On Binscatter

Matias D. Cattaneo[1], Richard K. Crump[2], Max H. Farrell[3] and Yingjie Feng[4]

November 2022

[1] Princeton University.
[2] Federal Reserve Bank of New York. The views expressed here are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.
[3] University of Chicago.
[4] Tsinghua University.

Outline

Introduction

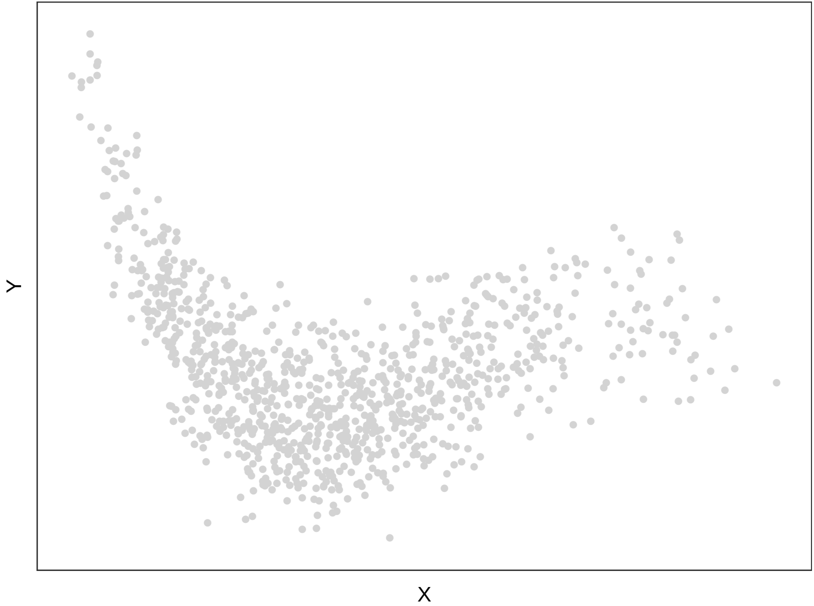
Binscatter is widely used in economics and other disciplines.

- Popularized by Chetty, Friedman, Rockoff, Saez, many others.
- Previous incarnations:
  - *Regressogram* (Tukey, 1961).
  - *Subclassification* (Cochran, 1968).
  - *Portfolio Sorting* (Fama, 1976).
  - *Regression Trees* (Friedman, 1977).
  - you tell me...
- Today: foundational, thorough study of Binscatter.
  - *Methodology*: guidance on valid and invalid current practices, and more.
  - *Theory*: novel strong approximation approach, and more.
  - *Practice*: new `Python`, `R` and `Stata` software (Binsreg package):

`https://nppackages.github.io/binsreg/`

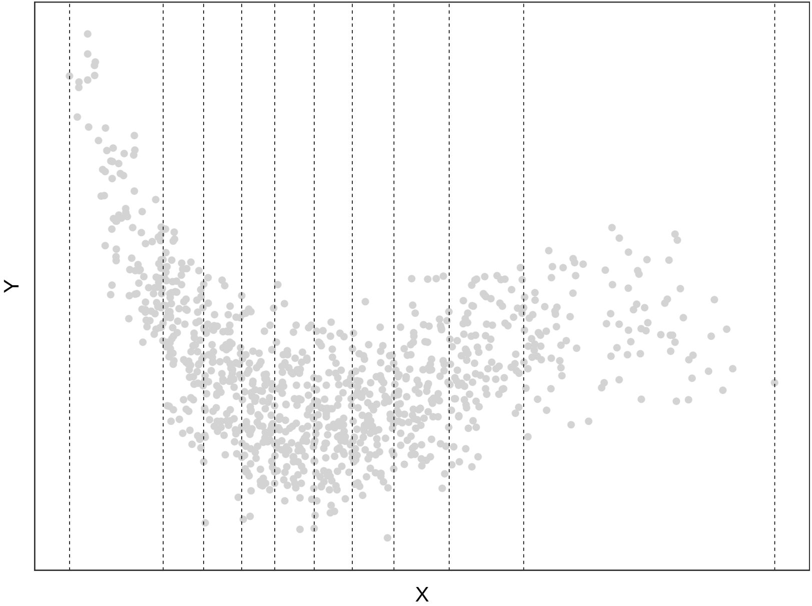
What is a binned scatter plot?

Step 1: Start with a familiar scatter plot



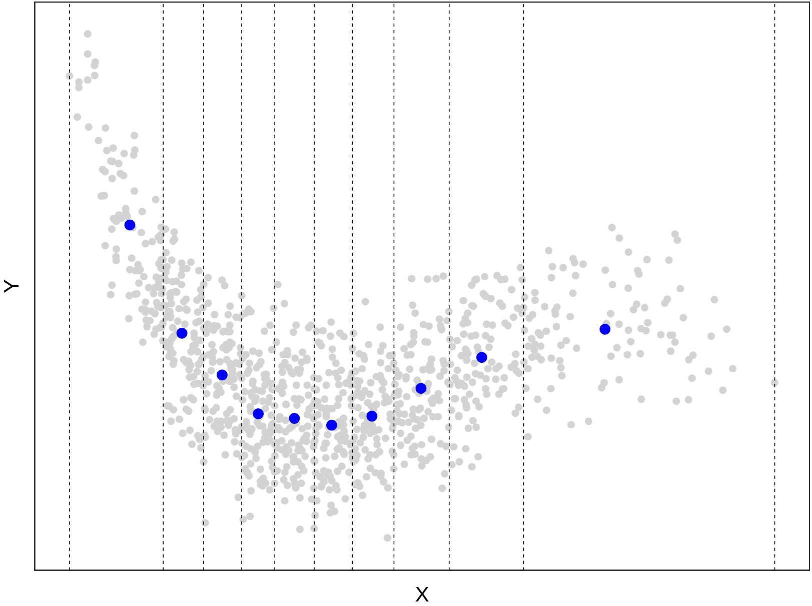
What is a binned scatter plot?

Step 2: Partition the support of X into bins



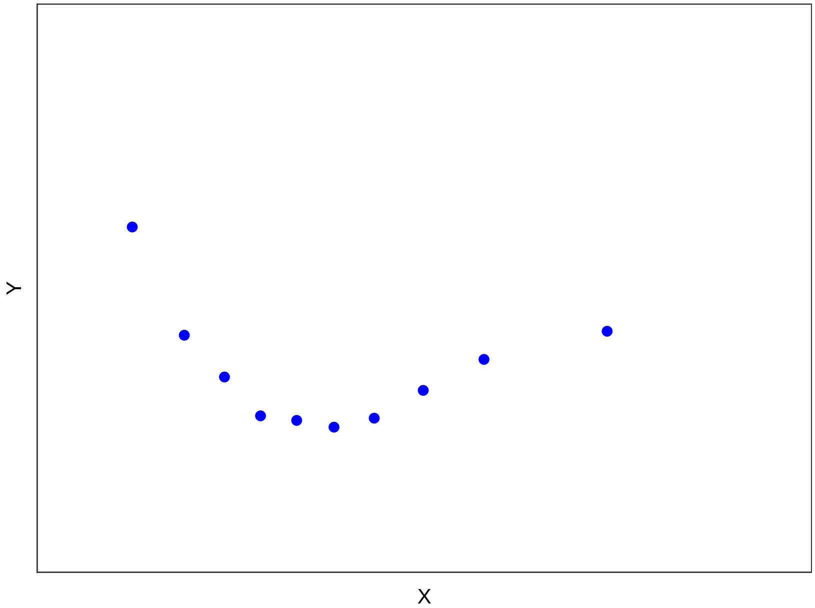
What is a binned scatter plot?

Step 3: Find the average Y in each bin



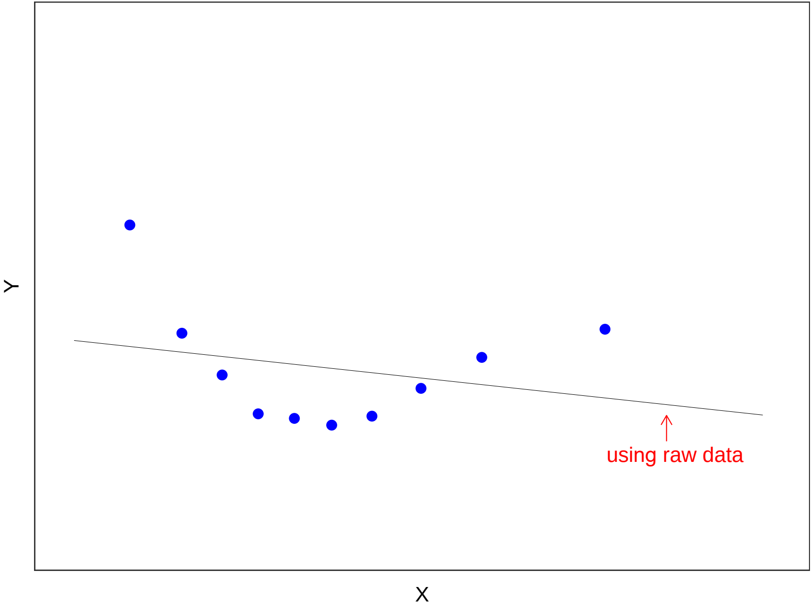
What is a binned scatter plot?

Step 4: Plot only bin means

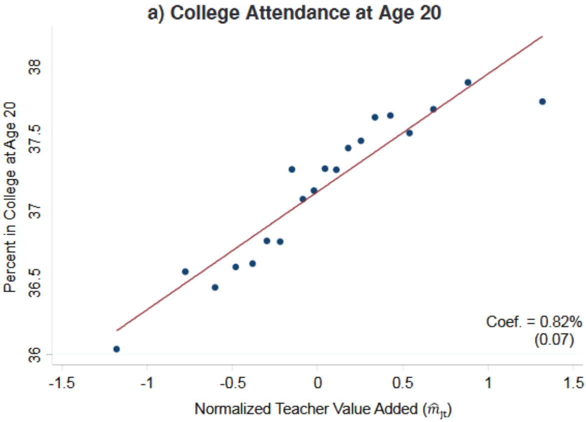


What is a binned scatter plot?

