

# Homogeneity Pursuit

**Jianqing Fan**

Princeton University

with **Tracy Ke and Yichao Wu**

<http://www.princeton.edu/~jqfan>

June 5, 2014

# Amazing Grace Wahba

**Math Genealogy:** 34 students and 204 descendants.

Grace Wahba  
Professor of Statistics, University of Wisconsin-Madison  
[Machine Learning - Statistical Model Building](#)  
Verified email at stat.wisc.edu  
[Homepage](#)

| Citation indices          |       |            |
|---------------------------|-------|------------|
|                           | All   | Since 2009 |
| <a href="#">Citations</a> | 25284 | 7881       |
| <a href="#">h-index</a>   | 60    | 30         |
| <a href="#">i10-index</a> | 130   | 70         |



Show: 20 1-20 [Next >](#)

[Title / Author](#)

[Spline models for observational data](#)

G Wahba

Siam

[Smoothing noisy data with spline functions](#)

P Craven, G Wahba

Numerische Mathematik 31 (4), 377-403

Cited  
by

[Year](#)

[5396](#) 1990

[2759](#) 1978



# Outline

- 1 Introduction
- 2 Clustering Algorithm in Regression via Data-driven Segmentation (CARDS)
- 3 Theoretical results: bCARDS & aCARDS
- 4 Numerical studies

# Introduction

# Linear regression

$$\mathbf{y} = \mathbf{X}\beta^0 + \varepsilon$$

Estimability: When  $p > n$ , structure of  $\beta$  is simple.

- ★ Sparsity of  $\beta^0$  (**known** atom 0).
- ★ Smoothness of  $\beta_i$  (against a variable): Nonparametric reg
- ★ Piecewise constant – Fused lasso (*Tibshirani et al, 05*)
- ★ **Homogeneity** (*Shen & Huang, 10*), e.g.

$$Y = \beta_1(X_1 + X_3) + \beta_2(X_2 + X_4 + X_5) + \beta_3 X_6 + \varepsilon.$$

# Linear regression

$$\mathbf{y} = \mathbf{X}\beta^0 + \varepsilon$$

Estimability: When  $p > n$ , structure of  $\beta$  is simple.

- ★ Sparsity of  $\beta^0$  (**known** atom 0).
- ★ Smoothness of  $\beta_i$  (against a variable): Nonparametric reg
- ★ Piecewise constant – Fused lasso (*Tibshirani et al, 05*)
- ★ **Homogeneity** (*Shen & Huang, 10*), e.g.

$$Y = \beta_1(X_1 + X_3) + \beta_2(X_2 + X_4 + X_5) + \beta_3X_6 + \varepsilon.$$

# Linear regression

$$\mathbf{y} = \mathbf{X}\beta^0 + \varepsilon$$

Estimability: When  $p > n$ , structure of  $\beta$  is simple.

- ★ Sparsity of  $\beta^0$  (**known** atom 0).
- ★ Smoothness of  $\beta_i$  (against a variable): Nonparametric reg
- ★ Piecewise constant – Fused lasso (*Tibshirani et al, 05*)
- ★ **Homogeneity** (*Shen & Huang, 10*), e.g.

$$Y = \beta_1(X_1 + X_3) + \beta_2(X_2 + X_4 + X_5) + \beta_3X_6 + \varepsilon.$$

# Homogeneity

**Homogeneity:**  $\beta_j^0 = \beta_{j'}^0 \ \forall j, j' \in A_k$  with  $A_1 \cup \dots \cup A_K = \{j\}_{j=1}^p$ .

**Motivation:** Reduce variance of estimators:  $\text{MSE} = \mathbf{O}(\mathbf{K}/\mathbf{n})$ .

## Examples:

- ★ Diagnostic lab tests and counting numbers of positives.
- ★ Groups of genes play similar roles in biological process.
- ★ Neighboring geographic locations share similar coefficients.
- ★ The same sector of finance share similar risk loadings

**Related Literature:** (*Park et al, 07; Friedman et al, 07; Bondell & Reich, 08; Zhu, et al, 13; Yang & He, 12.*)

# Homogeneity

Homogeneity:  $\beta_j^0 = \beta_{j'}^0 \ \forall j, j' \in A_k$  with  $A_1 \cup \dots \cup A_K = \{j\}_{j=1}^p$ .

Motivation: Reduce variance of estimators:  $\text{MSE} = \mathbf{O}(\mathbf{K}/\mathbf{n})$ .

