Inference for Non-linear Effects in High-dimensional Additive Models

Zijian Guo

Rutgers University

UW-Madison BMI Seminar
Feb 18, 2022

Paper and code

Guo, Z., Yuan, W., & Zhang, C. (2022). Decorrelated Local Linear Estimator: Inference for Non-linear Effects in High-dimensional Additive Models. arXiv preprint arXiv:1907.12732.

DLL

CRAN 0.1.0

The goal of DLL is to implement the Decorrelated Local Linear estimator proposed in <arxiv:1907.12732>. It constructs the confidence interval for the derivative of the function of interest under the high-dimensional sparse additive model.

Installation

You can install the released version of DLL from CRAN with:

```
install.packages("DLL")
```

Example

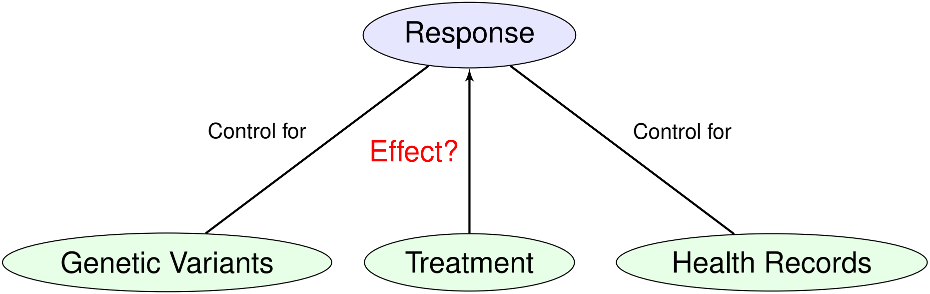
This is a basic example which shows you how to solve a common problem:

Available at `https://github.com/zijguo/HighDim-Additive-Inference`

Overview of talk

1. Motivation and Formulation
2. Decorrelation Idea
3. Decorrelated Local Linear Estimator
4. Theoretical Justification
5. Numerical Results

Treatment Effect in Observational Study



- Observational study: unmeasured confounders
- Solution: conditioning on a large set of covariates

A Convenient Solution

High-dimensional sparse linear model

$$Y_i = D_i\beta + \sum_{j=1}^{p} X_{ij}\tau_j + \epsilon_i, \quad \text{for } 1 \leq i \leq n.$$

- Number of covariates $p \gg$ sample size n.
- A few $\{\tau_j\}_{1\leq j\leq p}$ are non-zero.
- Inference for β

  Zhang & Zhang '14; Javanmard & Montanari '14; van de Geer, Bühlmann, Ritov & Dezeure '14; Chernozhukov, Hansen & Spindler '15.

MENU

nature

International journal of science

Letter | Published: 21 October 2015

Global non-linear effect of temperature on economic production

Marshall Burke, Solomon M. Hsiang & Edward Miguel

Nature **527**, 235–239 (12 November 2015) | Download Citation

Abstract

Growing evidence demonstrates that climatic conditions can have a profound impact on the functioning of modern human societies[1,2], but effects on economic activity appear inconsistent. Fundamental

学术搜索 nonlinear effects economics

找到约 827,000 条结果 （用时0.13秒）

Nonlinear effects of weather on corn yields — [PDF] oup.com
W Schlenker, MJ Roberts - Review of agricultural economics, 2006 - academic.oup.com
This paper examines the reduced-form relationship between weather and yields using a unique data set of corn yields and daily weather records covering the eastern United States for 1950–2004. Since weather variations in a fixed location are exogenous and random, the …
被引用次数：267 相关文章 所有 12 个版本 Web of Science: 100

Nonlinear regressions with integrated time series — [PDF] wiley.com, Full View
JY Park, PCB Phillips - Econometrica, 2001 - Wiley Online Library
… Crossref. Konrad Zolna, Phong B. Dao, Wieslaw J. Staszewski and Tomasz Barszcz, Nonlinear Cointegration Approach for Condition … and Ruben Schoonackers, Fiscal policy and TFP in the OECD: measuring direct and indirect effects, Empirical Economics, 49, 2, (605 …
被引用次数：404 相关文章 所有 18 个版本 Web of Science: 161

