LASSO-Driven Inference in Time and Space

Victor Chernozhukov
Wolfgang K. Härdle
Chen Huang
Weining Wang

MIT Center for Statistics
Humboldt–Universität zu Berlin
University of St.Gallen
City, University of London
http://stat.mit.edu
http://lvb.wiwi.hu-berlin.de
http://mathstat.unisg.ch
http://www.city.ac.uk

TENET

Tail-Event-driven NETwork Risk: Härdle et al. (2016)

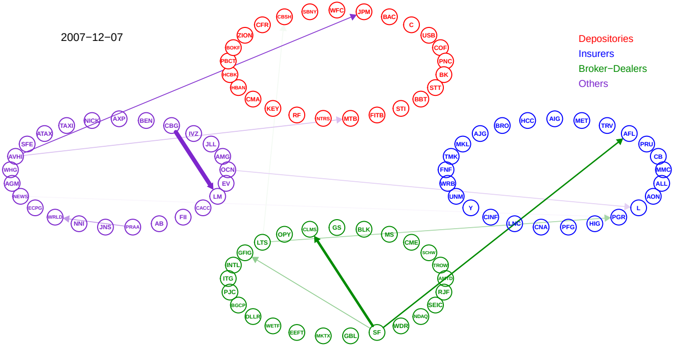


Figure 1: Financial risk network dynamics.

Financial Risk Meter

- Averaged penalty levels in dynamic network analysis
- Systemic risk level in the financial market over time
- Simultaneous inference between sectors

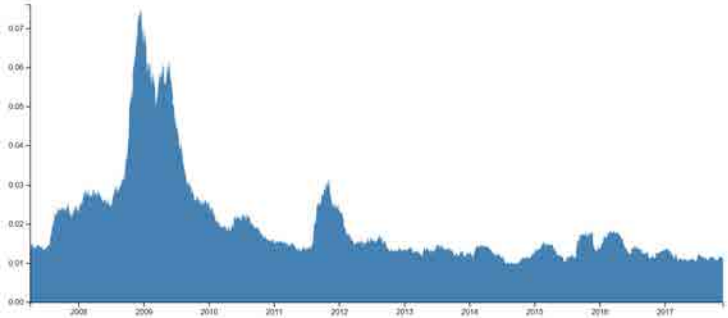


Figure 2: FRM over time `frm.wiwi.hu-berlin.de`

LOB Network

Time-varying Limit Order Book Networks: Härdle et al. (2018)

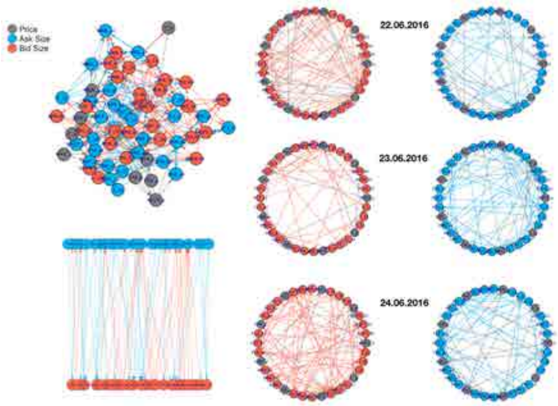


Figure 3: Plots of LOB networks from 22.06.2016-24.06.2016

The System of Regression Equations

$$Y_{j,t} = X_{j,t}^{\top}\beta_j^0 + \varepsilon_{j,t}, \quad \mathsf{E}\, \varepsilon_{j,t} X_{j,t} = 0, \quad j = 1, ..., J, \quad t = 1, \ldots, n,$$

- The dimension $K_j = \dim(X_{j,t}) \equiv K$ and the number of equations J are large (potentially larger than n)
- Allowing for temporal and spatial dependency

The System of Regression Equations

Example 1: Large VAR

$$Y_t = \sum_{l=1}^{p} \Phi_l^0 Y_{t-l} + \varepsilon_t, \quad \mathsf{E}\, \varepsilon_{j,t} Y_{t-l} = 0.$$

Example 2: Simultaneous equations systems

$$Y_{j,t} = Y_{-j,t}\delta_j^0 + X_t^\top \gamma_j^0 + \varepsilon_{j,t},$$

with the reduced form given by

$$Y_{j,t} = X_t^\top \beta_j^0 + \nu_{j,t}, \quad \mathsf{E}\, \nu_{j,t} X_t = 0.$$

Effective Prediction with Sparsity Method

- Exact sparsity assumption $|\beta_j^0|_0 = s_j \leq s = o(n),\ j = 1, \ldots J$
- ℓ_1-penalized estimator of β_j^0

$$\widehat{\beta}_j = \arg \min_{\beta \in \mathbb{R}^{K_j}} \frac{1}{n} \sum_{t=1}^{n} (Y_{j,t} - X_{j,t}^\top \beta)^2 + \frac{\lambda}{n} \sum_{k=1}^{K_j} |\beta_{jk}| \Psi_{jk}$$

- Select a joint penalty λ, which accounts for the dependency and aggregate the effects over equations
- Prediction performance bound (oracle inequalities)
- Individual and simultaneous inference on the coefficients

Practical Examples

Example 3: Cross-sectional Asset Pricing

$$Y_{j,t} = \beta_{j0} + \sum_{k=1}^{K} \beta_{jk} X_{jk,t} + \varepsilon_{j,t},$$

$Y_{j,t}$: excess returns for asset j, $X_{jk,t}$ are the factor returns and one is interested in testing: $H_0 : \beta_{j0} = 0, \forall j = 1, \ldots, J$.
Examples 4: Network formation and spillover effects:

$$Y_{j,t} = \beta_j D_{j,t} + \sum_{i \neq j} \omega_{ij} D_{i,t} + \gamma_j^\top X_{j,t} + \varepsilon_{j,t},$$

$Y_{j,t}$: log output for firm j, $D_{j,t}$: capital stock, and one is interested in testing the spillover parameters ω_{ij}.

Fundamental Results

- Oracle error bounds of ℓ_1-penalized estimator: Bickel et al. (2009), Belloni and Chernozhukov (2013)
- Ideal penalty level - max of sum of high-dim random vectors
  - Gaussian approximation and (block) multiplier bootstrap
  - Chernozhukov et al. (2013), Zhang and Wu (2017)
- Uniformly valid inference on target coefficients:
  - Post-selection inference (Neyman orthogonality or double selection): Belloni et al. (2014, 2015)
  - De-sparsified (de-biased) LASSO: Zhang and Zhang (2014), Van de Geer et al. (2014)

Our Contributions

- More general time dependence measure (Wu, 2005) ▸ Definition
- General Bahadur representation for the Z-estimators with dependent data
- Simultaneous confidence region via multiplier bootstrap
- Application: textual sentiment spillover effects

Outline

1. Motivation ✓
2. Estimation and Theoretical Results
3. Simulation Study
4. Application

"Ideal" Choice of λ

- Suppose we observe $\varepsilon_{j,t} = Y_{j,t} - X_{j,t}^\top \beta_j^0$, set

$$S_{jk} \stackrel{\text{def}}{=} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_{j,t} X_{jk,t}, \ \Psi_{jk} \stackrel{\text{def}}{=} \sqrt{\operatorname{avar}(S_{jk})}$$

$$\lambda^0(1-\alpha) \stackrel{\text{def}}{=} (1-\alpha) - \text{quantile of } 2c\sqrt{n} \max_{j \le J, k \le K} |S_{jk}/\Psi_{jk}|,$$

  where $c > 1$, e.g. $c = 1.1$, $\alpha = 0.1$.

