Section 1

ZTS (2016)

Lasso

- Tibshirani (1996)

$$\min_{\beta} n^{-1} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1$$

```
begin_cvx
        variable b(p);
        objective = sum_square( y - X * b )/n ...
                + lambda * norm(b, 1);
        minimize(objective)
end_cvx
```

GMM-Lasso

- Caner (2009)

$$\left\| n^{-1} Z' (y - X\beta) \right\|_2^2 + \lambda \|\beta\|_1 .$$

- Shi (2016) establishes an *oracle inequality* that implies the optimal rate of convergence under sparsity.
- Question: How to control too many IVs, or too many moments in nonlinear GMM?

GMM-Lasso

- Caner (2009)

$$\left\| n^{-1} Z' (y - X\beta) \right\|_2^2 + \lambda \left\| \beta \right\|_1 .$$

- Shi (2016) establishes an *oracle inequality* that implies the optimal rate of convergence under sparsity.
- Question: How to control too many IVs, or too many moments in nonlinear GMM?

Section 2

Su, ZTS, Phillips (2016)

Application: Saving Rate

- Edwards (1996)

$$S_{it} = \beta_1 S_{it-1} + \beta_2 I_{it} + \beta_3 R_{it} + \beta_4 G_{it} + \mu_i + \varepsilon_{it}.$$

- We: group pattern in

$$S_{it} = \beta_{1i} S_{it-1} + \beta_{2i} I_{it} + \beta_{3i} R_{it} + \beta_{4i} G_{it} + \mu_i + \varepsilon_{it}.$$

  - Group 1: (31 countries) Armenia, Australia, Bahamas, Belarus, Bolivia, Botswana, Cape Verde, China, Czech, Guatemala, Honduras, Hungary, Indonesia, Israel, Italy, Japan, Jordan, Latvia, Malawi, Malaysia, Mauritius, Mexico, Mongolia, Panama, Paraguay, Philippines, Romania, South Africa, Sri Lanka, Thailand, Ukraine
  - Group 2: (25 countries) Bangladesh, Canada, Costa Rica, Dominican, Egypt, Guyana, Iceland, India, Kenya, South Korea, Lithuania, Malta, Netherlands, Papua New Guinea, Peru, Russian, Singapore, Swaziland, Switzerland, Syrian, Tanzania, Uganda, United Kingdom, United States, Uruguay

Application: Saving Rate

- Edwards (1996)

$$S_{it} = \beta_1 S_{it-1} + \beta_2 I_{it} + \beta_3 R_{it} + \beta_4 G_{it} + \mu_i + \varepsilon_{it}.$$

- We: group pattern in

$$S_{it} = \beta_{1i} S_{it-1} + \beta_{2i} I_{it} + \beta_{3i} R_{it} + \beta_{4i} G_{it} + \mu_i + \varepsilon_{it}.$$

  - Group 1: (31 countries) Armenia, Australia, Bahamas, Belarus, Bolivia, Botswana, Cape Verde, China, Czech, Guatemala, Honduras, Hungary, Indonesia, Israel, Italy, Japan, Jordan, Latvia, Malawi, Malaysia, Mauritius, Mexico, Mongolia, Panama, Paraguay, Philippines, Romania, South Africa, Sri Lanka, Thailand, Ukraine
  - Group 2: (25 countries) Bangladesh, Canada, Costa Rica, Dominican, Egypt, Guyana, Iceland, India, Kenya, South Korea, Lithuania, Malta, Netherlands, Papua New Guinea, Peru, Russian, Singapore, Swaziland, Switzerland, Syrian, Tanzania, Uganda, United Kingdom, United States, Uruguay

Application: Saving Rate

- Edwards (1996)

$$S_{it} = \beta_1 S_{it-1} + \beta_2 I_{it} + \beta_3 R_{it} + \beta_4 G_{it} + \mu_i + \varepsilon_{it}.$$