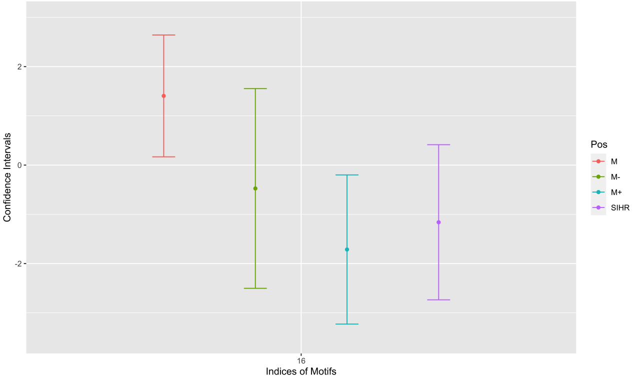
Nonlinear effects of inflation on economic growth — [PDF] jstor.org, Full View
M Sarel - Staff Papers, 1996 - Springer
… Nonlinear Effects of Inflation on Economic Growth MICHAEL SAREL* This paper examines the possibility of nonlinear effects of inflation on eco- nomic growth. It finds evidence of a significant structural break in the func- tion that relates economic growth to inflation …
被引用次数：804 相关文章 所有 11 个版本 Web of Science: 146

Non-linear Effects!

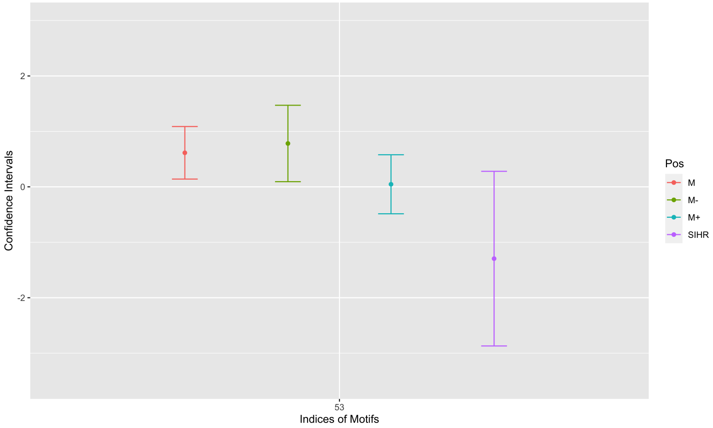
Example: Motif 16

The effect of the motifs' matching scores on the gene expression level.

- Motifs: the DNA sequences bound to transcription factors, which control the transcription activities.



Example: Motif 53



High-dimensional Additive Model

For $1 \leq i \leq n$,

$$Y_i = f(D_i) + g(X_i) + \epsilon_i \quad \text{with} \quad g(X_i) = \sum_{j=1}^{p} g_j(X_{i,j}),$$

- $Y_i \in \mathbb{R}$: outcome variable
- $D_i \in \mathbb{R}$: variable of interest
- $X_i \in \mathbb{R}^p$: high-dimensional baseline covariates

Definition of the Treatment Effect

For a pre-specified $a_0 \in \mathbb{R}$ and a small $\tau > 0$,

$$\lim_{\tau \to 0} \frac{\mathbb{E}(Y_i \mid D_i = a_0 + \tau, X_i) - \mathbb{E}(Y_i \mid D_i = a_0, X_i)}{\tau} = f'(a_0)$$

Research Problem

Inference for $f'(a_0)$ in the high-dim sparse additive model

Overview of talk

Review: Local Linear Estimator (Fan 1993)

$$Y_i = f(D_i) + \epsilon_i, \quad \text{for } 1 \leq i \leq n.$$

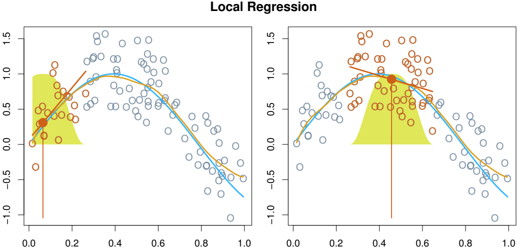


Figure: Local linear estimator from Elements of Statistical Learning

$$\left(\widehat{\beta}_0, \widehat{\beta}_1\right) = \underset{(\beta_0,\beta_1)}{\arg\min} \sum_{i=1}^{n} \left(Y_i - \beta_0 - \beta_1(D_i - a_0)\right)^2 K\left(\frac{a_0 - D_i}{h}\right)$$

Review: Weighted Average

For a pre-specified bandwidth $h > 0$, define the kernel

$$K_h(d) = \frac{1}{2h} \cdot \mathbf{1}\left(|D_i - a_0| \leq h\right).$$

The local linear estimator is expressed as

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} W_i^0 Y_i K_h(D_i)}{\sum_{i=1}^{n} W_i^0 (D_i - a_0) K_h(D_i)},$$

where

$$W_i^0 = (D_i - a_0) - \frac{\sum_{j=1}^{n} (D_j - a_0) K_h(D_j)}{\sum_{j=1}^{n} K_h(D_j)}.$$

A Plug-in Estimator?

