Inference for High-Dimensional Local Projection

Jiti Gao[1], Fei Liu[2] and Bin Peng[1]

[1]Monash University and [2]The University of Bath

Venice Time Series Workshop on April 22–23, 2026

Overview

Literature of Local Projection (LP)

References:

Jordá (2005);
Montiel Olea and Plagborg-Møller (2021);
Plagborg-Møller and Wolf (2021);
Gonçalves et al. (2024);
Inoue et al. (2024); etc

See Jordá and Taylor (2025) for a comprehensive review.

Given the extensive literature, we are selective. *We highlight key references throughout the presentation.*

What is LP about ?

Suppose that we observe

$$\{\mathbf{x}_t = (x_{1t}, \ldots, x_{Nt})^\top \mid t \in [T]\},$$

which can be represented by a high dimensional moving average infinity (HDMA(∞)) process:

$$\mathbf{x}_t = \sum_{\ell=0}^{\infty} \mathbf{B}_\ell \boldsymbol{\varepsilon}_{t-\ell}, \tag{1}$$

where $\mathbf{B}_\ell = \{b_{\ell,ij}\}_{N\times N}$ and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})^\top$ with ε_{it} being i.i.d. over both dimensions.

Key interest: Impulse response (i.e., $\mathbf{B}_\ell$)

Approach: LP (i.e., multiple OLS regressions)

Beveridge–Nelson (BN) Decomposition

The corresponding BN decomposition is given as follows:

$$\mathbf{x}_t = \mathbb{B}(1)\,\boldsymbol{\varepsilon}_t + \widetilde{\mathbb{B}}(L)\,\boldsymbol{\varepsilon}_{t-1} - \widetilde{\mathbb{B}}(L)\,\boldsymbol{\varepsilon}_t,$$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t = \frac{1}{T}\sum_{t=1}^{T}\mathbb{B}(1)\,\boldsymbol{\varepsilon}_t + O_P(T^{-1}),$$

where $\mathbb{B}(1) = \sum_{j=0}^{\infty}\mathbf{B}_j$ and $\widetilde{\mathbb{B}}(L) = \sum_{j=0}^{\infty}\widetilde{\mathbb{B}}_j L^j$ with $\widetilde{\mathbb{B}}_j = \sum_{k=j+1}^{\infty}\mathbf{B}_k$.

Under certain conditions, it can be shown that the HDMA(∞) representation of (1) covers many commonly used time–series models:

- HD–AR(1) model of the form:

$$\mathbf{x}_t = \mathbf{\Phi}\,\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}^j\,\boldsymbol{\varepsilon}_{t-j}, \tag{2}$$

  and many general HD–ARMA models.

- Model (2) implies that for each given $h \geq 1$,

$$\begin{aligned} \mathbf{x}_{t+h} &= \mathbf{\Phi}\,\mathbf{x}_{t+h} + \boldsymbol{\varepsilon}_{t+h} \\ &= \mathbf{\Phi}\,\left(\mathbf{\Phi}\,\mathbf{x}_{t+h-1} + \boldsymbol{\varepsilon}_{t+h-1}\right) + \boldsymbol{\varepsilon}_{t+h} = \cdots \\ &= \mathbf{\Phi}^h\,\mathbf{x}_t + \sum_{k=0}^{h-1} \mathbf{\Phi}^k\,\boldsymbol{\varepsilon}_{t+h-k} = \mathbf{\Phi}^h\,\mathbf{x}_t + \mathbf{u}_{t,h}, \end{aligned} \tag{3}$$

  where $\mathbf{u}_{t,h} = \sum_{k=0}^{h-1} \mathbf{\Phi}^k\,\boldsymbol{\varepsilon}_{t+h-k}$.

- Under $\lambda_{\max}(\mathbf{\Phi}) < 1$ and $\mathrm{E}\|\mathbf{x}_s\|^2 < \infty$, model (3) implies that as $h \to \infty$,

$$\mathbf{x}_s = \mathbf{\Phi}^h \mathbf{x}_{s-h} + \sum_{k=0}^{h-1} \mathbf{\Phi}^k \boldsymbol{\varepsilon}_{s-k} \to \sum_{k=0}^{\infty} \mathbf{\Phi}^k \boldsymbol{\varepsilon}_{s-k}, \tag{4}$$

  where $\lambda_{\max}(A)$ denotes the largest eigenvalue of matrix A.

- It is therefore that the LP representation of the form:

$$\mathbf{x}_{t+h} = \mathbf{\Phi}^h \mathbf{x}_t + \sum_{k=0}^{h-1} \mathbf{\Phi}^k \boldsymbol{\varepsilon}_{t+h-k} = \mathbf{\Phi}^h \mathbf{x}_t + \mathbf{u}_{t,h} \tag{5}$$

  may be considered as a long–horizon approximation of the MA(∞) process of the form (4).

Impulse Response

Definition 1

For $\forall h \geq 1$ and $\forall j \in [N]$, let

$$\mathbf{IR}_{h,j} = E[\mathbf{x}_{t+h} \mid \boldsymbol{\varepsilon}_{t|j}(1), \boldsymbol{\varepsilon}_{t-1}, \ldots] - E[\mathbf{x}_{t+h} \mid \boldsymbol{\varepsilon}_{t|j}(0), \boldsymbol{\varepsilon}_{t-1}, \ldots], \tag{6}$$

where $\boldsymbol{\varepsilon}_{t|j}(a) \coloneqq (\varepsilon_{1t}, \ldots, \varepsilon_{j-1,t}, a, \varepsilon_{j+1,t}, \ldots, \varepsilon_{Nt})^\top$.