- We: group pattern in

$$S_{it} = \beta_{1i} S_{it-1} + \beta_{2i} I_{it} + \beta_{3i} R_{it} + \beta_{4i} G_{it} + \mu_i + \varepsilon_{it}.$$

  - Group 1: (31 countries) Armenia, Australia, Bahamas, Belarus, Bolivia, Botswana, Cape Verde, China, Czech, Guatemala, Honduras, Hungary, Indonesia, Israel, Italy, Japan, Jordan, Latvia, Malawi, Malaysia, Mauritius, Mexico, Mongolia, Panama, Paraguay, Philippines, Romania, South Africa, Sri Lanka, Thailand, Ukraine
  - Group 2: (25 countries) Bangladesh, Canada, Costa Rica, Dominican, Egypt, Guyana, Iceland, India, Kenya, South Korea, Lithuania, Malta, Netherlands, Papua New Guinea, Peru, Russian, Singapore, Swaziland, Switzerland, Syrian, Tanzania, Uganda, United Kingdom, United States, Uruguay

Classifier-Lasso

- Bonhomme and Manresa (2015)
- Slope coefficient heterogeneity
  - Penalized least squares (PLS) for illustration

$$\min_{(\beta_i)_{i=1}^N, (\alpha_k)_{k=1}^K} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\tilde{y}_{it} - \tilde{x}_{it}' \beta_i\right)^2 + \frac{\lambda}{N} \sum_{i=1}^{N} \prod_{k=1}^{K} \|\beta_i - \alpha_k\|_2 .$$

- Insight
  - Consistent initial estimator of β_i when T is large
  - Multiplying L-1 penalty pushes the individual parameters to one and only one center
  - k-means classification and within-group estimation are two extreme special cases

Classifier-Lasso

- Bonhomme and Manresa (2015)
- Slope coefficient heterogeneity
  - Penalized least squares (PLS) for illustration

$$\min_{(\beta_i)_{i=1}^N,(\alpha_k)_{k=1}^K} \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left(\tilde{y}_{it}-\tilde{x}_{it}'\beta_i\right)^2+\frac{\lambda}{N}\sum_{i=1}^N\prod_{k=1}^K\|\beta_i-\alpha_k\|_2 .$$

- Insight
  - Consistent initial estimator of β_i when T is large
  - Multiplying L-1 penalty pushes the individual parameters to one and only one center
  - k-means classification and within-group estimation are two extreme special cases

Classifier-Lasso (continue)

- Iterative algorithm.
- In the k-th sub-step of the r-th iteration, we choose $(\boldsymbol{\beta}, \alpha_{\tilde{k}})$ to minimize

$$\min_{\boldsymbol{\beta}, \alpha_{\tilde{k}}} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (\tilde{y}_{it} - \tilde{x}_{it}' \beta_i)^2 + \frac{\lambda}{N} \sum_{i=1}^{N} \|\beta_i - \alpha_{\tilde{k}}\|_2 \gamma_i$$

where $\gamma_i = \prod_{k=1}^{\tilde{k}-1} \left\| \hat{\beta}_i^{(r,k)} - \hat{\alpha}_k^{(r)} \right\|_2 \prod_{k=\tilde{k}+1}^{K} \left\| \hat{\beta}_i^{(r-1,k)} - \hat{\alpha}_k^{(r-1)} \right\|_2$.

Findings

- Big N, big T
- Likelihood function framework
- Oracle property (as if the group identity is known)
- Consistent classification
- Post-selection inference
- Also applies to GMM and MLE

Section 3

Lee, ZTS and Gao (2018)

LASSO Family

$$\hat{\theta} = \arg\min_{\beta \in \mathbb{R}^p} \|y - W\theta\|_2^2 + \lambda_n \sum_{j=1}^{p} \hat{\tau}_j |\theta_j|,$$

- plain LASSO (plasso): $\hat{\tau}_j = 1$
- standardized LASSO (slasso): $\hat{\tau}_j =$ sample sd of j-th regr
- adaptive LASSO (alasso): $\hat{\tau}_j = 1/|\hat{\theta}_j^{init}|$.

