Nonparametric Modal Regression in the Presence of Measurement Error

Xianzheng Huang

Department of Statistics, University of South Carolina

August 15, 2016
(Joint work with Zhou, Haiming)

Problem of interest

- Modal model: For $x \in \mathscr{X}$,

$$\begin{aligned}\text{Mode}(Y|X=x) &= M(x) \\ &= \{y \in \mathscr{Y} : p_y(y|x) = 0,\ p_{yy}(y|x) < 0\}.\end{aligned}$$

Problem of interest

- Modal model: For $x \in \mathscr{X}$,

$$\text{Mode}(Y|X=x) = M(x) = \{y \in \mathscr{Y} : p_y(y|x) = 0,\ p_{yy}(y|x) < 0\}.$$

- Measurement error model:

$$W_j = X_j + U_j,\ \ j = 1, \ldots, n,$$

— $W_j \sim f_W(w)$: the observed covariate,
— $X_j \sim f_X(x)$: the underlying true covariate,
— $U_j \sim f_U(u)$: the measurement error (m.e.),
$U_j \perp (X_j, Y_j)$.

Problem of interest

- Modal model: For $x \in \mathscr{X}$,

$$\text{Mode}(Y|X=x) = M(x) = \{y \in \mathscr{Y} : p_y(y|x) = 0,\ p_{yy}(y|x) < 0\}.$$

- Measurement error model:

$$W_j = X_j + U_j,\ \ j = 1, \ldots, n,$$

— $W_j \sim f_W(w)$: the observed covariate,
— $X_j \sim f_X(x)$: the underlying true covariate,
— $U_j \sim f_U(u)$: the measurement error (m.e.),
$U_j \perp (X_j, Y_j)$.

- Goal: Estimate $M(x)$ based on $\{(Y_j, W_j),\ j = 1, \ldots, n\}$.

Road map of the talk

- Motivation and relevant literature
- Methodology
- Asymptotic properties
- Implementation
- To be resolved

Motivation: Speed-flow data

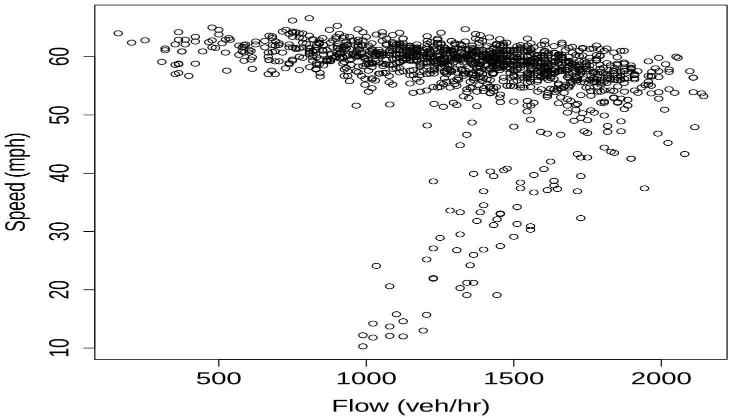


Fig. 1: Speed-flow diagram

Motivation: Temperature data

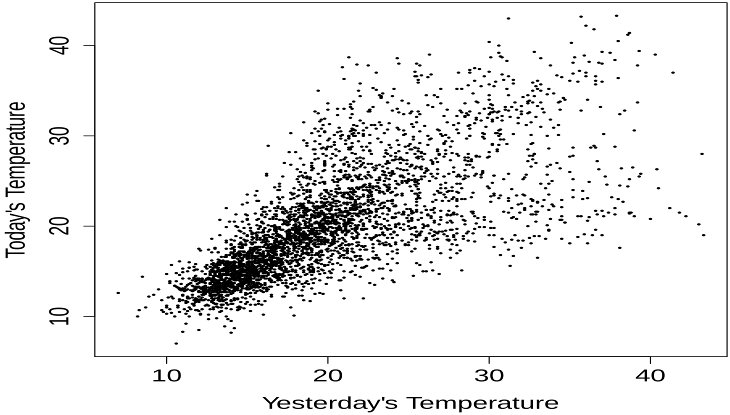


Fig. 2: Each day's temperature vs. the previous day's.

Relevant literature

- Einbeck & Tutz (2006): Base on kernel density estimation.

Relevant literature

- Einbeck & Tutz (2006): Base on kernel density estimation.
- Yao, Lindsay & Li (2012): Relate to local polynomial estimation for the mean function.

