Standard Errors for Two-Way Clustering with Serially Correlated Time Effects

Harold D. Chiang[1] Bruce E. Hansen[1] Yuya Sasaki[2]

[1]University of Wisconsin-Madison

[2]Vanderbilt University

Winter School 2022 University of Delhi

Abstract

- 2-way clustering with
  untruncated serial dependence in time effects
- Robust standard errors
- Supporting theory
- Simulations: superior performance to existing alternatives
- Empirical application: asset pricing model

Introduction

- Panel data: i – firm
  t – time
- Data generating process: $D_{it} = f(\alpha_i, \gamma_t, \varepsilon_{it})$
  - α_i – firm effect
  - γ_t – time effect
  - ε_{it} – idiosyncratic effect
- 2-way clustering by α_i & γ_t.
- γ_t captures omitted macro variables, e.g., business cycle.
  - Reasonable to assume γ_t is serially correlated.
  - Additional source of dependence beyond 2-way clustering.

Relation to the Literature

- 1-way clustering within i (Liang and Zeger, 1986; Arellano, 1987)
- 2-way clustering (Cameron, Gelbach, and Miller, 2011) cited by 3,361
  - No supporting theory.
  - Bootstrap by Menzel (2021).
  - Empirical process by Davezies, D'Haultfœuille, and Guyonvarch (2021).
  - High dimensions by Chiang, Kato, and Sasaki (2021).
- 2-way clustering + M-dependence with uniform weights(Thompson, 2011) cited by 1,546
  - No supporting theory.
  - Citing papers do not account for serial dependence.
- 2-way clustering + untruncated serial dependence (this paper)
  - Supporting theory.

Least Squares Estimation

- $(Y_{it}, X'_{it}) \quad i = 1, \ldots, N$ and $t = 1, \ldots, T$.
  - Y_{it}: scalar
  - X_{it}: $k \times 1$
- Linear regression model

$$Y_{it} = X'_{it}\beta + U_{it} \qquad E[U_{it}|X_{it}] = 0.$$

- Least squares

$$\hat{\beta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} X'_{it} \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} Y_{it} \right).$$

- Least squares residuals

$$\hat{U}_{it} = Y_{it} - X'_{it}\hat{\beta}.$$

Variance of $\hat{\beta}$

$V_{NT} = var(\hat{\beta}) = \hat{Q}^{-1}\Omega_{NT}\hat{Q}^{-1}$, where $\hat{Q} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}'$ and

$$\begin{aligned}
\Omega_{NT} &= \frac{1}{(NT)^2}\sum_{i=1}^{N} E\left[R_i R_i'\right] && \left(R_i = \sum_{t=1}^{T} X_{it}U_{it}\right)\\
&+ \frac{1}{(NT)^2}\sum_{t=1}^{T} E\left[S_t S_t'\right] && \left(S_t = \sum_{i=1}^{N} X_{it}U_{it}\right)\\
&- \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T} E[X_{it}X_{it}'U_{it}^2]\\
&+ \frac{1}{(NT)^2}\sum_{m=1}^{T-1} E\left[G_m + G_m'\right] && \left(G_m = \sum_{t=1}^{T-m} S_t S_{t+m}'\right)\\
&- \frac{1}{(NT)^2}\sum_{m=1}^{T-1} E\left[H_m + H_m'\right] && \left(H_m = \sum_{i=1}^{N}\sum_{t=1}^{T-m} X_{it}U_{it}X_{i,t+m}'U_{i,t+m}\right)
\end{aligned}$$

Interpreting the Decomposition

$$\Omega_{NT} = \underbrace{\frac{1}{(NT)^2}\sum_{i=1}^{N} E\left[R_i R_i'\right]}_{\text{Variance of firm sums}} \qquad \underbrace{\left(R_i = \sum_{t=1}^{T} X_{it}U_{it}\right)}_{i\text{-th firm cluster}}$$

$$+ \underbrace{\frac{1}{(NT)^2}\sum_{t=1}^{T} E\left[S_t S_t'\right]}_{\text{Variance of time sums}} \qquad \underbrace{\left(S_t = \sum_{i=1}^{N} X_{it}U_{it}\right)}_{t\text{-th time cluster}}$$