(1) and (6) jointly infer

$$\mathbf{IR}_{h,j} = \mathbf{B}_{h,j}, \tag{7}$$

where $\mathbf{B}_{h,j}$ is the j^{th} column of $\mathbf{B}_h$.

Some naive questions are

1. how many components of $\mathbf{B}_h$ can be recovered ? (i.e., long–horizon analysis)
2. under what conditions ?

Example 1

If $\{\mathbf{x}_t\}$ are generated by a VAR(1) process $\mathbf{x}_t = \mathbf{a}_1\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t$, the corresponding h–step ahead forecasting model is

$$\mathbf{x}_{t+h} = \mathbf{a}_1^h \mathbf{x}_t + \mathbf{u}_{t,h}, \tag{8}$$

where $\mathbf{u}_{t,h} \coloneqq \mathbf{a}_1^{h-1}\boldsymbol{\varepsilon}_{t+1} + \cdots + \mathbf{a}_1\boldsymbol{\varepsilon}_{t+h-1} + \boldsymbol{\varepsilon}_{t+h}$. Therefore,

(i) $\mathbf{u}_{t,h}$ is $(h-1)$–dependent (i.e., $\mathbf{u}_{t,h}$ and $\mathbf{u}_{s,h}$ are independent for $|t-s| \geq h$);

(ii) for $\forall t$, $\{\mathbf{u}_{t,h} \mid h \geq 1\}$ and $\{\mathbf{x}_s \mid s \leq t\}$ are independent of each other.

Example 1 (Revised I)

One can include (conditional) heteroskedasticity. Say, ARCH structure:

$$\mathbf{x}_t = \mathbf{a}_1\mathbf{x}_{t-1} + \tilde{\boldsymbol{\varepsilon}}_t, \quad \tilde{\boldsymbol{\varepsilon}}_t = \boldsymbol{\sigma}_t^{1/2}\boldsymbol{\varepsilon}_t, \quad \text{and} \quad \boldsymbol{\sigma}_t = \mathbf{b}_0 + \mathbf{b}_1^\top\boldsymbol{\sigma}_{t-1}\mathbf{b}_1,$$

where $\mathbf{b}_0$ and $\mathbf{b}_1$ need to fulfil certain conditions. In this case,

$$\mathbf{x}_{t+h} = \mathbf{a}_1^h\mathbf{x}_t + \tilde{\mathbf{u}}_{t,h}, \tag{9}$$

where $\tilde{\mathbf{u}}_{t,h} \coloneqq (\mathbf{a}_1^{h-1}\boldsymbol{\sigma}_{t+1}^{1/2}, \ldots, \mathbf{a}_1\boldsymbol{\sigma}_{t+h-1}^{1/2}, \boldsymbol{\sigma}_{t+h}^{1/2})\mathbf{v}_{t,h}$ with $\mathbf{v}_{t,h} \coloneqq (\boldsymbol{\varepsilon}_{t+1}^\top, \ldots, \boldsymbol{\varepsilon}_{t+h}^\top)^\top$.
Therefore, we have

(i) $\mathbf{v}_{t,h}$ is $(h-1)$–dependent;

(ii) for $\forall t$, $\{\mathbf{v}_{t,h} \mid h \geq 1\}$ and $\{\mathbf{x}_s \mid s \leq t\}$ are independent of each other.

Example 1 (Revised II)

Partition $\mathbf{x}_t$ of Example 1 as follows:

$$\begin{pmatrix} \mathbf{x}_t^* \\ \mathbf{x}_t^\dagger \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1^* & \mathbf{a}_1^\dagger \\ \mathbf{a}_2^* & \mathbf{a}_2^\dagger \end{pmatrix} \begin{pmatrix} \mathbf{x}_{t-1}^* \\ \mathbf{x}_{t-1}^\dagger \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_t^* \\ \boldsymbol{\varepsilon}_t^\dagger \end{pmatrix}, \tag{10}$$

wherein $\{\boldsymbol{\varepsilon}_t^*\}$ and $\{\boldsymbol{\varepsilon}_t^\dagger\}$ are independent of each other.

1. Equation (10) is similar to Eq. (2) of Gonçalves et al. (2024).
2. If $\mathbf{a}_2^* \equiv \mathbf{0}$, $\{\mathbf{x}_t^\dagger\}$ become exogenous. Let further $\mathbf{a}_2^\dagger \equiv \mathbf{0}$, $\{\mathbf{x}_t^\dagger \ (\equiv \boldsymbol{\varepsilon}_t^\dagger)\}$ are white noises. The setting is in the same sprit as in Inoue et al. (2024, Eq. (3)) and Gonçalves et al. (2024, Eq. (3)).
3. The first part of (10) gives a typical dynamic model with a large number of exogenous regressors: $\mathbf{x}_t^* = \mathbf{a}_1^* \mathbf{x}_{t-1}^* + \mathbf{a}_1 \mathbf{x}_{t-1}^\dagger + \varepsilon_t^*$, which aligns with Adamek et al. (2024, Eq. (2.4)) and Cha (2024, Eq. (2.2)) in the case where $\mathbf{x}_t^*$ is a scalar.

A System of Regression Models

Following our discussion about the link between HDMA(∞) models and HD–AR(p), it is reasonable to assume that $\{\mathbf{x}_t\}$ admits a system of regression models:

$$\mathbf{x}_{t+h} = \mathbf{A}_1\mathbf{x}_t + \cdots + \mathbf{A}_p\mathbf{x}_{t-p+1} + \mathbf{u}_{t,h} = (\mathbf{X}_t^\top \otimes \mathbf{I}_N)\,\mathrm{vec}(\tilde{\mathbf{A}}) + \mathbf{u}_{t,h}, \tag{11}$$

where p is fixed,

- $\tilde{\mathbf{A}} \coloneqq (\mathbf{A}_1, \ldots, \mathbf{A}_p)$, $\mathbf{X}_t = \mathrm{vec}(\mathbf{x}_{t-1}, \ldots, \mathbf{x}_{t-p})$;
- $\mathbf{u}_{t,h}$ has an $(h-1)$–dependent structure.

