Localizing Changes in High-Dimensional Regression Models

Alessandro Rinaldo,[1] Daren Wang,[2] Qin Wen,[3] Rebecca Willett,[3] and Yi Yu[4]

[1] Carnegie Mellon University, [2] University of Notre Dame, [3] University of Chicago, and [4] University of Warwick

Introduction

We observe (y_t, x_t) for $t = 1, \dots, T$, where $y_t = x_t\beta_t^* + \epsilon_t$. The βs are piecewise constant over t, and we want to find the times where the βs change.

Problem Formulation

- **Model:** $y_t = x_t^\top \beta_t^* + \varepsilon_t, \quad t = 1, \dots, n$, where $\{(x_t, y_t)\}_{t=1}^n \subset \mathbb{R}^p \times \mathbb{R}$. $\{\varepsilon_t\}_{t=1}^n \sim$ sub Gaussian(σ_ε^2).
  $\exists\ 1 = \eta_0 < \eta_1 < \dots < \eta_{K+1} = n+1$ such that $\beta_t^* \neq \beta_{t-1}^*$, if and only if $t \in \{\eta_k\}_{k=1}^K$.
- **Goal:** *Consistent* estimators $\{\hat{\eta}_k\}_{k=1}^{\widehat{K}}$, s.t. as the sample size $n \to \infty$, it holds with probability $\to 1$ that

$$\widehat{\mathbf{K}} = \mathbf{K} \quad \text{and} \quad \boldsymbol{\epsilon} = \max_{k=1,\dots,K} |\hat{\eta}_k - \eta_k| = o(\Delta),$$

where $\Delta = \min_{k=1,\dots,K+1} \eta_k - \eta_{k-1}$.

Dynamic Programming Approach

Let $\mathcal{P}$ be an integer interval partition of $\{1, \dots, n\}$ into $K_{\mathcal{P}}$ intervals, For a positive tuning parameter $\gamma > 0$, let

$$\widehat{\mathcal{P}} \in \arg\min_{\mathcal{P}} \left\{ \sum_{I \in \mathcal{P}} \sum_{t \in I} (y_t - x_t^\top \widehat{\beta}_I^\lambda)^2 + \gamma|\mathcal{P}| \right\}, \qquad (1)$$

where for another tuning parameter $\lambda \geq 0$,

$$\widehat{\beta}_I^\lambda = \arg\min_{v \in \mathbb{R}^p} \left\{ \sum_{t \in I} (y_t - x_t^\top v)^2 + \lambda\sqrt{\max\{|I|, \log(n \vee p)\}}\|v\|_1 \right\} \qquad (2)$$

In Theorem 1 and Corollary 2, we assume that

- **Sparsity**. $d_0 =$ the number of nonzero elements in β_t.
- $\boldsymbol{\Delta\kappa^2 \gtrsim d_0^2 K \sigma_\varepsilon^2 \log^{1+\xi}(n \vee p)}$, where $\xi > 0$, $\kappa = \min_{k=1,\dots,K+1} \|\beta_{\eta_k}^* - \beta_{\eta_{k-1}}^*\|_2$.

Theorem 1: Let $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{K}}$ be the solution to (1) and (2) with $\lambda = \mathcal{O}\left(\sqrt{d_0 \log(n \vee p)}\right)$, $\gamma = \mathcal{O}\left((K+1) d_0^2 \log(n \vee p)\right)$, then

$$\mathbb{P}\left\{ \widehat{K} = K, \max_{k=1,\dots,K} |\widetilde{\eta}_k - \eta_k| = \mathcal{O}\left(\frac{\mathbf{K d_0^2} \log(n \vee p)}{\kappa^2}\right) \right\} \geq 1 - C(n \vee p)^{-c},$$

The localization error converges to zero in probability since

$$\max_{k=1,\dots,K} \frac{|\hat{\eta}_k - \eta_k|}{\Delta} = \mathcal{O}\left(\frac{K d_0^2 \log(n \vee p)}{\kappa^2 \Delta}\right) = \mathcal{O}\left(1/\log^\xi(n \vee p)\right) \to 0.$$

Local Refinement

The localization error of the previous method can be improved using a second, more refined estimator.

Algorithm 1 Local Refinement

1: **INPUT:** Data $\{(x_t, y_t)\}_{t=1}^n$, $\{\widetilde{\eta}_k\}_{k=1}^{\widetilde{K}}$, $\zeta > 0$, $(\widetilde{\eta}_0, \widetilde{\eta}_{\widetilde{K}+1}) \leftarrow (0, n)$
2: **for** $k = 1, \dots, \widetilde{K}$ **do**
3: $(s_k, e_k) \leftarrow (2\widetilde{\eta}_{k-1}/3 + \widetilde{\eta}_k/3, \widetilde{\eta}_k/3 + 2\widetilde{\eta}_{k+1}/3)$

$$\left(\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\eta}_k\right) \leftarrow \arg\min_{\substack{\eta \in \{s_k+1, \dots, e_k-1\} \\ \beta_1, \beta_2 \in \mathbb{R}^p, \beta_1 \neq \beta_2}} \left\{ \sum_{t=s_k+1}^{\eta} \|y_t - \beta_1^\top x_t\|_2^2 + \sum_{t=\eta+1}^{e_k} \|y_t - \beta_2 x_t\|_2^2 + \zeta \sum_{i=1}^{p} \sqrt{(\eta - s_k)(\beta_1)_i^2 + (e_k - \eta)(\beta_2)_i^2} \right\} \qquad (3)$$

4: **end for**
5: **OUTPUT:** $\{\widehat{\eta}_k\}_{k=1}^{\widetilde{K}}$.

