LASSO-Driven Inference in Time and Space

Victor Chernozhukov
Wolfgang K. Härdle
Chen Huang
Weining Wang

MIT Center for Statistics
Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt–Universität zu Berlin
University of St.Gallen
City, University of London
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://irtg1792.hu-berlin.de

Many High-dim Regression Equations

- Predictive graphical model or causal inference
- Dimension reduction and effective prediction with LASSO
- Joint penalty level over equations
- Temporal and cross sectional dependence (time and space)
- Individual and simultaneous inference on the coefficients

TENET

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

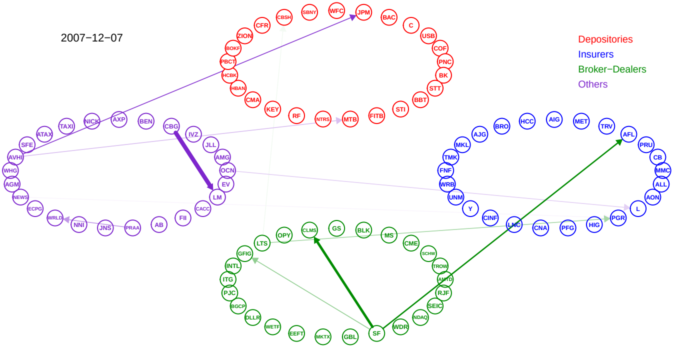


Figure 1: Financial risk network dynamics.

Financial Risk Meter (FRM)

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

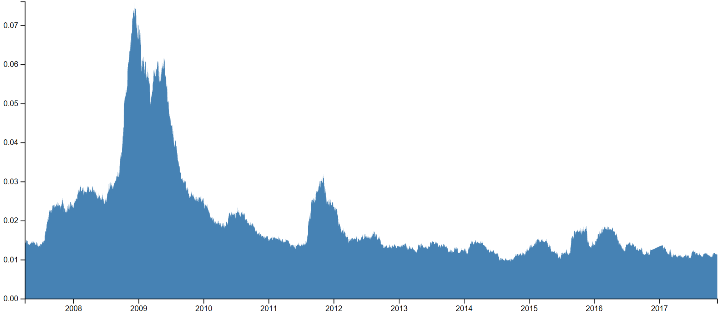


Figure 2: FRM over time `frm.wiwi.hu-berlin.de`

LOB Network

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

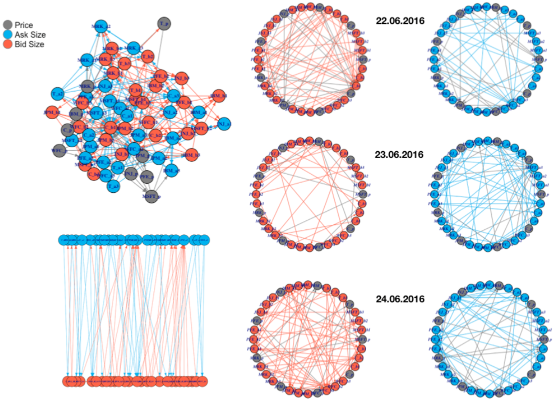


Figure 3: Plots of LOB networks from 22.06.2016-24.06.2016

Many High-dim Regression Equations

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

allowing for temporal and spatial dependency.

Example 1: Inference in high-dim system

$$Y_{j,t} = X_{k,t}\beta_{jk}^0 + X_{-k,t}^\top \beta_{j(-k)}^0 + \varepsilon_{j,t}, \quad \mathrm{E}\,\varepsilon_{j,t} X_t = 0,$$

$$X_{k,t} = X_{-k,t}^\top \delta_k^0 + \nu_{k,t}, \quad \mathrm{E}\,\nu_{k,t} X_{-k,t} = 0,$$

with the orthogonal moment conditions $\mathrm{E}\,\varepsilon_{j,t}\nu_{k,t} = 0$.

