High-dimensional Granger causality tests with an application to VIX and news

Andrii Babii[1] Eric Ghysels[1] Jonas Striaukas[2]

[1]UNC Chapel Hill

[2]UC Louvain

2021 SoFiE ML Conference

Motivation

1. High-dimensional Granger causality tests with sparse-group LASSO (sg-LASSO).
2. More generally: valid HAC-based post model selection inference.
3. Empirical example: Granger causal relationship between VIX and financial news.

Related literature

LASSO

Chernozhukov, Härdle, Huang, and Wang (2021), Feng, Giglio, and Xiu (2020), Belloni, Chernozhukov, and Hansen (2014), van de Geer, Bühlmann, Ritov, and Dezeure (2014), Simon, Friedman, Hastie, and Tibshirani (2013), Bickel, Ritov, and Tsybakov (2009), Tibshirani (1996).

Granger causality

Ghysels, Hill, and Motegi (2020), Granger and Newbold (1986), Granger (1969).

(low-dimensional) HAC-based time series inference

Andrews (1991), Newey and West (1987), Gallant (1987), Parzen (1957), Bartlett (1948), and Daniell (1946).

Contributions

New methodology

Structured high-dimensional Granger causality tests via sparse-group LASSO regularization (sg-LASSO).

New concentration inequality

Fuk-Nagaev inequalities for heavy-tailed τ-mixing processes.

Inference for high-dimensional time series regressions

Debiased CLT with explicit bias correction, and convergence rate of HAC estimator based on sg-LASSO residuals.

Granger causality between VIX and financial news:

The topic of financial crisis Granger causes VIX (and not the other way round).

Related work:

- Andrii Babii, Eric Ghysels, and Jonas Striaukas (2021) **Machine learning time series regressions with an application to nowcasting**, Journal of Business & Economic Statistics (forthcoming).

- Andrii Babii, Eric Ghysels, and Jonas Striaukas (2020) **Machine learning panel data regressions with an application to nowcasting price earnings ratios**, arXiv:2008.03600

Table of Contents

1. Granger causality
2. Sparse-group LASSO regularization
3. New concentration inequality
4. Debiased CLT and HAC-based inference
5. VIX and financial news

Granger causality, Granger (1969)

$(z_t)_{t\in\mathbb{Z}}$ does not Granger cause $(y_t)_{t\in\mathbb{Z}}$ if

$$P(y_{t+1}|x_t) = P(y_{t+1}|x_t - z_t),$$

where

- x_t is all the information in the universe available at time t;
- $x_t - z_t$ all this information apart from z_t;
- $P(y_{t+1}|x_t)$ is the least-squares projection of y_{t+1} on x_t.
- Can be generalized to conditional expectations/distributions.

Marginal improvement in projections due to the series $(z_t)_{t\in\mathbb{Z}}$.

Granger causality: a testable implication

Linear projection

$$y_{t+1} = \mathbf{z}_t^\top \alpha + \underbrace{\mathbf{w}_t^\top \beta}_{\text{controls}} + u_{t+1},$$

1. controls $\mathbf{w}_t$ = "all the information at time t" (high-dimensional).
2. where $\mathbf{z}_t = (z_t, z_{t-1}, \dots, z_{t-K})$ and $\alpha \in \mathbf{R}^{K+1}$.
3. Granger causality test:

$$H_0: \ \alpha = 0 \qquad vs. \qquad H_1: \alpha \neq 0.$$

High-dimensional linear projection

1. More general linear projection model

$$y_{t+1} = \sum_{j=1}^{\infty} \beta_j x_{t,j} + u_{t+1}, \qquad \mathbb{E}[u_{t+1}x_{t,j}] = 0, \ \forall j \geq 1.$$

2. Test

$$H_0 : \ \beta_G = 0 \qquad \text{vs.} \qquad H_1 : \ \beta_G \neq 0,$$

where $\beta_G = \{\beta_j : j \in G\}$ and $G \subset \{1, 2, \ldots, \}$ is low-dimensional.

Sparse-group structures

1. Example 1: news grouped by topics.
2. Example 2: Lags of a single covariate

$$\sum_{j=1}^{m} \beta_j x_{t-j}$$

- Dimensionality reduction: $\beta_j = \omega(j/m;\theta)$ is MIDAS weights;
- Sparse approximation is some dictionary $w(u) = (w_1(u), \dots, w_L(u))^\top$

$$\omega(u;\theta) \approx w(u)^\top \theta, \qquad \forall u \in [0,1],$$

  where and $\theta \in \mathbf{R}^L$ is approximately sparse.
- Example of dictionary: Legendre polynomials.

