Inference in High-Dimensional Panel Models:
Two-Way Dependence and Unobserved Heterogeneity

Kaicheng Chen

Department of Economics
Michigan State University

CES NA Annual Conference
Mar 23, 2025

Table of Contents

Preview of Results

- **Model**: a high-dimensional (regression) model for panel data. E.g., $Y_{it} = \theta_0 D_{it} + g_0(X_{it}, c_i, d_t) + U_{it}$.
- **Target**: inference for low-dim. parameters in the presence of high-dim. nuisance parameters.
- **Challenges**: unit and time cluster dependence as well as weak dependence across clusters.
- **Main contribution i**: a variant of (post) **LASSO**, robust to **two-way cluster-dependence** in panel data.
- **Main contribution ii**: a clustered-panel **cross-fitting** approach.

Preview of Results

- Both the variant of LASSO and panel cross-fitting are of **independent interest**.
- Together, they allow for **consistent estimation** and **valid inference** about the low-dim. parameter.
- **Main contribution iii**: valid inference using the full sample in the partial linear model.
- **Application**: hidden dimensionality in estimating government spending multiplier.

Example: Hidden High Dimensionality

- **Estimation of the multiplier**: the percentage increase in output that results from the 1 percent increase in government spending.
- Researchers often start with a **baseline** model:

$$Y_{it} = \theta_0 D_{it} + X_{it}\pi_0 + c_i + d_t + U_{it},\ \mathrm{E}[Z_{it}U_{it}] = 0$$

- A small number of control variables are considered: Why is high dimensionality relevant here?
- **Robustness check**: to avoid endogeneity caused by potential misspecification,

$$Y_{it} = \theta_0 D_{it} + g_0(X_{it}, c_i, d_t) + U_{it}.$$

- **Cost**: noisy or infeasible estimation with limited sample sizes (51 states with 39 periods).

Table of Contents

Challenge One

- To reduce dimensionality: sparse method, regularized estimator, e.g. LASSO.
- Consider the simplified model using the pooled panel:

$$\begin{aligned} Y_{it} =& \theta_0 D_{it} + g_0(X_{it}) + U_{it} \\ =& \theta_0 D_{it} + f_{it}\beta_0 + r_{it} + U_{it} \text{ by sparse approximation} \end{aligned}$$

- Obtain $(\tilde{\theta}, \tilde{\beta})$ by running penalized least square of Y_{it} on (D_{it}, f_{it}).
- Two-way cluster-dependence : Graphic illustration

Challenge One (continued)

- Is it valid under two-way dependence? Any adjustment needed? Rate of convergence?
- Approach 1: Assuming the stochastic error is **conditionally normal** (Bickel et al., 2009, AOS).
- Approach 2: **Self-normalizing** the non-Gaussian errors (Belloni et al., 2012, ECTA, Belloni et al., 2016, JBES)
- Approach 3: Deriving **new concentration inequalities** allowing for dependent error process (Babii et al., 2023, JOE, Gao et al., 2024, WP).
- **My proposal**: weighted LASSO using regressor-specific penalty weights robust to two-way cluster dependence.
- My construction of penalty level and weights.

Consistency and convergence rate results

- **Theorem**: Given the AHK, approximate sparsity, feasible weights, and regularity conditions, with some $C_\lambda = O(1)$ and $\gamma = o(1)$, we have the number of selected regressors be $O(s)$ and the l^2 rate of convergence for the (post) two-way cluster-LASSO is $O_P\left(\sqrt{\frac{s\log(p/\gamma)}{N\wedge T}}\right)$.
- **Comparison**: $O_P\left(\sqrt{\frac{s\log p}{NT}}\right)$ under random sampling as in Bickel et al., 2009, AOS; $O_P\left(\sqrt{\frac{s\log(p\vee NT)}{NT}}\right)$ under random sampling in Belloni et al., 2012, ECTA; $O_P\left(\sqrt{\frac{s\log(p\vee NT)}{Nl_T}}\right)$ under cross-sectional independence in Belloni et al. (2016) where $l_T \in [1, T]$.
- Oracle case

Table of Contents

Challenge Two

- Consider a semiparametric approach:

$$\widehat{\theta} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}' D_{it}\right]^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}'(Y_{it} - f_{it}\widehat{\zeta}).$$

- $\widehat{\zeta}$ can be noisy due to two-way cluster dependence and high dimensionality.
- **A better second-step** estimator: Let $\ddot{D}_{it} := D_{it} - \widehat{\mathrm{E}}[D_{it}|X_{it}]$,