Relevant literature

- Einbeck & Tutz (2006): Base on kernel density estimation.
- Yao, Lindsay & Li (2012): Relate to local polynomial estimation for the mean function.
- Yao & Li (2014): Assume Mode$(Y|X = x) = \beta_0 + \beta_1 x$.

Relevant literature

- Einbeck & Tutz (2006): Base on kernel density estimation.
- Yao, Lindsay & Li (2012): Relate to local polynomial estimation for the mean function.
- Yao & Li (2014): Assume Mode$(Y|X = x) = \beta_0 + \beta_1 x$.
- Chen, Genovese, Tibshirani & Wasserman (2016): Nonparametric modal regression (relating to mixture regression, clustering, ridge regression)

Relevant literature

- Einbeck & Tutz (2006): Base on kernel density estimation.
- Yao, Lindsay & Li (2012): Relate to local polynomial estimation for the mean function.
- Yao & Li (2014): Assume Mode$(Y|X = x) = \beta_0 + \beta_1 x$.
- Chen, Genovese, Tibshirani & Wasserman (2016): Nonparametric modal regression (relating to mixture regression, clustering, ridge regression)
- In the presence of m.e.: ???

Nonparametric modal regression accounting for m.e.

(I) Local constant estimator of $M(x)$:

$\hookrightarrow$ Construct the local constant **k**ernel **d**ensity **e**stimator (KDE) of $p(x, y)$;

(II) Local linear estimator of $M(x)$:

Nonparametric modal regression accounting for m.e.

(I) Local constant estimator of $M(x)$:

- $\hookrightarrow$ Construct the local constant **k**ernel **d**ensity **e**stimator (KDE) of $p(x, y)$;
- $\hookrightarrow$ obtain an estimate of $p_y(x, y)$, $\hat{p}_y(x, y)$;

(II) Local linear estimator of $M(x)$:

Nonparametric modal regression accounting for m.e.

(I) Local constant estimator of $M(x)$:

$\hookrightarrow$ Construct the local constant **k**ernel **d**ensity **e**stimator (KDE) of $p(x, y)$;

$\hookrightarrow$ obtain an estimate of $p_y(x, y)$, $\hat{p}_y(x, y)$;

$\hookrightarrow$ find solutions to $\hat{p}_y(x, y) = 0$.

(II) Local linear estimator of $M(x)$:

Nonparametric modal regression accounting for m.e.

(I) Local constant estimator of $M(x)$:

- $\hookrightarrow$ Construct the local constant **k**ernel **d**ensity **e**stimator (KDE) of $p(x, y)$;
- $\hookrightarrow$ obtain an estimate of $p_y(x, y)$, $\hat{p}_y(x, y)$;
- $\hookrightarrow$ find solutions to $\hat{p}_y(x, y) = 0$.

(II) Local linear estimator of $M(x)$:

- $\hookrightarrow$ Construct the local linear KDE of $p(y|x)$;

Nonparametric modal regression accounting for m.e.

(I) Local constant estimator of $M(x)$:

- $\hookrightarrow$ Construct the local constant **k**ernel **d**ensity **e**stimator (KDE) of $p(x, y)$;
- $\hookrightarrow$ obtain an estimate of $p_y(x, y)$, $\hat{p}_y(x, y)$;
- $\hookrightarrow$ find solutions to $\hat{p}_y(x, y) = 0$.

(II) Local linear estimator of $M(x)$:

- $\hookrightarrow$ Construct the local linear KDE of $p(y|x)$;
- $\hookrightarrow$ obtain an estimate of $p_y(y|x)$, $\hat{p}_y(y|x)$;

Nonparametric modal regression accounting for m.e.

(I) Local constant estimator of $M(x)$:

$\hookrightarrow$ Construct the local constant **k**ernel **d**ensity **e**stimator (KDE) of $p(x, y)$;

$\hookrightarrow$ obtain an estimate of $p_y(x, y)$, $\hat{p}_y(x, y)$;

$\hookrightarrow$ find solutions to $\hat{p}_y(x, y) = 0$.

(II) Local linear estimator of $M(x)$:

$\hookrightarrow$ Construct the local linear KDE of $p(y|x)$;

$\hookrightarrow$ obtain an estimate of $p_y(y|x)$, $\hat{p}_y(y|x)$;

$\hookrightarrow$ find solutions to $\hat{p}_y(y|x) = 0$.

