U(t).C$ for $j = 1, 2$
- $C = \int_{-\infty}^{\infty} J^2(v)dv = (2\pi c^2)^{-1} \int |w|^{2\beta}\phi_L^2(w)dw$
- R_1^2, R_2^2 are defined as in Assumption 8 also, $(R_1 R_2)(t) = \mathbb{E}[\{\pi_0(t, X)Y - \mu(t)\}\{\mu(t) - \pi_0(t, X)m(t, X)\}|T = t]$ and $v_h(t) = \mathbb{E}\{L_{U,h}^2(t - S)\}$

Asymptotics for unconditional ADRF

Theorem 4.2 Suppose that the error U is ordinary smooth of order β satisfying (18) and Assumption O holds. Under Assumptions $1-8$ and $\frac{(K^{-l}+h_0^2)(K^{-\alpha}+h_0^2)}{h^2} + \frac{(h\wedge h_0)^{1+2\beta}}{h_0^{1+2\beta}}\frac{K}{\sqrt{N}} \to 0$, then where

$$\begin{aligned}
\hat{\mu}_t - \mu_t =& \frac{\kappa_{21}}{2}\Big[\frac{f_T(t)\Phi_1(t) - \mu(t)\partial_t^2 f_T(t)}{f_T(t)}\Big]h^2 + o(h^2) \\
&+ \frac{\kappa_{21}}{2}\Big[\frac{\mu(t)\partial_t^2 f_T(t) - f_T(t)\Phi_2(t)}{f_T(t)}\Big]h_0^2 + o(h_0^2) \\
&+ \sum_{i=1}^{N}\frac{\eta_{h,h_0}(S_i, X_i, Y_i; t)}{N f_T(t)} + o_P\Bigg(\frac{1}{\sqrt{N(h\wedge h_0)^{1+2\beta}}}\Bigg)
\end{aligned}$$

where,$\kappa_{21} = \int u^2 L(u)du$, $\Phi_1(t) = \mathbb{E}[\{Y\partial_t^2 f_{T|X}(t|x)\}/\{f_{T|X}(t|X)\}]$ and $\Phi_2(t) = \mathbb{E}[\{m(t,X)\partial_t^2 f_{T|X}(t|X)\}/\{f_{T|X}(t|X)\}]$

Asymptotics for unconditional ADRF

Infact,

- if $h = o(h_0)$ then $\sqrt{h^{1+2\beta}/N} \sum_{i=1}^{N} \frac{\eta_{h,h_0}(S_i,X_i,Y_i;t)}{f_T(t)} \to N(0, V_1)$
- if $h_0 = o(h)$ then $\sqrt{h_0^{1+2\beta}/N} \sum_{i=1}^{N} \frac{\eta_{h,h_0}(S_i,X_i,Y_i;t)}{f_T(t)} \to N(0, V_2)$
- if $h_0 = \tilde{c}h$ then $\sqrt{h^{1+2\beta}/N} \sum_{i=1}^{N} \frac{\eta_{h,h_0}(S_i,X_i,Y_i;t)}{f_T(t)} \to N(0, V_3)$ where,

$$V_3 = \frac{(R_1^2 f_T) * f_U(t)}{f_T^2(t)} \int_{-\infty}^{\infty} J^2(v)dv + \frac{(R_2^2 f_T) * f_U(t)}{\tilde{c}^{2+2\beta} f_T^2(t)} \int_{-\infty}^{\infty} J^2(v/\tilde{c})dv + \frac{2(R_1 R_2 f_T) * f_U(t)}{\tilde{c}^{1+\beta} f_T^2(t)} \int_{-\infty}^{\infty} J(v)J(v/\tilde{c})dv$$

As long as $(K^{-l} + h_0)(K^{-\alpha} + h_0) = o(h^2)$, the error arising from the sieve approximation is asymptotically negligible. For example, $\hat{\mu}(t) - \mu(t)$ achieves the optimal convergence rate, $N^{-2/(2\beta+5)}$, if $h_0 \sim h \sim N^{-1/(2\beta+5)}$.Also we need $K = o(h^{-2}), \alpha + l > 1$ and

- $\alpha > 1/2$ if spline basis is used
- $\alpha > 1$ if a power series is used

Asymptotics for unconditional ADRF

Theorem 4.4 Suppose that the error U is super smooth of order β satisfying (19) and Assumption S holds.Letting $e(h) = h^{1/2}e^{-\frac{h^{-\beta}}{\gamma}}$ we have $v_h(t) = O(e^{-2}(h))$.If Assumptions $1-8$ holds and for $h \to \infty, v_h(t) \to \infty$ and $\frac{(K^{-l}+h_0^2)(K^{-\alpha}+h_0^2)}{h^2} + \frac{1}{\{e(h)\wedge e(h_0)\}}\frac{K}{\sqrt{N}} \to 0$ as $N \to \infty$, then for fixed t

$$\begin{aligned}
\hat{\mu}_t - \mu_t =& \frac{\kappa_{21}}{2}\Big[\frac{f_T(t)\Phi_1(t) - \mu(t)\partial_t^2 f_T(t)}{f_T(t)}\Big]h^2 + o(h^2) \\
&+ \frac{\kappa_{21}}{2}\Big[\frac{\mu(t)\partial_t^2 f_T(t) - f_T(t)\Phi_2(t)}{f_T(t)}\Big]h_0^2 + o(h_0^2) \\
&+ \sum_{i=1}^{N}\frac{\eta_{h,h_0}(S_i, X_i, Y_i; t)}{N f_T(t)}\{1 + o_P(1)\}
\end{aligned}$$

