Heteroskedasticity Consistent Standard Errors for Linear Models with Many Covariates*

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

January 29, 2015

Abstract

The linear regression model is widely used in empirical work in Economics. Researchers often include many covariates in their linear model specification in an attempt to control for observed and unobserved counfounders. In this paper, we show that all of the usual versions of Eicker-White heteroskedasticity consistent standard error estimators for linear models are inconsistent when many covariates are included. We then propose a new heteroskedasticity consistent standard error formula that is fully automatic and robust to both (conditional) heteroskedasticity of unknown form and the inclusion of possibly many covariates. We illustrate our findings in three distinct settings: (i) parametric linear models with many covariates, (ii) semiparametric semilinear models with many technical regressors, and (iii) linear panel models with many fixed effects. In the case of the third example, we also find that our general standard error formula reduces exactly to the one proposed in Stock and Watson (2008), up to an asymptotically negligible degrees-of-freedom correction, despite our result being derived from a completely different perspective.

Keywords: linear regression models, many regressors, heteroskedasticity, standard errors.

PRELIMINARY AND INCOMPLETE – COMMENTSWELCOME

***** PLEASE DO NOT DISTRIBUTE *****

*We thank Urlich Müller and seminar participants at Miami, Montreal and Princeton for their comments.

†Department of Economics, University of Michigan.

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

§Department of Economics, MIT.

1 Introduction

In some important applications of the linear regression model, the object of interest is β in a model of the form

$$y_{i,n} = \beta' x_{i,n} + \gamma_n' w_{i,n} + u_{i,n}, \qquad i = 1, \dots, n, \tag{1}$$

where $y_{i,n}$ is a scalar outcome variable, $x_{i,n}$ is a regressor of fixed dimension d, $w_{i,n}$ is a covariate of growing dimension K_n, and $u_{i,n}$ is an unobserved error term. Two canonical examples, discussed in more detail below, are the fixed effects panel data regression model and a series-based formulation of the partially linear regression model. In both examples, conducting OLS-based inference on β in (1) is straightforward when the error is (conditionally) homoskedastic and/or the dimension of the covariate is modeled as a vanishing fraction of the sample size. The latter assumption is violated in the fixed effects regression model and is somewhat unpalatable in the partially linear regression model. Thus, we study in this paper the consequences of allowing the error $u_{i,n}$ in (1) to be (conditionally) heteroskedastic in a setting where the dimension of the covariate $w_{i,n}$ is modeled as a non-vanishing fraction of the sample size.

- TO BE COMPLETED.

Our results contribute to the already sizeable literature on heteroskedasticity-robust variance estimators for linear regression models, a recent review of which is given by MacKinnon (2012). Important papers whose results are related to ours include White (1980), MacKinnon and White (1985), Wu (1986), Chesher and Jewitt (1987), Shao and Wu (1987), Chesher (1989), Cribari-Neto, Ferrari, and Cordeiro (2000), Bera, Suprayitno, and Premaratne (2002), Stock and Watson (2008), Cribari-Neto and da Gloria A. Lima (2011), and Müller (2013). In particular, Bera, Suprayitno, and Premaratne (2002) analyze some finite sample properties of a variance estimator similar to the one whose asymptotic properties are studied herein. Also, as further discussed below, the variance estimator proposed herein essentially reduces to the bias corrected variance estimator of Stock and Watson (2008) in the special case of the fixed effects model.

The rest of this paper is organized as follows. Section 2 presents the variance estimators we study and gives a heuristic description of their main properties. Section 3 introduces the three leading examples covered by our results, including easy-to-verify primitive conditions. Section 4 introduces a general framework that unifies the examples, gives the main results of the paper, and discusses their implications for the three examples we consider. Finally, Section 6 concludes. The supplemental appendix contains the proofs of all our results, not included in the paper to conserve space.