(I) Local constant estimator

- In the absence of m.e.: KDE of $p(x, y)$ is

$$\tilde{p}(x, y) = \frac{1}{nh_1h_2} \sum_{j=1}^{n} K_1\left(\frac{X_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right).$$

(I) Local constant estimator

- In the absence of m.e.: KDE of $p(x, y)$ is

$$\tilde{p}(x, y) = \frac{1}{nh_1h_2}\sum_{j=1}^{n} K_1\left(\frac{X_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right).$$

- In the presence of m.e.: We consider

$$\hat{p}(x, y) = \frac{1}{nh_1h_2}\sum_{j=1}^{n} K_{U,0}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right),$$

(I) Local constant estimator

- In the absence of m.e.: KDE of $p(x, y)$ is

$$\tilde{p}(x, y) = \frac{1}{nh_1h_2} \sum_{j=1}^{n} K_1\left(\frac{X_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right).$$

- In the presence of m.e.: We consider

$$\hat{p}(x, y) = \frac{1}{nh_1h_2} \sum_{j=1}^{n} K_{U,0}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right),$$

  - $K_{U,0}(t) = \frac{1}{2\pi} \int e^{-its} \frac{\phi_{K_1}(s)}{\phi_U(s/h_1)} ds$, which satisfies

(I) Local constant estimator

- In the absence of m.e.: KDE of $p(x, y)$ is

$$\tilde{p}(x, y) = \frac{1}{nh_1h_2} \sum_{j=1}^{n} K_1\left(\frac{X_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right).$$

- In the presence of m.e.: We consider

$$\hat{p}(x, y) = \frac{1}{nh_1h_2} \sum_{j=1}^{n} K_{U,0}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right),$$

  - $K_{U,0}(t) = \frac{1}{2\pi} \int e^{-its} \frac{\phi_{K_1}(s)}{\phi_U(s/h_1)} ds$, which satisfies
  - $E\left\{ K_{U,0}\left(\frac{W - x}{h_1}\right) \middle| X \right\} = K_1\left(\frac{X - x}{h_1}\right)$ (Stefanski & Carroll, 1990).

- Choosing $K_2(\cdot)$ in certain ways (e.g., normal kernel) yields

$$\hat{p}_y(x, y) = \frac{1}{nh_1h_2^2}\sum_{j=1}^{n} K_{U,0}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right)\left(\frac{Y_j - y}{h_2}\right).$$

- Choosing $K_2(\cdot)$ in certain ways (e.g., normal kernel) yields

$$\hat{p}_y(x,y) = \frac{1}{nh_1h_2^2}\sum_{j=1}^{n} K_{U,0}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right)\left(\frac{Y_j - y}{h_2}\right).$$

- An estimator of $y_M(x)$, $\hat{y}_{M0}(x)$, solves

$$\sum_{j=1}^{n} K_{U,0}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right)(Y_j - y) = 0.$$

- Choosing $K_2(\cdot)$ in certain ways (e.g., normal kernel) yields

$$\hat{p}_y(x,y) = \frac{1}{nh_1h_2^2}\sum_{j=1}^{n} K_{U,0}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right)\left(\frac{Y_j - y}{h_2}\right).$$

- An estimator of $y_M(x)$, $\hat{y}_{M0}(x)$, solves

$$\sum_{j=1}^{n} K_{U,0}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right)(Y_j - y) = 0.$$

- **Mean-shift algorithm**:

$$y^{(k+1)} = \frac{\sum_{j=1}^{n} K_{U,0}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y^{(k)}}{h_2}\right) Y_j}{\sum_{j=1}^{n} K_{U,0}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y^{(k)}}{h_2}\right)}.$$

(II) Local linear estimator

In the absence of m.e. (Fan, Yao & Tong, 1996):

- Inspired by $E\left\{\frac{1}{h_2}K_2\left(\frac{Y-y}{h_2}\right)\Big|X=x\right\} \approx p(y|x)$, $h_2 \to 0$,

(II) Local linear estimator

In the absence of m.e. (Fan, Yao & Tong, 1996):

- Inspired by $E\left\{\frac{1}{h_2}K_2\left(\frac{Y-y}{h_2}\right)\Big|X=x\right\} \approx p(y|x)$, $h_2 \to 0$,
- local linear estimators of the "*mean function*" $p(y|x)$,

$$\tilde{p}(y|x) = \boldsymbol{e}_1^{\mathrm{T}}\mathbf{S}_n^{-1}(x)\mathbf{T}_n(x,y),$$

(II) Local linear estimator

In the absence of m.e. (Fan, Yao & Tong, 1996):

