HETEROSKEDASTICITY-ROBUST INFERENCE IN LINEAR REGRESSION MODELS WITH MANY COVARIATES

SUPPLEMENTAL APPENDIX

KOEN JOCHMANS*
UNIVERSITY OF CAMBRIDGE

This supplemental appendix contains the proof of Theorem 1 (in Section A.1), derivations underlying the primitive conditions in Section 4 (in Section A.2), and additional simulations results (in Section B).

A Technical details

A.1 Proof of Theorem 1

We need to show that

$$\frac{\sum_{i=1}^{n} \hat{\boldsymbol{v}}_{i,n}\hat{\boldsymbol{v}}_{i,n}'\,(y_{i,n}\acute{u}_{i,n})}{n} = \frac{\sum_{i=1}^{n} \hat{\boldsymbol{v}}_{i,n}\hat{\boldsymbol{v}}_{i,n}'\,\sigma_{i,n}^2}{n} + o_p(1).$$

As Cattaneo, Jansson and Newey (2018), to ease notation, we set $r = 1$ without loss of generality.

Add and subtract $\varepsilon_{i,n}$ to get

$$\frac{\sum_{i=1}^{n} \hat{v}_{i,n}^2\,(y_{i,n}\acute{u}_{i,n} - \sigma_{i,n}^2)}{n} = \frac{\sum_{i=1}^{n} \hat{v}_{i,n}^2\,(\varepsilon_{i,n}^2 - \sigma_{i,n}^2)}{n} + \frac{\sum_{i=1}^{n} \hat{v}_{i,n}^2\,(y_{i,n}\acute{u}_{i,n} - \varepsilon_{i,n}^2)}{n}. \tag{A.1}$$

Consider the first term on the right-hand side. Because $\sigma_{i,n}^2 = \mathbb{E}(\varepsilon_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n)$ by definition,

$$\mathbb{E}\left(\left. \frac{\sum_{i=1}^{n} \hat{v}_{i,n}^2\,(\varepsilon_{i,n}^2 - \sigma_{i,n}^2)}{n} \right| \mathcal{X}_n, \mathcal{W}_n \right) = 0.$$

*Address: University of Cambridge, Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom. E-mail: kj345@cam.ac.uk.
Revision history: September 5, 2018 (first version); August 28, 2019 (revision); April 27, 2020 (this version).

Next, letting $c_{i,j} := \mathbb{E}(\varepsilon_{i,n}^2\, \varepsilon_{j,n}^2 | \mathcal{X}_n, \mathcal{W}_n) - \sigma_{i,n}^2 \sigma_{j,n}^2$ and noting that

$$\|c_{i,j}\| \leq \max_i \mathbb{E}(\varepsilon_{i,n}^4 | \mathcal{X}_n, \mathcal{W}_n) + \max_i \sigma_{i,n}^4 =: c_n,$$

Assumptions 1–3 imply that

$$\begin{aligned}\mathbb{E}\left(\left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2\,(\varepsilon_{i,n}^2 - \sigma_{i,n}^2)}{n} \right)^2 \middle| \mathcal{X}_n, \mathcal{W}_n \right) &= \frac{\sum_{g=1}^{G_n} \sum_{i\in N_g} \sum_{j \in N_g} \hat{v}_{i,n}^2 \hat{v}_{j,n}^2\, c_{i,j}}{n^2} \\ &\leq c_n \left(\max_g |N_g| \right) \left(\max_i \frac{\|\hat{v}_{i,n}\|}{\sqrt{n}} \right)^2 \frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n} \\ &= o_p(1)\end{aligned}$$

because $c_n = O_p(1)$, $\max_g |N_g| = O(1)$, $\max_i \|\hat{v}_{i,n}\|/\sqrt{n} = o_p(1)$, and

$$\frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n} \leq \frac{\sum_{i=1}^n v_{i,n}^2}{n} \leq 2\frac{\sum_{i=1}^n Q_{i,n}^2}{n} + 2\frac{\sum_{i=1}^n V_{i,n}^2}{n} = O_p(\chi_n) + O_p(1) = O_p(1); \tag{A.2}$$

here, the first inequality follows from the fact that $\hat{v}_{i,n}$ is a least-squares residual—and thus has minimal variance—and the second is an application of the well-known inequality $\frac{1}{2}(a_1 + a_2) \leq \sqrt{\frac{1}{2}(a_1^2 + a_2^2)}$. Consequently,

$$\frac{\sum_{i=1}^n \hat{v}_{i,n}^2\,(\varepsilon_{i,n}^2 - \sigma_{i,n}^2)}{n} = o_p(1),$$

and (A.1) reduces to

$$\frac{\sum_{i=1}^n \hat{v}_{i,n}^2\,(y_{i,n}\acute{u}_{i,n} - \sigma_{i,n}^2)}{n} = \frac{\sum_{i=1}^n \hat{v}_{i,n}^2\,(y_{i,n}\acute{u}_{i,n} - \varepsilon_{i,n}^2)}{n} + o_p(1). \tag{A.3}$$

We turn to the sample average on the right-hand side of this expression next.

