Quantifying lagged effects with a fused lasso penalty

Ilsang Ohn

January 10, 2017

Regression models with the lagged variable

- For $t = 1, \ldots, T$, let Y_t be a response variable at time t and x_t an exposure in which we are interested. We have pairs of the observations

$$\{(Y_t, x_t, x_{t-1}, \ldots, x_{t-L})\}_{t=L+1,\ldots,T}$$

- We want to estimate a regression function

$$g(\mathbb{E}(Y_t)) = h(x_t, \ldots, x_{t-L})$$

where g is a link function.

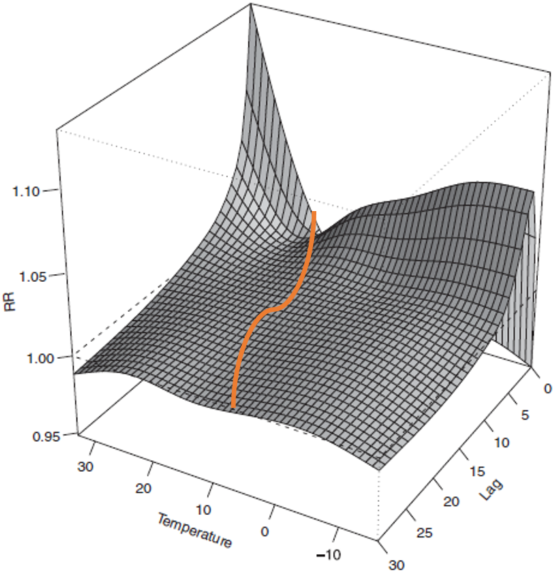
- Impose additivity:

$$g(\mathbb{E}(Y_t)) = h_0(x_t) + h_1(x_{t-1}) + \cdots + h_L(x_{t-L})$$

- Distributed lag nonlinear model: set $h_l(x) = s(x, l)$

$$g(\mathbb{E}(Y_t)) = s(x_t, 0) + s(x_{t-1}, 1) + \cdots + s(x_{t-L}, L)$$

for a bi-dimensional function s.



1.10
1.05
RR
1.00
0.95
30
20
10
Temperature
0
−10
0
5
10
15
Lag
20
25
30

Distributed lag nonlinear models

DLNMs use tensor-product splines. For two sets of basis functions $\{b_j\}_{j=1}^{J}$, $\{c_k\}_{k=1}^{K}$,

$$
\begin{aligned}
g(\mathbb{E}(Y_t)) &= \sum_{l=0}^{L} s(x_{t-l}, l) \\
&= \sum_{l=0}^{L} \sum_{j=1}^{J} \sum_{k=1}^{K} c_k(l) b_j(x_{t-l}) \tilde{\beta}_{kj} \\
&= \mathbf{x}\tilde{\boldsymbol{\beta}}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{x} &= \left(\sum_{l=0}^{L} c_1(l) b_1(x_{t-l}), \sum_{l=0}^{L} c_2(l) b_1(x_{t-l}), \ldots, \sum_{l=0}^{L} c_K(l) b_J(x_{t-l}) \right) \\
\tilde{\boldsymbol{\beta}} &= (\tilde{\beta}_{11}, \tilde{\beta}_{21}, \ldots, \tilde{\beta}_{KJ})
\end{aligned}
$$

Trial

Back to the additive structure:

$$g(\mathbb{E}(Y_t)) = h_0(x_t) + h_1(x_{t-1}) + \cdots + h_L(x_{t-L}) = \sum_{l=0}^{L} \sum_{j=1}^{J} b_j(x_{t-l})\beta_{lj}$$