$$-\frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T} E[X_{it}X_{it}'U_{it}^2]$$

$$+ \underbrace{\frac{1}{(NT)^2}\sum_{m=1}^{T-1} E\left[G_m + G_m'\right]}_{\text{Autocovariance of time sums}} \qquad \left(G_m = \sum_{t=1}^{T-m} S_t S_{t+m}'\right)$$

$$-\frac{1}{(NT)^2}\sum_{m=1}^{T-1} E\left[H_m + H_m'\right] \qquad \left(H_m = \sum_{i=1}^{N}\sum_{t=1}^{T-m} X_{it}U_{it}X_{i,t+m}'U_{i,t+m}\right)$$

Variance Estimation

We propose $\hat{V}_{NT} = \hat{Q}^{-1}\hat{\Omega}_{NT}\hat{Q}^{-1}$, where

$$
\begin{aligned}
\hat{\Omega}_{NT} = & \frac{1}{(NT)^2}\sum_{i=1}^{N}\hat{R}_i\hat{R}_i' \\
& + \frac{1}{(NT)^2}\sum_{t=1}^{T}\hat{S}_t\hat{S}_t' \\
& - \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'\hat{U}_{it}^2 \\
& + \frac{1}{(NT)^2}\sum_{m=1}^{M} w(m, M)\left(\hat{G}_m + \hat{G}_m' - \hat{H}_m - \hat{H}_m'\right)
\end{aligned}
$$

– $\hat{R}_i$, $\hat{S}_t$, $\hat{G}_m$ & $\hat{H}_m$ replace U_{it} by $\hat{U}_{it}$ in R_i, S_t, G_m & H_m, respectively.

– $w(\,\cdot\,, M)$ is a weight function, and M is a tuning parameter.

Comparison with Existing Variance Estimators 1

- Liang-Zeger-Arellano Estimator:

$$\hat{\Omega}_{NT}^{\mathrm{LZA}} = \frac{1}{(NT)^2}\sum_{i=1}^{N}\hat{R}_i\hat{R}_i'$$

$$\cancel{+\frac{1}{(NT)^2}\sum_{t=1}^{T}\hat{S}_t\hat{S}_t'}$$

$$\cancel{-\frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'\hat{U}_{it}^2}$$

$$\cancel{+\frac{1}{(NT)^2}\sum_{m=1}^{M}w(m,M)\left(\hat{G}_m+\hat{G}_m'-\hat{H}_m-\hat{H}_m'\right)}$$

Comparison with Existing Variance Estimators 2

- "Cluster within t" Estimator:

$$
\begin{aligned}
\hat{\Omega}_{NT}^{\text{Cw}t} = & \frac{1}{(NT)^2} \sum_{i=1}^{N} \hat{R}_i \hat{R}_i' \\
& + \frac{1}{(NT)^2} \sum_{t=1}^{T} \hat{S}_t \hat{S}_t' \\
& - \frac{1}{(NT)^2} \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} X_{it}' \hat{U}_{it}^2 \\
& + \frac{1}{(NT)^2} \sum_{m=1}^{M} w(m, M) \left(\hat{G}_m + \hat{G}_m' - \hat{H}_m - \hat{H}_m' \right)
\end{aligned}
$$

Comparison with Existing Variance Estimators 3

- Cameron-Gelbach-Miller Estimator:

$$\hat{\Omega}_{NT}^{\text{CGM}} = \frac{1}{(NT)^2}\sum_{i=1}^{N}\hat{R}_i\hat{R}_i' + \frac{1}{(NT)^2}\sum_{t=1}^{T}\hat{S}_t\hat{S}_t' - \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'\hat{U}_{it}^2 + \frac{1}{(NT)^2}\sum_{m=1}^{M}w(m,M)\left(\hat{G}_m + \hat{G}_m' - \hat{H}_m - \hat{H}_m'\right)$$