  ▸ Foundations

- Use Gaussian approx. or multiplier block bootstrap

Error Bounds for the Prediction Norm

Lemma 1

Suppose the uniform RE condition [A2] *holds with probability* $1 - o(1)$*, then under the exact sparsity assumption, we have*

$$|\widehat{\beta}_j - \beta_j^0|_{j,\mathrm{pr}} \stackrel{\mathrm{def}}{=} \left[n^{-1} \sum_{t=1}^{n} \left\{ X_{j,t}^{\top} (\widehat{\beta}_j - \beta_j^0) \right\}^2 \right]^{1/2} \leq C \lambda^0 (1-\alpha) \frac{\sqrt{s}}{n} \max_k \Psi_{jk}, \quad \text{for all } j = 1, \ldots, J, \tag{1}$$

with probability greater than $1 - \alpha - o(1)$*, where* C *depends on the [RE] coefficients*

Nagaev Type of Inequality

Theorem 2

Under [A1] *and* [A3], *we have*

$$\mathrm{P}\left(2c\sqrt{n}\max_{j,k}|S_{jk}/\Psi_{jk}| \geq r\right) \leq C_1 \varpi_n n r^{-q} \sum_{j=1}^{J}\sum_{k=1}^{K} \frac{\|X_{jk,\cdot}\varepsilon_{j,\cdot}\|_{q,\varsigma}^{q}}{\Psi_{jk}^{q}} + C_2 \sum_{j=1}^{J}\sum_{k=1}^{K} \exp\left(\frac{-C_3 r^2 \Psi_{jk}^2}{n\|X_{jk,\cdot}\varepsilon_{j,\cdot}\|_{2,\varsigma}^{2}}\right), \quad (2)$$

where for $\varsigma > 1/2 - 1/q$ *(weak dependence case),* $\varpi_n = 1$*; for* $\varsigma < 1/2 - 1/q$ *(strong dependence case),* $\varpi_n = n^{q/2-1-\varsigma q}$*.*
C_1, C_2, C_3 *are constants depending* q *and* ς*.*

Oracle Inequalities under $\lambda^0(1-\alpha)$

Corollary 3

Under [A1] *and* [A3]*, given*

$$\lambda^0(1-\alpha) \lesssim \max_{j,k} \Big(\|X_{jk,\cdot}\varepsilon_{j,\cdot}\|_{2,\varsigma}\{n \log(KJ/\alpha)\}^{1/2} \vee \|X_{jk,\cdot}\varepsilon_{j,\cdot}\|_{q,\varsigma}(n\varpi_n KJ/\alpha)^{1/q} \Big),$$

additionally suppose [A2] *holds with probability* $1 - \mathcal{o}(1)$*, then under the exact sparsity assumption,*

$$|\widehat{\beta}_j - \beta_j^0|_{j,\mathrm{pr}} \lesssim C\sqrt{s} \max_k \Psi_{jk} \max_j \Big\{ \|X_{jk,\cdot}\varepsilon_{j,\cdot}\|_{2,\varsigma} n^{-1/2}\{\log(KJ/\alpha)\}^{1/2} \vee \|X_{jk,\cdot}\varepsilon_{j,\cdot}\|_{q,\varsigma} n^{1/q-1}(\varpi_n KJ/\alpha)^{1/q} \Big\},$$

with probability $1 - \alpha - \mathcal{o}(1)$*.*

Empirical Choices of λ

- Gaussian Approximation:
  $Q(1-\alpha) \stackrel{\text{def}}{=} 2c\sqrt{n}\Phi^{-1}\{1-\alpha/(2JK)\}$
- Multiplier Bootstrap: selected by an algorithm
- Dependency over time: groups the data into blocks and resample the blocks

Gaussian Approximation

The Kolmogorov distance between two rv X and Y:

$$\rho(X, Y) = \sup_{r \geq 0} \left| \mathsf{P}(|X|_\infty \leq r) - \mathsf{P}(|Y|_\infty \leq r) \right|.$$

Theorem 4

Let $\widetilde{\mathcal{X}}_t \stackrel{\text{def}}{=} \operatorname{vec}\{(X_{jk,t}\varepsilon_{j,t})_{jk}\}$, $\widetilde{\mathcal{S}} \stackrel{\text{def}}{=} \operatorname{vec}\{(S_{jk})_{jk}\} = n^{-1/2} \sum_{t=1}^{n} \widetilde{\mathcal{X}}_t$, *and define the aggregated* ▸ *dependence adjusted norm over* j *and* k*, under* ▸ *[A1] and* ▸ *[A3]-[A4], we have*

$$\rho\left(D^{-1}\widetilde{\mathcal{S}}, D^{-1}\widetilde{\mathcal{Z}}\right) \to 0, \qquad \text{as } n \to \infty, \tag{3}$$

where $\widetilde{\mathcal{Z}} \sim \mathsf{N}(0, \Sigma_{\widetilde{\mathcal{X}}})$, *and* $\Sigma_{\widetilde{\mathcal{X}}}$ *is the* $JK \times JK$ *long run variance-covariance matrix of* $\widetilde{\mathcal{X}}_t$, D *is a diagonal matrix with the square root of the diagonal elements of* $\Sigma_{\widetilde{\mathcal{X}}}$.