To do so it is useful to work with the decomposition

$$\begin{aligned} y_{i,n}\acute{u}_{i,n} - \varepsilon_{i,n}^2 &= \sum_{j\neq i} \varepsilon_{i,n} (\boldsymbol{A}_n)_{i,j}\, \varepsilon_{j,n} + ((\boldsymbol{A}_n)_{i,i} - 1)\, \varepsilon_{i,n}^2 \\ &\quad + \sum_{j=1}^n \mu_{i,n} (\boldsymbol{A}_n)_{i,j}\, \varepsilon_{j,n} + \sum_{j=1}^n \varepsilon_{i,n} (\boldsymbol{A}_n)_{i,j}\, e_{j,n} \\ &\quad + \sum_{j=1}^n \mu_{i,n} (\boldsymbol{A}_n)_{i,j}\, e_{j,n}, \end{aligned} \tag{A.4}$$

where

$$(\boldsymbol{A}_n)_{i,j} := \frac{(\boldsymbol{H}_n)_{i,j}}{(\boldsymbol{M}_n)_{i,i}}, \qquad (\boldsymbol{H}_n)_{i,j} := (\boldsymbol{M}_n)_{i,j} - \left(\frac{\sum_{k=1}^{n} \hat{v}_{k,n}^2}{n}\right)^{-1} \frac{\hat{v}_{i,n}\,\hat{v}_{j,n}}{n}.$$

Using standard formulae for partitioned regression, $\boldsymbol{H}_n$ can be seen to be the annihilator matrix of a regression on both $x_{i,n}$ and $\boldsymbol{w}_{i,n}$, whereas $\boldsymbol{M}_n$ follows from a projection on $\boldsymbol{w}_{i,n}$ alone. Observe that $(\boldsymbol{A}_n)_{i,j} \neq (\boldsymbol{A}_n)_{j,i}$. Equation (A.4) follows from recalling that $y_{i,n} = \mu_{i,n} + \varepsilon_{i,n}$ and that

$$\acute{u}_{i,n} = \frac{\hat{u}_{i,n}}{(\boldsymbol{M}_n)_{i,i}} = \sum_{j=1}^{n} \frac{(\boldsymbol{H}_n)_{i,j}}{(\boldsymbol{M}_n)_{i,i}}\, u_{j,n} = \sum_{j=1}^{n} (\boldsymbol{A}_n)_{i,j}\,\varepsilon_{j,n} + \sum_{j=1}^{n} (\boldsymbol{A}_n)_{i,j}\, e_{j,n},$$

which itself is a consequence of $u_{i,n} = \varepsilon_{i,n} + e_{i,n}$ and the fact that $\hat{u}_{i,n} = \sum_{j=1}^{n} (\boldsymbol{H}_n)_{i,j}\, u_{j,n}$. Using (A.4) we have

$$\begin{aligned} \frac{\sum_{i=1}^{n} \hat{v}_{i,n}^2\,(y_{i,n}\acute{u}_{i,n} - \varepsilon_{i,n}^2)}{n} &= \frac{\sum_{i=1}^{n}\sum_{j\neq i} \hat{v}_{i,n}^2\,\varepsilon_{i,n}(\boldsymbol{A}_n)_{i,j}\,\varepsilon_{j,n}}{n} + \frac{\sum_{i=1}^{n} \hat{v}_{i,n}^2\,((\boldsymbol{A}_n)_{i,i} - 1)\,\varepsilon_{i,n}^2}{n} \\ &+ \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \hat{v}_{i,n}^2\,\mu_{i,n}(\boldsymbol{A}_n)_{i,j}\,\varepsilon_{j,n}}{n} \\ &+ \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \hat{v}_{i,n}^2\,\varepsilon_{i,n}(\boldsymbol{A}_n)_{i,j}\,e_{j,n}}{n} \\ &+ \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} \hat{v}_{i,n}^2\,\mu_{i,n}(\boldsymbol{A}_n)_{i,j}\,e_{j,n}}{n}. \end{aligned} \tag{A.5}$$

We will handle each of these five terms in turn.

For the first right-hand side term in (A.5),

$$\mathbb{E}\left(\left.\frac{\sum_{i=1}^{n}\sum_{j\neq i} \hat{v}_{i,n}^2\varepsilon_{i,n}\,(\boldsymbol{A}_n)_{i,j}\,\varepsilon_{j,n}}{n}\right| \mathcal{X}_n, \mathcal{W}_n\right) = 0$$

because the $\varepsilon_{i,n}$ are (conditionally) uncorrelated by Assumption 1. Next, to see that also

$$\mathbb{E}\left(\left.\left(\frac{\sum_{i=1}^{n}\sum_{j\neq i} \hat{v}_{i,n}^2\varepsilon_{i,n}\,(\boldsymbol{A}_n)_{i,j}\,\varepsilon_{j,n}}{n}\right)^2\right| \mathcal{X}_n, \mathcal{W}_n\right) = o_p(1), \tag{A.6}$$

first expand the square to get

$$\frac{\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{j_1\neq i_1}\sum_{j_2\neq i_2} \hat{v}_{i_1,n}^2\,(\boldsymbol{A}_n)_{i_1,j_1}\,\mathbb{E}(\varepsilon_{i_1,n}\varepsilon_{j_1,n}, \varepsilon_{i_2,n}\varepsilon_{j_2,n}|\mathcal{X}_n, \mathcal{W}_n)\,(\boldsymbol{A}_n)_{i_2,j_2}\,\hat{v}_{i_2,n}^2}{n^2}.$$