Structured sparsity

1. High-dimensional model selection problems:
   - covariate selection and learning the dictionary size;
   - selection of topics and selection of news within a topic.
2. Hierarchical model selection problem: sparse-group LASSO regularization.

Sparse-group LASSO regression/projection

1. sg-LASSO solves

$$\text{argmin}_{b\in\mathbf{R}^p}\|\mathbf{y}-\mathbf{X}b\|_T^2+2\lambda\Omega_\gamma(b)$$

   with regularization parameter $\lambda \geq 0$ and penalty

$$\Omega_\gamma(b)=\gamma|b|_1+(1-\gamma)\|b\|_{2,1},$$

   where $\gamma \in [0,1]$ is a relative weight, $|.|_1$ is ℓ_1 norm, $\|.\|_{2,1}$ is the group-LASSO norm.
2. Convex optimization problem.
3. Easy and fast to compute via proximal gradient descent.

sg-LASSO and dependent data

Weak dependence coefficients

1. α-mixing processes do not apply to the linear projection model with autoregressive lags.
2. τ-mixing coefficient of a stationary process $\xi_t \in \mathbf{R}$ with history $\mathcal{F}_t$

$$\tau_k = \sup_{j \geq 1} \frac{1}{j} \sup_{t+k \leq t_1 < \cdots < t_j} \tau(\mathcal{F}_t, (\xi_{t_1}, \ldots, \xi_{t_j})),$$

   where $\tau(\mathcal{F}, \zeta) = \mathbb{E}\left|\sup_{f \in \mathrm{Lip}_1} |\mathbb{E}(f(\zeta)|\mathcal{F}) - \mathbb{E}(f(\zeta))|\right|$.
3. Dedecker-Prieur coupling: there exists measurable $\zeta^* =_d \zeta$ with $\zeta^* \perp\!\!\!\perp \mathcal{F}$ such that

$$\mathbb{E}|\zeta^* - \zeta| = \tau(\mathcal{F}, \zeta).$$

Comparison to mixingale and α-mixing coefficient

1. The process is called τ-mixing if $\tau_k \downarrow 0$ as $k \uparrow \infty$.
2. The τ-mixing coefficient satisfies

$$\gamma_k \leq \tau_k \leq 2 \int_0^{2\alpha_k} Q(u)\mathrm{d}u$$

- γ_k is mixingale coefficient;
- α_k is α-mixing coefficient;
- Q is a tail function – the generalized inverse of $x \mapsto \Pr(|\xi_0| > u)$.

3. α-mixing processes are also τ-mixing.

sg-LASSO and dependent data

New Fuk-Nagaev type inequality

Theorem

Suppose that $(\xi_t)_{t\in\mathbb{Z}}$ is a centered stochastic process in $\mathbf{R}^p$ such that each coordinate $j \geq 1$ has

(i) *tails:* $\sup_t \mathbb{E}|\xi_{t,j}|^q = O(1)$ *with* $q > 2$;

(ii) *τ-mixing coefficients:* $\tau_k \leq dk^{-a}$ *with* $a > (q-1)/(q-2)$.

Then there exist $c_1, c_2 > 0$ *such that for every* $u > 0$

$$\Pr\left(\max_{1\leq j\leq p}\left|\frac{1}{T}\sum_{t=1}^{T}\xi_{t,j}\right| > u\right) \leq \frac{c_1 p}{u^{\kappa}T^{\kappa-1}} + 4pe^{-c_2 T u^2}.$$

sg-LASSO and dependent data

New Fuk-Nagaev inequality: comments

$$\Pr\left(\max_{1\leq j\leq p}\left|\frac{1}{T}\sum_{t=1}^{T}\xi_{t,j}\right| > u\right) \leq \frac{c_1 p}{u^{\kappa}T^{\kappa-1}} + 4pe^{-c_2 Tu^2}$$

1. polynomial + exponential bound on tail bound of high-dimensional means, cf. Fuk and Nagaev (1971) for independent data;
2. Intuition: exponential tail due to the CLT and polynomial tail due to heavy-tails.
3. κ measures persistence and tails;
4. sharp, unlike Markov's + Rosenthal's/Marcinkiewicz-Zygmund inequalities.