Many High-dim Regression Equations

Example 2: Large VAR

$$Y_t = \sum_{l=1}^{p} \Phi_l^0 Y_{t-l} + \varepsilon_t.$$

Example 3: 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

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

Many High-dim Regression Equations

Practical Examples: 1. Identification tests for large SVAR:

$$AY_t = BY_{t-1} + \varepsilon_t, \; Y_t = DY_{t-1} + \nu_t.$$

External instruments are required for the identification of A (Stock and Watson, 2012):

$$\begin{aligned} \mathsf{E}(\varepsilon_{j,t} z_{j,t}) &\neq 0, \\ \mathsf{E}(\varepsilon_{j',t} z_{j,t}) &= 0, \quad \text{for } j' \neq j. \end{aligned}$$

Run LASSO of

$$z_{j,t} = \widehat{\nu}_{j,t}^{\top} \delta_j + e_{j,t},$$

$\widehat{\nu}_{j,t}$ are obtained from a large VAR reduced form regression.

Many High-dim Regression Equations

Practical Examples: 2. Cross-sectional Asset Pricing:

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

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

Many High-dim Regression Equations

Practical Examples: 3. Network formation and spillover effects:

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

$Y_{j,t}$ is the log output for firm j, $D_{j,t}$ is the capital stock, $X_{j,t}$ includes other covariates (log labor, log capital etc.). One is interested in testing the spillover parameters ω_{ij}, Manresa (2013).

Effective Prediction with Sparsity

- Exact sparsity (ES) assumption
  $|\beta_j^0|_0 = s_j \le s = o(n),\ j = 1, \dots J$
- LASSO-penalized estimator of β_j^0

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

  where λ is the joint penalty, Ψ_{jk}'s are the penalty loadings.
- Prediction norm

$$\|\widehat{\beta}_j - \beta_j\|_{j,\mathrm{pr}} \stackrel{\mathrm{def}}{=} \left\{\frac{1}{n}\sum_{t=1}^{n}\left[X_{j,t}^\top(\widehat{\beta}_j - \beta_j)\right]^2\right\}^{1/2}.$$

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 (IV or double selection): Belloni et al. (2014, 2015)
  - De-sparsified (de-biased) LASSO: Zhang and Zhang (2014), Van de Geer et al. (2014)

Contributions

- A very general time dependence measure (Wu, 2005) ▸ Definition
- Aggregation of the effects over equations
- Easily implemented algo for effective estimation and inference
- Simultaneous confidence region for joint test
- 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

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

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

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

Foundations

- Theoretically, characterize the rate of $\lambda^0(1-\alpha)$ by the tail probability of S_{jk}
- Empirically, use Gaussian approx. or multiplier block bootstrap

Error Bounds for the Prediction Norm

Theorem 1

Suppose the uniform RE condition [A2] *holds with probability* $1 - o(1)$*, then with ES:*

$$|\widehat{\beta}_j - \beta_j^0|_{j,\mathrm{pr}} \leq C\lambda^0(1-\alpha)\frac{\sqrt{s}}{n}\max_k \Psi_{jk}, \quad \textit{for all } j = 1, \ldots, J, \qquad (1)$$

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

Nagaev Type of Inequality

Theorem 2

Under [A1] and [A3], we have

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

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 with ES:*

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

with probability $1-\alpha-\mathcal{o}(1)$. $x_n \lesssim y_n$ *means there exists constant* $C > 0$ *such that* $x_n/y_n \leq C$.

Empirical Choices of λ

- Gaussian Approximation:
  $Q(1-\alpha) \stackrel{\text{def}}{=} 2c\sqrt{n}\Phi^{-1}\{1-\alpha/(2JK)\}$
- Multiplier Bootstrap: $\Lambda(1-\alpha)$ 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:

$$\rho(X,Y) = \sup_{r\in\mathbb{R}} \Big| \mathsf{P}(|X|_\infty \le r) - \mathsf{P}(|Y|_\infty \le r) \Big|.$$

Theorem 4

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

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

where $\widetilde{\mathcal{Z}} \sim \mathrm{N}(0, \Sigma_{\widetilde{\mathcal{X}}})$, *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:

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

for sufficiently large n.