Gaussian Approximation

Corollary 5

Under the conditions of Theorem 4, we have

$$\sup_{\alpha \in (0,1)} |\,\mathrm{P}\{\max_{j,k} 2c\sqrt{n}|S_{jk}/\Psi_{jk}| \leq \widetilde{Q}(1-\alpha)\} - (1-\alpha)| \to 0, \quad (4)$$

for sufficiently large n, where $\widetilde{Q}(1-\alpha) = (1-\alpha)$ *quantile of* $2c\sqrt{n}\max_{j,k}|Z_{jk}/\Psi_{jk}|$*, with* Z_{jk} *is Gaussian centered rv with the same long run variance-covariance structure as* S_{jk}*.*

Algorithm for Multiplier Bootstrap

- 1. LASSO for each equation

$$\widetilde{\beta}_j = \arg \min_{\beta \in \mathbb{R}^{K_j}} \frac{1}{n} \sum_{t=1}^{n} (Y_{j,t} - X_{j,t}^\top \beta)^2 + \frac{\lambda_j}{n} \sum_{k=1}^{K_j} |\beta_{jk}| \Psi_{jk},$$

with $\lambda_j = 2c'\sqrt{n}\Phi^{-1}(1 - \alpha'/(2K_j))$, $\alpha' = 0.1$, $c' = 0.5$, $\Psi_{jk} = \sqrt{\mathrm{Var}(X_{jk,t}\breve{\varepsilon}_{j,t})}$, and $\breve{\varepsilon}_{j,t}$ are some preliminary estimates of the errors.

- 2. Keep $\widetilde{\varepsilon}_{j,t} = Y_{j,t} - X_{j,t}^\top \widetilde{\beta}_j$ and update Ψ_{jk} with $\widetilde{\varepsilon}_{j,t}$.

Algorithm for Multiplier Bootstrap

- 3. Divide $\{\widetilde{\varepsilon}_{j,t}\}$ into l_n blocks, each contains $b_n = n/l_n$ observations. $\Lambda(1-\alpha) \stackrel{\text{def}}{=} 2c\sqrt{n}q^{[B]}_{(1-\alpha)}$, $c > 1$, $\alpha = 0.1$, where $q^{[B]}_{(1-\alpha)}$ is the $(1-\alpha)$ quantile of $\max_{j,k} |Z^{[B]}_{jk}/\Psi_{jk}|$, and

$$Z^{[B]}_{jk} = \frac{1}{\sqrt{n}} \sum_{i=1}^{l_n} e_{j,i} \sum_{l=(i-1)b_n+1}^{ib_n} \widetilde{\varepsilon}_{j,l} X_{jk,l}, \qquad (5)$$

where e_i are drawn from i.i.d. $\mathrm{N}(0,1)$.

Multiplier Bootstrap for b)

Theorem 6 (Validity of Multiplier Bootstrap)

Under [A1], [A3], *and assume* $\Phi_{2q,\varsigma} < \infty$ *with* $q > 4$, $b_n = \mathcal{O}(n^\eta)$ *for some* $0 < \eta < 1$, *let* $\widetilde{\mathcal{Z}}^{[B]} \stackrel{\text{def}}{=} \text{vec}\{(Z_{jk}^{[B]})_{jk}\}$, *and* $\widetilde{\Psi} \stackrel{\text{def}}{=} \text{vec}\{(\Psi_{jk})_{jk}\}$, *then*

$$\widetilde{\rho}_n \stackrel{\text{def}}{=} \sup_{r \in \mathbb{R}} |\,\mathsf{P}(|\widetilde{\mathcal{Z}}^{[B]}/\widetilde{\Psi}|_\infty \leq r | \mathcal{X}_\cdot, \varepsilon_\cdot) - \mathsf{P}(|\widetilde{\mathcal{Z}}/\widetilde{\Psi}|_\infty \leq r)| \to 0, \text{ as } n \to \infty,$$

$$\sup_{\alpha \in (0,1)} |\,\mathsf{P}(|\widetilde{\mathcal{S}}/\widetilde{\Psi}|_\infty \leq q_{(1-\alpha)}^{[B]}) - (1-\alpha)| \to 0, \text{ as } n \to \infty.$$

Coefficient Estimation Procedure

- **Step** 1: LASSO for $Y_{j,t} = X_{jk,t}\beta^0_{jk} + X^\top_{j(-k),t}\beta^0_{j(-k)} + \varepsilon_{j,t}$, keep $\widehat{\beta}^{[1]}_{j(-k)}$
- **Step** 2: LASSO for $X_{jk,t} = X^\top_{j(-k),t}\gamma^0_{j(-k)} + v_{jk,t}$, keep the residuals $\widehat{v}_{jk,t} = X_{jk,t} - X^\top_{j(-k),t}\widehat{\gamma}_{j(-k)}$
- **Step** 3: Regress $Y_{j,t} - X^\top_{j(-k),t}\widehat{\beta}^{[1]}_{j(-k)}$ on $X_{jk,t}$ using $\widehat{v}_{jk,t}$ as IV, finally achieve $\widehat{\beta}^{[2]}_{jk}$

Orthogonality

Uniformly Valid Inference

Let $\psi_{jk}(Z_{j,t}, \beta_{jk}, h_{jk})$ denote the score, where $Z_{j,t} = (Y_{j,t}, X_{j,t}^\top)^\top$, $h_{jk}(X_{j(-k),t}) = (X_{j(-k),t}^\top \beta_{j(-k)}, X_{j(-k),t}^\top \gamma_{j(-k)})^\top$, for $(j,k) \in G$.

Theorem 7

Under conditions*, let* $\omega_{jk} \stackrel{\text{def}}{=} \mathsf{E}[\{\frac{1}{\sqrt{n}} \sum_{t=1}^n \psi_{jk}(Z_{j,t}, \beta_{jk}^0, h_{jk}^0)\}^2]$, $\phi_{jk} \stackrel{\text{def}}{=} \frac{\partial}{\partial \beta} \mathsf{E}\{\psi_{jk}(Z_{j,t}, \beta, h_{jk}^0)\}\big|_{\beta=\beta_{jk}^0}$*, we have*

$$\max_{(j,k)\in G} \Big|\sqrt{n}\sigma_{jk}^{-1}(\widehat{\beta}_{jk} - \beta_{jk}^0) - n^{-1/2}\sum_{t=1}^n \zeta_{jk,t}\Big| = o(g_n^{-1}), \text{ as } n \to \infty$$

with probability $1 - o(1)$*, where* $\sigma_{jk}^2 \stackrel{\text{def}}{=} \phi_{jk}^{-2}\omega_{jk}$, $\zeta_{jk,t} \stackrel{\text{def}}{=} -\phi_{jk}^{-1}\sigma_{jk}^{-1}\psi_{jk}(Z_{j,t}, \beta_{jk}^0, h_{jk}^0)$, $g_n \stackrel{\text{def}}{=} \{\log(e|G|)\}^{1/2}$.