Comparison with Existing Variance Estimators 4

- Thompson's Estimator:

$$\hat{\Omega}^{\mathrm{T}}_{NT} = \frac{1}{(NT)^2}\sum_{i=1}^{N}\hat{R}_i\hat{R}_i'$$

$$+\frac{1}{(NT)^2}\sum_{t=1}^{T}\hat{S}_t\hat{S}_t'$$

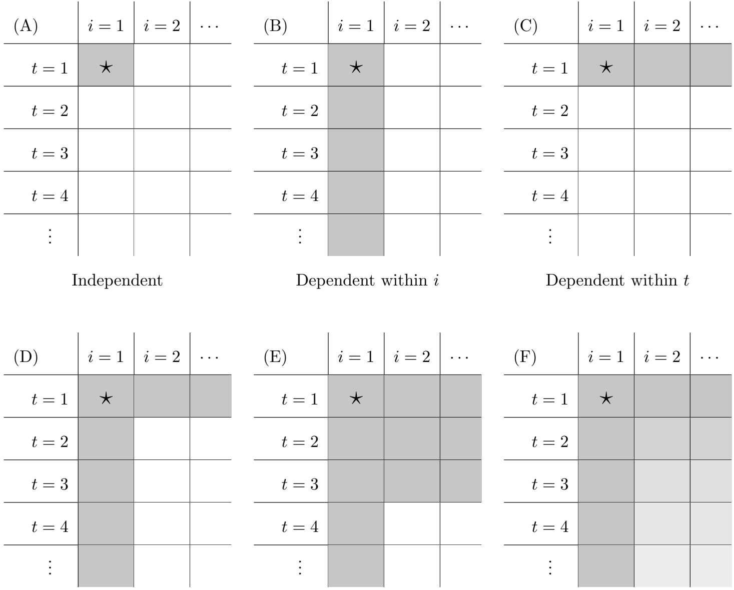
$$-\frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T}X_{it}X_{it}'\hat{U}_{it}^2$$

$$\cancel{+\frac{1}{(NT)^2}\sum_{m=1}^{M}w(m,M)\left(\hat{G}_m+\hat{G}_m'-\hat{H}_m-\hat{H}_m'\right)}$$

$$+\frac{1}{(NT)^2}\sum_{m=1}^{M}\left(\hat{G}_m+\hat{G}_m'-\hat{H}_m-\hat{H}_m'\right)$$

where M is fixed by model assumption, e.g., $M = 2$ under 2-dependence.

Alternative Models of Dependence



(A) Independent; (B) Dependent within i; (C) Dependent within t; (D) Two-way Dependent; (E) 2-Dependent Time Effects; (F) Untruncated Dependence

Correlated Time Effects

Two key variables in a standard market value equation:

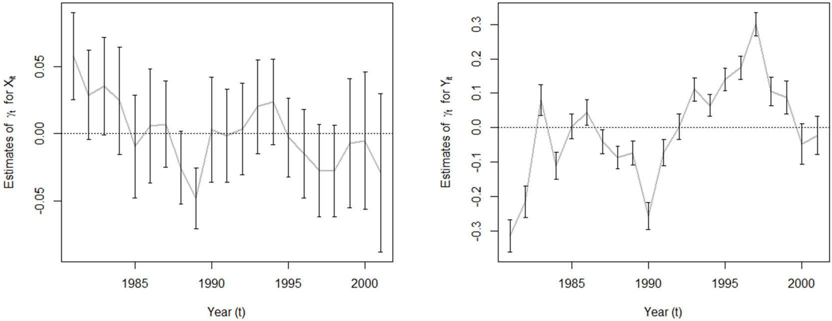


Figure: Estimates of the common time effects for log R&D stock (left) and log Tobin's average Q (right). The vertical lines indicate pointwise 95% confidence intervals.

Correlated Time Effects

It is useful to consider the model of the form $D_{it} = \alpha_i + \gamma_t + \varepsilon_{it}$. Two key variables in a standard market value equation:

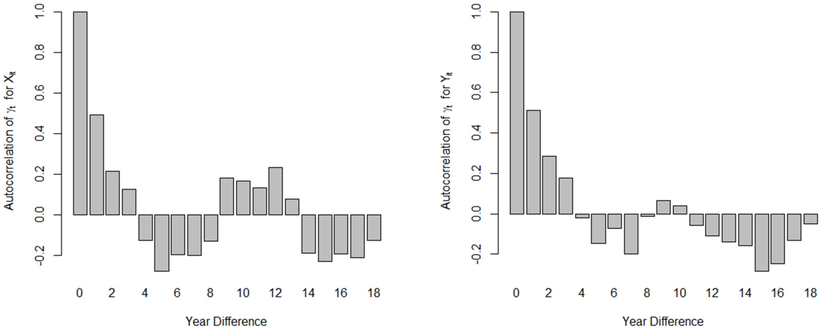


Figure: Autocorrelograms of the common time effects for log R&D stock (left) and log Tobin's average Q (right).

