Inference in Linear Regression Models with Many Covariates and Heteroskedasticity*

Supplemental Appendix

Matias D. Cattaneo[†] Michael Jansson[‡] Whitney K. Newey[§]

May 7, 2018

Abstract

This supplemental appendix contains proofs of the main theorems presented in the paper as well as other related technical results that may be of independent interest. Specific examples of linear regression models covered by our general framework are also discussed in detail, including the role of regularity conditions. Finally, complete results from a simulation study are reported.

*The first author gratefully acknowledges financial support from the National Science Foundation (SES 1459931). The second author gratefully acknowledges financial support from the National Science Foundation (SES 1459967) and the research support of CREATES (funded by the Danish National Research Foundation under grant no. DNRF78). The third author gratefully acknowledges financial support from the National Science Foundation (SES 1132399).

[†]Department of Economics and Department of Statistics, University of Michigan.

[‡]Department of Economics, UC Berkeley and *CREATES*.

[§]Department of Economics, MIT.

Contents

1 General Framework

Suppose $\{(y_{i,n}, \mathbf{x}'_{i,n}, \mathbf{w}'_{i,n}) : 1 \leq i \leq n\}$ is generated by

$$y_{i,n} = \boldsymbol{\beta}'\mathbf{x}_{i,n} + \boldsymbol{\gamma}'_n\mathbf{w}_{i,n} + u_{i,n}, \qquad i = 1, \ldots, n, \tag{SA-1}$$

Let $\|\cdot\|$ denote the Euclidean norm, set $\mathcal{X}_n = (\mathbf{x}_{1,n}, \ldots, \mathbf{x}_{n,n})$ and for a collection $\mathcal{W}_n$ of random variables satisfying $\mathbb{E}[\mathbf{w}_{i,n}|\mathcal{W}_n] = \mathbf{w}_{i,n}$ define the constants

$$\begin{aligned}
\varrho_n &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[R_{i,n}^2], \qquad R_{i,n} = \mathbb{E}[u_{i,n}|\mathcal{X}_n, \mathcal{W}_n], \\
\rho_n &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[r_{i,n}^2], \qquad r_{i,n} = \mathbb{E}[u_{i,n}|\mathcal{W}_n], \\
\chi_n &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\|\mathbf{Q}_{i,n}\|^2], \qquad \mathbf{Q}_{i,n} = \mathbb{E}[\mathbf{v}_{i,n}|\mathcal{W}_n],
\end{aligned}$$

where $\mathbf{v}_{i,n} = \mathbf{x}_{i,n} - (\sum_{j=1}^{n}\mathbb{E}[\mathbf{x}_{j,n}\mathbf{w}'_{j,n}])(\sum_{j=1}^{n}\mathbb{E}[\mathbf{w}_{j,n}\mathbf{w}'_{j,n}])^{-1}\mathbf{w}_{i,n}$ is the population counterpart of $\hat{\mathbf{v}}_{i,n}$. Also, letting $\lambda_{\min}(\cdot)$ denote the minimum eigenvalue of its argument, define

$$\mathcal{C}_n = \max_{1\leq i\leq n}\{\mathbb{E}[U_{i,n}^4|\mathcal{X}_n, \mathcal{W}_n] + \mathbb{E}[\|\mathbf{V}_{i,n}\|^4|\mathcal{W}_n] + 1/\mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n]\} + 1/\lambda_{\min}(\mathbb{E}[\tilde{\boldsymbol{\Gamma}}_n|\mathcal{W}_n]),$$

where $U_{i,n} = y_{i,n} - \mathbb{E}[y_{i,n}|\mathcal{X}_n, \mathcal{W}_n]$, $\mathbf{V}_{i,n} = \mathbf{x}_{i,n} - \mathbb{E}[\mathbf{x}_{i,n}|\mathcal{W}_n]$, $\tilde{\boldsymbol{\Gamma}}_n = \sum_{i=1}^{n}\tilde{\mathbf{V}}_{i,n}\tilde{\mathbf{V}}'_{i,n}/n$, and $\tilde{\mathbf{V}}_{i,n} = \sum_{j=1}^{n}M_{ij,n}\mathbf{V}_{j,n}$. Recall that

$$\mathbf{M}_n = \begin{pmatrix} M_{11,n} & \cdots & M_{1n,n} \\ \vdots & \ddots & \vdots \\ M_{n1,n} & \cdots & M_{nn,n} \end{pmatrix}, \qquad M_{ij,n} = \mathbb{1}(i=j) - \mathbf{w}'_{i,n}(\sum_{k=1}^{n}\mathbf{w}_{k,n}\mathbf{w}'_{k,n})^{-1}\mathbf{w}_{j,n},$$

$1 \leq i, j \leq n$. That is, $\mathbf{M}_n$ is the orthogonal projection matrix onto the complement of the column space of $\mathbf{w}_{i,n}$.

We impose the following three high-level conditions.

Assumption 1 *$\mathbb{C}[U_{i,n}, U_{j,n}|\mathcal{X}_n, \mathcal{W}_n] = 0$ for $i \neq j$ and $\max_{1\leq i\leq N_n}\#\mathcal{T}_{i,n} = O(1)$, where $\#\mathcal{T}_{i,n}$ is the cardinality of $\mathcal{T}_{i,n}$ and where $\{\mathcal{T}_{i,n} : 1 \leq i \leq N_n\}$ is a partition of $\{1, \ldots, n\}$ such that $\{(U_{t,n}, \mathbf{V}'_{t,n}) : t \in \mathcal{T}_{i,n}\}$ are independent over i conditional on $\mathcal{W}_n$.*

Assumption 2 *$\mathbb{P}[\lambda_{\min}(\sum_{i=1}^{n}\mathbf{w}_{i,n}\mathbf{w}'_{i,n}) > 0] \to 1$, $\overline{\lim}_{n\to\infty}K_n/n < 1$, and $\mathcal{C}_n = O_p(1)$.*

Assumption 3 *$\chi_n = O(1)$, $\varrho_n + n(\varrho_n - \rho_n) + n\chi_n\varrho_n = o(1)$, and $\max_{1\leq i\leq n}\|\hat{\mathbf{v}}_{i,n}\|/\sqrt{n} = o_p(1)$.*

2 Technical Lemmas

Our main results (Theorems 1–4 in the paper) are obtained by working with the representation

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \hat{\boldsymbol{\Gamma}}_n^{-1}\mathbf{S}_n,$$

where

$$\hat{\boldsymbol{\Gamma}}_n = \frac{1}{n}\sum_{1\le i\le n}\hat{\mathbf{v}}_{i,n}\hat{\mathbf{v}}_{i,n}' \qquad \text{and} \qquad \mathbf{S}_n = \frac{1}{\sqrt{n}}\sum_{1\le i\le n}\hat{\mathbf{v}}_{i,n}u_{i,n}.$$

Strictly speaking, the displayed representation is valid only when $\lambda_{\min}(\sum_{i=1}^n \mathbf{w}_{i,n}\mathbf{w}_{i,n}') > 0$ and $\lambda_{\min}(\hat{\boldsymbol{\Gamma}}_n) > 0$. Both events occur with probability approaching one under our assumptions and our main results are valid no matter which definitions (of $\hat{\beta}_n$ and $\hat{\Sigma}_n$) are employed on the complement of the union of these events, but for specificity we let $\bar{M}_{ij,n} = \omega_n M_{ij,n}$, where $\omega_n = \mathbb{1}\{\lambda_{\min}(\sum_{k=1}^n \mathbf{w}_{k,n}\mathbf{w}_{k,n}') > 0\}$, and, in a slight abuse of notation, we define

$$\hat{\boldsymbol{\Gamma}}_n = \frac{1}{n}\sum_{1\le i\le n}\hat{\mathbf{v}}_{i,n}\hat{\mathbf{v}}_{i,n}', \qquad \mathbf{S}_n = \frac{1}{\sqrt{n}}\sum_{1\le i\le n}\hat{\mathbf{v}}_{i,n}u_{i,n}, \qquad \hat{\mathbf{v}}_{i,n} = \sum_{1\le j\le n}\bar{M}_{ij,n}\mathbf{x}_{j,n},$$

and

$$\hat{\boldsymbol{\beta}}_n = \mathbb{1}\{\lambda_{\min}(\hat{\boldsymbol{\Gamma}}_n) > 0\}\hat{\boldsymbol{\Gamma}}_n^{-1}\Big(\frac{1}{n}\sum_{1\le i\le n}\hat{\mathbf{v}}_{i,n}y_{i,n}\Big).$$

We first present seven technical lemmas, some of which may be of independent interest. These technical lemmas are used to establish our main results.

The first lemma can be used to bound $\hat{\boldsymbol{\Gamma}}_n^{-1}$.

Lemma SA-1 *If Assumptions 1–3 hold, then* $\hat{\boldsymbol{\Gamma}}_n^{-1} = O_p(1)$.

Let $\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_n(\mathcal{X}_n, \mathcal{W}_n) = \mathbb{V}[\mathbf{S}_n|\mathcal{X}_n, \mathcal{W}_n]$. The second lemma can be used to bound $\boldsymbol{\Sigma}_n^{-1}$ and to show asymptotic normality of $\mathbf{S}_n$.

Lemma SA-2 *If Assumptions 1–3 hold, then* $\boldsymbol{\Sigma}_n^{-1} = O_p(1)$ *and* $\boldsymbol{\Sigma}_n^{-1/2}\mathbf{S}_n \to_d \mathcal{N}(\mathbf{0}, \mathbf{I})$.

The third lemma can be used to approximate $\hat{\sigma}_n^2$ by means of $\tilde{\sigma}_n^2$, where

$$\hat{\sigma}_n^2 = \frac{1}{n-d-K_n}\sum_{1\le i\le n}\hat{u}_{i,n}^2, \qquad \tilde{\sigma}_n^2 = \frac{1}{n-K_n}\sum_{1\le i\le n}\tilde{U}_{i,n}^2,$$

with $\hat{u}_{i,n} = \sum_{1\le j\le n}\bar{M}_{ij,n}(y_{j,n} - \hat{\boldsymbol{\beta}}_n'\mathbf{x}_{j,n})$ and $\tilde{U}_{i,n} = \sum_{1\le j\le n}\bar{M}_{ij,n}U_{j,n}$.

Lemma SA-3 *If Assumptions 1–3 hold, then* $\hat{\sigma}_n^2 = \mathbb{E}[\tilde{\sigma}_n^2|\mathcal{X}_n, \mathcal{W}_n] + o_p(1)$.

The fourth lemma can be used to approximate $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\kappa}_n)$ by means of $\tilde{\boldsymbol{\Sigma}}_n(\boldsymbol{\kappa}_n)$, where

$$\hat{\boldsymbol{\Sigma}}_n(\kappa_n) = \frac{1}{n}\sum_{1\le i,j\le n}\kappa_{ij,n}\hat{\mathbf{v}}_{i,n}\hat{\mathbf{v}}_{i,n}'\hat{u}_{j,n}^2, \qquad \tilde{\boldsymbol{\Sigma}}_n(\boldsymbol{\kappa}_n) = \frac{1}{n}\sum_{1\le i,j\le n}\kappa_{ij,n}\hat{\mathbf{v}}_{i,n}\hat{\mathbf{v}}_{i,n}'\tilde{U}_{j,n}^2.$$

Lemma SA-4 *Suppose Assumptions 1–3 hold.*
If $\|\boldsymbol{\kappa}_n\|_\infty = \max_{1\leq i\leq n}\sum_{1\leq j\leq n}|\kappa_{ij,n}| = O_p(1)$, *then* $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\kappa}_n) = \mathbb{E}[\tilde{\boldsymbol{\Sigma}}_n(\boldsymbol{\kappa}_n)|\mathcal{X}_n,\mathcal{W}_n] + o_p(1)$.

The fifth lemma can be combined with the third lemma to show consistency of $\hat{\boldsymbol{\Sigma}}_n^{\mathtt{HO}} = \hat{\sigma}_n^2\hat{\boldsymbol{\Gamma}}_n$ under homoskedasticity.

Lemma SA-5 *Suppose Assumption 1 holds.*
If $\mathbb{E}[U_{i,n}^2|\mathcal{X}_n,\mathcal{W}_n] = \sigma_n^2$, *then* $\mathbb{E}[\tilde{\sigma}_n^2|\mathcal{X}_n,\mathcal{W}_n] = \sigma_n^2\omega_n$ *and* $\boldsymbol{\Sigma}_n = \sigma_n^2\hat{\boldsymbol{\Gamma}}_n$.

The sixth lemma can be combined with the fourth lemma to show consistency of $\hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\kappa}_n)$. Part (a) is a general result stated under a high-level condition. Part (b) gives sufficient conditions for the condition of part (a) for estimators of HCk type and part (c) does likewise for $\hat{\boldsymbol{\Sigma}}_n^{\mathtt{HC}} = \hat{\boldsymbol{\Sigma}}_n(\boldsymbol{\kappa}_n^{\mathtt{HC}})$, $\boldsymbol{\kappa}_n^{\mathtt{HC}} = (\mathbf{M}_n \odot \mathbf{M}_n)^{-1}$. With a slight abuse of notation, let

$$\mathcal{M}_n = 1 - \min_{1\leq i\leq n}\bar{M}_{ii,n}.$$

Lemma SA-6 *Suppose Assumption 2 holds.*

(a) If

$$\max_{1\leq i\leq n}\left\{\left|\sum_{1\leq k\leq n}\kappa_{ik,n}\bar{M}_{ik,n}^2 - 1\right| + \sum_{1\leq j\leq n, j\neq i}\left|\sum_{1\leq k\leq n}\kappa_{ik,n}\bar{M}_{jk,n}^2\right|\right\} = o_p(1),$$

then $\mathbb{E}[\tilde{\boldsymbol{\Sigma}}_n(\boldsymbol{\kappa}_n)|\mathcal{X}_n,\mathcal{W}_n] = \boldsymbol{\Sigma}_n + o_p(1)$.

(b) Suppose $\kappa_{ij,n} = \omega_n\mathbb{1}(i=j)\Upsilon_{i,n}M_{ii,n}^{-\xi_{i,n}}$, *where* $0 \leq \xi_{i,n} \leq 4$ *and* $\Upsilon_{i,n} \geq 0$.
If $\max_{1\leq i\leq n}\{|1-\Upsilon_{i,n}|\} = o_p(1)$ *and if* $\mathcal{M}_n = o_p(1)$, *then* $\mathbb{E}[\tilde{\boldsymbol{\Sigma}}_n(\boldsymbol{\kappa}_n)|\mathcal{X}_n,\mathcal{W}_n] = \boldsymbol{\Sigma}_n + o_p(1)$ *and* $\|\boldsymbol{\kappa}_n\|_\infty = O_p(1)$.

(c) Suppose $\boldsymbol{\kappa}_n = \omega_n\boldsymbol{\kappa}_n^{\mathtt{HC}}$, *where*

$$\boldsymbol{\kappa}_n^{\mathtt{HC}} = \begin{pmatrix} \kappa_{11,n}^{\mathtt{HC}} & \cdots & \kappa_{1n,n}^{\mathtt{HC}} \\ \vdots & \ddots & \vdots \\ \kappa_{n1,n}^{\mathtt{HC}} & \cdots & \kappa_{nn,n}^{\mathtt{HC}} \end{pmatrix} = \begin{pmatrix} M_{11,n}^2 & \cdots & M_{1n,n}^2 \\ \vdots & \ddots & \vdots \\ M_{n1,n}^2 & \cdots & M_{nn,n}^2 \end{pmatrix}^{-1} = (\mathbf{M}_n \odot \mathbf{M}_n)^{-1}.$$

If $\mathbb{P}[\mathcal{M}_n < 1/2] \to 1$ *and if* $1/(1/2 - \mathcal{M}_n) = O_p(1)$, *then* $\mathbb{E}[\tilde{\boldsymbol{\Sigma}}_n(\boldsymbol{\kappa}_n)|\mathcal{X}_n,\mathcal{W}_n] = \boldsymbol{\Sigma}_n + o_p(1)$ *and* $\|\boldsymbol{\kappa}_n\|_\infty = O_p(1)$.

Finally, the seventh lemma can be used to formulate primitive sufficient conditions for the last part of Assumption 3.

Lemma SA-7 *Suppose Assumptions 1 and 2 hold and suppose that*

$$\frac{1}{n}\sum_{1\leq i\leq n}\mathbb{E}[\|\mathbf{Q}_{i,n}\|^{2+\theta}] = O(1)$$

for some $\theta \geq 0$. *If either (i)* $\theta > 0$ *and* $\mathcal{M}_n = o_p(1)$; *or (ii)* $\chi_n = o(1)$; *or (iii)* $\theta > 0$ *and* $\max_{1\leq i\leq n}\sum_{1\leq j\leq n}\mathbb{1}(M_{ij,n}\neq 0) = o_p(n^{\theta/(2\theta+2)})$, *then* $\max_{1\leq i\leq n}\|\hat{\mathbf{v}}_{i,n}\|/\sqrt{n} = o_p(1)$.

3 Properties of $\mathbf{M}_n \odot \mathbf{M}_n$

Because $\mathbf{M}_n$ is symmetric, so is $\mathbf{M}_n \odot \mathbf{M}_n$ and it follows from the Gerschgorin circle theorem (see, e.g., Barnes and Hoffman (1981) for an interesting discussion) that

$$\lambda_{\min}(\mathbf{M}_n \odot \mathbf{M}_n) \geq \min_{1\leq i\leq n}\{M_{ii,n}^2 - \sum_{1\leq j\leq n, j\neq i}|M_{ij,n}^2|\} = \min_{1\leq i\leq n}\{2M_{ii,n}^2 - \sum_{1\leq j\leq n} M_{ij,n}^2\},$$

where, using the fact that $\sum_{1\leq j\leq n} M_{ij,n}^2 = M_{ii,n}$ because $\mathbf{M}_n$ is idempotent,

$$\min_{1\leq i\leq n}\{2M_{ii,n}^2 - \sum_{1\leq j\leq n} M_{ij,n}^2\} = \min_{1\leq i\leq n}\{2M_{ii,n}^2 - M_{ii,n}\} = 2\min_{1\leq i\leq n}\{M_{ii,n}(M_{ii,n} - 1/2)\}.$$

Thus, $\lambda_{\min}(\mathbf{M}_n \odot \mathbf{M}_n) > 0$ (i.e., $\mathbf{M}_n \odot \mathbf{M}_n$ is positive definite) whenever $\mathcal{M}_n < 1/2$.

Under the same condition, $\mathbf{M}_n \odot \mathbf{M}_n$ is diagonally dominant and it follows from Theorem 1 of Varah (1975) that

$$\|(\mathbf{M}_n \odot \mathbf{M}_n)^{-1}\|_\infty \leq \frac{1}{1/2 - \mathcal{M}_n}.$$

4 Motivating Examples

This section discusses technical details and main results for three type of examples covered by our framework: (i) linear regression models with increasing dimension, (ii) semiparametric partially linear models, and (iii) fixed effects panel data regression models. Recall that the objective is to find an estimator $\hat{\mathbf{\Sigma}}_n$ of the variance of $\mathbf{S}_n = \sum_{i=1}^n \hat{\mathbf{v}}_{i,n}u_{i,n}/\sqrt{n}$ such that

$$\hat{\mathbf{\Omega}}_n^{-1/2}\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \to_d \mathcal{N}(\mathbf{0}, \mathbf{I}), \qquad \hat{\mathbf{\Omega}}_n = \hat{\mathbf{\Gamma}}_n^{-1}\hat{\mathbf{\Sigma}}_n\hat{\mathbf{\Gamma}}_n^{-1}, \tag{SA-2}$$

in which case asymptotically valid inference on $\boldsymbol{\beta}$ can be conducted in the usual way by employing the distributional approximation $\hat{\boldsymbol{\beta}}_n \overset{a}{\sim} \mathcal{N}(\boldsymbol{\beta}, \hat{\mathbf{\Omega}}_n/n)$.

4.1 Linear Regression Model with Increasing Dimension

Our first example corresponds to the classical linear regression model characterized by (SA-1) and the following assumptions.

Assumption LR1 $\{(y_{i,n}, \mathbf{x}_{i,n}', \mathbf{w}_{i,n}') : 1 \leq i \leq n\}$ are i.i.d. over i.

Assumption LR2 $\mathbb{P}[\lambda_{\min}(\sum_{i=1}^{n}\mathbf{w}_{i,n}\mathbf{w}_{i,n}')>0]\to 1$, $\overline{\lim}_{n\to\infty}K_n/n<1$, and $\mathcal{C}_n^{\texttt{LR}}=O_p(1)$, where

$$\begin{aligned}\mathcal{C}_n^{\texttt{LR}} &= \max_{1\leq i\leq n}\{\mathbb{E}[u_{i,n}^4|\mathbf{x}_{i,n},\mathbf{w}_{i,n}]+\mathbb{E}[\|\mathbf{V}_{i,n}\|^4|\mathbf{w}_{i,n}]\} \\ &\quad+\max_{1\leq i\leq n}\{1/\mathbb{E}[u_{i,n}^2|\mathbf{x}_{i,n},\mathbf{w}_{i,n}]+1/\lambda_{\min}(\mathbb{E}[\mathbf{V}_{i,n}\mathbf{V}_{i,n}'|\mathbf{w}_{i,n}])\},\end{aligned}$$

with $\mathbf{V}_{i,n}=\mathbf{x}_{i,n}-\mathbb{E}[\mathbf{x}_{i,n}|\mathbf{w}_{i,n}]$.

Assumption LR3 $\mathbb{E}[\|\mathbf{x}_{i,n}\|^2]=O(1)$, $n\mathbb{E}[(\mathbb{E}[u_{i,n}|\mathbf{x}_{i,n},\mathbf{w}_{i,n}])^2]=o(1)$, and $\max_{1\leq i\leq n}\|\hat{\mathbf{v}}_{i,n}\|/\sqrt{n}=o_p(1)$.

The main difference between Assumptions LR1-LR3 and those familiar from the fixed-K_n case is the presence of the conditions $n\mathbb{E}[(\mathbb{E}[u_{i,n}|\mathbf{x}_{i,n},\mathbf{w}_{i,n}])^2]=o(1)$ and $\max_{1\leq i\leq n}\|\hat{\mathbf{v}}_{i,n}\|/\sqrt{n}=o_p(1)$ in Assumption LR3. The first condition is implied by the classical "exogeneity" assumption $\mathbb{E}[u_{i,n}|\mathbf{x}_{i,n},\mathbf{w}_{i,n}]=0$, but is somewhat weaker because it allows for a (vanishing) misspecification error in the linear model specification. As for the second condition, at the present level of generality it seems difficult to formulate primitive sufficient conditions for $\max_{1\leq i\leq n}\|\hat{\mathbf{v}}_{i,n}\|/\sqrt{n}=o_p(1)$ that cover all cases of interest, but for completeness we mention that under mild moment conditions it follows from Lemma SA-8 below that it suffices to require that one of the following conditions hold:

(i) $\mathcal{M}_n=o_p(1)$, or

(ii) $\chi_n^{\texttt{LR}}=\min_{\boldsymbol{\delta}\in\mathbb{R}^{K_n\times d}}\mathbb{E}[\|\mathbb{E}(\mathbf{x}_{i,n}|\mathbf{w}_{i,n})-\boldsymbol{\delta}'\mathbf{w}_{i,n}\|^2]=o(1)$, or

(iii) $\max_{1\leq i\leq n}\sum_{j=1}^{n}\mathbb{1}(M_{ij,n}\neq 0)=o_p(n^{1/3})$.