CI for Individual Inference

- $H_0 : \beta^0_{jk} = 0$
- CI by asymptotic normality:
  $[\widehat{\beta}^{[2]}_{jk} - \widehat{\sigma}_{jk} n^{-1/2} \Phi^{-1}(1-\alpha/2), \widehat{\beta}^{[2]}_{jk} + \widehat{\sigma}_{jk} n^{-1/2} \Phi^{-1}(1-\alpha/2)]$
- Multiplier block bootstrap:
  - $T^*_{jk} = \frac{1}{\sqrt{n}} \sum_{i=1}^{l_n} e_{j,i} \sum_{l=(i-1)b_n+1}^{ib_n} \widehat{\zeta}_{jk,l}$, $T_{jk} = \frac{\sqrt{n}(\widehat{\beta}^{[2]}_{jk} - \beta^{0}_{jk})}{\widehat{\sigma}_{jk}}$
  - $[\widehat{\beta}^{[2]}_{jk} - \widehat{\sigma}_{jk} n^{-1/2} q^*_{(1-\alpha)}, \widehat{\beta}^{[2]}_{jk} + \widehat{\sigma}_{jk} n^{-1/2} q^*_{(1-\alpha)}]$, $q^*_{(1-\alpha)}$ is the $(1-\alpha)$ quantile of $|T^*_{jk}|$

Confidence Region (CR) for Simult. Inference

- $H_0 : \beta^0_{jk} = 0, \forall (j,k) \in G$
- Define $q^*_G(1-\alpha)$ as the $(1-\alpha)$ quantile of $\max_{(j,k)\in G} |T^*_{jk}|$
- Simultaneous confidence region: $\{\beta \in \mathbb{R}^{|G|} : \max_{(j,k)\in G} T_{jk} \leq q^*_G(1-\alpha) \text{ and } \min_{(j,k)\in G} T_{jk} \geq -q^*_G(1-\alpha)\}$
- For each component $(j,k) \in G$: $\widetilde{\mathrm{CI}}^*_{jk}(\alpha) = [\widehat{\beta}^{[2]}_{jk} - \widehat{\sigma}_{jk} n^{-1/2} q^*_G(1-\alpha), \widehat{\beta}^{[2]}_{jk} + \widehat{\sigma}_{jk} n^{-1/2} q^*_G(1-\alpha)]$

Consistency of the Bootstrapped CR

Corollary 8

Under conditions*, we have*

$$\sup_{\alpha\in(0,1)} \left| \mathsf{P}(\beta^0_{jk} \in \widetilde{\mathsf{CI}}^*_{jk}(\alpha),\ \forall (j,k) \in G) - (1-\alpha) \right| = o(1), \text{ as } n \to \infty,$$

with probability $1 - o(1)$.

Predictive Performance

DGP 1:

$$Y_{j,t} = X_t^\top \beta_j^0 + \varepsilon_{j,t}, \quad t = 1, \dots, n, \; j = 1, \dots, J$$

- $X_t \in \mathbb{R}^K \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \Sigma)$, with $\Sigma_{k_1,k_2} = \gamma^{|k_1 - k_2|}$, $\gamma = 0.5$, $\varepsilon_{j,t} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, 1)$
- divide $\{1, \dots, K\}$ evenly into blocks with fixed block size 5, $\beta^0_{jk} = 10$ if k and j belong to one block and 0 otherwise
- $n = 100$, take 5000 bootstrap replications

Predictive Performance

	$J = K = 50$	$J = K = 100$	$J = K = 150$
		Prediction norm	
Mean	0.96	0.95	0.93
Median	0.97	0.95	0.94
Std.	0.03	0.03	0.03
		Euclidian norm	
Mean	0.96	0.94	0.93
Median	0.97	0.95	0.93
Std.	0.04	0.03	0.03

Table 1: Prediction norm and Euclidean norm ratios (overall λ relative to single λ_j's, average over equations). Results are computed over 1000 repeats of simulations.

Predictive Performance

DGP 2:

$$Y_{j,t} = X_t^\top \beta_j^0 + \varepsilon_{j,t}, \quad t = 1, \ldots, n, \ j = 1, \ldots, J, \ J = K$$

- $X_t = \sum_{\ell=0}^{\infty} A_\ell \xi_{t-\ell}$, $A_\ell = (\ell + 1)^{-\rho-1} M_\ell$, where the entries of M_ℓ are i.i.d. N(0, 1). In practice truncate the sum to $\ell = 1000$.
- $\xi_{k,t} = e_{k,t}(0.8e_{k,t-1}^2 + 0.2)^{1/2}$, where $e_{k,t}$ are i.i.d. from $t(d)/\sqrt{d/(d-2)}$ with $d = 8$
- ε_t are generated by the same fashion independently

Predictive Performance

Optimal Choice of b_n:

- Theoretical bias-variance trade-off results in an admissible range of the rate
- Depends on the dependency and the dimensionality
- In practice, take the one giving the lowest prediction norm on a grid search

Predictive Performance

	$\rho = 0.1$ (stronger dependency)			$\rho = 1.0$ (weaker dependency)		
	$J = 50$	$J = 100$	$J = 150$	$J = 50$	$J = 100$	$J = 150$
$b_n = 2$	2.07	2.91	3.59	**2.02**	2.63	3.23
$b_n = 4$	2.06	2.89	3.56	2.03	**2.62**	3.223
$b_n = 6$	2.05	2.90	3.52	2.08	2.63	**3.220**
$b_n = 8$	**2.04**	2.8841	**3.51**	2.21	2.65	3.23
$b_n = 10$	2.05	**2.8836**	3.53	2.36	2.71	3.30
$b_n = 12$	2.06	2.91	3.57	2.56	2.83	3.39

Table 2: The prediction norm (average over equations) using several choices of b_n. Results are computed over 1000 simulations.

Predictive Performance

	$\rho = 0.1$ (stronger dependency)			$\rho = 1.0$ (weaker dependency)		
	$J = 50$	$J = 100$	$J = 150$	$J = 50$	$J = 100$	$J = 150$
	Prediction norm					
Mean	0.91	0.85	0.83	0.94	0.88	0.83
Median	0.92	0.85	0.83	0.94	0.88	0.83
Std.	0.04	0.04	0.03	0.04	0.03	0.03
	Euclidean norm					
Mean	0.90	0.84	0.81	0.93	0.86	0.82
Median	0.91	0.85	0.81	0.93	0.87	0.82
Std.	0.05	0.04	0.03	0.05	0.04	0.03

Table 3: Prediction norm and Euclidean norm ratios (overall λ relative to single λ_j's, average over equations, $J = K$). Results are computed over 1000 repeats of simulations.