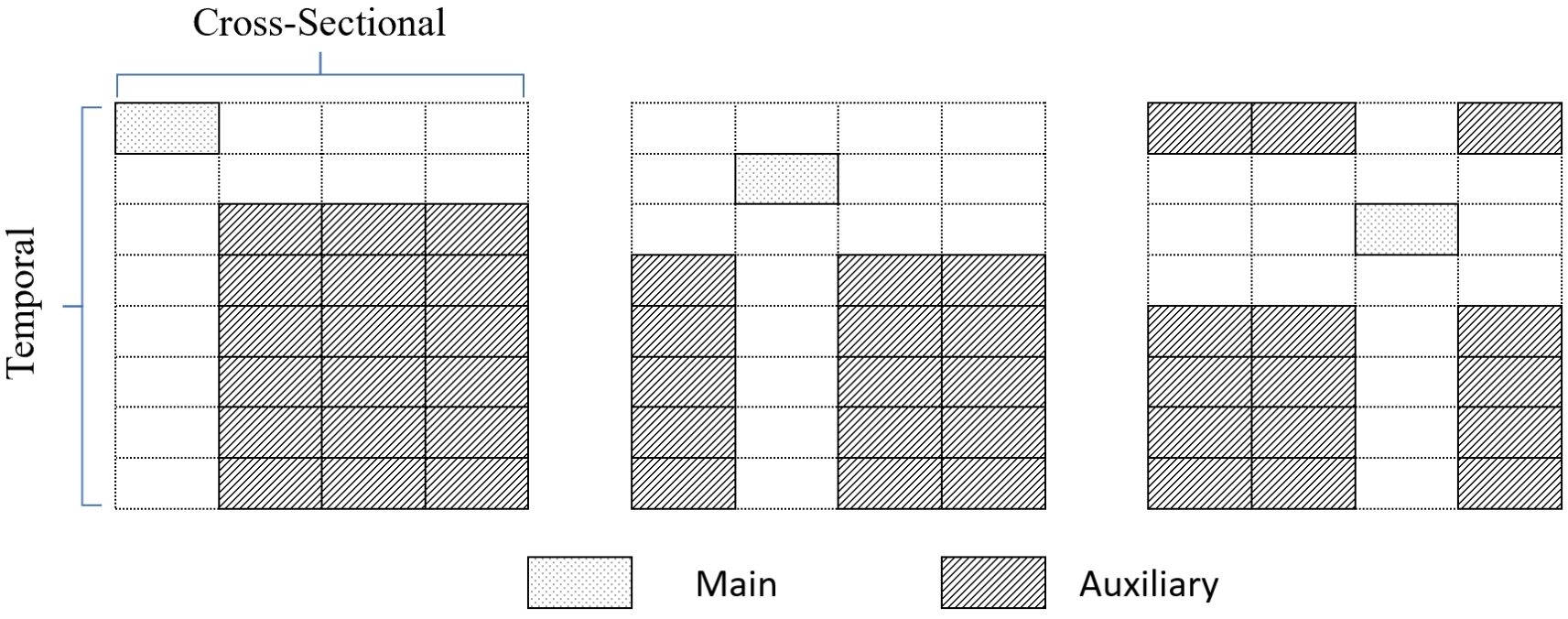
$$\widehat{\theta} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{D}_{it}' D_{it}\right]^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{D}_{it}' \left(Y_{it} - f_{it}\widehat{\zeta}\right).$$

- But there is still a **problematic error term** in $\widehat{\theta} - \theta_0$:

$$\sum_{i=1}^{N}\sum_{t=1}^{T} V_{it}^{D} f_{it} \left(\zeta_0 - \widehat{\zeta}\right), \quad V_{it}^{D} := D_{it} - \mathrm{E}[D_{it}|X_{it}].$$

- **Cross-fitting:** split the sample for the two-step estimations.

Clustered-Panel Cross-Fitting



Lemma (validity of the cross-fitting): Under Assumptions 1 (AHK) and 2 (beta-mixing), the cross-fitting sub-samples are "approximately" independent as $N, T \to \infty$ with $\log(N)/T \to 0$.

Asymptotic Normality

- **Theorem**: Given rates of convergence for the first-step and regularity conditions, $\sqrt{N \wedge T}\left(\widehat{\theta} - \theta_0\right) \Rightarrow \mathrm{N}(0, V)$ where $V := A_0^{-1}\Omega A_0^{-1'}$, $\Omega := \Lambda_a \Lambda_a' + c\Lambda_g \Lambda_g'$.
- A sufficient L^2 rate of convergence for η_0 is $o\left((N \wedge T)^{-1/4}\right)$.
- Cluster-robust variance estimators

Table of Contents

Challenge Three

- Consider the following partial linear model:

$$Y_{it} = D_{it}\theta_0 + g(X_{it}, c_i, d_t) + U_{it}, \ \mathrm{E}[U_{it}|X_{it}, c_i, d_t] = 0.$$

- Z_{it} has the same dimension of D_{it}; $\mathrm{E}[Z_{it}U_{it}] = 0$. As a special case, $Z_{it} = D_{it}$.
- Correlated random effects: generalized Mundlak device.
- Common fixed-effect or random-effect approaches **may not be compatible** with cross-fitting.
- Without cross-fitting, hard to establish inferential theory with **growing dimensions** in general.
- Inference using **full sample** is considered in this model.

Asymptotic Normality without Cross-Fitting

- Under sparse approximation and Mundlak device, the (near) Neyman-orthogonal moment function is given by

$$\psi(W_{it};\theta,\eta) := (Z_{it} - f_{it}\zeta_0)\,(Y_{it} - f_{it}\beta_0 - \theta_0\,(D_{it} - f_{it}\pi_0))\,.$$

  where f_{it} includes a constant and the polynomial transformation of $(X_{it}, \bar{X}_i, \bar{X}_t, \bar{D}_i, \bar{D}_t)$.

- **Theorem**: Under Assumptions AHK, generalized Mundlak device, regularity conditions and sparse approximation with $s = o\left(\frac{\sqrt{N\wedge T}}{\log(p/\gamma)}\right)$, $\|r_{it}^{l}\|_{NT,2} = o_P\left(\sqrt{\frac{1}{N\wedge T}}\right)$ for $l = Y, D$, as $N, T \to \infty$ and $N/T \to c$ where $0 < c < \infty$, the full-sample two-step estimator is asymptotically normal.