Each of these conditions is interpretable. First, $\mathcal{M}_n\geq K_n/n$ because $\sum_{i=1}^{n}M_{ii,n}=n-K_n$ and a necessary condition for **(i)** is therefore that $K_n/n\to 0$. Conversely, because

$$\mathcal{M}_n\leq\frac{K_n}{n}\frac{1-\min_{1\leq i\leq n}M_{ii,n}}{1-\max_{1\leq i\leq n}M_{ii,n}},$$

the condition $K_n/n\to 0$ is sufficient for **(i)** whenever the design is "approximately balanced" in the sense that $(1-\min_{1\leq i\leq n}M_{ii,n})/(1-\max_{1\leq i\leq n}M_{ii,n})=O_p(1)$. In other words, **(i)** requires and effectively covers the case where it is assumed that K_n is a vanishing fraction of n. In contrast, conditions **(ii)** and **(iii)** can hold also when K_n is a non-vanishing fraction of n, which is the case of primary interest in this paper.

Because **(ii)** is a requirement on the accuracy of the approximation

$$\mathbb{E}[\mathbf{x}_{i,n}|\mathbf{w}_{i,n}]\approx\boldsymbol{\delta}_n'\mathbf{w}_{i,n},\qquad\boldsymbol{\delta}_n=\mathbb{E}[\mathbf{w}_{i,n}\mathbf{w}_{i,n}']^{-1}\mathbb{E}[\mathbf{w}_{i,n}\mathbf{x}_{i,n}'],$$

primitive conditions for it are available when the elements of $\mathbf{w}_{i,n}$ are approximating functions, as in the partially linear model example discussed next. Indeed, in such cases one typically has $\chi_n^{\texttt{LR}}=O(K_n^{-\alpha})$ for some $\alpha>0$, so condition **(ii)** not only accommodates $K_n/n\nrightarrow 0$, but actually places no upper bound on the magnitude of K_n in important special cases.

Finally, condition **(iii)** is useful to handle cases where $\mathbf{w}_{i,n}$ cannot be interpreted as approximating functions, but rather just many different covariates included in the linear model specification. This condition is a "sparsity" condition on the matrix $\mathbf{M}_n$, which allows for $K_n/n \nrightarrow 0$. Although somewhat stronger than needed, the condition is easy to verify in certain cases, including linear regression models with dummy variables, the partially linear model with "locally bounded" bases of approximation, and linear panel data models with fixed effect, as further illustrated below.

Specializing Theorems 2–4 in the paper to this linear regression model, we obtain the following result.

Theorem LR Suppose Assumptions LR1–LR3 hold.

(a) If $\mathbb{E}[u_{i,n}^2|\mathbf{x}_{i,n},\mathbf{z}_{i,n}] = \sigma_n^2$, then (SA-2) holds with $\hat{\mathbf{\Sigma}}_n = \hat{\mathbf{\Sigma}}_n^{\mathtt{HO}}$.

(b) If $\mathcal{M}_n \to_p 0$, then (SA-2) holds with $\hat{\mathbf{\Sigma}}_n = \hat{\mathbf{\Sigma}}_n^{\mathtt{EW}}$.

(c) If $\mathbb{P}[\mathcal{M}_n < 1/2] \to 1$ and if $1/(1/2 - \mathcal{M}_n) = O_p(1)$, then (SA-2) holds with $\hat{\mathbf{\Sigma}}_n = \hat{\mathbf{\Sigma}}_n^{\mathtt{HC}}$.

This theorem gives a formal justification for employing $\hat{\mathbf{\Sigma}}_n^{\mathtt{HC}}$ as the variance estimator when forming confidence intervals for $\boldsymbol{\beta}$ in linear models with possibly many nuisance covariates and heteroskedasticity. The resulting confidence intervals for $\boldsymbol{\beta}$ will remain consistent even when K_n is proportional to n, provided the technical conditions given in part (c) are satisfied.

4.1.1 Gaussian Regressors

The linear regression model discussed above, and each of the other models discussed in the upcoming sections, distinguishes between the main covariates of interest $\mathbf{x}_{i,n}$ and the additional nuisance covariates $\mathbf{w}_{i,n}$. This distinction is important not only conceptually but also technically because the high-level restrictions imposed on the distribution of these covariates are indeed quite different in general (see, for example, Assumptions LR1–LR3 above). However, an important exception occurs when the "long" vector of covariates $(\mathbf{x}_{i,n}', \mathbf{w}_{i,n}')$ is assumed to have a Gaussian distribution. In this exceptional case, the role of the covariates $\mathbf{x}_{i,n}$ and $\mathbf{w}_{i,n}$ is indeed interchangeable and therefore our results apply to *any* finite dimensional subvector of $(\mathbf{x}_{i,n}', \mathbf{w}_{i,n}')$. The main goal of this subsection is to illustrate this finding.

To begin, we assume that $\{(y_{i,n}, \mathbf{x}_{i,n}', \mathbf{w}_{i,n}') : 1 \leq i \leq n\}$ are i.i.d., and furthermore let $(\mathbf{x}_{i,n}', \mathbf{w}_{i,n}')' \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The mean-zero, variance-identity assumption is a useful normalization. Under the Gaussian distributional assumption on the covariates,

$$\mathbf{w}_{i,n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \qquad \text{and} \qquad \mathbf{W}_n = \sum_{i=1}^{n} \mathbf{w}_{i,n}\mathbf{w}_{i,n}' \sim \mathcal{W}_{K_n}(n, \mathbf{I})$$

with $\mathcal{W}_K(a, \mathbf{B})$ denoting the Wishart distribution. These conditions immediately imply Assumptions LR1–LR3 provided that $n\mathbb{E}[(\mathbb{E}[u_{i,n}|\mathbf{x}_{i,n}, \mathbf{w}_{i,n}])^2] = o(1)$ and $\mathcal{C}_n^{\mathtt{LR}} = O_p(1)$. In particular, $\max_{1\leq i\leq n} \mathbb{E}[u_{i,n}^4|\mathbf{x}_{i,n}, \mathbf{w}_{i,n}] = O_p(1)$ and $\max_{1\leq i\leq n} 1/\mathbb{E}[u_{i,n}^2|\mathbf{x}_{i,n}, \mathbf{w}_{i,n}] = O_p(1)$ can be verified, for example, if the conditional heteroskedasticity is multiplicative, as well as bounded and bounded away from zero. Similarly, $\max_{1\leq i\leq n}\mathbb{E}[\|\mathbf{V}_{i,n}\|^4|\mathbf{w}_{i,n}]=O_p(1)$ and $\max_{1\leq i\leq n}1/\lambda_{\min}(\mathbb{E}[\mathbf{V}_{i,n}\mathbf{V}_{i,n}'|\mathbf{w}_{i,n}])=O_p(1)$ can be verified by restricting the covariance matrix of $(\mathbf{x}_{i,n}',\mathbf{w}_{i,n}')$. The other conditions in Assumptions LR1–LR3 are easy to verify due to the joint Gaussian distributional assumption on the vector of covariates $(\mathbf{x}_{i,n}',\mathbf{w}_{i,n}')$. Thus, asymptotic normality of the least squares estimator of any finite subvector of the parameters accompanying $(\mathbf{x}_{i,n}',\mathbf{w}_{i,n}')$ holds (Theorem 1 in the paper).

Finally, to obtain the conclusions of Theorem LR it remains to study the properties of

$$\mathcal{M}_n=1-\min_{1\leq i\leq n}M_{ii,n}=\max_{1\leq i\leq n}P_{ii,n},\qquad P_{ii,n}=\mathbf{w}_{i,n}'\mathbf{W}_n^{-1}\mathbf{w}_{i,n},$$

and, in particular, to give conditions under which either $\mathcal{M}_n\to_p 0$ for part (b), or $\mathbb{P}[\mathcal{M}_n<1/2]\to 1$ and $1/(1/2-\mathcal{M}_n)=O_p(1)$ for part (c).

For all $i=1,2,\cdots,n$, we have

$$P_{ii,n}=\frac{P_{ii,n}^{(i)}}{1+P_{ii,n}^{(i)}},\qquad P_{ii,n}^{(i)}=\mathbf{w}_{i,n}'\mathbf{W}_{(i),n}^{-1}\mathbf{w}_{i,n},\qquad \mathbf{W}_{(i),n}=\sum_{j=1,j\neq i}^{n}\mathbf{w}_{j,n}\mathbf{w}_{j,n}',$$

where $\mathbf{w}_{i,n}\perp\!\!\!\perp\mathbf{W}_{(i),n}$. Standard properties about quadratic forms of normal random variables imply that

$$P_{ii,n}^{(i)}\sim\frac{K_n}{n-K_n}\mathcal{F}(K_n,n-K_n),$$

with $\mathcal{F}(a,b)$ denoting the F distribution, and hence

$$P_{ii,n}=\frac{P_{ii,n}^{(i)}}{1+P_{ii,n}^{(i)}}\sim\frac{\frac{K_n}{n-K_n}\mathcal{F}(K_n,n-K_n)}{1+\frac{K_n}{n-K_n}\mathcal{F}(K_n,n-K_n)}\sim\mathcal{B}\left(\frac{K_n}{2},\frac{n-K_n}{2}\right),$$

with $\mathcal{B}(a,b)$ denoting the Beta distribution. It follows that

$$\mathbb{E}[P_{ii,n}^{\vartheta}]=\prod_{\ell=0}^{\vartheta-1}\frac{K_n+2\ell}{n+2\ell},\qquad \vartheta=1,2,\cdots$$

and, in particular,

$$\mathbb{E}[P_{ii,n}]=\frac{K_n}{n},\qquad \mathbb{E}[P_{ii,n}^2]=\frac{K_n(K_n+2)}{n(n+2)},$$

$$\mathbb{E}[P_{ii,n}^3]=\frac{K_n(K_n+2)(K_n+4)}{n(n+2)(n+4)},\qquad \mathbb{E}[P_{ii,n}^4]=\frac{K_n(K_n+2)(K_n+4)(K_n+6)}{n(n+2)(n+4)(n+6)}.$$

Therefore, using these results, we obtain

$$\max_{1\leq i\leq n}\left|P_{ii,n}-\frac{K_n}{n}\right|=O_p(n^{-1/4})$$

because

$$\mathbb{E}\left[\max_{1\leq i\leq n}|P_{ii,n}-\mathbb{E}[P_{ii,n}]|\right]\leq n^{1/\vartheta}\left(\max_{1\leq i\leq n}\mathbb{E}\left[|P_{ii,n}-\mathbb{E}[P_{ii,n}]|^{\vartheta}\right]\right)^{1/\vartheta}, \qquad \vartheta>1,$$

and

$$\mathbb{E}\left[|P_{ii,n}-\mathbb{E}[P_{ii,n}]|^4\right]=\frac{12}{n^2}\frac{K^4n-12K^4-2K^3n^2+24K^3n+K^2n^3-16K^2n^2+4Kn^3}{n^5+12n^4+44n^3+48n^2}=O(n^{-2}).$$

In conclusion, the assumption $\mathbf{w}_{i,n}\sim\mathcal{N}(\mathbf{0},\mathbf{I})$ implies $\mathcal{M}_n\to_p 0$ if $K_n/n\to 0$, and $\mathbb{P}[\mathcal{M}_n<1/2]\to 1$ and $1/(1/2-\mathcal{M}_n)=O_p(1)$ if $\overline{\lim}_{n\to\infty}K_n/n<1/2$. This, in turn, implies that the conclusions in Theorem LR hold for the least squares estimator of any finite subvector of parameters accompanying the covariates $(\mathbf{x}'_{i,n},\mathbf{w}'_{i,n})$ in the linear regression model, under the joint Gaussian distributional assumption.

4.2 Fixed Effects Panel Data Regression Model

A second class of examples covered by our results are related to linear panel data models with multi-way fixed effects. For example, Stock and Watson (2008) consider heteroskedasticity-robust inference for the panel data regression model

$$Y_{it}=\alpha_i+\boldsymbol{\beta}'\mathbf{X}_{it}+U_{it}, \qquad i=1,\ldots,N, \qquad t=1,\ldots,T, \tag{SA-3}$$

where $\alpha_i\in\mathbb{R}$ is an individual-specific intercept, $\mathbf{X}_{it}\in\mathbb{R}^d$ is a regressor of dimension d, $U_{it}\in\mathbb{R}$ is an error term, and the following assumptions are satisfied.

Assumption FE1 $\{(U_{i1},\ldots,U_{iT},\mathbf{X}'_{i1}\ldots,\mathbf{X}'_{iT}):1\leq i\leq n\}$ are independent over i, $T\geq 3$ is fixed, and $\mathbb{E}[U_{it}U_{is}|\mathbf{X}_{i1}\ldots,\mathbf{X}_{iT}]=0$ for $t\neq s$.

Assumption FE2 $\mathcal{C}_N^{\mathtt{FE}}=O_p(1)$, where

$$\mathcal{C}_N^{\mathtt{FE}} = \max_{1\leq i\leq N,1\leq t\leq T}\{\mathbb{E}[U_{it}^4|\mathbf{X}_{i1}\ldots,\mathbf{X}_{iT}]+\mathbb{E}[\|\mathbf{X}_{it}\|^4]\} + \max_{1\leq i\leq N,1\leq t\leq T}\{1/\mathbb{E}[U_{it}^2|\mathbf{X}_{i1}\ldots,\mathbf{X}_{iT}]+1/\lambda_{\min}(\mathbb{E}[\tilde{\mathbf{V}}_{it}\tilde{\mathbf{V}}'_{it}])\},$$

with $\tilde{\mathbf{V}}_{it}=\mathbf{X}_{it}-\mathbb{E}[\mathbf{X}_{it}]-T^{-1}\sum_{s=1}^T(\mathbf{X}_{is}-\mathbb{E}[\mathbf{X}_{is}])$.

Assumption FE3 $\mathbb{E}[U_{it}|\mathbf{X}_{i1}\ldots,\mathbf{X}_{iT}]=0$.

Defining $n=NT$, $K_n=N$, $\boldsymbol{\gamma}_n=(\alpha_1,\ldots,\alpha_N)'$, and

$$(y_{(i-1)T+t,n},\mathbf{x}'_{(i-1)T+t,n},u_{(i-1)T+t,n},\mathbf{w}'_{(i-1)T+t,n})=(Y_{it},\mathbf{X}'_{it},U_{it},\mathbf{e}'_{i,N}), \qquad 1\leq i\leq N,\ 1\leq t\leq T,$$

where $\mathbf{e}_{i,N}\in\mathbb{R}^N$ is the i-th unit vector of dimension N, the model (SA-3) is also of the form (SA-1) and $\hat{\boldsymbol{\beta}}_n$ is the fixed effects estimator of $\boldsymbol{\beta}$. In general, this model does not satisfy Assumption

LR1, but Assumption FE1 enables us to employ results for independent random variables when developing asymptotics. In other respects this model is in fact more tractable than the previous models due to the special nature of the covariates $\mathbf{w}_{i,n}$.

One implication of Assumptions FE1 and FE3 is that $\mathbb{E}[Y_{it}|\mathbf{X}_{i1},\ldots,\mathbf{X}_{iT}] = \alpha_i + \boldsymbol{\beta}'\mathbf{X}_{it}$, where α_i can depend on i and the conditioning variables $(\mathbf{X}_{i1},\ldots,\mathbf{X}_{iT})$ in an arbitrary way. In the spirit of "fixed effects" (as opposed to "correlated random effects"), Assumptions FE1–FE3 further allow $\mathbb{V}[Y_{it}|\mathbf{X}_{i1},\ldots,\mathbf{X}_{iT}]$ to depend not only on $(\mathbf{X}_{i1},\ldots,\mathbf{X}_{iT})$, but also on i. In particular, unlike Stock and Watson (2008), we do not require $(U_{i1},\ldots,U_{iT},\mathbf{X}'_{i1}\ldots,\mathbf{X}'_{iT})$ to be i.i.d. over i. In addition, we do not require any kind of stationarity on the part of $(U_{it},\mathbf{X}'_{it})$. The amount of variance heterogeneity permitted is quite large, as Assumption FE2 basically only requires $\mathbb{V}[Y_{it}|\mathbf{X}_{i1},\ldots,\mathbf{X}_{iT}] = \mathbb{E}[U_{it}^2|\mathbf{X}_{i1},\ldots,\mathbf{X}_{iT}]$ to be bounded and bounded away from zero. (On the other hand, serial correlation is assumed away because Assumptions FE1 and FE3 imply that $\mathbb{C}[Y_{it},Y_{is}|\mathbf{X}_{i1},\ldots,\mathbf{X}_{iT}] = 0$ for $t \neq s$.)

Because $K_n/n = 1/T$ is fixed this model does not admit an analog of Theorem 3 in the paper. On the other hand, it does admit an analog of Theorems 2 and 4 in the paper.

Theorem FE Suppose Assumptions FE1–FE3 hold. Then (SA-2) holds with $\hat{\boldsymbol{\Sigma}}_n = \hat{\boldsymbol{\Sigma}}_n^{\texttt{HC}}$. If also $\mathbb{E}[U_{it}^2|\mathbf{X}_{i1},\ldots,\mathbf{X}_{iT}] = \sigma^2$, then (SA-2) holds with $\hat{\boldsymbol{\Sigma}}_n = \hat{\boldsymbol{\Sigma}}_n^{\texttt{HO}}$.

Consistency under homoskedasticity follows from Lemma SA-3: the fixed effects α_i are not consistently estimated, but the estimator $\hat{\sigma}_n^2$ averages over the noisy estimates, and hence $\hat{\sigma}_n^2 \to_p \sigma^2$. To see the connection between our results and those in Stock and Watson (2008), observe that $\mathbf{M}_n = \mathbf{I}_N \otimes [\mathbf{I}_T - \boldsymbol{\iota}_T\boldsymbol{\iota}'_T/T]$ for $\boldsymbol{\iota}_T \in \mathbb{R}^T$ a $T \times 1$ vector of ones. We then obtain $M_{ii,n} = 1 - 1/T$ (for $i = 1,\ldots,n$) and therefore $\mathcal{M}_n \leq 1/3$ because $T \geq 3$. More importantly, perhaps, we obtain a closed-form expression for $\kappa_n^{\texttt{HC}}$ given by

$$\boldsymbol{\kappa}_n^{\texttt{HC}} = \mathbf{I}_N \otimes \frac{T}{T-2}\left[\mathbf{I}_T - \frac{1}{(T-1)^2}\boldsymbol{\iota}_T\boldsymbol{\iota}'_T\right].$$

As a consequence,

$$\hat{\boldsymbol{\Sigma}}_n^{\texttt{HC}} = \frac{1}{N(T-2)}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{\mathbf{X}}_{it}\tilde{\mathbf{X}}'_{it}\hat{U}_{it}^2 - \frac{1}{N(T-2)}\sum_{i=1}^{N}\left(\frac{1}{T-1}\sum_{t=1}^{T}\tilde{\mathbf{X}}_{it}\tilde{\mathbf{X}}'_{it}\right)\left(\frac{1}{T-1}\sum_{t=1}^{T}\hat{U}_{it}^2\right),$$

where $\tilde{\mathbf{X}}_{it} = \mathbf{X}_{it} - T^{-1}\sum_{s=1}^{T}\mathbf{X}_{is}$ and $\hat{U}_{it} = Y_{it} - T^{-1}\sum_{s=1}^{T}Y_{is} - \hat{\boldsymbol{\beta}}'_n\tilde{\mathbf{X}}_{it}$. Apart from an asymptotically negligible degrees of freedom correction, this estimator coincides with the estimator $\hat{\Sigma}^{HR-FE}$ of Stock and Watson (2008, Eq. (6), p. 156).

The generic variance estimator $\hat{\boldsymbol{\Sigma}}_n^{\texttt{HC}}$ is not well defined when $T = 2$. Nevertheless, in the special case of the one-way fixed effects linear panel model a simple, case-specific alternative valid inference method is available if the model is transformed. To be more precise, heteroskedastic-robust inference for $\boldsymbol{\beta}$ is straightforward if the following first-differences model is considered:

$$\Delta Y_i = \boldsymbol{\beta}' \Delta \mathbf{X}_i + \Delta U_i, \qquad i = 1, \ldots, N,$$

where $\Delta Y_i = Y_{i2} - Y_{i1}$, $\Delta \mathbf{X}_i = \mathbf{X}_{i2} - \mathbf{X}_{i1}$ and $\Delta U_i = U_{i2} - U_{i1}$. Conventional standard errors can be used to conduct valid inference on the least-squares estimator of $\boldsymbol{\beta}$ in the first-differences model, which is robust to large K_n by construction. In fact, the resulting least-squares estimator is numerically equivalent to the one-way fixed effects estimator discussed above.

The results above not only highlight a tight connection between our general standard error estimator and the one in Stock and Watson (2008), but also indicate that our general formula $\hat{\boldsymbol{\Sigma}}_n^{\texttt{HC}}$ could be used to derive explicit, simple expressions in other contexts where multi-way fixed effects or similar discrete regressors are included. For a second concrete example, see the recent work of Verdier (2017) in the context of linear models with two-way unobserved heterogeneity and sparsely matched data.

4.3 Semiparametric Partially Linear Model

Another model covered by our results is the partially linear model

$$y_i = \boldsymbol{\beta}' \mathbf{x}_i + g(\mathbf{z}_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{SA-4}$$

where $\mathbf{x}_i$ and $\mathbf{z}_i$ are explanatory variables, ε_i is an error term, and the function $g(\mathbf{z})$ is unknown. Suppose $\{p^k(\mathbf{z}) : k = 1, 2, \cdots\}$ are functions having the property that linear combinations can approximate square-integrable functions of $\mathbf{z}$ well, in which case $g(\mathbf{z}_i) \approx \boldsymbol{\gamma}_n' \mathbf{p}_n(\mathbf{z}_i)$ for some $\boldsymbol{\gamma}_n$, where $\mathbf{p}_n(\mathbf{z}) = (p^1(\mathbf{z}), \ldots, p^{K_n}(\mathbf{z}))'$. Defining $y_{i,n} = y_i$, $\mathbf{x}_{i,n} = \mathbf{x}_i$, $\mathbf{w}_{i,n} = \mathbf{p}_n(\mathbf{z}_i)$, and $u_{i,n} = \varepsilon_i + g(\mathbf{z}_i) - \boldsymbol{\gamma}_n' \mathbf{w}_{i,n}$, the model (SA-4) is of the form (SA-1), and $\hat{\boldsymbol{\beta}}_n$ is the series estimator of $\boldsymbol{\beta}$. See Cattaneo, Jansson, and Newey (2018) for references.

Let $\mathbf{h}(\mathbf{z}_i) = \mathbb{E}[\mathbf{x}_i | \mathbf{z}_i]$. Our analysis of $\hat{\boldsymbol{\beta}}_n$ proceeds under the following assumptions.