Debiased CLT

Theorem

Suppose that $p^{2/(\kappa-2)}/T \to 0$*. Then under mild conditions*

$$\sqrt{T}(\hat{\beta}_G + b_G - \beta_G) \xrightarrow{d} N(0, \Xi_G).$$

Remarks:

1. Can handle $p > T$, unlike OLS/ridge regression.
2. b_G is estimated "bias" correction.
3. long run variance: $\Xi_G = \lim_{T\to\infty} \mathrm{Var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_t \Theta_G x_t\right)$.
4. Application: Wald test of Granger causality.

HAC estimator

1. Long run variance

$$\Xi_G = \sum_{k \in \mathbb{Z}} \Gamma_k$$

2. HAC estimator

$$\hat{\Xi}_G = \sum_{|k|<T} K\left(\frac{k}{M_T}\right)\hat{\Gamma}_k$$

with bandwidth $M_T \uparrow \infty$ and kernel $K : \mathbf{R} \to [-1, 1]$.

HAC estimator: MSE

Theorem

Under some conditions

$$\|\hat{\Xi}_G - \Xi_G\| = O_P\left(M_T\left(\frac{p^{1/\kappa}}{T^{1-1/\kappa}} \vee \sqrt{\frac{\log p}{T}} + \frac{p^{2/\kappa}}{T^{2-3/\kappa}} + \frac{p^{5/\kappa}}{T^{4-5/\kappa}}\right) + M_T^{-2}\right)$$

Remarks:

- The optimal bandwidth should scale with dimension p and κ (persistence and tails).
- Intuition: need to estimate regression errors in the long-run variance $\implies$ cannot debias/partial-out the high-dimensional part.
- The usual rules-of-thumbs do not apply.

Empirical illustration: VIX and financial news

Questions

Questions:

1. Which news topics from the Wall Street Journal Granger cause future VIX?
2. Are news topics Granger caused by VIX (reverse Granger causality)?

Methodology

Structured sg-LASSO (daily VIX lags + group news in topics)

$$\underbrace{y_{t+1}}_{\text{monthly VIX}} = \underbrace{\psi(L^{1/m};\theta)y_t}_{\text{daily VIX}} + \underbrace{\sum_{k=1}^{K}\beta_k x_{t,k}}_{\text{monthly news}} + u_{t+1}.$$

- monthly VIX = VIX at the end of month $t+1$ from FRED;
- daily VIX = 22 daily lags of VIX starting at the end of month t aggregated with Legendre polynomial;
- monthly news = 180 news series from Bybee, Kelly, Manela, and Xiu (2020).

Results I: financial news Granger cause VIX

Variable $\backslash M_T$	20	40	60	20	40	60
		Parzen			Quadratic Spectral	
			1% significance			
Daily VIX lags	0.000	0.000	0.001	0.000	0.001	0.002
Financial crisis	0.005	0.002	0.001	0.002	0.001	0.000
			5% significance			
Recession	0.011	0.008	0.009	0.008	0.008	0.013
Aerospace/defense	0.014	0.014	0.017	0.012	0.018	0.027

Table: VIX Granger causality results. We report p-values of series that are significant at 1% and 5% significance level for a range of M_T values and two kernel functions.

Results II: Reverse causality

$$x_{t+1,j} = \psi(L^{1/m};\theta)y_t + \sum_{k\neq j}\beta_k x_{t,k} + u_t,$$

where $x_{t,j}$ is Financial crisis news series.

Variable $\backslash M_T$	20	40	60	20	40	60
	Parzen			Quadratic Spectral		
Daily VIX lags	0.050	0.071	0.091	0.060	0.086	0.129

Table: Bi-directional Granger causality results. We report p-values for a range of M_T values and both kernel functions.

Concluding remarks

1. Debiased inference for structured high-dimensional time series regressions:
   - explicit bias correction;
   - valid HAC estimation based on sg-LASSO residuals.
2. ML methods may be useful for Granger causality when $p > T$.
3. Theory shows that we can handle heavy-tailed dependent data.
4. Financial crisis news seem to Granger cause VIX.
5. `midasml` package on CRAN.

Thank you!

ababii.github.io