- Inspired by $E\left\{\frac{1}{h_2}K_2\left(\frac{Y-y}{h_2}\right)\Big|X=x\right\} \approx p(y|x)$, $h_2 \to 0$,
- local linear estimators of the "*mean function*" $p(y|x)$,

$$\tilde{p}(y|x) = \boldsymbol{e}_1^{\mathrm{T}}\mathbf{S}_n^{-1}(x)\mathbf{T}_n(x,y),$$

— $\boldsymbol{e}_1 = (1,\ 0)^{\mathrm{T}}$,

$$\mathbf{S}_n(x) = \begin{bmatrix} S_{n,0}(x) & S_{n,1}(x) \\ S_{n,1}(x) & S_{n,2}(x) \end{bmatrix}, \quad \mathbf{T}_n(x,y) = \left[T_{n,0}(x,y),\ T_{n,1}(x,y)\right]^{\mathrm{T}},$$

(II) Local linear estimator

In the absence of m.e. (Fan, Yao & Tong, 1996):

- Inspired by $E\left\{\frac{1}{h_2}K_2\left(\frac{Y-y}{h_2}\right)\Big|X=x\right\} \approx p(y|x)$, $h_2 \to 0$,
- local linear estimators of the "*mean function*" $p(y|x)$,

$$\tilde{p}(y|x) = \boldsymbol{e}_1^{\mathrm{T}}\mathbf{S}_n^{-1}(x)\mathbf{T}_n(x,y),$$

— $\boldsymbol{e}_1 = (1,\, 0)^{\mathrm{T}}$,

$$\mathbf{S}_n(x) = \begin{bmatrix} S_{n,0}(x) & S_{n,1}(x) \\ S_{n,1}(x) & S_{n,2}(x) \end{bmatrix}, \quad \mathbf{T}_n(x,y) = [T_{n,0}(x,y),\, T_{n,1}(x,y)]^{\mathrm{T}},$$

— $S_{n,\ell}(x) = \frac{1}{nh_1}\sum_{j=1}^{n}\left(\frac{X_j-x}{h_1}\right)^{\ell}K_1\left(\frac{X_j-x}{h_1}\right)$,

$$T_{n,\ell}(x,y) = \frac{1}{nh_1}\sum_{j=1}^{n}\left(\frac{X_j-x}{h_1}\right)^{\ell}K_1\left(\frac{X_j-x}{h_1}\right)\frac{1}{h_2}K_2\left(\frac{Y_j-y}{h_2}\right).$$

In the presence of m.e., we consider

$$\hat{p}(y|x) = \boldsymbol{e}_1^{\mathrm{T}} \hat{\mathbf{S}}_n^{-1}(x) \hat{\mathbf{T}}_n(x, y),$$

— $\hat{S}_{n,\ell}(x) = \frac{1}{nh_1} \sum_{j=1}^{n} K_{U,\ell}\left(\frac{W_j - x}{h_1}\right),$

$$\hat{T}_{n,\ell}(x, y) = \frac{1}{nh_1h_2} \sum_{j=1}^{n} K_{U,\ell}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right),$$

In the presence of m.e., we consider

$$\hat{p}(y|x) = \boldsymbol{e}_1^{\mathrm{T}} \hat{\mathbf{S}}_n^{-1}(x) \hat{\mathbf{T}}_n(x, y),$$

- $\hat{S}_{n,\ell}(x) = \frac{1}{nh_1} \sum_{j=1}^{n} K_{U,\ell}\left(\frac{W_j - x}{h_1}\right),$

  $\hat{T}_{n,\ell}(x, y) = \frac{1}{nh_1h_2} \sum_{j=1}^{n} K_{U,\ell}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right),$

- $K_{U,\ell}(t) = i^{-\ell} \frac{1}{2\pi} \int e^{-its} \frac{\phi_{K_1}^{(\ell)}(s)}{\phi_U(s/h_1)} ds$, which satisfies

In the presence of m.e., we consider

$$\hat{p}(y|x) = \boldsymbol{e}_1^{\mathrm{T}} \hat{\mathbf{S}}_n^{-1}(x) \hat{\mathbf{T}}_n(x, y),$$

- $\hat{S}_{n,\ell}(x) = \frac{1}{nh_1} \sum_{j=1}^{n} K_{U,\ell}\left(\frac{W_j - x}{h_1}\right),$

  $\hat{T}_{n,\ell}(x, y) = \frac{1}{nh_1h_2} \sum_{j=1}^{n} K_{U,\ell}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right),$
- $K_{U,\ell}(t) = i^{-\ell} \frac{1}{2\pi} \int e^{-its} \frac{\phi_{K_1}^{(\ell)}(s)}{\phi_U(s/h_1)} ds$, which satisfies
- $E\left\{K_{U,\ell}\left(\frac{W - x}{h_1}\right) \middle| X\right\} = \left(\frac{X - x}{h_1}\right)^{\ell} K_1\left(\frac{X - x}{h_1}\right)$ (Delaigle, Fan & Carroll, 2009).