Our Proposed Variance Estimator, Once Again

$V_{NT} = \hat{Q}^{-1}\hat{\Omega}_{NT}\hat{Q}^{-1}$, where $\hat{Q} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}'$ and

$$\hat{\Omega}_{NT} = \frac{1}{(NT)^2}\sum_{i=1}^{N}\hat{R}_i\hat{R}_i' \qquad \left(\hat{R}_i = \sum_{t=1}^{T} X_{it}\hat{U}_{it}\right)$$

$$+ \frac{1}{(NT)^2}\sum_{t=1}^{T}\hat{S}_t\hat{S}_t' \qquad \left(\hat{S}_t = \sum_{i=1}^{N} X_{it}\hat{U}_{it}\right)$$

$$- \frac{1}{(NT)^2}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{it}X_{it}'\hat{U}_{it}^2$$

$$+ \frac{1}{(NT)^2}\sum_{m=1}^{M} w(m,M)\left(\hat{G}_m + \hat{G}_m'\right) \qquad \left(\hat{G}_m = \sum_{t=1}^{T-m}\hat{S}_t\hat{S}_{t+m}'\right)$$

$$- \frac{1}{(NT)^2}\sum_{m=1}^{M} w(m,M)\left(\hat{H}_m + \hat{H}_m'\right) \qquad \left(\hat{H}_m = \sum_{i=1}^{N}\sum_{t=1}^{T-m} X_{it}\hat{U}_{it}X_{i,t+m}'\hat{U}_{i,t+m}\right),$$

where we recommend the triangular weights $w(m,M) = 1 - m/(M+1)$.

Choosing the Tuning Parameter M

- Andrews (1991): minimizes the asymptotic MSE of the time-series variance estimator.
- We adapt Andrews (1991) by treating a time sum as a unit of observation in time series.
- Define the time-sums $S_{jt} = \sum_{i=1}^{n} X_{jit}\hat{U}_{it}$.
- Obtain the OLS $\hat{\rho}$ in the AR(1) equation $S_{jt} = \hat{\rho}_j S_{j,t-1} + \hat{e}_{jt}$.
- Under the triangular weights, we use

$$\hat{M} = 1.8171 \cdot \left(\frac{\sum_{j=1}^{k} \frac{\hat{\rho}_j^2}{(1-\hat{\rho}_j)^4}}{\sum_{j=1}^{k} \frac{\left(1-\hat{\rho}_j^2\right)^2}{(1-\hat{\rho}_j)^4}} \right)^{1/3} T^{1/3}.$$

Basic Setup

- Panel data: $\{D_{it} : 1 \leq i \leq N; 1 \leq t \leq T\}$
  - e.g., for linear regression model: $D_{it} = (Y_{it}, X'_{it}, U_{it})'$
- Data generating process:

$$D_{it} = f(\alpha_i, \gamma_t, \varepsilon_{it}),$$

  where
  - $\{\alpha_i\}$, $\{\gamma_t\}$ and $\{\varepsilon_{it}\}$ mutually independent,
  - α_i is i.i.d. across i.
  - ε_{it} is i.i.d. across (i, t).
  - γ_t is a strictly stationary serially correlated process.
- Generalization of Aldous-Hoover-Kallenberg representation

Estimation of the Mean

- First consider $D_{it} = X_{it}$.
- Estimand: $\theta = E[X_{it}]$.
  - W.O.L.G., $\theta = E[X_{it}] = 0$.
- Sample mean estimator: $\hat{\theta} = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} X_{it}$.
- Hoeffding-type decomposition

$$X_{it} = a_i + b_t + e_{it},$$

where

- $a_i = E[X_{it} \mid \alpha_i]$
- $b_t = E[X_{it} \mid \gamma_t]$
- $e_{it} = X_{it} - a_i - b_t$