- Consistent variance estimators using full sample

Table of Contents

Simulation: DGP(i)

- DGP(i) - Linear model:

$$\begin{aligned} Y_{it} &= D_{it}\theta_0 + X_{it}\beta_0 + U_{it}, \\ D_{it} &= X_{it}\pi_0 + V_{it}, \end{aligned}$$

  where β_0 and π_0 are sparse in a cut-off design.

- DGP(i) - Additive components:

$$\begin{aligned} X_{it,j} &= w_1\alpha_{i,j} + w_2\gamma_{t,j} + w_3\varepsilon_{it,j}, \\ U_{it} &= w_1\alpha_i^u + w_2\gamma_t^u + w_3\varepsilon_{it}^u, \\ V_{it} &= w_1\alpha_i^v + w_2\gamma_t^v + w_3\varepsilon_{it}^v, \end{aligned}$$

Simulation: DGP(ii)

- DGP(ii) - Partial linear model:

$$\begin{aligned} Y_{it} &= D_{it}\theta_0 + (X_{it}\beta_0 + c_i + d_t)^2 + U_{it}, \\ D_{it} &= \frac{\exp(X_{it}\pi_0)}{1+\exp(X_{it}\pi_0)} + V_{it}, \\ c_i &= \bar{D}_i + \bar{X}_i\xi_0 + \epsilon_i^c, \quad d_t = \bar{D}_t + \bar{X}_t\zeta_0 + \epsilon_t^d, \end{aligned}$$

  where β_0, π_0, ξ_0, and ζ_0 are sparse in a polynomial-decay design;

- DGP(ii) - Multiplicative components:

$$\begin{aligned} X_{it,j} &= w_1\alpha_{i,j} + w_2\gamma_{t,j} + w_3\varepsilon_{it,j}, \\ U_{it} &= \frac{w_4}{c_p}\sum_{j=1}^{p}[\alpha_i^u\gamma_{t,j} + \alpha_{i,j}\gamma_t^u] + w_5\varepsilon_{it}^u, \\ V_{it} &= \frac{w_4}{c_p}\sum_{j=1}^{p}[\alpha_i^v\gamma_{t,j} + \alpha_{i,j}\gamma_t^v] + w_5\varepsilon_{it}^v, \end{aligned}$$

Simulation results

Table 1: DGP(i) with $N = T = 25$, $s = 5$, $p = 200$, $\iota = 0.5$, $\rho = 0.5$, $c_\beta = c_\pi = 0.5$

Cross Fitting	First-Step Estimator	First-Step Ave. Sel. Y	First-Step Ave. Sel. D	Second-Step Bias	Second-Step SD	Second-Step RMSE	Coverage (%) CHS	Coverage (%) DKA
No	POLS	200	200	0.003	0.053	0.053	78.9	95.1
	H LASSO	26.0	26.0	0.062	0.065	0.090	58.5	78.7
	R LASSO	17.6	17.6	0.070	0.067	0.097	65.2	79.5
	C LASSO	8.6	8.9	0.036	0.095	0.101	80.0	87.5
	TW LASSO	6.7	6.9	0.023	0.096	0.099	84.3	90.4
Yes	POLS	200	200	0.006	0.113	0.113	98.2	99.4
	H LASSO	16.9	16.6	0.053	0.131	0.141	96.0	97.6
	R LASSO	9.5	9.5	0.054	0.130	0.141	96.0	98.2
	C LASSO	8.0	8.1	0.041	0.130	0.136	96.2	97.4
	TW LASSO	6.7	6.4	0.057	0.126	0.138	95.8	97.2

Simulation results

Table 2: DGP(i) with $N = T = 25$, $s = 5$, $p = 600$, $\iota = 0.5$, $\rho = 0.5$, $c_\beta = c_\pi = 0.5$

Cross Fitting	First-Step Estimator	First-Step Ave. Sel. Y	First-Step Ave. Sel. D	Second-Step Bias	Second-Step SD	Second-Step RMSE	Coverage (%) CHS	Coverage (%) DKA
No	POLS	600	600	0.008	0.221	0.221	26.6	38.6
	H LASSO	39.5	39.8	0.073	0.049	0.087	51.2	78.9
	R LASSO	25.1	25.3	0.079	0.055	0.097	52.4	79.1
	C LASSO	14.0	15.2	0.058	0.096	0.112	68.8	78.4
	TW LASSO	6.9	7.5	0.033	0.098	0.103	81.6	88.1
Yes	H LASSO	24.8	24.7	0.056	0.134	0.146	94.5	98.4
	R LASSO	12.1	12.1	0.054	0.137	0.147	94.5	96.1
	C LASSO	10.7	11.6	0.043	0.139	0.145	95.1	96.1
	TW LASSO	6.8	7.6	0.065	0.140	0.154	90.7	95.1

Simulation results

Table 3: DGP(ii) with $N = T = 25$, $s = p = 10$, $\iota = 0.5$, $\rho = 0.5$, $c_\beta = 1, c_\pi = 4, c_\xi = c_\zeta = 1/4$; 2nd-order polynomial series are used for approximation