## Examples:

- ★ Diagnostic lab tests and counting numbers of positives.
- ★ Groups of genes play similar roles in biological process.
- ★ Neighboring geographic locations share similar coefficients.
- ★ The same sector of finance share similar risk loadings

Related Literature: (*Park et al, 07; Friedman et al, 07; Bondell & Reich, 08; Zhu, et al, 13; Yang & He, 12.*)

# Homogeneity

Homogeneity:  $\beta_j^0 = \beta_{j'}^0 \ \forall j, j' \in A_k$  with  $A_1 \cup \dots \cup A_K = \{j\}_{j=1}^p$ .

Motivation: Reduce variance of estimators:  $\text{MSE} = \mathbf{O}(\mathbf{K}/\mathbf{n})$ .

## Examples:

- ★ Diagnostic lab tests and counting numbers of positives.
- ★ Groups of genes play similar roles in biological process.
- ★ Neighboring geographic locations share similar coefficients.
- ★ The same sector of finance share similar risk loadings

Related Literature: (*Park et al, 07; Friedman et al, 07; Bondell & Reich, 08; Zhu, et al, 13; Yang & He, 12.*)

# Homogeneity

Homogeneity:  $\beta_j^0 = \beta_{j'}^0 \quad \forall j, j' \in A_k$  with  $A_1 \cup \dots \cup A_K = \{j\}_{j=1}^p$ .

Motivation: Reduce variance of estimators:  $\text{MSE} = \mathbf{O}(\mathbf{K}/\mathbf{n})$ .

## Examples:

- ★ Diagnostic lab tests and counting numbers of positives.
- ★ Groups of genes play similar roles in biological process.
- ★ Neighboring geographic locations share similar coefficients.
- ★ The same sector of finance share similar risk loadings

Related Literature: (*Park et al, 07; Friedman et al, 07; Bondell & Reich, 08; Zhu, et al, 13; Yang & He, 12.*)

# Challenges

■ No prior information on grouping. (sparsity w/o known atom)

## A naive approach:

- ◆ Obtain a preliminary estimate  $\tilde{\beta}$  and sort it
- ◆ Group coefficients that are close to each other.
- ◆ Force each estimated group to share a common coef & refit.

But how to group?

Wrong grouping can not be corrected!

# Challenges

■ No prior information on grouping. (sparsity w/o known atom)

## A naive approach:

- ◆ Obtain a preliminary estimate  $\tilde{\beta}$  and sort it
- ◆ Group coefficients that are close to each other.
- ◆ Force each estimated group to share a common coef & refit.

**But how to group?**

**Wrong grouping can not be corrected!**

# CARDS

# Basic version of CARDS

★ Preordering: Construct the rank statistics  $\{\tau(j)\}_{j=1}^p$  such that  $\tilde{\beta}_{\tau(1)} \leq \tilde{\beta}_{\tau(2)} \leq \cdots \leq \tilde{\beta}_{\tau(p)}$ .

★ Estimation: Fit penalized least-squares

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \sum_{j=1}^{p-1} p_{\lambda}(|\beta_{\tau(j+1)} - \beta_{\tau(j)}|) \right\}.$$

Consistency condition:  $\beta_{\tau(1)}^0 \leq \beta_{\tau(2)}^0 \leq \cdots \leq \beta_{\tau(p)}^0$ ,

**much weaker** than knowing the group.

- Fuse lasso assume  $\tau(i) = i$  is known. Results applied to fused lasso.
- Implemented by LLA (*Zou & Li, 08*) or CCCP (*Kim et al. 08*)
- Fuse penalty expedites the computation

# Ordered segmentation

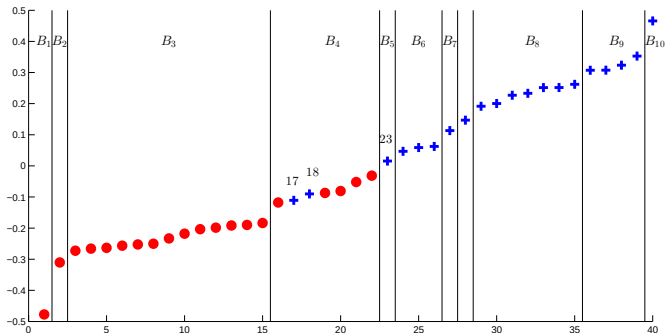
Ordered segmentation: The sets  $\{B_l\}_{l=1}^L$  form a partition of  $\{1, \dots, p\}$  and orderable.

