Supplement to "Linear regression with many controls of limited explanatory power"

(*Quantitative Economics*, Vol. 12, No. 2, May 2021, 405–442)

Chenchuan (Mark) Li
Department of Economics, Princeton University

Ulrich K. Müller
Department of Economics, Princeton University

S.1. Additional results in heteroskedastic models

Consider the linear model (1) with nonstochastic regressors, so that in vector form

$$\begin{aligned}\mathbf{y} &= \mathbf{x}\beta + \mathbf{Q}\boldsymbol{\delta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \\ &= \mathbf{R}\boldsymbol{\alpha} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},\end{aligned}$$

where $\mathbf{R} = (\mathbf{x}, \mathbf{Q})$, $\boldsymbol{\alpha} = (\beta, \boldsymbol{\delta}')'$, and as in the main text, $\mathbf{Q}'\mathbf{x} = \mathbf{0}$ and $\mathbf{Q}'\mathbf{Z} = \mathbf{0}$.

We consider a set-up with $M \to \infty$ clusters of not necessarily equal size. Write

$$\mathbf{y}_j = \mathbf{R}_j\boldsymbol{\alpha} + \mathbf{Z}_j\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_j$$

for the observations in the jth cluster (so that the sum of the lengths of the $\mathbf{y}_j$ vectors over $j = 1, \dots, M$ equals n, and n is implicitly a function of M). We allow the sequence of regressors $\mathbf{R}$ and $\mathbf{Z}$, the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, the number of observations per cluster, and the distribution of $\boldsymbol{\varepsilon}_j$ to depend on M in a double array fashion. In particular, this allows for the number of regressors p and/or m to be proportional to the sample size. To ease notation, we do no make this dependence on M explicit.

Define the $n \times 2$ matrix $\mathbf{v} = (\mathbf{v}_1', \dots, \mathbf{v}_M')'$. Let $\|\cdot\|$ be the spectral norm.

Condition 1. (a) $\boldsymbol{\varepsilon}_j$, $j = 1, \dots, M$ are independent with $E[\boldsymbol{\varepsilon}_j] = 0$ and $E[\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j'] = \boldsymbol{\Sigma}_j$.

(b) $\|(M^{-1}\sum_{j=1}^M \mathbf{v}_j'\boldsymbol{\Sigma}_j\mathbf{v}_j)^{-1}\| = O(1)$, $\max_j \|\mathbf{v}_j\|^4 \cdot \sum_{j=1}^M E[\|\boldsymbol{\varepsilon}_j\|^4] = o(M^2)$.

(c) $\|M^{-1}\sum_{j=1}^M \mathbf{R}_j\mathbf{R}_j'\| = O(1)$, $\|(M^{-1}\sum_{j=1}^M \mathbf{R}_j\mathbf{R}_j')^{-1}\| = O(1)$, $\max_j \|\boldsymbol{\Sigma}_j\| = o(M)$, $\max_j \|\boldsymbol{\Sigma}_j\| \cdot \max_j \|\mathbf{v}_j\|^4 = O(M)$, and $\max_j \|\mathbf{v}_j\|^2 = O(M)$.

(d) $\max_j \|\mathbf{v}_j\|^2 \cdot \kappa^2 = o(M/n)$ and $\max_j \|\boldsymbol{\Sigma}_j\| \cdot \max_j \|\mathbf{v}_j\|^4 \cdot \kappa^2 = o(M^2/n)$, where $\kappa^2 = \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\gamma}/n$.

Chenchuan (Mark) Li: chenchuanli@gmail.com
Ulrich K. Müller: umueller@princeton.edu

TABLE S.1. Properties of 95% Armstrong and Kolsar inference.