Inference Performance

$$Y_{j,t} = d_{j,t}\alpha_j^0 + X_t^\top \beta_j^0 + \varepsilon_{j,t},\ d_{j,t} = X_t^\top \theta_j^0 + v_{j,t},\ t = 1, \ldots, n,\ j = 1, \ldots, J$$

- $\alpha_j^0 = \alpha^0$ for $j = 1, \ldots, J$
- Block diagonal structure in $\{\beta_{jk}^0\}$ and $\{\theta_{jk}^0\}$:
  - divide $\{1, \ldots, K\}$ evenly into blocks with fixed block size 5
  - if k and j belong to one block $\beta_{jk}^0 \sim \text{Unif}[0, 5]$, $\theta_{jk}^0 \sim \text{Unif}[0, 0.25]$
- X_t, ε_t, v_t are generated as dependent data by the same way

Inference Performance

	$\rho = 0.1$ (stronger dependency)			$\rho = 1.0$ (weaker dependency)		
	$J = 50$	$J = 100$	$J = 150$	$J = 50$	$J = 100$	$J = 150$
			$\alpha^0 = 0$			
Ind. Asym.	0.017	0.013	0.013	0.024	0.015	0.012
Ind. Boot.	0.030	0.020	0.016	0.022	0.017	0.014
Simult. Boot.	0.026	0.047	0.053	0.052	0.055	0.059
			$\alpha^0 \sim \text{Unif}[0, 2.5]$			
Ind. Asym.	0.871	0.856	0.855	0.876	0.862	0.857
Ind. Boot.	0.875	0.857	0.857	0.876	0.863	0.858
Mult. Boot.	0.841	0.803	0.800	0.844	0.825	0.809
			$\alpha^0 \sim \text{Unif}[0, 5]$			
Ind. Asym.	0.938	0.925	0.928	0.938	0.932	0.927
Ind. Boot.	0.939	0.925	0.933	0.929	0.933	0.927
Mult. Boot.	0.928	0.907	0.907	0.926	0.918	0.908

Table 4: Average rejection rate of $H_0^j : \alpha_j^0 = 0$ over j for the ind. (or mult.) inference and the rejection rate of $H_0 : \alpha_1^0 = \cdots = \alpha_J^0 = 0$ for simult. inference (significance level = 0.05).

Data Source

- Textual sentiment effect on financial variables
- Financial news articles on NASDAQ community platform
- Unsupervised learning approach to extract sentiment variable
- Sentiment words lists - BL option lexicon and LM financial sentiment dictionary
- Bullishness indicator based on the average proportion of positive/negative words (Zhang et al. 2016)

Data Source

- 63 S&P 500 constituents stocks from 9 GICS sectors
- Response: stock returns and volatilities
- Controls: S&P 500 index returns and CBOE VIX index
- Daily data from January 2, 2015 to December 31, 2015
- Spillover effects over individual stocks and sectors

Model Setting

$$r_{j,t} = c_j + B_t^\top \beta_j + z_t^\top \gamma_j + r_{j,t-1}\delta_j + \varepsilon_{j,t},$$

or

$$\log \sigma^2_{j,t} = c_j + B_t^\top \beta_j + z_t^\top \gamma_j + \log \sigma^2_{j,t-1}\delta_j + \varepsilon_{j,t},$$

where the sentiment variables and control variables are included in $B_t = (B_{1,t}, \ldots, B_{J,t})^\top$ and z_t.

Model Setting - ctd

- Bullishness for stock j on day t with the related article i:

$$B_{j,t} = \log \left[\frac{\{1 + m^{-1} \sum_{i=1}^{m} \mathbf{I}(Pos_{j,i,t} > Neg_{j,i,t})\}}{\{1 + m^{-1} \sum_{i=1}^{m} \mathbf{I}(Pos_{j,i,t} < Neg_{j,i,t})\}} \right].$$

  $Pos_{j,i,t}$, $Neg_{j,i,t}$ are the average proportion of positive/negative words based on the lexicon

- Response variables

$$r_{j,t} = \log(P^C_{j,t}) - \log(P^O_{j,t}),$$

$$\sigma^2_{j,t} = 0.511(u_{j,t} - d_{j,t})^2 - 0.019\{r_{j,t}(u_{j,t} + d_{j,t}) - 2u_{j,t}d_{j,t}\} - 0.383r^2_{j,t},$$

  $u_{j,t} = \log(P^H_{j,t}) - \log(P^O_{j,t})$, $d_{j,t} = \log(P^L_{j,t}) - \log(P^O_{j,t})$, with $P^H_{j,t}$, $P^L_{j,t}$, $P^O_{j,t}$, and $P^C_{j,t}$ are the highest, lowest, opening and closing prices. Garman and Klass (1980)

Graphical network - Individual Inference

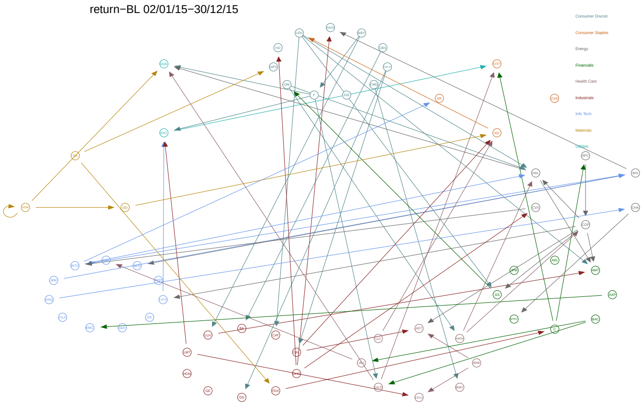


Figure 4: Graphical network among individual stocks (return - BL)

Graphical network - Individual Inference

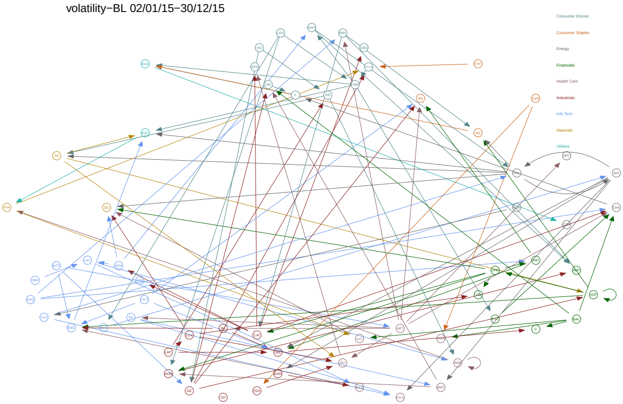


Figure 5: Graphical network among individual stocks (volatility - BL)

Graphical network - Individual Inference

Example: dependency between two stocks

- textual sentiment effect on stock return $H_0^{jk}: \beta_{jk} = 0$
- directional edge from "DOW" to "DD"
- self effect of "DOW"

Figure 6: Dependency between DOW and DD (return - BL)

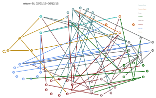
Graphical network - Simultaneous Inference

- Joint sentiment effect from sector S_1 on returns of sector S_2
- Simult. inference on $H_0 : \beta_{jk} = 0, \; \forall j \in S_1, \, k \in S_2$
- Conclusions:
  - returns: energy$\rightarrow$health care
  - volatility: financials$\rightarrow$health care, IT$\rightarrow$energy, consumer discretionary$\rightarrow$utilities