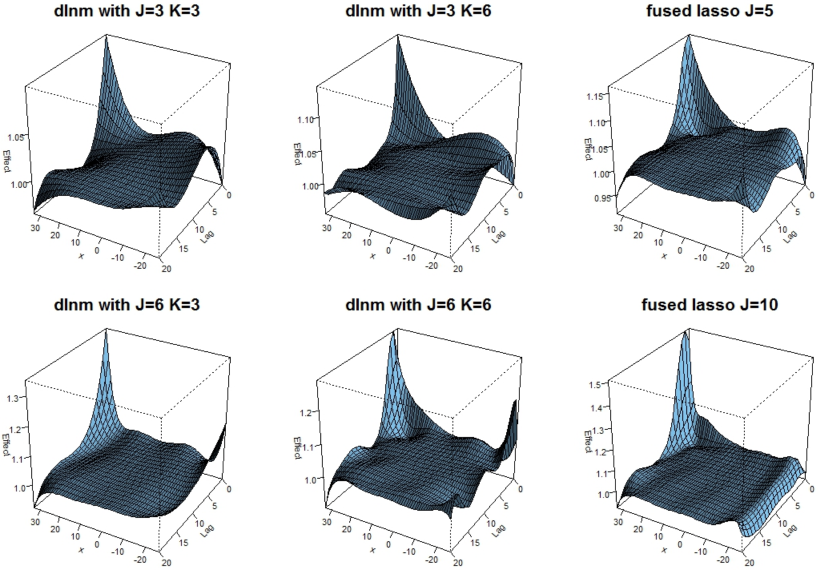
Let $\boldsymbol{\beta}_{(l)} = (\beta_{1l}, \ldots, \beta_{Jl})^\top$, $l = 0, \ldots, L$ and $\boldsymbol{\beta}^{(j)} = (\beta_{j0}, \ldots, \beta_{jL})^\top$, $j = 1, \ldots, J$

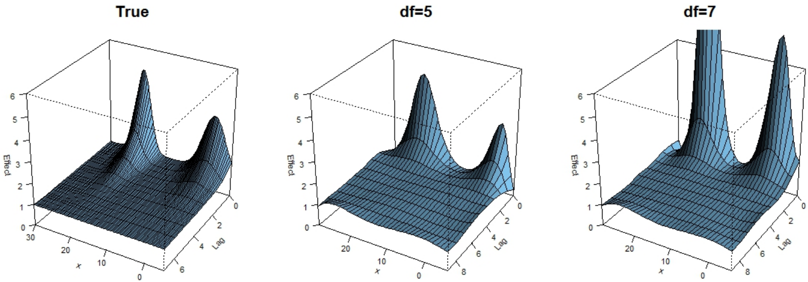
$$\begin{array}{ccccc|l} \boldsymbol{\beta}^{(1)} & \ldots & \boldsymbol{\beta}^{(j)} & \ldots & \boldsymbol{\beta}^{(J)} & \\ \beta_{01} & \ldots & \beta_{0j} & \ldots & \beta_{0J} & \boldsymbol{\beta}_{(0)} \text{ for } h_0 \\ \beta_{11} & \ldots & \beta_{1j} & \ldots & \beta_{1J} & \boldsymbol{\beta}_{(1)} \text{ for } h_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{l1} & \ldots & \beta_{lj} & \ldots & \beta_{lJ} & \boldsymbol{\beta}_{(l)} \text{ for } h_l \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{L1} & \ldots & \beta_{Lj} & \ldots & \beta_{LJ} & \boldsymbol{\beta}_{(L)} \text{ for } h_L \end{array}$$

Obtain $\hat{\boldsymbol{\beta}}$ by solving

$$\underset{\beta}{\operatorname{argmin}} - \log L(\beta) + \lambda_1 \sum_{l=0}^{L} ||D_1^{(k_1)} \boldsymbol{\beta}_{(l)}|| + \lambda_2 \sum_{j=1}^{J} ||D_2^{(k_2)} \boldsymbol{\beta}^{(j)}||$$

where $D_1^{(k+1)} \in \mathbb{R}^{(J-k_1)\times J}$, $D_2^{(k_2)} \in \mathbb{R}^{(L+1-k_2)\times(L+1)}$ are the discrete difference operators of orders k_1 and k_2 respectively.


dlnm with J=3 K=3
dlnm with J=3 K=6
fused lasso J=5
dlnm with J=6 K=3
dlnm with J=6 K=6
fused lasso J=10
Effect
x
Lag



True
df=5
df=7
Effect
x
Lag

On asymptotically optimal confidence regions and tests for high-dimensional models.