- Differentiating $\hat{p}(y|x)$ w.r.t. y gives

$$\hat{p}_y(y|x) = \boldsymbol{e}_1^{\mathrm{T}} \hat{\mathbf{S}}_n^{-1}(x) \hat{\mathbf{T}}'_n(x, y),$$

- Differentiating $\hat{p}(y|x)$ w.r.t. y gives

$$\hat{p}_y(y|x) = \boldsymbol{e}_1^{\mathrm{T}} \hat{\mathbf{S}}_n^{-1}(x) \hat{\mathbf{T}}'_n(x, y),$$

  — $\hat{\mathbf{T}}'_n(x, y) = [\hat{T}'_{n,0}(x, y), \hat{T}'_{n,1}(x, y)]^{\mathrm{T}}$, in which, for $\ell = 0, 1$,

- Differentiating $\hat{p}(y|x)$ w.r.t. y gives

$$\hat{p}_y(y|x) = \boldsymbol{e}_1^{\mathrm{T}} \hat{\mathbf{S}}_n^{-1}(x) \hat{\mathbf{T}}'_n(x, y),$$

  - $\hat{\mathbf{T}}'_n(x, y) = [\hat{T}'_{n,0}(x, y), \hat{T}'_{n,1}(x, y)]^{\mathrm{T}}$, in which, for $\ell = 0, 1$,
  -

$$\hat{T}'_{n,\ell}(x, y) = \frac{1}{nh_1h_2^2} \sum_{j=1}^{n} K_{U,\ell}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right)\left(\frac{Y_j - y}{h_2}\right).$$

- Differentiating $\hat{p}(y|x)$ w.r.t. y gives

$$\hat{p}_y(y|x) = \boldsymbol{e}_1^{\mathrm{T}} \hat{\mathbf{S}}_n^{-1}(x) \hat{\mathbf{T}}_n'(x, y),$$

— $\hat{\mathbf{T}}_n'(x, y) = [\hat{T}_{n,0}'(x, y), \hat{T}_{n,1}'(x, y)]^{\mathrm{T}}$, in which, for $\ell = 0, 1$,

$$\hat{T}_{n,\ell}'(x, y) = \frac{1}{n h_1 h_2^2} \sum_{j=1}^{n} K_{U,\ell}\left(\frac{W_j - x}{h_1}\right) K_2\left(\frac{Y_j - y}{h_2}\right)\left(\frac{Y_j - y}{h_2}\right).$$

- Setting $\hat{p}_y(y|x) = 0$, a mode estimator $\hat{y}_{M1}(x)$ solves

$$\sum_{j=1}^{n} \left\{ K_{U,0}\left(\frac{W_j - x}{h_1}\right) \hat{S}_{n,2}(x) - K_{U,1}\left(\frac{W_j - x}{h_1}\right) \hat{S}_{n,1}(x) \right\} \times$$
$$K_2\left(\frac{Y_j - y}{h_2}\right)(Y_j - y) = 0.$$

- **Mean-shift algorithm**:

$$y^{(k+1)} = \frac{\sum_{j=1}^{n} \omega_j^{(k)} Y_j}{\sum_{j=1}^{n} \omega_j^{(k)}},$$

where

$$\omega_j^{(k)} = \left\{ K_{U,0}\left(\frac{W_j - x}{h_1}\right) \hat{S}_{n,2}(x) - K_{U,1}\left(\frac{W_j - x}{h_1}\right) \hat{S}_{n,1}(x) \right\} \times K_2\left(\frac{Y_j - y^{(k)}}{h_2}\right).$$

Asymptotic properties of $\hat{y}_M(x)$

Consider

- Pointwise error rate:

$$\Delta_n(x) = |\hat{y}_M(x) - y_M(x)|, \text{ for } x \in \mathscr{X}.$$

- Uniform error rate:

$$\Delta_n = \sup_{x \in \mathscr{X}} \Delta_n(x).$$

- Mean integrated squared error:

$$\text{MISE} = E\left\{\int_{\mathscr{X}} \Delta_n^2(x)dx\right\}.$$

Relating $\hat{y}_{M}(x)$ to $\hat{p}_y(x, y_{M})$

Because $\hat{p}_y(x, \hat{y}_M) = 0$, one has

Relating $\hat{y}_{M}(x)$ to $\hat{p}_y(x, y_M)$

Because $\hat{p}_y(x, \hat{y}_M) = 0$, one has

$$\begin{aligned} \hat{p}_y(x, y_M) &= \hat{p}_y(x, y_M) - 0 \\ &= \hat{p}_y(x, y_M) - \hat{p}_y(x, \hat{y}_M) \\ &= (y_M - \hat{y}_M)\hat{p}_{yy}(x, y^*), \end{aligned}$$

where y^* lies between y_M and $\hat{y}_M$.