LASSO-Driven Inference in Time and Space

Victor Chernozhukov
Wolfgang K. Härdle
Chen Huang
Weining Wang

MIT Center for Statistics
Humboldt–Universität zu Berlin
University of St.Gallen
City, University of London
http://stat.mit.edu
http://lvb.wiwi.hu-berlin.de
http://mathstat.unisg.ch
http://www.city.ac.uk



Single Equation LASSO Performance

Theorem 1 of Belloni and Chernozhukov (2013)

Suppose the restricted eigenvalue condition holds, under the exact sparsity assumption and given the event $\lambda_j \geq 2c\sqrt{n} \max_{1\leq k\leq K} |S_{jk}/\Psi_{jk}|$, *then* $\widetilde{\beta}_j$ *obtained from single equation LASSO satisfy*

$$
\begin{aligned}
|\widetilde{\beta}_j - \beta_j^0|_{j,pr} &= \left[\frac{1}{n} \sum_{t=1}^{n} \left\{ X_{j,t}^\top (\widetilde{\beta}_j - \beta_j^0) \right\}^2 \right]^{1/2} \\
&\leq (1 + 1/c) \frac{\lambda_j \sqrt{s_j}}{n \kappa_j(\overline{c})} \max_{1\leq k\leq K} \Psi_{jk}. \qquad (6)
\end{aligned}
$$

▸ Ideal λ

Measure of Dependence [by Wu (2005)]

[A1] Assume $X_{jk,t} = g_{jk}(\dots, \xi_{t-1}, \xi_t)$, where ξ_t are i.i.d. random elements (innovations or shocks) across t and $g_{jk}(\cdot)$ are measurable functions (filters).

- Replace ξ_0 by an i.i.d. copy of ξ_0^*, and $X_{jk,t}^* = g_{jk}(\dots, \xi_0^*, \dots, \xi_t)$
- Functional dependence measure $\delta_{q,j,k,t} \stackrel{\text{def}}{=} \|X_{jk,t} - X_{jk,t}^*\|_q$, $q \geq 1$, which measures the effect of ξ_0 on $X_{jk,t}$;
  $\Delta_{m,q,j,k} \stackrel{\text{def}}{=} \sum_{t=m}^{\infty} \delta_{q,j,k,t}$, which measures the cumulative effect of ξ_0 on $X_{jk,t \geq m}$
- Dependence adjusted norm of $X_{jk,t}$:
  $\|X_{jk,\cdot}\|_{q,\varsigma} = \sup_{m \geq 0}(m+1)^{\varsigma} \Delta_{m,q,j,k}$, $\varsigma > 0$

Measure of Temporal Dependency

Example: AR(1) process

$$X_t = \alpha X_{t-1} + \xi_t = \sum_{\ell=0}^{\infty} \alpha^\ell \xi_{t-\ell},\ |\alpha| < 1.$$

- $\delta_{q,t} = \|X_t^* - X_t\|_q = \|\alpha^t \xi_0^* - \alpha^t \xi_0\|_q = |\alpha|^t \|\xi_0^* - \xi_0\|_q$,
  $\Delta_{m,q} = \sum_{t=m}^{\infty} \delta_{q,t} = \|\xi_0^* - \xi_0\|_q \sum_{t=m}^{\infty} |\alpha|^t \propto |\alpha|^m$
- $\|X_\cdot\|_{q,\varsigma} = \sup_{m\geq 0}(m+1)^\varsigma \Delta_{m,q} < \infty$

Measure of Spatial Dependency

Example: Spatial MA structure in the errors

$$\varepsilon_t = \rho W \varepsilon_t = \sum_{\ell=0}^{\infty} \rho^\ell W^\ell \eta_{t-\ell}, \ \max_j |[\rho^t W^t]_j|_1 \le |c|^t, \ |c| < 1.$$

- $\delta_{q,t} = \|\varepsilon^*_{j,t} - \varepsilon_{j,t}\|_q = \|[\rho^t W^t]_j(\eta_0^* - \eta_0)\|_q \le |[\rho^t W^t]_j|_1 \max_j \|\eta^*_{j,0} - \eta_{j,0}\|_q$,
  $\Delta_{m,q} = \sum_{t=m}^{\infty} \delta_{q,t} = \max_j \|\eta^*_{j,0} - \eta_{j,0}\|_q \sum_{t=m}^{\infty} |c|^t \propto |c|^m$
- $\|\varepsilon_{j,\cdot}\|_{q,\varsigma} = \sup_{m \ge 0}(m+1)^\varsigma \Delta_{m,q} < \infty$

▸ Contributions

Restricted Eigenvalue (RE) Condition

[A2] (RE uniformly) Given $c \geq 1$, for $\eta \in \mathbb{R}^K$,

$$\kappa_j(c) \stackrel{\text{def}}{=} \min_{|\eta_{T_j^c}|_1 \leq c|\eta_{T_j}|_1, \eta \neq 0} \frac{\sqrt{s_j}|\eta|_{j,\text{pr}}}{|\eta_{T_j}|_1} > 0,$$

holds uniformly over $j = 1, \ldots, J$ with probability $1 - o(1)$, where $T_j \stackrel{\text{def}}{=} \{k : \beta^0_{jk} \neq 0\}$ and $s_j = |T_j| = o(n)$, $\eta_{T_j k} = \eta_k$ if $k \in T_j$, $\eta_{T_j k} = 0$ if $k \notin T_j$.

▸ Error Bounds for Prediction Norm

Moment Conditions

[A3] $\|\varepsilon_{j,\cdot}\|_{q,\varsigma} < \infty$, and $\|X_{jk,\cdot}\|_{q,\varsigma} < \infty$.

▸ Nagaev Inequality

Aggregation over High Dimensions

For single equation j, let

- $\Phi_{j,q,\varsigma} = 2\max_k \|X_{jk,\cdot}\|_{q,\varsigma}\|\varepsilon_{j,\cdot}\|_{q,\varsigma}$
- $\Gamma_{j,q,\varsigma} = 2\|\varepsilon_{j,\cdot}\|_{q,\varsigma}(\sum_k \|X_{jk,\cdot}\|_{q,\varsigma}^{q/2})^{2/q}$
- $\Theta_{j,q,\varsigma} = \Gamma_{j,q,\varsigma} \wedge \{2\||X_{j,\cdot}|_\infty\|_{q,\varsigma}\|\varepsilon_{j,\cdot}\|_{q,\varsigma}(\log KJ)^{3/2}\}$, where $\||X_{j,\cdot}|_\infty\|_{q,\varsigma} = \sup_{m\geq 0}(m+1)^\varsigma \sum_{t=m}^\infty \||X_{j,t} - X_{j,t}^*|_\infty\|_q$

Over all equations, let $\mathcal{X}_t \overset{\text{def}}{=} \text{vec}\{(X_{jk,t})_{jk}\}$