Assumption PL1 $\{(y_i, \mathbf{x}_i', \mathbf{z}_i') : 1 \le i \le n\}$ are i.i.d. over i.

Assumption PL2 $\mathbb{P}[\lambda_{\min}(\sum_{i=1}^n \mathbf{p}_n(\mathbf{z}_i)\mathbf{p}_n(\mathbf{z}_i)') > 0] \to 1$, $\overline{\lim}_{n\to\infty} K_n/n < 1$, and $\mathcal{C}_n^{\texttt{PL}} = O_p(1)$, where

$$\mathcal{C}_n^{\texttt{PL}} = \max_{1 \le i \le n} \{\mathbb{E}[\varepsilon_i^4 | \mathbf{x}_i, \mathbf{z}_i] + \mathbb{E}[\|\boldsymbol{\nu}_i\|^4 | \mathbf{z}_i] + 1/\mathbb{E}[\varepsilon_i^2 | \mathbf{x}_i, \mathbf{z}_i] + 1/\lambda_{\min}(\mathbb{E}[\boldsymbol{\nu}_i \boldsymbol{\nu}_i' | \mathbf{z}_i])\},$$

with $\boldsymbol{\nu}_i = \mathbf{x}_i - \mathbb{E}[\mathbf{x}_i | \mathbf{z}_i]$.

Assumption PL3 $\mathbb{E}[\varepsilon_i | \mathbf{x}_i, \mathbf{z}_i] = 0$, $\varrho_n^{\texttt{PL}} = o(1)$, $\chi_n^{\texttt{PL}} = O(1)$, and $n \varrho_n^{\texttt{PL}} \chi_n^{\texttt{PL}} = o(1)$, where

$$\varrho_n^{\texttt{PL}} = \min_{\boldsymbol{\gamma} \in \mathbb{R}^{K_n}} \mathbb{E}[|g(\mathbf{z}_i) - \boldsymbol{\gamma}' \mathbf{p}_n(\mathbf{z}_i)|^2], \qquad \chi_n^{\texttt{PL}} = \min_{\boldsymbol{\delta} \in \mathbb{R}^{K_n \times d}} \mathbb{E}[\|\mathbf{h}(\mathbf{z}_i) - \boldsymbol{\delta}' \mathbf{p}_n(\mathbf{z}_i)\|^2],$$

and $\max_{1 \le i \le n} \|\hat{\mathbf{v}}_{i,n}\| / \sqrt{n} = o_p(1)$.

Because $g(\mathbf{z}_i) \neq \boldsymbol{\gamma}_n'\mathbf{p}_n(\mathbf{z}_i)$ in general, the partially linear model does not (necessarily) satisfy $\mathbb{E}[u_{i,n}|\mathbf{x}_{i,n}, \mathbf{w}_{i,n}] = 0$. The approach taken here, made precise in Assumption PL3, is motivated by the fact that linear combinations of $\{\mathbf{p}^k(\mathbf{z})\}$ are assumed to be able to approximate the functions $g(\mathbf{z})$ and possibly $\mathbf{h}(\mathbf{z})$ well. For further technical details see, for example, Newey (1997), Chen (2007), Cattaneo and Farrell (2013), and Belloni, Chernozhukov, Chetverikov, and Kato (2015).

In particular, under standard smoothness conditions, and for standard choices of basis functions, we have $\varrho_n^{\texttt{PL}} = O(K_n^{-\alpha_g})$ and $\chi_n^{\texttt{PL}} = O(K_n^{-\alpha_h})$ for some pair (α_g, α_h) of positive constants, in which case Assumption PL2 holds provided $K_n^{\alpha_g+\alpha_h}/n \to \infty$. Furthermore, in this case, $\chi_n^{\texttt{PL}} = O(K_n^{-\alpha_h}) = o(1)$ and it therefore follows from Lemma SA-7 that $\max_{1\leq i\leq n} \|\hat{\mathbf{v}}_{i,n}\|/\sqrt{n} = o_p(1)$.

Overall, our results impose only weak smoothness conditions on the underlying unknown functions $g(\mathbf{z})$ and $\mathbf{h}(\mathbf{z})$. Furthermore, our results do not even require $\chi_n^{\texttt{PL}} = o(1)$ in general. To illustrate this point, suppose only $\chi_n^{\texttt{PL}} \leq \mathbb{E}[\|\mathbf{h}(\mathbf{z}_i)\|^2] = O(1)$ is assumed, and consider the special case of partitioning estimators over evenly spaced blocks; see Cattaneo and Farrell (2013) for details. To be more precise, setting $\text{supp}(\mathbf{z}_i) = [0,1]^d$ for simplicity, partition the support of $\mathbf{z}_i$ into evenly spaced disjoint hyper-cubes, subclasses or blocks: $\text{supp}(\mathbf{z}_i) = \bigcup_{\ell=1}^{K_n} \mathcal{B}_{\ell,n}$ with $\mathcal{B}_{\ell,n} \cap \mathcal{B}_{\ell,n} = \emptyset$ and $|\mathcal{B}_{\ell,n}| = 1/K_n$. Partitioning estimators approximate the unknown function by fitting a p-th degree polynomial regression within each block. In this case, $\mathbf{M}_n$ becomes a simple banded matrix: defining $\mathbf{p}_{n,\ell}(\mathbf{z}_i) = \mathbb{1}(\mathbf{z}_i \in \mathcal{B}_{\ell,n})\mathbf{r}(\mathbf{z}_i)$ with $\mathbf{r}(\mathbf{z}) = (1, \mathbf{z}, \cdots, \mathbf{z}^p)'$, we obtain

$$\sum_{1\leq j\leq n} \mathbb{1}(M_{ij,n} \neq 0) = p+1,$$

and hence the conditions of Lemma SA-7 can be verified under simple regularity conditions when $\overline{\lim}_{n\to\infty} K_n/n < 1$. This discussion extends to other "locally supported" basis (i.e., any basis that generates a banded projection matrix), under appropriate regularity conditions. In particular, for example, note that zero order partitioning estimators over evenly spaced blocks are numerically equal to uniform splines of order zero.

The results for the partially linear model (SA-4) are in perfect analogy with those for the linear regression model.

Theorem PL Suppose Assumptions PL1–PL3 hold.

(a) If $\mathbb{E}[\varepsilon_i^2|\mathbf{x}_i, \mathbf{z}_i] = \sigma^2$, then (SA-2) holds with $\hat{\boldsymbol{\Sigma}}_n = \hat{\boldsymbol{\Sigma}}_n^{\texttt{HO}}$.

(b) If $\mathcal{M}_n \to_p 0$, then (SA-2) holds with $\hat{\boldsymbol{\Sigma}}_n = \hat{\boldsymbol{\Sigma}}_n^{\texttt{EW}}$.

(c) If $\mathbb{P}[\mathcal{M}_n < 1/2] \to 1$ and if $1/(1/2 - \mathcal{M}_n) = O_p(1)$, then (SA-2) holds with $\hat{\boldsymbol{\Sigma}}_n = \hat{\boldsymbol{\Sigma}}_n^{\texttt{HC}}$.

A result similar to Theorem PL(a) was previously reported in Cattaneo, Jansson, and Newey (2018) under strictly stronger assumptions relative to those used herein. Furthermore, parts (b) and (c) of Theorem PL are new to the literature, providing in particular valid inference in (saturated) semi-linear models with possibly many basis functions of approximations.

5 Proofs of Main Results

Theorem 1 in the paper follows from Lemmas SA-1 and SA-2. Theorem 2 in the paper follows from Theorem 1 combined with Lemmas SA-3 and SA-5. Theorems 3 and 4 in the paper follow from Theorem 1 combined with Lemmas SA-4 and SA-6.

5.1 Linear Regression Model with Increasing Dimension

If Assumption LR1 holds, then Assumption 1 holds with $\mathcal{W}_n = (\mathbf{w}_{1,n}, \ldots, \mathbf{w}_{n,n})$, $N_n = n$, $\mathcal{T}_{i,n} = \{i\}$, and $\max_{1\leq i\leq N_n} \#\mathcal{T}_{i,n} = 1$. Moreover, $\chi_n \leq \max_{1\leq i\leq n} \mathbb{E}[\|\mathbf{x}_{i,n}\|^2]$ and $\rho_n \leq \varrho_n = \mathbb{E}[(\mathbb{E}[u_{i,n}|\mathbf{x}_{i,n}, \mathbf{w}_{i,n}])^2]$, so Assumption 3 holds if Assumptions LR1 and LR3 hold. Finally, Assumption 2 is implied by Assumptions LR1–LR3. In particular,

$$\begin{aligned}
\lambda_{\min}(\mathbb{E}[\tilde{\boldsymbol{\Gamma}}_n|\mathcal{W}_n]) &= \lambda_{\min}(\frac{1}{n}\sum_{1\leq i\leq n} \mathbb{E}[\tilde{\mathbf{V}}_{i,n}\tilde{\mathbf{V}}_{i,n}'|\mathbf{w}_{i,n}]) \\
&= \omega_n \lambda_{\min}(\frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n}\mathbb{E}[\mathbf{V}_{i,n}\mathbf{V}_{i,n}'|\mathbf{w}_{i,n}]) \\
&\geq \omega_n \frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n}\lambda_{\min}(\mathbb{E}[\mathbf{V}_{i,n}\mathbf{V}_{i,n}'|\mathbf{w}_{i,n}]) \\
&\geq \omega_n (\frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n}) \min_{1\leq i\leq n} \lambda_{\min}(\mathbb{E}[\mathbf{V}_{i,n}\mathbf{V}_{i,n}'|\mathbf{w}_{i,n}]) \\
&\geq \omega_n (1 - K_n/n)/\mathcal{C}_n^{\mathtt{LR}},
\end{aligned}$$

so $1/\lambda_{\min}(\mathbb{E}[\tilde{\boldsymbol{\Gamma}}_n|\mathcal{W}_n]) = O_p(1)$ because $\mathbb{P}[\omega_n = 1] \to 1$, $\overline{\lim}_{n\to\infty} K_n/n < 1$, and $\mathcal{C}_n^{\mathtt{LR}} = O_p(1)$.

Under Assumptions LR1 and LR2, we have

$$\chi_n = \chi_n^{\mathtt{LR}} = \min_{\delta\in\mathbb{R}^{K_n\times d}} \mathbb{E}[\|\mathbb{E}(\mathbf{x}_{i,n}|\mathbf{w}_{i,n}) - \boldsymbol{\delta}'\mathbf{w}_{i,n}\|^2] = \mathbb{E}[\|\mathbf{Q}_{i,n}\|^2],$$

where

$$\mathbf{Q}_{i,n} = \mathbb{E}[\mathbf{v}_{i,n}|\mathbf{w}_{i,n}], \qquad \mathbf{v}_{i,n} = \mathbf{x}_{i,n} - \mathbb{E}[\mathbf{x}_{i,n}\mathbf{w}_{i,n}']\mathbb{E}[\mathbf{w}_{i,n}\mathbf{w}_{i,n}']^{-1}\mathbf{w}_{i,n}.$$

Setting $\theta = 2$ in Lemma SA-7 and specializing it to the linear regression model with increasing dimension we therefore obtain the following lemma, whose conditions were discussed above.

Lemma SA-8 *Suppose Assumptions LR1 and LR2 hold and suppose that* $\mathbb{E}[\|\mathbf{x}_{i,n}\|^2] = O(1)$, $\mathbb{E}[u_{i,n}|\mathbf{x}_{i,n}, \mathbf{w}_{i,n}] = 0$, *and* $\mathbb{E}[\|\mathbf{Q}_{i,n}\|^4] = O(1)$. *If either (i)* $\mathcal{M}_n = o_p(1)$; *or (ii)* $\chi_n^{\mathtt{LR}} = o(1)$; *or (iii)* $\max_{1\leq i\leq n}\sum_{1\leq j\leq n} \mathbb{1}(M_{ij,n} \neq 0) = o_p(n^{1/3})$, *then* $\max_{1\leq i\leq n} \|\hat{\mathbf{v}}_{i,n}\|/\sqrt{n} = o_p(1)$.

5.2 Fixed Effects Panel Data Regression Model

If Assumption FE1 holds, then Assumption 1 holds with $\mathcal{W}_n = (\mathbf{w}_{1,n}, \ldots, \mathbf{w}_{n,n})$, $N_n = N = n/T$, $\mathcal{T}_{i,n} = \{T(i-1)+1, \ldots, Ti\}$, and $\max_{1\leq i\leq N_n} \#\mathcal{T}_{i,n} = T$. Moreover, $\chi_n \leq \max_{1\leq i\leq N, 1\leq t\leq T} \mathbb{E}[\|\mathbf{X}_{it}\|^2]$, so Assumption 3 holds (with $\varrho_n = \rho_n = 0$) when Assumptions FE1–FE3 hold, the condition $\max_{1\leq i\leq n} \|\hat{\mathbf{v}}_{i,n}\|/\sqrt{n} = o_p(1)$ holding by Lemma SA-7 because $\sum_{1\leq j\leq n} \mathbb{1}(M_{ij,n} \neq 0) = T$. Finally, Assumption 2 is implied by Assumptions FE1–FE3. In particular,

$$\begin{aligned}\lambda_{\min}(\mathbb{E}[\tilde{\mathbf{\Gamma}}_n|\mathcal{W}_n]) &= \lambda_{\min}(\frac{1}{NT}\sum_{1\leq i\leq N,1\leq t\leq T}\mathbb{E}[\tilde{\mathbf{V}}_{it}\tilde{\mathbf{V}}_{it}']) \\ &\geq \min_{1\leq i\leq N,1\leq t\leq T}\lambda_{\min}(\mathbb{E}[\tilde{\mathbf{V}}_{it}\tilde{\mathbf{V}}_{it}']) \geq 1/\mathcal{C}_n^{\mathtt{FE}},\end{aligned}$$

so $1/\lambda_{\min}(\mathbb{E}[\tilde{\mathbf{\Gamma}}_n|\mathcal{W}_n]) = O_p(1)$ because $\mathcal{C}_n^{\mathtt{FE}} = O_p(1)$.

5.3 Semiparametric Partially Linear Model

If Assumption PL1 holds, then Assumption 1 holds with $\mathcal{W}_n = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$, $N_n = n$, $\mathcal{T}_{i,n} = \{i\}$, and $\max_{1\leq i\leq N_n} \#\mathcal{T}_{i,n} = 1$. Moreover, in this case we have

$$\chi_n = \min_{\delta\in\mathbb{R}^{K_n\times d}} \mathbb{E}[\|\mathbb{E}(\mathbf{x}_i|\mathbf{z}_i) - \boldsymbol{\delta}'\mathbf{p}_n(\mathbf{z}_i)\|^2] = \chi_n^{\mathtt{PL}}$$

and, using $\mathbb{E}[\mathbf{y}_i - \boldsymbol{\beta}'\mathbf{x}_i|\mathbf{x}_i,\mathbf{z}_i] = g(\mathbf{z}_i) = \mathbb{E}[\mathbf{y}_i - \boldsymbol{\beta}'\mathbf{x}_i|\mathbf{z}_i]$,

$$\varrho_n = \min_{\gamma\in\mathbb{R}^{K_n}} \mathbb{E}[|\mathbb{E}[\mathbf{y}_i - \boldsymbol{\beta}'\mathbf{x}_i|\mathbf{z}_i] - \boldsymbol{\gamma}'\mathbf{p}_n(\mathbf{z}_i)|^2] = \min_{\gamma\in\mathbb{R}^{K_n}} \mathbb{E}[|\mathbb{E}[y_i - \boldsymbol{\beta}'\mathbf{x}_i|\mathbf{x}_i,\mathbf{z}_i] - \boldsymbol{\gamma}'\mathbf{p}_n(\mathbf{z}_i)|^2] = \rho_n = \varrho_n^{\mathtt{PL}},$$

so Assumption 3 holds when Assumptions PL1 and PL3 hold. Finally, Assumption 2 is implied by Assumptions PL1–PL3. In particular,

$$\begin{aligned}\lambda_{\min}(\mathbb{E}[\tilde{\mathbf{\Gamma}}_n|\mathcal{W}_n]) &= \omega_n\lambda_{\min}(\frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n}\mathbb{E}[\boldsymbol{\nu}_i\boldsymbol{\nu}_i'|\mathbf{z}_i]) \\ &\geq \omega_n\frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n}\lambda_{\min}(\mathbb{E}[\boldsymbol{\nu}_i\boldsymbol{\nu}_i'|\mathbf{z}_i]) \\ &\geq \omega_n(\frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n})\min_{1\leq i\leq n}\lambda_{\min}(\mathbb{E}[\boldsymbol{\nu}_i\boldsymbol{\nu}_i'|\mathbf{z}_i]) \\ &\geq \omega_n(1-K_n/n)/\mathcal{C}_n^{\mathtt{PL}},\end{aligned}$$

so $1/\lambda_{\min}(\mathbb{E}[\tilde{\mathbf{\Gamma}}_n|\mathcal{W}_n]) = O_p(1)$ because $\mathbb{P}[\omega_n = 1] \to 1$, $\overline{\lim}_{n\to\infty} K_n/n < 1$, and $\mathcal{C}_n^{\mathtt{PL}} = O_p(1)$.

6 Proofs of Technical Lemmas

Throughout the proofs we simplify the notation by assuming without loss of generality that $d = 1$. In Lemma SA-2 the case where $d > 1$ can be handled by means of the Cramér-Wold device and simple bounding arguments.

6.1 Proof of Lemma SA-1

It suffices to show that $\tilde{\Gamma}_n = \mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n] + o_p(1)$ and that $\hat{\Gamma}_n - \tilde{\Gamma}_n \geq o_p(1)$.

First,

$$\tilde{\Gamma}_n = \frac{1}{n}\sum_{1\leq i\leq N_n} a_{ii,n} + \frac{2}{n}\sum_{1\leq i,j\leq N_n, i<j} a_{ij,n}, \qquad a_{ij,n} = \sum_{s\in\mathcal{T}_{i,n}, t\in\mathcal{T}_{j,n}} \bar{M}_{st,n}V_{s,n}V_{t,n},$$

where $\sum_{1\leq i,j\leq N_n}\mathbb{V}[a_{ij,n}|\mathcal{W}_n] = O_p(n)$ because

$$\mathbb{V}[a_{ij,n}|\mathcal{W}_n] \leq (\#\mathcal{T}_{i,n})(\#\mathcal{T}_{j,n})\sum_{s\in\mathcal{T}_{i,n}, t\in\mathcal{T}_{j,n}} \bar{M}_{st,n}^2\mathbb{V}[V_{s,n}V_{t,n}|\mathcal{W}_n] \leq \mathcal{C}_{\mathcal{T},n}^2\mathcal{C}_{V,n}\sum_{s\in\mathcal{T}_{i,n}, t\in\mathcal{T}_{j,n}} \bar{M}_{st,n}^2,$$

where $\mathcal{C}_{\mathcal{T},n} = \max_{1\leq i\leq N_n}\#(\mathcal{T}_{i,n}) = O_p(1)$, $\mathcal{C}_{V,n} = 1 + \max_{1\leq i\leq n}\mathbb{E}[\|V_{i,n}\|^4|\mathcal{W}_n] = O_p(1)$ and

$$\sum_{1\leq i,j\leq N_n}\sum_{s\in\mathcal{T}_{i,n}, t\in\mathcal{T}_{j,n}} \bar{M}_{st,n}^2 = \sum_{1\leq i,j\leq n}\bar{M}_{ij,n}^2 = \sum_{1\leq i\leq n}\bar{M}_{ii,n} \leq n.$$

As a consequence,

$$\mathbb{V}[\frac{1}{n}\sum_{1\leq i\leq N_n} a_{ii,n}|\mathcal{W}_n] = \frac{1}{n^2}\sum_{1\leq i\leq N_n}\mathbb{V}[a_{ii,n}|\mathcal{W}_n] \leq \frac{1}{n^2}\sum_{1\leq i,j\leq N_n}\mathbb{V}[a_{ij,n}|\mathcal{W}_n] = o_p(1)$$

and

$$\mathbb{V}[\frac{1}{n}\sum_{1\leq i,j\leq N_n, i<j} a_{ij,n}|\mathcal{W}_n] = \frac{1}{n^2}\sum_{1\leq i,j\leq N_n, i<j}\mathbb{V}[a_{ij,n}|\mathcal{W}_n] \leq \frac{1}{n^2}\sum_{1\leq i,j\leq N_n}\mathbb{V}[a_{ij,n}|\mathcal{W}_n] = o_p(1),$$

implying in particular that $\tilde{\Gamma}_n = \mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n] + o_p(1)$.

Next, defining $\tilde{Q}_{i,n} = \sum_{1\leq j\leq n}\bar{M}_{ij,n}Q_{j,n}$, we have

$$\hat{\Gamma}_n - \tilde{\Gamma}_n = \frac{1}{n}\sum_{1\leq i\leq n}\tilde{Q}_{i,n}^2 + \frac{2}{n}\sum_{1\leq i\leq n}\tilde{Q}_{i,n}\tilde{V}_{i,n} \geq \frac{2}{n}\sum_{1\leq i\leq n}\tilde{Q}_{i,n}\tilde{V}_{i,n} = \frac{2}{n}\sum_{1\leq i\leq n}\tilde{Q}_{i,n}V_{i,n} = o_p(1),$$

the last equality using the fact that $\mathbb{E}[\tilde{Q}_{i,n}V_{i,n}|\mathcal{W}_n] = 0$ and

$$\begin{aligned}
\mathbb{V}[\frac{1}{n}\sum_{1\leq i\leq n}\tilde{Q}_{i,n}V_{i,n}|\mathcal{W}_n] &= \frac{1}{n^2}\sum_{1\leq i\leq N_n}\mathbb{V}[\sum_{s\in\mathcal{T}_{i,n}}\tilde{Q}_{s,n}V_{s,n}|\mathcal{W}_n] \\
&\leq \frac{1}{n^2}\sum_{1\leq i\leq N_n}(\#\mathcal{T}_{i,n})\sum_{s\in\mathcal{T}_{i,n}}\mathbb{V}[\tilde{Q}_{s,n}V_{s,n}|\mathcal{W}_n] \leq \frac{1}{n}\mathcal{C}_{\mathcal{T},n}\mathcal{C}_{V,n}(\frac{1}{n}\sum_{1\leq i\leq N_n}\sum_{s\in\mathcal{T}_{i,n}}\tilde{Q}_{s,n}^2) \\
&\leq \frac{1}{n}\mathcal{C}_{\mathcal{T},n}\mathcal{C}_{V,n}(\frac{1}{n}\sum_{1\leq i\leq n}Q_{i,n}^2) = \frac{1}{n}O_p(\chi_n) = o_p(1).
\end{aligned}$$

6.2 Proof of Lemma SA-2

Defining $\tilde{S}_n = S_n - \mathbb{E}[S_n|\mathcal{X}_n, \mathcal{W}_n] = \sum_{1\leq i\leq n} \hat{v}_{i,n}U_{i,n}/\sqrt{n}$ and employing the decomposition

$$S_n - \tilde{S}_n = \frac{1}{\sqrt{n}}\sum_{1\leq i\leq n} \tilde{V}_{i,n}r_{i,n} + \frac{1}{\sqrt{n}}\sum_{1\leq i\leq n} \tilde{Q}_{i,n}r_{i,n} + \frac{1}{\sqrt{n}}\sum_{1\leq i\leq n} \hat{v}_{i,n}(R_{i,n} - r_{i,n}),$$

we begin by showing that $S_n = \tilde{S}_n + o_p(1)$.