2 Variance Estimators

For the purposes of discussing variance estimators associated with the OLS estimator $\hat{\beta}_n$ of β in (1) it is convenient to write $\hat{\beta}_n$ in "partialled out" form as

$$\hat{\beta}_n = \left(\sum_{i=1}^n \hat{v}_{i,n}\hat{v}_{i,n}'\right)^{-1}\left(\sum_{i=1}^n \hat{v}_{i,n}y_{i,n}\right), \qquad \hat{v}_{i,n} = \sum_{j=1}^n M_{ij,n}x_{j,n},$$

where $M_{ij,n} = \mathbb{1}(i=j) - w_{i,n}'(\sum_{k=1}^n w_{k,n}w_{k,n}')^{-1}w_{j,n}$, $\mathbb{1}(\cdot)$ denotes the indicator function, and the relevant inverses are assumed to exist. Defining $\hat{\Gamma}_n = \sum_{i=1}^n \hat{v}_{i,n}\hat{v}_{i,n}'/n$, the objective is to find an estimator $\hat{\Sigma}_n$ of the variance of $\sum_{i=1}^n \hat{v}_{i,n}u_{i,n}/\sqrt{n}$ such that

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

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

Defining $\hat{u}_{i,n} = \sum_{j=1}^n M_{ij,n}(y_{j,n} - \hat{\beta}_n' x_{j,n})$, standard choices of $\hat{\Sigma}_n$ in the fixed-K_n case include the homoskedasticity-only estimator

$$\hat{\Sigma}_n^{\mathtt{HO}} = \hat{\sigma}_n^2\hat{\Gamma}_n, \qquad \hat{\sigma}_n^2 = \frac{1}{n-d-K_n}\sum_{i=1}^n \hat{u}_{i,n}^2,$$

and the Eicker-White-type estimator

$$\hat{\Sigma}_n^{\mathtt{EW}} = \frac{1}{n}\sum_{i=1}^n \hat{v}_{i,n}\hat{v}_{i,n}'\hat{u}_{i,n}^2.$$

Perhaps not too surprisingly, we find that consistency of $\hat{\Sigma}_n^{\mathtt{HO}}$ under homoskedasticity holds quite generally even for models with many covariates. In contrast, construction of a heteroskedasticity-robust estimator of Σ_n is more challenging, as it turns out that consistency of $\hat{\Sigma}_n^{\mathtt{EW}}$ generally requires K_n to be a vanishing fraction of n.

To fix ideas, suppose $(y_{i,n}, x_{i,n}', w_{i,n}')$ are i.i.d. over i and define $V_{i,n} = x_{i,n} - \mathbb{E}[x_{i,n}|w_{i,n}]$ and $U_{i,n} = y_{i,n} - \mathbb{E}[y_{i,n}|x_{i,n}, w_{i,n}]$. It turns out that (under certain regularity conditions)

$$\hat{\Sigma}_n^{\mathtt{EW}} = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\left[\sum_{k=1}^n M_{ik,n}^2 M_{jk,n}^2\right]V_{i,n}V_{i,n}'\mathbb{E}[U_{j,n}^2|x_{j,n}, w_{j,n}] + o_p(1),$$

whereas a requirement for (2) to hold is that the estimator $\hat{\Sigma}_n$ satisfies

$$\hat{\Sigma}_n = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n M_{ij,n}^2 V_{i,n}V_{i,n}'\mathbb{E}[U_{j,n}^2|x_{j,n}, w_{j,n}] + o_p(1). \tag{3}$$

The difference between the leading terms in the expansions is non-negligible in general unless $K_n/n \to 0$. In recognition of this problem with $\hat{\Sigma}_n^{\texttt{EW}}$, we study the more general class of estimators of the form

$$\hat{\Sigma}_n(\kappa_n) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_{ij,n}\hat{v}_i\hat{v}_i'\hat{u}_j^2,$$

where $\kappa_{ij,n}$ denotes element (i,j) of a symmetric matrix $\kappa_n = \kappa_n(w_{1,n},\ldots,w_{n,n})$. Estimators that can be written in this fashion include $\hat{\Sigma}_n^{\texttt{EW}}$ (which corresponds to $\kappa_n = I_n$) as well as variants of the so-called HCk estimators, $k = 1,2,3,4$, discussed by MacKinnon (2012), among others.[1]