Outcome proxy

$$\widehat{Y}_i = Y_i - \widehat{g}(X_i) = f(D_i) + \left[g(X_i) - \widehat{g}(X_i)\right] + \epsilon_i.$$

A natural plug-in estimator,

$$\widetilde{f'(a_0)} = \frac{\sum_{i=1}^{n} W_i^0 \widehat{Y}_i K_h(D_i)}{\sum_{i=1}^{n} W_i^0 (D_i - a_0) K_h(D_i)}.$$

- The local linear estimator applied to the data $\{D_i, \widehat{Y}_i\}_{1 \leq i \leq n}$
- A large bias and not ready for statistical inference

Generic Estimator

Our proposed estimator is of the form,

$$\widehat{f'(a_0)} = \frac{\frac{1}{n}\sum_{i=1}^{n} W_i \widehat{Y}_i K_h(D_i)}{\frac{1}{n}\sum_{i=1}^{n} W_i (D_i - a_0) K_h(D_i)}.$$

Generic Estimator

Our proposed estimator is of the form,

$$\widehat{f'(a_0)} = \frac{\frac{1}{n}\sum_{i=1}^{n} W_i \widehat{Y}_i K_h(D_i)}{\frac{1}{n}\sum_{i=1}^{n} W_i (D_i - a_0) K_h(D_i)}.$$

$$\widehat{f'(a_0)} - f'(a_0) = \mathrm{Err}_L + \mathrm{Err}_H$$

where

$$\mathrm{Err}_L = \frac{\frac{1}{n}\sum_{i=1}^{n} W_i [f(a_0) + r(D_i) + \epsilon_i] K_h(D_i)}{\frac{1}{n}\sum_{i=1}^{n} W_i (D_i - a_0) K_h(D_i)}$$

$$\mathrm{Err}_H = \frac{\frac{1}{n}\sum_{i=1}^{n} W_i [\widehat{g}(X_i) - g(X_i)] K_h(D_i)}{\frac{1}{n}\sum_{i=1}^{n} W_i (D_i - a_0) K_h(D_i)}.$$

Population decorrelation weight

Goal: construct the weights $\{W_i\}_{1 \le i \le n}$ such that

- Err_L is similar to that in the univariate case.
- Err_H is significantly reduced!

Define the population decorrelation weights

$$W_i = (D_i - a_0) - l(X_i) \quad \text{with} \quad l(X_i) := \frac{\mathbb{E}\left([D_i - a_0]K_h(D_i)|X_i\right)}{\mathbb{E}\left(K_h(D_i)|X_i\right)}.$$

Decorrelation property

$$\mathbb{E}\left[W_i K_h(D_i) \mid X_i\right] = 0$$

$$\mathbb{E}\left[W_i(\widehat{g}(X_i) - g(X_i))K_h(D_i) \mid X_i\right] = 0$$

Take home message

$$\widehat{f'(a_0)} = \frac{\frac{1}{n}\sum_{i=1}^{n} W_i \widehat{Y}_i K_h(D_i)}{\frac{1}{n}\sum_{i=1}^{n} W_i (D_i - a_0) K_h(D_i)},$$

with

$$W_i = (D_i - a_0) - l(X_i) \quad \text{with} \quad l(X_i) := \frac{\mathbb{E}\left([D_i - a_0] K_h(D_i) | X_i\right)}{\mathbb{E}\left(K_h(D_i) | X_i\right)}.$$

- nearly unbiased
- similar to the oracle estimator

$$\frac{\sum_{i=1}^{n} W_i^0 Y_i^{\mathrm{ora}} K_h(D_i)}{\sum_{i=1}^{n} W_i^0 (D_i - a_0) K_h(D_i)} \quad \text{with} \quad Y_i^{\mathrm{ora}} = Y_i - g(X_i) = f(D_i) + \epsilon_i.$$