Estimation of the Mean

- Hoeffding-type decomposition

$$X_{it} = a_i + b_t + e_{it}.$$

- Averages:

$$\hat{\theta} = \frac{1}{N}\sum_{i=1}^{N} a_i + \frac{1}{T}\sum_{t=1}^{T} b_t + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}.$$

Assumption

For some $r > 1$ and $\delta > 0$, **(i)** $X_{it} = f(\alpha_i, \gamma_t, \varepsilon_{it})$ where $\{\alpha_i\}$, $\{\gamma_t\}$, and $\{\varepsilon_{it}\}$ are mutually independent sequences, α_i is i.i.d across i, ε_{it} is i.i.d across (i, t), and γ_t is strictly stationary.

(ii) $E[||X_{it}||^{4(r+\delta)}] < \infty$.*

(iii) γ_t is an α-mixing sequence with size $2r/(r-1)$, that is, $\alpha(\ell) = O(\ell^{-\lambda})$ for a $\lambda > 2r/(r-1)$.*

* cf. Newey and West (1987) and Hansen (1992)

Variance of the Sample Mean

- Define

$$\Sigma_a = E[a_i a_i']$$
$$\Sigma_b = \sum_{\ell=-\infty}^{\infty} E[b_t b_{t+\ell}']$$
$$\Sigma_e = \sum_{\ell=-\infty}^{\infty} E[e_{it} e_{i,t+\ell}']$$

Theorem

Suppose that the assumption holds. Then $||\Sigma_a|| < \infty$, $||\Sigma_b|| < \infty$, *and* $||\Sigma_e|| < \infty$, *and as* $(N, T) \to \infty$,

$$var(\hat{\theta}) = \frac{1}{N}\Sigma_a + \frac{1}{T}\Sigma_b\left(1 + o(1)\right) + \frac{1}{NT}\Sigma_e\left(1 + o(1)\right).$$

Furthermore, $\hat{\theta} \xrightarrow{p} \theta$ *as* $N, T \to \infty$.

Convergence Rates of the Sample Mean

$$var(\hat{\theta}) = \frac{1}{N}\Sigma_a + \frac{1}{T}\Sigma_b\left(1 + o(1)\right) + \frac{1}{NT}\Sigma_e\left(1 + o(1)\right).$$

- Case: $\Sigma_a > 0$ or $\Sigma_b > 0$
  - non-degenerate firm or time effect.
  - $\frac{1}{NT}\Sigma_e\left(1 + o(1)\right) \ll_p \frac{1}{N}\Sigma_a + \frac{1}{T}\Sigma_b\left(1 + o(1)\right)$.
- Case: X_{it} is i.i.d.
  - $\Sigma_a = 0$ and $\Sigma_b = 0$.
  - $var(\hat{\theta}) = \frac{1}{NT}\Sigma_e$

Asymptotic Distribution for the Sample Mean

Assumption

One of the the following two conditions holds.

(i) Either $\Sigma_a > 0$ or $\Sigma_b > 0$, and $N/T \to c \in (0, \infty)$ as $(N, T) \to \infty$.

or

(ii) X_{it} are independent and identically distributed across i and t, and $var(X_{it}) > 0$.

Theorem

Suppose that the assumptions hold. Then

$$var(\hat{\theta})^{-1/2}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I_m).$$

- The self-normalization accommodates differing convergence rates – robustness against unknown clustering structures.

Linear Regression

- Linear regression model

$$
\begin{aligned}
Y_{it} &= X_{it}'\beta + U_{it} \\
E[U_{it}|X_{it}] &= 0
\end{aligned}
$$

- Least squares

$$
\hat{\beta} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} X_{it}' \right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{T} X_{it} Y_{it} \right).
$$

- Set $D_{it} = (Y_{it}, X_{it}', U_{it})'$, so

$$
(Y_{it}, X_{it}', U_{it})' = f(\alpha_i, \gamma_t, \varepsilon_{it}).
$$

- We now let $a_i = E[X_{it}U_{it} \mid \alpha_i]$ and $b_t = E[X_{it}U_{it} \mid \gamma_t]$.
(Previously, $a_i = E[X_{it} \mid \alpha_i]$ and $b_t = E[X_{it} \mid \gamma_t]$.)