Step 5: Add a polynomial fit to raw data



Typical Example: Chetty, Friedman and Rockoff (2014, AER)

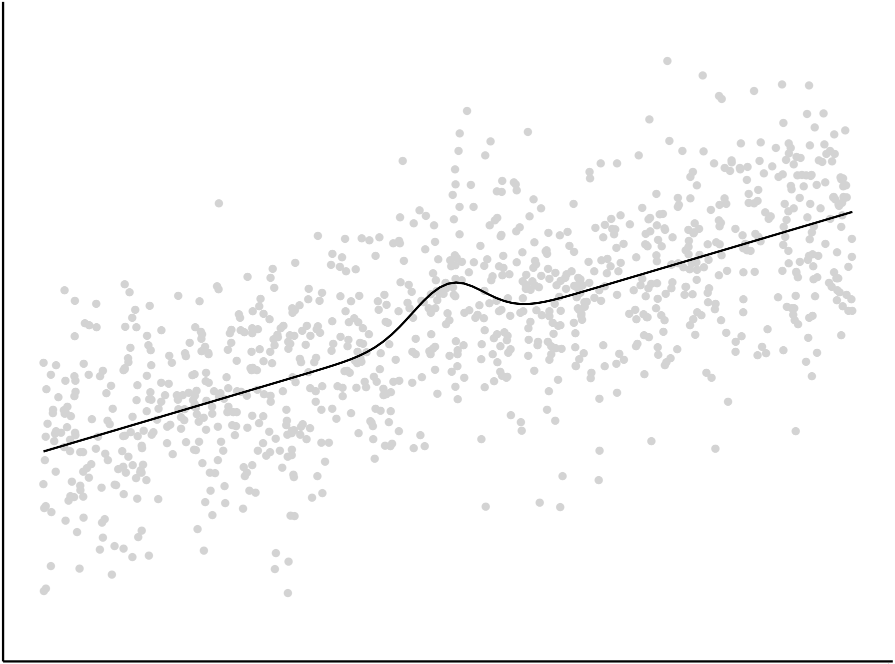


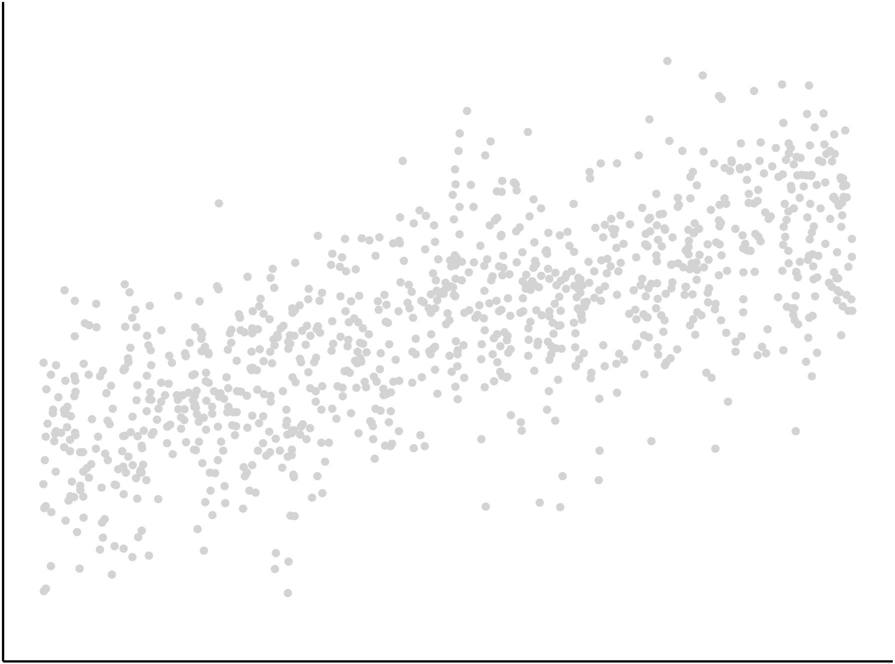
Note: $n = 4,170,905$ with # of bins $J = 20$

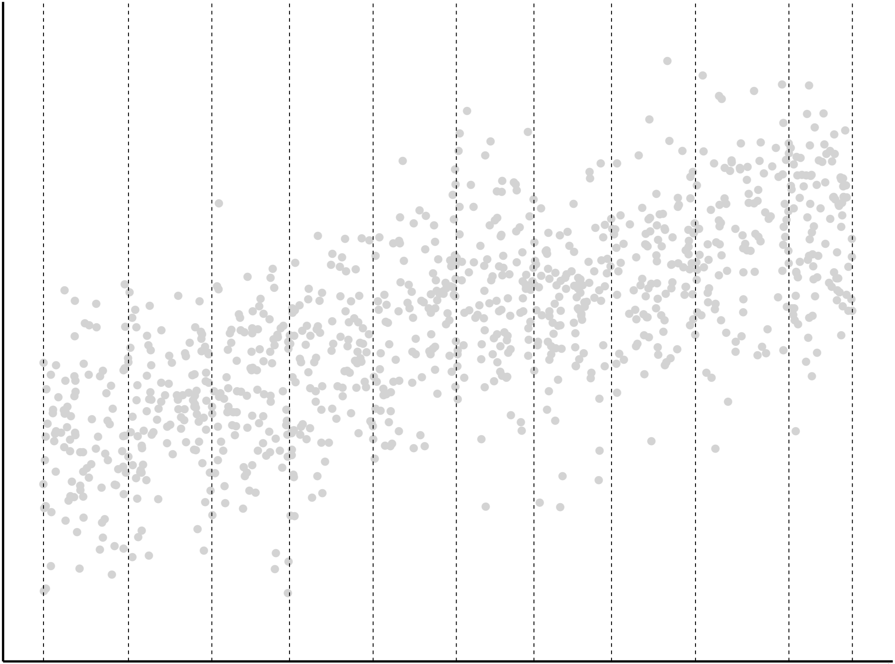
Outline

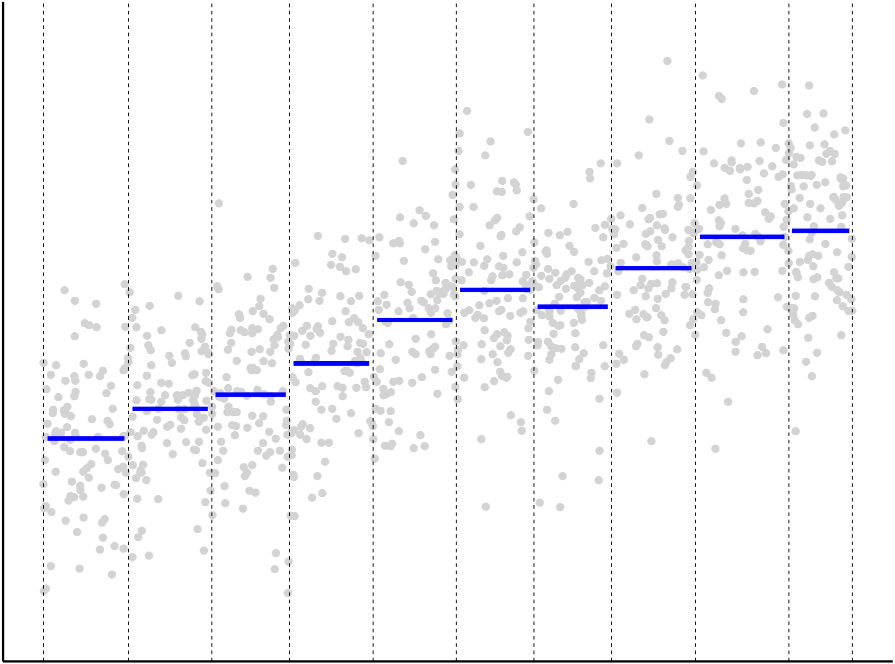
Overview: Contributions

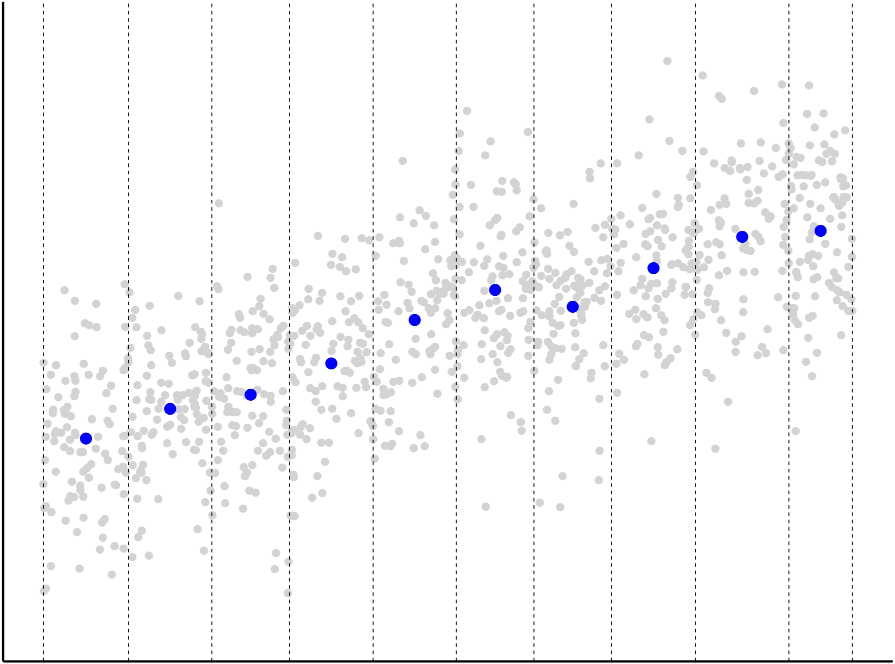
1. Set up formal, general framework for studying **Binscatter**.
   - *Respects practice*: quantile-spaced binning, covariate adjustment.
   - *Extensions*: higher-order polynomial, smoothness-restricted approximations.
   - *Generalizations*: semi-linear QMLE (quantiles, logistic, etc.).
2. IMSE-Optimal choice of binning structure.
3. Valid point estimators, confidence intervals, and confidence bands.
4. Valid hypothesis testing of parametric specification and shape restrictions.
5. Novel theoretical results specifically developed for binscatter.
6. `Python`, `R`, and `Stata` software resolving valid and invalid current practices.