Overview of talk

Treatment model

$$D_i = X_i^{\intercal}\gamma + \delta_i, \quad \text{for } 1 \leq i \leq n.$$

- γ is a sparse vector and δ_i is independent of X_i
- Let $\phi(\delta)$ denote the density function of δ_i.

$$W_i = (D_i - a_0) - l(X_i)$$

with

$$l(X_i) = \frac{\int_{\mu_i - h}^{\mu_i + h} (\delta - \mu_i)\,\phi(\delta) d\delta}{\int_{\mu_i - h}^{\mu_i + h} \phi(\delta) d\delta} \quad \text{with} \quad \mu_i = a_0 - X_i^{\intercal}\gamma.$$

Cross Fitting

Randomly split $\{1, 2, \cdots, n\}$ into two disjoint subsets $\mathcal{I}_a$ and $\mathcal{I}_b$, with $\mathcal{I}_a \cup \mathcal{I}_b = \{1, 2, \cdots, n\}$, $|\mathcal{I}_a| = \lfloor n/2 \rfloor$, and $|\mathcal{I}_b| = n - \lfloor n/2 \rfloor$. We estimate γ by

$$\widehat{\gamma}^a = \arg\min_{\gamma \in \mathbb{R}^p} \frac{1}{2|\mathcal{I}_a|} \sum_{i \in \mathcal{I}_a} \left(D_i - X_i^{\intercal}\gamma\right)^2 + \lambda_1 \sum_{j=1}^{p} \frac{\|X_{\mathcal{I}_a, j}\|_2}{\sqrt{n_a}} |\gamma_j|.$$

Estimate $\{\mu_i = a_0 - X_i^{\intercal}\gamma\}_{i \in \mathcal{I}_b}$ and $\{\delta_i = D_i - X_i^{\intercal}\gamma\}_{i \in \mathcal{I}_b}$ by

$$\widehat{\mu}_i = a_0 - X_i^{\intercal}\widehat{\gamma}^a \quad \text{and} \quad \widehat{\delta}_i = D_i - X_i^{\intercal}\widehat{\gamma}^a \quad \text{for} \quad i \in \mathcal{I}_b.$$

Kernel estimators of the weights

For $i \in \mathcal{I}_b$, we estimate $l(X_i) = \frac{\int_{\mu_i - h}^{\mu_i + h} (\delta - \mu_i)\phi(\delta) d\delta}{\int_{\mu_i - h}^{\mu_i + h} \phi(\delta) d\delta}$ by

$$\widehat{l}(X_i, \widehat{\gamma}^a) = \frac{\sum_{j \in \mathcal{I}_b} (\widehat{\delta}_j - \widehat{\mu}_i) \mathbf{1}(|\widehat{\delta}_j - \widehat{\mu}_i| \leq h)}{\sum_{j \in \mathcal{I}_b} \mathbf{1}(|\widehat{\delta}_j - \widehat{\mu}_i| \leq h)} \quad \text{for} \quad i \in \mathcal{I}_b. \qquad (1)$$

Construct the estimators of $\{l(X_i)\}_{i \in \mathcal{I}_a}$ in a similar way to (1) by switching the roles of $\mathcal{I}_a$ and $\mathcal{I}_b$.

Decorrelated Local Linear Estimator

We define the estimated weights as

$$\widetilde{W}_i = (D_i - a_0) - \widehat{l}(X_i) \quad \text{with} \quad \widehat{l}(X_i) = \begin{cases} \widehat{l}(X_i, \widehat{\gamma}^b) & \text{for } i \in \mathcal{I}_a \\ \widehat{l}(X_i, \widehat{\gamma}^a) & \text{for } i \in \mathcal{I}_b \end{cases}$$

Construct the decorrelation weights as

$$\widehat{W}_i = \widetilde{W}_i - [\sum_{j=1}^{n} \widetilde{W}_j K_h(D_j)]/[\sum_{j=1}^{n} K_h(D_j)] \quad \text{for} \quad 1 \leq i \leq n.$$

DLL

$$\widehat{f'(a_0)} = \frac{\sum_{i=1}^{n} \widehat{W}_i \widehat{Y}_i K_h(D_i)}{\sum_{i=1}^{n} \widehat{W}_i (D_i - a_0) K_h(D_i)}.$$

Initial estimators: a review

- $\Psi_{i,0} = (\phi_{0,1}(D_i), \cdots, \phi_{0,M}(D_i)) \in \mathbb{R}^M$
- $\Psi_{i,j} = (\phi_{j,1}(X_{i,j}), \cdots, \phi_{j,M}(X_{i,j})) \in \mathbb{R}^M$ for $1 \leq j \leq p$.