Linear Regression

Assumption

For some $\delta > 0$ and $r > 1$, **(i)** $\{(Y_{it}, X'_{it}, U_{it}) : 1 \leq i \leq N, 1 \leq t \leq T\}$ are generated following $(Y_{it}, X'_{it}, U_{it})' = f(\alpha_i, \gamma_t, \varepsilon_{it})$, where $\{\alpha_i\}$, $\{\gamma_t\}$, and $\{\varepsilon_{it}\}$ are mutually independent sequences, α_i is i.i.d across i, ε_{it} is i.i.d across (i, t), & γ_t is strictly stationary.

(ii) $Q = E[X_{it}X'_{it}] > 0$, $E[\|X_{it}\|^{8(r+\delta)}] < \infty$, & $E[\|U_{it}\|^{8(r+\delta)}] < \infty$.

(iii) γ_t is a β-mixing sequence with size $2r/(r-1)$, that is, $\beta(\ell) = O(\ell^{-\lambda})$ for a $\lambda > 2r/(r-1)$.

(iv) One of the following two conditions hold: **(1)** Either $\Sigma_a > 0$ or $\Sigma_b > 0$, and $N/T \to c \in (0, \infty)$ as $(N, T) \to \infty$; or **(2)** (X_{it}, U_{it}) are independent and identically distributed across i and t, & $var(X_{it}U_{it}) > 0$.

(v) For each $M \geq 1$ and $1 \leq m \leq M$, $w(m, M) = 1 - [m/(M+1)]$.

(vi) $M/\min\{N, T\}^{1/2} = o(1)$.

Asymptotic Variances and Variance Estimators

- Asymptotic variance of $\hat{\beta}$: $V_{NT} = Q^{-1}\Omega_{NT}Q^{-1}$.
- Variance estimator for $\hat{\beta}$: $\hat{V}_{NT} = \hat{Q}^{-1}\hat{\Omega}_{NT}\hat{Q}^{-1}$.

- Linear transformation: $\theta = R'\beta$.
- Estimator of θ: $\hat{\theta} = R'\hat{\beta}$.
- Asymptotic variance of $\hat{\theta}$: $\Sigma_{NT} = R'V_{NT}R$.
- Variance estimator for $\hat{\theta}$: $\hat{\Sigma}_{NT} = R'\hat{V}_{NT}R$.

The Main Theoretical Results

Theorem

If the assumption holds, then

$$\Sigma_{NT}^{-1}\hat{\Sigma}_{NT} \xrightarrow{p} I_k.$$

Theorem

If the assumption holds, then

$$\Sigma_{NT}^{-1/2}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I_m)$$

and

$$\hat{\Sigma}_{NT}^{-1/2}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I_m).$$

Simulation Design

- Linear model:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + U_{it},$$

- The right-hand side variables $(X_{it}, U_{it})'$ are generated through

$$X_{it} = w_\alpha \alpha_i^x + w_\gamma \gamma_t^x + w_\varepsilon \varepsilon_{it}^x \qquad \text{and}$$
$$U_{it} = w_\alpha \alpha_i^u + w_\gamma \gamma_t^u + w_\varepsilon \varepsilon_{it}^u.$$

- Set $(\beta_0, \beta_1) = (1, 1)$ throughout.
- Vary the weights $(w_\alpha, w_\gamma, w_\varepsilon)$ across sets of simulations.
- 10,000 MC iterations.

Simulation Design, Continued

- $(\alpha_i^x, \alpha_i^u, \varepsilon_{it}^x, \varepsilon_{it}^u)$ are all mutually independent $N(0,1)$.
- (γ_t^x, γ_t^u) – AR design:

$$\gamma_t^x = \rho\gamma_{t-1}^x + \tilde{\gamma}_t^x \text{ where } \tilde{\gamma}_t^x \text{ are independent from } N(0, 1-\rho^2),$$
$$\gamma_t^u = \rho\gamma_{t-1}^u + \tilde{\gamma}_t^u \text{ where } \tilde{\gamma}_t^u \text{ are independent from } N(0, 1-\rho^2),$$

  where $\rho = 0.425$.
- (γ_t^x, γ_t^u) – MA design:

$$\gamma_t^x = \sum_{h=0}^{5} \varphi_h \tilde{\gamma}_{t-h}^x \text{ where } \tilde{\gamma}_t^x \text{ are independent from } N(0,1),$$
$$\gamma_t^u = \sum_{h=0}^{5} \varphi_h \tilde{\gamma}_{t-h}^u \text{ where } \tilde{\gamma}_t^u \text{ are independent from } N(0,1),$$

  where $(\varphi_0, \varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_t) = (0.5, 0.1, 0.1, 0.1, 0.1, 0.1)$.