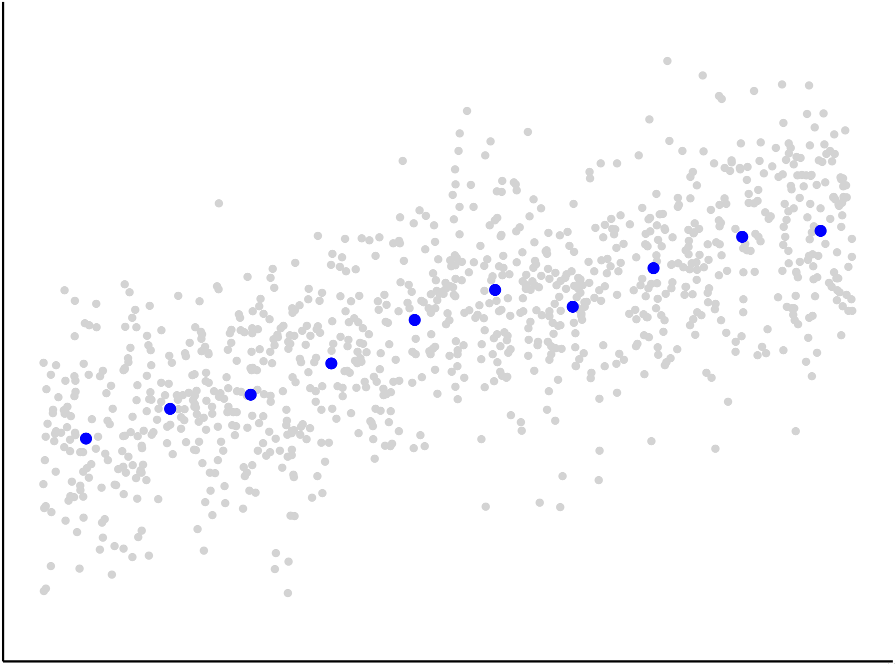


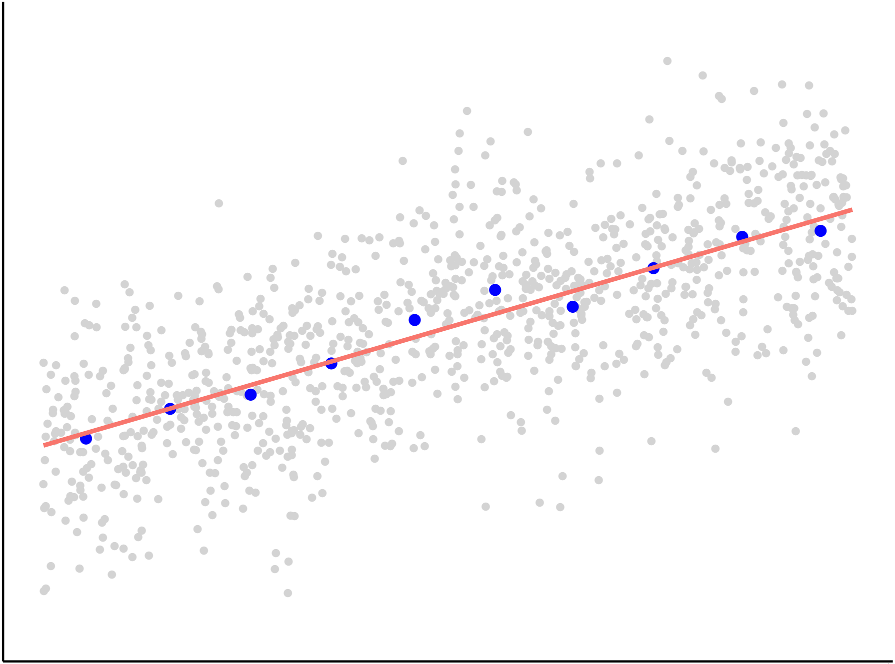


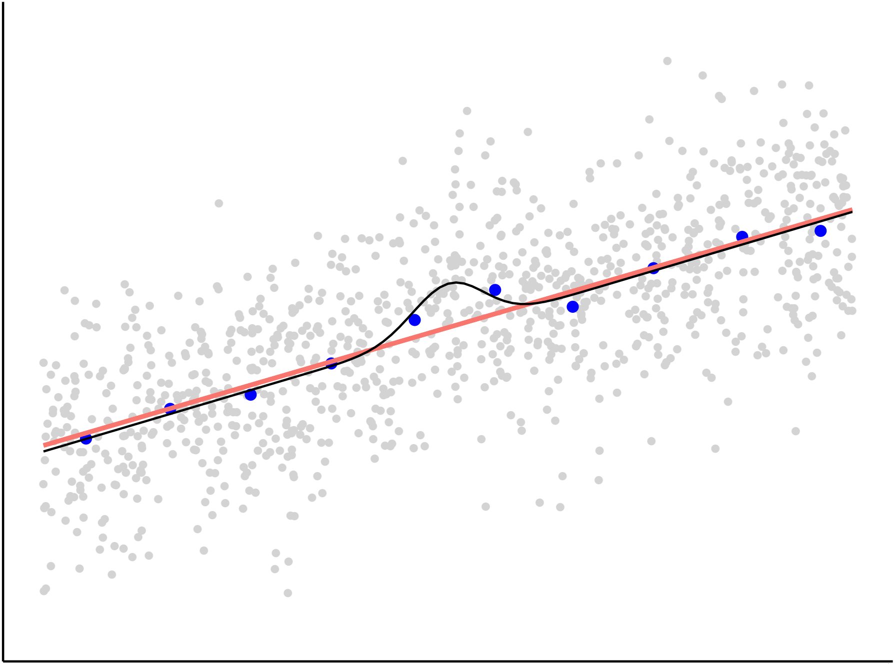


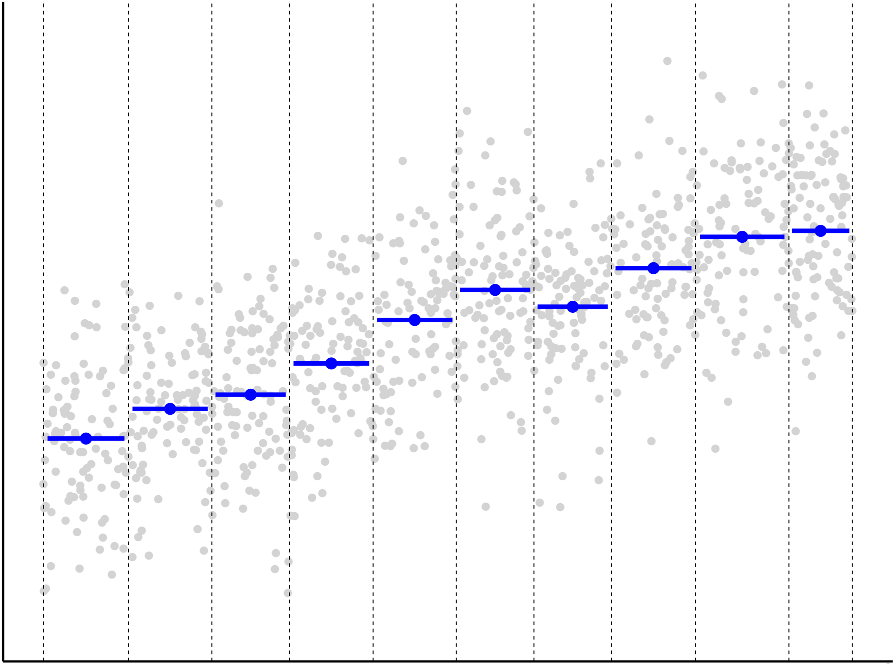












Framework: Canonical Binscatter

$$y_i = \mu(x_i) + \varepsilon_i, \qquad \mathbb{E}[\varepsilon_i | x_i] = 0$$

Binscatter:

$$\widehat{\mu}(x) = \widehat{\mathbf{b}}(x)'\widehat{\boldsymbol{\beta}}, \qquad \widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \widehat{\mathbf{b}}(x_i)'\boldsymbol{\beta})^2$$

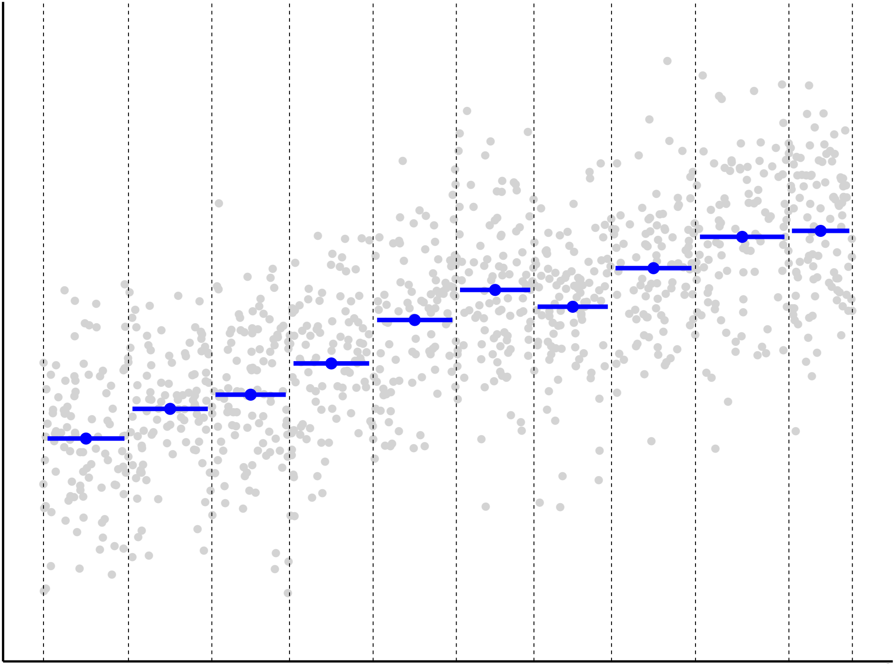
- Partitioning/Binning:

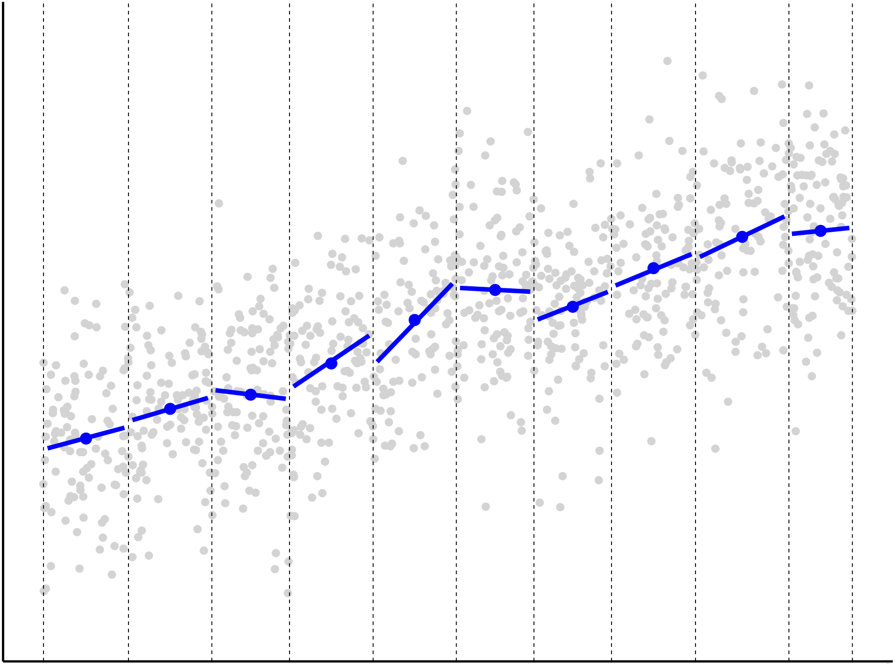
$$\widehat{\Delta} = \{\widehat{\mathcal{B}}_1, \dots, \widehat{\mathcal{B}}_J\}, \qquad \widehat{\mathcal{B}}_j = \begin{cases} \big[x_{(1)}, x_{(\lfloor n/J \rfloor)}\big) & \text{if } j = 1 \\ \big[x_{(\lfloor n(j-1)/J \rfloor)}, x_{(\lfloor nj/J \rfloor)}\big) & \text{if } j = 2, \dots, J-1 \\ \big[x_{(\lfloor n(J-1)/J \rfloor)}, x_{(n)}\big] & \text{if } j = J \end{cases}$$

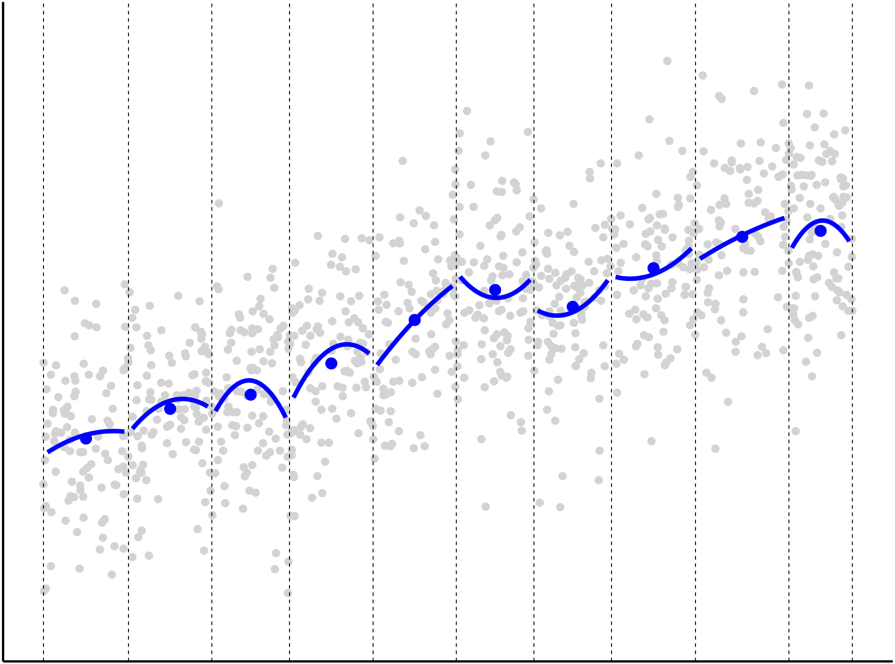
- Within-Bin Constant Approximation:

$$\widehat{\mathbf{b}}(x) = \left[\ \mathbb{1}_{\widehat{\mathcal{B}}_1}(x) \quad \mathbb{1}_{\widehat{\mathcal{B}}_2}(x) \quad \cdots \quad \mathbb{1}_{\widehat{\mathcal{B}}_J}(x)\ \right]'$$

  - Dimension: J.