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

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

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* $\widetilde{\mathcal{Z}}^{[B]} \stackrel{\text{def}}{=} \operatorname{vec}\{(Z_{jk}^{[B]})_{jk}\}$, *and* $\widetilde{\Psi} \stackrel{\text{def}}{=} \operatorname{vec}\{(\Psi_{jk})_{jk}\}$, *then*

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

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

Single 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: LAD regression of $Y_{j,t} - X^\top_{j(-k),t}\widehat{\beta}^{[1]}_{j(-k)}$ on $X_{jk,t}$ using $\widehat{v}_{jk,t}$ as IV, finally achieve $\widehat{\beta}^{[2]}_{jk}$ and $\widehat{\sigma}_{jk}$

Orthogonality

Uniform Bahadur Representation

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*

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

with probability $1 - \mathcal{O}(1)$, *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:
  - $T^*_{jk} = \frac{1}{\sqrt{n}} \sum_{i=1}^{l_n} e_{j,i} \sum_{l=(i-1)b_n+1}^{ib_n} \widehat{\zeta}_{jk,l}$, $T_{jk} = \frac{\sqrt{n}(\widehat{\beta}^{[2]}_{jk} - \beta^0_{jk})}{\widehat{\sigma}_{jk}}$
  - $[\widehat{\beta}^{[2]}_{jk} - \widehat{\sigma}_{jk} n^{-1/2} q^*_{(1-\alpha/2)}, \widehat{\beta}^{[2]}_{jk} + \widehat{\sigma}_{jk} n^{-1/2} q^*_{(1-\alpha/2)}]$, $q^*_{(1-\alpha/2)}$ is the $(1-\alpha/2)$ quantile of $|T^*_{jk}|$

Confidence Region (CR) for Simult. Inference

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

Consistency of the Bootstrapped CR

Corollary 8

Under conditions*, we have*

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

with probability $1 - o(1)$*.*

Predictive Performance

DGP 1:

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

- $X_{j,t} \in \mathbb{R}^K \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \Sigma)$, with $\Sigma_{k_1,k_2} = \rho^{|k_1-k_2|}$, $\rho = 0.5$, $\varepsilon_{j,t} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, 1)$
- divide $\{1, \ldots, K\}$ evenly into blocks with fixed block size 5, $\beta_{jk}^0 = 10$ if k and j belong to one block and 0 otherwise
- $n = 100$, take 500 bootstrap replications

Predictive Performance

	$J = K = 50$	$J = K = 100$	$J = K = 150$
		Prediction norm	
Mean	0.89	0.84	0.79
Median	0.91	0.87	0.84
		Euclidian norm	
Mean	0.90	0.85	0.79
Median	0.89	0.85	0.81

Table 1: Prediction norm and Euclidean norm ratios (overall λ relative to single λ_j's, mean or median over equations). Results are averaged over 100 repeats of simulations.

Predictive Performance

DGP 2:

$$Y_t = \Phi^0 Y_{t-1} + \varepsilon_t, \quad Y_t \in \mathbb{R}^K \quad t = 1, \ldots, n,$$

- Φ^0 has a block diagonal structure where the blocks are 5×5 matrices with all entries in each block equal ϕ
- $\varepsilon_t \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, \mathbf{I}_K)$
- $n = 100$, take 500 bootstrap replications

Predictive Performance

		$\phi = 0.05$			$\phi = 0.15$		
		$K = 50$	100	150	$K = 50$	100	150
		Prediction norm					
$b_n = 4$	Mean	0.88	0.85	0.76	1.05	1.04	1.02
	Median	0.97	0.96	0.95	1.05	1.04	1.02
$b_n = 10$	Mean	0.89	0.84	0.75	1.08	1.06	1.04
	Median	0.97	0.96	0.95	1.07	1.05	1.04
$b_n = 20$	Mean	0.89	0.85	0.75	1.10	1.06	1.05
	Median	0.97	0.96	0.95	1.09	1.06	1.05
$b_n = 25$	Mean	0.89	0.85	0.74	1.10	1.07	1.04
	Median	0.97	0.97	0.95	1.09	1.06	1.04
		Euclidean norm					
$b_n = 4$	Mean	0.84	0.79	0.57	0.99	0.96	0.94
	Median	1	1	1	1.00	1	1
$b_n = 10$	Mean	0.84	0.79	0.56	0.98	0.95	0.92
	Median	1	1	1	1.00	1	1
$b_n = 20$	Mean	0.85	0.80	0.57	0.96	0.94	0.92
	Median	1	1	1	1.00	1	1
$b_n = 25$	Mean	0.85	0.80	0.53	0.96	0.94	0.92
	Median	1	1	1	1.00	1	1

Table 2: Prediction norm and Euclidean norm ratios (overall λ relative to single λ_j's, mean or median over equations).