Cross Fitting	First-Step Estimator	First-Step Ave. Sel. Y	First-Step Ave. Sel. D	Second-Step Bias	Second-Step SD	Second-Step RMSE	Coverage (%) CHS	Coverage (%) DKA
No	POLS	560	560	0.012	0.173	0.173	54.4	67.4
	H LASSO	12.2	3.4	0.032	0.126	0.130	87.2	90.8
	R LASSO	11.0	3.3	0.030	0.127	0.130	86.2	91.0
	C LASSO	12.3	24.7	0.030	0.127	0.130	87.8	91.8
	TW LASSO	9.3	3.1	0.023	0.127	0.129	87.8	93.6
Yes	H LASSO	9.0	2.6	0.015	0.156	0.157	95.6	98.8
	R LASSO	6.9	2.0	0.010	0.157	0.158	95.8	98.8
	C LASSO	9.1	3.1	0.003	0.153	0.153	96.6	99.0
	TW LASSO	6.8	1.2	0.020	0.151	0.152	97.2	98.8

Table of Contents

Example (continued): Hidden High Dimensionality

- A partial linear model with non-additive unobserved heterogeneity:

$$Y_{it} = \theta_0 D_{it} + g(X_{it}, W_t, c_i, d_t) + U_{it},\ \mathrm{E}[U_{it}|X_{it}, W_t, c_i, d_t] = 0$$

- Y_{it}: state gross output; D_{it} state military spending; W_t: real interest rate and national oil price. X_{it} state population.
- Every variable is in terms of the change instead of the level;
- Due to the endogeneity in the variation of the regional military procurement, the identification is through a Bartik IV: $E[Z_{it} U_{it}] = 0$.

Table 4: Multiplier estimates from the original model

(1) Unobs. Heterog.	(2) Oil Price	(3) Real Int.	(4) Pop.	(5) First Step	(6) IV 1 $\widehat{\theta}$	(7) CHS s.e.	(8) DKA s.e.
	No	No	No	POLS	1.43	0.68	0.81
	Yes	No	No	POLS	1.30	0.56	0.72
Fixed Effects	No	Yes	No	POLS	1.40	0.57	0.70
	Yes	Yes	No	POLS	1.27	0.45	0.71
	Yes	Yes	Yes	POLS	1.36	0.43	0.56

Table 5: Estimates of the open economy relative multiplier from the extended model.

(1) Cross-Fitting	(2) Poly. Trans.	(3) Param. Gen.	(4) First Stage	(5) Z: Param. Sel.	(6) $\widehat{\theta}$	(7) CHS s.e.	(8) DKA s.e.
No	None	7	POLS	7	1.51	0.66	0.82
			H LASSO	2	1.43	0.66	0.81
			C LASSO	4	1.43	0.66	0.81
			TW LASSO	2	1.43	0.70	0.84
No	2nd	35	POLS	35	1.73	0.99	1.15
			H LASSO	6	1.73	1.01	1.17
			CR LASSO	5	1.75	1.02	1.19
			TW LASSO	3	1.47	0.62	0.77
No	3rd	119	POLS	119	2.20	1.19	1.37
			H LASSO	10	1.97	1.16	1.38
			CR LASSO	6	0.98	0.66	0.82
			TW LASSO	5	1.47	0.61	0.76

Table 6: Estimates of the open economy relative multiplier from the extended model.

(1) Cross-Fitting	(2) Poly. Trans.	(3) Param. Gen.	(4) First Stage	(5) Z: Param. Ave. Sel.	(6) $\widehat{\theta}$	(7) CHS s.e.	(8) DKA s.e.
Yes	None	7	H LASSO	2.0	1.28	1.73	2.00
			C LASSO	2.0	1.32	1.75	2.03
			TW LASSO	2.6	1.18	1.77	2.05
Yes	2nd	35	H LASSO	5.2	1.12	2.18	2.52
			C LASSO	5.8	1.46	1.95	2.24
			TW LASSO	4.1	1.20	1.42	1.70
Yes	3rd	119	H LASSO	8.3	1.81	3.17	3.47
			C LASSO	6.5	1.25	1.59	1.91
			TW LASSO	5.3	1.50	1.18	1.44

Table of Contents

Summary

- The inferential theory for high-dim. models is particularly **relevant in panel settings**.
- This paper **enriches the toolbox** of researchers in dealing with **high-dim. panel models**.
- I develop a **LASSO-based** estimator for a high-dimensional regression model and a **clustered-panel cross-fitting** for valid inference.
- **Unobserved heterogeneity** complicates the inference. I propose a flexible **correlated random effect** approach and **inference using the full sample**.
- I illustrate in a panel data application that **high dimensionality can be hidden** and how proposed approaches allow for a **robustness check**.

Thank you for
listening and comments!

Two-way cluster dependence

- **Assumption AHK** Random elements $W_{it} = (Y_{it}, X_{it}, U_{it})$ are generated by the underlying process:

$$W_{it} = \mu + h(\alpha_i, \gamma_t, \varepsilon_{it}), \quad \forall i \geq 1, t \geq 1,$$

  where $\mu = E[W_{it}]$; h is unknown; vector components $(\alpha_i)_{i\geq 1}$, $(\gamma_t)_{t\geq 1}$, and $(\varepsilon_{it})_{i\geq 1, t\geq 1}$ are mutually independent; α_i is i.i.d across i, ε_{it} is i.i.d across i and t, and γ_t is strictly stationary.
  - Common in cluster-robust inference literature.
- **Assumption AR** (*beta*-mixing of $\{\gamma_t\}_{t>1}$)
  - A generalization of Aldous-Huber-Kallenberg (AHK) representation (Chiang et al., 2024, REStat).