Conditional independence of the $\varepsilon_{i,n}$ across groups g, g' in the partition of the observations and zero correlation within each group g implies that the summand will be zero unless either (i) $i_1 = i_2$ and $j_1 = j_2$, with $i_1 \in N_g$ and $j_1 \in N_{g'}$; or (ii) $i_1 = j_2$ and $i_2 = j_1$, with $i_1 \in N_g$ and $i_2 \in N_{g'}$; or (iii) $(i_1, i_2, j_1, j_2) \in N_g$, where $g \neq g'$. The contribution of Case (i) terms equals

$$\begin{aligned}
\frac{\sum_{g=1}^{G_n} \sum_{g' \neq g} \sum_{i \in N_g} \sum_{j \in N_{g'}} \hat{v}_{i,n}^4\, \sigma_{i,n}^2 (\boldsymbol{A}_n)_{i,j}^2\, \sigma_{j,n}^2}{n^2} &\leq \frac{(\max_i \sigma_{i,n}^2)^2\ \sum_{i=1}^n \hat{v}_{i,n}^4 \sum_{j \neq i} (\boldsymbol{A}_n)_{i,j}^2}{n^2} \\
&\leq \frac{(\max_i \sigma_{i,n}^2)^2\ \sum_{i=1}^n \hat{v}_{i,n}^4\, (\boldsymbol{M}_n)_{i,i}^{-2}}{n^2} \\
&\leq \frac{(\max_i \sigma_{i,n}^2)^2\, (\min_i (\boldsymbol{M}_n)_{i,i})^{-2} \sum_{i=1}^n \hat{v}_{i,n}^4}{n^2} \\
&\leq \frac{(\max_i \sigma_{i,n}^2)^2}{(\min_i (\boldsymbol{M}_n)_{i,i})^2} \left(\max_i \frac{\|\hat{v}_{i,n}\|}{\sqrt{n}} \right)^2 \frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n} \\
&= o_p(1),
\end{aligned}$$

where we use that $\sum_{j \neq i} (\boldsymbol{A}_n)_{i,j}^2 \leq \sum_{j=1}^n (\boldsymbol{A}_n)_{i,j}^2 = (\boldsymbol{H}_n)_{i,i}\, (\boldsymbol{M}_n)_{i,i}^{-2}$—which follows from the fact that $\boldsymbol{H}_n$ is a projection matrix, and so $\sum_{j=1}^n (\boldsymbol{H}_n)_{i,j}^2 = (\boldsymbol{H}_n)_{i,i} \in [0,1]$ by idempotency—and envoke Assumptions 2 and 4 for $(\max_i \sigma_{i,n}^2)(\min_i (\boldsymbol{M}_n)_{i,i})^{-1} = O_p(1)$. Similarly, the contribution of Case (ii) terms equals

$$\frac{\sum_{g=1}^{G_n} \sum_{g' \neq g} \sum_{i \in N_g} \sum_{j \in N_{g'}} \hat{v}_{i,n}^2\, \sigma_{i,n}^2\, (\boldsymbol{A}_n)_{i,j}\, (\boldsymbol{A}_n)_{j,i}\, \hat{v}_{j,n}^2 \sigma_{j,n}^2}{n^2}$$

and is bounded by

$$(\max_i \sigma_{i,n}^2)^2\, \frac{\sum_{i=1}^n \sum_{j \neq i} \hat{v}_i^2 \hat{v}_j^2\, (\boldsymbol{A}_n)_{i,j}\, (\boldsymbol{A}_n)_{j,i}}{n^2} \leq \frac{(\max_i \sigma_{i,n}^2)^2 \left(\max_i \frac{\|\hat{v}_{i,n}\|}{\sqrt{n}} \right)^2}{(\min_i (\boldsymbol{M}_n)_{i,i})^2}\ \frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n} = o_p(1),$$

where, now, we use that $(\boldsymbol{A}_n)_{i,j} (\boldsymbol{A}_n)_{j,i} \geq 0$ to validate the first upper bound, and that $\sum_{j \neq i} (\boldsymbol{A}_n)_{i,j} (\boldsymbol{A}_n)_{j,i} \leq \sum_{j=1}^n (\boldsymbol{A}_n)_{i,j} (\boldsymbol{A}_n)_{j,i} \leq (\min_i (\boldsymbol{M}_n)_{i,i})^{-2}$ for the second upper bound. Finally, the contribution of Case (iii) terms equals

$$\sum_{g=1}^{G_n} \sum_{i_1, i_2 \in N_g} \sum_{\substack{j_1, j_2 \in N_g \\ j_1 \neq i_1 \\ j_2 \neq i_2}} \frac{\hat{v}_{i_1,n}^2\, (\boldsymbol{A}_n)_{i_1,j_1}\, \mathbb{E}(\varepsilon_{i_1,n} \varepsilon_{j_1,n}, \varepsilon_{i_2,n} \varepsilon_{j_2,n} | \mathcal{X}_n, \mathcal{W}_n)\, (\boldsymbol{A}_n)_{i_2,j_2}\, \hat{v}_{i_2,n}^2}{n^2}$$

and, using the Cauchy-Schwarz inequality, is bounded by

$$\frac{(\max_i \mathbb{E}(\varepsilon_{i,n}^4|\mathcal{X}_n,\mathcal{W}_n))\,(\max_g|N_g|)^3}{(\min_i(\boldsymbol{M}_n)_{i,i})^2}\left(\max_i \frac{\|\hat{v}_{i,n}\|}{\sqrt{n}}\right)^2 \frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n} = o_p(1),$$

which follows by the same arguments. Equation (A.6) has been shown.