Van de Geer, S. et al., 2014, AoS

Introduction

- Consider a high dimensional regression model

$$Y = X\beta^0 + \epsilon$$

  with $n \times p$ fixed or random design matrix X and $\epsilon \sim \mathrm{N}(0, \sigma_\epsilon^2 I)$ independent of X.
- Want uncertainty quantification for Lasso estimator

$$\hat{\beta}_\lambda = \operatorname*{argmin}_{\beta \in \mathbb{R}^p} \frac{1}{2n} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

Desparsified lasso

- It is well-known that the lasso estimator fulfills the KKT conditions:

$$-\frac{1}{n}X^\top(Y - \mathbf{X}\hat{\beta}) + \lambda\hat{\kappa} = 0$$
$$\|\hat{\kappa}\|_\infty \leq 1 \text{ and } \hat{\kappa}_j = \mathrm{sign}(\hat{\beta}_j) \text{ if } \hat{\beta}_j \neq 0$$

  which can be rewritten with the notation $\widehat{\Sigma} = X^\top X/n$:

$$\widehat{\Sigma}(\hat{\beta} - \beta^0) + \lambda\hat{\kappa} = \frac{1}{n}X^\top\epsilon$$

- Suppose that $\widehat{\Theta}$ is a reasonable approximation for an inverse of $\widehat{\Sigma}$, then

$$\hat{\beta} - \beta^0 + \widehat{\Theta}\lambda\hat{\kappa} = \frac{1}{n}\widehat{\Theta}X^\top\epsilon - \frac{1}{\sqrt{n}}\Delta$$

  where

$$\Delta := \sqrt{n}(\widehat{\Theta}\widehat{\Sigma} - I)(\hat{\beta} - \beta^0)$$

- Δ is asymptotically negligible under certain sparsity assumptions. This suggests the following estimator

$$\hat{b} = \hat{\beta} + \widehat{\Theta}\lambda\hat{\kappa} = \hat{\beta} + \frac{1}{n}\widehat{\Theta}X^\top(Y - \mathbf{X}\hat{\beta})$$

Desparsified lasso

- With a suitable $\widehat{\Theta}$, we can obtain

$$\sqrt{n}(\hat{b} - \beta^0) = W + o_{\mathbb{P}}(1), \quad W|X \sim \mathrm{N}\left(0, \sigma_\epsilon^2 \widehat{\Theta}\widehat{\Sigma}\widehat{\Theta}^\top\right)$$

- An asymptotic pointwise confidence interval for β_j^0 is then given by

$$[\hat{b}_j - c(\alpha, n, \sigma_\epsilon), \hat{b}_j + c(\alpha, n, \sigma_\epsilon)]$$
$$c(\alpha, n, \sigma_\epsilon) := \Phi^{-1}(1 - \alpha/2)\sigma_\epsilon \sqrt{(\widehat{\Theta}\widehat{\Sigma}\widehat{\Theta}^\top)_{j,j}/n}$$

Desparsified lasso

- To construct the approximate inverse $\widehat{\Theta}$, we use the nodewise lasso. Let

$$\hat{\gamma}_j := \operatorname*{argmin}_{\gamma \in \mathbb{R}^{p-1}} \frac{1}{2n} \|X_j - X_{-j}\gamma\|_2^2 + \lambda_j \|\gamma\|_1$$

$$\hat{\tau}_j^2 := \frac{1}{n} \|X_j - X_{-j}\hat{\gamma}_j\|_2^2 + \lambda_j \|\hat{\gamma}_j\|_1$$

- Define

$$\widehat{\Theta}^{\text{lasso}} := \hat{T}^{-2}\hat{\Gamma}$$

where

$$\hat{\Gamma} := \begin{pmatrix} 1 & -\hat{\gamma}_{1,2} & \cdots & -\hat{\gamma}_{1,p} \\ -\hat{\gamma}_{2,1} & 1 & \cdots & -\hat{\gamma}_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ -\hat{\gamma}_{p,1} & -\hat{\gamma}_{p,2} & \cdots & 1 \end{pmatrix}$$

$$\hat{T}^2 := \mathrm{diag}(\hat{\tau}_1^2, \ldots, \hat{\tau}_p^2)$$

- From the KKT conditions, $\|\widehat{\Sigma}(\widehat{\Theta}_j^{\text{lasso}})^\top - e_j\|_\infty \leq \lambda_j/\hat{\tau}_j^2$.
- Let

$$\hat{b}^{\text{lasso}} = \hat{\beta} + \frac{1}{n}\widehat{\Theta}^{\text{lasso}} X^\top (Y - \mathbf{X}\hat{\beta})$$