First, defining $\tilde{r}_{i,n} = \sum_{1\leq j\leq n} \bar{M}_{ij,n}r_{j,n}$ and using $\mathbb{E}[\tilde{r}_{i,n}V_{i,n}|\mathcal{W}_n] = 0$ and

$$\begin{aligned}\mathbb{V}[\frac{1}{\sqrt{n}}\sum_{1\leq i\leq n} \tilde{r}_{i,n}V_{i,n}|\mathcal{W}_n] &= \frac{1}{n}\sum_{1\leq i\leq N_n} \mathbb{V}[\sum_{s\in\mathcal{T}_{i,n}} \tilde{r}_{s,n}V_{s,n}|\mathcal{W}_n] \leq \frac{1}{n}\sum_{1\leq i\leq N_n}(\#\mathcal{T}_{i,n})\sum_{s\in\mathcal{T}_{i,n}} \mathbb{V}[\tilde{r}_{s,n}V_{s,n}|\mathcal{W}_n] \\ &\leq \mathcal{C}_{\mathcal{T},n}\mathcal{C}_{V,n}\frac{1}{n}\sum_{1\leq i\leq n}\tilde{r}_{i,n}^2 \leq \mathcal{C}_{\mathcal{T},n}\mathcal{C}_{V,n}\frac{1}{n}\sum_{1\leq i\leq n} r_{i,n}^2 = O_p(\rho_n) = O_p(\varrho_n) = o_p(1),\end{aligned}$$

we have

$$\frac{1}{\sqrt{n}}\sum_{1\leq i\leq n} \tilde{V}_{i,n}r_{i,n} = \frac{1}{\sqrt{n}}\sum_{1\leq i\leq n} \tilde{r}_{i,n}V_{i,n} = o_p(1).$$

Also, using the Cauchy-Schwarz inequality,

$$|\frac{1}{\sqrt{n}}\sum_{1\leq i\leq n} \tilde{Q}_{i,n}r_{i,n}|^2 \leq n(\frac{1}{n}\sum_{1\leq i\leq n} Q_{i,n}^2)(\frac{1}{n}\sum_{1\leq i\leq n} r_{i,n}^2) = O_p(n\chi_n\rho_n) = o_p(1)$$

and

$$|\frac{1}{\sqrt{n}}\sum_{1\leq i\leq n} \hat{v}_{i,n}(R_{i,n} - r_{i,n})|^2 \leq n(\frac{1}{n}\sum_{1\leq i\leq n} \hat{v}_{i,n}^2)(\frac{1}{n}\sum_{1\leq i\leq n} |R_{i,n} - r_{i,n}|^2) = O_p[n(\varrho_n - \rho_n)] = o_p(1),$$

where the penultimate equality uses

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

and $\mathbb{E}[|R_{i,n} - r_{i,n}|^2] = \mathbb{E}[R_{i,n}^2] - \mathbb{E}[r_{i,n}^2]$. As a consequence, $S_n = \tilde{S}_n + o_p(1)$.

Next, using Assumption 1,

$$\begin{aligned}\Sigma_n &= \frac{1}{n}\sum_{1\leq i\leq N_n}\sum_{t\in\mathcal{T}_i} \hat{v}_{t,n}^2\mathbb{E}[U_{t,n}^2|\mathcal{X}_n, \mathcal{W}_n] = \frac{1}{n}\sum_{1\leq i\leq n} \hat{v}_{i,n}^2\mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n] \\ &\geq \hat{\Gamma}_n \min\nolimits_{1\leq i\leq n} \mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n],\end{aligned}$$

so $\Sigma_n^{-1} = O_p(1)$. The proof can therefore be completed by showing that $\Sigma_n^{-1/2}\tilde{S}_n \to_d \mathcal{N}(0, 1)$.

We shall do so assuming without loss of generality that $\lambda_{\min}(\Sigma_n) > 0$ (a.s.). Because

$$\Sigma_n^{-1/2}\tilde{S}_n = \frac{1}{\sqrt{n}} \sum_{1 \le i \le N_n} \eta_{i,n}, \qquad \eta_{i,n} = \Sigma_n^{-1/2} \sum_{t \in \mathcal{T}_i} \hat{v}_{t,n} U_{t,n},$$

where, conditional on $(\mathcal{X}_n, \mathcal{W}_n)$, $\eta_{i,n}$ are mean zero independent random variables with

$$\frac{1}{n} \sum_{1 \le i \le N_n} \mathbb{V}[\eta_{i,n} | \mathcal{X}_n, \mathcal{W}_n] = 1,$$

it follows from the Berry-Esseen inequality that

$$\sup_{z \in \mathbb{R}} |\mathbb{P}(\Sigma_n^{-1/2}\tilde{S}_n \le z | \mathcal{X}_n, \mathcal{W}_n) - \Phi(z)| \le \min(\frac{\sum_{1 \le i \le N_n} \mathbb{E}[|\eta_{i,n}|^3 | \mathcal{X}_n, \mathcal{W}_n]}{n^{3/2}}, 1),$$

where $\Phi(\cdot)$ is the standard normal cdf. It therefore suffices to show that

$$\frac{1}{n^{3/2}} \sum_{1 \le i \le N_n} \mathbb{E}[|\eta_{i,n}|^3 | \mathcal{X}_n, \mathcal{W}_n] = o_p(1).$$

Now,

$$\begin{aligned}
\frac{1}{n^{3/2}} \sum_{1 \le i \le N_n} \mathbb{E}[|\eta_{i,n}|^3 | \mathcal{X}_n, \mathcal{W}_n] &\le \lambda_{\min}(\Sigma_n)^{-3/2} \frac{1}{n^{3/2}} \sum_{1 \le i \le N_n} \mathbb{E}[| \sum_{t \in \mathcal{T}_i} \hat{v}_{t,n} U_{t,n}|^3 | \mathcal{X}_n, \mathcal{W}_n] \\
&\le \lambda_{\min}(\Sigma_n)^{-3/2} \frac{1}{n^{3/2}} \sum_{1 \le i \le N_n} (\#\mathcal{T}_i)^2 \sum_{t \in \mathcal{T}_i} |\hat{v}_{t,n}|^3 \mathbb{E}[|U_{t,n}|^3 | \mathcal{X}_n, \mathcal{W}_n] \\
&\le \mathcal{C}_{\mathcal{T},n}^2 \mathcal{C}_{U,n} \lambda_{\min}(\Sigma_n)^{-3/2} \frac{1}{n^{3/2}} \sum_{1 \le i \le N_n} \sum_{t \in \mathcal{T}_i} |\hat{v}_{t,n}|^3 = o_p(1),
\end{aligned}$$

where $\mathcal{C}_{U,n} = 1 + \max_{1 \le i \le n} \mathbb{E}[U_{i,n}^4 | \mathcal{X}_n, \mathcal{W}_n] = O_p(1)$ and where the last equality uses the fact that

$$\frac{1}{n^{3/2}} \sum_{1 \le i \le N_n} \sum_{t \in \mathcal{T}_i} |\hat{v}_{t,n}|^3 = \frac{1}{n^{3/2}} \sum_{1 \le i \le n} |\hat{v}_{i,n}|^3 \le (\frac{\max_{1 \le i \le n} |\hat{v}_{i,n}|}{\sqrt{n}})(\frac{1}{n} \sum_{1 \le i \le n} \hat{v}_{i,n}^2) = o_p(1).$$

6.3 Proof of Lemma SA-3

It suffices to show that $\tilde{\sigma}_n^2 = \mathbb{E}[\tilde{\sigma}_n^2 | \mathcal{X}_n, \mathcal{W}_n] + o_p(1) = O_p(1)$ and that $\sum_{1 \le i \le n} (\hat{u}_{i,n} - \tilde{U}_{i,n})^2 = o_p(n)$.

First,

$$\tilde{\sigma}_n^2 = \frac{1}{n - K_n} \sum_{1 \le i \le N_n} b_{ii,n} + \frac{2}{n - K_n} \sum_{1 \le i,j \le N_n, i<j} b_{ij,n}, \qquad b_{ij,n} = \sum_{s \in \mathcal{T}_{i,n}, t \in \mathcal{T}_{j,n}} \bar{M}_{st,n} U_{s,n} U_{t,n},$$

where $\sum_{1\leq i,j\leq N_n}\mathbb{V}[b_{ij,n}|\mathcal{X}_n,\mathcal{W}_n]=O_p(n)$ because

$$\begin{aligned}\mathbb{V}[b_{ij,n}|\mathcal{X}_n,\mathcal{W}_n] &\leq (\#\mathcal{T}_{i,n})(\#\mathcal{T}_{j,n})\sum_{s\in\mathcal{T}_{i,n},t\in\mathcal{T}_{j,n}}\bar{M}_{st,n}^2\mathbb{V}[U_{s,n}U_{t,n}|\mathcal{X}_n,\mathcal{W}_n]\\ &\leq \mathcal{C}_{\mathcal{T},n}^2\mathcal{C}_{U,n}\sum_{s\in\mathcal{T}_{i,n},t\in\mathcal{T}_{j,n}}\bar{M}_{st,n}^2\leq \mathcal{C}_{\mathcal{T},n}^2\mathcal{C}_{U,n}n.\end{aligned}$$

As a consequence,

$$\begin{aligned}\mathbb{V}[\frac{1}{n-K_n}\sum_{1\leq i\leq N_n}b_{ii,n}|\mathcal{X}_n,\mathcal{W}_n] &= \frac{1}{(n-K_n)^2}\sum_{1\leq i\leq N_n}\mathbb{V}[b_{ii,n}|\mathcal{X}_n,\mathcal{W}_n]\\ &\leq \frac{1}{(n-K_n)^2}\sum_{1\leq i,j\leq N_n}\mathbb{V}[b_{ij,n}|\mathcal{X}_n,\mathcal{W}_n]=o_p(1)\end{aligned}$$

and

$$\begin{aligned}\mathbb{V}[\frac{1}{n-K_n}\sum_{1\leq i,j\leq N_n,i<j}b_{ij,n}|\mathcal{X}_n,\mathcal{W}_n] &= \frac{1}{(n-K_n)^2}\sum_{1\leq i,j\leq N_n,i<j}\mathbb{V}[b_{ij,n}|\mathcal{X}_n,\mathcal{W}_n]\\ &\leq \frac{1}{(n-K_n)^2}\sum_{1\leq i,j\leq N_n}\mathbb{V}[b_{ij,n}|\mathcal{X}_n,\mathcal{W}_n]=o_p(1),\end{aligned}$$

implying in particular that $\tilde{\sigma}_n^2=\mathbb{E}[\tilde{\sigma}_n^2|\mathcal{X}_n,\mathcal{W}_n]+o_p(1)$, where

$$\mathbb{E}[\tilde{\sigma}_n^2|\mathcal{X}_n,\mathcal{W}_n]=\frac{1}{n-K_n}\sum_{1\leq i\leq n}\bar{M}_{ii,n}\mathbb{E}[U_{i,n}^2|\mathcal{X}_n,\mathcal{W}_n]\leq\mathcal{C}_{U,n}=O_p(1).$$

Next, by Lemmas SA-1 and SA-2 and their proofs, $\tilde{\Gamma}_n(\hat{\beta}_n-\beta)^2=o_p(1)$. Also, using $\varrho_n\to 0$, we have

$$\frac{1}{n}\sum_{1\leq i\leq n}\tilde{R}_{i,n}^2\leq\frac{1}{n}\sum_{1\leq i\leq n}R_{i,n}^2=O_p(\varrho_n)=o_p(1).$$

As a consequence, using $\hat{u}_{i,n}-\tilde{U}_{i,n}=\tilde{R}_{i,n}-\tilde{V}_{i,n}(\hat{\beta}_n-\beta)$,

$$\sum_{1\leq i\leq n}(\hat{u}_{i,n}-\tilde{U}_{i,n})^2\leq 2n[\frac{1}{n}\sum_{1\leq i\leq n}\tilde{R}_{i,n}^2+\tilde{\Gamma}_n(\hat{\beta}_n-\beta)^2]=o_p(n).$$

6.4 Proof of Lemma SA-4

It suffices to show that $\hat{\Sigma}_n(\kappa_n)=\tilde{\Sigma}_n(\kappa_n)+o_p(1)$ and that $\tilde{\Sigma}_n(\kappa_n)=\mathbb{E}[\tilde{\Sigma}_n(\kappa_n)|\mathcal{X}_n,\mathcal{W}_n]+o_p(1)$.

First,

$$\tilde{\Sigma}_n(\kappa_n)=\frac{1}{n}\sum_{1\leq i\leq N_n}c_{ii,n}+\frac{2}{n}\sum_{1\leq i,j\leq N_n,i<j}c_{ij,n},$$

$$c_{ij,n}=\sum_{s\in\mathcal{T}_{i,n},t\in\mathcal{T}_{j,n}}\sum_{1\leq k,l\leq n}\kappa_{kl,n}\hat{v}_{k,n}^2\bar{M}_{sl,n}\bar{M}_{tl,n}U_{s,n}U_{t,n},$$

where $\sum_{1\leq i,j\leq N_n}\mathbb{V}[c_{ij,n}|\mathcal{X}_n,\mathcal{W}_n]=o_p(n^2)$ because

$$\begin{aligned}
\mathbb{V}[c_{ij,n}|\mathcal{X}_n,\mathcal{W}_n] &\leq (\#\mathcal{T}_{i,n})(\#\mathcal{T}_{j,n})\sum_{s\in\mathcal{T}_{i,n},t\in\mathcal{T}_{j,n}}(\sum_{1\leq k,l\leq n}\kappa_{kl,n}\hat{v}_{k,n}^2\bar{M}_{sl,n}\bar{M}_{tl,n})^2\mathbb{V}[U_{s,n}U_{t,n}|\mathcal{X}_n,\mathcal{W}_n] \\
&\leq \mathcal{C}_{\mathcal{T},n}^2\mathcal{C}_{U,n}\sum_{s\in\mathcal{T}_{i,n},t\in\mathcal{T}_{j,n}}\sum_{1\leq k,l,K,L\leq n}\kappa_{kl,n}\kappa_{KL,n}\hat{v}_{k,n}^2\hat{v}_{K,n}^2\bar{M}_{sl,n}\bar{M}_{tl,n}\bar{M}_{sL,n}\bar{M}_{tL,n} \\
&= \mathcal{C}_{\mathcal{T},n}^2\mathcal{C}_{U,n}\sum_{1\leq i,j\leq n}\sum_{1\leq k,l,K,L\leq n}\kappa_{kl,n}\kappa_{KL,n}\hat{v}_{k,n}^2\hat{v}_{K,n}^2\bar{M}_{il,n}\bar{M}_{jl,n}\bar{M}_{iL,n}\bar{M}_{jL,n} \\
&= \mathcal{C}_{\mathcal{T},n}^2\mathcal{C}_{U,n}\sum_{1\leq k,l,K,L\leq n}\kappa_{kl,n}\kappa_{KL,n}\hat{v}_{k,n}^2\hat{v}_{K,n}^2\bar{M}_{lL,n}^2 \\
&\leq \mathcal{C}_{\mathcal{T},n}^2\mathcal{C}_{U,n}\sum_{1\leq k,l,K,L\leq n}|\kappa_{kl,n}||\kappa_{KL,n}|\hat{v}_{k,n}^2\hat{v}_{K,n}^2\bar{M}_{lL,n}^2
\end{aligned}$$

and

$$\begin{aligned}
\sum_{1\leq k,l,K,L\leq n}|\kappa_{kl,n}||\kappa_{KL,n}|\hat{v}_{k,n}^2\hat{v}_{K,n}^2\bar{M}_{lL,n}^2 &\leq (\max_{1\leq i\leq n}\hat{v}_{i,n}^2)\sum_{1\leq k,l,K,L\leq n}|\kappa_{kl,n}||\kappa_{KL,n}|\hat{v}_{K,n}^2\bar{M}_{lL,n}^2 \\
&\leq (\max_{1\leq i\leq n}\hat{v}_{i,n}^2)\|\kappa_n\|_\infty\sum_{1\leq l,K,L\leq n}|\kappa_{KL,n}|\hat{v}_{K,n}^2\bar{M}_{lL,n}^2 \\
&\leq (\max_{1\leq i\leq n}\hat{v}_{i,n}^2)\|\kappa_n\|_\infty\sum_{1\leq K,L\leq n}|\kappa_{KL,n}|\hat{v}_{K,n}^2 \\
&\leq n^2(\frac{\max_{1\leq i\leq n}|\hat{v}_{i,n}|}{\sqrt{n}})^2\|\kappa_n\|_\infty^2(\frac{1}{n}\sum_{1\leq i\leq n}\hat{v}_{i,n}^2)=o_p(n^2).
\end{aligned}$$

As a consequence,

$$\mathbb{V}[\frac{1}{n}\sum_{1\leq i\leq N_n}c_{ii,n}|\mathcal{X}_n,\mathcal{W}_n]=\frac{1}{n^2}\sum_{1\leq i\leq N_n}\mathbb{V}[c_{ii,n}|\mathcal{X}_n,\mathcal{W}_n]\leq\frac{1}{n^2}\sum_{1\leq i,j\leq N_n}\mathbb{V}[c_{ij,n}|\mathcal{X}_n,\mathcal{W}_n]=o_p(1)$$

and

$$\mathbb{V}[\frac{1}{n}\sum_{1\leq i,j\leq N_n,i<j}c_{ij,n}|\mathcal{X}_n,\mathcal{W}_n]=\frac{1}{n^2}\sum_{1\leq i,j\leq N_n,i<j}\mathbb{V}[c_{ij,n}|\mathcal{X}_n,\mathcal{W}_n]\leq\frac{1}{n^2}\sum_{1\leq i,j\leq N_n}\mathbb{V}[c_{ij,n}|\mathcal{X}_n,\mathcal{W}_n]=o_p(1).$$

In particular, $\tilde{\Sigma}_n(\kappa_n)=\mathbb{E}[\tilde{\Sigma}_n(\kappa_n)|\mathcal{X}_n,\mathcal{W}_n]+o_p(1)$, where

$$\begin{aligned}
|\mathbb{E}[\tilde{\Sigma}_n(\kappa_n)|\mathcal{X}_n,\mathcal{W}_n]| &\leq \frac{1}{n}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}|\hat{v}_{j,n}^2\mathbb{E}[\tilde{U}_{i,n}^2|\mathcal{X}_n,\mathcal{W}_n]\leq\mathcal{C}_{U,n}\frac{1}{n}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}|\hat{v}_{j,n}^2 \\
&\leq \mathcal{C}_{U,n}\|\kappa_n\|_\infty(\frac{1}{n}\sum_{1\leq i\leq n}\hat{v}_{i,n}^2)=O_p(1).
\end{aligned}$$

To complete the proof it suffices to show that

$$\hat{\Sigma}_n(\kappa_n) - \tilde{\Sigma}_n(\kappa_n) = \frac{1}{n}\sum_{1\le i,j\le n} \kappa_{ij,n}\hat{v}_{j,n}^2([\tilde{U}_{i,n} + \tilde{R}_{i,n} - \tilde{V}_{i,n}(\hat{\beta}_n - \beta)]^2 - \tilde{U}_{i,n}^2) = o_p(1).$$

To do so, it suffices (by the Cauchy-Schwarz inequality and using $\hat{v}_{j,n} = \tilde{V}_{j,n} + \tilde{Q}_{j,n}$, $\tilde{R}_{i,n} = \tilde{r}_{i,n} + (\tilde{R}_{i,n} - \tilde{r}_{i,n})$, and $\tilde{\Sigma}_n(\kappa_n) = O_p(1)$) to show that

$$\frac{1}{n}\sum_{1\le i,j\le n} |\kappa_{ij,n}|\tilde{V}_{j,n}^2\tilde{r}_{i,n}^2 = o_p(1), \qquad \frac{1}{n}\sum_{1\le i,j\le n} |\kappa_{ij,n}|\tilde{V}_{j,n}^2|\tilde{R}_{i,n} - \tilde{r}_{i,n}|^2 = o_p(1),$$

$$\frac{1}{n}\sum_{1\le i,j\le n} |\kappa_{ij,n}|\tilde{Q}_{j,n}^2\tilde{R}_{i,n}^2 = o_p(1), \qquad (\hat{\beta}_n - \beta)^2\frac{1}{n}\sum_{1\le i,j\le n} |\kappa_{ij,n}|\hat{v}_{j,n}^2\tilde{V}_{i,n}^2 = o_p(1).$$

First, $n^{-1}\sum_{1\le i,j\le n} |\kappa_{ij,n}|\tilde{V}_{j,n}^2\tilde{r}_{i,n}^2 = o_p(1)$ because

$$\begin{aligned}\mathbb{E}[\frac{1}{n}\sum_{1\le i,j\le n} |\kappa_{ij,n}|\tilde{V}_{j,n}^2\tilde{r}_{i,n}^2|\mathcal{W}_n] &= \frac{1}{n}\sum_{1\le i\le n}\tilde{r}_{i,n}^2\sum_{1\le j\le n}|\kappa_{ij,n}|\mathbb{E}[\tilde{V}_{j,n}^2|\mathcal{W}_n] \\ &\le \mathcal{C}_{\mathcal{T},n}\mathcal{C}_{V,n}\|\kappa_n\|_\infty\frac{1}{n}\sum_{1\le i\le n}\tilde{r}_{i,n}^2 = O_p(\rho_n) = o_p(1),\end{aligned}$$

where the inequality uses

$$\begin{aligned}\mathbb{E}[\tilde{V}_{i,n}^2|\mathcal{W}_n] &= \mathbb{E}[(\sum_{1\le j\le n}\bar{M}_{ij,n}V_{j,n})^2|\mathcal{W}_n] = \mathbb{E}[(\sum_{1\le j\le N_n}\sum_{s\in\mathcal{T}_{j,n}}\bar{M}_{is,n}V_{s,n})^2|\mathcal{W}_n] \\ &= \sum_{1\le j\le N_n}\mathbb{E}[(\sum_{s\in\mathcal{T}_{j,n}}\bar{M}_{is,n}V_{s,n})^2|\mathcal{W}_n] \\ &\le \sum_{1\le j\le N_n}(\#\mathcal{T}_{j,n})\sum_{s\in\mathcal{T}_{j,n}}\bar{M}_{is,n}^2\mathbb{E}[V_{s,n}^2|\mathcal{W}_n] \le \mathcal{C}_{\mathcal{T},n}\mathcal{C}_{V,n}\sum_{1\le j\le N_n}\sum_{s\in\mathcal{T}_{j,n}}\bar{M}_{is,n}^2 \\ &= \mathcal{C}_{\mathcal{T},n}\mathcal{C}_{V,n}\sum_{1\le j\le n}\bar{M}_{ij,n}^2 = \mathcal{C}_{\mathcal{T},n}\mathcal{C}_{V,n}\bar{M}_{ii,n} \le \mathcal{C}_{\mathcal{T},n}\mathcal{C}_{V,n}.\end{aligned}$$