Asymptotics for unconditional ADRF

- $\sum_{i=1}^{N} \frac{\eta_{h,h_0}(S_i,X_i,Y_i;t)}{N f_T(t)} = O_P(N^{-1/2}\{e(h) \wedge e(h_0)\}^{-1})$

Moreover, if $v_h(t) \geq d_1 f_S(t) h^{d_3} e^{(2h^{-\beta}/\gamma - d_2 h^{-d_4\beta})}$ for constants $d_2 > 0$, $0 < d_4 < 1$ and d_1, d_3 we have

- $[\mathrm{var}\{\eta_{h,h_0}(S_i, X_i, Y_i; t)\}]^{-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^{N}\{\eta_{h,h_0}(S_i, X_i, Y_i; t)\} \to N(0,1)$

As in the ordinary smooth case,as long as $\pi_0(t,)$is sufficiently smooth or K grows sufficiently fast,and h_0 decays fast enough, the sieve approximation error of our estimator is asymptotically negligible and the dominating bias term is the same as that in the ordinary smooth case.

- From the theorem, when $h \sim h_0$ and $h \wedge h_0 = d(\log N)^{-1/\beta}$ for a constant $d > (2/\gamma)^{1/\beta}$, one finds that the rate of variance, $N^{-1}\{e(h) \wedge e(h_0)\}^{-2} = o(h^4 + h_0^4)$, is negligible compared to the asymptotic bias, and the convergence rate of $\hat{\mu}(t) - \mu(t)$ is $(\log N)^{-2/\beta}$.

Preliminaries

Lets assume K and h_0 are given,then CV criteria to this context to choose h, we see $CV(h) = \sum_{i=1}^{N}\{\hat{\pi}(T_i, X_i)Y_i - \hat{\mu}^{-i}(T_i)\}^2 w(T_i)$ is difficult to compute since we do not observe T. To tackle this,

- generate two additional sets of contaminated data,$S^*_{i,d} = S_i + U^*_{id}$ and $S^{**}_{i,d} = S^*_{i,d} + U^{**}_{i,d}$, with $d = 1(1)D$ large and i.i.d $U^{**}_{i,d}, U^{**}_{i,d} \sim U$
- Based on this we first obtain $(\hat{\pi}^*_d, \hat{\mu}^*_d)$ and then $(\hat{\pi}^{**}_d, \hat{\mu}^{**}_d)$
- compute $\hat{h}^*$ and $\hat{h}^{**}$ which minimises $\sum_{i=1}^{D} CV^*_d(h)/D$ and $\sum_{i=1}^{D} CV^{**}_d(h)/D$ respectively
- Intuitively, we then expect the relationship between $\hat{h}^*$ and our target bandwidth h to be similar to that between $\hat{h}^*$ and $\hat{h}^{**}$. Thus, the authors considered that $h/\hat{h}^* \approx \hat{h}^*/\hat{h}^{**}$ and used a linear back-extrapolation procedure that would give the bandwidth $\hat{h}_{DH} = (\hat{h}^*)^2/\hat{h}^{**}$

Two-step procedure and local constant extrapolation

Extending the SIMEX method to choose three parameters simultaneously is computationally unstable thus the paper proposes

- Plug-in bandwidth h_{PI} for the kernel de convolution estimator with bandwidth h of the density of T proposed by Delaigle and Gijbels (2002) minimises the asymptotic MSE of the estimator,whose bias is of rate of h^2 and standard deviation $\sqrt{v_h(t)/N}$
- set $h_0 = h_{PI}$ and choose $K = \lfloor \tilde{c} h_{PI}^{-2} \log(h_{PI} + 1) \rfloor$ such that $K \geq 2$, where $\tilde{c}$ minimises the generalised CV criterion (Craven and Wahba, 1978)
- The linear back-extrapolation sometimes gave highly unstable results as which extrapolant function should be used in practice is unknown hence they used local constant estimator(Fan and Gijbels,1996), with $h_d^* = c_d^* h_{PI}$ and $h_d^{**} = c_d^{**} h_{PI}$ where c_d^*, c_d^{**} minimise $CV_d^*(h)$ and $CV_d^{**}(h)$,respectively and we choose $\hat{h} = \sum_{d=1}^{D} h_d^* \phi\{(\hat{h}^* - h_d^{**})/b\} / \sum_{d=1}^{D} \phi\{(\hat{h}^* - h_d^{**})/b\}$

Thank You