Predictive Regression Model

- Cointegration and Mixed Roots
  - Time series $i = 1, \ldots, n$. Linear model

$$\begin{aligned} y_i &= \theta_n^{*\prime} W + u_i \\ &= \alpha^{*\prime} Z_i + \phi_n^{*\prime} X_i^c + \beta_n^{*\prime} X_i + u_i \\ &= \sum_{j=1}^{p_z} z_{ij} \alpha_j^* + \sum_{j=1}^{p_1} v_{1ij} \phi_{1j}^* + \sum_{j=1}^{p_x} x_{ij} \beta_{jn}^* + u_i \end{aligned}$$

  - Z_i: (p_z) stationary
  - X_i^c: $(p_c = p_1 + p_2)$ cointegrated vector with rank p_1

$$\underset{p_1 \times p_c}{A} X_i^c = X_{1i}^c - \underset{p_1 \times p_2}{A_1} X_{2i}^c = v_{1i} \; ,$$

$$\triangle X_{2i}^c = v_{2i} \; ,$$

  - X_i: (p_x) unit root.

OLS Theory

- "extend" the I(0) regressors as $Z^+ = [Z, v_1]$ and the I(1) regressors as $X^+ := [X_2^c, X]$.

$$R_n = \begin{pmatrix} \sqrt{n} \cdot I_{p_z+p_1} & 0 \\ 0 & n \cdot I_{p_2+p_x} \end{pmatrix},$$

- OLS estimator

$$R_n Q \left(\widehat{\theta}_n^{ols} - \theta_n^* \right) \Longrightarrow \left(\Omega^+ \right)^{-1} \xi^+,$$

where the right-hand side is a non-degenerate continuous random variable.

Intuition

	coef.	OLS	plasso	slasso	ada ($\neq 0$)	ada ($= 0$)
			order of $\widehat{\tau}_j/n$			
z_j	α^*_{0j}	$n^{-1/2}$	n^{-1}	n^{-1}	n^{-1}	$n^{-1/2}$
x^c	ϕ^*_{1j}	$n^{-1/2}$	n^{-1}	$n^{-1/2}$	n^{-1}	$n^{-1/2}$
x_j	$\beta^*_{0j}/\sqrt{n}$	n^{-1}	n^{-1}	$n^{-1/2}$	$n^{-1/2}$	1

- Findings:
  - Alasso retains oracle property in pure I(0) and I(1)
  - Plasso and Slasso fail to maintain consistency and variable screening in all components

Intuition

	coef.	OLS	plasso	slasso	ada ($\neq 0$)	ada ($= 0$)
			order of $\widehat{\tau}_j/n$			
z_j	α^*_{0j}	$n^{-1/2}$	n^{-1}	n^{-1}	n^{-1}	$n^{-1/2}$
x^c	ϕ^*_{1j}	$n^{-1/2}$	n^{-1}	$n^{-1/2}$	n^{-1}	$n^{-1/2}$
x_j	$\beta^*_{0j}/\sqrt{n}$	n^{-1}	n^{-1}	$n^{-1/2}$	$n^{-1/2}$	1

- Findings:
  - Alasso retains oracle property in pure I(0) and I(1)
  - Plasso and Slasso fail to maintain consistency and variable screening in all components