All of the HCk-type estimators (correspond to a diagonal choice of κ_n and) share with $\hat{\Sigma}_n^{\texttt{EW}}$ the shortcoming that they do not satisfy (3). On the other hand, it turns out that a certain non-diagonal choice of κ_n makes it possible to satisfy (3) even if K_n is a non-vanishing fraction of n. To be specific, it turns out that (under regularity conditions and) under mild conditions under the weights $\kappa_{ij,n}$, $\hat{\Sigma}_n(\kappa_n)$ satisfies

$$\hat{\Sigma}_n(\kappa_n) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\sum_{k=1}^{n}\sum_{l=1}^{n}\kappa_{kl,n}M_{ik,n}^2M_{jl,n}^2\right]V_{i,n}V_{i,n}'\mathbb{E}[U_{j,n}^2|x_{j,n},w_{j,n}] + o_p(1),$$

suggesting that (3) holds with $\hat{\Sigma}_n = \hat{\Sigma}_n(\kappa_n)$ provided κ_n is chosen in such a way that

$$M_{ij,n}^2 = \sum_{k=1}^{n}\sum_{l=1}^{n}\kappa_{kl,n}M_{ik,n}^2M_{jl,n}^2, \qquad 1 \le i,j \le n.$$

Accordingly, we define

$$\hat{\Sigma}_n^{\texttt{HC}} = \hat{\Sigma}_n(\kappa_n^{\texttt{HC}}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_{ij,n}^{\texttt{HC}}\hat{v}_{i,n}\hat{v}_{i,n}'\hat{u}_{j,n}^2,$$

where, with M_n denoting the matrix with element (i,j) given by $M_{ij,n}$ and $\odot$ denoting the Hadamard product,

$$\kappa_n^{\texttt{HC}} = \begin{pmatrix}\kappa_{11,n}^{\texttt{HC}} & \cdots & \kappa_{1n,n}^{\texttt{HC}}\\ \vdots & \ddots & \vdots\\ \kappa_{n1,n}^{\texttt{HC}} & \cdots & \kappa_{nn,n}^{\texttt{HC}}\end{pmatrix} = \begin{pmatrix}M_{11,n}^2 & \cdots & M_{1n,n}^2\\ \vdots & \ddots & \vdots\\ M_{n1,n}^2 & \cdots & M_{nn,n}^2\end{pmatrix}^{-1} = (M_n \odot M_n)^{-1}.$$

The estimator $\hat{\Sigma}_n^{\texttt{HC}}$ is well defined whenever $M_n \odot M_n$ is invertible, a simple sufficient condition for which is that $\min_{1\le i\le n} M_{ii,n} > 1/2$.[2] More importantly, a slight strengthening of this condition will be shown to be sufficient for (2) and (3) to hold with $\hat{\Sigma}_n = \hat{\Sigma}_n^{\texttt{HC}}$.

[1] To be specific, a natural variant of HCk is obtained by choosing κ_n to be diagonal with $\kappa_{ii,n} = d_n M_{ii,n}^{-\xi_{i,n}}$, where $(d_n,\xi_{i,n}) = (n/(n-K_n),0)$ for HC1, $(d_n,\xi_{i,n}) = (1,1)$ for HC2, $(d_n,\xi_{i,n}) = (1,2)$ for HC3, and $(d_n,\xi_{i,n}) = (1,\min(4,nM_{ii,n}/K_n))$ for HC4.

[2] This is a consequence of the Gershgorin circle theorem. For details, see the supplemental appendix.

3 Examples

The heuristics of the preceding section will be made precise in the next section. Before doing so, we present three leading examples, all of which are covered by the results developed in Section 4: (i) linear regression models with increasing dimension, (ii) semiparametric partially linear models, and (iii) fixed effects panel data regression models.

Let $\lambda_{\min}(A)$ denote the minimum eigenvalue of a square matrix A and let $\|\cdot\|$ denote the Euclidean norm.

3.1 Linear Regression Model with Increasing Dimension

The model of main interest is the linear regression model characterized by (1) and the following assumptions.

Assumption LR1 $\{(y_{i,n}, x'_{i,n}, w'_{i,n}) : 1 \leq i \leq n\}$ are i.i.d. over i and $\mathbb{E}[u_{i,n}|x_{i,n}, w_{i,n}] = 0$.

Assumption LR2 $\rho_n^{\mathtt{LR}} \to 0$, where

$$\rho_n^{\mathtt{LR}} = \min_{\delta \in \mathbb{R}^{K_n \times d}} \mathbb{E}[\|\mathbb{E}[x_{i,n}|w_{i,n}] - \delta' w_{i,n}\|^2].$$

Assumption LR3 $\mathbb{P}[\lambda_{\min}(\sum_{i=1}^n w_{i,n}w'_{i,n}) > 0] \to 1$ and $\mathcal{C}_n^{\mathtt{LR}} = O_p(1)$, where

$$\mathcal{C}_n^{\mathtt{LR}} = \max_{1\leq i\leq n} \left\{\mathbb{E}[u_{i,n}^4|x_{i,n}, w_{i,n}] + \mathbb{E}[\|V_{i,n}\|^4|w_{i,n}] + 1/\mathbb{E}[u_{i,n}^2|x_{i,n}, w_{i,n}] + 1/\lambda_{\min}(\mathbb{E}[V_{i,n}V'_{i,n}|w_{i,n}])\right\},$$

with $V_{i,n} = x_{i,n} - \mathbb{E}[x_{i,n}|w_{i,n}]$.