The second right-hand side term in (A.5) has mean

$$\mathbb{E}\left(\left\|\frac{\sum_{i=1}^n \hat{v}_{i,n}^2\,((\boldsymbol{A}_n)_{i,i}-1)\,\varepsilon_{i,n}^2}{n}\right\|\middle|\,\mathcal{X}_n,\mathcal{W}_n\right) = \left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n}\right)^{-1}\frac{\sum_{i=1}^n \hat{v}_{i,n}^4\sigma_{i,n}^2\,(\boldsymbol{M}_n)_{i,i}^{-1}}{n^2},$$

where we use that

$$((\boldsymbol{A}_n)_{i,i}-1) = -\left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n}\right)^{-1}\left(\frac{\hat{v}_{i,n}^2}{n}\frac{1}{(\boldsymbol{M}_n)_{i,i}}\right). \tag{A.7}$$

This vanishes because $\left(n^{-1}\sum_{i=1}^n \hat{v}_{i,n}^2\right)^{-1} = O_p(1)$ (Cattaneo, Jansson and Newey 2018, Lemma SA-1) and

$$\frac{\sum_{i=1}^n \hat{v}_{i,n}^4\sigma_{i,n}^2\,(\boldsymbol{M}_n)_{i,i}^{-1}}{n^2} \leq (\min_i(\boldsymbol{M}_n)_{i,i})^{-1}\,(\max_i \sigma_{i,n}^2)\left(\max_i \frac{\|\hat{v}_{i,n}\|}{\sqrt{n}}\right)^2 \frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n} = o_p(1).$$

To see that

$$\mathbb{E}\left(\left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2\,((\boldsymbol{A}_n)_{i,i}-1)\,\varepsilon_{i,n}^2}{n}\right)^2\middle|\,\mathcal{X}_n,\mathcal{W}_n\right) = o_p(1),$$

first expand the square and once again use (A.7) to see that the second moment can be written as

$$\left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n}\right)^{-2}\left(\frac{\sum_{i=1}^n\sum_{j=1}^n \hat{v}_{i,n}^4\hat{v}_{j,n}^4(\boldsymbol{M}_n)_{i,i}^{-1}\,(\boldsymbol{M}_n)_{j,j}^{-1}\,\mathbb{E}(\varepsilon_{i,n}^2\varepsilon_{j,n}^2|\mathcal{X}_n,\mathcal{W}_n)}{n^4}\right).$$

Applying the Cauchy-Schwarz inequality to $\mathbb{E}(\varepsilon_{i,n}^2\varepsilon_{j,n}^2|\mathcal{X}_n,\mathcal{W}_n)$ then yields the upper bound

$$\left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n}\right)^{-2}\left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^4\,(\boldsymbol{M}_n)_{i,i}^{-1}\sqrt{\mathbb{E}(\varepsilon_{i,n}^4|\mathcal{X}_n,\mathcal{W}_n)}}{n^2}\right)^2 = O_p(1)\,o_p(1) = o_p(1);$$

this follows by the same argument as that just used for the first moment, only now using the fact that the fourth moment is uniformly bounded.

The third right-hand side term in (A.5) is mean zero because $\mathbb{E}(\varepsilon_{i,n}|\mathcal{X}_n, \mathcal{W}_n) = 0$ by construction. Its variance is

$$\mathbb{E}\left(\left(\frac{\sum_{i=1}^n \sum_{j=1}^n \hat{v}_{i,n}^2\, \mu_{i,n}(\boldsymbol{A}_n)_{i,j}\, \varepsilon_{j,n}}{n}\right)^2 \middle| \mathcal{X}_n, \mathcal{W}_n\right) = \sum_{j=1}^n \sigma_{j,n}^2 \left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2\, \mu_{i,n}(\boldsymbol{A}_n)_{i,j}}{n}\right)^2.$$

To show that this vanishes first use

$$\sum_{j=1}^n (\boldsymbol{A}_n)_{i,j}\, (\boldsymbol{A}_n)_{k,j} = \sum_{j=1}^n \frac{(\boldsymbol{H}_n)_{i,j}\, (\boldsymbol{H}_n)_{j,k}}{(\boldsymbol{M}_n)_{i,i}\, (\boldsymbol{M}_n)_{k,k}} = \frac{(\boldsymbol{H}_n)_{i,k}}{(\boldsymbol{M}_n)_{i,i}\, (\boldsymbol{M}_n)_{k,k}},$$

which follows from idempotency of $\boldsymbol{H}_n$, to see that

$$\sum_{j=1}^n \left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2\, \mu_{i,n}(\boldsymbol{A}_n)_{i,j}}{n}\right)^2 = \frac{\sum_{i=1}^n \sum_{k=1}^n \hat{v}_{i,n}^2\, \mu_{i,n} \left(\sum_{j=1}^n (\boldsymbol{A}_n)_{i,j}\, (\boldsymbol{A}_n)_{k,j}\right)\, \hat{v}_{k,n}^2\, \mu_{k,n}}{n^2}$$