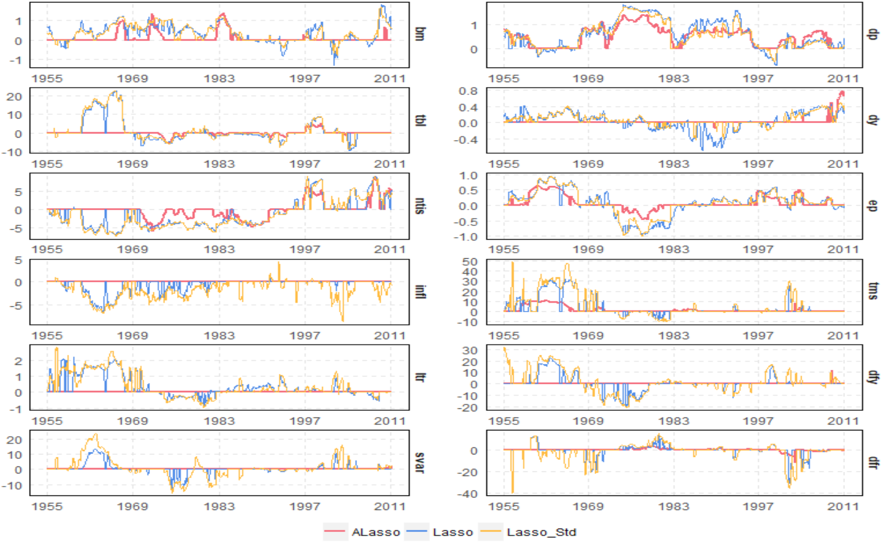
Application

- Welch and Goyal (2008)
- Koo, Anderson, Seo and Yao (2016)
- Monthly data: January 1945—December 2012
- Dependent variable: S&P 500 excess return
- 12 predictors
- 9 variables' AR(1) coefficients bigger than 0.95

MPSE

h	OLS	RWwD	Alasso	Plasso	Slasso
			10-year rolling window		
$\frac{1}{12}$	0.00209	0.00188	0.00187	**0.00186**	0.00187
$\frac{1}{4}$	0.00936	0.00663	**0.00615**	0.00834	0.00758
$\frac{1}{2}$	0.01835	0.01644	**0.01316**	0.01608	0.01534
1	0.03404	0.04292	**0.02882**	0.03084	0.02951
2	0.07708	0.12968	**0.05398**	0.07248	0.06261
3	0.20066	0.27608	**0.12125**	0.15875	0.17730
			15-year rolling window		
$\frac{1}{12}$	0.00203	0.00196	**0.00182**	0.00186	0.00187
$\frac{1}{4}$	0.00826	0.00692	**0.00605**	0.00656	0.00654
$\frac{1}{2}$	0.02009	0.01714	**0.01548**	0.01846	0.01697
1	0.03996	0.04449	0.03013	**0.02940**	0.03686
2	0.05947	0.13694	**0.03887**	0.05240	0.05392
3	0.11166	0.29198	**0.08014**	0.10578	0.11163

Figure: Estimated Coefficients (10-year rolling window, $h = 1$)



Section 4

ZTS and Huang (2019)

Panel data approach (PDA)

- Hsiao, Ching and Wan (2012): Hong Kong-Mainland Economic Integration
- Event: The effect of the Closer Economic Partnership Arrangement (CEPA): Jan 1, 2004
- Pre-treatment: 1993:Q1—2003:Q4 (44 observations)
- Post-treatment: 2004:Q1—2008:Q1 (17 observations)
- Estimate

$$y_{0t}^{0} = Y_t'\beta + \varepsilon_t$$

  where the dependent variable is Hong Kong's GDP growth rate, and control units are 24 countries' GDP growth rate.
- Many potential control units: $2^{24} \approx 17,000,000$, $2^{87} = 1.6 \times 10^{26}$

Panel data approach (PDA)

- Hsiao, Ching and Wan (2012): Hong Kong-Mainland Economic Integration
- Event: The effect of the Closer Economic Partnership Arrangement (CEPA): Jan 1, 2004
- Pre-treatment: 1993:Q1—2003:Q4 (44 observations)
- Post-treatment: 2004:Q1—2008:Q1 (17 observations)
- Estimate

$$y_{0t}^{0} = Y_t'\beta + \varepsilon_t$$

  where the dependent variable is Hong Kong's GDP growth rate, and control units are 24 countries' GDP growth rate.
- Many potential control units: $2^{24} \approx 17,000,000$, $2^{87} = 1.6 \times 10^{26}$

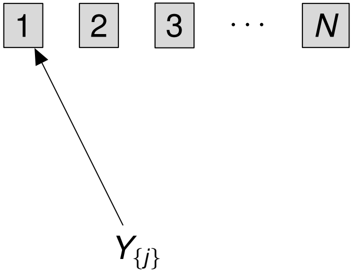
Forward Selection

$$\boxed{1} \quad \boxed{2} \quad \boxed{3} \quad \cdots \quad \boxed{N}$$

- A greedy algorithm. Computational efficiency.