- Similar to assign “letter” grades to a class
- Given preliminary rank, look for the gap at least  $\delta$

Consistency condition:  $\max_{j \in B_l} \beta_j^0 \leq \min_{j \in B_{l+1}} \beta_j^0, \quad \forall l \leq L$   
(weaker)

# A toy example

- $n = 100$ ,  $p = 40$  predictors from two groups.
- $\beta_j^0 = -0.2$  for **Group 1** and  $\beta_j^0 = 0.2$  for **Group 2**.
- $Y_i = \mathbf{X}_i^T \beta^0 + \varepsilon_i$  with  $\mathbf{X}_i \sim N_p(\mathbf{0}, \mathbf{I})$  and  $\varepsilon_i \sim N(0, 1)$ .



# Hybrid pairwise penalty

$$P_{\tau, \lambda_1, \lambda_2}(\beta) = \sum_{l=1}^{L-1} \sum_{i \in \mathbf{B}_l, j \in \mathbf{B}_{l+1}} \rho_{\lambda_1}(|\beta_i - \beta_j|) + \sum_{l=1}^L \sum_{i, j \in \mathbf{B}_l} \rho_{\lambda_2}(|\beta_i - \beta_j|).$$

## Special cases:

- $L = p$  or  $\delta = 0 \implies \sum_{j=1}^{p-1} \rho_{\lambda}(|\beta_{\tau(j+1)} - \beta_{\tau(j)}|)$  in bCARDS.
- $L = 1$  or  $\delta = \infty \implies P_{\lambda}^{TV}(\beta) = \sum_{1 \leq i, j \leq p} \rho_{\lambda}(|\beta_i - \beta_j|)$ . More computation (Shen & Huang, 2010)

# Hybrid pairwise penalty

$$P_{\tau, \lambda_1, \lambda_2}(\beta) = \sum_{l=1}^{L-1} \sum_{i \in \mathbf{B}_l, j \in \mathbf{B}_{l+1}} p_{\lambda_1}(|\beta_i - \beta_j|) + \sum_{l=1}^L \sum_{i, j \in \mathbf{B}_l} p_{\lambda_2}(|\beta_i - \beta_j|).$$

## Special cases:

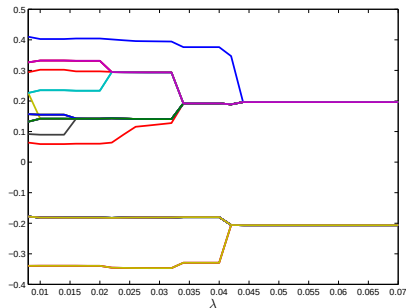
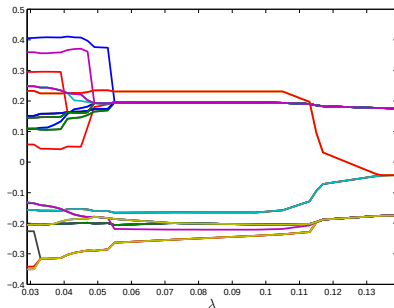
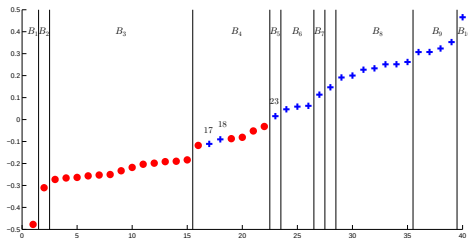
- $L = p$  or  $\delta = 0 \implies \sum_{j=1}^{p-1} p_{\lambda}(|\beta_{\tau(j+1)} - \beta_{\tau(j)}|)$  in bCARDS.
- $L = 1$  or  $\delta = \infty \implies P_{\lambda}^{TV}(\beta) = \sum_{1 \leq i, j \leq p} p_{\lambda}(|\beta_i - \beta_j|)$ . More computation (Shen & Huang, 2010)

# Advanced version of CARDS

- Preliminary Ranking: Obtain a preliminary est and sort it.
- Segmentation: Given gap  $\delta > 0$ , construct an ordered segmentation.
- Estimation: Minimize

$$Q_n(\beta) = \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|^2 + P_{\Upsilon, \lambda_1, \lambda_2}(\beta).$$

# How does it work? bCARD vs aCARD



# Theoretical results

## Basic CARDS

# Properties of CARDS: heuristics

Showcase: orthogonal design  $\mathbf{X}^T \mathbf{X} = n \mathbf{I}_p$

OLS estimator:  $\hat{\beta}^{ols} = n^{-1} \mathbf{X}^T \mathbf{y}$  follows

$$\hat{\beta}_j^{ols} = \beta_j^0 + \varepsilon_j, \quad \varepsilon_j \stackrel{i.i.d.}{\sim} N(0, n^{-1}), \quad j = 1, \dots, p.$$

$$\Rightarrow \|\hat{\beta}^{ols} - \beta^0\| = O_P(\sqrt{p/n})$$

# Properties of basic CARDS: heuristics

Oracle: **knows**  $(A_1, A_2, \dots, A_K)$ .