Remark:

1. $\tilde{\mathbf{A}}$ varies with respect to h. For $h = 1$ only (i.e., HD Var models), refer to Basu and Michailidis (2015), and Miao et al. (2023).
2. N^2p number of parameters, so sparsity is required to balance feasibility and generality in practical analysis.

Our Focus

Recall models (1) and (11):

$$\mathbf{x}_t = \sum_{\ell=0}^{\infty} \mathbf{B}_\ell \boldsymbol{\varepsilon}_{t-\ell}$$

and

$$\mathbf{x}_{t+h} = (\mathbf{X}_t^\top \otimes \mathbf{I}_N)\,\mathrm{vec}(\tilde{\mathbf{A}}) + \mathbf{u}_{t,h},$$

we focus on

$$\text{long–horizon analysis} + \text{HD analysis}$$

in what follows.

Assumptions 1 and 2

Assumption 1

Let $\{\varepsilon_{it}\}$ *be i.i.d. over* (i,t) *and satisfy* $E[\varepsilon_{11}]=0$, $E[\varepsilon_{11}^2]=1$, *and* $E|\varepsilon_{11}|^J<\infty$ *with* $J\,(\geq 4)$ *being fixed and finite.*

Assumption 2

1. *Let* $g(\cdot)$ *be a smooth function such that* $\mathbf{u}_{t,h}=\mathbf{g}(\boldsymbol{\varepsilon}_{t+h},\ldots,\boldsymbol{\varepsilon}_{t+1})$ *satisfies* $E[\mathbf{u}_{t,h}]=\mathbf{0}$, $E[\mathbf{u}_{t,h}\mathbf{u}_{t,h}^\top]=\boldsymbol{\Sigma}_h$, *and* $\max_{i,h}\|\mathbf{e}_i^\top\mathbf{u}_{t,h}\|_J<\infty$, *where* J *is given in Assumption 1.*
2. *Let (i).* $\max_j\sum_{\ell=0}^\infty \ell^{\frac{1}{2}-\frac{1}{2(J+1)}}\|\boldsymbol{\beta}_{\ell,j}\|_2^{\frac{J}{J+1}}<\infty$; *(ii).* $\frac{d_{\boldsymbol{\beta}}^4\log N}{T}\to 0$ *with* $d_{\boldsymbol{\beta}}:=\sum_{\ell=0}^\infty\|\mathbf{B}_\ell\|_\infty$; *(iii).* $h/T\to 0$ *as* $(h,T)\to(\infty,\infty)$, *where* $\mathbf{B}_\ell=(\boldsymbol{\beta}_{\ell,1},\ldots,\boldsymbol{\beta}_{\ell,N})^\top$.

Proposition 1

Suppose that $\{\mathbf{x}_t\}$ *admit both representations* (1) *and* (11)*, Assumptions 1 and 2.1 hold, and* $\mathbf{B}_0 = \mathbf{I}_N$*. For* $\forall h \geq 1$*,*

$$\mathbf{A}_1 = \mathbf{B}_h = (\mathbf{IR}_{h,1}, \ldots, \mathbf{IR}_{h,N}).$$

Remark:

1. Without $\mathbf{B}_0 = \mathbf{I}_N$, we have $\mathbf{A}_1\mathbf{B}_0 = \mathbf{B}_h$.
2. $\mathbf{A}_j$ for $j \geq 2$, while being essential components of (11), are not directly connected to the impulse responses.

Concentration Inequality for Long–Horizon Analysis (I)

Let $u_{is,h}$ and $x_{j,s}$ be the i^{th} and j^{th} elements of $\mathbf{u}_{s,h}$ and $\mathbf{x}_s$ respectively.

Theorem 2

Under Assumptions 1-2, there exit positive constants c_1, c_2, c_3, c_4 *such that for* $\forall i, j \in [N]$, $\forall \ell \in 0 \cup [p-1]$, *and* $\delta > c_4 \sqrt{T} \mu_h^{1+1/J}$

$$\Pr\left(\max_{t \in [T-h]} \left| \sum_{s=1}^{t} u_{is,h} x_{j,s-\ell} \right| \geq \delta \right) \leq c_1 \frac{T}{\delta^J} \mu_h^{J+1} + c_2 \exp\left(- \frac{c_3 \delta^2}{T \mu_h^{2+2/J}} \right),$$

where $\mu_h = \max_{j \in [N]} \sum_{\ell=0}^{\infty} [(\ell + h)^{\frac{J}{2}-1} \|\boldsymbol{\beta}_{\ell,j}\|_2^J]^{\frac{1}{J+1}} \lesssim h^{\frac{J-2}{2(J+1)}}$.

Concentration Inequality for Long–Horizon Analysis (II)

Remark:

1. The term μ_h is simply due to lack of structure in $\mathbf{u}_{t,h}$ and (1).
2. Since $J \geq 4$, $h^{\frac{J-2}{2(J+1)}} \in [h^{1/5}, h^{1/2}]$. Thus, in the worst case scenario, μ_h diverges at the rate $h^{1/2}$.
3. Theorem 2 explains the loss of accuracy in forecasting models as $h \uparrow$.