Define $\{\widehat{\beta}_j^a\}_{0\leq j\leq p}$ as the minimizers of

$$\underset{\beta_j\in\mathbb{R}^M,\ 0\leq j\leq p}{\arg\min} \frac{1}{2|\mathcal{I}_a|}\sum_{i\in\mathcal{I}_a}(Y_i-\sum_{j=0}^{p}\Psi_{i,j}^{\intercal}\beta_j)^2+\lambda\sum_{j=0}^{p}\sqrt{\beta_j^{\intercal}\left(\frac{1}{|\mathcal{I}_a|}\sum_{i\in\mathcal{I}_a}\Psi_{i,j}\Psi_{i,j}^{\intercal}\right)\beta_j}.$$

$$\widehat{g}^a(X_i) = \sum_{j=1}^{p}\Psi_{i,j}^{\intercal}\widehat{\beta}_j^a \quad \text{and} \quad \widehat{g}(X_i) = \begin{cases} \widehat{g}^b(X_i) & \text{for } i \in \mathcal{I}_a \\ \widehat{g}^a(X_i) & \text{for } i \in \mathcal{I}_b \end{cases}$$

Estimate σ^2 by residual sum of squares.

Statistical inference

Estimate the variance of $\widehat{f'(a_0)}$ by

$$\widehat{\mathrm{V}} = \frac{\widehat{\sigma}^2}{n^2 \widehat{S}_n^2} \sum_{i=1}^{n} \widehat{W}_i^2 K_h^2(D_i), \quad \widehat{S}_n = \frac{1}{n} \sum_{i=1}^{n} \widehat{W}_i (D_i - a_0) K_h(D_i).$$

Construct the following $1-\alpha$ confidence interval for $f'(a_0)$,

$$\mathrm{CI}[f'(a_0)] = \left(\widehat{f'(a_0)} - z_{\alpha/2}\sqrt{\widehat{\mathrm{V}}}, \widehat{f'(a_0)} + z_{\alpha/2}\sqrt{\widehat{\mathrm{V}}} \right)$$

where $z_{\alpha/2}$ denotes the upper $\alpha/2$ quantile of N(0,1).

Overview of talk

Theorem 1.

Under regularity conditions,

$$\frac{1}{\sqrt{\mathrm{V}}}\left(\widehat{f'(a_0)} - f'(a_0)\right) \xrightarrow{d} N(0,1),$$

where

$$\mathrm{V} := \frac{\sigma^2}{n^2 \widehat{S}_n^2} \sum_{i=1}^{n} \widehat{W}_i^2 K_h^2(D_i) \xrightarrow{p} \frac{3\sigma^2}{nh^3 \cdot \pi(a_0)}.$$

- Same rate as the univariate case
- Sparse additive model $Y_i = f(D_i) + g(X_i) + \epsilon_i$
- Sparse linear model $D_i = X_i^{\intercal}\gamma + \delta_i$
- A consistent initial estimator $\widehat{g}$
- Linear treatment model and independent δ_i

Overview of talk

Simulation settings

Set $p = 1500$ and generate

$$f(d) = 1.5\sin(d) \quad g_1(x) = 2\exp(-x/2) \quad g_2(x) = (x-1)^2 - 25/12$$
$$g_3(x) = x - 1/3 \quad g_4(x) = 0.75x \quad g_5(x) = 0.5x.$$

- Exactly sparse: $g_j = 0$ for $6 \le j \le p$.
- Approximately sparse:

$$g_6(x) = 0.5x \quad g_7(x) = 0.4x \quad g_8(x) = 0.3x \quad g_9(x) = 0.2x \quad g_{10}(x) = 0.1\sin(2\pi x)$$

$$g_{11}(x) = 0.2\cos(2\pi x) \quad g_{12}(x) = 0.3\sin^2(2\pi x) \quad g_{13}(x) = 0.4\cos^3(2\pi x)$$

$$g_{14}(x) = 0.5\sin^3(2\pi x) \quad g_j(x) = x/(j-1), \quad \text{for} \quad 15 \le j \le p$$

Normal. We generate $(D_i, X_i^{\intercal})^{\intercal}$ following the multivariate Normal distribution $N(\mu, \Sigma)$, where $\mu_j = -0.25$ for $1 \leq j \leq p+1$ and $\Sigma \in \mathbb{R}^{(p+1)\times(p+1)}$ is a toeplitz covariance matrix.