Next,

$$\begin{aligned}\frac{1}{n}\sum_{1\le i,j\le n}|\kappa_{ij,n}|\tilde{V}_{j,n}^2|\tilde{R}_{i,n} - \tilde{r}_{i,n}|^2 &\le n\mathcal{C}_{\kappa,n}(\frac{1}{n}\sum_{1\le i\le n}\tilde{V}_{i,n}^2)(\frac{1}{n}\sum_{1\le i\le n}|\tilde{R}_{i,n} - \tilde{r}_{i,n}|^2) \\ &= O_p[n(\varrho_n - \rho_n)] = o_p(1)\end{aligned}$$

and

$$\frac{1}{n}\sum_{1\le i,j\le n}|\kappa_{ij,n}|\tilde{Q}_{j,n}^2\tilde{R}_{i,n}^2 \le n\|\kappa_n\|_\infty(\frac{1}{n}\sum_{1\le i\le n}\tilde{Q}_{i,n}^2)(\frac{1}{n}\sum_{1\le i\le n}\tilde{R}_{i,n}^2) = O_p(n\chi_n\varrho_n) = o_p(1).$$

Finally,

$$(\hat{\beta}_n - \beta)^2 \frac{1}{n} \sum_{1\leq i,j\leq n} |\kappa_{ij,n}| \hat{v}_{j,n}^2 \tilde{V}_{i,n}^2 = o_p(1)$$

because $\sqrt{n}(\hat{\beta}_n - \beta) = O_p(1)$ and

$$\begin{aligned}\frac{1}{n^2} \sum_{1\leq i,j\leq n} |\kappa_{ij,n}| \hat{v}_{j,n}^2 \tilde{V}_{i,n}^2 &\leq (\max_{1\leq i\leq n} \hat{v}_{i,n}^2) \frac{1}{n^2} \sum_{1\leq i,j\leq n} |\kappa_{ij,n}| \tilde{V}_{i,n}^2 \\ &\leq (\frac{\max_{1\leq i\leq n} |\hat{v}_{i,n}|}{\sqrt{n}})^2 \|\kappa_n\|_\infty (\frac{1}{n} \sum_{1\leq i\leq n} \tilde{V}_{i,n}^2) = o_p(1).\end{aligned}$$

6.5 Proof of Lemma SA-5

Because $\mathbb{E}[\tilde{U}_{j,n}^2|\mathcal{X}_n, \mathcal{W}_n] = \sum_{1\leq i\leq n} \bar{M}_{ij,n}^2 \mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n]$,

$$\begin{aligned}\mathbb{E}[\tilde{\sigma}_n^2|\mathcal{X}_n, \mathcal{W}_n] &= \frac{1}{n-K_n} \sum_{1\leq i,j\leq n} \bar{M}_{ij,n}^2 \mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n] = \frac{1}{n-K_n} \sum_{1\leq i\leq n} \bar{M}_{ii,n} \mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n] \\ &= \frac{1}{n-K_n} \sum_{1\leq i\leq n} \bar{M}_{ii,n} \mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n],\end{aligned}$$

so if $\mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n] = \sigma_n^2$, then

$$\mathbb{E}[\tilde{\sigma}_n^2|\mathcal{X}_n, \mathcal{W}_n] = \sigma_n^2 \frac{\sum_{1\leq i\leq n} \bar{M}_{ii,n}}{n-K_n} = \sigma_n^2 \omega_n$$

and

$$\Sigma_n = \frac{1}{n} \sum_{1\leq i\leq n} \hat{v}_{i,n}^2 \mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n] = \sigma_n^2 \hat{\Gamma}_n.$$

6.6 Proof of Lemma SA-6

Defining $d_{ij,n} = \sum_{1\leq k\leq n} \kappa_{ik,n} \bar{M}_{jk,n}^2 - \mathbb{1}(i=j)$, we have

$$\mathbb{E}[\tilde{\Sigma}_n(\kappa_n)|\mathcal{X}_n, \mathcal{W}_n] - \Sigma_n = \frac{1}{n} \sum_{1\leq i,j\leq n} d_{ij,n} \hat{v}_{i,n}^2 \mathbb{E}[U_{j,n}^2|\mathcal{X}_n, \mathcal{W}_n],$$

so if $\max_{1\leq i\leq n} \sum_{1\leq j\leq n} |d_{ij,n}| = o_p(1)$, then

$$\begin{aligned}|\mathbb{E}[\tilde{\Sigma}_n(\kappa_n)|\mathcal{X}_n, \mathcal{W}_n] - \Sigma_n| &\leq \frac{1}{n} \sum_{1\leq i,j\leq n} |d_{ij,n}| \hat{v}_{i,n}^2 \mathbb{E}[U_{j,n}^2|\mathcal{X}_n, \mathcal{W}_n] \leq \mathcal{C}_{U,n} \frac{1}{n} \sum_{1\leq i,j\leq n} |d_{ij,n}| \hat{v}_{i,n}^2 \\ &\leq \mathcal{C}_{U,n} (\frac{1}{n} \sum_{1\leq i\leq n} \hat{v}_{i,n}^2)(\max_{1\leq i\leq n} \sum_{1\leq j\leq n} |d_{ij,n}|) = o_p(1).\end{aligned}$$

This establishes part (a).

Next, if $\lambda_{\min}(\sum_{k=1}^n w_{k,n} w_{k,n}') > 0$ and if $\kappa_{ij,n} = \mathbb{1}\{i=j\} \Upsilon_{i,n} M_{ii,n}^{-\xi_{i,n}}$ (with $0 \leq \xi_{i,n} \leq 4$ and

$\Upsilon_{i,n} \geq 0$), then

$$\begin{aligned}\sum_{1\leq j\leq n}|d_{ij,n}| &= |\Upsilon_{i,n}M_{ii,n}^{2-\xi_{i,n}}-1|+\sum_{1\leq j\leq n,j\neq i}\Upsilon_{i,n}M_{ii,n}^{-\xi_{i,n}}M_{ji,n}^2 \\ &\leq |\Upsilon_{i,n}-1|+\Upsilon_{i,n}|M_{ii,n}^{2-\xi_{i,n}}-1|+\Upsilon_{i,n}M_{ii,n}^{1-\xi_{i,n}}(1-M_{ii,n}) \\ &\leq |\Upsilon_{i,n}-1|+\Upsilon_{i,n}(M_{ii,n}^{-2}-M_{ii,n}^2)+\Upsilon_{i,n}M_{ii,n}^{-3}(1-M_{ii,n}) \\ &= |\Upsilon_{i,n}-1|+\Upsilon_{i,n}[(1+M_{ii,n}^2)(1+M_{ii,n})+M_{ii,n}^{-3}](1-M_{ii,n}).\end{aligned}$$

Part (b) follows from this inequality and the fact that $\mathbb{P}[\lambda_{\min}(\sum_{k=1}^n w_{k,n}w_{k,n}')>0]\to 1$.

Finally, if $\mathcal{M}_n < 1/2$, then

$$|\sum_{1\leq k\leq n}\kappa_{ik,n}^{\mathtt{HC}}M_{ik,n}^2-1|+\sum_{1\leq j\leq n,j\neq i}|\sum_{1\leq k\leq n}\kappa_{ik,n}^{\mathtt{HC}}M_{jk,n}^2|=0$$

and, by Theorem 1 of Varah (1975),

$$\|\kappa_n^{\mathtt{HC}}\|_\infty \leq \frac{1}{1/2-\mathcal{M}_n}.$$

Part (c) follows from these displays and the fact that $\mathbb{P}[\mathcal{M}_n < 1/2]\to 1$.

6.7 Proof of Lemma SA-7

Because $\hat{v}_{i,n} = \tilde{V}_{i,n} + \tilde{Q}_{i,n}$, we have

$$\frac{\max_{1\leq i\leq n}|\hat{v}_{i,n}|}{\sqrt{n}} \leq \frac{\max_{1\leq i\leq n}|\tilde{V}_{i,n}|}{\sqrt{n}}+\frac{\max_{1\leq i\leq n}|\tilde{Q}_{i,n}|}{\sqrt{n}}=\frac{\max_{1\leq i\leq n}|\tilde{Q}_{i,n}|}{\sqrt{n}}+o_p(1),$$

the equality using the fact that

$$\mathbb{P}[\frac{\max_{1\leq i\leq n}|\tilde{V}_{i,n}|}{\sqrt{n}}>\varepsilon|\mathcal{W}_n]\leq \min(\sum_{1\leq i\leq n}\mathbb{P}[|\tilde{V}_{i,n}|>\varepsilon\sqrt{n}|\mathcal{W}_n],1)\leq \min(\frac{1}{\varepsilon^4 n^2}\sum_{1\leq i\leq n}\mathbb{E}[\tilde{V}_{i,n}^4|\mathcal{W}_n],1)$$

and

$$\begin{aligned}
\mathbb{E}[\tilde{V}_{i,n}^4|\mathcal{W}_n] &= \mathbb{E}[(\sum_{1\le j\le n}\bar{M}_{ij,n}V_{j,n})^4|\mathcal{W}_n] = \mathbb{E}[(\sum_{1\le j\le N_n}\sum_{s\in\mathcal{T}_{j,n}}\bar{M}_{is,n}V_{s,n})^4|\mathcal{W}_n] \\
&= \sum_{1\le j\le N_n}\mathbb{E}[(\sum_{s\in\mathcal{T}_{j,n}}\bar{M}_{is,n}V_{s,n})^4|\mathcal{W}_n] \\
&\quad +3\sum_{1\le j,k\le N_n,k\neq j}\mathbb{E}[(\sum_{s\in\mathcal{T}_{j,n}}\bar{M}_{is,n}V_{s,n})^2(\sum_{t\in\mathcal{T}_{k,n}}\bar{M}_{it,n}V_{t,n})^2|\mathcal{W}_n] \\
&\le \sum_{1\le j\le N_n}(\#\mathcal{T}_{j,n})^3\sum_{s\in\mathcal{T}_{j,n}}\bar{M}_{is,n}^4\mathbb{E}[V_{s,n}^4|\mathcal{W}_n] \\
&\quad +3\sum_{1\le j,k\le N_n,k\neq j}(\#\mathcal{T}_{j,n})(\#\mathcal{T}_{k,n})\sum_{s\in\mathcal{T}_{j,n},t\in\mathcal{T}_{k,n}}\bar{M}_{is,n}^2\bar{M}_{it,n}^2\mathbb{E}[V_{s,n}^2V_{t,n}^2|\mathcal{W}_n] \\
&\le 3\mathcal{C}_{\mathcal{T},n}^3\mathcal{C}_{V,n}\sum_{1\le j,k\le N_n}\sum_{s\in\mathcal{T}_{j,n},t\in\mathcal{T}_{k,n}}\bar{M}_{is,n}^2\bar{M}_{it,n}^2 \\
&= 3\mathcal{C}_{\mathcal{T},n}^3\mathcal{C}_{V,n}\sum_{1\le j,k\le n}\bar{M}_{ij,n}^2\bar{M}_{ik,n}^2 = 3\mathcal{C}_{\mathcal{T},n}^3\mathcal{C}_{V,n}\bar{M}_{ii,n}^2 \le 3\mathcal{C}_{\mathcal{T},n}^3\mathcal{C}_{V,n} = O_p(1).
\end{aligned}$$

It therefore suffices to show that $\max_{1\le i\le n}|\tilde{Q}_{i,n}|/\sqrt{n} = o_p(1)$.

Defining $\bar{M}_{ij,n}^{\perp} = \omega_n\mathbb{1}(i=j) - \bar{M}_{ij,n}$, we have

$$\begin{aligned}
\frac{\max_{1\le i\le n}|\tilde{Q}_{i,n}|}{\sqrt{n}} &= \frac{\omega_n}{\sqrt{n}}\max_{1\le i\le n}|Q_{i,n} - \sum_{1\le j\le n}\bar{M}_{ij,n}^{\perp}Q_{j,n}| \\
&\le \frac{\max_{1\le i\le n}|Q_{i,n}|}{\sqrt{n}} + \frac{1}{\sqrt{n}}\max_{1\le i\le n}|\sum_{1\le j\le n}\bar{M}_{ij,n}^{\perp}Q_{j,n}| \\
&= \frac{1}{\sqrt{n}}\max_{1\le i\le n}|\sum_{1\le j\le n}\bar{M}_{ij,n}^{\perp}Q_{j,n}| + o_p(1),
\end{aligned}$$

where the last equality uses the fact that $n^{-1}\sum_{1\le i\le n}\mathbb{E}[|Q_{i,n}|^{2+\theta}] = o(n^{\theta/2})$ if $\theta > 0$ or if $\chi_n = o(1)$. It therefore suffices to show that $\max_{1\le i\le n}|\sum_{1\le j\le n}\bar{M}_{ij,n}^{\perp}Q_{j,n}|/\sqrt{n} = o_p(1)$.

In cases (i) and (ii) the desired conclusion follows from $\sum_{1\le j\le n}(\bar{M}_{ij,n}^{\perp})^2 = \bar{M}_{ii,n}^{\perp} \le \mathcal{M}_n$ because, by the Cauchy-Schwarz inequality,

$$\begin{aligned}
(\frac{\max_{1\le i\le n}|\sum_{1\le j\le n}\bar{M}_{ij,n}^{\perp}Q_{j,n}|}{\sqrt{n}})^2 &\le (\max_{1\le i\le n}\bar{M}_{ii,n}^{\perp})(\frac{1}{n}\sum_{1\le i\le n}Q_{i,n}^2) \\
&\le \mathcal{M}_n(\frac{1}{n}\sum_{1\le i\le n}Q_{i,n}^2) = \mathcal{M}_nO_p(\chi_n).
\end{aligned}$$

Finally, in case (iii) the desired conclusion follows from $n^{-1}\sum_{1\le i\le n}\mathbb{E}[|Q_{i,n}|^{2+\theta}] = O(1)$ because, by the Hölder inequality,

$$\begin{aligned}\frac{1}{\sqrt{n}}\Big|\sum_{1\le j\le n}\bar{M}_{ij,n}^{\perp}Q_{j,n}\Big| &\le \sqrt[(2+\theta)/(1+\theta)]{\frac{1}{n^{\theta/(2\theta+2)}}\sum_{1\le j\le n}|\bar{M}_{ij,n}^{\perp}|^{(2+\theta)/(1+\theta)}}\ \sqrt[2+\theta]{\frac{1}{n}\sum_{1\le j\le n}|Q_{j,n}|^{2+\theta}} \\ &\le \sqrt[(2+\theta)/(1+\theta))]{\frac{1}{n^{\theta/(2\theta+2)}}[1+\sum_{1\le j\le n}\mathbb{1}(M_{ij,n}\neq 0)]}O_p(1).\end{aligned}$$

7 Simulations

We conducted a simulation study to assess the finite sample properties of our proposed inference methods as well as those of other standard inference methods available in the literature. Based on the generic linear regression model (SA-1), we consider data generating processes (DGPs) motivated by the three examples discussed above. To be more precise, our simulation study includes (i) a linear regression model with increasing dimension, (ii) one-way and two-way fixed effects panel data regression models, and (iii) a semiparametric partially linear model.

For each model, we employed three distinct distributions for the unobservable error terms: Gaussian, Asymmetric, and Bimodal. The Lebesgue densities corresponding to each of these distributions are plotted in Figure 1. For each case, we investigate both homoskedastic as well as (conditional on $\mathbf{x}_{i,n}$ and/or $\mathbf{w}_{i,n}$) heteroskedastic errors entering the corresponding linear model. Finally, for each model, we also varied the distribution generating the possibly high-dimensional covariates $\mathbf{w}_{i,n}$, covering both discrete covariates and uniformly distributed covariates. Putting all together, the Monte Carlo experiment includes a total of 15 distinct DGPs. A synopsis of each DGP is given in Table 1, while the specification details and the results for each example are discussed in the following subsections.

For each DGP, we conducted $S = 5,000$ simulations with $n = 700$, and $\dim(\mathbf{x}_i) = 1$ with $\beta = 1$. In each replication, we constructed eight Gaussian-Based and eight bootstrap-based confidence intervals. Our paper presents theory for Gaussian-based inference methods, but we also included bootstrap-based inference methods for completeness. In fact, as discussed in the main paper, it is known that the bootstrap is invalid when $K \propto n$ in linear regression models. For each inference method we report both empirical coverage error of 95% nominal confidence intervals and their average length. The latter provides a summary of efficiency/power for each inference method.

The eight Gaussian-based confidence intervals take the form:

$$\mathsf{I}_\ell = \left[\ \hat{\beta}_n - \Phi^{-1}(1-\alpha/2)\cdot\sqrt{\frac{\hat{\Omega}_{n,\ell}}{n}}\ ,\ \hat{\beta}_n - \Phi^{-1}(\alpha/2)\cdot\sqrt{\frac{\hat{\Omega}_{n,\ell}}{n}}\ \right], \qquad \hat{\Omega}_{n,\ell} = \hat{\Gamma}_n^{-1}\hat{\Sigma}_{n,\ell}\hat{\Gamma}_n^{-1},$$

where Φ^{-1} denotes the inverse of the standard normal cdf, and $\hat{\Sigma}_{n,\ell}$ with $\ell \in \{$HO0,HO1,HC0,HC1,HC2,HC3,HC4,H corresponds to each of the variance estimators discussed in the main paper. To be precise, $\ell =$HO0 and $\ell =$HO1 give the standard homoskedasticity-consistent standard errors without and with degrees of freedom correction, respectively, while $\ell =$HC0 through $\ell =$HC4 corresponds to the HCK

class of heteroskedasticity-consistent standard errors discussed in, e.g., MacKinnon (2012). Last but not least, $\ell =$HCK uses $\hat{\Sigma}_{n,\ell} = \hat{\Sigma}_n^{\texttt{HC}}$, our proposed new inference method that is valid under conditional heteroskedasticity and many covariates. Finally, the length of the confidence intervals is:

$$\mathsf{L}_\ell = [\Phi^{-1}(1-\alpha/2) - \Phi^{-1}(\alpha/2)] \cdot \sqrt{\frac{\hat{\Omega}_{n,\ell}}{n}} = 2 \cdot \Phi^{-1}(1-\alpha/2) \cdot \sqrt{\frac{\hat{\Omega}_{n,\ell}}{n}},$$

due to the symmetry of the Gaussian distribution.

To construct the eight bootstrap-based confidence intervals we employ the standard nonparametric bootstrap applied to each possible t-statistic:

$$\mathsf{T}_\ell = \frac{\hat{\beta}_n - \beta}{\sqrt{\hat{\Omega}_{n,\ell}/n}},$$

indexed by the choice of standard error estimator, $\ell \in \{$HO0,HO1,HC0,HC1,HC2,HC3,HC4,HC$K\}$. This bootstrap method is usually called the t-percentile bootstrap and the corresponding bootstrap-based confidence intervals take the form:

$$\mathsf{I}_\ell^* = \left[\ \hat{\beta}_n - Q_\ell^{-1}(1-\alpha/2) \cdot \sqrt{\hat{\Omega}_{n,\ell}} \ , \ \hat{\beta}_n - Q_\ell^{-1}(\alpha/2) \cdot \sqrt{\hat{\Omega}_{n,\ell}} \ \right],$$

where $Q_\ell^{-1}(a)$ denotes the a-th quantile of the (approximation to the) bootstrap distribution of T_ℓ. Finally, the length of the confidence intervals is:

$$\mathsf{L}_\ell = [Q_\ell^{-1}(1-\alpha/2) - Q_\ell^{-1}(\alpha/2)] \cdot \sqrt{\frac{\hat{\Omega}_{n,\ell}}{n}},$$

where $Q_\ell^{-1}(1-\alpha/2) \neq -Q_\ell^{-1}(\alpha/2)$ due to the possibly asymmetry of the bootstrap distribution. For the bootstrap-based inference procedures, we employ $B = 500$ replications.

For each of the 15 DGPs considered, we report one table including two panels. The first panel presents empirical coverage of each of the 16 confidence intervals, while the second panel reports the average length for each of the 16 confidence intervals. In the remaining of this section we discuss the details underlying each of DGP and present the numerical results.

7.1 Linear Regression Model with Increasing Dimension

To facilitate comparability, we employed a DGP that is as similar as possible to those employed in the literature before. In particular, we considered the following model (we drop the subindex n for notation simplicity):

$$\begin{aligned} y_i &= \beta x_i + \boldsymbol{\gamma}_n' \mathbf{w}_i + u_i, & u_i|(x_i, \mathbf{w}_i) &\sim \text{i.i.d. } (0, \sigma_{ui}^2), & \sigma_{ui}^2 &= \varkappa_u (1 + (t(x_i) + \boldsymbol{\iota}' \mathbf{w}_i)^2)^\vartheta, \\ x_i &= v_i, & v_i|\mathbf{w}_i &\sim \text{i.i.d. } (0, \sigma_{vi}^2), & \sigma_{vi}^2 &= \varkappa_v (1 + (\boldsymbol{\iota}' \mathbf{w}_i)^2)^\vartheta, \end{aligned}$$

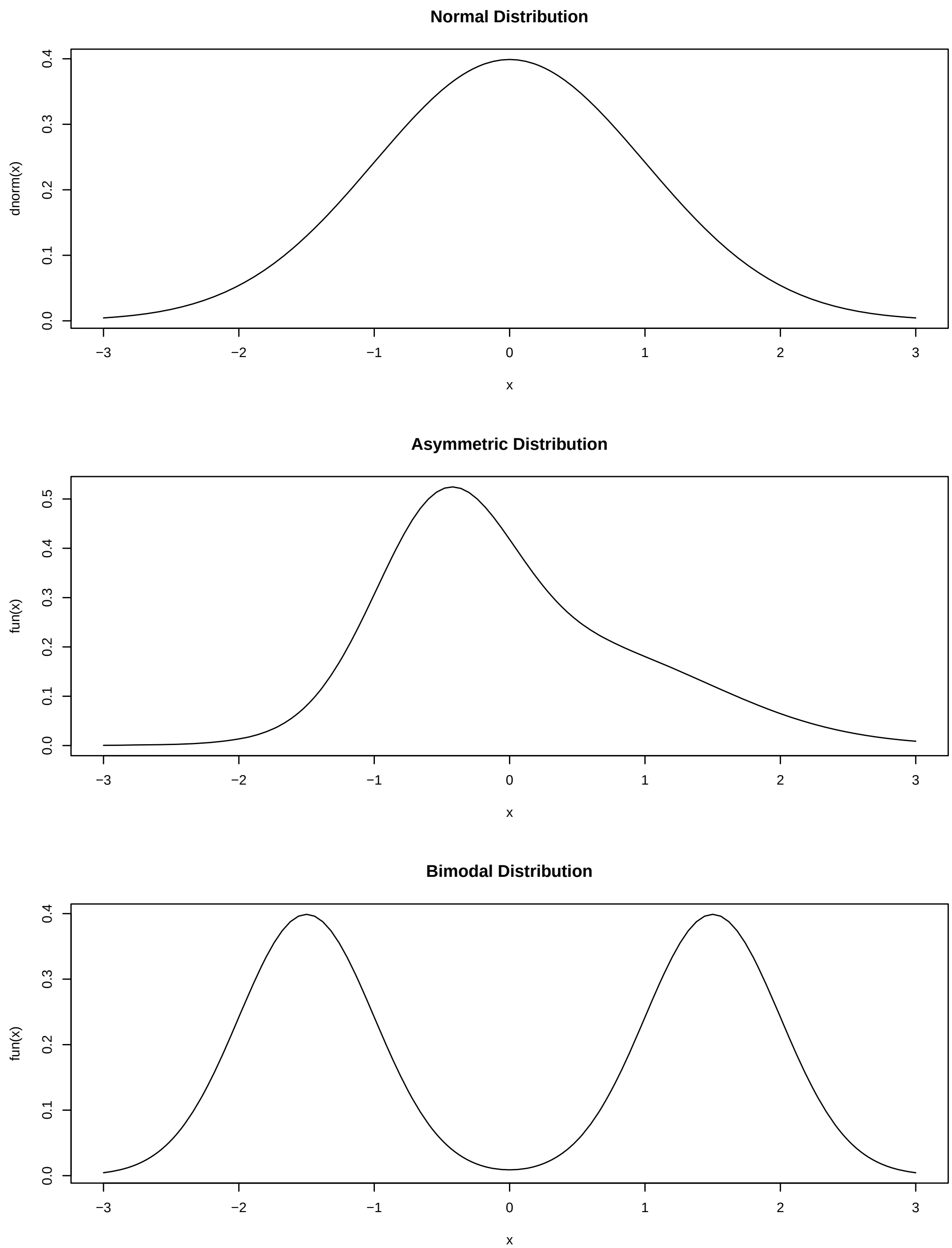

Figure 1: Lebesgue Densities of Error Terms Distributions.