Corollary 2: Let $\{\widetilde{\eta}_k\}_{k=1}^K$ be a set of time points satisfying

$$\max_{k=1,\dots,K} |\widetilde{\eta}_k - \eta_k| \leq \Delta/7,$$

$\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ be the change point estimators generated from Algorithm 1 with $\zeta = \mathcal{O}\left(\sqrt{\log(n \vee p)}\right)$ as inputs. Then,

$$\mathbb{P}\left\{ \widehat{K} = K, \max_{k=1,\dots,K} |\hat{\eta}_k - \eta_k| = \mathcal{O}\left(\frac{\mathbf{d_0} \log(n \vee p)}{\kappa^2}\right) \right\} \geq 1 - Cn^{-c},$$

Compared to Theorem 1, in Algorithm 1

- The use of the **group Lasso-type penalty** brings down the error from d_0^2 to d_0;
- The refinement is done locally within each disjoint interval, so the procedure is **parallelizable**.

Lower Bounds

Lower Bounds:

- No algorithm is guaranteed to be consistent in the regime $\boldsymbol{\kappa^2\Delta \lesssim d_0\sigma_\varepsilon^2}$
- In the consistency regime, a minimax lower bound on the localization errors is $\mathbf{d_0}\boldsymbol{\sigma_\varepsilon^2\kappa^{-2}}$.

- The change point localization task is impossible when $\boldsymbol{\kappa^2\Delta \lesssim d_0\sigma_\varepsilon^2}$
- When $\boldsymbol{\Delta\kappa^2 \gtrsim d_0^2 K \sigma_\varepsilon^2 \log^{1+\xi}(n \vee p)}$, it can be solved by our algorithms at a nearly minimax optimal rate.

Numerical Experiments

Simulations: Methods: dynamic programming (**DP**), efficient binary segmentation algorithm (**EBSA**)[1], local refinement initialized by DP (**DP.LR**), and initialized by EBSA (**EBSA.LR**).