Comparing Alternative Estimators

- EHW – Eicker-Huber-White; HC0.
- CRi – 1-way CR within i; (Liang and Zeger, 1986; Arellano, 1987)
- CRt – 1-way CR within t.
- CGM – 2-way CR (Cameron et al., 2011).
- M – 2-way CR bootstrap (Menzel, 2021).
- T – 2-way CR with 2-dependence (Thompson, 2011).
- CHS – This paper (Chiang, Hansen, Sasaki).

I.I.D. Design: Nominal Probability = 95%

	N	T	w_α	w_γ	w_ε	EHW	CRi	CRt	CGM	M	T	CHS
(I)	25	25	0.00	0.00	1.00	0.95	0.93	0.93	0.91	1.00	0.78	0.90
(II)	50	50	0.00	0.00	1.00	0.95	0.94	0.94	0.93	1.00	0.87	0.94
(III)	75	75	0.00	0.00	1.00	0.95	0.94	0.94	0.94	1.00	0.91	0.95

AR Design: Nominal Probability = 95%

	N	T	w_α	w_γ	w_ε	EHW	CRi	CRt	CGM	M	T	CHS
(IV)-AR	25	25	0.50	0.25	0.25	0.40	0.89	0.38	0.89	0.91	0.89	0.90
(V)-AR	50	50	0.50	0.25	0.25	0.30	0.91	0.37	0.92	0.93	0.92	0.93
(VI)-AR	75	75	0.50	0.25	0.25	0.25	0.93	0.36	0.93	0.94	0.94	0.94
(VII)-AR	25	25	0.25	0.25	0.50	0.78	0.85	0.84	0.89	0.98	0.84	0.92
(VIII)-AR	50	50	0.25	0.25	0.50	0.66	0.84	0.83	0.91	0.97	0.91	0.94
(IX)-AR	75	75	0.25	0.25	0.50	0.59	0.83	0.82	0.92	0.96	0.93	0.94
(X)-AR	25	25	0.25	0.50	0.25	0.37	0.39	0.84	0.84	0.88	0.79	0.89
(XI)-AR	50	50	0.25	0.50	0.25	0.27	0.35	0.86	0.87	0.88	0.88	0.91
(XII)-AR	75	75	0.25	0.50	0.25	0.23	0.35	0.88	0.89	0.89	0.90	0.92

MA Design: Nominal Probability = 95%

	N	T	w_α	w_γ	w_ε	EHW	CRi	CRt	CGM	M	T	CHS
(IV)-MA	25	25	0.50	0.25	0.25	0.37	0.90	0.25	0.89	0.91	0.89	0.90
(V)-MA	50	50	0.50	0.25	0.25	0.28	0.93	0.20	0.93	0.93	0.93	0.93
(VI)-MA	75	75	0.50	0.25	0.25	0.23	0.93	0.17	0.93	0.93	0.93	0.93
(VII)-MA	25	25	0.25	0.25	0.50	0.84	0.92	0.82	0.91	0.99	0.86	0.92
(VIII)-MA	50	50	0.25	0.25	0.50	0.75	0.93	0.76	0.93	0.99	0.93	0.95
(IX)-MA	75	75	0.25	0.25	0.50	0.67	0.92	0.70	0.93	0.98	0.93	0.95
(X)-MA	25	25	0.25	0.50	0.25	0.52	0.74	0.77	0.87	0.92	0.84	0.90
(XI)-MA	50	50	0.25	0.50	0.25	0.39	0.73	0.78	0.90	0.92	0.90	0.93
(XII)-MA	75	75	0.25	0.50	0.25	0.32	0.73	0.79	0.90	0.91	0.91	0.93