Estimation Theory (I)

Recall that we need sparsity for $\widetilde{\mathbf{A}} = \{\widetilde{A}_{ij}\}_{N\times Np}$ of (11). Therefore, we introduce

$$\mathbf{S}_{\mathscr{A}} \coloneqq \{I(|\widetilde{A}_{ij}| > 0)\}_{N\times Np} \quad \text{and} \quad \mathbf{S}_{\mathscr{A}^{\bar{}}} \coloneqq \{I(\widetilde{A}_{ij} = 0)\}_{N\times Np},$$

so $d_{\mathscr{A}} \coloneqq \|\mathbf{S}_{\mathscr{A}}\|^2$ gives the total number of nonzero elements in $\widetilde{\mathbf{A}}$.

Accordingly, define the following objective function:

$$Q_0(\tilde{\mathbf{a}}) = \frac{1}{T-h}\sum_{t=1}^{T-h} \|\mathbf{x}_{t+h} - (\mathbf{X}_t^\top \otimes \mathbf{I}_N)\,\mathrm{vec}(\widetilde{\mathbf{a}})\|_2^2, \tag{12}$$

where $\tilde{\mathbf{a}}$ is a generic $N \times Np$ matrix.

Estimation Theory (II)

We then introduce the following procedure to recover $\widetilde{\mathbf{A}}$.

Step 1 : Initial estimation via LASSO:

$$\widehat{\widetilde{\mathbf{a}}} = \underset{\widetilde{\mathbf{a}}}{\operatorname{argmin}}(Q_0(\widetilde{\mathbf{a}}) + \gamma \| \operatorname{vec}(\widetilde{\mathbf{a}})\|_1), \tag{13}$$

where γ is a tuning parameter.

Step 2 : Refined estimation via adaptive LASSO:

$$\widehat{\widetilde{\mathbf{a}}}_\phi = \underset{\widetilde{\mathbf{a}}}{\operatorname{argmin}}(Q_0(\widetilde{\mathbf{a}}) + \gamma \| \operatorname{vec}(\widetilde{\mathbf{a}}) \circ \boldsymbol{\phi}\|_1), \tag{14}$$

where $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_{N^2p})^\top$ is a vector of predetermined weights.

The numerical algorithm has been well developed. See McIlhagga (2016) for instance.

Selection Results (I)

Theorem 3

Let $\gamma \asymp \frac{\mu_h^{1+1/J}\sqrt{\log(N)}}{\sqrt{T}}$. *Under some regularity conditions, we then have the following results:*

1. $\| \operatorname{vec}(\widetilde{\mathbf{A}} - \widehat{\widetilde{\mathbf{a}}})\|_2 = O_P\big(\frac{\mu_h^{1+1/J}\sqrt{d_{\mathscr{A}} \log N}}{\sqrt{T}}\big)$;
2. $\| \operatorname{vec}(\widetilde{\mathbf{A}} - \widehat{\widetilde{\mathbf{a}}})\|_1 = O_P\big(\frac{\mu_h^{1+1/J} d_{\mathscr{A}}\sqrt{\log N}}{\sqrt{T}}\big)$.

Selection of the Lags (I)

Define the following information criterion:

$$\mathrm{IC}(\mathfrak{p}) = \frac{1}{T-h}\sum_{t=1}^{T-h} \|\mathbf{x}_{t+h} - (\mathbf{X}_t^\top \otimes \mathbf{I}_N)\,\mathrm{vec}(\widehat{\widetilde{\mathbf{a}}}_{\mathfrak{p}})\|_2^2 + \mathfrak{p}\cdot\xi,$$

where $\widehat{\widetilde{\mathbf{a}}}_{\mathfrak{p}}$ is obtained via (13) using $\mathfrak{p}$ lags, and ξ is a tuning parameter. The optimal lag is then estimated by

$$\widehat{p} = \underset{\mathfrak{p}\leq\mathfrak{p}^*}{\mathrm{argmin}}\,\mathrm{IC}(\mathfrak{p}), \tag{15}$$

where $\mathfrak{p}^*$ is chosen by the user for specified large and fixed integer.

Selection of the Lags (II)

Theorem 4

Let $\frac{\mu_h^{2+2/J} d_{\mathscr{A}} \log N}{T\xi} \to 0$ *and* $\xi \to 0$. *Then under some regularity conditions, we have* $\Pr(\widehat{p} = p) \to 1$.

Remark:

1. Theorem 4 holds for $\forall h$.
2. From a practical standpoint, we suggest that using a small horizon (e.g., $h = 1$ or 2) is preferable for selecting the optimal lag, as this significantly simplifies the required restrictions.

Covariance Matrix Estimation (I)

Our goal is to estimate $\mathbf{\Omega}_h \coloneqq E[\widetilde{\mathbf{\Omega}}_h]$, where

$$\widetilde{\mathbf{\Omega}}_h \coloneqq \frac{1}{T-h} \sum_{|t-k|<h}^{T-h} (\mathbf{X}_t \otimes \mathbf{I}_N)\mathbf{u}_{t,h}\mathbf{u}_{k,h}^\top(\mathbf{X}_k^\top \otimes \mathbf{I}_N). \tag{16}$$

Note that if $\widetilde{\mathbf{A}}$ is sparse, it will naturally pass the sparsity to $\mathbf{\Omega}_h$.

Example 1 (Cont.) If $\mathbf{a}_1 = \mathbf{0}$, then $\mathbf{\Omega}_h = \mathbf{I}_{N^2}$ for $\forall h \geq 1$.

Therefore, we impose the densest sparse condition (Bickel and Levina, 2008). Specifically, $\mathbf{\Omega}_h \in \mathbb{U}(c_a, c_N, \bar{c})$, where

$$\mathbb{U}(c_a, c_N, \bar{c}) = \big\{\mathbf{\Omega} = \{\omega_{ij}\} \mid \omega_{ii} \leq \bar{c} \text{ and } \sum_j |\omega_{ij}|^{c_a} \leq c_N \text{ for all } i\big\}, \tag{17}$$

$0 < c_a < 1$, and $\bar{c}$ is a fixed number.