Framework: Within-Bin Polynomial Approximation

$$y_i = \mu(x_i) + \varepsilon_i, \qquad \mathbb{E}[\varepsilon_i | x_i] = 0$$

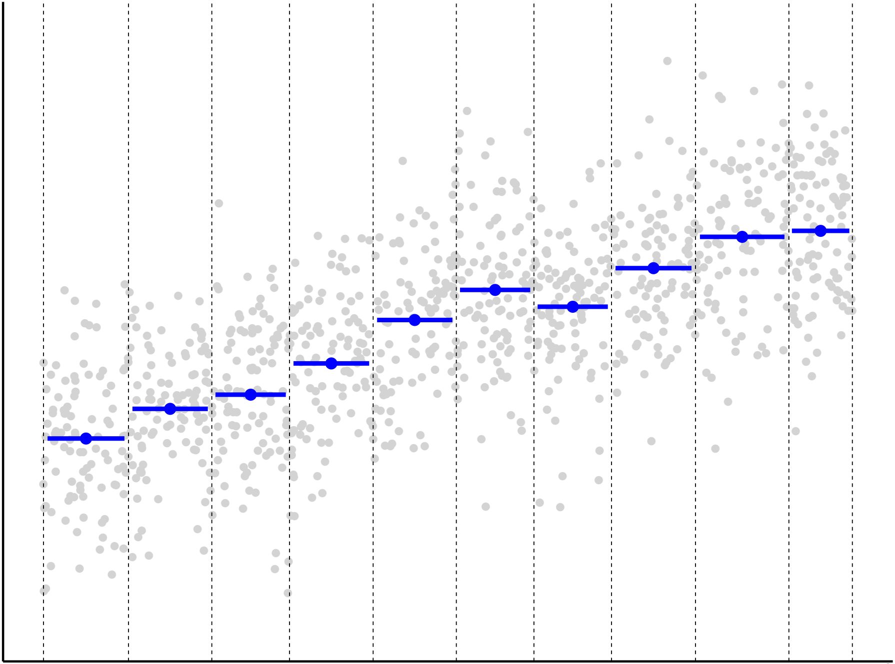
Binscatter:

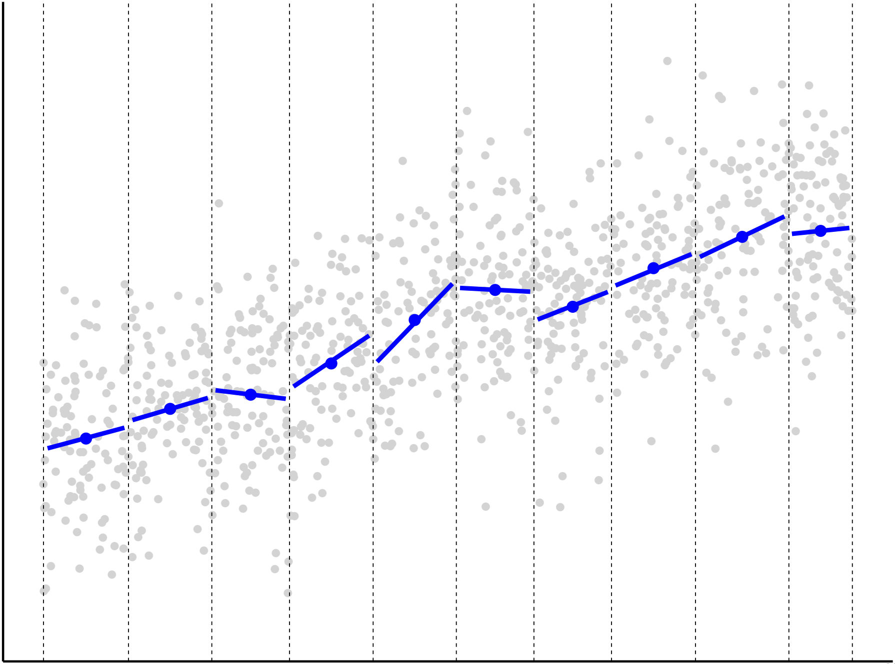
$$\widehat{\mu}^{(v)}(x) = \widehat{\mathbf{b}}^{(v)}(x)'\widehat{\boldsymbol{\beta}}, \qquad \widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \widehat{\mathbf{b}}(x_i)'\boldsymbol{\beta})^2$$

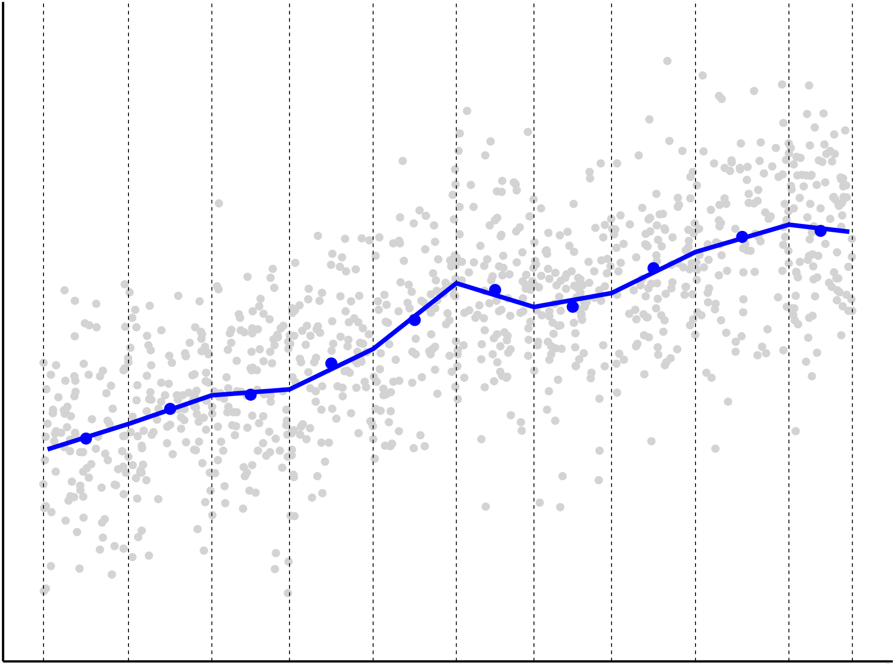
- Partitioning/Binning: $\widehat{\Delta} = \{\widehat{\mathcal{B}}_1, \ldots, \widehat{\mathcal{B}}_J\}$.
- Within-Bin Polynomial Approximation:

$$\widehat{\mathbf{b}}(x) = \left[\ \mathbb{1}_{\widehat{\mathcal{B}}_1}(x) \quad \mathbb{1}_{\widehat{\mathcal{B}}_2}(x) \quad \cdots \quad \mathbb{1}_{\widehat{\mathcal{B}}_J}(x)\ \right]' \otimes \left[\ 1 \quad x \quad \cdots \quad x^p\ \right]'$$

  - Dimension: $(p+1) \cdot J$.
- Restrictions: $0 \leq v \leq p$.







Framework: Across-Bins Smoothness Restriction

$$y_i = \mu(x_i) + \varepsilon_i, \qquad \mathbb{E}[\varepsilon_i | x_i] = 0$$

Binscatter:

$$\widehat{\mu}^{(v)}(x) = \widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\boldsymbol{\beta}}, \qquad \widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} (y_i - \widehat{\mathbf{b}}_s(x_i)'\boldsymbol{\beta})^2$$

- Partitioning/Binning: $\widehat{\Delta} = \{\widehat{\mathcal{B}}_1, \ldots, \widehat{\mathcal{B}}_J\}$.
- Across-Bins Smoothness Restriction:

$$\widehat{\mathbf{b}}_s(x) = \widehat{\mathbf{T}}_s \widehat{\mathbf{b}}(x), \qquad \widehat{\mathbf{b}}(x) = \begin{bmatrix} \mathbb{1}_{\widehat{\mathcal{B}}_1}(x) & \cdots & \mathbb{1}_{\widehat{\mathcal{B}}_J}(x) \end{bmatrix}' \otimes \begin{bmatrix} 1 & \cdots & x^p \end{bmatrix}'$$

  - Dimension $\widehat{\mathbf{T}}_s$: $[(p+1)J - (J-1)s] \times (p+1)J$.
- Restrictions: $0 \leq s, v \leq p$.

Framework: Covariate Adjustment

$$y_i = \mu(x_i) + \mathbf{w}_i'\boldsymbol{\gamma} + \epsilon_i, \qquad \mathbb{E}[\epsilon_i|x_i, \mathbf{w}_i] = 0$$

Covariate-Adjusted Binscatter:

$$\widehat{\mu}^{(v)}(x) = \widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\boldsymbol{\beta}}, \qquad \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \underset{\boldsymbol{\beta},\boldsymbol{\gamma}}{\arg\min} \sum_{i=1}^{n} (y_i - \widehat{\mathbf{b}}_s(x_i)'\boldsymbol{\beta} - \mathbf{w}_i'\boldsymbol{\gamma})^2$$

- Partitioning/Binning: $\{\widehat{\mathcal{B}}_1, \ldots, \widehat{\mathcal{B}}_J\}$ — Binscatter Basis: $\widehat{\mathbf{b}}_s(x)$.
- Dimension: $[(p+1)J - (J-1)s] + d$ — Restrictions: $0 \leq s, v \leq p$.

Framework: Covariate Adjustment

$$y_i = \mu(x_i) + \mathbf{w}_i'\boldsymbol{\gamma} + \epsilon_i, \qquad \mathbb{E}[\epsilon_i | x_i, \mathbf{w}_i] = 0$$

Covariate-Adjusted Binscatter:

$$\widehat{\mu}^{(v)}(x) = \widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\boldsymbol{\beta}}, \qquad \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \underset{\boldsymbol{\beta},\boldsymbol{\gamma}}{\arg\min} \sum_{i=1}^{n} (y_i - \widehat{\mathbf{b}}_s(x_i)'\boldsymbol{\beta} - \mathbf{w}_i'\boldsymbol{\gamma})^2$$

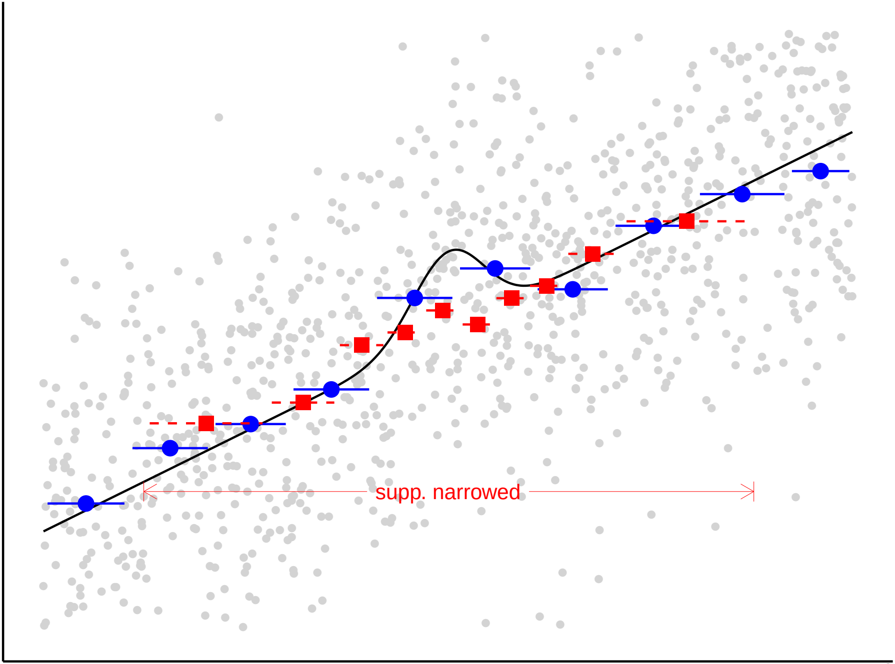
- Partitioning/Binning: $\{\widehat{\mathcal{B}}_1, \ldots, \widehat{\mathcal{B}}_J\}$ — Binscatter Basis: $\widehat{\mathbf{b}}_s(x)$.
- Dimension: $[(p+1)J - (J-1)s] + d$ — Restrictions: $0 \leq s, v \leq p$.