Forward Selection

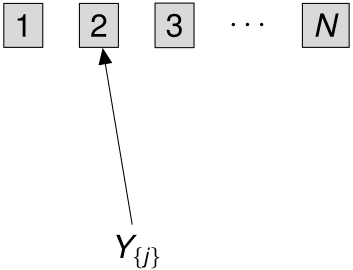
$r = 1$



- A greedy algorithm. Computational efficiency.

Forward Selection

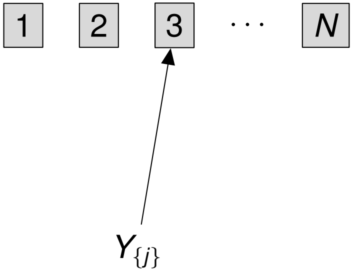
$r = 1$



- A greedy algorithm. Computational efficiency.

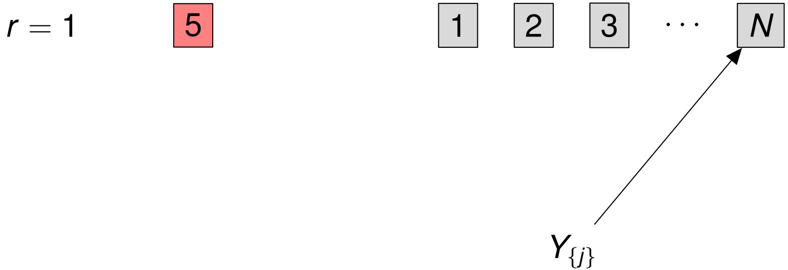
Forward Selection

$r = 1$



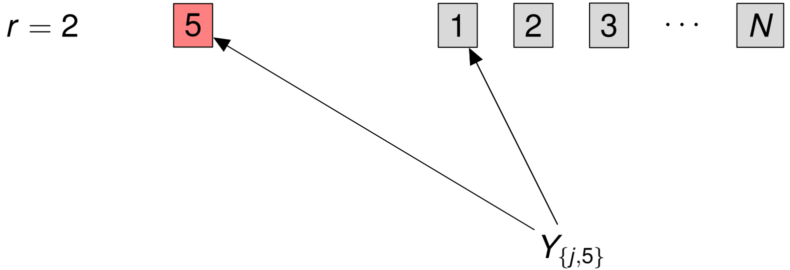
- A greedy algorithm. Computational efficiency.

Forward Selection



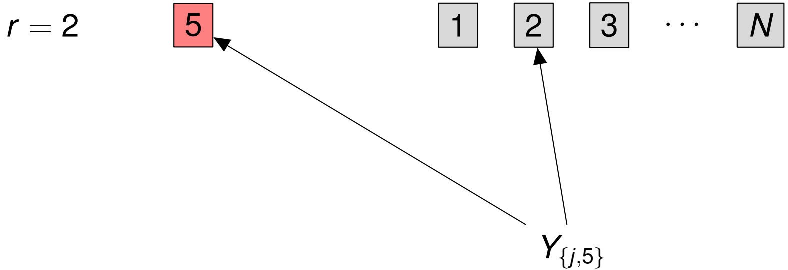
- A greedy algorithm. Computational efficiency.

Forward Selection



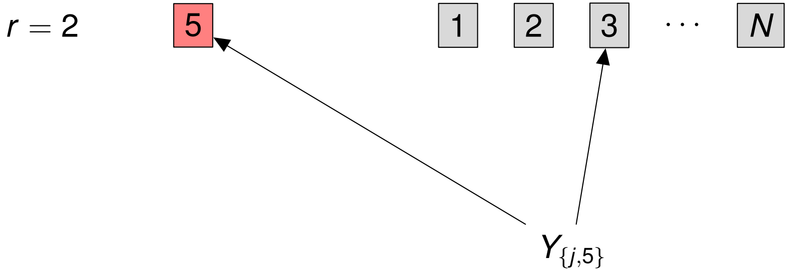
- A greedy algorithm. Computational efficiency.