Table 1: Synopsis of Data Generating Processes.

DGP	Setting	Distributions		
		$\mathbf{w}_{i,n}$	$\mathbf{V}_{i,n}$	$U_{i,n}$
Model 1	Linear Regression	Discrete	Gaussian	Gaussian
Model 2	Linear Regression	Discrete	Asymmetric	Asymmetric
Model 3	Linear Regression	Discrete	Bimodal	Bimodal
Model 4	Linear Regression	Uniform	Gaussian	Gaussian
Model 5	Linear Regression	Uniform	Asymmetric	Asymmetric
Model 6	Linear Regression	Uniform	Bimodal	Bimodal
Model 7	One-way Fixed Effects	Dummy Variables	Gaussian	Gaussian
Model 8	One-way Fixed Effects	Dummy Variables	Asymmetric	Asymmetric
Model 9	One-way Fixed Effects	Dummy Variables	Bimodal	Bimodal
Model 10	Two-way Fixed Effects	Dummy Variables	Gaussian	Gaussian
Model 11	Two-way Fixed Effects	Dummy Variables	Asymmetric	Asymmetric
Model 12	Two-way Fixed Effects	Dummy Variables	Bimodal	Bimodal
Model 13	Partially Linear	Power Series	Gaussian	Gaussian
Model 14	Partially Linear	Power Series	Asymmetric	Asymmetric
Model 15	Partially Linear	Power Series	Bimodal	Bimodal

Note: See upcoming subsections for precise definitions and other parameter values.

where $\boldsymbol{\iota} = (1, 1, \cdots, 1)'$, $\beta = 1$ and $\boldsymbol{\gamma} = \mathbf{0}$, the constants $\varkappa_u$ and $\varkappa_v$ are chosen so that $\mathbb{V}[u_i] = \mathbb{V}[v_i] = 1$, and $t(a) = a\mathbb{1}(-2 \leq a \leq 2) + 2\,\text{sgn}(a)(1 - \mathbb{1}(-2 \leq a \leq 2))$. In the absence of the additional covariates $\mathbf{w}_{i,n}$, and when the distribution of the error terms is Gaussian, the design essentially coincides with the one in Stock and Watson (2008), and is very similar to the one considered in MacKinnon (2012). More generally, we consider three different distributions for the error terms and also include possibly many additional covariates $\mathbf{w}_i$. We impose random sampling across $i = 1, 2, \cdots, n$.

We consider five dimensions for $\mathbf{w}_i : K \in \{1, 71, 141, 211, 281\}$, where in all cases the first covariate is an intercept. Given the above, the two main parameters varying in the Monte Carlo experiments are: the constant ϑ and the distribution of the covariates $\mathbf{w}_{i,n}$. The first parameter controls the degree of heteroskedasticity: $\vartheta = 0$ corresponds to homoskedasticity, and $\vartheta = 1$ corresponds to moderate heteroskedasticity, as classified by MacKinnon (2012). For the distribution of the covariates $\mathbf{w}_i$ we consider two cases: (i) independent, sparse and discrete covariates constructed as $\mathbb{1}(\mathcal{N}(0, 1) \geq 2.5)$, giving Models 1–3 in Table 1, and (ii) independent uniformly distributed on $(-1, 1)$, giving Models 4–6 in Table 1. For each case, the unobserved errors (u_i, v_i) are taken to be independent and with the same distribution (each model corresponds to one distribution depicted in Figure 1). See Table 1 for details on the labelling.

The results for this example are given in Tables 2–7. These tables report empirical coverage rates for the 16 nominal 95% confidence intervals for β, across the range of K, for both the Homoskedastic Model and the Heteroskedastic model. The main findings from the small simulation study are in line with our theoretical results. We find that the confidence interval estimators constructed our proposed standard errors formula $\hat{\Sigma}_n^{\mathtt{HC}}$, denoted HC$K$, offer close-to-correct empirical coverage in all cases. The alternative heteroskedasticity consistent standard errors currently available in the literature lead to confidence intervals that could deliver substantial under or over coverage depending on the design and degree of heteroskedasticity considered. We also found that inference based on HC3 standard errors is conservative. The bootstrap-based confidence intervals seem to perform better but they still do not deliver close-to-correct empirical coverage in all cases. Finally, our proposed inference method exhibits good average interval length properties, when compared to the other procedures (many of which are in fact far from their nominal coverage target).

7.2 Fixed Effects Panel Data Regression Model

We also consider the following fixed effects panel data model:

$$y_{it} = \alpha_i + e_{g_{it}} + \beta x_{it} + u_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T,$$

where α_i are unobserved unit-specific time invariant factors, and $e_{g_{it}}$ are unobserved factors common to all observations sharing the same index $g_{it} \in \{1, 2, \cdots, G\}$. When $e_{g_{it}} = 0$, this model reduces to the one-way fixed effects model studied in Stock and Watson (2008), while otherwise this model coincides with the one studied in Verdier (2017).

In this simulation example, we also set $d = 1$ and $\beta = 1$, and consider i.i.d. sampling across i and t, which coincides with the first simulation setup in Stock and Watson (2008). To generate heteroskedasticity, we let $\mathbf{z}_{it} = (z_{1it}, \cdots, z_{d_z it})'$ with $z_{\ell it} \sim$ i.i.d. $\mathsf{Uniform}(-1, 1)$, $\ell = 1, \cdots, d_z$, and employ:

$$
\begin{aligned}
u_{it}|(x_{it}, \mathbf{z}_{it}) &\sim \text{i.i.d. } (0, \sigma^2_{\varepsilon i}), \qquad & \sigma^2_{\varepsilon i} &= \varkappa_\varepsilon(1 + (t(x_{it}) + \boldsymbol{\iota}'\mathbf{z}_{it})^2)^\vartheta, \\
x_{it}|\mathbf{z}_{it} &\sim \text{i.i.d. } (0, \sigma^2_{vi}), \qquad & \sigma^2_{vi} &= \varkappa_v(1 + (\boldsymbol{\iota}'\mathbf{z}_{it})^2)^\vartheta,
\end{aligned}
$$

where once again the constants $\varkappa_u$ and $\varkappa_v$ are chosen so that $\mathbb{V}[u_i] = \mathbb{V}[v_i] = 1$, and the Homoskedastic Model sets $\vartheta = 0$ while the Heteroskedastic Model sets $\vartheta = 1$. The distributions of u_{it} and x_{it} are again chosen to be each of the ones depicted in Figure 1, leading to fixed effects Models 10–15 in Table 1. Specifically, when it is assumed that $e_{g_{it}} = 0$ for estimation we obtain the one-way fixed Models 10–12 in Table 1, while when it is assumed that $e_{g_{it}} \neq 0$ for estimation we obtain the two-way fixed Models 13–15 in Table 1. In the former case, the model includes $\dim(\mathbf{w}_i) = N$ fixed effects, while in the latter case we assume $G_N = \lfloor N/3 \rfloor$ groups in the population, leading to $\dim(\mathbf{w}_i) = N + G_N$ indicator variables. For all models, we set $\alpha_i = e_{g_{it}} = 0$ when generating the data (i.e., $\boldsymbol{\gamma}_n = \mathbf{0}$ in the DGP), and consider $T \in \{700, 10, 5, 4, 3\}$ and $N = \lfloor 700/T \rfloor$ so that the total sample size is always roughly $n = 700$.

Two key differences between these DGPs and the ones considered in the context of linear regression models with increasing dimension are: (i) the heteroskedasticity does not change with $\dim(\mathbf{w}_i)$, and (ii) the design matrix induced by the fixed effects is truly sparse. Note that Models 1–3 are similar to those considered here, as both include only dummy variables, but a key distinctive feature of multi-way fixed effect models is the lack of potential overlap across groups and the lack of randomness in the construction of $\mathbf{w}_i$.

The results for the one-way fixed effects model are given in Tables 8–10, while the results for the two-way fixed effects model are given in Tables 11–13. In all cases, the numerical findings are consistent with our theoretical results, and in line with the results obtained from the other simulation models. One important feature of these DGPs is that the failure of the bootstrap is more extreme. This is due, in part, to the one-way and two-way fixed effect structure that induces a highly sparse covariates vector $\mathbf{w}_i$, leading to singularity and other numerical issues when employing the a bootstrap distributional approximation. In contrast, our analytic distribution theory performs remarkably well.

7.3 Semiparametric Partially Linear Model

Finally, to complement our simulation study, we also conducted a Monte Carlo experiment using a semiparametric semilinear model. In this case the model takes the form:

$$
\begin{aligned}
y_i &= \beta x_i + g(\mathbf{z}_i) + \varepsilon_i, \qquad & \varepsilon_i|(x_i, \mathbf{z}_i) &\sim \text{i.i.d. } (0, \sigma^2_{\varepsilon i}), \qquad & \sigma^2_{\varepsilon i} &= \varkappa_\varepsilon(1 + (t(x_i) + \boldsymbol{\iota}'\mathbf{z}_i)^2)^\vartheta, \\
x_i &= h(\mathbf{z}_i) + v_i, \qquad & v_i|\mathbf{z}_i &\sim \text{i.i.d. } (0, \sigma^2_{vi}), \qquad & \sigma^2_{vi} &= \varkappa_v(1 + (\boldsymbol{\iota}'\mathbf{z}_i)^2)^\vartheta,
\end{aligned}
$$

where $d = 1$, $\beta = 1$, $\dim(\mathbf{z}_i) = 6$, $\mathbf{z}_i = (z_{1i}, \cdots, z_{6i})'$ with $z_{\ell i} \sim$ i.i.d. $\mathsf{Uniform}(-1, 1)$, $\ell = 1, \cdots, 6$, giving Models 7–9 in Table 1. The unknown regression functions are set to $g(\mathbf{z}_i) = \exp(-\|\mathbf{z}_i\|^{1/2})$ and $h(\mathbf{z}_i) = \exp(\|\mathbf{z}_i\|^{1/2})$, which are non-linear and non-additive-separable in the covariates $\mathbf{z}_i$. Note that here $\boldsymbol{\gamma}_n \neq \mathbf{0}$ by construction. The constants $\varkappa_u$ and $\varkappa_v$ are again chosen so that $\mathbb{V}[u_i] = \mathbb{V}[v_i] = 1$, and we continue to consider two models: Homoskedastic ($\vartheta = 0$) and Heteroskedastic ($\vartheta = 1$). We impose random sampling across $i = 1, 2, \cdots, n$, and for each simulation model we set the marginal distribution of the unobserved independent errors (ε_i, v_i) according to each of the three distributions depicted in Figure 1. In the absence of heteroskedasticity, the simulation models coincide with the ones considered in Cattaneo, Jansson, and Newey (2018).

To construct the covariates $\mathbf{w}_{i,n}$ entering linear regression model (SA-1), we employ power series expansions. The following table gives a summary of the expansions considered, where $\mathbf{w}_{i,n} = \mathbf{p}_n(\mathbf{z}_i) = \mathbf{p}(\mathbf{z}_i; K)$ for $K \in \{7, 13, 28, 34, 84, 90, 210, 216\}$ are defined as follows.

Polynomial Basis Expansion: $\dim(\mathbf{z}_i) = 6$ and $n = 700$

K	$\mathbf{p}(\mathbf{z}_i; K)$	K/n
1	1	0.001
7	$(1, z_{1i}, z_{2i}, z_{3i}, z_{4i}, z_{5i}, z_{6i})'$	0.010
13	$(\mathbf{p}(\mathbf{z}_i; 7)', z_{1i}^2, z_{2i}^2, z_{3i}^2, z_{4i}^2, z_{5i}^2, z_{6i}^2)'$	0.019
28	$\mathbf{p}(\mathbf{z}_i; 13)$ + first-order interactions	0.040
34	$(\mathbf{p}(\mathbf{z}_i; 28)', z_{1i}^3, z_{2i}^3, z_{3i}^3, z_{4i}^3, z_{5i}^3, z_{6i}^3)'$	0.049
84	$\mathbf{p}(\mathbf{z}_i; 34)$ + second-order interactions	0.120
90	$(\mathbf{p}(\mathbf{z}_i; 84)', z_{1i}^4, z_{2i}^4, z_{3i}^4, z_{4i}^4, z_{5i}^4, z_{6i}^4)'$	0.129
210	$\mathbf{p}(\mathbf{z}_i; 90)$ + third-order interactions	0.300
216	$(\mathbf{p}(\mathbf{z}_i; 210)', z_{1i}^5, z_{2i}^5, z_{3i}^5, z_{4i}^5, z_{5i}^5, z_{6i}^5)'$	0.309

The results for this example are given in Tables 14–16, which report only $K \in \{1, 13, 34, 90, 216\}$ to conserve same space. The numerical findings are in perfect agreement with those reported previously for the linear model with increasing dimension, with the only exception that now the model clearly exhibits misspecification error (i.e., for $K = 1$ all methods are affected by misspecification bias). The key distinctive features of this setting, relative to the linear regression model with increasing dimension considered previously, are: (i) $\boldsymbol{\gamma}_n \neq \mathbf{0}$ by construction, (ii) heteroskedasticity does not change as the dimension of the model changes (i.e., it only depends on x_i and $\mathbf{z}_i$, but not $\mathbf{w}_{i,n}$), and (iii) the covariates $\mathbf{w}_{i,n}$ are correlated and dependent through the polynomial expansion.

Table 2: Simulation Results, Model 1, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.949	0.950	0.948	0.948	0.948	0.948	0.948	0.948	0.946	0.946	0.943	0.943	0.943	0.943	0.943	0.943
$K/n = 0.101$	0.939	0.956	0.939	0.952	0.952	0.962	0.980	0.951	0.951	0.951	0.947	0.947	0.948	0.949	0.947	0.942
$K/n = 0.201$	0.916	0.947	0.919	0.947	0.946	0.968	0.989	0.945	0.965	0.965	0.950	0.950	0.949	0.946	0.944	0.939
$K/n = 0.301$	0.900	0.950	0.904	0.954	0.951	0.977	0.983	0.949	0.980	0.980	0.961	0.961	0.949	0.931	0.948	0.933
$K/n = 0.401$	0.881	0.954	0.884	0.955	0.952	0.989	0.972	0.949	0.989	0.989	0.976	0.976	0.956	0.928	0.967	0.944
Heteroskedastic Model																
$K/n = 0.001$	0.880	0.880	0.945	0.945	0.945	0.945	0.946	0.945	0.939	0.939	0.937	0.937	0.937	0.937	0.937	0.937
$K/n = 0.101$	0.725	0.750	0.885	0.904	0.926	0.957	0.989	0.948	0.897	0.897	0.916	0.906	0.907	0.909	0.902	0.919
$K/n = 0.201$	0.762	0.804	0.853	0.901	0.924	0.973	0.995	0.945	0.919	0.919	0.919	0.909	0.908	0.907	0.908	0.920
$K/n = 0.301$	0.784	0.856	0.837	0.903	0.926	0.981	0.977	0.947	0.944	0.944	0.936	0.926	0.919	0.903	0.920	0.920
$K/n = 0.401$	0.758	0.875	0.792	0.908	0.929	0.990	0.950	0.948	0.975	0.975	0.962	0.962	0.936	0.900	0.953	0.926

(b) Interval Length

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149	0.149
$K/n = 0.101$	0.148	0.156	0.148	0.157	0.156	0.165	0.186	0.156	0.161	0.161	0.158	0.158	0.158	0.158	0.158	0.157
$K/n = 0.201$	0.148	0.166	0.149	0.167	0.165	0.185	0.225	0.165	0.180	0.180	0.170	0.170	0.169	0.167	0.166	0.164
$K/n = 0.301$	0.148	0.177	0.150	0.179	0.177	0.212	0.219	0.177	0.210	0.210	0.189	0.189	0.182	0.172	0.180	0.174
$K/n = 0.401$	0.148	0.192	0.150	0.194	0.191	0.247	0.213	0.190	0.260	0.260	0.223	0.223	0.200	0.174	0.212	0.189
Heteroskedastic Model																
$K/n = 0.001$	0.148	0.148	0.186	0.186	0.186	0.186	0.187	0.186	0.186	0.186	0.188	0.188	0.188	0.188	0.188	0.188
$K/n = 0.101$	0.148	0.156	0.213	0.225	0.241	0.273	0.357	0.254	0.243	0.243	0.264	0.264	0.266	0.268	0.273	0.269
$K/n = 0.201$	0.148	0.166	0.187	0.209	0.226	0.276	0.353	0.244	0.243	0.243	0.252	0.252	0.251	0.248	0.251	0.249
$K/n = 0.301$	0.148	0.177	0.170	0.203	0.219	0.287	0.278	0.240	0.259	0.259	0.254	0.254	0.244	0.232	0.247	0.239
$K/n = 0.401$	0.148	0.191	0.159	0.206	0.220	0.310	0.239	0.241	0.300	0.300	0.276	0.276	0.248	0.218	0.269	0.243

Table 3: Simulation Results, Model 2, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.944	0.944	0.942	0.942	0.942	0.942	0.942	0.942	0.937	0.937	0.939	0.939	0.939	0.939	0.939	0.939
$K/n = 0.101$	0.930	0.945	0.930	0.946	0.945	0.957	0.977	0.945	0.947	0.947	0.941	0.941	0.941	0.941	0.940	0.942
$K/n = 0.201$	0.916	0.948	0.918	0.951	0.948	0.967	0.989	0.946	0.959	0.959	0.949	0.949	0.948	0.944	0.943	0.940
$K/n = 0.301$	0.900	0.948	0.898	0.948	0.945	0.979	0.981	0.945	0.969	0.969	0.955	0.955	0.948	0.936	0.946	0.942
$K/n = 0.401$	0.868	0.950	0.872	0.950	0.947	0.986	0.969	0.944	0.988	0.988	0.968	0.968	0.947	0.910	0.958	0.932
Heteroskedastic Model																
$K/n = 0.001$	0.818	0.818	0.940	0.941	0.941	0.941	0.941	0.941	0.940	0.940	0.939	0.939	0.939	0.939	0.939	0.939
$K/n = 0.101$	0.681	0.712	0.882	0.899	0.924	0.954	0.986	0.945	0.896	0.896	0.910	0.910	0.913	0.915	0.916	0.927
$K/n = 0.201$	0.738	0.791	0.858	0.902	0.924	0.969	0.993	0.947	0.915	0.915	0.920	0.920	0.920	0.914	0.915	0.924
$K/n = 0.301$	0.747	0.822	0.818	0.892	0.919	0.974	0.971	0.944	0.938	0.938	0.925	0.925	0.914	0.900	0.918	0.921
$K/n = 0.401$	0.743	0.859	0.787	0.890	0.912	0.987	0.934	0.943	0.969	0.969	0.951	0.951	0.920	0.878	0.935	0.911

(b) Interval Length

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149	0.149
$K/n = 0.101$	0.148	0.156	0.148	0.157	0.156	0.165	0.185	0.156	0.162	0.162	0.159	0.159	0.159	0.159	0.159	0.157
$K/n = 0.201$	0.148	0.166	0.149	0.167	0.165	0.185	0.225	0.165	0.180	0.180	0.170	0.170	0.168	0.166	0.165	0.163
$K/n = 0.301$	0.148	0.177	0.149	0.179	0.177	0.211	0.218	0.176	0.209	0.209	0.189	0.189	0.181	0.172	0.180	0.174
$K/n = 0.401$	0.148	0.192	0.150	0.194	0.191	0.246	0.213	0.190	0.260	0.260	0.223	0.223	0.200	0.174	0.212	0.189
Heteroskedastic Model																
$K/n = 0.001$	0.148	0.148	0.214	0.214	0.214	0.214	0.214	0.214	0.215	0.215	0.218	0.218	0.218	0.218	0.218	0.218
$K/n = 0.101$	0.149	0.157	0.231	0.244	0.261	0.296	0.386	0.279	0.263	0.263	0.292	0.292	0.295	0.298	0.304	0.299
$K/n = 0.201$	0.148	0.166	0.196	0.220	0.238	0.290	0.371	0.260	0.253	0.253	0.269	0.269	0.267	0.265	0.267	0.267
$K/n = 0.301$	0.149	0.178	0.176	0.211	0.228	0.298	0.289	0.252	0.267	0.267	0.267	0.267	0.257	0.245	0.260	0.254
$K/n = 0.401$	0.148	0.191	0.162	0.210	0.225	0.316	0.244	0.249	0.303	0.303	0.283	0.283	0.254	0.223	0.275	0.250