Residualized Binscatter (a No, No!):

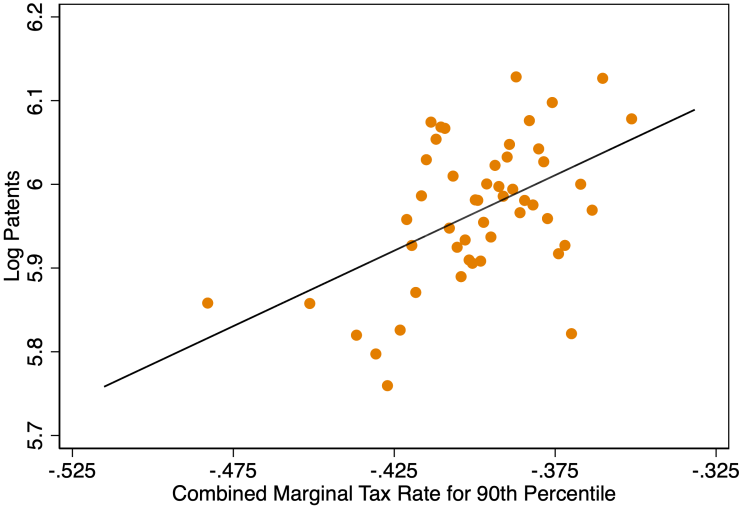
$$\widetilde{\mu}(x) = \widehat{\mathbf{b}}(x)'\widetilde{\boldsymbol{\beta}}, \qquad \widetilde{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\min} \sum_{i=1}^{n} (\widetilde{y}_i - \widehat{\mathbf{b}}(\widetilde{x}_i)'\boldsymbol{\beta})^2$$

where

$$\widetilde{y}_i = y_i - (1, \mathbf{w}_i)'\widehat{\boldsymbol{\delta}}_{y.w} \qquad \text{and} \qquad \widetilde{x}_i = x_i - (1, \mathbf{w}_i)'\widehat{\boldsymbol{\delta}}_{x.w}$$

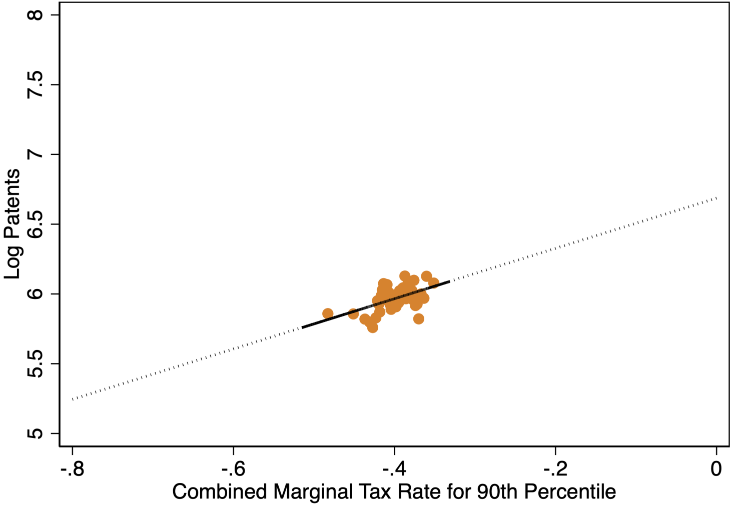

supp. narrowed

Example: Akcigit, Grigsby, Nicholas, and Stantcheva (2022, QJE)



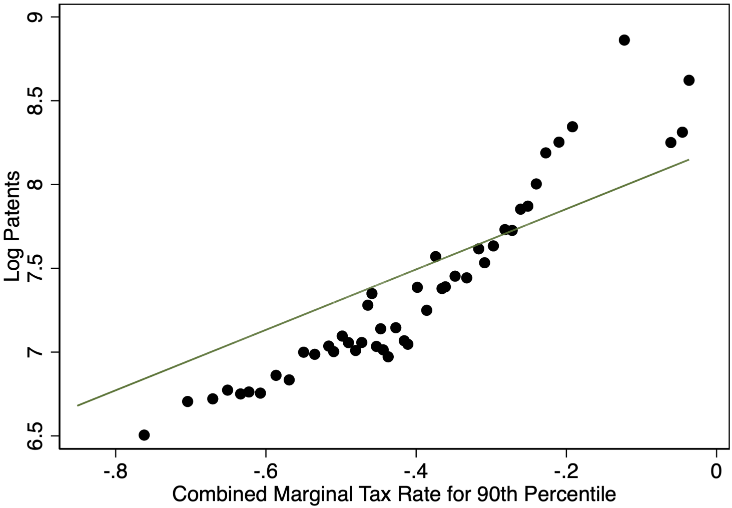
Method: Residualized binscatter (`binscatter`) – original.

Example: Akcigit, Grigsby, Nicholas, and Stantcheva (2022, QJE)



Method: Residualized binscatter (`binscatter`) – original + true data scale.

Example: Akcigit, Grigsby, Nicholas, and Stantcheva (2022, QJE)



Method: Semi-linear binscatter (`binsreg`).

Framework: Uncertainty Quantification

$$y_i = \mu(x_i) + \mathbf{w}_i'\boldsymbol{\gamma} + \epsilon_i, \qquad \mathbb{E}[\epsilon_i | x_i, \mathbf{w}_i] = 0$$

Covariate-Adjusted Binscatter:

$$\widehat{\mu}^{(v)}(x) = \widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\boldsymbol{\beta}}, \qquad \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{i=1}^{n} (y_i - \widehat{\mathbf{b}}_s(x_i)'\boldsymbol{\beta} - \mathbf{w}_i'\boldsymbol{\gamma})^2$$

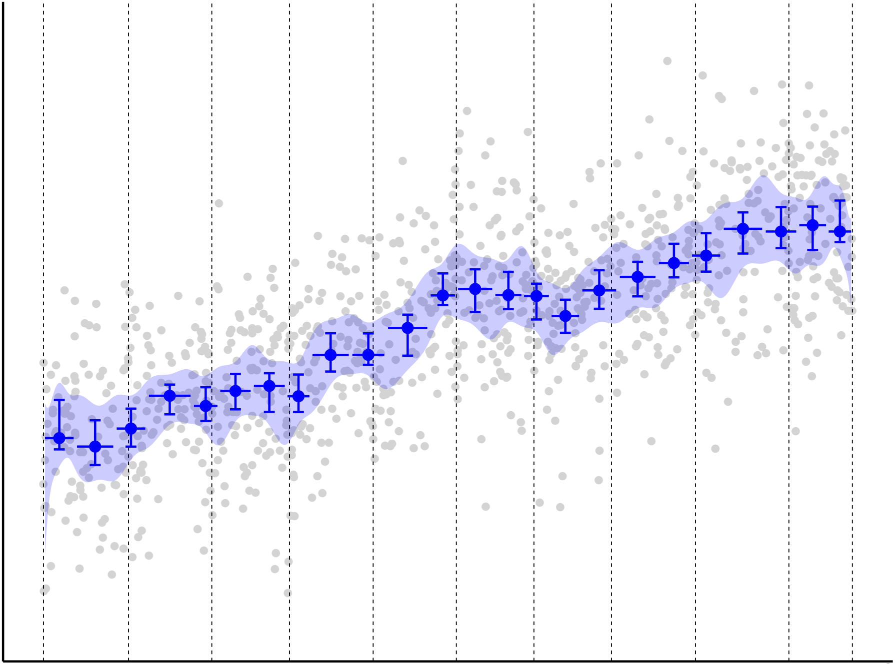
- Partitioning/Binning: $\{\widehat{\mathcal{B}}_1, \ldots, \widehat{\mathcal{B}}_J\}$ — Binscatter Basis: $\widehat{\mathbf{b}}_s(x)$.
- Dimension: $[(p+1)J - (J-1)s] + d$ — Restrictions: $0 \leq s, v \leq p$.

Confidence Intervals vs. Confidence Bands:

$$\widehat{I}_p(x) = \Big[\, \widehat{\mu}^{(v)}(x) \pm \mathfrak{c} \cdot \sqrt{\widehat{\Omega}(x)/n} \,\Big]$$

$$\text{CI} \implies \mathfrak{c} = \Phi^{-1}(1 - \alpha/2)$$

$$\text{CB} \implies \mathfrak{c} = \inf\Big\{ c \in \mathbb{R}_+ : \mathbb{P}^*\Big[\sup_{x \in \mathcal{X}} \big|\widehat{Z}_p(x)\big| \leq c \Big] \geq 1 - \alpha \Big\}$$



Framework: Specification and Shape Testing

$$y_i = \mu(x_i) + \mathbf{w}_i'\boldsymbol{\gamma} + \epsilon_i, \qquad \mathbb{E}[\epsilon_i | x_i, \mathbf{w}_i] = 0$$

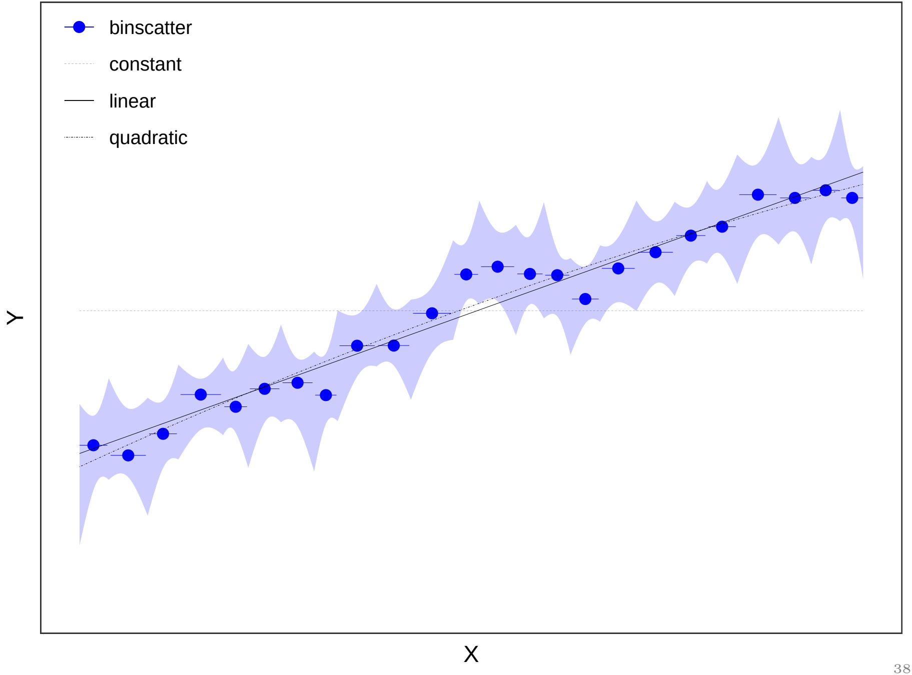
Covariate-Adjusted Binscatter:

$$\widehat{\mu}^{(v)}(x) = \widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\boldsymbol{\beta}}, \qquad \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{i=1}^{n} (y_i - \widehat{\mathbf{b}}_s(x_i)'\boldsymbol{\beta} - \mathbf{w}_i'\boldsymbol{\gamma})^2$$

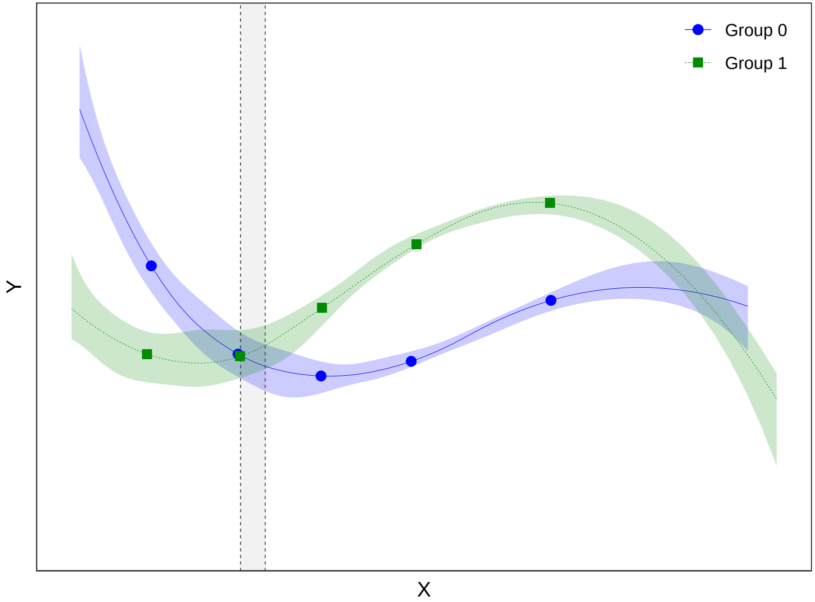
- Partitioning/Binning: $\{\widehat{\mathcal{B}}_1, \ldots, \widehat{\mathcal{B}}_J\}$ — Binscatter Basis: $\widehat{\mathbf{b}}_s(x)$.
- Dimension: $[(p+1)J - (J-1)s] + d$ — Restrictions: $0 \le s, v \le p$.