$$\hat{\beta}_{A,k}^{ols} = \beta_{A,k}^0 + \bar{\epsilon}_k, \quad \bar{\epsilon}_k \sim N(0, n^{-1}|A_k|^{-1}).$$

$$\begin{aligned} \|\hat{\beta}^{oracle} - \beta^0\|^2 &= \sum_{k=1}^K |A_k| |\hat{\beta}_{A,k}^{ols} - \beta_{A,k}^0|^2 \\ &= O_p\left(\sum_{k=1}^K |A_k| \cdot n^{-1}|A_k|^{-1}\right) \\ &= \mathbf{O_p(K/n)}. \end{aligned}$$

Sparsity:  $K = s + 1$ .

# Properties of basic CARDS

Oracle estimator:  $\hat{\beta}^{oracle} = \arg \min_{\beta \in \mathcal{M}_A} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|^2 \right\}$

where  $\mathcal{M}_A = \{\beta : \beta_i = \beta_j, \quad \forall i, j \in A_k\}$

## Oracle property of bCARDS

*If  $K = o(n)$ , group gaps sufficiently large, and ranks of  $\tilde{\beta}$  and  $\beta^0$  are consistent with prob  $\geq 1 - \varepsilon_0$ , then with probability  $\geq 1 - \varepsilon_0 - n^{-1}K - (n \vee p)^{-1}$ , bCARDS has a strictly local minimizer  $\hat{\beta}$  such that*

$$\star \quad \hat{\beta} = \hat{\beta}^{oracle},$$

$$\star \quad \|\hat{\beta} - \beta^0\| = O_p(\sqrt{K/n}).$$

# bCARDS and LLA algorithm

- Set an initial solution  $\hat{\beta}^{(0)} = \hat{\beta}^{initial}$ .
- Update the solution by

$$\hat{\beta}^{(m)} = \arg \min_{\beta} \left\{ \frac{1}{2n} \|\mathbf{y} - \mathbf{X}\beta\|^2 + \sum_{j=1}^{p-1} p'_\lambda (|\hat{\beta}_{\tau(j+1)}^{(m-1)} - \hat{\beta}_{\tau(j)}^{(m-1)}|) \cdot |\beta_{\tau(j+1)} - \beta_{\tau(j)}| \right\}.$$

## Oracle property of bCARD-LLA (Fan, Xue, Zou, 14)

If  $\|\hat{\beta}^{initial} - \beta^0\|_\infty \leq \lambda_n/2$ , then with probability at least  $1 - \varepsilon_0 - n^{-1}K - (n \vee p)^{-1}$ , the LLA algorithm yields  $\hat{\beta}^{oracle}$  after one iteration, and it converges to  $\hat{\beta}^{oracle}$  after two iterations.

# Consistent and robustness of rank mapping

## Consistent rank mapping by OLS

If  $p = O(n^\alpha)$  ( $0 < \alpha < 1$ ) and  $\lambda_{\min}(\frac{1}{n}\mathbf{X}^T\mathbf{X}) \geq c > 0$ , then with prob  $\geq 1 - O(n^{-\alpha})$ , the **ranks of  $\hat{\beta}^{ols}$  and  $\beta^0$  are the same.**

Severity of misranking:  $K^*(\tau) = \sum_{j=1}^{p-1} 1\{\beta_{\tau(j)}^0 \neq \beta_{\tau(j+1)}^0\}$ .

## Robustness to rank mapping

With prob  $\geq 1 - \varepsilon_0 - n^{-1}K - (n \vee p)^{-1}$ , bCARDS has a minimizer  $\hat{\beta}$  such that  $\|\hat{\beta} - \beta^0\| = O_p(\sqrt{K^*(\tau)/n})$ .

# Consistent and robustness of rank mapping

## Consistent rank mapping by OLS

If  $p = O(n^\alpha)$  ( $0 < \alpha < 1$ ) and  $\lambda_{\min}(\frac{1}{n}\mathbf{X}^T\mathbf{X}) \geq c > 0$ , then with  $\text{prob} \geq 1 - O(n^{-\alpha})$ , the **ranks of  $\hat{\beta}^{ols}$  and  $\beta^0$  are the same.**

Severity of misranking:  $K^*(\tau) = \sum_{j=1}^{p-1} 1\{\beta_{\tau(j)}^0 \neq \beta_{\tau(j+1)}^0\}$ .