is a quadratic form in the matrix $\boldsymbol{H}_n$. Because $\boldsymbol{H}_n$ is a projection matrix we have that $\sum_{i=1}^n \sum_{j=1}^n a_i (\boldsymbol{H}_n)_{i,j} a_j \leq \sum_{i=1}^n a_i^2$ for any $(a_1, \ldots, a_n)$. Consequently, the variance satisfies

$$\begin{aligned}
\sum_{j=1}^n \sigma_{j,n}^2 \left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2\, \mu_{i,n}(\boldsymbol{A}_n)_{i,j}}{n}\right)^2 &\leq (\max_i \sigma_{i,n}^2) \sum_{j=1}^n \left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2\, \mu_{i,n}(\boldsymbol{A}_n)_{i,j}}{n}\right)^2 \\
&\leq (\max_i \sigma_{i,n}^2) \sum_{i=1}^n \left(\frac{\hat{v}_{i,n}^2\, \mu_{i,n}\, (\boldsymbol{M}_n)_{i,i}^{-1}}{n}\right)^2 \\
&\leq \frac{(\max_i \sigma_{i,n}^2)}{(\min_i (\boldsymbol{M}_n)_{i,i})^2} \frac{\sum_{i=1}^n \hat{v}_{i,n}^4\, \mu_{i,n}^2}{n^2} \\
&\leq \frac{(\max_i \sigma_{i,n}^2)}{(\min_i (\boldsymbol{M}_n)_{i,i})^2} \left(\max_i \frac{\|\mu_{i,n}\|}{\sqrt{n}}\right)^2 \frac{\sum_{i=1}^n \hat{v}_{i,n}^4}{n} = o_p(1),
\end{aligned}$$

where the final transition uses the fact that $\hat{v}_{i,n} = \tilde{Q}_{i,n} + \tilde{V}_{i,n}$ together with the inequality

$$\frac{\sum_{i=1}^n \hat{v}_{i,n}^4}{n} \leq 4\frac{\sum_{i=1}^n \tilde{Q}_{i,n}^4}{n} + 4\frac{\sum_{i=1}^n \tilde{V}_{i,n}^4}{n} = O_p(1);$$

the last equality following from Assumption 4 and the fact that $\max_i \mathbb{E}(\tilde{V}_{i,n}^4|\mathcal{W}_n) = O_p(1)$ by Assumption 1 and Assumption 2, a detailed derivation of this last result is available in the proof of Lemma SA-7 in the supplementary material to Cattaneo, Jansson and Newey (2018).

The fourth right-hand side term in (A.5), is zero mean for the same reason as the third. Its variance is

$$\mathbb{E}\left(\left(\frac{\sum_{i=1}^n \sum_{j=1}^n \hat{v}_{i,n}^2\, \varepsilon_{i,n}(\boldsymbol{A}_n)_{i,j}\, e_{j,n}}{n}\right)^2 \middle| \mathcal{X}_n, \mathcal{W}_n\right) = \frac{\sum_{i=1}^n \sigma_{i,n}^2 \hat{v}_{i,n}^4 \left(\sum_{j=1}^n (\boldsymbol{A}_n)_{i,j}\, e_{j,n}\right)^2}{n^2}$$

and is bounded by

$$\frac{(\max_i \sigma_{i,n}^2)}{(\min_i (\boldsymbol{M}_n)_{i,i})^2} \left(\max_i \frac{\|\hat{v}_{i,n}\|}{\sqrt{n}}\right)^2 \left(\frac{\sum_{i=1}^n \hat{v}_{i,n}^2}{n}\right) n \left(\frac{\sum_{i=1}^n e_{i,n}^2}{n}\right) = o_p(n\varrho_n) = o_p(1),$$

where we use $(\sum_{j=1}^n (\boldsymbol{A}_n)_{i,j}\, e_{j,n})^2 \le (\sum_{j=1}^n (\boldsymbol{A}_n)_{i,j}^2)\,(\sum_{j=1}^n e_{j,n}^2) \le (\boldsymbol{M}_n)_{i,i}^{-2}\,(\sum_{j=1}^n e_{j,n}^2)$ and rely on $n\varrho_n = O(1)$, as stated in Assumption 4 to reach the desired conclusion.

The fifth right-hand side term in (A.5), finally, is the bias term. The Cauchy-Schwarz inequality gives

$$\left(\frac{\sum_{i=1}^n \sum_{j=1}^n \hat{v}_{i,n}^2 \mu_{i,n}\,(\boldsymbol{A}_n)_{i,j}\, e_{j,n}}{n}\right)^2 \le \frac{\sum_{i=1}^n \hat{v}_{i,n}^4\,(\boldsymbol{M}_n)_{i,i}^{-2}}{n}\; \frac{\sum_{i=1}^n \mu_{i,n}^2 (\sum_{j=1}^n (\boldsymbol{H}_n)_{i,j}\, e_{j,n})^2}{n},$$

where

$$\frac{\sum_{i=1}^n \hat{v}_{i,n}^4\,(\boldsymbol{M}_n)_{i,i}^{-2}}{n} \le \frac{(\min_i (\boldsymbol{M}_n)_{i,i}^{-2})\, \sum_{i=1}^n \hat{v}_{i,n}^4}{n} = O_p(1),$$