Inference Performance

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

- $\alpha_j^0 = \alpha^0$ for $j = 1, \ldots, J$
- $X_t \in \mathbb{R}^K$ follow VAR(1) with normal errors, $\varepsilon_{j,t}$ follow AR(1), $v_{j,t} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0,1)$, $n = 100$
- Block diagonal structure in $\{\beta_{jk}^0\}$ and $\{\gamma_{jk}^0\}$:
  - divide $\{1, \ldots, K\}$ evenly into blocks with fixed block size 5
  - if k and j belong to one block
    $\beta_{jk}^0 = 0.5/(k - \lfloor \frac{k}{5} \rfloor \times 5), \theta_{jk}^0 = 0.25(k - \lfloor \frac{k}{5} \rfloor \times 5)$

Power Curve

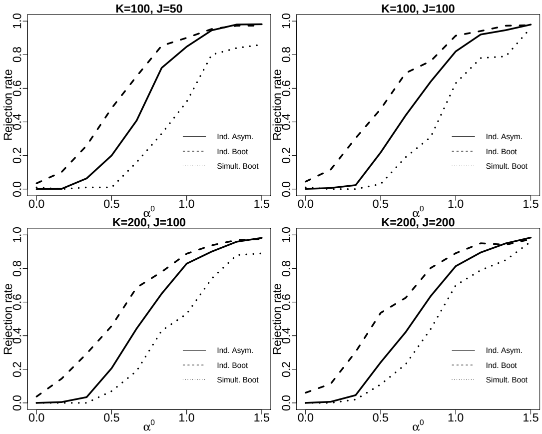


Figure 4: Average rejection rate of $H_0^j : \alpha_j^0 = 0$ over j for individ. test and $H_0 : \alpha_1^0 = \cdots = \alpha_J^0 = 0$ for simult. test (nominal level $= 0.05$).

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

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

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:

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

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

- Response variables

$$r_{j,t} = \log(P_{j,t}^{C}) - \log(P_{j,t}^{O}),$$

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

  $u_{j,t} = \log(P_{j,t}^{H}) - \log(P_{j,t}^{O})$, $d_{j,t} = \log(P_{j,t}^{L}) - \log(P_{j,t}^{O})$, with $P_{j,t}^{H}$, $P_{j,t}^{L}$, $P_{j,t}^{O}$, and $P_{j,t}^{C}$ are the highest, lowest, opening and closing prices.
  Garman and Klass (1980)

Graphical network - Individual Inference

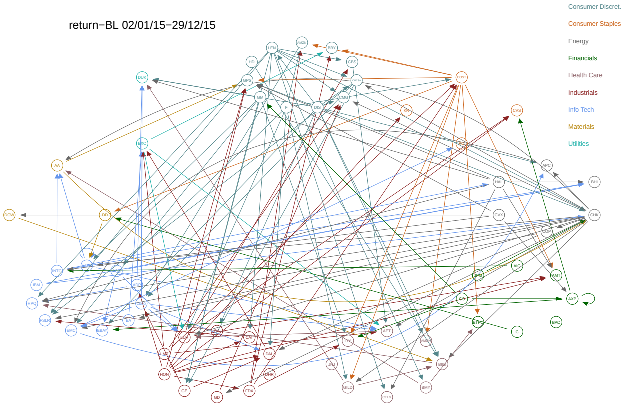


Figure 5: Graphical network among individual stocks (return - BL)

Graphical network - Individual Inference

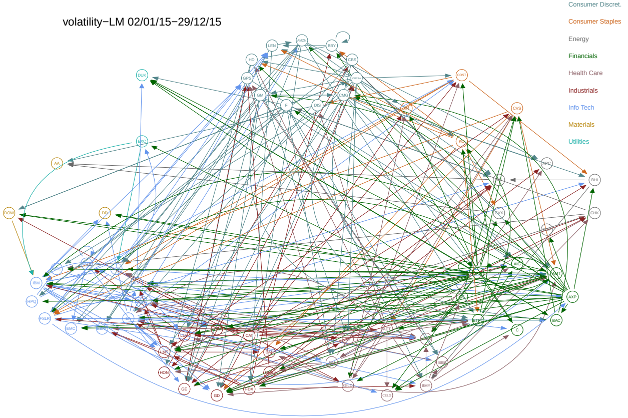


Figure 6: Graphical network among individual stocks (volatility - LM)