- $\Phi_{q,\varsigma} = \max_{j,k} 2\|X_{jk,\cdot}\|_{q,\varsigma}\|\varepsilon_{j,\cdot}\|_{q,\varsigma}$
- $\Gamma_{q,\varsigma} = 2(\sum_j \|\varepsilon_{j,\cdot}\|_{q,\varsigma}^{q/2})^{2/q}(\sum_{j,k} \|X_{jk,\cdot}\|_{q,\varsigma}^{q/2})^{2/q}$
- $\Theta_{q,\varsigma} = \Gamma_{q,\varsigma} \wedge \{\||\mathcal{X}_\cdot|_\infty\|_{q,\varsigma}\|\varepsilon_{j,\cdot}\|_{q,\varsigma}(\log KJ)^{3/2}\}$, where $\||\mathcal{X}_\cdot|_\infty\|_{q,\varsigma} = \sup_{m\geq 0}(m+1)^\varsigma \sum_{t=m}^\infty \||\mathcal{X}_t - \mathcal{X}_t^*|_\infty\|_q$

▸ Gaussian Approximation

More Assumptions

[A4] i)(weak dependency case) Given $\Theta_{2q,\varsigma} < \infty$ with $q \geq 2$ and $\varsigma > 1/2 - 1/q$, then $\Theta_{2q,\varsigma} n^{1/q-1/2}\{\log(KJn)\}^{3/2} \to 0$ and $L_1 \max(W_1, W_2) = o(1) \min(N_1, N_2)$;
ii) (strong dependency case) given $0 < \varsigma < 1/2 - 1/q$, then $\Theta_{2q,\varsigma}\{\log(KJ)\}^{1/2} = o(n^{\varsigma})$ and $L_1 \max(W_1, W_2, W_3) = o(1) \min(N_2, N_3)$;
where $L_1 = [\Phi_{4,\varsigma}\Phi_{4,0}\{\log(KJ)\}^2]^{1/\varsigma}$,
$W_1 = (\Phi_{6,0}^6 + \Phi_{8,0}^4)\{\log(KJn)\}^7$, $W_2 = \Phi_{4,\varsigma}^2\{\log(KJn)\}^4$,
$W_3 = [n^{-\varsigma}\{\log(KJn)\}^{3/2}\Theta_{2q,\varsigma}]^{1/(1/2-\varsigma-1/q)}$,
$N_1 = \{n/\log(KJ)\}^{q/2}\Theta_{2q,\varsigma}^{q}$, $N_2 = n\{\log(KJ)\}^{-2}\Phi_{4,\varsigma}^{-2}$,
$N_3 = [n^{1/2}\{\log(KJ)\}^{-1/2}\Theta_{2q,\varsigma}^{-1}]^{1/(1/2-\varsigma)}$.

▸ Gaussian Approximation

Orthogonality Property

Use $v_{jk,t}$ as an instrument in the following moment equation (e.g. mean regression case) for the target coefficient β^0_{jk}

$$\mathsf{E}[\psi_{jk}\{Z_{j,t}, \beta^0_{jk}, h^0_{jk}(X_{j(-k),t})\}] = \mathsf{E}(\varepsilon_{j,t} v_{jk,t}) = 0,$$

which has the orthogonality property

$$\frac{\partial}{\partial \beta_{j(-k)}} \mathsf{E}[\psi_{jk}\{Z_{j,t}, \beta^0_{jk}, h_{jk}(X_{j(-k),t})\}]\Big|_{\beta_{j(-k)}=\beta^0_{j(-k)}} = 0,$$
$$\frac{\partial}{\partial \gamma_{j(-k)}} \mathsf{E}[\psi_{jk}\{Z_{j,t}, \beta^0_{jk}, h_{jk}(X_{j(-k),t})\}]\Big|_{\gamma_{j(-k)}=\gamma^0_{j(-k)}} = 0.$$

▸ Estimation

Conditions for Theorem 7

- **Properties of ψ_{jk}:** the map $(\beta, h) \mapsto \mathsf{E}\{\psi_{jk}(Z_{j,t}, \beta, h) | X_{j(-k),t}\}$ is twice continuously differentiable, and for every $\vartheta \in \{\beta, h_1, \ldots, h_M\}$, $\mathsf{E}(\sup_{\beta \in \mathcal{B}_{jk}} |\partial_\vartheta \mathsf{E}[\psi_{jk}\{Z_{j,t}, \beta, h^0_{jk}(X_{j(-k),t})\} | X_{j(-k),t}]|^2) \leq C_1$; moreover, there exist constants $L_{1n}, L_{2n} \geq 1$, $\nu > 0$ and a cube $\mathcal{T}_{jk}(X_{j(-k),t}) = \times_{m=1}^{M} \mathcal{T}_{jk,m}(X_{j(-k),t})$ in $\mathbb{R}^M$ with center $h^0_{jk}(X_{j(-k),t})$ such that for every $\vartheta, \vartheta' \in \{\beta, h_1, \ldots, h_M\}$, $\sup_{(\beta,h) \in \mathcal{B}_{jk} \times \mathcal{T}_{jk}(X_{j(-k),t})} |\partial_\vartheta \partial_{\vartheta'} \mathsf{E}\{\psi_{jk}(Z_{j,t}, \beta, h) | X_{j(-k),t}\}| \leq \ell_1(X_{j(-k),t})$, $\mathsf{E}\{|\ell_1(X_{j(-k),t})|^4\} \leq L_{1n}$, and for every $\beta, \beta' \in \mathcal{B}_{jk}$, $h, h' \in \mathcal{T}_{jk}(X_{j(-k),t})$, $\mathsf{E}[\{\psi_{jk}(Z_{j,t}, \beta, h) - \psi_{jk}(Z_{j,t}, \beta', h')\}^2 | X_{j(-k),t}] \leq \ell_2(X_{j(-k),t})(|\beta - \beta'|^\nu + |h - h'|_2^\nu)$, $\mathsf{E}\{|\ell_2(X_{j(-k),t})|^4\} \leq L_{2n}$.
- The 2nd-order moments of scores are bounded away from zero, $\omega_{jk} = \mathsf{E}\{(\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \psi^0_{jk,t})^2\} \geq c_1$, $\psi^0_{jk,t} = \psi_{jk}\{Z_{j,t}, \beta^0_{jk}, h^0_{jk}(X_{j(-k),t})\}$.
- $f_{\varepsilon_j}(\cdot)$ is continuously differentiable, $f_{\varepsilon_j}(\cdot)$ and $f'_{\varepsilon_j}(\cdot)$ are bounded from the above.