Covariance Matrix Estimation (II)

Let

$$\widehat{\mathbf{\Omega}}_h \coloneqq \frac{1}{T-h} \sum_{|t-k|<h}^{T-h} (\mathbf{X}_t \otimes \mathbf{I}_N)\widehat{\mathbf{u}}_{t,h}\widehat{\mathbf{u}}_{k,h}^{\top}(\mathbf{X}_k^{\top} \otimes \mathbf{I}_N), \tag{18}$$

where $\widehat{\mathbf{u}}_{t,h} \coloneqq \mathbf{x}_{t+h} - (\mathbf{X}_t^{\top} \otimes \mathbf{I}_N)\,\mathrm{vec}(\widehat{\widetilde{\mathbf{a}}}_{\phi})$.

The estimator of $\mathbf{\Omega}_h$ is given by

$$\mathscr{G}_{\eta}(\widehat{\mathbf{\Omega}}_h), \tag{19}$$

where $\eta \asymp \sqrt{h \log N/T}$, and $\mathscr{G}_{\eta}(\mathbf{B}) \coloneqq \{b_{ij} I(|b_{ij}| \geq \eta, i \neq j) + b_{ij} I(i = j)\}$ for $\forall \mathbf{B} = \{b_{ij}\}_{N \times N}$ (i.e., *penalize off-diagonal elements*).

Asymptotic Distribution (I)

First, we conduct the node–wise LASSO estimation:

$$\widehat{\mathbf{b}}_i = \underset{\mathbf{b}_i}{\operatorname{argmin}} \Big\{ \frac{1}{T-h} \sum_{t=1}^{T-h} \|\mathbf{Z}_{i,t}^\top - \mathbf{Z}_{-i,t}^\top \mathbf{b}_i\|_2^2 + 2\widetilde{\gamma}_i \|\mathbf{b}_i\|_1 \Big\}, \tag{20}$$

where $\widetilde{\gamma}_i \asymp \frac{\mu_h^{1+1/J}\sqrt{\log(N)}}{\sqrt{T}}$. Let $\widehat{\tau}_i^2 = \frac{1}{T-h}\sum_{t=1}^{T-h} \|\mathbf{Z}_{i,t}^\top - \mathbf{Z}_{-i,t}^\top \widehat{\mathbf{b}}_i\|_2^2 + \widetilde{\gamma}_i \|\hat{\mathbf{b}}_i\|_1$, and then define the debiased estimator $\widehat{\widetilde{\mathbf{a}}}_{\text{bc}}$ as

$$\operatorname{vec}(\widehat{\widetilde{\mathbf{a}}}_{\text{bc}}) = \operatorname{vec}(\hat{\tilde{\mathbf{a}}}) + \frac{1}{T-h} \widehat{\mathbf{\Omega}}_z \sum_{t=1}^{T-h} \mathbf{Z}_t \big(\mathbf{x}_{t+h} - \mathbf{Z}_t^\top \operatorname{vec}(\widehat{\widetilde{\mathbf{a}}}) \big), \tag{21}$$

where $\widehat{\mathbf{\Omega}}_z = \hat{\mathbf{T}}_z^{-1} \widehat{\mathbf{C}}_z$, $\widehat{\mathbf{T}}_z = \operatorname{diag}(\widehat{\tau}_1^2, \ldots, \widehat{\tau}_{N^2p}^2)$, and $\widehat{\mathbf{C}}_z = (\widehat{\mathbf{C}}_{z,1}, \ldots, \widehat{\mathbf{C}}_{z,N^2p})^\top$. Each row $\widehat{\mathbf{C}}_{z,i}^\top$ is constructed by placing a 1 in the i^{th} position and filling the remaining N^2p-1 entries with the elements of $-\widehat{\mathbf{b}}_i$ in their natural order.

Asymptotic Distribution (II)

Theorem 5

Let $\|\boldsymbol{\rho}_{\mathbf{a}}\|_1 < \infty$ *and* $(d_{\mathscr{A}} + \max_i s_{\mathscr{A},i})(\log N)T^{-1/2}\mu_h^{2+2/J} \to 0$. *Then under some regularity conditions, we have*

$$\sqrt{T}\left(\boldsymbol{\rho}_{\mathbf{a}}^{\top}\widehat{\boldsymbol{\Omega}}_z\,\mathscr{G}_\eta(\widehat{\boldsymbol{\Omega}}_h)\,\widehat{\boldsymbol{\Omega}}_z\boldsymbol{\rho}_{\mathbf{a}}\right)^{-1/2}\boldsymbol{\rho}_{\mathbf{a}}^{\top}\ \mathrm{vec}(\widehat{\widetilde{\mathbf{a}}}_{\mathrm{bc}} - \widetilde{\mathbf{A}}) \xrightarrow{D} \mathcal{N}(0,1).$$

Remark: Under the HD LP framework, the node-wise LASSO estimation can be much simplified numerically in practical analysis.

We offer a detailed algorithm with theoretical justification in the paper.