Assumption

For some $s > 1$ and $\delta > 0$,

1. $E[X_{it}' U_{it}] = 0$, $E[\|X_{it}\|^{8(s+\delta)}] < \infty$, $E[\|U_{it}\|^{8(s+\delta)}] < \infty$.
2. *Either* $\Lambda_a \Lambda_a' > 0$ *or* $\Lambda_g \Lambda_g' > 0$, *and* $N/T \to c$ *as* $(N, T) \to \infty$ *for some constant* c.

Back

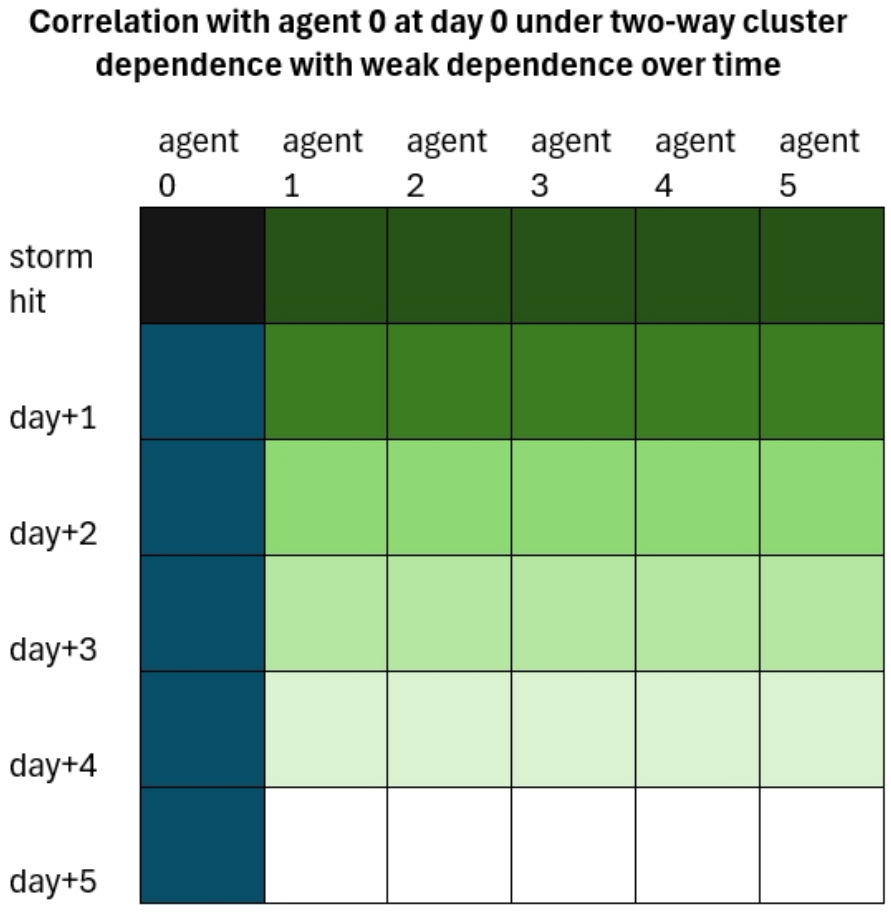
High Dimensionality from Flexible Modeling

- Suppose X is $k \times 1$. Let L^τ be $\tau-$th order polynomial transformation and let r denote the approximation error.
- Then, the high dimensionality is realized as follows:

model	sparse approx.	dim. of unknown param.
$Y = f(X) + U$	no approx.	∞,
$Y = L^\tau(X)\beta + r + U$	$\tau = 2$	$k^2/2 + 3k/2$
$Y = L^\tau(X)\beta + r + U$	$\tau = 3$	$k^3/6 + k^2 + 11k/6$

Back

Graphic illustration of two-way dependence



Back

Two-way cluster dependence

- **Assumption 1** Random elements $W_{it} = (Y_{it}, X_{it}, V_{it})$ are generated by the underlying process:

$$W_{it} = \mu + h(\alpha_i, \gamma_t, \varepsilon_{it}), \quad \forall i \geq 1, t \geq 1,$$

  where $\mu = E[W_{it}]$; h is unknown; vector components $(\alpha_i)_{i\geq 1}$, $(\gamma_t)_{t\geq 1}$, and $(\varepsilon_{it})_{i\geq 1, t\geq 1}$ are mutually independent; α_i is i.i.d across i, ε_{it} is i.i.d across i and t, and γ_t is strictly stationary.
  - Common in cluster-robust inference literature.
- **Assumption 2** (*beta*-mixing of $\{\gamma_t\}_{t>1}$)
  - A generalization of Aldous-Huber-Kallenberg (AHK) representation (Chiang et al., 2024, REStat).

Back

Absolute Regularity

Let $\|\nu\|_{TV}$ denote the total variation norm of a signed measure ν on a measurable space (S, Σ) where Σ is a σ-algebra on S:

$$\|\nu\|_{TV} = \sup_{A \in \Sigma} \nu(A) - \nu(A^c)$$

Define the dependence coefficient of X and Y as:

$$\beta(X, Y) = \frac{1}{2}\|P_{X,Y} - P_X \times P_Y\|_{TV}$$

Assumption (Absolute Regularity of $\{\gamma_t\}_{t \geq 1}$)

The sequence $\{\gamma_t\}_{t \geq 1}$ is beta-mixing at a geometric rate:

$$\beta_\gamma(q) = \sup_{s \leq T} \beta\left(\{\gamma_t\}_{t \leq s}, \{\gamma_t\}_{t \geq s+q}\right) \leq c_\kappa exp(-\kappa q), \forall q \in \mathbb{Z}^+,$$

for some constants $\kappa > 0$ and $c_\kappa \geq 0$.