Conditions for Theorem 7

- **Properties of the nuisance function:** with probability $1 - \mathcal{o}(1)$, $\widehat{h}_{jk} \in \mathcal{H}_{jk}$, where $\mathcal{H}_{jk} = \times_{m=1}^{M} \mathcal{H}_{jk,m}$ with each $\mathcal{H}_{jk,m}$ being the class of functions of the form $\widetilde{h}_{jk,m}(X_{j(-k),t}) = X_{j(-k),t}^{\top}\theta_{jk,m}$, $\|\theta_{jk,m}\|_0 \leq s$, $\widetilde{h}_{jk,m} \in \mathcal{T}_{jk,m}$. There exists sequence of constants $\rho_n \downarrow 0$ such that $\mathsf{E}[\{\widetilde{h}_{jk,m}(X_{j(-k),t}) - h^0_{jk,m}(X_{j(-k),t})\}^2] \lesssim \rho_n^2$.
- The true parameter β^0_{jk} satisfies $\mathsf{E}[\psi_{jk}\{Z_{j,t}, \beta_{jk}, h^0_{jk}(X_{j(-k),t})\}] = 0$. Let $\mathcal{B}_{jk}$ be a fixed and closed interval and $\widehat{\mathcal{B}}_{jk}$ be a possibly stochastic interval such that with probability $1 - \mathcal{o}(1)$, $[\beta^0_{jk} \pm c_1 r_n] \subset \widehat{\mathcal{B}}_{jk} \subset \mathcal{B}_{jk}$, $r_n = n^{-1/2}(\log a_n)^{1/2} \max\limits_{(j,k)\in G} \|\psi^0_{jk,\cdot}\|_{2,\varsigma} + n^{-1} r_\varsigma (\log a_n)^{3/2} \big\| \max\limits_{(j,k)\in G} |\psi^0_{jk,\cdot}| \big\|_{q,\varsigma}$, $r_n \lesssim \rho_n$, where $a_n = \max(JK, n, e)$, $r_\varsigma = n^{1/q}$ for $\varsigma > 1/2 - 1/q$ and $r_\varsigma = n^{1/2-\varsigma}$ for $\varsigma < 1/2 - 1/q$.

Conditions for Theorem 7 ▸ Theorem 7

- **Identifiability:**
  $2|\mathrm{E}[\psi_{jk}\{Z_{j,t},\beta,h^0_{jk}(X_{j(-k),t})\}]| \geq |\phi_{jk}(\beta-\beta^0_{jk})| \wedge c_1$ holds for all $\beta \in \mathcal{B}_{jk}$, and $|\phi_{jk}| \geq c_1$.
- $\mathcal{F}_{jk} = \{z \mapsto \psi_{jk}\{z,\beta,\widetilde{h}(x_{j(-k)})\} : \beta \in \mathcal{B}_{jk}, \widetilde{h} \in \mathcal{H}_{jk} \cup \{h^0_{jk}\}\}$ is pointwise measurable and has envelope $F_{jk} \geq \sup_{f\in\mathcal{F}_{jk}} |f|$, such that $F = \max_{(j,k)\in G} F_{jk}$ satisfies $\mathrm{E}\{F^q(z)\} < \infty$ for some $q \geq 4$.
- **Dimension growth rates:**
  $\rho_{n,\upsilon}(L_{2n} s \log a_n)^{1/2} + n^{-1/2} r_\varsigma (s \log a_n)^{3/2} \|F(z_t)\|_q + \rho_n^2 n^{1/2} = \scriptstyle{o}(g_n^{-1})$ (for smooth case $\rho_{n,\upsilon} = \rho_n s$, $\rho_{n,\upsilon} = \rho_n^{1/2}$ for non-smooth case).
  $n^{-1/2}(s \log a_n)^{1/2} \max_{f\in\mathcal{F}'} \|f(z_t)\|_2 + n^{-1} r_\varsigma (s \log a_n)^{3/2} \|\bar{F}'(z_t)\|_q = \mathcal{O}(\rho_n)$, where $\mathcal{F}' = \{z \mapsto \psi_{jk}\{z,\beta,\widetilde{h}(x_{j(-k)})\} : (j,k) \in G, \beta \in \mathcal{B}_{jk}, \widetilde{h} \in \mathcal{H}_{jk} \cup \{h^0_{jk}\}\}$ with $\bar{F}' = \sup_{f\in\mathcal{F}'} |f|$.

References

Belloni, A. and Chernozhukov, V. (2013)
Least Squares after Model Selection in High-dimensional Sparse Models
Bernoulli, 19(2), 521-547

Belloni, A., Chernozhukov, V. and Kato, K. (2014)
Inference on Treatment Effects after Selection among High-dimensional Controls
The Review of Economic Studies, 81(2), 608-650

Belloni, A., Chernozhukov, V. and Kato, K. (2015)
Uniform Post Selection Inference for Least Absolute Deviation Regression and Other Z-estimation Problems
Biometrika, 102(1), 77-94

References

Bickel, P. J., Ritov, Y. and Tsybakov, A. B. (2009)
Simultaneous Analysis of Lasso and Dantzig Selector
The Annals of Statistics, 37(4), 1705-1732

Chernozhukov, V., Chetverikov, D. and Kato, K. (2013)
Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-dimensional Random Vectors
The Annals of Statistics, 41(6), 2786-2819

Garman, M. B. and Klass, M. J. (1980)
On the Estimation of Security Price Volatilities from Historical Data
The Journal of Business, 53(1), 67-78

References

Härdle, W. K. Chen, S., Liang, C. and Schienle, M. (2018)
Time-varying Limit Order Book Networks
IRTG 1792 Discussion Paper 2018-016

Härdle, W. K., Wang, W. and Yu, L. (2016)
TENET: Tail-event Driven Network Risk
Journal of Econometrics, 192(2), 499-513

Manresa, E. (2013)
Estimating the Structure of Social Interactions Using Panel Data
Unpublished Manuscript, CEMFI

References

Stock, J. H. and Watson, M. W. (2012)
Disentangling the Channels of the 2007-2009 Recession
Brookings Panel on Economic Activity

Van de Geer, S., Bühlmann, P., Ritov, Y. and Dezeure, R. (2014)
On Asymptotically Optimal Confidence Regions and Tests for High-dimensional Models
Annals of Statistics, 42(3), 1166-1202

Wu, W. B. (2005)
Nonlinear System Theory: Another Look at Dependence
Proceedings of the National Academy of Sciences of the United States of America, 102, 14150-14154

References

Zhang, C.-H. and Zhang, S. S. (2014)
Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76(1), 217-242

Zhang, D. and Wu, W. B. (2017)
Gaussian Approximation for High Dimensional Time Series
The Annals of Statistics, 45(5), 2895-1919

Zhang, J. L., Härdle, W. K., Chen, C. Y. and Bommes, E. (2016)
Distillation of News Flow into Analysis of Stock Reactions
Journal of Business & Economic Statistics, 34(4), 547-563