Simulation Design (I)

Let the DGP be as follows:

$$\mathbf{x}_t = \mathbf{a}_1\mathbf{x}_{t-1} + \mathbf{a}_2\mathbf{x}_{t-2} + \boldsymbol{\varepsilon}_t, \tag{22}$$

where $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{I}_N)$. Thus, $p = 2$. $\mathbf{a}_1 = \{a_{1,ij}\}_{N\times N}$ and $\mathbf{a}_2 = \{a_{2,ij}\}_{N\times N}$ are designed as follows:

$$a_{1,ij} = \begin{cases} 0.25 & \text{if } i = j \text{ and } i \leq N/2 \\ 0.35 & \text{if } i - j = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad a_{2,ij} = \begin{cases} 0.35 & \text{if } i = j \text{ and } i > N/2 \\ -0.25 & \text{if } j - i = 1 \\ 0 & \text{otherwise} \end{cases}.$$

The model (22) admits an HDMA(∞) representation $\mathbf{x}_t = \sum_{\ell=0}^{\infty} \mathbf{B}_\ell \boldsymbol{\varepsilon}_{t-\ell}$, in which

$$\mathbf{B}_\ell = \mathbf{S}\mathbf{C}^\ell\mathbf{S}^\top, \quad \mathbf{S} = (\mathbf{I}_N, \mathbf{0}_N), \quad \text{and} \quad \mathbf{C} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 \\ \mathbf{I}_n & \mathbf{0} \end{pmatrix}.$$

Simulation Design (II)

For each generated dataset, we fit the observations $\{\mathbf{x}_t\}$ to the following set of h–step ahead forecasting models:

$$\mathbf{x}_{t+h} = \mathbf{A}_1\mathbf{x}_t + \mathbf{A}_2\mathbf{x}_{t-1} + \mathbf{u}_{t,h}, \tag{23}$$

where $h \geq 1$.
We consider the cases with $N \in \{20, 30, 40\}$ and $T \in \{300, 400, 500\}$, so *the number of parameters involved ranges from* 800 *to* 3200.

Simulation Results (I)

The estimation procedure is

1. estimating the number of lags based on ℓ-step ahead forecasting models;
2. detecting the sparsity, and estimating $\mathbf{B}_h = \{b_{h,ij}\}$ under the selected optimal lag.

Thus, the first measure is

$$\underbrace{S_\ell^- = \frac{1}{R}\sum_{r=1}^{R} I(\widehat{p}_{\ell,r} < p)}_{\text{under-selection}} \quad \underbrace{S_\ell = \frac{1}{R}\sum_{r=1}^{R} I(\widehat{p}_{\ell,r} = p)}_{\text{correct selection}} \quad \underbrace{S_\ell^+ = \frac{1}{R}\sum_{r=1}^{R} I(\widehat{p}_{\ell,r} > p)}_{\text{over-selection}},$$

where $\ell \in \{1, 2\}$, and $\widehat{p}_{\ell,r}$ denotes the estimated lag in the r^{th} replication for $\forall \ell \in \{1, 2\}$.

Simulation Results (II)

For each $\widehat{p}_\ell$ with $\ell \in \{1, 2\}$, we proceed, and quantify the sparsity detection as:

$$\text{SL}_{|\widehat{p}_\ell} = \frac{1}{R}\sum_{r=1}^{R}\frac{1}{N^2}\sum_{i,j}|I(\widehat{a}_{a,ij,r} = 0) - I(b_{h,ij} = 0)|,$$

where the subscript $|\ \widehat{p}_\ell$ indicates the results are c based on the estimated lag order $\hat{p}_\ell$, $\widehat{\mathbf{A}}_a = \{\hat{a}_{a,ij}\}$ defines the adaptive LASSO estimate, and r indexes the r^{th} replication.

Finally, based on $\widehat{p}_\ell$ and the estimated sparsity, we introduce the following measure:

$$\text{AD}_{\star|\widehat{p}_\ell} = \frac{1}{R}\sum_{r=1}^{R}\|\widehat{\mathbf{A}}_{\star,r} \circ \hat{\mathbf{S}}_r - \mathbf{B}_h\|_2,$$

where $\star \in \{a, d\}$, a and d stand for adaptive LASSO and debiased LASSO respectively, and $\widehat{\mathbf{S}}_r \coloneqq \{I(\widehat{a}_{a,ij,r} \neq 0)\}$.

Simulation Results (III)

Table: Results of Selection and Estimation (Conditional on $\widehat{p}_2$)

		Panel A						Panel B					
					$\mathrm{SL}_{\vert\widehat{p}_2}$			$\mathrm{AD}_{a\vert\widehat{p}_2}$			$\mathrm{AD}_{d\vert\widehat{p}_2}$		
N	T	S_2^-	S_2	S_2^+	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$	$h=1$	$h=5$	$h=10$
20	300	0.000	0.932	0.068	0.019	0.308	0.620	0.465	0.410	0.376	0.406	0.396	0.369
	400	0.000	0.978	0.022	0.013	0.304	0.619	0.408	0.393	0.373	0.343	0.375	0.367
	500	0.000	0.994	0.006	0.009	0.301	0.618	0.364	0.379	0.370	0.307	0.357	0.363
30	300	0.000	0.966	0.034	0.023	0.216	0.448	0.563	0.453	0.412	0.543	0.446	0.410
	400	0.000	0.992	0.008	0.018	0.214	0.448	0.536	0.445	0.412	0.499	0.435	0.410
	500	0.000	0.998	0.002	0.015	0.213	0.448	0.497	0.437	0.411	0.426	0.424	0.409
40	300	0.000	0.974	0.026	0.024	0.166	0.349	0.591	0.471	0.428	0.587	0.468	0.427
	400	0.000	0.996	0.004	0.020	0.165	0.349	0.583	0.467	0.428	0.573	0.463	0.427
	500	0.000	0.992	0.002	0.017	0.165	0.349	0.568	0.465	0.428	0.546	0.458	0.427

- For $\forall \ell \in \{1, 2\}$, we observe that the values of S_ℓ are close to 1, which aligns with theoretical expectations.
- When $\ell = 2$, we have a small proportion over-section (i.e., $S_\ell^+ > 0$), and S_ℓ moves towards 1 as T increases.
- As expected, the coefficients have the strongest signal for $h = 1$ when implementing LASSO procedure.
- Hence, it is reasonable to expect $h = 1$ offers better finite sample performance from different perspectives such as those presented by Panel A of Table 1.