Relating $\hat{y}_{M}(x)$ to $\hat{p}_y(x, y_{M})$

Because $\hat{p}_y(x, \hat{y}_M) = 0$, one has

$$\begin{aligned} \hat{p}_y(x, y_M) &= \hat{p}_y(x, y_M) - 0 \\ &= \hat{p}_y(x, y_M) - \hat{p}_y(x, \hat{y}_M) \\ &= (y_M - \hat{y}_M)\hat{p}_{yy}(x, y^*), \end{aligned}$$

where y^* lies between y_M and $\hat{y}_M$.

Hence,

$$\hat{y}_M(x) - y_M(x) \approx -p_{yy}^{-1}(x, y_M)\hat{p}_y(x, y_M).$$

Asymptotic error rates

$$\Delta_n(x) = O(h_1^2 + h_2^2) + \begin{cases} O_P\left(\sqrt{\dfrac{1}{nh_1^{1+2b}h_2^3}}\right) & \text{for ordinary smooth } U, \\ O_P\left\{\dfrac{\exp(h_1^{-b}/d_2)}{\sqrt{nh_1^{1-2b_2}h_2^3}}\right\} & \text{for super smooth } U, \end{cases}$$

- ordinary smooth U: $|\phi_U(t)| = O(t^{-b})$, as $t \to \infty$, for some $b > 0$,

Asymptotic error rates

$$\Delta_n(x) = O(h_1^2 + h_2^2) + \begin{cases} O_P\left(\sqrt{\dfrac{1}{nh_1^{1+2b}h_2^3}}\right) & \text{for ordinary smooth } U, \\ O_P\left\{\dfrac{\exp(h_1^{-b}/d_2)}{\sqrt{nh_1^{1-2b_2}h_2^3}}\right\} & \text{for super smooth } U, \end{cases}$$

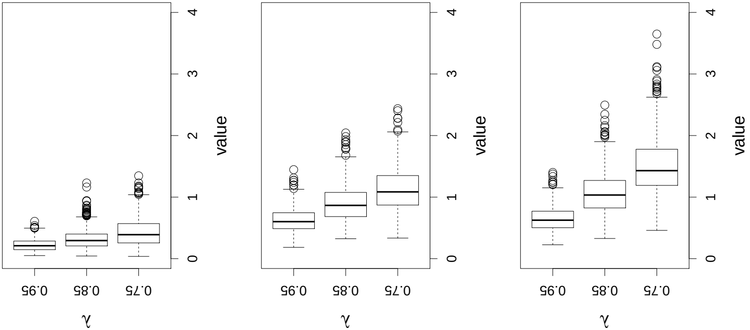
- ordinary smooth U: $|\phi_U(t)| = O(t^{-b})$, as $t \to \infty$, for some $b > 0$,
- super smooth U: $|\phi_U(t)| = O\{t^{b_0}\exp(-t^b/d_2)\}$, as $t \to \infty$, for some $b_0, b, d_2 > 0$.

Simulation study—Case I

- $[Y|X = x] \sim$
  $0.5N\left(m(x) - 2\sigma(x), 2.5^2\sigma^2(x)\right) + 0.5N\left(m(x), 0.5^2\sigma^2(x)\right)$,
- $m(x) = x + x^2$,
- $\sigma(x) = 0.5 + e^{-x^2}$,
- $X \sim \text{Uniform}(-2, 2)$, and
- $U \sim \text{Laplace}(0, \sigma_u/\sqrt{2})$,
- σ_u takes values that lead to $\lambda = 0.75, 0.85, 0.95$, where

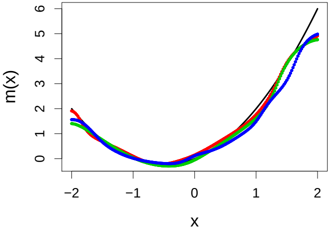
$$\lambda = \frac{\text{Var}(X)}{\text{Var}(X) + \sigma_u^2} = \text{the reliability ratio.}$$

Here, $M(x) = \{y_M(x)\}$ and $y_M(x) \approx m(x)$.