Questions:

- Is $\mu(x)$ constant, linear or quadratic?
- Is $\mu(x)$ positive, increasing or convex?
- What about $\mathbb{E}[y_i | x_i = x, \mathbf{w}_i = \mathbf{w}]$?
- What about more general regression-like models?


binscatter
constant
linear
quadratic
Y
X

Application: Treatment Effect Heterogeneity



Framework: Other Parameters & QMLE

QMLE Binscatter:

$$\widehat{\mu}^{(v)}(x) = \widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\boldsymbol{\beta}}, \qquad \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \underset{\boldsymbol{\beta},\boldsymbol{\gamma}}{\arg\min} \sum_{i=1}^{n} \rho(y_i - \eta(\widehat{\mathbf{b}}_s(x_i)'\boldsymbol{\beta} + \mathbf{w}_i'\boldsymbol{\gamma})).$$

- $\rho(u) = u^2 \implies$ Binscatter ($\eta(u) = u$), GLM Binscatter ($\eta(u) = \Lambda(u)$).
- $\rho(u;\tau) = (2\tau - 1)(y - u) + |y - u| \implies \tau$-th Quantile Binscatter.
- Huber loss, MLE, etc.

Parameters of interest:

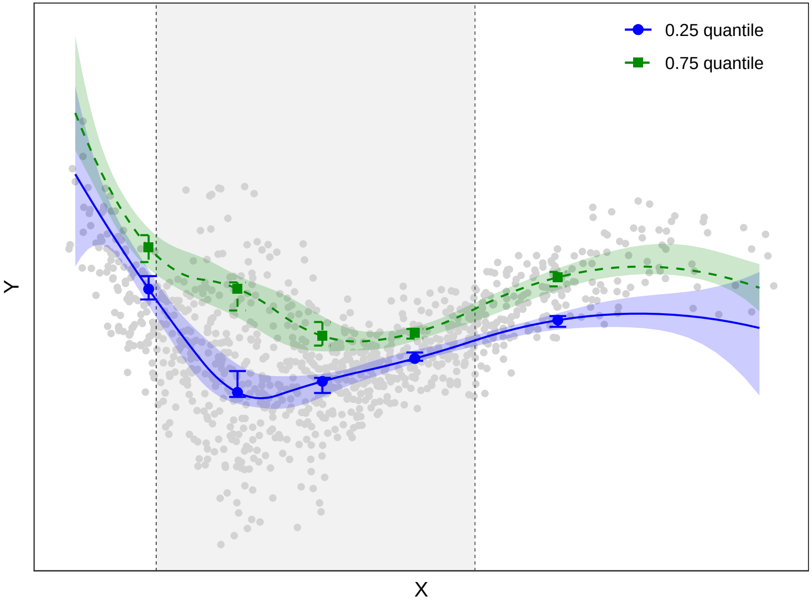
$$(\mu_0(\cdot), \boldsymbol{\gamma}_0) = \underset{\mu \in \mathcal{M}, \boldsymbol{\gamma} \in \mathbb{R}^d}{\arg\min} \ \mathbb{E}[\rho(y_i; \eta(\mu(x_i) + \mathbf{w}_i'\boldsymbol{\gamma}))]$$

$$\vartheta(x, \boldsymbol{a}_w) = \eta(\mu_0(x) + \boldsymbol{a}_w'\boldsymbol{\gamma}_0) \quad \text{and} \quad \vartheta_x^{(1)}(x, \boldsymbol{a}_w) = \frac{\partial}{\partial x}\vartheta(x, \mathbf{w})\Big|_{\mathbf{w}=\boldsymbol{a}_w}$$

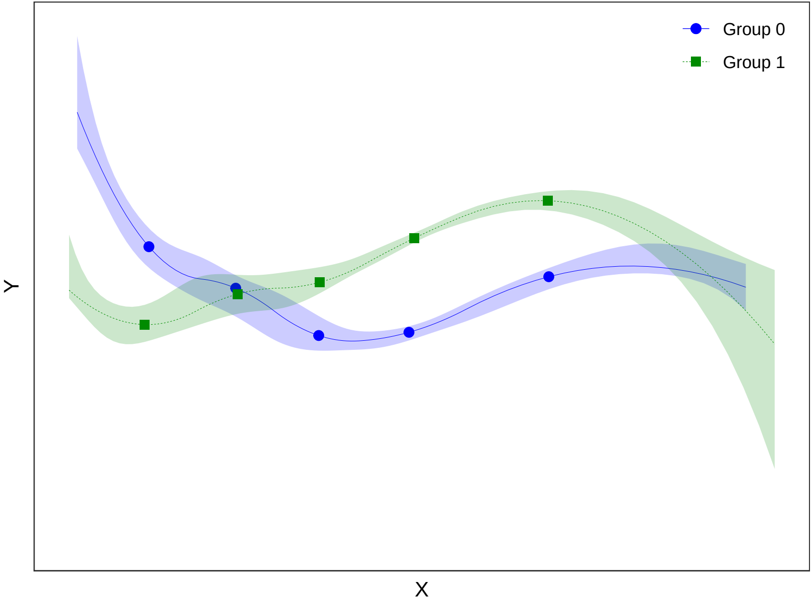
Generalized Binscatter:

$$\widehat{\vartheta}(x, \widehat{\boldsymbol{a}}_w) = \eta(\widehat{\mu}(x) + \widehat{\boldsymbol{a}}_w'\widehat{\boldsymbol{\gamma}}) \quad \text{and} \quad \widehat{\vartheta}_x^{(1)}(x, \widehat{\boldsymbol{a}}_w) = \eta^{(1)}(\widehat{\mu}(x) + \widehat{\boldsymbol{a}}_w'\widehat{\boldsymbol{\gamma}})\widehat{\mu}^{(1)}(x)$$

Application: Quantile Semi-Parametric Regression



Application: Treatment Effect Heterogeneity



Outline

IMSE-Optimal Partitioning/Binning

$$\widehat{\mu}^{(v)}(x) = \widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\boldsymbol{\beta}}, \qquad \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \underset{\boldsymbol{\beta},\boldsymbol{\gamma}}{\arg\min} \sum_{i=1}^{n} (y_i - \widehat{\mathbf{b}}_s(x_i)'\boldsymbol{\beta} - \mathbf{w}_i'\boldsymbol{\gamma})^2$$

- Partitioning/Binning: $\{\widehat{\mathcal{B}}_1, \ldots, \widehat{\mathcal{B}}_J\}$, with $\widehat{\mathcal{B}}_j = \big[x_{(\lfloor n(j-1)/J \rfloor)}, x_{(\lfloor nj/J \rfloor)}\big)$.

- IMSE Expansion:

$$\int \left(\widehat{\mu}^{(v)}(x) - \mu^{(v)}(x)\right)^2 f(x)dx \;\approx_{\mathbb{P}}\; \frac{J^{1+2v}}{n}\mathscr{V}_n(p,s,v) + J^{-2(p+1-v)}\mathscr{B}_n(p,s,v)$$

- IMSE-optimal choice:

$$J_{\texttt{IMSE}} = \left\lceil \left(\frac{2(p-v+1)\mathscr{B}_n(p,s,v)}{(1+2v)\mathscr{V}_n(p,s,v)}\right)^{\frac{1}{2p+3}} \; n^{\frac{1}{2p+3}} \right\rceil$$

- Result handles estimated quantiles. Evenly-Spaced binning also studied.

IMSE-Optimal Partitioning/Binning

$$\widehat{\mu}^{(v)}(x) = \widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\boldsymbol{\beta}}, \qquad \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \operatorname*{arg\,min}_{\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{i=1}^{n} (y_i - \widehat{\mathbf{b}}_s(x_i)'\boldsymbol{\beta} - \mathbf{w}_i'\boldsymbol{\gamma})^2$$

- IMSE-optimal choice (fixed p and s):

$$J_{\texttt{IMSE}}(p,s) = \left\lceil \left(\frac{2(p-v+1)\mathscr{B}_n(p,s,v)}{(1+2v)\mathscr{V}_n(p,s,v)} \right)^{\frac{1}{2p+3}} n^{\frac{1}{2p+3}} \right\rceil$$

- Alternative: set $J = \texttt{J}$ ($\texttt{J} = 20$, say) $\implies$ choose p (and s):

$$p_{\texttt{IMSE}} = \operatorname*{arg\,min}_{p\in\mathbb{N}_0} \left| J_{\texttt{IMSE}}(p,p) - \texttt{J} \right|$$

- Implementations: set $J = \texttt{J}$ ($\texttt{J} = 20$, say) $\implies$ choose p (and s):

$$\widehat{J}_{\texttt{IMSE}}(p,s) = \left\lceil \widehat{\mathscr{C}}_n(p,s,v)^{\frac{1}{2p+3}} \; n^{\frac{1}{2p+3}} \right\rceil, \qquad \widehat{p}_{\texttt{IMSE}} = \operatorname*{arg\,min}_{p\in\mathbb{N}_0} \left| \widehat{J}_{\texttt{IMSE}}(p,p) - \texttt{J} \right|$$

Pointwise Inference: Confidence Intervals

$$\widehat{T}_p(x) = \frac{\widehat{\mu}^{(v)}(x) - \mu^{(v)}(x)}{\sqrt{\widehat{\Omega}(x)/n}}, \qquad 0 \leq v, s \leq p$$

$$\widehat{\Omega}(x) = \widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\mathbf{Q}}^{-1}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{Q}}^{-1}\widehat{\mathbf{b}}_s^{(v)}(x), \quad \widehat{\mathbf{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{b}}_s(x_i)\widehat{\mathbf{b}}_s(x_i)'(y_i - \widehat{\mathbf{b}}_s(x_i)'\widehat{\boldsymbol{\beta}} - \mathbf{w}_i'\widehat{\boldsymbol{\gamma}})^2$$

- Distributional Approximation:

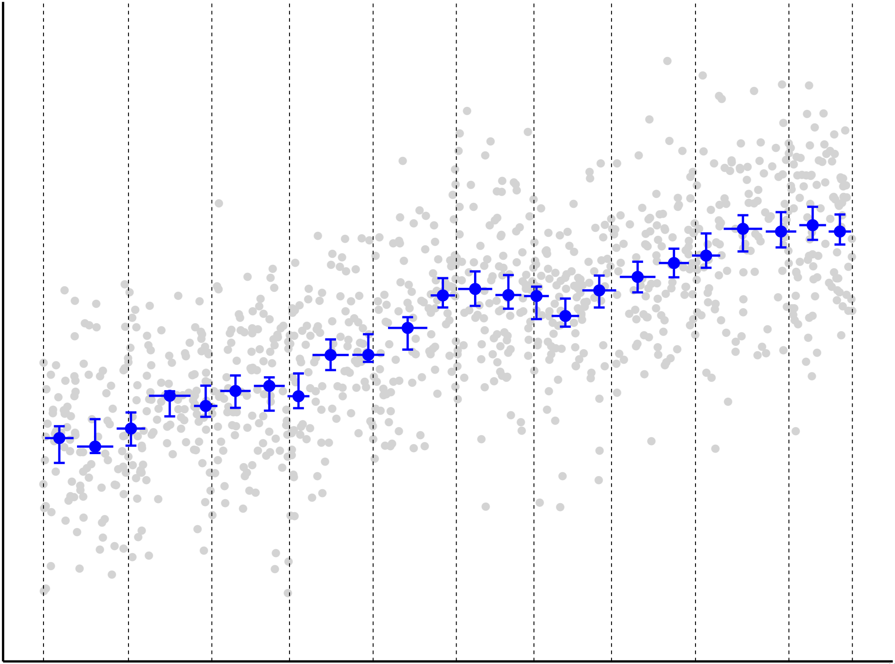
$$\sup_{u\in\mathbb{R}} \left| \mathbb{P}\big[\widehat{T}_p(x) \leq u\big] - \Phi(u) \right| \to 0, \qquad \text{for each } x \in \mathcal{X}$$

- Valid Confidence Intervals: $J = J_{\texttt{IMSE}}$ for p, then for $q \geq 1$,

$$\mathbb{P}\Big[\mu^{(v)}(x) \in \widehat{I}_{p+q}(x)\Big] \to 1 - \alpha, \qquad \text{for all } x \in \mathcal{X},$$

where

$$\widehat{I}_p(x) = \Big[\ \widehat{\mu}^{(v)}(x) \pm \mathfrak{c} \cdot \sqrt{\widehat{\Omega}(x)/n}\ \Big], \qquad \mathfrak{c} = \Phi^{-1}(1 - \alpha/2).$$



Uniform Inference

Main Goal: Approximate the "distribution" of the stochastic process

$$\left\{ \widehat{T}_p(x) = \frac{\widehat{\mu}^{(v)}(x) - \mu^{(v)}(x)}{\sqrt{\widehat{\Omega}(x)/n}} : x \in \mathcal{X} \right\}, \qquad 0 \leq v, s \leq p$$

- Useful to approximate distribution of statistics such as

$$\sup_{x \in \mathcal{X}} |\widehat{T}_p(x)|, \qquad \sup_{x \in \mathcal{X}} \widehat{T}_p(x), \qquad \inf_{x \in \mathcal{X}} \widehat{T}_p(x), \qquad \text{etc.}$$

- New strong approximation approach (based on Hungarian construction):

$$\sup_{x \in \mathcal{X}} \left| \widehat{T}_p(x) - Z_p(x) \right| = o_{\mathbb{P}}(r_n), \qquad Z_p(x) = \frac{\widehat{\mathbf{b}}_0^{(v)}(x)' \mathbf{T}_s' \mathbf{Q}^{-1} \mathbf{\Sigma}^{1/2} \mathbf{N}_K}{\sqrt{\Omega(x)}},$$

where

$$\mathbf{N}_K \sim \mathcal{N}(\mathbf{0}, \mathbf{I_K}), \qquad \widehat{\mathbf{Q}} \approx_{\mathbb{P}} \mathbf{Q}, \qquad \widehat{\mathbf{T}}_s \approx_{\mathbb{P}} \mathbf{T}_s, \qquad \widehat{\Omega}(x) \approx_{\mathbb{P}} \Omega(x), \qquad \text{etc.}$$

Uniform Inference: Heuristics of Technical Idea (4 Steps)

1. Hats off, except non-uniform-controlled partitioning scheme:

$$\sup_{x\in\mathcal{X}} |\widehat{T}_p(x) - t_p(x)| = o_{\mathbb{P}}(r_n), \qquad t_p(x) = \frac{\widehat{\mathbf{b}}_0^{(v)}(x)'\mathbf{T}_s'\mathbf{Q}^{-1}\mathbb{G}_n[\mathbf{b}_s(x_i)\epsilon_i]}{\sqrt{\Omega(x)}}$$

Uniform Inference: Heuristics of Technical Idea (4 Steps)

1. Hats off, except non-uniform-controlled partitioning scheme:

$$\sup_{x\in\mathcal{X}} |\widehat{T}_p(x) - t_p(x)| = o_{\mathbb{P}}(r_n), \qquad t_p(x) = \frac{\widehat{\mathbf{b}}_0^{(v)}(x)'\mathbf{T}_s'\mathbf{Q}^{-1}\mathbb{G}_n[\mathbf{b}_s(x_i)\epsilon_i]}{\sqrt{\Omega(x)}}$$

2. Coupling to conditional Gaussian Process (Hungarian construction):

$$\sup_{x\in\mathcal{X}} |t_p(x) - z_p(x)| = o_{\mathbb{P}}(r_n), \qquad z_p(x) = \frac{\widehat{\mathbf{b}}_0^{(v)}(x)'\mathbf{T}_s'\mathbf{Q}^{-1}\mathbb{G}_n[\mathbf{b}_s(x_i)\sigma(x_i)\zeta_i]}{\sqrt{\Omega(x)}}$$

Uniform Inference: Heuristics of Technical Idea (4 Steps)

1. Hats off, except non-uniform-controlled partitioning scheme:

$$\sup_{x \in \mathcal{X}} |\widehat{T}_p(x) - t_p(x)| = o_{\mathbb{P}}(r_n), \qquad t_p(x) = \frac{\widehat{\mathbf{b}}_0^{(v)}(x)'\mathbf{T}_s'\mathbf{Q}^{-1}\mathbb{G}_n[\mathbf{b}_s(x_i)\epsilon_i]}{\sqrt{\Omega(x)}}$$

2. Coupling to conditional Gaussian Process (Hungarian construction):

$$\sup_{x \in \mathcal{X}} |t_p(x) - z_p(x)| = o_{\mathbb{P}}(r_n), \qquad z_p(x) = \frac{\widehat{\mathbf{b}}_0^{(v)}(x)'\mathbf{T}_s'\mathbf{Q}^{-1}\mathbb{G}_n[\mathbf{b}_s(x_i)\sigma(x_i)\zeta_i]}{\sqrt{\Omega(x)}}$$

3. Coupling to unconditional (up to non-uniform partitioning) Gaussian Process:

$$\sup_{x \in \mathcal{X}} |z_p(x) - Z_p(x)| = o_{\mathbb{P}}(r_n), \qquad Z_p(x) = \frac{\widehat{\mathbf{b}}_0^{(v)}(x)'\mathbf{T}_s'\mathbf{Q}^{-1}\boldsymbol{\Sigma}\boldsymbol{\zeta}}{\sqrt{\Omega(x)}}, \qquad \boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I_K})$$

Uniform Inference: Heuristics of Technical Idea (4 Steps)

1. Hats off, except non-uniform-controlled partitioning scheme:

$$\sup_{x\in\mathcal{X}} |\widehat{T}_p(x) - t_p(x)| = o_{\mathbb{P}}(r_n), \qquad t_p(x) = \frac{\widehat{\mathbf{b}}_0^{(v)}(x)'\mathbf{T}_s'\mathbf{Q}^{-1}\mathbb{G}_n[\mathbf{b}_s(x_i)\epsilon_i]}{\sqrt{\Omega(x)}}$$

2. Coupling to conditional Gaussian Process (Hungarian construction):

$$\sup_{x\in\mathcal{X}} |t_p(x) - z_p(x)| = o_{\mathbb{P}}(r_n), \qquad z_p(x) = \frac{\widehat{\mathbf{b}}_0^{(v)}(x)'\mathbf{T}_s'\mathbf{Q}^{-1}\mathbb{G}_n[\mathbf{b}_s(x_i)\sigma(x_i)\zeta_i]}{\sqrt{\Omega(x)}}$$

3. Coupling to unconditional (up to non-uniform partitioning) Gaussian Process:

$$\sup_{x\in\mathcal{X}} |z_p(x) - Z_p(x)| = o_{\mathbb{P}}(r_n), \qquad Z_p(x) = \frac{\widehat{\mathbf{b}}_0^{(v)}(x)'\mathbf{T}_s'\mathbf{Q}^{-1}\boldsymbol{\Sigma}\boldsymbol{\zeta}}{\sqrt{\Omega(x)}}, \qquad \boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I_K})$$

4. For example, supremum approximation (with hats back on):

$$\sup_{u\in\mathbb{R}} \left| \mathbb{P}\left[\sup_{x\in\mathcal{X}} |\widehat{T}_p(x)| \leq u\right] - \mathbb{P}^*\left[\sup_{x\in\mathcal{X}} |\widehat{Z}_p(x)| \leq u\right] \right| = o_{\mathbb{P}}(1)$$

Uniform Inference: Confidence Bands

$$\sup_{u\in\mathbb{R}}\left|\mathbb{P}\Big[\sup_{x\in\mathcal{X}}|\widehat{T}_p(x)|\leq u\Big]-\mathbb{P}^*\Big[\sup_{x\in\mathcal{X}}|\widehat{Z}_p(x)|\leq u\Big]\right|=o_{\mathbb{P}}(1)$$

$$\widehat{Z}_p(x)=\frac{\widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\mathbf{Q}}^{-1}\widehat{\mathbf{\Sigma}}^{1/2}}{\sqrt{\widehat{\Omega}(x)}}\mathbf{N}_K,\qquad \mathbf{N}_K\sim\mathcal{N}(\mathbf{0},\mathbf{I_K})$$

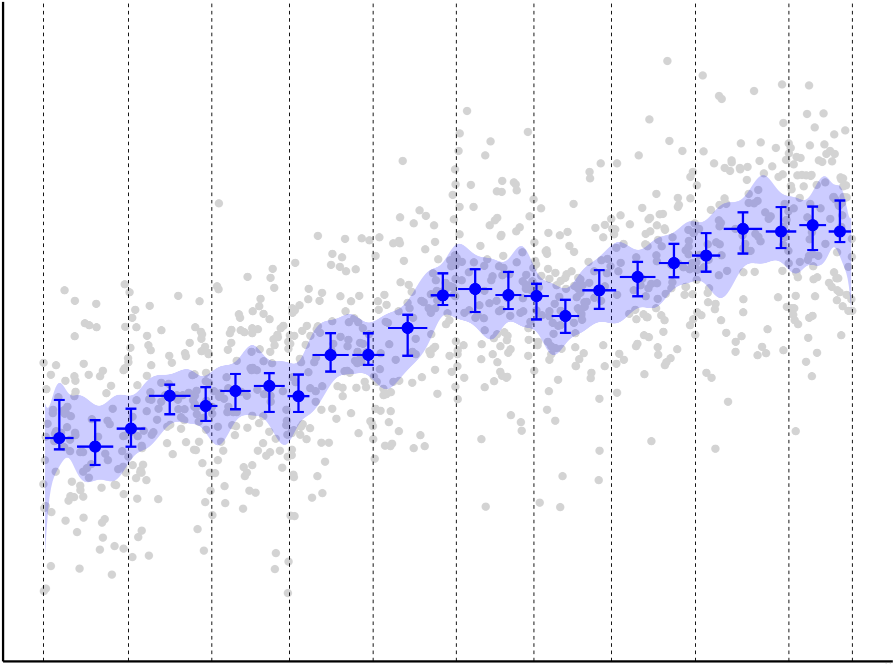
- Valid Confidence Band: $J=J_{\mathtt{IMSE}}$ for p, then for $q\geq 1$,

$$\mathbb{P}\Big[\mu^{(v)}(x)\in\widehat{I}_{p+q}(x),\ \text{for all } x\in\mathcal{X}\Big]\to 1-\alpha,$$

where

$$\widehat{I}_p(x)=\Big[\ \widehat{\mu}^{(v)}(x)\pm\mathfrak{c}\cdot\sqrt{\widehat{\Omega}(x)/n}\ \Big],$$

$$\mathfrak{c}=\inf\Big\{c\in\mathbb{R}_+:\mathbb{P}^*\Big[\sup_{x\in\mathcal{X}}\big|\widehat{Z}_p(x)\big|\leq c\Big]\geq 1-\alpha\Big\}$$