- We observe in Panel A of Table 1 that the SL value tends to become relatively large as the horizon h increases.
- As h increases, the number of non-zero elements in the true matrix $\mathbf{B}_h$ may rise; however, many of these elements become negligible (i.e., close to zero) due to weak signal propagation.
- The adaptive LASSO procedure may incorrectly detect these negligible elements as zero, thereby leading to a higher overall SL score.

- Panel B of Table 1 summarises the relevant results.
- The results conditional on $\widehat{p}_1$ and $\widehat{p}_2$ are very similar, which should be expected.
- Despite the relatively large number of parameters, which ranges from 800 to 3200, both the adaptive LASSO estimation error (AD_a) and the debiased LASSO estimation error (AD_d) are observed to be small.
- Also, both AD_a and AD_d move towards 0 as T increases for all (N, h).
- Overall, the debiased LASSO exhibits superior finite sample performance, which is consistent with the theoretical expectation that debiasing mitigates the shrinkage bias inherent in the standard LASSO procedure.

Conclusions

- We establish large–sample properties for the LP methodology within a HD framework, with a central focus on achieving robust long-horizon inference.
- We integrate a general dependence structure into h–step ahead forecasting models via a flexible specification of the residual terms.
- Additionally, we study the corresponding HD covariance matrix estimation, explicitly addressing the complexity arising from the long-horizon setting.
- Extensive Monte Carlo simulations are conducted to substantiate the derived theoretical findings.
- In an empirical study, we utilise the proposed HD+LP framework to revisit the impact of business news attention on U.S. industry-level stock volatility.

Notational appendix

- For a positive integer L, we let $[L] \coloneqq \{1, \ldots, N\}$. Let $\mathbf{e}_i$ be a $N \times 1$ selection vector with the i^{th} element being 1 and the others being 0, and $i \in [N]$.
- For a vector $\mathbf{v} = (v_1, \ldots, v_N)^\top$, we let $\|\mathbf{v}\|_1 \coloneqq \sum_{i=1}^N |v_i|$, and $\|\mathbf{v}\|_2 \coloneqq \{\sum_{i=1}^N |v_i|^2\}^{1/2}$.
- For a matrix $\mathbf{B} \coloneqq \{b_{ij}\}$, we let $\|\mathbf{B}\|$, $\|\mathbf{B}\|_2$, $\|\mathbf{B}\|_{\max}$, $\|\mathbf{B}\|_1$, and $\|\mathbf{B}\|_\infty$ define its Frobenius norm, spectral norm, max norm, column norm, and row norm respectively.
- Additionally, we define the following operator for a symmetric square matrix $\mathbf{B}$:

$$\mathscr{G}_\eta(\mathbf{B}) \coloneqq \{b_{ij} I(|b_{ij}| \geq \eta) \quad \text{for} \quad i \neq j\}, \tag{24}$$

where η is a user-specified threshold parameter. For a random variable z, let $\|z\|_p \coloneqq (E|z|_p^p)^{1/p}$.

Empirical Study

- We utilize the proposed high–dimensional local projection framework to revisit the impact of business news attention on U.S. industry-level stock volatility.

  See Baker et al. (2016) and Bybee et al. (2024) for relevant literature for instance.
- We include lagged volatilities from all industries as regressors to model cross–sector spillovers. The spillover of volatility across assets and sectors has long been a focus in finance (Diebold and Yilmaz, 2009; Diebold and Yılmaz, 2014).

Takeaway Message:
Recession-related news attention continues to exert a positive effect on volatility across most industries, and the presence of volatility spillovers from the financial sector to other sectors is again confirmed.

Data (I)

U.S. industry-level volatility measures — We collect industry portfolio returns for 37 industry categories from Kenneth French's data library, and monthly industry volatility is computed as the standard deviation of daily returns within each calendar month.

News attention data — We rely on the monthly topic-level attention indices constructed by Bybee et al. (2024), which cover 180 distinct news topics over the period 1984-2017. From this universe, we select 21 topics that are most closely related to economic growth and financial markets.

Eventually, we obtain a multivariate time-series dataset spanning January 1984 to June 2017 ($T = 402$), comprising a total of 58 variables: 37 industry-level volatility series and 21 news topic attention measures. All volatility series and news attention indices are subsequently standardized to have zero mean and a variance of one.

Data (II)

Panel A: Industry Classification		Panel B: News Topics		
	Industry		News Topic	Category
AG	Agriculture, forestry, and fishing	**REC**	**Recession**	Economic Growth
MN	Mining	RHI	Record High	Economic Growth
OG	Oil and Gas Extraction	EGR	Economic Growth	Economic Growth
⋮	⋮	⋮	⋮	⋮
AP	Apparel and other Textile Products	IPO	IPOs	Financial Market
WD	Lumber and Wood Products	BYD	Bond Yields	Financial Market
FU	Furniture and Fixtures	SRT	Short Sales	Financial Market
PA	Paper and Allied Products	SML	Small Caps	Financial Market
PR	Printing and Publishing	TBY	Treasury Bonds	Financial Market
⋮	⋮	⋮	⋮	⋮

Selected Results (I)

Regressions — We consider three LP specifications:

- **Model 1**: 37 industry-level volatility variables and the **recession-related news attention index**;
- **Model 2**: 37 industry-level volatility variables and 8 news attention indices related to economic growth;
- **Model 3**: 37 industry-level volatility variables, 8 news attention indices related to economic growth, and 13 news attention indices related to financial markets.