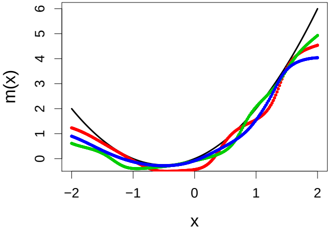


(a) $\hat{y}_{M1}(x)$ (b) $\hat{y}_{M0}(x)$ (c) Naive estimate

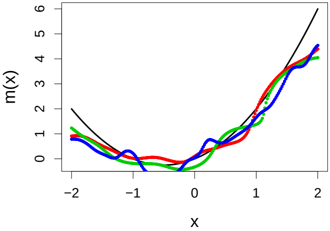
Fig. 3: Boxplots of ISEs versus λ.



(a) $\hat{y}_{M1}(x)$



(b) $\hat{y}_{M0}(x)$



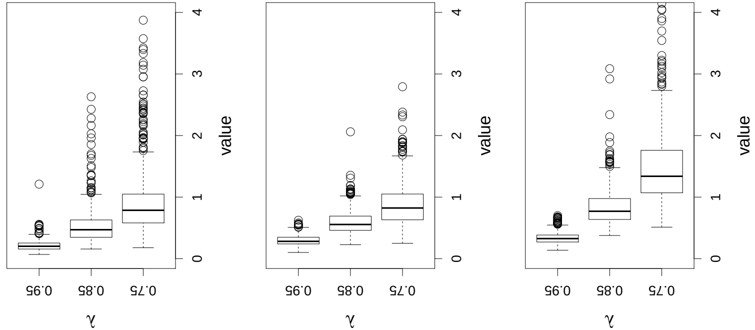
(c) Naive estimate

Fig. 4: Quantile curves. Q_1: red; Q_2: green; Q_3: blue; truth: black.

Simulation study—Case II

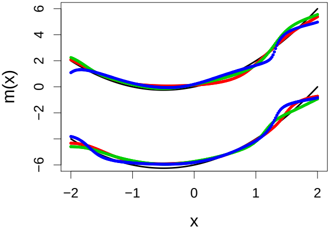
- $[Y|X = x] \sim 0.5N\left(m_1(x), 0.5^2\right) + 0.5N\left(m_2(x), 0.5^2\right)$,
- $m_1(x) = x + x^2$, $m_2(x) = m_1(x) - 6$,
- $X \sim \text{Uniform}(-2, 2)$, and
- $U \sim \text{Laplace}(0, \sigma_u/\sqrt{2})$.

Here, $M(x) = \{y_M^{(1)}(x),\ y_M^{(2)}(x)\}$, where $y_M^{(k)}(x) \approx m_k(x)$, for $k = 1, 2$.

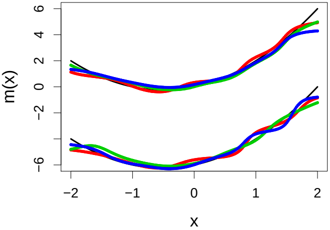


(a) $\hat{y}_{M1}(x)$ (b) $\hat{y}_{M0}(x)$ (c) Naive estimate

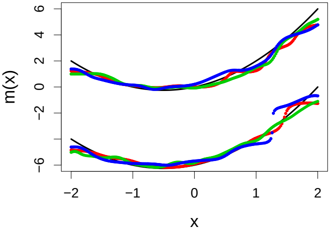
Fig. 5: Boxplots of ISEs versus λ.



(a) $\hat{y}_{M1}(x)$



(b) $\hat{y}_{M0}(x)$



(c) Naive estimate

Fig. 6: Quantile curves. Q_1: red; Q_2: green; Q_3: blue; truth: black.

Bandwidth selection (ideal/unrealistic)

We used the theoretical optimal bandwidths (h_1, h_2) that minimize

$$\text{ISE} = \int_{x \in \mathscr{X}} \left\{ \text{Haus}(\hat{M}(x),\ M(x)) \right\}^2 dx,$$

where

- $\text{Haus}(A, B) = \inf\{r : A \subset B \oplus r, B \subset A \oplus r\}$ is the Hausdorff distance between two sets, A and B,

Bandwidth selection (ideal/unrealistic)

We used the theoretical optimal bandwidths (h_1, h_2) that minimize

$$\text{ISE} = \int_{x \in \mathscr{X}} \left\{ \text{Haus}(\hat{M}(x),\ M(x)) \right\}^2 dx,$$

where

- $\text{Haus}(A, B) = \inf\{r : A \subset B \oplus r, B \subset A \oplus r\}$ is the Hausdorff distance between two sets, A and B,
- $A \oplus r = \{x : \inf_{y \in A} |x - y| \le r\}$.