and

$$\begin{aligned}
\frac{\sum_{i=1}^n \mu_{i,n}^2 (\sum_{j=1}^n (\boldsymbol{H}_n)_{i,j}\, e_{j,n})^2}{n} &\le \left(\max_i \frac{\|\mu_{i,n}\|}{\sqrt{n}}\right)^2 \sum_{i=1}^n \left(\sum_{j=1}^n (\boldsymbol{H}_n)_{i,j}\, e_{j,n}\right)^2 \\
&= \left(\max_i \frac{\|\mu_{i,n}\|}{\sqrt{n}}\right)^2 \left(\sum_{j=1}^n \sum_{k=1}^n e_{j,n} \left(\sum_{i=1}^n (\boldsymbol{H}_n)_{i,j}\,(\boldsymbol{H}_n)_{i,k}\right) e_{k,n}\right) \\
&= \left(\max_i \frac{\|\mu_{i,n}\|}{\sqrt{n}}\right)^2 \left(\sum_{j=1}^n \sum_{k=1}^n e_{j,n} (\boldsymbol{H}_n)_{j,k}\, e_{k,n}\right) \\
&\le \left(\max_i \frac{\|\mu_{i,n}\|}{\sqrt{n}}\right)^2 \left(\sum_{j=1}^n e_{j,n}^2\right) = o_p(n\varrho_n) = o_p(1),
\end{aligned}$$

by the same arguments as before.

Collecting results for the five right-hand side terms in (A.5) implies that (A.3) becomes

$$\frac{\sum_{i=1}^n \hat{v}_{i,n}^2\,(y_{i,n} \acute{u}_{i,n})}{n} = \frac{\sum_{i=1}^n \hat{v}_{i,n}^2\, \sigma_{i,n}^2}{n} + o_p(1),$$

which is what we wanted to show. □

A.2 Sufficient conditions for Assumption 4.

We continue to work with the case where $r = 1$. We provide primitive conditions for the requirements

$$\text{(i)} \quad \frac{\sum_{i=1}^n \tilde{Q}_{i,n}^4}{n} = O_p(1) \qquad \text{and} \qquad \text{(ii)} \quad \max_i \frac{\|\mu_{i,n}\|}{\sqrt{n}} = o_p(1),$$

in turn.

Condition (i). As shown in the proof of Lemma SA-7 in the supplementary material to Cattaneo, Jansson and Newey (2018), $\chi_n = o(1)$ implies that

$$\max_i \frac{\|\tilde{Q}_{i,n}\|}{\sqrt{n}} = o_p(1).$$

Consequently,

$$\frac{\sum_{i=1}^n \tilde{Q}_{i,n}^4}{n} \leq \left(\max_i \frac{\|\tilde{Q}_{i,n}\|}{\sqrt{n}}\right)^2 \sum_{i=1}^n \tilde{Q}_{i,n}^2 \leq \left(\max_i \frac{\|\tilde{Q}_{i,n}\|}{\sqrt{n}}\right)^2 n \left(\frac{\sum_{i=1}^n Q_{i,n}^2}{n}\right) = o_p(n\chi_n),$$

where

$$\sum_{i=1}^n \tilde{Q}_{i,n}^2 = \sum_{i=1}^n \left(\sum_{j=1}^n (\boldsymbol{M}_n)_{i,j} Q_{j,n}\right)^2 = \sum_{i=1}^n \sum_{j=1}^n Q_{i,n} (\boldsymbol{M}_n)_{i,j} Q_{j,n} \leq \sum_{i=1}^n Q_{i,n}^2$$

was used. Hence, $n\chi_n = O(1)$ is sufficient for Condition (i) to hold.

Next, recall that

$$n_{i,n} = \sum_{j=1}^n \{(\boldsymbol{M}_n)_{i,j} \neq 0\},$$

and let $[i]_n := \{j : (\boldsymbol{M}_n)_{i,j} \neq 0\}$. If $\max_i n_{i,n} = O_p(1)$ and $\sum_{i=1}^n Q_{i,n}^4 = O_p(n)$, then we obtain

$$\begin{aligned}
\frac{\sum_{i=1}^n \tilde{Q}_{i,n}^4}{n} = \frac{\sum_{i=1}^n \left(\sum_{j=1}^n (\boldsymbol{M}_n)_{i,j} Q_{j,n}\right)^4}{n} &\leq \frac{\sum_{i=1}^n \left(\sum_{j=1}^n (\boldsymbol{M}_n)_{i,j}^{4/3}\right)^3 \left(\sum_{j\in[i]_n} Q_{j,n}^4\right)}{n} \\
&\leq \frac{\sum_{i=1}^n n_{i,n}^3 \left(\sum_{j\in[i]_n} Q_{j,n}^4\right)}{n} \\
&\leq (\max_i n_{i,n})^3 \frac{\sum_{i=1}^n \sum_{j\in[i]_n} Q_{j,n}^4}{n} \\
&\leq (\max_i n_{i,n})^4 \frac{\sum_{j=1}^n Q_{j,n}^4}{n} = O_p(1)
\end{aligned}$$

by an application of Hölder's inequality.