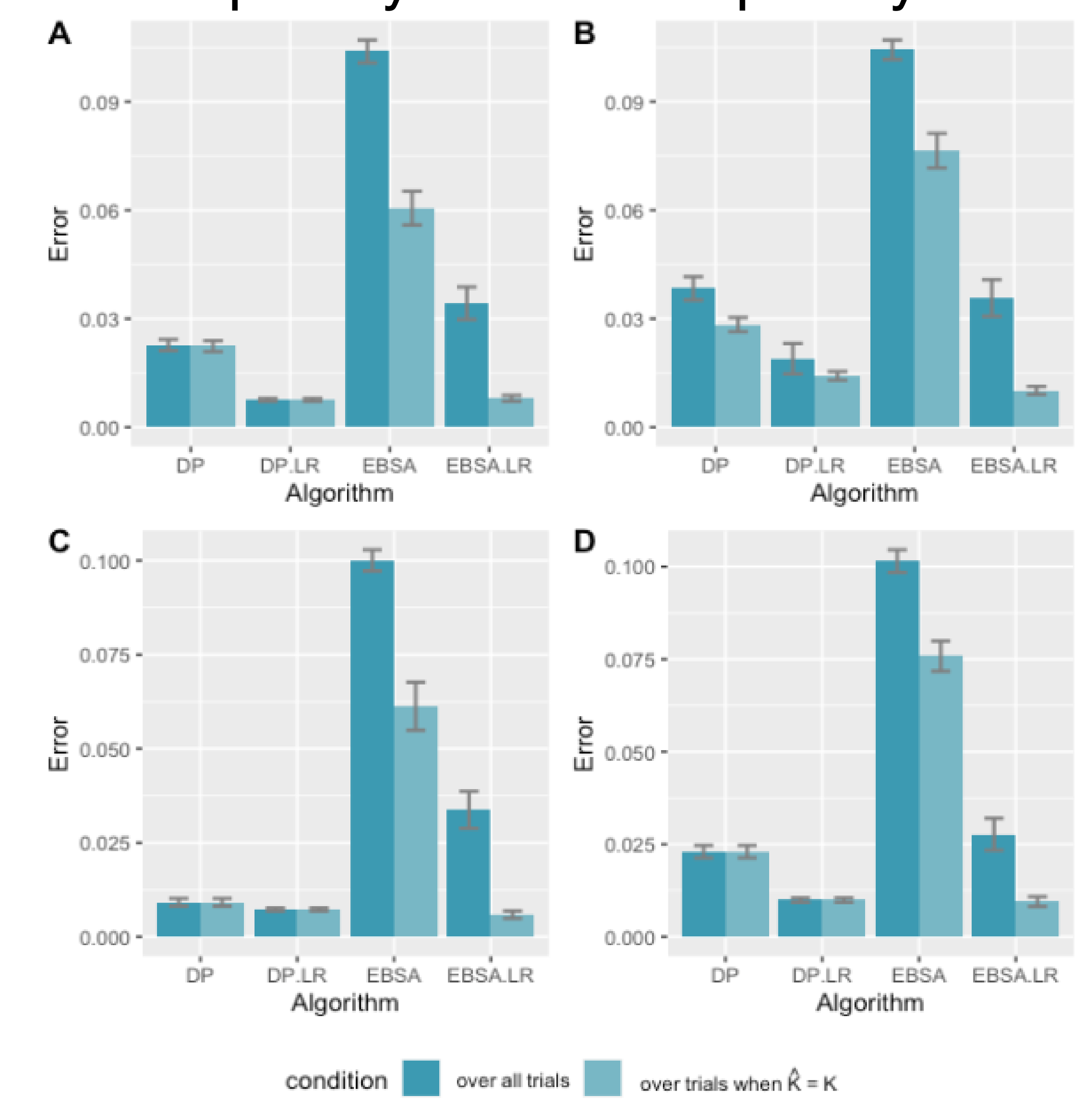
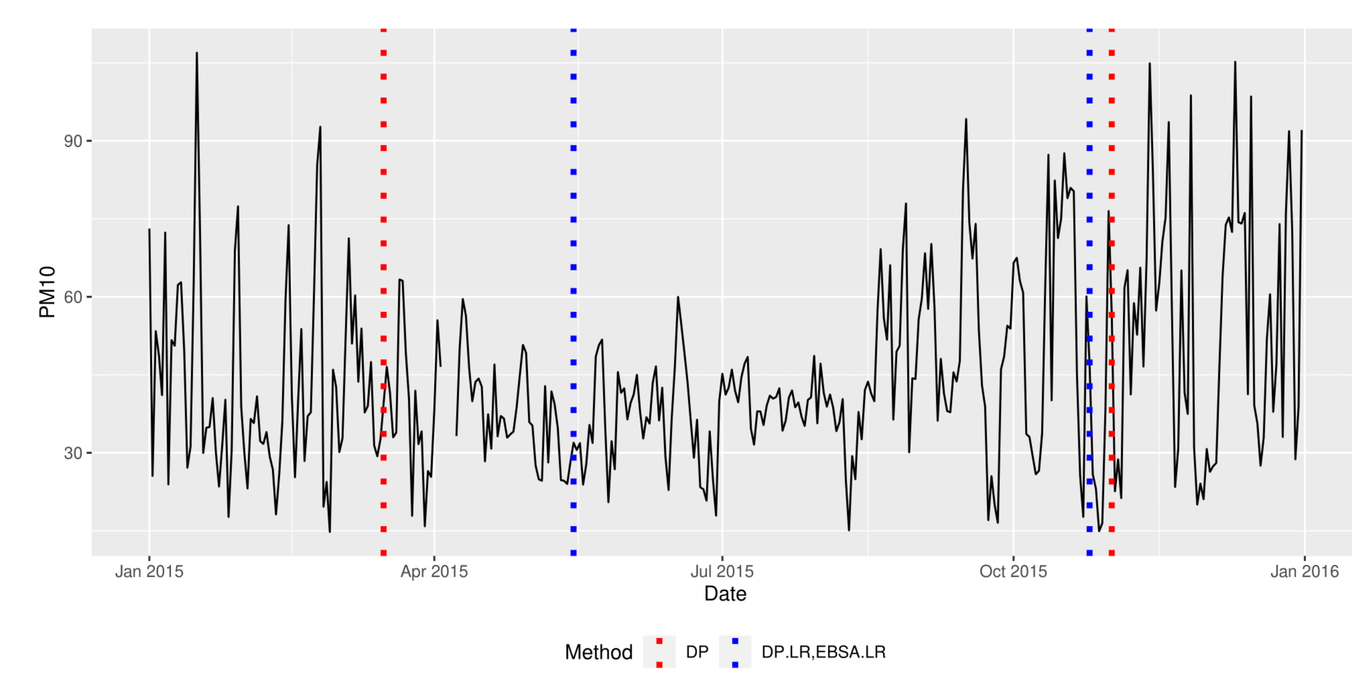


Fig. 1: Plots A,B: $\kappa = 4$, $d_0 \in \{10, 20\}$; Plots C,D: $\kappa = 6$, $d_0 \in \{10, 20\}$.

Air Quality Data in Northern Taiwan, 2015:

- Response variable: PM10 in Banqiao
- Covariates: environment factors; PM10 in other districts



- EBSA detects no changes
- DP.LR and EBSA.LR both detect May 15th (typhoon) and October 25th, 2015 as the change points (severe air pollution reaching "purple alert")

References

[1] Florencia Leonardi and Peter Bühlmann. "Computationally efficient change point detection for high-dimensional regression". In: *arXiv preprint arXiv:1601.03704* (2016).