Graphical network - Individual Inference

Example: dependency between two stocks

- textual sentiment effect on stock volatility $H_0^{jk} : \beta_{jk} = 0$
- directional edge from "BBY" (Best Buy) to "LEN" (Lennar)
- self effect of "BBY"

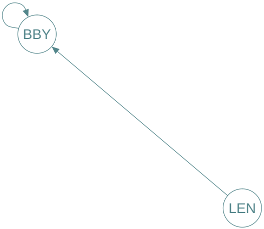


Figure 7: Dependency between BBY and LEN (volatility - LM)

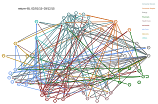
Graphical network - Simultaneous Inference

- 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:
  - denser spillover effects among individ. stocks when the outcome is volatility
  - sector connections (return): industrials$\rightarrow$consumer discretionary, health care$\rightarrow$utilities
  - sector connections (volatility): utilities$\rightarrow$IT, financials$\rightarrow$consumer staples

LASSO-Driven Inference in Time and Space

Victor Chernozhukov
Wolfgang K. Härdle
Chen Huang
Weining Wang

MIT Center for Statistics
Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt–Universität zu Berlin
University of St.Gallen
City, University of London
http://lvb.wiwi.hu-berlin.de
http://www.case.hu-berlin.de
http://irtg1792.hu-berlin.de



Single Equation LASSO Performance

Theorem 1 of Belloni and Chernozhukov (2013)

Suppose the RE 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* $\widetilde{\beta}_j$ *from single equation LASSO satisfy:*

$$|\widetilde{\beta}_j - \beta_j^0|_{j,pr} \leq (1 + 1/c)\frac{\lambda_j \sqrt{s_j}}{n\kappa_j(\overline{c})} \max_{1\leq k\leq K} \Psi_{jk}. \qquad (6)$$

▸ Ideal λ

Measure of Dependence [by Wu (2005)]

[A1] Assume $X_{jk,t} = g_{jk}(\ldots, \xi_{t-1}, \xi_t)$, 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 $X^*_{jk,t} = g_{jk}(\ldots, \xi_0^*, \ldots, \xi_t)$
- Functional dependence measure $\delta_{q,j,k,t} \stackrel{\text{def}}{=} \|X_{jk,t} - X^*_{jk,t}\|_q$, $q \geq 1$, which measures the dependency 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 Dependency

Example: AR(1) process

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

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

Nagaev Inequality

Restricted Eigenvalue (RE) Condition

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

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

holds uniformly over $j = 1, \ldots, J$, where $T_j \overset{\text{def}}{=} \{k : \beta^0_{jk} \neq 0\}$ 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] $\varepsilon_{j,t}$ and $X_{jk,t}$ have finite moments up to the $q-$th order.

▸ 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 $\mathcal{X}_t \stackrel{\text{def}}{=} \text{vec}\{(X_{jk,t})_{jk}\}$

- $\Phi_{q,\varsigma} = \max_{j,k} 2\|X_{jk,\cdot}\|_{q,\varsigma}\|\varepsilon_{j,\cdot}\|_{q,\varsigma}$
- $\Gamma_{q,\varsigma} = 2(\sum_j \|\varepsilon_{j,\cdot}\|_{q,\varsigma}^{q/2})^{2/q}(\sum_{j,k} \|X_{jk,\cdot}\|_{q,\varsigma}^{q/2})^{2/q}$
- $\Theta_{q,\varsigma} = \Gamma_{q,\varsigma} \wedge \{\||\mathcal{X}_\cdot|_\infty\|_{q,\varsigma}\|\varepsilon_{j,\cdot}\|_{q,\varsigma}(\log KJ)^{3/2}\}$, 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 for the target coefficient β^0_{jk}

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

which has the orthogonality property

$$\frac{\partial}{\partial \beta_{j(-k)}} \mathsf{E}(\varepsilon_{j,t} v_{jk,t}) \Big|_{\beta_{j(-k)} = \beta^0_{j(-k)}} = 0.$$