Approximately sparse

			Bias			RMSE			SE			Coverage			CI Length		
a_0	True	n	DLL	Plug	Orac	DLL	Plug	Orac	DLL	Plug	Orac	DLL	Plug	Orac	DLL	Plug	Orac
0.10	1.49	500	0.21	0.46	0.02	0.44	0.60	0.39	0.39	0.38	0.39	0.91	0.72	0.94	1.51	1.45	1.49
		1000	0.07	0.31	0.00	0.35	0.45	0.33	0.34	0.33	0.33	0.93	0.83	0.93	1.31	1.27	1.27
		1500	0.05	0.26	0.01	0.31	0.39	0.29	0.31	0.30	0.29	0.94	0.86	0.95	1.18	1.15	1.15
0.25	1.45	500	0.20	0.45	0.00	0.45	0.60	0.39	0.41	0.39	0.39	0.91	0.77	0.94	1.56	1.50	1.55
		1000	0.07	0.31	0.01	0.36	0.46	0.35	0.35	0.34	0.35	0.94	0.83	0.94	1.35	1.32	1.32
		1500	0.07	0.27	0.03	0.32	0.41	0.30	0.31	0.31	0.30	0.96	0.84	0.96	1.22	1.19	1.18

Uniform. We generate $(D_i^0, (X_i^0)^\intercal)^\intercal$ following $N(\mu, \Sigma)$ with the same μ and Σ as in Setting 1. We define $D_i = 5(G(D_i^0) - 0.5)$ and $X_{i,j} = 5(G(X_{i,j}^0) - 0.5)$ for $1 \leq j \leq p$, with G denoting the CDF of $N(-0.25, 1)$. The marginal distributions of D_i and $X_{i,j}$ are Uniform$(-2.5, 2.5)$ and D_i is correlated with $\{X_{i,j}\}_{1 \leq j \leq p}$.

Exactly sparse

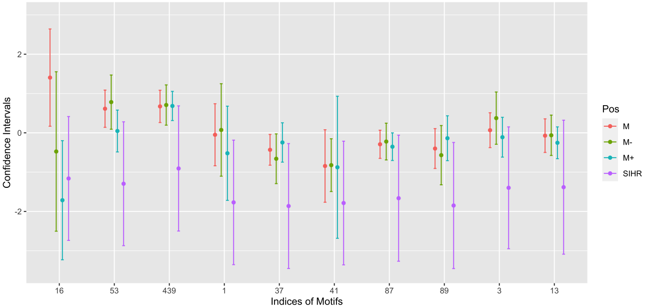
			Bias			RMSE			SE			Coverage			CI Length		
a_0	True	n	DLL	Plug	Orac	DLL	Plug	Orac	DLL	Plug	Orac	DLL	Plug	Orac	DLL	Plug	Orac
0.10	1.49	500	0.12	0.24	0.01	0.72	0.74	0.69	0.71	0.70	0.69	0.94	0.92	0.93	2.81	2.73	2.60
		1000	0.05	0.19	0.05	0.64	0.66	0.60	0.63	0.63	0.60	0.95	0.93	0.95	2.42	2.39	2.26
		1500	0.04	0.15	0.02	0.57	0.58	0.55	0.57	0.56	0.55	0.96	0.94	0.95	2.19	2.16	2.05
0.25	1.45	500	0.08	0.21	0.00	0.73	0.74	0.68	0.72	0.71	0.68	0.94	0.93	0.94	2.79	2.73	2.59
		1000	0.04	0.17	0.03	0.62	0.64	0.58	0.62	0.62	0.58	0.95	0.94	0.95	2.41	2.38	2.25
		1500	0.03	0.15	0.02	0.56	0.57	0.52	0.56	0.55	0.52	0.95	0.93	0.94	2.18	2.14	2.04

Real Data

Motif Regression studies the effect of the motifs' matching scores on the gene expression level

- Motifs: the DNA sequences bound to transcription factors, which control the transcription activities.
- A gene's expression level can be well-predicted by the matching scores of a set of motifs.
- Consists of the expression values of $n = 2587$ genes and the scores of $p + 1 = 666$ motifs.