Condition (ii). We first note that

$$\max_i \|\mu_{i,n}\| \leq \max_i \|x_{i,n}\| \, \|\beta\| + \max_i \|\boldsymbol{w}'_{i,n}\boldsymbol{\gamma}_n + e_{i,n}\|,$$

and that the first term on the right-hand side is easily handled. For any $\epsilon > 0$ and $\theta > 0$, we have

$$\Pr\left(\max_i \frac{\|x_{i,n}\|}{\sqrt{n}} > \epsilon\right) \leq \sum_{i=1}^{n} \Pr\left(\|x_{i,n}\| > \epsilon\sqrt{n}\right) \leq \left(\frac{n^{-\theta/2}}{\epsilon^{2+\theta}}\right) \frac{\sum_{i=1}^{n} \mathbb{E}\left(\|x_{i,n}\|^{2+\theta}\right)}{n}.$$

Consequently, $\max_i \|x_{i,n}\|/\sqrt{n} = o_p(1)$ follows from $\sum_{i=1}^{n} \mathbb{E}(\|x_{i,n}\|^{2+\theta}) = O(n)$, which is a conventional requirement.

The same argument can be used for the second term in cases where $\boldsymbol{w}'_{i,n}\boldsymbol{\gamma}_n$ is a series approximation to a well-behaved function $\varphi(\boldsymbol{z}_{i,n})$. In that case, $\boldsymbol{w}'_{i,n}\boldsymbol{\gamma}_n + e_{i,n} = \varphi(\boldsymbol{z}_{i,n})$, and the requirement that $\sum_{i=1}^{n} \mathbb{E}(\|\varphi(\boldsymbol{z}_{i,n})\|^{2+\theta}) = O(n)$ again does not appear overly strong to impose.

We can also tackle the problem by first noting that

$$\max_i \|\boldsymbol{w}'_{i,n}\boldsymbol{\gamma}_n + e_{i,n}\| \leq \max_i \|\boldsymbol{w}'_{i,n}\boldsymbol{\gamma}_n\| + \max_i \|e_{i,n}\| \leq \max_i \|\boldsymbol{w}'_{i,n}\boldsymbol{\gamma}_n\| + o_p(\sqrt{n}),$$

using the fact that

$$\Pr\left(\max_i \frac{\|e_{i,n}\|}{\sqrt{n}} \geq \epsilon\right) \leq \frac{1}{\epsilon^2} \frac{\sum_{i=1}^{n} \mathbb{E}(\|e_{i,n}\|^2)}{n} = O(\varrho_n) = o(1)$$

by Assumption 3, and then imposing a growth rate on the number of parameters that affect each observation, together with a moment condition on $\boldsymbol{w}_{i,n}$. Moreover, writing $\boldsymbol{w}_{i,n} = (w_{i,n,1}, \ldots, w_{i,n,q_n})'$ and $\boldsymbol{\gamma}_n = (\gamma_{n,1}, \ldots, \gamma_{n,q_n})'$,

$$\boldsymbol{w}'_{i,n}\boldsymbol{\gamma}_n = \sum_{j=1}^{q_n} w_{i,n,j}\,\gamma_{n,j} = \sum_{j\in(i)_n} w_{i,n,j}\,\gamma_{n,j},$$

where the index set $(i)_n$ has cardinality $\kappa_{i,n} := |(i)_n|$. In grouped data with κ-way fixed effects, for example, $\kappa_{i,n} = \kappa$ for all i, but we will also cover the case where $\kappa_{i,n}$ grows with n. By Hölder's inequality,

$$\mathbb{E}(\|\boldsymbol{w}'_{i,n}\boldsymbol{\gamma}_n\|^{2+\theta}) \leq \left(\sum_{j\in(i)_n} \mathbb{E}(\|w_{i,n,j}\|^{2+\theta})\right) \left(\sum_{j\in(i)_n} \gamma_{n,j}^{(2+\theta)/(1+\theta)}\right)^{1+\theta}.$$

Consequently, if $\sum_{j\in(i)_n}\mathbb{E}(\|w_{i,n,j}\|^{2+\theta})=O(\kappa_i)$ for all i and $\gamma_{n,i}=O(1)$ for all i we obtain that

$$\Pr\left(\max_i \frac{\|\boldsymbol{w}'_{i,n}\boldsymbol{\gamma}_n\|}{\sqrt{n}}\right)\le\left(\frac{n^{-\theta/2}}{\epsilon^{2+\theta}}\right)\frac{\sum_{i=1}^n\mathbb{E}\left(\|w'_{i,n}\boldsymbol{\gamma}_n\|^{2+\theta}\right)}{n}=O\left(\frac{(\max_i\kappa_{i,n})^{2+\theta}}{n^{\theta/2}}\right),$$

which vanishes provided that $\max_i\kappa_{i,n}=O(n^{\frac{1}{2}\frac{\theta}{2+\theta}})$. The same rate requirement can equally be obtained under the alternative condition that

$$\frac{\sum_{i=1}^n\sum_{j\in(i)_n}\gamma_{n,j}^{2+\theta}}{n}=O(1),$$

together with the assumption that $\max_i\max_j\mathbb{E}(\|w_{i,n,j}\|^{2+\theta})=O(1)$, by another application of Hölder's inequality.