Back

Assumption (Approximate Sparse Model)

The unknown function f can be well-approximated by a dictionary of transformations $f_{it} = F(X_{it})$ where f_{it} is a $p \times 1$ vector and F is a measurable map, such that

$$f(X_{it}) = f_{it}\zeta_0 + r_{it}$$

where the coefficients ζ_0 and the approximation error r_{it} satisfy

$$\|\zeta_0\|_0 \leq s = o(N \wedge T), \ \|r_{it}\|_{2,NT} \equiv R = O_P\left(\sqrt{\frac{s}{N \wedge T}}\right).$$

Back

My Construction of Weights

- I consider the following choice of penalty level λ and penalty weights ω: for each $j = 1, ..., p$,

$$\lambda = C_\lambda \frac{NT}{(N \wedge T)^{1/2}} \Phi^{-1}\left(1 - \frac{\gamma}{2p}\right),$$

$$\omega_j = \frac{N \wedge V}{N^2} \sum_{i=1}^{N} a_{i,j}^2 + \frac{N \wedge V}{T^2} \sum_{b=1}^{B} \left(\sum_{t \in H_b} g_{t,j} \right)^2 .$$

- Extra Tuning Parameters: C_λ, γ, B.
- Feasible weights: replace $a_{i,j}$ by $\frac{1}{T} \sum_{t=1}^{T} f_{it,j} \widehat{V}_{it}$ and replace $g_{t,j}$ by $\frac{1}{N} \sum_{i=1}^{N} f_{it,j} \widehat{V}_{it}$.

Back

Tuning Parameters

- Tuning parameters for λ: $C_\lambda = O(1)$ and $\gamma = o(1)$. In practice, $C_\lambda = 2$ and $\gamma = \log(p \vee N \vee T)$.
- Tuning parameters for ω: $B = \text{round}(T/h)$, $h = \text{round}(T^{1/5}) + 1$, and $H_b = \{t : h(b-1) + 1 \leq t \leq hb\}$

Back

Valid feasible weights: There exist l, u such that $l\omega_j^{1/2} \leq \widehat{\omega}_j^{1/2} \leq u\omega_j^{1/2}$, uniformly over $j = 1, ..., p$ where $0 < l \leq 1$ and $1 \leq u < \infty$ such that $l \to 1$. Back

- As we allow the dimension of f_{it} to be larger than the sample size, the empirical Gram matrix $M_f = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} f_{it} f_{it}'$ is singular.
- However, it turns out we only need its certain sub-matrices to be non-singular.

Assumption (Sparse Eigenvalues)

For any $C > 0$, there exists constants $0 < \kappa_1 < \kappa_2 < \infty$ such that with probability approaching one as $(N, T) \to \infty$ jointly,

$$\kappa_1 \leq \min_{\delta \in \Delta(m)} \delta' M_f \delta < \max_{\delta \in \Delta(m)} \delta' M_f \delta \leq \kappa_2,$$

where $\Delta(m) = \{\delta : \|\delta\|_0 = m, \|\delta\|_2 = 1\}$.

Back

Assumption (Regularity Conditions)

(i) $\log(p/\gamma) = o\left(T^{1/6}/(\log T)^2\right)$ *and* $p = o\left(T^{7/6}/(\log T)^2\right)$. *(ii) For some* $\mu > 1, \delta > 0$, $\max_{j \leq p} E[|f_{it,j}|^{8(\mu+\delta)}] < \infty$, $E[|V_{it}|^{8\mu+\delta)}] < \infty$. *(iii)* $\min_{j \leq p} E(a_{i,j}^2) > 0$ *and* $\min_{j \leq p} E(g_{t,j}^2) > 0$.

Back

Rate of Convergence in the Oracle Case

- Consider the sample mean estimator $\widehat{\theta} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Y_{it}$. Rewrite the estimator through a Hajek projection:

$$\widehat{\theta} - \theta_0 = \frac{1}{N}\sum_{i=1}^{N} a_i + \frac{1}{T}\sum_{t=1}^{T} g_t + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it},$$

  where $a_i := \mathrm{E}[Y_{it} - \theta_0|\alpha_i]$, $g_t := \mathrm{E}[Y_{it} - \theta_0|\gamma_t]$, and $e_{it} := Y_{it} - \theta_0 - a_i - g_t$.

- Under some regularity conditions, for each j, $\widehat{\theta}_j = O_P\left(\frac{1}{\sqrt{N \wedge T}}\right)$ and $\|\widehat{\theta} - \theta_0\|_2 = O_P\left(\sqrt{\frac{s}{N \wedge T}}\right)$.

Back

Panel-DML: Orthogonalized Moment Condition

- Let $\varphi(W_{it}; \theta, \eta)$ be an identifying moment condition:

$$E[\varphi(W_{it}; \theta_0, \eta_0)] = 0$$

  where W_{it} are random elements; θ are the low-dim. parameters of interest and η are nuisance parameters.