Theoretical result

(A1) Let $S_0 = \{j : \beta_j^0 \neq 0\}$ be the active set of variables and $s_0 = |S_0|$ be its cardinality. Assume the sparsity

$$s_0 = o(\sqrt{n}/\log(p_n))$$

(A2) The rows of X are i.i.d. realizations from a Gaussian distribution whose p-dimensional inner product matrix Σ has strictly positive smallest eigenvalue $\Lambda^2_{\min}$ satisfying $1/\Lambda^2_{\min} = O(1)$. Furthermore, $\max_j \Sigma_{j,j} = O(1)$.

(A3) $\Theta := \Sigma^{-1}$ is low sparse. i.e, letting $s_j = \{k \neq j : \Theta_{j,k} \neq 0\}|$, then

$$\max_j s_j = o(n/\log(p_n))$$

Theoretical result

THEOREM. Consider the linear model with Gaussian error with $\sigma_\epsilon^2 = O(1)$, and assume (A1)-(A3). Consider a suitable choice of the regularization parameters $\lambda \asymp \sqrt{\log(p)/n}$ and $\lambda_j \asymp \sqrt{\log(p)/n}$ uniformly in j for the lasso for nodewise regression. Then

$$
\begin{aligned}
&\sqrt{n}(\hat{b}^{\mathrm{lasso}} - \beta^0) = W + \Delta \\
&W = \frac{1}{\sqrt{n}} \widehat{\Theta}^{\mathrm{lasso}} X^\top \epsilon \sim \mathrm{N}_n(0, \sigma_\epsilon^2 \widehat{\Omega}), \quad \widehat{\Omega} = (\widehat{\Theta}^{\mathrm{lasso}})^\top \widehat{\Sigma} \widehat{\Theta}^{\mathrm{lasso}} \\
&\|\Delta\|_\infty = o_{\mathbb{P}}(1).
\end{aligned}
$$

- For a one-dimensional component β_j^0 (with j fixed), we obtain for all $z \in \mathbb{R}$

$$\mathbb{P}\left[\frac{\sqrt{n}(\hat{b}_j^{\text{lasso}} - \beta_j^0)}{\sigma_\epsilon \sqrt{\widehat{\Omega}_{j,j}}} \leq z \Big| X\right] - \Phi(z) = o_{\mathbb{P}}(1).$$

- For any fixed group $G \subset \{1, \ldots, p\}$, we obtain for all $z \in \mathbb{R}$

$$\mathbb{P}\left[\max_{j \in G} \frac{\sqrt{n}|\hat{b}_j^{\text{lasso}} - \beta_j^0|}{\sigma_\epsilon \sqrt{\widehat{\Omega}_{j,j}}} \leq z \Big| X\right] - \mathbb{P}\left[\max_{j \in G} \frac{|W_j|}{\sigma_\epsilon \sqrt{\widehat{\Omega}_{j,j}}} \leq z \Big| X\right] = o_{\mathbb{P}}(1).$$

Conditionally on X, the asymptotic distribution of $\max_{j \in G} \frac{\sqrt{n}|\hat{b}_j^{\text{lasso}}|}{\sigma_\epsilon \sqrt{\widehat{\Omega}_{j,j}}}$ under the null-hypothesis $H_0 : \beta_j^0 = 0 \; \forall j \in G$ is asymptotically equal tot eh maximum of dependent $\chi^2(1)$ variables $\max_{j \in G} \frac{|W_j|}{\sigma_\epsilon \sqrt{\widehat{\Omega}_{j,j}}}$ whose distribution can be easily simulated since $\widehat{\Omega}$ is known.