Panel A: Weighted expected length of CI for b under $d \sim U[-\bar{k}, \bar{k}]$

	AK($\bar{k}$) Interval					Lower Bound				
$\rho \backslash \bar{k}$	0	1	3	10	30	0	1	3	10	30
0.50	3.9	4.2	4.5	4.5	4.5	3.9	4.2	4.4	4.5	4.5
0.90	3.9	5.0	7.4	9.0	9.0	3.9	5.0	6.9	8.2	8.7
0.99	3.9	5.2	9.1	21.5	27.8	3.9	5.2	8.9	17.4	23.9

Panel B: Expected length of CI for b, maximized over $|d| \leq \bar{k}$

	AK($\bar{k}$) Interval					Lower Bound				
$\rho \backslash \bar{k}$	0	1	3	10	30	0	1	3	10	30
0.50	3.9	4.2	4.5	4.5	4.5	3.9	4.2	4.4	4.5	4.5
0.90	3.9	5.0	7.4	9.0	9.0	3.9	5.0	7.1	8.4	8.9
0.99	3.9	5.2	9.1	21.5	27.8	3.9	5.2	8.9	18.4	25.1

Panel C: Ratio of expected length of AK CI for b relative to long regression interval

	Minimized over $|d| \leq \bar{k}$					Maximized over $|d| \leq \bar{k}$				
$\rho \backslash \bar{k}$	0	1	3	10	30	0	1	3	10	30
0.50	0.87	0.92	0.98	1.00	1.00	.87	0.92	0.98	1.00	1.00
0.90	0.44	0.55	0.83	1.00	1.00	0.44	0.55	0.83	1.00	1.00
0.99	0.14	0.19	0.33	0.77	1.00	0.14	0.19	0.33	0.77	1.00

Panel D: Median of $\bar{k}^*_\phi$ under $b = 0$, $P(d = d_0) = P(d = -d_0) = 1/2$

	$\bar{k}^*_{\text{AK}}$					Upper Bound				
$\rho \backslash d_0$	0	1	3	10	30	0	1	3	10	30
0.50	0.0	0.0	0.0	0.7	1.5	0.0	0.0	0.7	4.2	14.3
0.90	0.0	0.0	0.8	2.9	4.7	0.0	0.0	1.2	7.6	25.8
0.99	0.0	0.0	1.3	7.6	11.7	0.0	0.0	1.4	8.4	28.4

Panel E: Weighted average MSE of equivariant estimators of b under $d \sim U[-\bar{k}, \bar{k}]$

	$\hat{b}_{\text{AK}}$					Lower Bound				
$\rho \backslash \bar{k}$	0	1	3	10	30	0	1	3	10	30
0.50	1.00	1.09	1.26	1.33	1.33	1.00	1.07	1.22	1.30	1.32
0.90	1.00	1.27	2.78	5.26	5.26	1.00	1.25	2.53	4.38	4.97
0.99	1.00	1.33	3.79	22.8	50.2	1.00	1.32	3.77	20.4	39.7

Note: See Table 1.

THEOREM 3. (a) *Under Condition* 1(a) *and* (b), *as* $M \to \infty$,

$$\boldsymbol{\Omega}_n^{-1/2} M^{-1} \sum_{j=1}^{M} \mathbf{v}_j' \mathbf{y}_j \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{I}_2),$$

where $\boldsymbol{\Omega}_n = M^{-2} \sum_{j=1}^{M} \mathbf{v}_j' \boldsymbol{\Sigma}_j \mathbf{v}_j$;

(b) *Under Condition* 1(a)–(d), $\boldsymbol{\Omega}_n^{-1} \hat{\boldsymbol{\Omega}}_n \xrightarrow{p} \mathbf{I}_2$, *where*

$$\hat{\boldsymbol{\Omega}}_n = M^{-2} \sum_{j=1}^{M} \mathbf{v}_j' \hat{\boldsymbol{\varepsilon}}_j \hat{\boldsymbol{\varepsilon}}_j' \mathbf{v}_j \quad \text{and} \quad \hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \mathbf{R}(\mathbf{R}'\mathbf{R})^{-1} \mathbf{R}'\mathbf{y}. \tag{S.1}$$

This result immediately implies the following.