## Robustness to rank mapping

With  $\text{prob} \geq 1 - \varepsilon_0 - n^{-1}K - (n \vee p)^{-1}$ , bCARDS has a minimizer  $\hat{\beta}$  such that  $\|\hat{\beta} - \beta^0\| = \mathbf{O}_p(\sqrt{K^*(\tau)/n})$ .

# bCARDS with $L_1$ penalty

Under irrerepresentable condition, the bCARDS with  $\rho(t) = |t|$  has a unique global minimizer  $\hat{\beta}$  such that

★  $\hat{\beta} \in \mathcal{M}_A$ ;

★  $\text{sgn}(\hat{\beta}_{A,k+1} - \hat{\beta}_{A,k}) = \text{sgn}(\beta_{A,k+1}^0 - \beta_{A,k}^0)$ ,  $k = 1, \dots, K-1$ ;

★  $\|\hat{\beta} - \beta^0\| = O_p(\sqrt{K/n} + \gamma_n)$ , where  $\gamma_n = \lambda_n(\sum_{k=1}^K \frac{1}{|A_k|})^{1/2}$ .

■ The bias is of order  $\sqrt{K(\log p)/n}$ .

# Theoretical results

## Advanced CARDS

# Properties of advanced CARDS

Assume  $P(\max_{j \in B_l} \beta_j^0 \leq \min_{j \in B_{l+1}} \beta_j^0, \quad \forall l \leq L) > 1 - \varepsilon_0$

## Properties of aCARDS

With prob  $\geq 1 - \varepsilon_0 - n^{-1}K - 2(n \vee p)^{-1}$ , aCARDS has a minimizer  $\hat{\beta}$  such that

- $\hat{\beta} = \hat{\beta}^{oracle}$ ,
- $\|\hat{\beta} - \beta^0\| = O_p(\sqrt{K/n})$ .

Asymptotic normality:  $\mathbf{b}_n^T (\mathbf{X}_A^T \mathbf{X}_A)^{1/2} (\hat{\beta}_A - \beta_A^0) \xrightarrow{d} N(0, 1)$

(**smaller than OLS**), where  $\mathbf{x}_{A,k} = \sum_{j \in A_k} \mathbf{x}_j$ .

# Sparse CARDS

Explore **homogeneity** and **sparsity** simultaneously.

sCARDS: Preliminary estimate  $\tilde{S} \supset S_0$

$$Q_n^{\text{sparse}}(\beta) = \frac{1}{2n} \|\mathbf{y} - \mathbf{x}_{\tilde{S}} \beta_{\tilde{S}}\|^2 + P_{\gamma, \lambda_1, \lambda_2}(\beta_{\tilde{S}}) + \sum_{j \in \tilde{S}} p_{\lambda}(|\beta_j|),$$

Local oracle properties can be extended to sCARDS.

# Simulation studies

# Normalized mutual information (NMI)

NMI of two partitions  $\mathbb{C} = \{C_k\}$  and  $\mathbb{D} = \{D_j\}$  of  $\{1, \dots, p\}$ :

$$\text{NMI}(\mathbb{C}, \mathbb{D}) = \frac{I(\mathbb{C}; \mathbb{D})}{[H(\mathbb{C}) + H(\mathbb{D})]/2},$$

where  $I(\mathbb{C}; \mathbb{D}) = \sum_{k,j} (|C_k \cap D_j|/p) \log(p|C_k \cap D_j|/|C_k||D_j|)$ , and  $H(\mathbb{C}) = \sum_k (|C_k|/p) \log(|C_k|/p)$  is the entropy of  $\mathbb{C}$ .

- ★  $\text{NMI}(\mathbb{C}, \mathbb{D})$  takes values on  $[0, 1]$ .
- ★ A larger NMI implies the two partitions are closer.
- ★  $\text{NMI} = 1$  means that the two groupings are the same.

# Normalized mutual information (NMI)

NMI of two partitions  $\mathbb{C} = \{C_k\}$  and  $\mathbb{D} = \{D_j\}$  of  $\{1, \dots, p\}$ :

$$\text{NMI}(\mathbb{C}, \mathbb{D}) = \frac{I(\mathbb{C}; \mathbb{D})}{[H(\mathbb{C}) + H(\mathbb{D})]/2},$$

where  $I(\mathbb{C}; \mathbb{D}) = \sum_{k,j} (|C_k \cap D_j|/p) \log(p|C_k \cap D_j|/|C_k||D_j|)$ , and  $H(\mathbb{C}) = \sum_k (|C_k|/p) \log(|C_k|/p)$  is the entropy of  $\mathbb{C}$ .