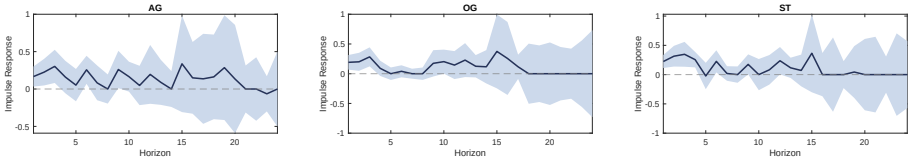
Selected Results (II)

IC suggests $\widehat{p} = 3$.

Table: Estimated Impulse Responses of Volatilities to Recession Attention

Model 1						Model 2						Model 3					
Ind.	IR	StD	Ind.	IR	StD	Ind.	IR	StD	Ind.	IR	StD	Ind.	IR	StD	Ind.	IR	StD
AG	0.167	0.081	FM			FM	0.188	0.096	FM	0.102	0.095	AG	0.203	0.104	FM	0.124	0.125
MN			MC	0.191	0.101	MC			MC	0.198	0.097	MN			MC	0.195	0.120
OG	0.191	0.075	EL	0.184	0.108	EL	0.222	0.088	EL			OG	0.234	0.133	EL	0.151	0.119
ST	0.224	0.067	CA	0.211	0.079	CA	0.188	0.046	CA	0.194	0.043	ST	0.217	0.093	CA	0.197	0.115
CN	0.212	0.072	IN			IN	0.188	0.024	IN			CN	0.265	0.055	IN		
FD			MF	0.154	0.084	MF	0.098	0.104	MF	0.120	0.074	FD	0.092	0.134	MF		
TB			TS	0.169	0.091	TS			TS	0.150	0.071	TB			TS	0.143	0.050
TX	0.193	0.078	PH	0.173	0.081	PH	0.182	0.067	PH	0.148	0.065	TX	0.189	0.109	PH	0.126	0.115
AP	0.164	0.086	TV	0.192	0.079	TV	0.132	0.076	TV	0.150	0.068	AP			TV	0.146	0.136
WD	0.108	0.072	UT	0.227	0.090	UT	0.074	0.065	UT	0.216	0.094	WD	0.127	0.109	UT	0.224	0.149
FU	0.092	0.067	GB			GB	0.083	0.074	GB			FU	0.106	0.141	GB		
PA	0.089	0.077	SM			SM			SM			PA	0.123	0.123	SM		
PR	0.136	0.063	WT			WT	0.115	0.059	WT			PR	0.152	0.106	WT		
CH	0.166	0.096	WS			WS	0.174	0.092	WS	0.102	0.074	CH	0.143	0.086	WS		
PE			RT	0.235	0.070	RT			RT	0.185	0.064	PE			RT	0.235	0.141
RB	0.162	0.091	FI	0.183	0.072	FI	0.112	0.089	FI	0.167	0.077	RB	0.128	0.139	FI	0.162	0.116
LE	0.231	0.111	SV	0.173	0.100	SV	0.213	0.125	SV			LE	0.186	0.165	SV	0.149	0.103
GL	0.152	0.075	GV			GV	0.126	0.055	GV			GL	0.121	0.106	GV		
MT	0.162	0.068				MT	0.161	0.061				MT	0.165	0.104			

Selected Results (III)



(a) AG (b) OG (c) ST

(d) CN (e) MF (f) FI

Figure: Estimated Impulse Responses of Industry Volatilities to Recession News Attention.

References (I)

Adamek, R., S. Smeekes, and I. Wilms (2024). Local projection inference in high dimensions. *The Econometrics Journal 27*(3), 323–342.

Baker, S. R., N. Bloom, and S. J. Davis (2016). Measuring economic policy uncertainty. *The quarterly journal of economics 131*(4), 1593–1636.

Basu, S. and G. Michailidis (2015). Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics 43*(4), 1535–1567.

Bickel, P. J. and E. Levina (2008). Covariance regularization by thresholding. *The Annals of Statistics 36*(6), 2577–2604.

Bybee, L., B. Kelly, A. Manela, and D. Xiu (2024). Business news and business cycles. *The Journal of Finance 79*(5), 3105–3147.

Cha, J. (2024). Local projections inference with high-dimensional covariates without sparsity.

Diebold, F. X. and K. Yilmaz (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal 119*(534), 158–171.

References (II)

Diebold, F. X. and K. Yılmaz (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics 182*(1), 119–134.

Gonçalves, S., A. M. Herrera, L. Kilian, and E. Pesavento (2024). State-dependent local projections. *Journal of Econometrics 244*(2), 105702.

Inoue, A., B. Rossi, and Y. Wang (2024). Local projections in unstable environments. *Journal of Econometrics 244*(2), 105726.

Jordá, Ó. (2005). Estimation and inference of impulse responses by local projections. *American Economic Review 95*(1), 161–182.

Jordá, Ó. and A. M. Taylor (2025). Local projections. *Journal of Economic Literature 63*(1), 59–110.

McIlhagga, W. (2016). Penalized: A MATLAB toolbox for fitting generalized linear models with penalties. *Journal of Statistical Software 72*, 1–21.

Miao, K., P. C. Phillips, and L. Su (2023). High-dimensional VARs with common factors. *Journal of Econometrics 233*(1), 155–183.

References (III)

Montiel Olea, J. L. and M. Plagborg-Møller (2021). Local projection inference is simpler and more robust than you think. *Econometrica 89*(4), 1789–1823.

Plagborg-Møller, M. and C. K. Wolf (2021). Local projections and vars estimate the same impulse responses. *Econometrica 89*(2), 955–980.