COROLLARY 2. (a) *Let* $\mathbf{v}$ *be such that* $M^{-1}\mathbf{v}'\mathbf{y} = M^{-1} \sum_{j=1}^{M} \mathbf{v}_j' \mathbf{y}_j = (\hat{\beta}_{\text{long}}, \hat{\beta}_{\text{short}})'$, *and assume that Condition* 1 *holds. Then* (15) *holds, and* $\boldsymbol{\Omega}_n^{-1} \hat{\boldsymbol{\Omega}}_n \xrightarrow{p} \mathbf{I}_2$ *with* $\hat{\boldsymbol{\Omega}}_n$ *defined in* (S.1).

(b) *Let the* j*th row of* $\mathbf{v}$ *be equal to* $(\hat{w}_j^z, w_j)$ *as defined in Section* 4, *and assume that Condition* 1 *holds. Then under* $\beta = 0$, (21) *holds, and* $\boldsymbol{\Omega}_n^{-1} \hat{\boldsymbol{\Omega}}_n \xrightarrow{p} \mathbf{I}_2$ *with* $\hat{\boldsymbol{\Omega}}_n$ *defined in* (S.1).

REMARK 4. Since $(\hat{\beta}_{\text{long}}, \hat{\beta}_{\text{short}})$ and the IV score in (21) are $\mathbf{v}_j$-weighted averages of $\boldsymbol{\varepsilon}_j$, some bound on the relative magnitude of $\|\mathbf{v}_j\|$ is necessary to obtain asymptotic normality. The bounds in Condition 1 are relatively weak, allowing for $\max_j \|\mathbf{v}_j\| = o(M^{1/4})$ (if $\sum_{j=1}^{M} E[\|\boldsymbol{\varepsilon}_j\|^4] = O(M)$, $\max_j \|\boldsymbol{\Sigma}_j\| = O(1)$ and $\kappa^2 = o(M^{1/2})$). At the same time, one could also imagine that $\max_j \|\mathbf{v}_j\| = O(1)$, which would then allow for $E[\|\boldsymbol{\varepsilon}_j\|^4] = o(M)$, either because of increasingly fat tails, or because the number of observations per cluster is growing.

The result in part (a) makes no assumptions on $\boldsymbol{\gamma}$, so no restrictions are put on the asymptotic behavior of κ_n or τ_n.

The definition of $\hat{\boldsymbol{\Omega}}_n$ in part (b) for $M^{-1}\mathbf{v}'\mathbf{y} = (\hat{\beta}_{\text{long}}, \hat{\beta}_{\text{short}})'$ is the standard clustered variance estimator, except that the regression residuals are computed from the short regression. Under $\max_j \|\mathbf{v}_j\| = O(1)$ and $\max_j \|\boldsymbol{\Sigma}_j\| = O(1)$, $\kappa^2 = o(M/n)$ is enough to obtain consistency of $\hat{\boldsymbol{\Omega}}_n$. The important special case of independent but heteroskedastic disturbances ε_i (so that $\hat{\boldsymbol{\Omega}}_n$ reduces to the White (1980) standard errors based on short regression residuals), is obtained for $M = n$.

PROOF. (a) By the Cramér–Wold device, it suffices to show that $M^{-1}\boldsymbol{\upsilon}'\mathbf{v}'\boldsymbol{\varepsilon}/\sqrt{\boldsymbol{\upsilon}'\boldsymbol{\Omega}_n\boldsymbol{\upsilon}} \Rightarrow \mathcal{N}(0, 1)$ for all 2×1 vectors $\boldsymbol{\upsilon}$ with $\boldsymbol{\upsilon}'\boldsymbol{\upsilon} = 1$. This follows from the (triangular array version of the) Lyapunov central limit theorem applied to the M independent variables $\boldsymbol{\upsilon}'\mathbf{v}_j'\boldsymbol{\varepsilon}_j \sim$