We shall consider this model in some detail because it is important in its own right and because the insights obtained for it can be used constructively in other cases, including the partially linear model (4) and the fixed effects panel data regression model (5) presented below. Linear regression models with (possibly) increasing dimension have a long tradition in Econometrics and Statistics (see, e.g., Koenker (1988)), and we consider them here as a theoretical device to obtain asymptotic approximations that better represent the finite-sample behavior of the statistics of interest.

The main difference between Assumptions LR1-LR3 and those familiar from the fixed-K_n case is the presence of the condition $\rho_n^{\mathtt{LR}} \to 0$ in Assumption LR2. The purpose of this condition is to ensure that the approximation

$$\mathbb{E}[x_{i,n}|w_{i,n}] \approx \delta_n' w_{i,n}, \qquad \delta_n = \mathbb{E}[w_{i,n}w'_{i,n}]^{-1}\mathbb{E}[w_{i,n}x'_{i,n}],$$

is sufficiently accurate to allow $v_{i,n} = x_{i,n} - \delta_n' w_{i,n}$ to be replaced by $V_{i,n} = x_{i,n} - \mathbb{E}[x_{i,n}|w_{i,n}]$ when analyzing $\hat{\beta}_n$ and the associated variance estimators.

It seems difficult to formulate primitive sufficient conditions for $\rho_n^{\mathtt{LR}} \to 0$ at the present level of generality, but three cases where the assumption is satisfied are the following. First, regardless of the dependence between $x_{i,n}$ and $w_{i,n}$, we have $\mathbb{E}[x_{i,n}|w_{i,n}] = \delta_n' w_{i,n}$ (implying $\rho_n^{\texttt{LR}} = 0$) when $w_{i,n}$ is discrete and satisfies the "saturation" condition that any function of $w_{i,n}$ can be written as a linear function of $w_{i,n}$. Second, $\rho_n^{\texttt{LR}} = 0$ when $w_{i,n}$ satisfies the exogeneity condition $\mathbb{E}[x_{i,n}|w_{i,n}] = 0$ (or when $\mathbb{E}[x_{i,n}|w_{i,n}]$ does not depend on $w_{i,n}$, and $w_{i,n}$ contains an intercept). A variant of this phenomenon occurs in the panel data model (5) discussed below. Third, one typically has $\rho_n^{\texttt{LR}} = O(K_n^{-\alpha})$ for some $\alpha > 0$ when the elements of $w_{i,n}$ are approximating functions, as in the partially linear model (4) discussed next.

3.2 Semiparametric Partially Linear Model

A related econometric model is the partially linear model

$$y_i = \beta' x_i + g(z_i) + \varepsilon_i, \qquad i = 1, \dots, n, \tag{4}$$

where x_i and z_i are explanatory variables, ε_i is an error term, and the function $g(z)$ is unknown. Suppose $\{p^k(z) : k = 1, 2, \cdots, K_n\}$ are functions having the property that linear combinations can approximate square-integrable functions of z well, in which case $g(z_i) \approx \gamma_n' p_n(z_i)$ for some γ_n, where $p_n(z) = (p^1(z), \dots, p^{K_n}(z))'$. Defining $y_{i,n} = y_i$, $x_{i,n} = x_i$, $w_{i,n} = p_n(z_i)$ and $u_{i,n} = \varepsilon_i + g(z_i) - \gamma_n' w_{i,n}$, the model (4) is of the form (1), and $\hat{\beta}_n$ is the series estimator of β previously studied by Donald and Newey (1994) and Cattaneo, Jansson, and Newey (2014). In this case, our analysis of $\hat{\beta}_n$ will proceed under the following assumptions.

Assumption PL1 $\{(y_i, x_i', z_i') : 1 \leq i \leq n\}$ are i.i.d. over i and $\mathbb{E}[\varepsilon_i|x_i, z_i] = 0$.