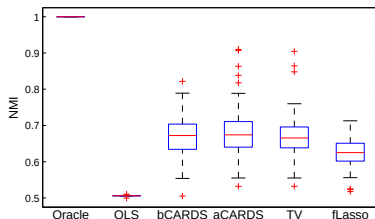
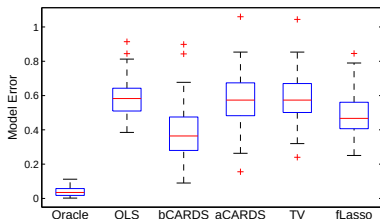
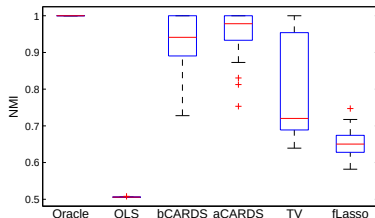
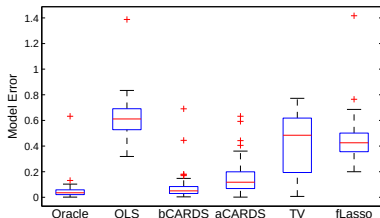
- ★  $\text{NMI}(\mathbb{C}, \mathbb{D})$  takes values on  $[0, 1]$ .
- ★ A larger NMI implies the two partitions are closer.
- ★  $\text{NMI} = 1$  means that the two groupings are the same.

# Simulation 1: equal-size groups

$$Y = \mathbf{X}^T \beta^0 + \varepsilon, \quad \mathbf{X} \sim N(\mathbf{0}, \mathbf{I}), \quad \varepsilon \sim N(0, 1).$$

- $p = 60$  predictors consist of four groups with coefficients  $\beta_0$  takes  $-2r, -r, r$ , and  $2r$
- $n = 100$ , tuning via BIC

# Simulation 1: Model error and NMI



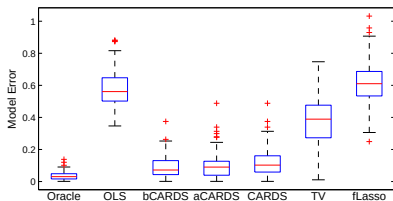
Top Panel:  $r = 1$  and bottom panel:  $r = 0.5$

## Simulation 2: unequal-size groups

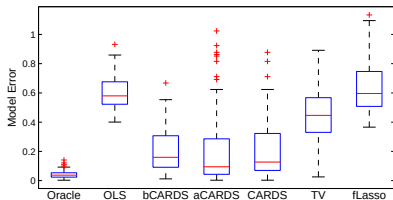
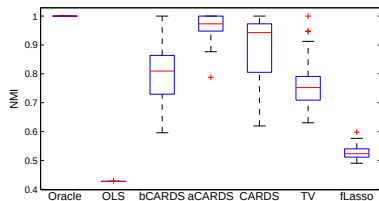
The same setting as **Simul 1** except group sizes are **unequal**:

- (2A) Four groups of size 1, 15, 15, 29 with coefficients  $-4r, -r, r$ , and  $2r$ .
- (2B) One dominating group of size 50 with coefficient  $-2r$  and the other 10 predictors have coefficients  $0, 2/9, 4/9, \dots, 2$ .

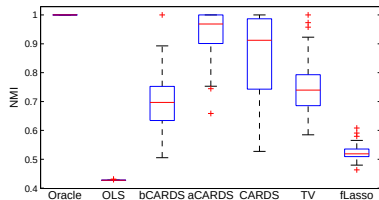
# Simulation 2A: Model errors and NMI



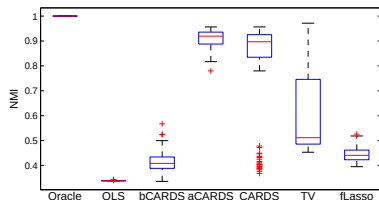
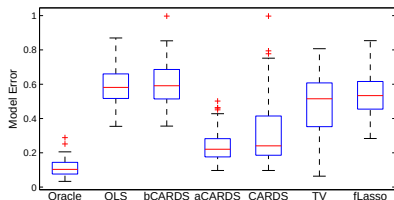
(a)  $r=1$