Bandwidth selection (practical)

In the absence of m.e.: Choose (h_1, h_2) to minimize ISE (Fan and Yim, 2004).

- $$\begin{aligned} \text{ISE} &= \int \{\tilde{p}(y|x) - p(y|x)\}^2 f_X(x)dxdy \\ &= \int_{\mathscr{Y}} \int_{\mathscr{X}} \tilde{p}(y|x)^2 f_X(x)dxdy \\ &\quad -2\int_{\mathscr{Y}} \int_{\mathscr{X}} \tilde{p}(y|x)p(y|x)f_X(x)dxdy + ... \end{aligned}$$

Bandwidth selection (practical)

In the absence of m.e.: Choose (h_1, h_2) to minimize ISE (Fan and Yim, 2004).

- $$\begin{aligned} \text{ISE} &= \int \{\tilde{p}(y|x) - p(y|x)\}^2 f_X(x) dx dy \\ &= \int_{\mathscr{Y}} \int_{\mathscr{X}} \tilde{p}(y|x)^2 f_X(x) dx dy \\ &\quad -2 \int_{\mathscr{Y}} \int_{\mathscr{X}} \tilde{p}(y|x) p(y|x) f_X(x) dx dy + \ldots \end{aligned}$$

- $$\begin{aligned} & \text{CV}_d(h_1, h_2) \\ = \ & \frac{1}{n} \sum_{j=1}^{n} \int_{\mathscr{Y}} \tilde{p}_{-j}(y|X_j)^2 dy - \frac{2}{n} \sum_{j=1}^{n} \tilde{p}_{-j}(Y_j|X_j). \end{aligned}$$

We propose...

$$\mathrm{CV}_m(h_1, h_2) = \frac{1}{n}\sum_{j=1}^{n} d^2\left(Y_j,\, \hat{M}_{-j}(X_j)\right)\tilde{N}^2_{-j}(X_j)\omega(X_j),$$

- $d(x, S) = \inf_{y\in S}|x-y|$ for a set S,
- $\hat{N}(x) =$ the number of distinct elements in $\hat{M}(x)$.

Case	Ref	Reg	Boot	CV_d	CV_m	Ideal
I	0.59	**0.50**	2.19	6.19	0.60	0.34
	(0.08)	(0.12)	(0.88)	(1.56)	(0.05)	(0.05)
II	3.46	1.22	0.70	2.16	**0.41**	0.36
	(1.84)	(0.66)	(0.45)	(0.73)	(0.10)	(0.05)

Table 1: Average of EISE across 500 MC replicates. Numbers in parentheses are $10\times$ Std. Err.

where the empirical integrated square error (EISE) is

$$\mathrm{EISE} = \sum_{k=0}^{\mathcal{M}} \left\{ \mathrm{Haus}(\hat{M}(x_k), M(x_k)) \right\}^2 \Delta.$$

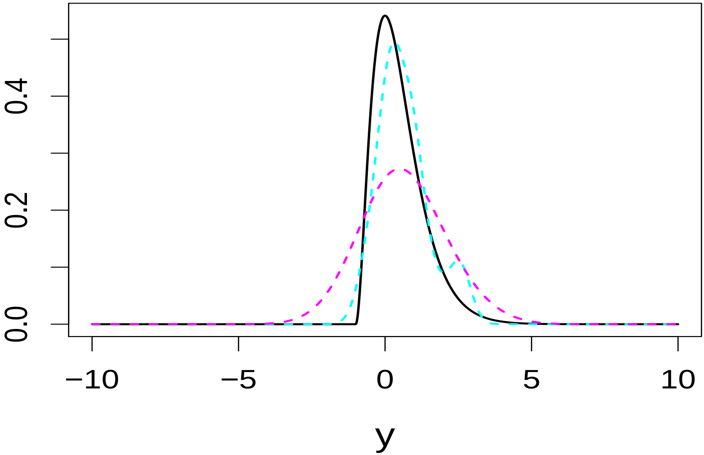


Fig. 7: Black line: the true $p(y|x)$; blue line: the estimate using CV_d; red line: the estimate using our CV_m.

To be resolved...

- Convergence of the mean-shift algorithm
- Bandwidth selection
- Conditional density estimation in the presence of m.e. in both Y and X