Asset Pricing Model

- Fama-French three-factor model:

$$R_{it} - R_{ft} = \beta_1(R_{Mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + e_{it},$$

where

  - R_{it}: total return of portfolio/stock i in month t
  - R_{ft}: risk-free rate of return in month t
  - R_{Mt}: total market portfolio return in month t
  - SMB_t: size premium (small−big)
  - HML_t: value premium (high−low)
  - $\beta = (\beta_1, \beta_2, \beta_3)'$: factor coefficients

Data

- Monthly: January 2000 – December 2009 ($T = 119$)
- Returns R_{it}:
  - (A) 44 industry portfolios excluding financial sectors ($N = 44$)
  - (B) individual stocks ($N = 779$)
- Risk-free rates R_{ft}: monthly 30-day T-bill yields
- Fixed effects:
  - $\ddot{Y}_{it}$ = within-transformation of $R_{it} - R_{ft}$
  - $\ddot{X}_{it}$ = within-transformation of $(R_{Mt} - R_{ft}, SMB_t, HML_t)'$

Estimation

- Within estimator:

$$\hat{\beta} = \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{X}_{it} \ddot{X}_{it}' \right)^{-1} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{X}_{it} \ddot{Y}_{it}$$

- Variance estimator: $\hat{V}_{NT} = \hat{Q}^{-1} \hat{\Omega}_{NT} \hat{Q}^{-1}$ with
  - Y_{it} replaced by $\ddot{Y}_{it}$
  - X_{it} replaced by $\ddot{X}_{it}$
  - $\hat{U}_{it}$ replaced by $\ddot{Y}_{it} - \ddot{X}_{it} \hat{\beta}$

(A) 44 Industry Portfolios. $(N, T) = (44, 119)$.

		Standard Errors						
	$\hat{\beta}$	EHW	CRi	CRt	CGM	M	T	CHS
MKT	0.959	0.022	0.055	0.030	0.059	0.021	0.059	0.059
SMB	0.076	0.029	0.035	0.041	0.045	0.029	0.056	0.052
HML	0.358	0.030	0.066	0.049	0.076	0.028	0.082	0.079

(B) Individual Stocks. $(N, T) = (779, 119)$.

		Standard Errors						
	$\hat{\beta}$	EHW	CRi	CRt	CGM	M	T	CHS
MKT	1.157	0.012	0.021	0.033	0.037	0.035	0.036	0.034
SMB	0.474	0.020	0.025	0.051	0.053	0.055	0.068	0.070
HML	0.173	0.016	0.029	0.053	0.058	0.054	0.055	0.053

Summary

- 2-way clustering with
  untruncated serial dependence in time effects
- Robust standard errors
- Supporting theory
- Simulations: superior performance to existing alternatives
- Empirical application: asset pricing model

Thank you!

Harold D. Chiang, Bruce E. Hansen, & Yuya Sasaki

References

ARELLANO, M. (1987): "Computing robust standard errors for within-groups estimators," *Oxford Bulletin of Economics and Statistics*, 49, 431–434.

CAMERON, C. A., J. B. GELBACH, AND D. L. MILLER (2011): "Robust inference with multiway clustering," *Journal of Business and Economic Statistics*, 29, 238–249.

CHIANG, H. D., K. KATO, AND Y. SASAKI (2021): "Inference for high-dimensional exchangeable arrays," *Journal of the American Statistical Association*, forthcoming.

DAVEZIES, L., X. D'HAULTFŒUILLE, AND Y. GUYONVARCH (2021): "Empirical process results for exchangeable arrays," *Annals of Statistics*, 49, 845–862.

HANSEN, B. E. (1992): "Consistent covariance matrix estimation for dependent heterogeneous processes," *Econometrica*, 967–972.

LIANG, K.-Y. AND S. L. ZEGER (1986): "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13–22.

MENZEL, K. (2021): "Bootstrap with cluster-dependence in two or more dimensions," *Econometrica*, forthcoming.

NEWEY, W. K. AND K. D. WEST (1987): "A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix," *Econometrica*, 55, 703.

THOMPSON, S. B. (2011): "Simple formulas for standard errors that cluster by both firm and time," *Journal of Financial Economics*, 99, 1–10.