B Additional simulation results

We next present simulation results for the other models considered in the supplementary material to Cattaneo, Jansson and Newey (2018).

B.1 One-way panel model

The first model considered is the standard fixed-effect model for panel data. The design is similar to the design used in the main text, although here there is no randomness in the dummies and the groups do no overlap. For double-indexed data $(y_{(g,m)},x_{(g,m)})$, the model is

$$y_{(g,m)}=x_{(g,m)}\,\beta+\alpha_g+\varepsilon_{(g,m)},\qquad g=1,\dots,G,\quad m=1\dots,M,$$

and α_g is a group-specific intercept. The within-group (fixed-effect) estimator equals the ordinary least-squares estimator of $y_{(g,m)}$ on $x_{(g,m)}$ and G dummy variables that capture group membership of the individual observations. We draw $x_{(g,m)}\sim$ i.i.d. $\boldsymbol{N}(0,1)$ and $\varepsilon_{(g,m)}\sim$ i.i.d. $\boldsymbol{N}(0,1)$, set $\beta=1$ and $\alpha_g=0$ for all groups g. The samples sizes considered have $G=\lfloor 700/M\rfloor$ for $M\in\{700,10,5,4,3,2\}$, which yields a total sample size of 700 (except when $M = 3$, in which case the sample size is 699). This is the same total sample size as in our simulations in the main text.

Table A.1 reports the simulation results for this model. In the case $M = 2$ HCK does not exist and the table reports results for the standard HC0 estimator applied to the first-differenced model.

B.2 Partially-linear model

We next provide simulation results for a series estimator of the partially-linear model

$$y_i = x_i\,\beta + \exp(-\sqrt{\|\boldsymbol{z}_i\|}) + \varepsilon_i, \qquad x_i = \exp(\sqrt{\|\boldsymbol{z}_i\|}) + V_i,$$

where, again $x_i \sim$ i.i.d. $\boldsymbol{N}(0,1)$ with $\beta = 1$, and we draw $(\varepsilon_i, V_i) \sim$ i.i.d. $\boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I}_2)$ and, for $\boldsymbol{z}_i := (z_{i,1}, \ldots, z_{i,6})'$, generate each $z_{i,j} \sim$ i.i.d. Uniform$[-1,1]$. We approximate the function $\exp(-\sqrt{\|\boldsymbol{z}_i\|})$ by a power-series expansion of order κ_n. Moreover, for a given κ_n, let $\boldsymbol{w}_{i,n}$ denote the vector that collects all (unique) terms of the form $z_{i,1}^{k_1} \times z_{i,2}^{k_2} \times \cdots \times z_{i,6}^{k_6}$ with $k_1 + \cdots + k_6 = \kappa_n$. This yields $q_n = (6+\kappa_n)!/(6! \times \kappa_n!)$ as the dimension of the nuisance parameter. We then estimate β by the least-squares estimator of y_i on x_i and $\boldsymbol{w}_{i,n}$, again maintaining a sample size of 700. Note that, here, the vector $\boldsymbol{\gamma}_n$ is non-zero, but $\|\boldsymbol{\gamma}_n\| = O(1)$ because the approximation converges.

The simulation results for $\kappa_n \in \{1,2,3,4,5\}$ are collected in Table A.2.

References

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

Table A.1: One-way panel model

G	1	70	140	175	233	350
M	700	10	5	4	3	2
q_n/n	.0014	.1000	.2000	.2500	.3333	.5000
Rejection frequency of 5%-level test						
HC0	.0529	.0667	.0835	.0971	.1155	.1665
HC1	.0526	.0518	.0519	.0552	.0539	.0525
HC2	.0527	.0519	.0519	.0555	.0541	.0526
HC3	.0526	.0419	.0315	.0286	.0187	.0069
HCK	.0527	.0527	.0533	.0560	.0555	*.0526*
HCA	.0526	.0544	.0540	.0576	.0557	.0550
Average width of 95% confidence interval						
HC0	.1478	.1478	.1478	.1477	.1479	.1479
HC1	.1480	.1559	.1654	.1708	.1813	.2095
HC2	.1479	.1558	.1653	.1706	.1811	.2092
HC3	.1480	.1642	.1848	.1970	.2218	.2958
HCK	.1479	.1557	.1651	.1704	.1808	*.2092*
HCA	.1478	.1556	.1651	.1704	.1809	.2089

Table A.2: Partially-linear model

κ_n	1	2	3	4	5
q_n	7	28	84	210	462
q_n/n	.01	.04	.12	.30	.66
Rejection frequency of 5%-level test					
HC0	.0649	.0536	.0650	.0927	.1823
HC1	.0634	.0493	.0497	.0451	.0218
HC2	.0636	.0494	.0511	.0521	.0508
HC3	.0630	.0449	.0366	.0197	.0012
HCK	.0636	.0493	.0512	.0523	.0605
HCA	.0678	.0571	.0568	.0603	.0700
Width of 95% confidence interval					
HC0	.1387	.1476	.1481	.1520	.1741
HC1	.1395	.1508	.1580	.1819	.2992
HC2	.1394	.1506	.1573	.1765	.2527
HC3	.1401	.1537	.1677	.2110	.4335
HCK	.1394	.1506	.1572	.1763	.2496
HCA	.1391	.1498	.1563	.1754	.2496