$(0, \boldsymbol{\upsilon}'\mathbf{v}_j'\boldsymbol{\Sigma}_j\mathbf{v}_j\boldsymbol{\upsilon})$ and Condition 1(b), since

$$\frac{\sum_{j=1}^{M} E\left[\left(\boldsymbol{\upsilon}'\mathbf{v}_j'\boldsymbol{\varepsilon}_j\right)^4\right]}{\left(\sum_{j=1}^{M}\boldsymbol{\upsilon}'\mathbf{v}_j'\boldsymbol{\Sigma}_j\mathbf{v}_j\boldsymbol{\upsilon}\right)^2} \leq \max_j \|\mathbf{v}_j\|^4 \cdot M^{-2}\sum_{j=1}^{M} E\left[\|\boldsymbol{\varepsilon}_j\|^4\right] \cdot \left\| \left(M^{-1}\sum_{j=1}^{M}\mathbf{v}_j'\boldsymbol{\Sigma}_j\mathbf{v}_j\right)^{-1}\right\|^2 \to 0$$

and $\mathrm{Var}[M^{-1}\boldsymbol{\upsilon}'\mathbf{v}'\boldsymbol{\varepsilon}/\sqrt{\boldsymbol{\upsilon}'\boldsymbol{\Omega}_n\boldsymbol{\upsilon}}] = 1$.

(b) We show convergence of $\boldsymbol{\upsilon}'\hat{\boldsymbol{\Omega}}_n\boldsymbol{\upsilon}/(\boldsymbol{\upsilon}'\boldsymbol{\Omega}_n\boldsymbol{\upsilon}) \xrightarrow{p} 1$ for all 2×1 vectors $\boldsymbol{\upsilon}$ with $\boldsymbol{\upsilon}'\boldsymbol{\upsilon} = 1$. Note that $\boldsymbol{\upsilon}'\hat{\boldsymbol{\Omega}}_n\boldsymbol{\upsilon} = M^{-2}\sum_{j=1}^{M}\hat{\boldsymbol{\varepsilon}}_j'\mathbf{V}_j\hat{\boldsymbol{\varepsilon}}_j = M^{-2}\hat{\boldsymbol{\varepsilon}}'\mathbf{V}\hat{\boldsymbol{\varepsilon}}$ with $\mathbf{V}_j = \mathbf{v}_j\boldsymbol{\upsilon}\boldsymbol{\upsilon}'\mathbf{v}_j'$ and $\mathbf{V} = \mathrm{diag}(\mathbf{V}_1, \ldots, \mathbf{V}_M)$, and

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}} &= \mathbf{M}_R\boldsymbol{\varepsilon} + \mathbf{M}_R\mathbf{Z}\boldsymbol{\gamma} \\ &= \boldsymbol{\varepsilon} - \mathbf{R}\left(\mathbf{R}'\mathbf{R}\right)^{-1}\mathbf{R}'\boldsymbol{\varepsilon} + \mathbf{M}_R\mathbf{Z}\boldsymbol{\gamma} \qquad \text{(S.2)}\end{aligned}$$

with $\mathbf{M}_R = \mathbf{I}_n - \mathbf{R}(\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'$, so that

$$\begin{aligned}\hat{\boldsymbol{\varepsilon}}'\mathbf{V}\hat{\boldsymbol{\varepsilon}} = \boldsymbol{\varepsilon}'\mathbf{V}\boldsymbol{\varepsilon} &+ \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{M}_R\mathbf{V}\mathbf{M}_R\mathbf{Z}\boldsymbol{\gamma} + 2\boldsymbol{\gamma}'\mathbf{Z}'\mathbf{M}_R\mathbf{V}\mathbf{M}_R\boldsymbol{\varepsilon} - 2\boldsymbol{\varepsilon}'\mathbf{V}\mathbf{R}\left(\mathbf{R}'\mathbf{R}\right)^{-1}\mathbf{R}'\boldsymbol{\varepsilon} \\ &+ \boldsymbol{\varepsilon}'\mathbf{R}\left(\mathbf{R}'\mathbf{R}\right)^{-\mathbf{1}}\mathbf{R}'\mathbf{V}\mathbf{R}\left(\mathbf{R}'\mathbf{R}\right)^{-1}\mathbf{R}'\boldsymbol{\varepsilon}.\end{aligned}$$