Uniform Inference: Parametric Specification Testing

$$\ddot{\mathsf{H}}_0 : \sup_{x\in\mathcal{X}} \left|\mu^{(v)}(x) - m^{(v)}(x,\boldsymbol{\theta})\right| = 0 \quad \text{for some } \boldsymbol{\theta}\in\boldsymbol{\Theta} \qquad \text{vs.} \qquad \ddot{\mathsf{H}}_\text{A} : \sup_{x\in\mathcal{X}} \left|\mu^{(v)}(x) - m^{(v)}(x,\boldsymbol{\theta})\right| > 0 \quad \text{for all } \boldsymbol{\theta}\in\boldsymbol{\Theta}$$

- Test statistic: for $\widehat{\boldsymbol{\theta}}$ and $m(\cdot)$ "well-behaved" under $\ddot{\mathsf{H}}_0$ and $\ddot{\mathsf{H}}_\text{A}$,

$$\ddot{T}_p(x) = \frac{\widehat{\mu}^{(v)}(x) - m^{(v)}(x,\widehat{\boldsymbol{\theta}})}{\sqrt{\widehat{\Omega}(x)/n}}, \qquad 0 \le v, s \le p,$$

  - For given p set $J = J_{\texttt{IMSE}}$, and for $q \ge 1$ set

$$\mathfrak{c} = \inf\left\{c\in\mathbb{R}_+ : \mathbb{P}^*\left[\sup_{x\in\mathcal{X}} |\widehat{Z}_{p+q}(x)| \le c\right] \ge 1-\alpha\right\}$$

  - Under $\ddot{\mathsf{H}}_0$, then

$$\lim_{n\to\infty} \mathbb{P}\left[\sup_{x\in\mathcal{X}} \left|\ddot{T}_{p+q}(x)\right| > \mathfrak{c}\right] = \alpha,$$

  - Under $\ddot{\mathsf{H}}_\text{A}$, then

$$\lim_{n\to\infty} \mathbb{P}\left[\sup_{x\in\mathcal{X}} \left|\ddot{T}_{p+q}(x)\right| > \mathfrak{c}\right] = 1.$$

Uniform Inference: Shape Restriction Testing

$$\dot{\mathsf{H}}_0 : \sup_{x\in\mathcal{X}} \mu^{(v)}(x) \leq 0 \qquad \text{vs.} \qquad \dot{\mathsf{H}}_\text{A} : \sup_{x\in\mathcal{X}} \mu^{(v)}(x) > 0$$

- Test statistic:

$$\dot{T}_p(x) = \frac{\widehat{\mu}^{(v)}(x)}{\sqrt{\widehat{\Omega}(x)/n}}, \qquad 0 \leq v, s \leq p,$$

  - For given p set $J = J_{\texttt{IMSE}}$, and for $q \geq 1$ set

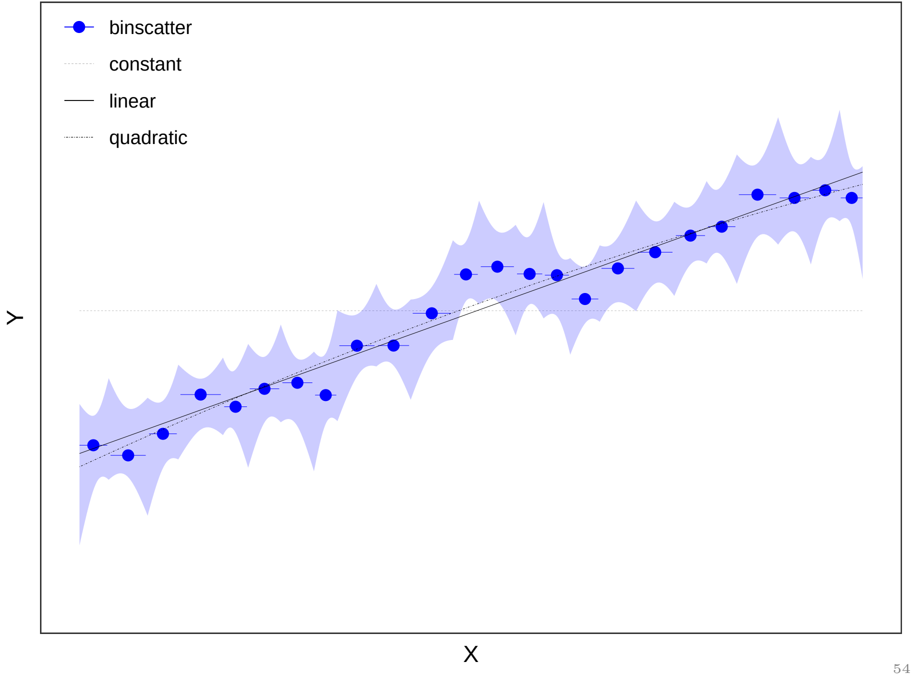
$$\mathfrak{c} = \inf\left\{c \in \mathbb{R}_+ : \mathbb{P}^*\Big[\sup_{x\in\mathcal{X}} \widehat{Z}_{p+q}(x) \leq c\Big] \geq 1-\alpha\right\}$$

  - Under $\dot{\mathsf{H}}_0$, then

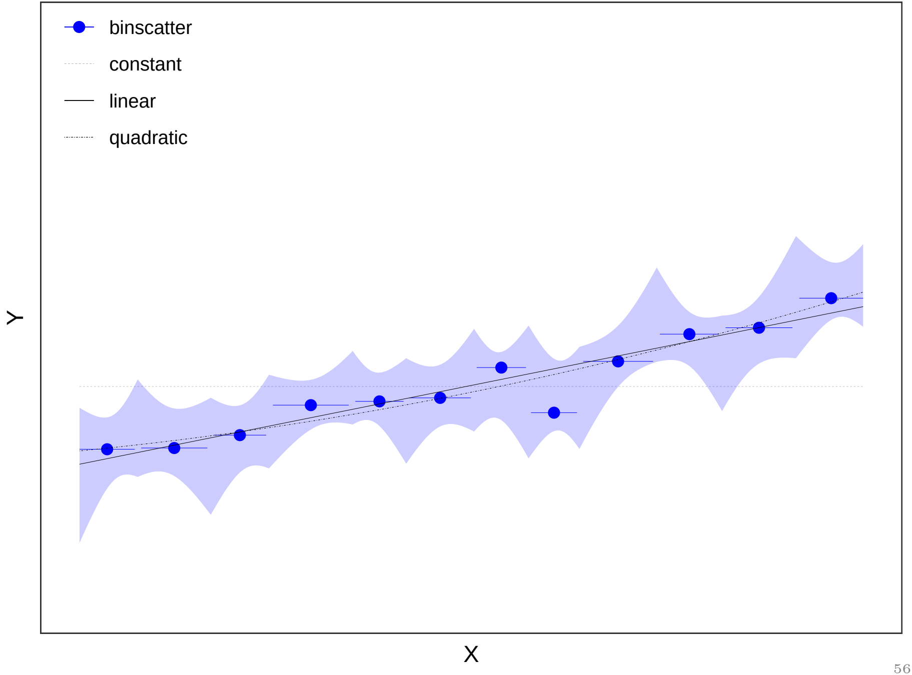
$$\lim_{n\to\infty} \mathbb{P}\Big[\sup_{x\in\mathcal{X}} \dot{T}_{p+q}(x) > \mathfrak{c}\Big] \leq \alpha,$$

  - Under $\dot{\mathsf{H}}_\text{A}$, then

$$\lim_{n\to\infty} \mathbb{P}\Big[\sup_{x\in\mathcal{X}} \dot{T}_{p+q}(x) > \mathfrak{c}\Big] = 1.$$


binscatter
constant
linear
quadratic
Y
X

	Half Support ($n = 482$)			Full Support ($n = 1000$)		
	Test Statistic	P-value	J	Test Statistic	P-value	J
Parametric Specification						
Constant	11.716	0.000	12	11.607	0.000	24
Linear	2.994	0.092	12	4.968	0.000	24
Quadratic	2.392	0.384	12	4.300	0.002	24
Shape Restrictions						
Negativity	4.069	0.000	12	12.226	0.000	24
Increasing	−1.964	0.536	13	−2.168	0.394	13
Concavity	2.269	0.316	14	2.544	0.180	14


binscatter
constant
linear
quadratic
Y
X

Uniform Inference: Generalized Binscatter

Generalized Binscatter:

$$\widehat{\mu}^{(v)}(x) = \widehat{\mathbf{b}}_s^{(v)}(x)'\widehat{\boldsymbol{\beta}}, \qquad \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \arg\min_{\boldsymbol{\beta},\boldsymbol{\gamma}} \sum_{i=1}^{n} \rho(y_i - \eta(\widehat{\mathbf{b}}_s(x_i)'\boldsymbol{\beta} + \mathbf{w}_i'\boldsymbol{\gamma})).$$

$$\widehat{\vartheta}(x, \widehat{\boldsymbol{a}}_w) = \eta(\widehat{\mu}(x) + \widehat{\boldsymbol{a}}_w'\widehat{\boldsymbol{\gamma}}) \qquad \widehat{\vartheta}_x(x, \widehat{\boldsymbol{a}}_w) = \eta^{(1)}(\widehat{\mu}(x) + \widehat{\boldsymbol{a}}_w'\widehat{\boldsymbol{\gamma}})\widehat{\mu}^{(1)}(x)$$

Uniform Bahadur Representation (up to bias of order J^{-m}):

$$\sup_{x\in\mathcal{X}} \left| \widehat{\mu}^{(v)}(x) - \mu_0^{(v)}(x) + \widehat{\mathbf{b}}_s^{(v)}(x)'\bar{\mathbf{Q}}^{-1}\mathbb{E}_n[\widehat{\mathbf{b}}_s(x_i)\eta_{i,1}\psi(\epsilon_i)] \right| \lesssim_{\mathbb{P}} J^v \left(\frac{J\log n}{n} \right)^{3/4} \sqrt{\log n}$$

$$\eta_i = \eta(\mu_0(x_i) + \mathbf{w}_i'\boldsymbol{\gamma}_0), \qquad \psi(u) = \text{weak derivative of } \rho(u), \qquad \epsilon_i = y_i - \eta_i$$

Key condition: $J^2\log(n)/n = o(1)$ — even $J\log(n)/n = o(1)$ when $s = 0$).

Outline

Overview

- Binscatter is widely used across disciplines.
- Methodological and formal results lagging behind its popularity.
- We offer a through treatment of canonical binscatter and its generalizations.
  - Formal framework: covariate-adjustment, smoothness restrictions, and more.
  - Optimal choice of partitioning/binning.
  - Confidence intervals and confidence bands.
  - Hypothesis testing for shape restrictions and for parametric specifications.
  - Quantile, non-linear least squares, and other QMLE estimation methods.
- New theoretical results for linear and non-linear partitioning-based estimators with random partitions.
- Binsreg package for `Python`, `R`, and `Stata`.

`https://nppackages.github.io/binsreg/`

References

1. Cattaneo, Farrell & Feng (2020): "Large Sample Properties of Partitioning-Based Series Estimators" *Annals of Statistics* 48(3): 1718-1741.

   - Strong approximations for least squares estimators with non-random partitions.
   - Software implementation: `https://nppackages.github.io/lspartition/`

2. Cattaneo, Crump, Farrell & Feng (2021+): "On Binscatter" *arXiv*: 1902.09608.

   - Strong approximations for QMLE partitioning-based semi-linear series estimators with random partitions.
   - Specification and shape restriction testing.
   - Software implementation: `https://nppackages.github.io/binsreg/`