▸ Estimation

Conditions for Theorem 7

- The dependence adjusted sub-Gaussian norm $\|v_{jk,\cdot}\|_{\psi_{1/2}} = \sup_{q\geq 2} q^{-1/2} \sum_{t=0}^{\infty} \|\operatorname{E}(v_{jk,t}|\mathcal{F}_0) - \operatorname{E}(v_{jk,t}|\mathcal{F}_{-1})\|_q < \infty$.
- **Properties of ψ_{jk}:** the map $(\beta, h) \mapsto \operatorname{E}\{\psi_{jk}(Z_{j,t}, \beta, h)|X_{j(-k),t}\}$ is twice continuously differentiable, and for every $\vartheta \in \{\beta, h_1, \ldots, h_M\}$, $\operatorname{E}[\sup_{\beta\in\mathcal{B}_{jk}} |\partial_\vartheta \operatorname{E}\{\psi_{jk}(Z_{j,t}, \beta, h)|X_{j(-k),t}\}|^2] \leq C_1$; 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 $\vartheta, \vartheta' \in \{\beta, h_1, \ldots, h_M\}$, $\sup_{(\beta,h)\in\mathcal{B}_{jk}\times\mathcal{T}_{jk}(X_{j(-k),t})} |\partial_\vartheta\partial_{\vartheta'} \operatorname{E}\{\psi_{jk}(Z_{j,t}, \beta, h)|X_{j(-k),t}\}| \leq L_{1n}$, and for every $\beta, \beta' \in \mathcal{B}_{jk}$, $h, h' \in \mathcal{T}_{jk}(X_{j(-k),t})$, $\operatorname{E}[\{\psi_{jk}(Z_{j,t}, \beta, h) - \psi_{jk}(Z_{j,t}, \beta', h')\}^2|X_{j(-k),t}] \leq L_{2n}(|\beta - \beta'|^\nu + |h - h'|_2^\nu)$.

Conditions for Theorem 7

- **Identifiability:**
  $2|\operatorname{E}[\psi_{jk}\{Z_{j,t},\beta,h^0_{jk}(X_{j(-k),t})\}]| \geq |\phi_{jk}(\beta-\beta^0_{jk})| \wedge c_1$ holds for all $\beta \in \mathcal{B}_{jk}$, and $|\phi_{jk}| \geq c_1$.
- **Dimension growth rates:** there exist sequences of constants $\rho_n \downarrow 0, \delta_n \downarrow 0$ such that
  $\rho_n^{\nu/2}(L_{2n} s \log a_n)^{1/2} + n^{-1/2} r_\varsigma (s \log a_n)^2 = o(g_n^{-1})$ and
  $n^{-1/2}(s \log a_n)^{1/2} + n^{-1} r_\varsigma (s \log a_n)^2 = \mathcal{O}(\rho_n)$, where $r_\varsigma = n^{1/q}$ for $\varsigma > 1/2 - 1/q$ and $r_\varsigma = n^{1/2-\varsigma}$ for $\varsigma < 1/2 - 1/q$,
  $a_n \stackrel{\text{def}}{=} \max(JK, n, e)$.
- The conditional density of Y_j given X_j is bounded and continuously differentiable.

Conditions for Theorem 7

- **Properties of the nuisance function:** with probability $1 - o(1)$, $\widehat{h}_{jk} \in \mathcal{H}_{jk}$, 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 $\widetilde{h}_{jk,m}(X_{j(-k),t}) = X_{j(-k),t}^{\top} \theta_{jk,m}$, $\|\theta_{jk,m}\|_0 \le s$, $\widetilde{h}_{jk,m} \in \mathcal{T}_{jk,m}$ and $\mathsf{E}[\{\widetilde{h}_{jk,m}(X_{j(-k),t}) - h^0_{jk,m}(X_{j(-k),t})\}^2] \le C_1 s(\log a_n)/n$.
- 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 $1 - o(1)$, $[\beta^0_{jk} \pm c_1 n^{-1/2} \log^2 a_n] \subset \widehat{\mathcal{B}}_{jk} \subset \mathcal{B}_{jk}$.

▸ Theorem 7

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