Now

$$\begin{aligned}\boldsymbol{\gamma}'\mathbf{Z}'\mathbf{M}_R\mathbf{V}\mathbf{M}_R\mathbf{Z}\boldsymbol{\gamma} &\leq \|\mathbf{V}\| \cdot \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{M}_R\mathbf{Z}\boldsymbol{\gamma} \\ &\leq \max_j \|\mathbf{v}_j\|^2 \cdot \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\gamma} = \max_j \|\mathbf{v}_j\|^2 \cdot n\kappa^2\end{aligned}$$

and, with $\boldsymbol{\Sigma} = \mathrm{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_M)$,

$$\begin{aligned}\mathrm{Var}\left[\boldsymbol{\gamma}'\mathbf{Z}'\mathbf{M}_R\mathbf{V}\mathbf{M}_R\boldsymbol{\varepsilon}\right] &= \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{M}_R\mathbf{V}\mathbf{M}_R\boldsymbol{\Sigma}\mathbf{M}_R\mathbf{V}\mathbf{M}_R\mathbf{Z}\boldsymbol{\gamma} \\ &\leq \max_j \|\boldsymbol{\Sigma}_j\| \cdot \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{M}_R\mathbf{V}\mathbf{M}_R\mathbf{V}\mathbf{M}_R\mathbf{Z}\boldsymbol{\gamma} \\ &\leq \max_j \|\boldsymbol{\Sigma}_j\| \cdot \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{M}_R\mathbf{V}^2\mathbf{M}_R\mathbf{Z}\boldsymbol{\gamma} \\ &\leq \max_j \|\boldsymbol{\Sigma}_j\| \cdot \max_j \|\mathbf{v}_j\|^4 \cdot n\kappa^2\end{aligned}$$

and

$$\left\|\mathrm{Var}\left[\mathbf{R}'\boldsymbol{\varepsilon}\right]\right\| = \left\|\mathbf{R}'\boldsymbol{\Sigma}\mathbf{R}\right\| \leq \max_j \|\boldsymbol{\Sigma}_j\| \cdot \sum_{j=1}^{M}\|\mathbf{R}_j\|,^2$$

$$\left\|\mathrm{Var}\left[\mathbf{R}'\mathbf{V}\boldsymbol{\varepsilon}\right]\right\| = \left\|\mathbf{R}'\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}\mathbf{R}\right\| \leq \max_j \|\boldsymbol{\Sigma}_j\| \cdot \max_j \|\mathbf{v}_j\|^4 \cdot \sum_{j=1}^{M}\|\mathbf{R}_j\|^2,$$

$$\left\|\mathbf{R}'\mathbf{V}\mathbf{R}\right\| \le \max_j \|\mathbf{v}_j\|^2 \cdot \sum_{j=1}^{M} \|\mathbf{R}_j\|^2.$$

Furthermore, $(\boldsymbol{\upsilon}'\boldsymbol{\Omega}_n\boldsymbol{\upsilon})^{-1} = (M^{-2}\sum_{j=1}^{M}\boldsymbol{\upsilon}'\mathbf{v}_j'\boldsymbol{\Sigma}_j\mathbf{v}_j\boldsymbol{\upsilon})^{-1} \le \|(M^{-2}\sum_{j=1}^{M}\mathbf{v}_j'\boldsymbol{\Sigma}_j\mathbf{v}_j)^{-1}\| = O(M^{-1})$, so that under Condition 1(b)–(d), $M^{-2}(\hat{\boldsymbol{\varepsilon}}'\mathbf{V}\hat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon}'\mathbf{V}\boldsymbol{\varepsilon})/(\boldsymbol{\upsilon}'\boldsymbol{\Omega}_n\boldsymbol{\upsilon}) \xrightarrow{p} 0$.