- Let $\psi(W_{it}; \theta, \eta)$ be a corresponding orthogonal moment condition such that

$$\begin{aligned} E[\psi(W; \theta_0, \eta_0)] &= 0, \\ \partial_\eta E[\psi(W; \theta_0, \eta_0)][\eta - \eta_0] &= 0. \end{aligned}$$

Back

Cross Fitting Validity

Lemma (Independent Coupling)

Consider the main sample $W(k,l)$ and auxiliary sample $W(-k,-l)$ for $k = 1, ..., K$ and $l = 1, ..., L$. Suppose Assumptions 1-2 hold for $\{W_{it}\}$. Then, if $\log N/T \to 0$ as $N, T \to \infty$, we can construct $\tilde{W}(k,l)$ and $\tilde{W}(-k,-l)$ such that:

- *They are independent of each other;*
- *They have the same marginal distribution as $W(k,l)$ and $W(-k,-l)$, respectively;*

and
$$Pr\left\{(W(k,l), W(-k,-l)) \neq \left(\tilde{W}(k,l), \tilde{W}(-k,-l)\right), \text{ for some } (k,l)\right\} = o(1)$$

Back

Assumption (Statistical Rates and Score Regularity)

For some positive sequence (Δ_{NT}) *that* $\Delta_{NT} \to 0$, *we have*

(i) *For each* (k, l), *the nuisance estimator* $\widehat{\eta}_{k,l}$ *belongs to the realization set* $\mathcal{T}_{NT}$ *with probability* $1 - \Delta_{NT}$, *where* $\mathcal{T}_{NT}$ *contains* η_0.

(ii) *For all* $i \geq 1$, $t \geq 1$, *and some* $q > 2$, *the following moment conditions hold:*

$$m_{NT} := \sup_{\eta \in \mathcal{T}_{NT}} \left(E_P \left\|\psi(W_{it}; \theta_0, \eta)\right\|^q\right)^{1/q} < \infty, \tag{1}$$

$$m'_{NT} := \sup_{\eta \in \mathcal{T}_{NT}} \left(E_P \left\|\psi^a(W_{it}; \eta)\right\|^q\right)^{1/q} < \infty. \tag{2}$$

Assumption (Statistical Rates and Score Regularity)

(iii) *The following conditions on the statistical rates r_{NT}, r'_{NT}, λ'_{NT} hold for all $i \geq 1$, $t \geq 1$:*

$$r_{NT} := \sup_{\eta \in \mathcal{T}_{NT}} \|E_P[\psi^a(W_{it}; \eta) - \psi^a(W_{it}; \eta_0)]\| \leq \delta_{NT},$$

$$r'_{NT} := \sup_{\eta \in \mathcal{T}_{NT}} \left(E_P \|\psi(W_{it}; \theta_0, \eta) - \psi(W_{it}; \theta_0, \eta_0)\|^2\right)^{1/2} \leq \delta_{NT},$$

$$\lambda'_{NT} := \sup_{r \in (0,1), \eta \in \mathcal{T}_{NT}} \left\|\partial_r^2 E_P[\psi(W_{it}; \theta_0, \eta_0 + r(\eta - \eta_0))]\right\| \leq \delta_{NT}/\sqrt{N}.$$

Back

Assumption (Linear Orthogonal Scores)

For any $P \in \mathcal{P}_{NT}$, the following conditions hold:

(i) $\psi(W;\theta,\eta) = \psi^a(W,\eta)\theta + \psi^b(W,\eta),\ \forall\ W \in \mathcal{W}, \theta \in \Theta, \eta \in \mathcal{T}$.

(ii) $\psi(W;\theta,\eta)$ *satisfy the Neyman orthogonality conditions, or more generally, by a* λ_{NT} *near-orthogonality condition:* $\lambda_{NT} := \sup_{\eta \in \mathcal{T}_{NT}} \|\partial_r E[\psi(W;\theta_0,\eta_0 + r(\eta - \eta_0))]|_{r=0}\| \leq \delta_{NT}/\sqrt{N}$, *where* $\mathcal{T}_{NT} \in \mathcal{T}$ *is a nuisance realization set.*

(iii) *The map* $\eta \to E_P[\psi(W_{it};\theta,\eta)]$ *is twice continuously Gateaux-differentiable on* $\mathcal{T}$.

(iv) *The singular values of the matrix* $A_0 := E_P[\psi^a(W_{it};\eta_0)]$ *are bounded between* a_0 *and* a_1.

(v) *Either* $\lambda_{min}[\Lambda_a\Lambda_a'] > 0$ *or* $\lambda_{min}[\Lambda_g\Lambda_g'] > 0$.