Assumption PL2 $\varrho_n^{\texttt{PL}} \to 0$, $\rho_n^{\texttt{PL}} \to 0$, and $n\varrho_n^{\texttt{PL}}\rho_n^{\texttt{PL}} \to 0$, where

$$\varrho_n^{\texttt{PL}} = \min_{\gamma \in \mathbb{R}^{K_n}} \mathbb{E}[|\mathbb{E}[y_i - \beta' x_i|x_i, z_i] - \gamma' p_n(z_i)|^2], \qquad \rho_n^{\texttt{PL}} = \min_{\delta \in \mathbb{R}^{K_n \times d}} \mathbb{E}[\|\mathbb{E}[x_i|z_i] - \delta' p_n(z_i)\|^2].$$

Assumption PL3 $\mathbb{P}[\lambda_{\min}(\sum_{i=1}^n p_n(z_i)p_n(z_i)') > 0] \to 1$ and $\mathcal{C}_n^{\texttt{PL}} = O_p(1)$, where

$$\mathcal{C}_n^{\texttt{PL}} = \max_{1 \leq i \leq n} \left\{ \mathbb{E}[\varepsilon_i^4|x_i, z_i] + \mathbb{E}[\|\nu_i\|^4|z_i] + 1/\mathbb{E}[\varepsilon_i^2|x_i, z_i] + 1/\lambda_{\min}(\mathbb{E}[\nu_i \nu_i'|z_i]) \right\},$$

with $\nu_i = x_i - \mathbb{E}[x_i|z_i]$.

In general, the partially linear model does not satisfy $\mathbb{E}[u_{i,n}|x_{i,n}, w_{i,n}] = 0$ because $g(z_i) \neq \gamma_n' p_n(z_i)$. To accommodate this failure a relaxation of Assumption LR1 is needed. The approach taken here, made precise in Assumption PL2, is motivated by the fact that linear combinations of $\{p^k(z)\}$ are assumed to be able to approximate the function $g(z)$ well. Under standard smoothness conditions, and for standard choices of basis functions, we have $\varrho_n^{\texttt{PL}} = O(K_n^{-\alpha_\varrho})$ and $\rho_n^{\texttt{PL}} = O(K_n^{-\alpha_\rho})$ for some pair $(\alpha_\varrho, \alpha_\rho)$ of positive constants, in which case Assumption PL2 holds provided $K_n^{\alpha_\varrho + \alpha_\rho}/n \to \infty$. For further technical details see, for example, Newey (1997) and Belloni, Chernozhukov, Chetverikov, and Kato (2014).

3.3 Fixed Effects Panel Data Regression Model

Stock and Watson (2008) consider heteroskedasticity-robust inference for the panel data regression model

$$Y_{it} = \alpha_i + \beta' X_{it} + U_{it}, \qquad i = 1, \ldots, N, \qquad t = 1, \ldots, T, \tag{5}$$

where $T \geq 3$ is fixed, $\alpha_i \in \mathbb{R}$ is an individual-specific intercept, $X_{it} \in \mathbb{R}^d$ is a regressor of dimension d, $U_{it} \in \mathbb{R}$ is an error term, and the following assumptions are satisfied.

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

Assumption FE2 $\mathbb{E}[U_{it}|X_{i1} \ldots, X_{iT}] = 0$ and $\mathbb{E}[\tilde{X}_{it}] = 0$, where $\tilde{X}_{it} = X_{it} - T^{-1} \sum_{s=1}^{T} X_{is}$.

Assumption FE3 $\mathcal{C}_N^{\mathtt{FE}} = O_p(1)$, where

$$\begin{aligned} \mathcal{C}_N^{\mathtt{FE}} &= \max_{1 \leq i \leq N, 1 \leq t \leq T} \left\{ \mathbb{E}[U_{it}^4|X_{i1} \ldots, X_{iT}] + \mathbb{E}[\|X_{it}\|^4] \right\} \\ &\quad + \max_{1 \leq i \leq N, 1 \leq t \leq T} \left\{ 1/\mathbb{E}[U_{it}^2|X_{i1} \ldots, X_{iT}] + 1/\lambda_{\min}\left(\mathbb{E}[\textstyle\sum_{s=1}^{T} \tilde{X}_{is}\tilde{X}'_{is}/T] \right) \right\}. \end{aligned}$$

The model (5) is also of the form (1), and $\hat{\beta}_n$ is the fixed effects estimator of β. To see this, define $n = NT$, $K_n = N$, $\gamma_n = (\alpha_1, \ldots, \alpha_N)'$, and

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

where $e_{i,N} \in \mathbb{R}^N$ is the i-th unit vector of dimension N. 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 $w_{i,n}$.