Finally, rewrite $\boldsymbol{\varepsilon}'\mathbf{V}\boldsymbol{\varepsilon} = \sum_{j=1}^{M}\boldsymbol{\upsilon}'\mathbf{v}_j'\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j'\mathbf{v}_j\boldsymbol{\upsilon}$. Then $E[M^{-1}\boldsymbol{\varepsilon}'\mathbf{V}\boldsymbol{\varepsilon} - M\boldsymbol{\upsilon}'\boldsymbol{\Omega}_n\boldsymbol{\upsilon}] = 0$, and

$$\begin{aligned}\operatorname{Var}\left[M^{-1}\boldsymbol{\varepsilon}'\mathbf{V}\boldsymbol{\varepsilon} - M\boldsymbol{\upsilon}'\boldsymbol{\Omega}_n\boldsymbol{\upsilon}\right] &= M^{-2}\sum_{j=1}^{M}\operatorname{Var}\left[\boldsymbol{\upsilon}'\mathbf{v}_j'\left(\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j' - \boldsymbol{\Sigma}_j\right)\mathbf{v}_j\boldsymbol{\upsilon}\right] \\ &\le M^{-2}\max_j\|\mathbf{v}_j\|^4 \cdot \sum_{j=1}^{M} E\left[\|\boldsymbol{\varepsilon}_j\|^4\right]\end{aligned}$$

and the result follows from $(\boldsymbol{\upsilon}'\boldsymbol{\Omega}_n\boldsymbol{\upsilon})^{-1} = O(M^{-1})$ and Condition 1(a). □

S.2. Asymptotics under double bounds

Let $\mathbf{S} = (\mathbf{Q}, \mathbf{Z})$, and the following treats $\mathbf{S}$ as nonstochastic (or conditions on its realization). Straightforward algebra yields that under $\beta = 0$,

$$\sum_{i=1}^{n}\hat{x}_i^z y_i = \hat{\mathbf{x}}^{z\prime}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}_x'\mathbf{M}_S\boldsymbol{\varepsilon}, \tag{S.3}$$

$$\begin{aligned}\sum_{i=1}^{n}x_i y_i &= \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\gamma}_x + \boldsymbol{\varepsilon}_x'\mathbf{Z}\boldsymbol{\gamma} + \mathbf{x}'\boldsymbol{\varepsilon} & \text{(S.4)} \\ &= \boldsymbol{\gamma}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\gamma}_x + \boldsymbol{\varepsilon}_x'\mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_x'\mathbf{M}_Q\boldsymbol{\varepsilon} + \boldsymbol{\gamma}_x'Z'\boldsymbol{\varepsilon}, & \text{(S.5)}\end{aligned}$$

where $\mathbf{M}_S$ and $\mathbf{M}_Q$ are the $n \times n$ projection matrices associated with $\mathbf{Q}$ and $\mathbf{S}$ with elements $M_{Q,ij}$ and $M_{S,ij}$, respectively. With $\Delta^{\mathrm{Dbl}} = n^{-1}\boldsymbol{\gamma}'\mathbf{Z}'\mathbf{Z}\boldsymbol{\gamma}_x$, the bound $|\Delta^{\mathrm{Dbl}}| \le \bar{\kappa} \cdot \bar{\kappa}_x$ follows from the Cauchy–Schwarz inequality.