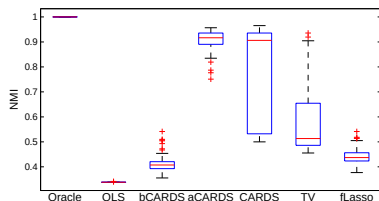
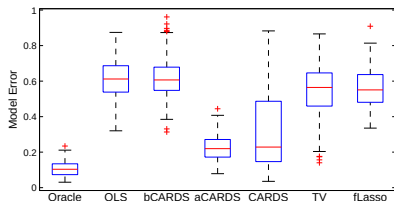
(b)  $r=0.7$



# Simulation 2B: Models and NMI



(c)  $r=1$



(d)  $r=0.7$

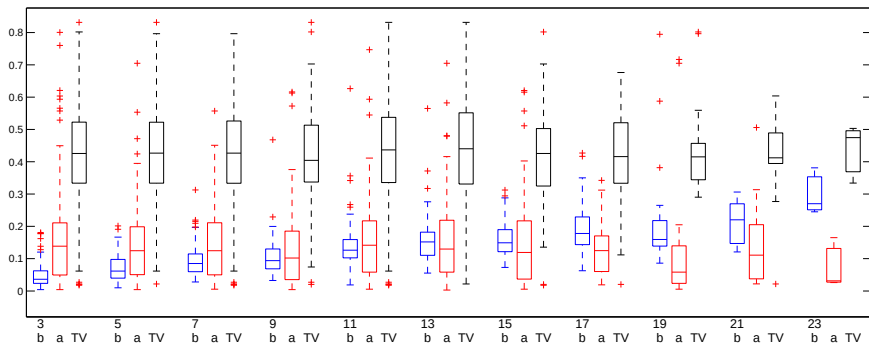
## Simulation 3: misranking

Model: The same as Experiment 1 with  $r = 1$

Preliminary rank: Based on OLS estimator from  $\mathbf{z} \sim N(\mathbf{X}\beta^0, \sigma^2 \mathbf{I}_n)$ , with 11 different  $\sigma$  in  $\{1, 1.2, 1.4, \dots, 3\}$ .

- A larger value of  $\sigma$  tends to yield a “worse” preliminary rank.
- Generate  $100 * 11$  data sets and classify results according to  $K^*$ , severity of misranking.

# Simulation 3: Result by degree of misranking



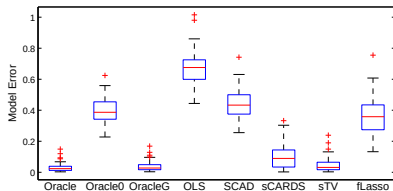
Average model error changes with  $K^*$ .

■ aCARDS is robust to misranking; outperform TV.

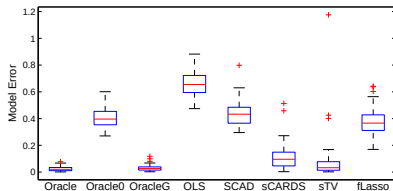
■ bCARDS perform the best when  $K^*$  is small

## Simulation 4: sparsity and homogeneity

**Model:** Adding 40 unimportant variables to Experiment 1,  
 $p = 100, n = 150$ .



(a)  $r=1$



(b)  $r=0.7$

## Simulation 5: a spatial-temporal model

$$Y_{it} = \mathbf{x}_t^T \beta_i + \varepsilon_{it}, \quad 1 \leq i \leq q$$

■  $q = 100$  locations and  $k = 5$  common predictors

Spatial Homogeneity: 4 spatial groups:

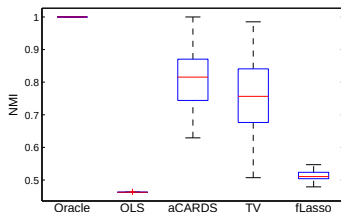
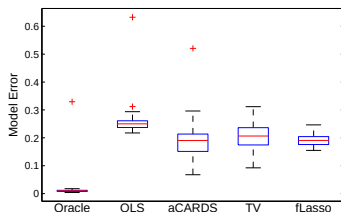
$$\beta_1 = \cdots = \beta_{25}, \quad \cdots, \quad \beta_{76} = \cdots = \beta_{100}$$

$$\beta_{i,j} = b_j - 2I_{1 \leq i \leq 25} - I_{26 \leq i \leq 50} + I_{51 \leq i \leq 75} + 2I_{76 \leq i \leq 100}$$

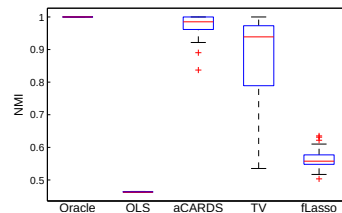
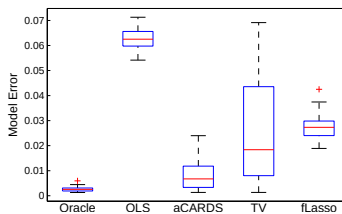
with  $b_j = 0.1 \times (j - 1), 1 \leq j \leq 5$ .