CIs for $f'(a_0)$ by `DLL` and `SIHR`



Highly non-linear relationship: the standard deviation of the regression error is about 2.5 by `SIHR` but 1.45 by `DLL`.

$$\sqrt{\widehat{\mathrm{V}}} = \widehat{\sigma}\sqrt{\frac{1}{n^2 \widehat{S}_n^2}\sum_{i=1}^{n} \widehat{W}_i^2 K_h^2(D_i)}.$$

Semi-real Analysis

We simulate the synthetic response variable

$$Y_i^{syn} = \widehat{f}(D_i) + \widehat{g}(X_i) + \bar{\epsilon}_i, \quad \text{with} \quad \bar{\epsilon}_i \sim N(0, \widehat{\sigma}^2).$$

- Same $\{D_i, X_i\}_{1 \leq i \leq 2587}$ as the real data.
- The noise level estimator $\widehat{\sigma}^2$ and $\widehat{f}$ and $\widehat{g}$.
- 500 simulations: evaluate `DLL` on the M, $M-$, $M+$.

	Bias				SE				Coverage				Length			
Motif	M	M+	M-	SIHR	M	M+	M-	SIHR	M	M+	M-	SIHR	M	M+	M-	SIHR
1	0.08	0.15	0.14	1.37	0.23	0.45	0.32	0.41	0.93	0.87	0.94	0.45	0.94	1.73	1.19	2.66
3	0.00	0.01	0.05	1.39	0.22	0.39	0.32	0.43	0.96	0.95	0.93	0.45	0.93	1.34	1.09	2.70
13	0.06	0.09	0.02	1.35	0.27	0.41	0.31	0.42	0.96	0.96	0.97	0.59	1.12	1.61	1.18	2.88
16	0.24	0.26	0.00	1.30	0.36	0.71	0.45	0.40	0.87	0.94	0.98	0.53	1.03	2.28	2.07	2.66
37	0.09	0.15	0.33	1.44	0.20	0.55	0.53	0.42	0.93	0.95	0.95	0.44	0.77	2.20	2.16	2.75
41	0.15	0.06	0.07	1.36	0.42	0.36	0.85	0.41	0.97	0.96	0.95	0.47	1.86	1.45	3.31	2.66
53	0.22	0.12	0.08	1.35	0.25	0.36	0.27	0.41	0.89	0.94	0.96	0.49	0.93	1.30	1.02	2.66
87	0.06	0.03	0.18	1.49	0.22	0.33	0.27	0.43	0.95	0.95	0.90	0.36	0.88	1.31	1.01	2.70
89	0.04	0.07	0.12	1.50	0.27	0.43	0.34	0.41	0.95	0.95	0.93	0.35	1.06	1.54	1.21	2.78
439	0.01	0.05	0.05	1.46	0.29	0.44	0.29	0.43	0.92	0.92	0.96	0.39	0.99	1.51	1.19	2.69

Conclusion and Discussion

- Inference for $f'(a_0)$ in high-dim sparse additive model.
- Inference in high-dimensional additive model
  - Test $f = 0$?
  - Inference for $f(a_0)$?
- Model complexity and interpretation
  - Inference for interaction effects?

Model checking in high dimensions!

Other interesting stuff

- High-dimensional inference
- Causal inference with hidden confounders
- Transfer learning and semi-supervised learning
- Non-standard and post-selection inference
- Causal + Machine Learning
- ...

Reference and Acknowledgement

Guo, Z., Yuan, W., & Zhang, C. (2022). Decorrelated Local Linear Estimator: Inference for Non-linear Effects in High-dimensional Additive Models. arXiv preprint arXiv:1907.12732.

Rakshit, P., Cai, T. T., & **Guo, Z.** (2021). `SIHR`: An R Package for Statistical Inference in High-dimensional Linear and Logistic Regression Models. arXiv preprint arXiv:2109.03365.

Acknowledgement to NIH and NSF for support.

Thank You!