For fixed and finite p, standard arguments yield a CLT (24) and an associate asymptotic covariance estimator. For diverging p, more careful arguments are required, as discussed in Cattaneo, Jansson, and Newey (2018a, 2018b). In particular, by the Cramér–Wold device, and arguments very similar to the ones employed in the proof of Lemma A.2 of Chao, Swanson, Hausman, Newey, and Woutersen (2012), one obtains the following result.

Lemma 6. *Suppose that* $(\varepsilon_{x,i}, \varepsilon_i)$ *are mean-zero independent across* i, $E[\varepsilon_{x,i}\varepsilon_i] = 0$, *and for some* C *that does not depend on* n, $E[\varepsilon_{x,i}^4] < C$, $E[\varepsilon_i^4] < C$ *and* $E[\varepsilon_{x,i}^4\varepsilon_i^4] < C$ *almost* *surely. If* $p \to \infty$, *then* (24) *holds with*

$$\begin{aligned}
&\boldsymbol{\Omega}^{\text{Dbl}} \\
&= n^{-2} \begin{pmatrix} \sum_{i,j} M_{S,ij}^2 E[\varepsilon_{x,i}^2 \varepsilon_j^2] & \sum_{i,j} M_{S,ij} M_{Q,ij} E[\varepsilon_{x,i}^2 \varepsilon_j^2] \\ \sum_{i,j} M_{S,ij} M_{Q,ij} E[\varepsilon_{x,i}^2 \varepsilon_j^2] & \sum_{i,j} M_{Q,ij}^2 E[\varepsilon_{x,i}^2 \varepsilon_j^2] + \sum_{i=1}^n E[(z_i'\gamma)^2 \varepsilon_{x,i}^2 + (z_i'\gamma_x)^2 \varepsilon_i^2] \end{pmatrix}.
\end{aligned}$$

In the high-dimensional case with $p/n \to c \in (0, 1)$, it is not obvious how one would obtain a consistent estimator of $n\boldsymbol{\Omega}^{\text{Dbl}}$ in general, because it is difficult to estimate $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}_x$ with sufficient precision. We leave this question for future research.

In order to make further progress, suppose that $\mathbf{S}$ is such that $\|\boldsymbol{\Omega}^{\text{Dbl}}\| = O(n)$ and $\|(\boldsymbol{\Omega}^{\text{Dbl}})^{-1}\| = O(n^{-1})$, where $\|\cdot\|$ is the spectral norm. Assume further that $\kappa = o(1)$. Then under the assumptions of Lemma 6, or other weak dependence assumptions, $\text{Var}[\boldsymbol{\varepsilon}_x'\mathbf{Z}\boldsymbol{\gamma}] = o(n)$. The term $\boldsymbol{\varepsilon}_x'\mathbf{Z}\boldsymbol{\gamma}$ in (S.4) thus no longer makes a contribution to the asymptotic distribution. Under these assumptions, one can therefore proceed as in Section S.1 with $\mathbf{v} = (\hat{\mathbf{x}}^z, \mathbf{x})$ in Condition 1 to obtain both an alternative CLT (24) under clustering, and an appropriate estimator $\hat{\boldsymbol{\Omega}}^{\text{Dbl}}$ conditional on $\mathbf{x}$.

References

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018a), "Alternative asymptotics and the partially linear model with many regressors." *Econometric Theory*, 34 (2), 277–301. [5]

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018b), "Inference in linear regression models with many covariates and heteroscedasticity." *Journal of the American Statistical Association*, 113 (523), 1350–1361. [5]

Chao, J. C., N. R. Swanson, J. A. Hausman, W. K. Newey, and T. Woutersen (2012), "Asymptotic distribution of JIVE in a heteroskedastic IV regression with many instruments." *Econometric Theory*, 28 (1), 42–86. [5]

White, H. (1980), "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica*, 48, 817–830. [3]

Co-editor Andres Santos handled this manuscript.

Manuscript received 15 March, 2020; final version accepted 19 November, 2020; available online 21 December, 2020.