# Simulation 5: Model error and MMI

$T = 20$



$T = 80$



# Applications

# Financial data and market beta

Fama-French model:  $Y_{it} = \alpha_i + \mathbf{X}_t^T \beta_i^0 + \varepsilon_{it}$ ,

- $\mathbf{X}_t$  are three Fama-French risk factors
- $Y_{it}$  excess return of asset  $i$ .
- $\{\alpha_i\}$  are sparse and penalized

Data: Daily returns of 410 stocks, the surviving components of the S&P500 index in 12/1/2010 – 12/1/2011 ( $T = 254$ ).

Market  $\beta$ : shorten from 5 years to 1 year by **homogeneity**.

# Financial data and market beta

Fama-French model:  $Y_{it} = \alpha_i + \mathbf{X}_t^T \beta_i^0 + \varepsilon_{it}$ ,

- $\mathbf{X}_t$  are three Fama-French risk factors
- $Y_{it}$  excess return of asset  $i$ .
- $\{\alpha_i\}$  are sparse and penalized

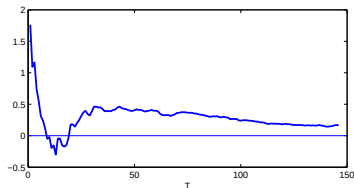
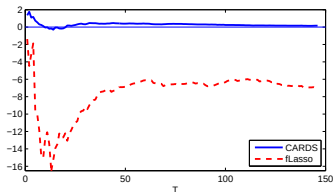
Data: Daily returns of 410 stocks, the surviving components of the S&P500 index in 12/1/2010 – 12/1/2011 ( $T = 254$ ).

Market  $\beta$ : shorten from 5 years to 1 year by **homogeneity**.

# Results S&P 500 returns

**Testing period:** 12/1/11 – 7/1/12 ( $T = 146$ )

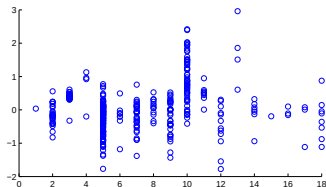
$$\text{cRSS}_t = \sum_{s=1}^t \rho^{\lfloor s/10 \rfloor} \sum_i (\hat{y}_{it} - y_{it})^2 \text{ with } \rho = 0.95.$$



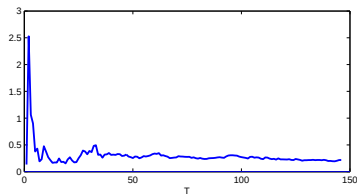
Present of PE from 12/1/11 – 7/1/12:  $100(\text{cRSS}_t^{\text{ols}} - \text{cRSS}_t^{\text{cards}}) / \text{cRSS}_t^{\text{ols}}$ .

The right panel is a zoom-in of the results for CARDS.

# Results of S&P 500 returns (I)



(c)



(d)

(a) OLS coefficients on the “book-to-market ratio” factor. The x axis is sectors.

(b) Percentage improvement in the 29 utility stocks.

## Results of S&P 500 returns (II)

| Fama-French factors     | No. of coef. groups |
|-------------------------|---------------------|
| “market return”         | 41                  |
| “market capitalization” | 32                  |
| “book-to-market ratio”  | 56                  |
| intercept               | 60                  |

Number of groups in fitting the S&P500 data.

# Summary

- ★ Important to explore homogeneity to reduce variance.
- ★ Propose bCARDS (fuse), aCARDS (hybrid), sCARDS (screening) to promote homogeneity; **no prior** group info.
- ★ Study various theoretical results, MSE reduces to  $O(K/n)$ .
- ★ Establish oracle properties, CARDS-LLA, and examine impact of misranking.

# Summary

- ★ Important to explore homogeneity to reduce variance.
- ★ Propose bCARDS (fuse), aCARDS (hybrid), sCARDS (screening) to promote homogeneity; **no prior** group info.
- ★ Study various theoretical results, MSE reduces to  $O(K/n)$ .
- ★ Establish oracle properties, CARDS-LLA, and examine impact of misranking.

*Happy 80th Birthday*



*Grace*