Table 4: Simulation Results, Model 3, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.954	0.954	0.954	0.955	0.955	0.955	0.955	0.955	0.951	0.951	0.950	0.950	0.950	0.950	0.950	0.950
$K/n = 0.101$	0.942	0.954	0.942	0.954	0.953	0.964	0.980	0.952	0.952	0.952	0.948	0.948	0.949	0.948	0.949	0.946
$K/n = 0.201$	0.923	0.954	0.924	0.954	0.953	0.972	0.992	0.952	0.961	0.961	0.950	0.950	0.946	0.942	0.939	0.937
$K/n = 0.301$	0.906	0.955	0.910	0.957	0.954	0.981	0.985	0.953	0.973	0.973	0.962	0.962	0.954	0.946	0.952	0.948
$K/n = 0.401$	0.872	0.951	0.876	0.952	0.950	0.988	0.970	0.948	0.986	0.986	0.975	0.975	0.950	0.920	0.965	0.939
Heteroskedastic Model																
$K/n = 0.001$	0.876	0.876	0.949	0.949	0.949	0.949	0.949	0.949	0.946	0.946	0.949	0.949	0.949	0.949	0.949	0.949
$K/n = 0.101$	0.713	0.739	0.892	0.908	0.926	0.957	0.991	0.945	0.900	0.900	0.916	0.916	0.917	0.918	0.933	0.931
$K/n = 0.201$	0.739	0.794	0.855	0.896	0.921	0.965	0.992	0.946	0.926	0.926	0.916	0.916	0.914	0.919	0.933	0.927
$K/n = 0.301$	0.750	0.829	0.821	0.893	0.918	0.977	0.973	0.946	0.943	0.943	0.928	0.928	0.914	0.898	0.927	0.920
$K/n = 0.401$	0.757	0.867	0.793	0.896	0.916	0.984	0.939	0.940	0.969	0.969	0.953	0.943	0.924	0.895	0.950	0.916

(b) Interval Length

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148
$K/n = 0.101$	0.148	0.156	0.149	0.157	0.156	0.165	0.186	0.156	0.161	0.161	0.157	0.157	0.158	0.158	0.158	0.156
$K/n = 0.201$	0.148	0.166	0.149	0.167	0.165	0.185	0.225	0.165	0.180	0.180	0.170	0.170	0.168	0.166	0.165	0.163
$K/n = 0.301$	0.148	0.177	0.150	0.179	0.177	0.211	0.219	0.177	0.209	0.209	0.188	0.188	0.181	0.171	0.179	0.173
$K/n = 0.401$	0.148	0.192	0.150	0.194	0.191	0.247	0.213	0.191	0.260	0.260	0.223	0.223	0.199	0.174	0.211	0.189
Heteroskedastic Model																
$K/n = 0.001$	0.148	0.148	0.187	0.187	0.187	0.187	0.187	0.187	0.188	0.188	0.189	0.189	0.189	0.189	0.189	0.189
$K/n = 0.101$	0.148	0.156	0.222	0.234	0.251	0.286	0.374	0.268	0.252	0.252	0.271	0.271	0.273	0.275	0.279	0.276
$K/n = 0.201$	0.148	0.166	0.192	0.215	0.233	0.285	0.364	0.254	0.250	0.250	0.259	0.259	0.258	0.255	0.257	0.255
$K/n = 0.301$	0.148	0.177	0.173	0.207	0.224	0.293	0.285	0.247	0.263	0.263	0.260	0.260	0.250	0.237	0.253	0.243
$K/n = 0.401$	0.148	0.191	0.161	0.208	0.223	0.314	0.243	0.246	0.303	0.303	0.281	0.281	0.252	0.221	0.273	0.246

Table 5: Simulation Results, Model 4, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.945	0.945	0.945	0.946	0.945	0.946	0.946	0.945	0.943	0.943	0.946	0.946	0.946	0.946	0.946	0.946
$K/n = 0.101$	0.939	0.953	0.938	0.951	0.951	0.964	0.978	0.951	0.956	0.956	0.954	0.954	0.953	0.953	0.952	0.953
$K/n = 0.201$	0.916	0.945	0.914	0.944	0.944	0.968	0.994	0.944	0.951	0.951	0.940	0.940	0.936	0.933	0.930	0.934
$K/n = 0.301$	0.904	0.955	0.903	0.952	0.952	0.979	0.985	0.950	0.968	0.968	0.951	0.951	0.942	0.930	0.940	0.936
$K/n = 0.401$	0.864	0.948	0.860	0.947	0.946	0.988	0.970	0.945	0.986	0.986	0.969	0.969	0.943	0.908	0.960	0.936
Heteroskedastic Model																
$K/n = 0.001$	0.874	0.875	0.946	0.946	0.946	0.946	0.947	0.946	0.944	0.944	0.944	0.944	0.944	0.944	0.944	0.944
$K/n = 0.101$	0.607	0.630	0.905	0.923	0.923	0.946	0.962	0.946	0.908	0.908	0.910	0.910	0.911	0.931	0.922	0.933
$K/n = 0.201$	0.582	0.636	0.872	0.910	0.910	0.942	0.980	0.942	0.896	0.896	0.921	0.921	0.930	0.935	0.933	0.941
$K/n = 0.301$	0.582	0.667	0.829	0.900	0.900	0.950	0.962	0.944	0.903	0.903	0.932	0.932	0.928	0.915	0.927	0.936
$K/n = 0.401$	0.568	0.684	0.770	0.875	0.875	0.958	0.920	0.947	0.921	0.921	0.951	0.951	0.921	0.884	0.941	0.931

(b) Interval Length

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148
$K/n = 0.101$	0.148	0.156	0.148	0.156	0.156	0.164	0.183	0.156	0.158	0.158	0.156	0.156	0.156	0.156	0.155	0.156
$K/n = 0.201$	0.148	0.166	0.148	0.166	0.166	0.185	0.231	0.166	0.176	0.176	0.168	0.168	0.166	0.163	0.161	0.164
$K/n = 0.301$	0.148	0.177	0.148	0.177	0.176	0.211	0.223	0.176	0.205	0.205	0.186	0.186	0.178	0.169	0.176	0.173
$K/n = 0.401$	0.148	0.192	0.148	0.191	0.191	0.246	0.216	0.190	0.256	0.256	0.222	0.222	0.196	0.172	0.211	0.188
Heteroskedastic Model																
$K/n = 0.001$	0.148	0.148	0.186	0.186	0.186	0.186	0.187	0.186	0.187	0.187	0.188	0.188	0.188	0.188	0.188	0.188
$K/n = 0.101$	0.148	0.156	0.301	0.318	0.318	0.336	0.375	0.332	0.315	0.315	0.369	0.369	0.369	0.368	0.367	0.371
$K/n = 0.201$	0.148	0.166	0.279	0.313	0.313	0.351	0.437	0.340	0.314	0.314	0.379	0.379	0.375	0.371	0.368	0.382
$K/n = 0.301$	0.148	0.177	0.253	0.303	0.303	0.363	0.384	0.343	0.321	0.321	0.386	0.386	0.370	0.354	0.368	0.380
$K/n = 0.401$	0.148	0.192	0.231	0.299	0.299	0.387	0.339	0.352	0.355	0.355	0.420	0.420	0.373	0.329	0.401	0.388

Table 6: Simulation Results, Model 5, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								**Bootstrap Distributional Approximation**							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.954	0.954	0.953	0.953	0.953	0.953	0.954	0.953	0.947	0.947	0.949	0.949	0.949	0.949	0.949	0.949
$K/n = 0.101$	0.933	0.948	0.930	0.943	0.943	0.958	0.976	0.942	0.943	0.943	0.939	0.939	0.939	0.939	0.938	0.939
$K/n = 0.201$	0.922	0.949	0.920	0.948	0.947	0.971	0.993	0.946	0.959	0.959	0.950	0.950	0.945	0.942	0.940	0.943
$K/n = 0.301$	0.898	0.950	0.896	0.947	0.947	0.981	0.988	0.944	0.975	0.975	0.958	0.958	0.950	0.937	0.949	0.944
$K/n = 0.401$	0.864	0.950	0.866	0.950	0.950	0.988	0.973	0.947	0.986	0.986	0.975	0.975	0.951	0.914	0.968	0.940
Heteroskedastic Model																
$K/n = 0.001$	0.820	0.820	0.946	0.947	0.947	0.947	0.948	0.947	0.940	0.940	0.941	0.941	0.941	0.941	0.941	0.941
$K/n = 0.101$	0.592	0.615	0.904	0.919	0.919	0.944	0.957	0.944	0.904	0.904	0.916	0.916	0.916	0.916	0.936	0.938
$K/n = 0.201$	0.590	0.641	0.874	0.914	0.914	0.945	0.984	0.945	0.900	0.900	0.928	0.928	0.926	0.924	0.932	0.942
$K/n = 0.301$	0.582	0.665	0.828	0.901	0.901	0.950	0.962	0.944	0.903	0.903	0.938	0.938	0.928	0.922	0.927	0.934
$K/n = 0.401$	0.560	0.687	0.776	0.883	0.883	0.955	0.921	0.948	0.925	0.925	0.952	0.952	0.927	0.888	0.943	0.930

(b) Interval Length

	Gaussian Distributional Approximation								**Bootstrap Distributional Approximation**							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149	0.149
$K/n = 0.101$	0.148	0.156	0.148	0.156	0.156	0.164	0.183	0.156	0.159	0.159	0.157	0.157	0.157	0.156	0.156	0.157
$K/n = 0.201$	0.148	0.166	0.148	0.166	0.165	0.185	0.230	0.165	0.176	0.176	0.168	0.168	0.166	0.163	0.161	0.164
$K/n = 0.301$	0.148	0.177	0.148	0.177	0.177	0.211	0.223	0.176	0.205	0.205	0.186	0.186	0.178	0.169	0.176	0.173
$K/n = 0.401$	0.148	0.192	0.148	0.191	0.191	0.247	0.216	0.190	0.256	0.256	0.221	0.221	0.196	0.172	0.211	0.188
Heteroskedastic Model																
$K/n = 0.001$	0.147	0.147	0.213	0.213	0.213	0.213	0.213	0.213	0.214	0.214	0.217	0.217	0.217	0.217	0.217	0.217
$K/n = 0.101$	0.148	0.156	0.305	0.322	0.322	0.340	0.380	0.336	0.318	0.318	0.376	0.376	0.375	0.375	0.374	0.378
$K/n = 0.201$	0.148	0.165	0.279	0.312	0.312	0.350	0.436	0.339	0.315	0.315	0.382	0.382	0.378	0.374	0.371	0.385
$K/n = 0.301$	0.149	0.179	0.255	0.305	0.305	0.366	0.387	0.345	0.327	0.327	0.397	0.397	0.381	0.364	0.378	0.391
$K/n = 0.401$	0.148	0.191	0.231	0.299	0.299	0.387	0.339	0.352	0.354	0.354	0.418	0.418	0.372	0.327	0.399	0.386

Table 7: Simulation Results, Model 6, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								**Bootstrap Distributional Approximation**							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.948	0.949	0.948	0.948	0.948	0.948	0.948	0.948	0.944	0.944	0.944	0.944	0.944	0.944	0.944	0.944
$K/n = 0.101$	0.937	0.951	0.937	0.950	0.950	0.963	0.979	0.950	0.948	0.948	0.947	0.947	0.947	0.946	0.945	0.946
$K/n = 0.201$	0.914	0.944	0.914	0.945	0.944	0.971	0.992	0.943	0.962	0.962	0.948	0.948	0.943	0.938	0.938	0.940
$K/n = 0.301$	0.905	0.951	0.904	0.949	0.949	0.980	0.986	0.948	0.977	0.977	0.963	0.963	0.950	0.942	0.948	0.945
$K/n = 0.401$	0.875	0.952	0.872	0.951	0.950	0.988	0.972	0.949	0.988	0.988	0.973	0.973	0.952	0.920	0.967	0.941
Heteroskedastic Model																
$K/n = 0.001$	0.880	0.880	0.949	0.949	0.949	0.949	0.949	0.949	0.947	0.947	0.950	0.950	0.950	0.950	0.950	0.950
$K/n = 0.101$	0.578	0.609	0.903	0.919	0.919	0.944	0.960	0.941	0.908	0.908	0.922	0.922	0.922	0.922	0.932	0.944
$K/n = 0.201$	0.575	0.630	0.872	0.911	0.911	0.948	0.980	0.947	0.903	0.903	0.917	0.917	0.915	0.913	0.921	0.938
$K/n = 0.301$	0.565	0.653	0.840	0.902	0.902	0.950	0.962	0.946	0.909	0.909	0.922	0.922	0.910	0.911	0.920	0.937
$K/n = 0.401$	0.566	0.689	0.797	0.895	0.895	0.961	0.934	0.948	0.931	0.931	0.959	0.959	0.936	0.899	0.943	0.940

(b) Interval Length

	Gaussian Distributional Approximation								**Bootstrap Distributional Approximation**							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149	0.149
$K/n = 0.101$	0.148	0.156	0.148	0.156	0.156	0.165	0.183	0.156	0.158	0.158	0.156	0.156	0.156	0.156	0.155	0.156
$K/n = 0.201$	0.148	0.166	0.148	0.166	0.166	0.185	0.231	0.165	0.176	0.176	0.168	0.168	0.165	0.163	0.161	0.164
$K/n = 0.301$	0.148	0.177	0.148	0.177	0.177	0.212	0.224	0.177	0.205	0.205	0.187	0.187	0.178	0.169	0.176	0.173
$K/n = 0.401$	0.148	0.191	0.148	0.191	0.191	0.246	0.216	0.190	0.256	0.256	0.222	0.222	0.196	0.172	0.211	0.187
Heteroskedastic Model																
$K/n = 0.001$	0.148	0.148	0.187	0.188	0.187	0.188	0.188	0.187	0.189	0.189	0.190	0.190	0.190	0.190	0.190	0.190
$K/n = 0.101$	0.148	0.156	0.311	0.328	0.329	0.347	0.387	0.343	0.324	0.324	0.368	0.368	0.367	0.367	0.366	0.369
$K/n = 0.201$	0.148	0.165	0.287	0.321	0.321	0.360	0.448	0.349	0.321	0.321	0.377	0.377	0.373	0.368	0.365	0.377
$K/n = 0.301$	0.148	0.177	0.260	0.311	0.311	0.373	0.394	0.354	0.329	0.329	0.388	0.388	0.372	0.354	0.369	0.377
$K/n = 0.401$	0.149	0.192	0.238	0.309	0.308	0.399	0.349	0.365	0.364	0.364	0.427	0.427	0.379	0.333	0.407	0.386

Table 8: Simulation Results, Model 7, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.951	0.952	0.950	0.950	0.950	0.950	0.950	0.950	0.952	0.952	0.952	0.952	0.952	0.952	0.952	0.952
$K/n = 0.100$	0.934	0.950	0.934	0.949	0.949	0.961	0.975	0.948	0.948	0.948	0.944	0.944	0.945	0.945	0.946	0.944
$K/n = 0.200$	0.919	0.951	0.918	0.949	0.949	0.971	0.993	0.948	0.949	0.949	0.932	0.932	0.931	0.930	0.709	0.031
$K/n = 0.250$	0.910	0.951	0.908	0.949	0.949	0.976	0.990	0.947	0.768	0.768	0.727	0.727	0.693	0.696	NA	NA
$K/n = 0.333$	0.892	0.953	0.890	0.951	0.951	0.983	0.983	0.950	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.629	0.630	0.934	0.935	0.935	0.945	0.946	0.945	0.929	0.929	0.931	0.931	0.931	0.931	0.931	0.931
$K/n = 0.100$	0.597	0.618	0.903	0.921	0.921	0.946	0.961	0.942	0.911	0.911	0.936	0.936	0.936	0.936	0.938	0.935
$K/n = 0.200$	0.606	0.664	0.878	0.917	0.917	0.946	0.981	0.946	0.916	0.916	0.915	0.915	0.916	0.920	0.686	0.026
$K/n = 0.250$	0.600	0.670	0.853	0.908	0.907	0.948	0.975	0.943	0.716	0.716	0.690	0.690	0.645	0.649	NA	NA
$K/n = 0.333$	0.594	0.691	0.822	0.900	0.900	0.956	0.956	0.944	NA	NA	NA	NA	NA	NA	NA	NA

(b) Interval Length

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149	0.149
$K/n = 0.100$	0.148	0.156	0.148	0.156	0.156	0.164	0.183	0.156	0.160	0.160	0.158	0.158	0.158	0.158	0.158	0.156
$K/n = 0.200$	0.148	0.166	0.148	0.165	0.165	0.185	0.231	0.165	0.178	0.178	0.166	0.166	0.164	0.164	0.109	0.004
$K/n = 0.250$	0.148	0.171	0.148	0.171	0.171	0.197	0.227	0.170	0.127	0.127	0.114	0.114	0.106	0.106	NA	NA
$K/n = 0.333$	0.148	0.182	0.148	0.181	0.181	0.222	0.222	0.181	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.136	0.136	0.292	0.292	0.292	0.292	0.292	0.292	0.293	0.293	0.322	0.322	0.322	0.322	0.322	0.322
$K/n = 0.100$	0.136	0.144	0.268	0.283	0.283	0.298	0.331	0.294	0.284	0.284	0.326	0.326	0.326	0.326	0.329	0.328
$K/n = 0.200$	0.136	0.152	0.246	0.275	0.275	0.308	0.385	0.296	0.289	0.289	0.310	0.310	0.307	0.308	0.204	0.006
$K/n = 0.250$	0.135	0.156	0.236	0.273	0.272	0.314	0.363	0.297	0.203	0.203	0.203	0.203	0.188	0.190	NA	NA
$K/n = 0.333$	0.136	0.167	0.226	0.277	0.277	0.339	0.339	0.305	NA	NA	NA	NA	NA	NA	NA	NA

Table 9: Simulation Results, Model 8, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								**Bootstrap Distributional Approximation**							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.951	0.951	0.951	0.951	0.951	0.951	0.951	0.951	0.950	0.950	0.949	0.949	0.949	0.949	0.949	0.949
$K/n = 0.100$	0.938	0.950	0.936	0.949	0.949	0.961	0.979	0.949	0.950	0.950	0.946	0.946	0.945	0.945	0.947	0.945
$K/n = 0.200$	0.922	0.955	0.921	0.951	0.951	0.972	0.994	0.951	0.949	0.949	0.937	0.937	0.936	0.936	0.699	0.010
$K/n = 0.250$	0.903	0.948	0.903	0.947	0.947	0.978	0.991	0.947	0.757	0.757	0.717	0.717	0.666	0.669	NA	NA
$K/n = 0.333$	0.886	0.948	0.886	0.944	0.943	0.982	0.982	0.942	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.601	0.601	0.931	0.932	0.931	0.952	0.962	0.952	0.915	0.915	0.926	0.926	0.926	0.926	0.926	0.926
$K/n = 0.100$	0.602	0.631	0.908	0.924	0.923	0.947	0.962	0.944	0.919	0.919	0.936	0.936	0.937	0.936	0.934	0.934
$K/n = 0.200$	0.610	0.655	0.879	0.916	0.915	0.946	0.982	0.946	0.910	0.910	0.918	0.918	0.917	0.918	0.700	0.015
$K/n = 0.250$	0.590	0.652	0.847	0.906	0.906	0.945	0.972	0.943	0.701	0.701	0.675	0.675	0.636	0.642	NA	NA
$K/n = 0.333$	0.593	0.691	0.832	0.900	0.899	0.952	0.952	0.947	NA	NA	NA	NA	NA	NA	NA	NA

(b) Interval Length

	Gaussian Distributional Approximation								**Bootstrap Distributional Approximation**							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149	0.149
$K/n = 0.100$	0.148	0.156	0.148	0.156	0.156	0.164	0.182	0.156	0.160	0.160	0.158	0.158	0.158	0.158	0.159	0.157
$K/n = 0.200$	0.148	0.166	0.148	0.166	0.165	0.185	0.231	0.165	0.178	0.178	0.166	0.166	0.164	0.165	0.109	0.003
$K/n = 0.250$	0.148	0.171	0.148	0.171	0.171	0.197	0.228	0.171	0.128	0.128	0.114	0.114	0.105	0.106	NA	NA
$K/n = 0.333$	0.148	0.182	0.148	0.181	0.181	0.222	0.222	0.181	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.139	0.139	0.305	0.305	0.305	0.305	0.306	0.305	0.304	0.304	0.337	0.337	0.337	0.337	0.337	0.337
$K/n = 0.100$	0.138	0.146	0.279	0.294	0.294	0.310	0.345	0.306	0.295	0.295	0.343	0.343	0.343	0.343	0.346	0.344
$K/n = 0.200$	0.138	0.154	0.257	0.287	0.287	0.321	0.401	0.309	0.304	0.304	0.333	0.333	0.331	0.332	0.221	0.004
$K/n = 0.250$	0.138	0.160	0.248	0.286	0.286	0.330	0.381	0.312	0.211	0.211	0.214	0.214	0.198	0.200	NA	NA
$K/n = 0.333$	0.138	0.169	0.235	0.288	0.288	0.353	0.353	0.317	NA	NA	NA	NA	NA	NA	NA	NA

Table 10: Simulation Results, Model 9, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								**Bootstrap Distributional Approximation**							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.951	0.952	0.951	0.951	0.951	0.951	0.951	0.951	0.948	0.948	0.950	0.950	0.950	0.950	0.950	0.950
$K/n = 0.100$	0.934	0.947	0.932	0.947	0.946	0.956	0.972	0.946	0.949	0.949	0.944	0.944	0.945	0.945	0.945	0.943
$K/n = 0.200$	0.916	0.948	0.916	0.947	0.947	0.970	0.993	0.947	0.954	0.954	0.939	0.939	0.935	0.935	0.705	0.018
$K/n = 0.250$	0.901	0.945	0.900	0.943	0.943	0.972	0.989	0.942	0.747	0.747	0.709	0.709	0.662	0.663	NA	NA
$K/n = 0.333$	0.887	0.949	0.884	0.948	0.948	0.982	0.982	0.948	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.614	0.614	0.945	0.945	0.945	0.945	0.946	0.945	0.943	0.943	0.957	0.957	0.957	0.957	0.957	0.957
$K/n = 0.100$	0.596	0.621	0.918	0.935	0.935	0.948	0.968	0.945	0.920	0.920	0.947	0.947	0.946	0.947	0.950	0.946
$K/n = 0.200$	0.591	0.646	0.883	0.919	0.919	0.950	0.985	0.949	0.909	0.909	0.919	0.919	0.913	0.916	0.681	0.022
$K/n = 0.250$	0.582	0.647	0.857	0.907	0.907	0.947	0.971	0.944	0.697	0.697	0.677	0.677	0.644	0.646	NA	NA
$K/n = 0.333$	0.588	0.684	0.836	0.910	0.910	0.958	0.958	0.945	NA	NA	NA	NA	NA	NA	NA	NA

(b) Interval Length

	Gaussian Distributional Approximation								**Bootstrap Distributional Approximation**							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149	0.149
$K/n = 0.100$	0.148	0.156	0.148	0.156	0.156	0.164	0.183	0.156	0.161	0.161	0.158	0.158	0.158	0.158	0.159	0.157
$K/n = 0.200$	0.148	0.166	0.148	0.165	0.165	0.185	0.231	0.165	0.177	0.177	0.164	0.164	0.162	0.163	0.109	0.002
$K/n = 0.250$	0.148	0.171	0.148	0.171	0.171	0.197	0.227	0.170	0.126	0.126	0.112	0.112	0.104	0.104	NA	NA
$K/n = 0.333$	0.148	0.182	0.148	0.181	0.181	0.222	0.222	0.181	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.134	0.134	0.299	0.299	0.299	0.299	0.299	0.299	0.298	0.298	0.318	0.318	0.318	0.318	0.318	0.318
$K/n = 0.100$	0.134	0.141	0.274	0.289	0.289	0.304	0.338	0.300	0.289	0.289	0.323	0.323	0.323	0.323	0.325	0.324
$K/n = 0.200$	0.135	0.151	0.254	0.284	0.284	0.317	0.396	0.306	0.298	0.298	0.314	0.314	0.311	0.312	0.210	0.004
$K/n = 0.250$	0.134	0.155	0.243	0.281	0.280	0.324	0.374	0.307	0.206	0.206	0.203	0.203	0.189	0.190	NA	NA
$K/n = 0.333$	0.135	0.166	0.232	0.285	0.285	0.349	0.349	0.314	NA	NA	NA	NA	NA	NA	NA	NA