Back

Variance Estimators

$$\widehat{V}_{CHS} = \widehat{A}^{-1}\widehat{\Omega}_{CHS}\widehat{A}^{-1'}, \qquad \widehat{V}_{DKA} = \widehat{A}^{-1}\widehat{\Omega}_{DKA}\widehat{A}^{-1'}$$
$$\widehat{\Omega}_{CHS} = \widehat{\Omega}_A + \widehat{\Omega}_{DK} - \widehat{\Omega}_{NW}, \quad \widehat{\Omega}_{DKA} = \widehat{\Omega}_A + \widehat{\Omega}_{DK}.$$

where $\widehat{A} = \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L}\frac{1}{N_k T_l}\sum_{i\in I_k, s\in S_l}\psi^a(W_{it};\widehat{\eta}_{kl})$, and

$$\widehat{\Omega}_A := \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L}\frac{1}{N_k T_l^2}\sum_{i\in I_k, t\in S_l, r\in S_l}\psi(W_{it};\widehat{\theta},\widehat{\eta}_{kl})\psi(W_{ir};\widehat{\theta},\widehat{\eta}_{kl})',$$

$$\widehat{\Omega}_{DK} := \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L}\frac{K/L}{N_k T_l^2}\sum_{t\in S_l, r\in S_l} k\left(\frac{|t-r|}{M}\right)\sum_{i\in I_k, j\in I_k}\psi(W_{it};\widehat{\theta},\widehat{\eta}_{kl})\psi(W_{jr};\widehat{\theta},\widehat{\eta}_{kl})',$$

$$\widehat{\Omega}_{NW} := \frac{1}{KL}\sum_{k=1}^{K}\sum_{l=1}^{L}\frac{K/L}{N_k T_l^2}\sum_{i\in I_k, t\in S_l, r\in S_l} k\left(\frac{|t-r|}{M}\right)\psi(W_{it};\widehat{\theta},\widehat{\eta}_{kl})\psi(W_{ir};\widehat{\theta},\widehat{\eta}_{kl})'.$$

where $k\left(\frac{m}{M}\right) = 1 - \frac{m}{M}$ is the Bartlett kernel and M is the bandwidth parameter.

Back

CRE approach: the generalized Mundlak device

- A generalized Mundlak device:

$$c_i = h_c(\bar{F}_i, \epsilon_i^c), \tag{3}$$

$$d_t = h_d(\bar{F}_t, \epsilon_t^d), \tag{4}$$

  where $\bar{F}_i = \frac{1}{T}\sum_{t=1}^{T} F_{it}$, $\bar{F}_t = \frac{1}{N}\sum_{i=1}^{N} F_{it}$, $F_{it} := (D_{it}, X_{it}')'$; h_c and h_d are unknown functions; the stochastic errors $(\epsilon_i^c, \epsilon_t^d)$ are each a random draw from some unknown distribution.

- **Not as strong as it seems**: the equivalence exists among within-transformation, fixed-effect, and Mundlak device with POLS in a linear model (Mundlak, 1978, ECTA, Wooldridge, 1997, JBES).
- Generalized by a flexible function. Also see Wooldridge and Zhu, 2020, JBES.

Back

Assumption (Regularity Conditions for the Partial Linear Model)

(i) *A_0 is non-singular.*

(ii) *For any ϵ, $h_c(F, \epsilon)$ and $h_d(F, \epsilon)$ are invertible in F.*

(iii) *For some $\mu > 1, \delta > 0$, $\max_{j \leq p} \mathrm{E}[|f_{it,j}|^{8(\mu+\delta)}] < \infty$ and $\mathrm{E}[|V_{it}^l|^{8(\mu+\delta)}] < \infty$ for $l = g, D, Y, Z$.*

(iv) *Either $\lambda_{min}[\Sigma_a] > 0$ or $\lambda_{min}[\Sigma_g] > 0$, and $\min_{j \leq p} E[a_{i,j}^l]^2 > 0$ and $\min_{j \leq p} E[g_{t,j}^l]^2 > 0$, $l = D, Y, Z$.*

(v) $\log(p/\gamma) = o\left(T^{1/6}/(\log T)^2\right)$ *and* $p = o(T^{7/6}/(\log T)^2)$.

(vi) *The feasible penalty weights $\widehat{\omega}_l$ satisfy the* condition *for $l = D, Y, Z$.*

(viii) *sparse eigenvalues* condition.

Back

Variance estimators using full sample

Theorem

Suppose assumptions for Theorem holds for $P = P_{NT}$ for each (N, T) with $r_{it}^D = r_{it}^Y = 0$ a.s., and $M/T^{1/2} = o(1)$. Then, $(N, T) \to \infty$ and $N/T \to c$ where $0 < c < \infty$,

$$\widehat{V}_{\mathrm{CHS}} = V + o_P(1),$$
$$\widehat{V}_{\mathrm{DKA}} = \widehat{V}_{\mathrm{CHS}} + o_P(1).$$

Back

Babii, A., Ball, R. T., Ghysels, E., and Striaukas, J. (2023). Machine learning panel data regressions with heavy-tailed dependent data: Theory and application. *Journal of Econometrics*, 237(2):105315.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6):2369–2429.

Belloni, A., Chernozhukov, V., Hansen, C., and Kozbur, D. (2016). Inference in high-dimensional panel models with an application to gun control. *Journal of Business & Economic Statistics*, 34(4):590–605.

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.

Chiang, H. D., Hansen, B. E., and Sasaki, Y. (2024). Standard errors for two-way clustering with serially correlated time effects. *Review of Economics and Statistics*, pages 1–40.

Gao, J., Peng, B., and Yan, Y. (2024). Robust inference for high-dimensional panel data models. *Available at SSRN 4825772*.

Mundlak, Y. (1978). On the pooling of cross-section and time-series data. *Econometrica*, 46(69):X6.

Wooldridge, J. M. (1997). Multiplicative panel data models without the strict exogeneity assumption. *Econometric Theory*, 13(5):667–678.

Wooldridge, J. M. and Zhu, Y. (2020). Inference in approximately sparse correlated random effects probit models with panel data. *Journal of Business & Economic Statistics*, 38(1):1–18.