Forward Selection



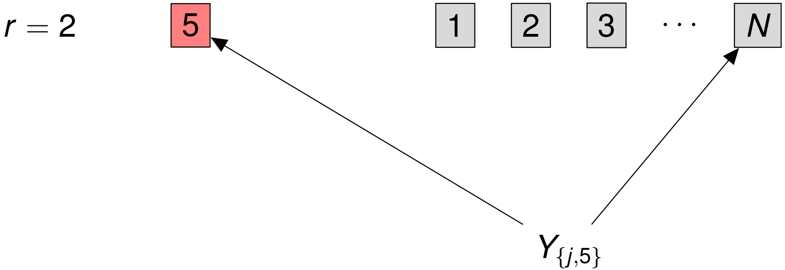
- A greedy algorithm. Computational efficiency.

Forward Selection



- A greedy algorithm. Computational efficiency.

Forward Selection



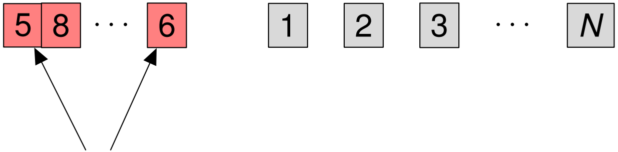
- A greedy algorithm. Computational efficiency.

Forward Selection

- A greedy algorithm. Computational efficiency.

Forward Selection

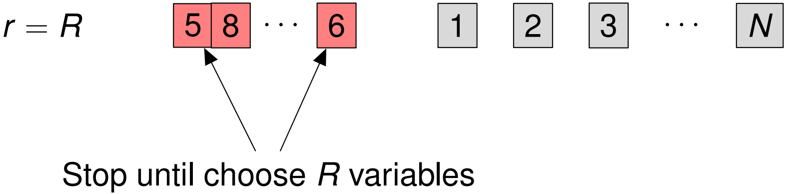
$r = R$



Stop until choose R variables

- A greedy algorithm. Computational efficiency.

Forward Selection



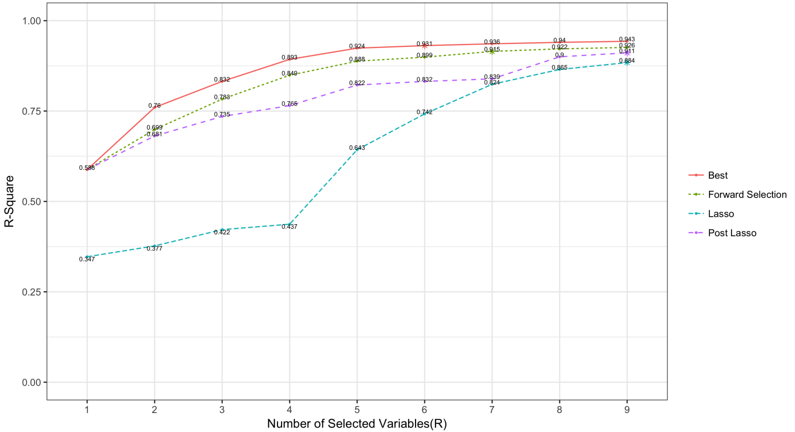
- A greedy algorithm. Computational efficiency.

Contributions

- Forward selection for control units to accommodate dense coefficient
- Standard procedure for inference
- Insight
  - Factor model helps justify statistical assumptions
  - Weak dependence facilitates post-selection inference

Effectiveness of Selection

- In-sample R-squared:



- Number of estimations: $\binom{24}{8} = 735471$ vs. $\sum_{j=0}^{7}(24-j) = 164$