Table 11: Simulation Results, Model 10, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								**Bootstrap Distributional Approximation**							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.946	0.946	0.944	0.944	0.944	0.944	0.945	0.944	0.944	0.944	0.944	0.944	0.944	0.944	0.944	0.944
$K/n = 0.133$	0.939	0.955	0.939	0.956	0.955	0.969	0.988	0.954	0.952	0.952	0.944	0.944	0.932	0.933	0.833	0.363
$K/n = 0.266$	0.909	0.952	0.911	0.952	0.950	0.980	0.988	0.949	0.723	0.723	0.692	0.692	0.554	0.550	NA	NA
$K/n = 0.333$	0.888	0.950	0.892	0.950	0.947	0.981	0.979	0.947	0.019	0.019	0.019	0.019	0.009	0.009	NA	NA
$K/n = 0.443$	0.869	0.954	0.869	0.954	0.951	0.990	0.964	0.949	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.607	0.607	0.933	0.934	0.933	0.944	0.944	0.944	0.929	0.929	0.931	0.931	0.931	0.931	0.931	0.931
$K/n = 0.133$	0.606	0.643	0.903	0.926	0.924	0.949	0.971	0.948	0.915	0.915	0.931	0.931	0.915	0.916	0.801	0.384
$K/n = 0.266$	0.582	0.655	0.853	0.912	0.908	0.947	0.968	0.941	0.657	0.657	0.650	0.650	0.529	0.526	NA	NA
$K/n = 0.333$	0.582	0.687	0.828	0.904	0.899	0.952	0.950	0.947	0.022	0.022	0.022	0.022	0.009	0.009	NA	NA
$K/n = 0.443$	0.568	0.720	0.789	0.909	0.906	0.972	0.926	0.944	NA	NA	NA	NA	NA	NA	NA	NA

(b) Interval Length

	Gaussian Distributional Approximation								**Bootstrap Distributional Approximation**							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149	0.149
$K/n = 0.133$	0.148	0.159	0.149	0.160	0.159	0.170	0.201	0.159	0.164	0.164	0.158	0.158	0.151	0.152	0.128	0.051
$K/n = 0.266$	0.148	0.173	0.149	0.175	0.173	0.202	0.222	0.173	0.122	0.122	0.109	0.109	0.084	0.084	NA	NA
$K/n = 0.333$	0.148	0.182	0.149	0.183	0.181	0.222	0.218	0.181	0.003	0.003	0.003	0.003	0.002	0.002	NA	NA
$K/n = 0.443$	0.148	0.199	0.149	0.200	0.198	0.265	0.212	0.197	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.147	0.148	0.324	0.324	0.324	0.324	0.324	0.324	0.323	0.323	0.353	0.353	0.353	0.353	0.353	0.353
$K/n = 0.133$	0.148	0.160	0.297	0.319	0.316	0.337	0.391	0.330	0.319	0.319	0.361	0.361	0.343	0.344	0.290	0.118
$K/n = 0.266$	0.148	0.172	0.268	0.313	0.308	0.356	0.394	0.337	0.211	0.211	0.216	0.216	0.170	0.169	NA	NA
$K/n = 0.333$	0.149	0.183	0.256	0.313	0.308	0.374	0.371	0.342	0.008	0.008	0.008	0.008	0.006	0.006	NA	NA
$K/n = 0.443$	0.147	0.198	0.234	0.315	0.309	0.411	0.333	0.347	NA	NA	NA	NA	NA	NA	NA	NA

Table 12: Simulation Results, Model 11, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.949	0.950	0.945	0.946	0.946	0.946	0.947	0.946	0.947	0.947	0.950	0.950	0.950	0.950	0.950	0.950
$K/n = 0.133$	0.930	0.950	0.932	0.951	0.951	0.965	0.985	0.950	0.935	0.935	0.930	0.930	0.912	0.914	0.823	0.379
$K/n = 0.266$	0.907	0.947	0.909	0.949	0.947	0.973	0.988	0.946	0.713	0.713	0.678	0.678	0.555	0.555	NA	NA
$K/n = 0.333$	0.894	0.949	0.896	0.951	0.949	0.982	0.981	0.948	0.029	0.029	0.029	0.029	0.016	0.016	NA	NA
$K/n = 0.443$	0.859	0.951	0.860	0.952	0.950	0.991	0.963	0.946	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.595	0.595	0.932	0.932	0.932	0.942	0.943	0.942	0.931	0.931	0.934	0.934	0.934	0.934	0.934	0.934
$K/n = 0.133$	0.582	0.614	0.900	0.922	0.918	0.946	0.967	0.944	0.908	0.908	0.926	0.926	0.912	0.914	0.798	0.346
$K/n = 0.266$	0.574	0.646	0.846	0.907	0.902	0.945	0.966	0.944	0.659	0.659	0.640	0.640	0.523	0.522	NA	NA
$K/n = 0.333$	0.575	0.674	0.839	0.910	0.904	0.959	0.958	0.947	0.029	0.029	0.023	0.023	0.013	0.013	NA	NA
$K/n = 0.443$	0.566	0.709	0.787	0.904	0.899	0.973	0.926	0.947	NA	NA	NA	NA	NA	NA	NA	NA

(b) Interval Length

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149	0.149
$K/n = 0.133$	0.148	0.159	0.148	0.160	0.159	0.170	0.201	0.159	0.164	0.164	0.159	0.159	0.151	0.152	0.129	0.052
$K/n = 0.266$	0.148	0.173	0.149	0.174	0.173	0.201	0.222	0.172	0.120	0.120	0.107	0.107	0.084	0.084	NA	NA
$K/n = 0.333$	0.148	0.181	0.149	0.183	0.181	0.221	0.218	0.180	0.005	0.005	0.004	0.004	0.003	0.003	NA	NA
$K/n = 0.443$	0.148	0.199	0.149	0.200	0.198	0.265	0.212	0.197	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.149	0.149	0.335	0.335	0.335	0.335	0.336	0.335	0.335	0.335	0.374	0.374	0.374	0.374	0.374	0.374
$K/n = 0.133$	0.148	0.159	0.304	0.326	0.323	0.346	0.401	0.339	0.326	0.326	0.377	0.377	0.359	0.359	0.301	0.119
$K/n = 0.266$	0.149	0.174	0.273	0.318	0.313	0.363	0.401	0.343	0.218	0.218	0.224	0.224	0.175	0.174	NA	NA
$K/n = 0.333$	0.147	0.181	0.259	0.317	0.312	0.378	0.376	0.347	0.008	0.008	0.007	0.007	0.005	0.005	NA	NA
$K/n = 0.443$	0.148	0.198	0.238	0.320	0.314	0.418	0.338	0.354	NA	NA	NA	NA	NA	NA	NA	NA

Table 13: Simulation Results, Model 12, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.948	0.948	0.948	0.948	0.948	0.948	0.948	0.948	0.945	0.945	0.947	0.947	0.947	0.947	0.947	0.947
$K/n = 0.133$	0.934	0.950	0.934	0.950	0.949	0.963	0.987	0.949	0.945	0.945	0.937	0.937	0.921	0.922	0.829	0.360
$K/n = 0.266$	0.913	0.951	0.914	0.952	0.950	0.979	0.988	0.950	0.706	0.706	0.674	0.674	0.551	0.551	NA	NA
$K/n = 0.333$	0.891	0.948	0.892	0.950	0.947	0.983	0.981	0.946	0.024	0.024	0.027	0.027	0.014	0.014	NA	NA
$K/n = 0.443$	0.854	0.947	0.858	0.948	0.946	0.991	0.958	0.944	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.599	0.599	0.933	0.933	0.933	0.947	0.934	0.943	0.918	0.918	0.933	0.933	0.933	0.933	0.933	0.933
$K/n = 0.133$	0.590	0.624	0.911	0.932	0.929	0.944	0.975	0.944	0.917	0.917	0.936	0.936	0.918	0.916	0.823	0.372
$K/n = 0.266$	0.568	0.648	0.853	0.909	0.903	0.946	0.967	0.946	0.680	0.680	0.671	0.671	0.540	0.540	NA	NA
$K/n = 0.333$	0.559	0.653	0.828	0.905	0.899	0.953	0.953	0.945	0.031	0.031	0.027	0.027	0.017	0.017	NA	NA
$K/n = 0.443$	0.557	0.705	0.790	0.904	0.898	0.966	0.920	0.949	NA	NA	NA	NA	NA	NA	NA	NA

(b) Interval Length

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.148	0.149	0.149	0.149	0.149	0.149	0.149
$K/n = 0.133$	0.148	0.159	0.149	0.160	0.159	0.171	0.201	0.159	0.164	0.164	0.159	0.159	0.152	0.152	0.129	0.053
$K/n = 0.266$	0.148	0.173	0.149	0.174	0.172	0.201	0.221	0.172	0.121	0.121	0.108	0.108	0.085	0.085	NA	NA
$K/n = 0.333$	0.148	0.182	0.149	0.183	0.181	0.221	0.218	0.181	0.005	0.005	0.004	0.004	0.003	0.003	NA	NA
$K/n = 0.443$	0.148	0.199	0.149	0.200	0.198	0.266	0.213	0.198	NA	NA	NA	NA	NA	NA	NA	NA
Heteroskedastic Model																
$K/n = 0.001$	0.148	0.148	0.338	0.338	0.338	0.338	0.338	0.338	0.340	0.340	0.362	0.362	0.362	0.362	0.362	0.362
$K/n = 0.133$	0.148	0.159	0.309	0.332	0.328	0.351	0.406	0.344	0.330	0.330	0.365	0.365	0.347	0.348	0.291	0.120
$K/n = 0.266$	0.148	0.173	0.279	0.326	0.320	0.370	0.409	0.351	0.222	0.222	0.222	0.222	0.174	0.173	NA	NA
$K/n = 0.333$	0.148	0.181	0.264	0.324	0.318	0.385	0.384	0.355	0.007	0.007	0.006	0.006	0.004	0.004	NA	NA
$K/n = 0.443$	0.148	0.198	0.242	0.325	0.319	0.423	0.344	0.360	NA	NA	NA	NA	NA	NA	NA	NA

Table 14: Simulation Results, Model 13, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.876	0.876	0.878	0.878	0.878	0.878	0.878	0.878	0.885	0.885	0.888	0.888	0.888	0.888	0.888	0.888
$K/n = 0.019$	0.948	0.950	0.947	0.950	0.950	0.951	0.956	0.949	0.949	0.949	0.949	0.949	0.949	0.949	0.949	0.950
$K/n = 0.049$	0.938	0.945	0.939	0.945	0.945	0.953	0.963	0.945	0.940	0.940	0.940	0.940	0.940	0.940	0.940	0.940
$K/n = 0.129$	0.936	0.953	0.935	0.951	0.950	0.965	0.984	0.950	0.954	0.954	0.950	0.950	0.950	0.950	0.949	0.949
$K/n = 0.309$	0.905	0.957	0.912	0.960	0.955	0.985	0.984	0.951	0.978	0.978	0.957	0.957	0.947	0.929	0.948	0.939
Heteroskedastic Model																
$K/n = 0.001$	0.408	0.409	0.450	0.451	0.450	0.451	0.452	0.450	0.450	0.450	0.459	0.459	0.459	0.459	0.459	0.459
$K/n = 0.019$	0.760	0.765	0.937	0.940	0.940	0.941	0.947	0.941	0.936	0.936	0.941	0.941	0.942	0.942	0.942	0.941
$K/n = 0.049$	0.752	0.769	0.925	0.932	0.933	0.943	0.958	0.946	0.928	0.928	0.933	0.933	0.932	0.932	0.934	0.933
$K/n = 0.129$	0.756	0.787	0.904	0.920	0.925	0.951	0.981	0.944	0.919	0.919	0.928	0.928	0.928	0.929	0.927	0.928
$K/n = 0.309$	0.723	0.807	0.869	0.930	0.926	0.976	0.972	0.947	0.945	0.945	0.946	0.946	0.939	0.926	0.943	0.933

(b) Interval Length

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.094	0.094	0.094	0.094	0.094	0.094	0.094	0.094	0.094	0.094	0.095	0.095	0.095	0.095	0.095	0.095
$K/n = 0.019$	0.148	0.149	0.148	0.149	0.149	0.150	0.153	0.149	0.149	0.149	0.150	0.150	0.150	0.150	0.150	0.150
$K/n = 0.049$	0.148	0.152	0.148	0.151	0.151	0.155	0.163	0.151	0.152	0.152	0.152	0.152	0.152	0.152	0.152	0.152
$K/n = 0.129$	0.148	0.158	0.148	0.159	0.158	0.169	0.197	0.158	0.163	0.163	0.159	0.159	0.158	0.158	0.156	0.157
$K/n = 0.309$	0.148	0.178	0.152	0.183	0.178	0.214	0.212	0.178	0.210	0.210	0.188	0.188	0.181	0.169	0.182	0.176
Heteroskedastic Model																
$K/n = 0.001$	0.052	0.052	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.057	0.057	0.057	0.057	0.057	0.057
$K/n = 0.019$	0.220	0.222	0.347	0.351	0.351	0.355	0.363	0.353	0.350	0.350	0.371	0.371	0.371	0.371	0.372	0.372
$K/n = 0.049$	0.221	0.226	0.342	0.351	0.353	0.364	0.388	0.357	0.348	0.348	0.372	0.372	0.372	0.373	0.374	0.374
$K/n = 0.129$	0.221	0.237	0.328	0.352	0.356	0.389	0.471	0.365	0.359	0.359	0.380	0.380	0.381	0.381	0.383	0.382
$K/n = 0.309$	0.230	0.277	0.327	0.394	0.392	0.485	0.474	0.411	0.445	0.445	0.454	0.454	0.440	0.415	0.445	0.432

Table 15: Simulation Results, Model 14, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.865	0.866	0.862	0.863	0.863	0.863	0.863	0.863	0.869	0.869	0.873	0.873	0.873	0.873	0.873	0.873
$K/n = 0.019$	0.954	0.955	0.949	0.951	0.951	0.954	0.960	0.951	0.948	0.948	0.950	0.950	0.950	0.950	0.951	0.951
$K/n = 0.049$	0.948	0.954	0.945	0.952	0.951	0.957	0.965	0.951	0.948	0.948	0.948	0.948	0.948	0.948	0.948	0.948
$K/n = 0.129$	0.939	0.953	0.940	0.953	0.953	0.963	0.983	0.953	0.951	0.951	0.948	0.948	0.948	0.946	0.945	0.946
$K/n = 0.309$	0.897	0.952	0.906	0.958	0.951	0.981	0.980	0.949	0.976	0.976	0.961	0.961	0.954	0.935	0.956	0.947
Heteroskedastic Model																
$K/n = 0.001$	0.395	0.396	0.436	0.437	0.436	0.437	0.438	0.436	0.452	0.452	0.486	0.486	0.486	0.486	0.486	0.486
$K/n = 0.019$	0.755	0.759	0.939	0.943	0.942	0.945	0.949	0.944	0.935	0.935	0.938	0.938	0.938	0.938	0.938	0.938
$K/n = 0.049$	0.759	0.770	0.928	0.934	0.935	0.946	0.959	0.945	0.932	0.932	0.938	0.938	0.938	0.938	0.939	0.938
$K/n = 0.129$	0.743	0.780	0.912	0.933	0.937	0.957	0.985	0.948	0.927	0.927	0.936	0.936	0.936	0.936	0.937	0.935
$K/n = 0.309$	0.703	0.789	0.862	0.927	0.928	0.974	0.969	0.946	0.943	0.943	0.938	0.938	0.931	0.915	0.933	0.925

(b) Interval Length

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.090	0.091	0.090	0.090	0.090	0.090	0.090	0.090	0.091	0.091	0.091	0.091	0.091	0.091	0.091	0.091
$K/n = 0.019$	0.148	0.149	0.147	0.149	0.149	0.150	0.153	0.149	0.149	0.149	0.150	0.150	0.150	0.150	0.150	0.150
$K/n = 0.049$	0.148	0.152	0.148	0.151	0.151	0.155	0.163	0.151	0.152	0.152	0.152	0.152	0.152	0.152	0.152	0.152
$K/n = 0.129$	0.148	0.158	0.148	0.159	0.158	0.169	0.197	0.158	0.162	0.162	0.159	0.159	0.158	0.158	0.156	0.158
$K/n = 0.309$	0.148	0.178	0.152	0.183	0.178	0.214	0.212	0.177	0.210	0.210	0.189	0.189	0.182	0.169	0.182	0.176
Heteroskedastic Model																
$K/n = 0.001$	0.052	0.053	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.057	0.057	0.057	0.057	0.057	0.057
$K/n = 0.019$	0.222	0.224	0.349	0.352	0.352	0.356	0.364	0.354	0.350	0.350	0.373	0.373	0.373	0.373	0.373	0.374
$K/n = 0.049$	0.220	0.225	0.339	0.348	0.350	0.361	0.384	0.354	0.346	0.346	0.372	0.372	0.372	0.373	0.374	0.374
$K/n = 0.129$	0.221	0.237	0.328	0.352	0.356	0.389	0.472	0.365	0.358	0.358	0.381	0.381	0.382	0.383	0.386	0.384
$K/n = 0.309$	0.230	0.277	0.325	0.391	0.390	0.484	0.472	0.409	0.444	0.444	0.452	0.452	0.437	0.413	0.443	0.431

Table 16: Simulation Results, Model 15, $n = 700$, $S = 5,000$, $B = 500$.

(a) Empirical Coverage

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.941	0.941	0.940	0.940	0.940	0.940	0.941	0.940	0.939	0.939	0.941	0.941	0.941	0.941	0.941	0.941
$K/n = 0.019$	0.946	0.949	0.946	0.949	0.949	0.950	0.955	0.949	0.945	0.945	0.949	0.949	0.949	0.949	0.949	0.949
$K/n = 0.049$	0.943	0.947	0.942	0.948	0.947	0.953	0.963	0.947	0.939	0.939	0.939	0.939	0.939	0.939	0.939	0.939
$K/n = 0.129$	0.931	0.948	0.932	0.948	0.946	0.963	0.984	0.946	0.949	0.949	0.943	0.943	0.943	0.943	0.940	0.941
$K/n = 0.309$	0.901	0.952	0.909	0.960	0.952	0.982	0.981	0.952	0.974	0.974	0.958	0.958	0.949	0.931	0.949	0.940
Heteroskedastic Model																
$K/n = 0.001$	0.407	0.407	0.451	0.452	0.452	0.452	0.453	0.452	0.470	0.470	0.484	0.484	0.484	0.484	0.484	0.484
$K/n = 0.019$	0.763	0.769	0.939	0.940	0.940	0.942	0.946	0.940	0.944	0.934	0.949	0.949	0.949	0.949	0.949	0.949
$K/n = 0.049$	0.760	0.771	0.932	0.937	0.939	0.946	0.958	0.942	0.945	0.935	0.944	0.944	0.945	0.945	0.945	0.945
$K/n = 0.129$	0.744	0.778	0.909	0.933	0.936	0.956	0.984	0.943	0.946	0.926	0.935	0.935	0.934	0.937	0.935	0.936
$K/n = 0.309$	0.710	0.793	0.865	0.923	0.924	0.971	0.967	0.934	0.949	0.949	0.946	0.946	0.940	0.923	0.940	0.932

(b) Interval Length

	Gaussian Distributional Approximation								Bootstrap Distributional Approximation							
	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK	HO0	HO1	HC0	HC1	HC2	HC3	HC4	HCK
Homoskedastic Model																
$K/n = 0.001$	0.120	0.121	0.120	0.120	0.120	0.120	0.121	0.120	0.120	0.120	0.121	0.121	0.121	0.121	0.121	0.121
$K/n = 0.019$	0.148	0.149	0.148	0.149	0.149	0.151	0.154	0.149	0.149	0.149	0.150	0.150	0.150	0.150	0.150	0.150
$K/n = 0.049$	0.148	0.152	0.148	0.152	0.152	0.155	0.163	0.152	0.152	0.152	0.152	0.152	0.152	0.152	0.152	0.152
$K/n = 0.129$	0.148	0.159	0.148	0.159	0.158	0.170	0.198	0.158	0.163	0.163	0.159	0.159	0.159	0.158	0.157	0.158
$K/n = 0.309$	0.148	0.178	0.152	0.184	0.178	0.214	0.213	0.178	0.209	0.209	0.188	0.188	0.180	0.168	0.181	0.175
Heteroskedastic Model																
$K/n = 0.001$	0.052	0.052	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.057	0.057	0.057	0.057	0.057	0.057
$K/n = 0.019$	0.218	0.221	0.350	0.354	0.354	0.358	0.366	0.356	0.352	0.352	0.367	0.367	0.367	0.367	0.367	0.367
$K/n = 0.049$	0.219	0.225	0.343	0.352	0.354	0.365	0.389	0.358	0.350	0.350	0.368	0.368	0.369	0.369	0.370	0.369
$K/n = 0.129$	0.222	0.238	0.333	0.358	0.362	0.395	0.480	0.372	0.365	0.365	0.383	0.383	0.383	0.384	0.386	0.384
$K/n = 0.309$	0.230	0.277	0.330	0.397	0.395	0.490	0.478	0.416	0.448	0.448	0.451	0.451	0.437	0.414	0.442	0.428

References

Barnes, E. R., and A. J. Hoffman (1981): "On Bounds for Eigenvalues of Real Symmetric Matrices," *Linear Algebra and its Applications*, 40, 217–223.

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2015): "On the Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*, 186(2), 345–366.

Cattaneo, M. D., and M. H. Farrell (2013): "Optimal Convergence Rates, Bahadur Representation, and Asymptotic Normality of Partitioning Estimators," *Journal of Econometrics*, 174(2), 127–143.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018): "Alternative Asymptotics and the Partially Linear Model with Many Regressors," *Econometric Theory*, 34(2), 277–301.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," in *Handbook of Econometrics, Volume VI*, ed. by J. J. Heckman, and E. E. Leamer, pp. 5549–5632. Elsevier Science B.V., Amsterdam, Netherlands.

MacKinnon, J. G. (2012): "Thirty Years of Heteroskedasticity-Robust Inference," in *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, ed. by X. Chen, and N. R. Swanson. Springer.

Newey, W. K. (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168.

Stock, J. H., and M. W. Watson (2008): "Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression," *Econometrica*, 76(1), 155–174.

Varah, J. M. (1975): "A Lower Bound for the Smallest Singular Value of a Matrix," *Linear Algebra and its Applications*, 11(1), 3–5.

Verdier, V. (2017): "Estimation and Inference for Linear Models with Two-Way Fixed Effects and Sparsely Matched Data," Working paper, UNC.