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

4 Results

The three models presented in the previous section are non-nested, but may be treated in a unified way by embedding them in a general framework. That general framework, which accommodates our motivating examples as well as others, is presented next.

4.1 General Framework

Suppose $\{(y_{i,n}, x'_{i,n}, w'_{i,n}) : 1 \leq i \leq n\}$ is generated by (1). Let $\{\mathcal{T}_{i,n} : 1 \leq i \leq N_n\}$ be a partition of $\{1, \ldots, n\}$ and define $\mathcal{C}_n^{\mathcal{T}} = \max_{1 \leq i \leq N_n}(\#\mathcal{T}_{i,n})$, where $\#\mathcal{T}_{i,n}$ is the cardinality of $\mathcal{T}_{i,n}$. Let $\mathcal{X}_n = (x_{1,n}, \ldots, x_{n,n})$ and for a set $\mathcal{W}_n$ of random variables satisfying $\mathbb{E}[w_{i,n}|\mathcal{W}_n] = w_{i,n}$, define the constants

$$\begin{aligned}
\varrho_n &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[|R_{i,n}^u|^2], & R_{i,n}^u &= \mathbb{E}[u_{i,n}|\mathcal{X}_n, \mathcal{W}_n], \\
\bar{\varrho}_n &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[|Q_{i,n}^u|^2], & Q_{i,n}^u &= \mathbb{E}[u_{i,n}|\mathcal{W}_n], \\
\rho_n &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\|R_{i,n}^v\|^2], & R_{i,n}^v &= \mathbb{E}[v_{i,n}|\mathcal{W}_n],
\end{aligned}$$

where $v_{i,n} = x_{i,n} - \delta'_n w_{i,n}$ with $\delta_n = (\sum_{i=1}^n \mathbb{E}[w_{i,n}w'_{i,n}])^{-1}\sum_{i=1}^n \mathbb{E}[w_{i,n}x'_{i,n}]$. Also define

$$\overline{\mathcal{C}}_n = \max_{1\leq i\leq n} \mathbb{E}[U_{i,n}^4|\mathcal{X}_n, \mathcal{W}_n] + \max_{1\leq i\leq n} \mathbb{E}[\|V_{i,n}\|^4|\mathcal{W}_n], \quad \underline{\mathcal{C}}_n = \min_{1\leq i\leq n} \mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n],$$

where $U_{i,n} = u_{i,n} - R_{i,n}^u = y_{i,n} - \mathbb{E}[y_{i,n}|\mathcal{X}_n, \mathcal{W}_n]$ and $V_{i,n} = v_{i,n} - R_{i,n}^v = x_{i,n} - \mathbb{E}[x_{i,n}|\mathcal{W}_n]$.

In the supplemental appendix we show how the three examples fit in this general framework and verify that Assumptions LR1–LR3, PL1–PL3 and FE1–FE3, respectively, imply the following three assumptions.

Assumption 1 *$\{(U_{t,n}, V_{t,n}) : t \in \mathcal{T}_{i,n}\}$ are independent over i conditional on $\mathcal{W}_n$, $\mathbb{C}[U_{i,n}, U_{j,n}|\mathcal{X}_n, \mathcal{W}_n] = 0$ for $i \neq j$, and $\mathcal{C}_n^{\mathcal{T}} = O(1)$.*

Assumption 2 *$\varrho_n + n(\varrho_n - \bar{\varrho}_n) \to 0$, $\rho_n \to 0$, and $n\varrho_n\rho_n \to 0$.*

Assumption 3 *$\mathbb{P}[\lambda_{\min}(\sum_{i=1}^n w_{i,n}w'_{i,n}) > 0] \to 1$, $\overline{\mathcal{C}}_n = O_p(1)$, and $1/\underline{\mathcal{C}}_n = O_p(1)$.*

4.2 General Results

As a means to the end of establishing (2), we give an asymptotic normality result for $\hat{\beta}_n$ which may be of interest in its own right. Let $\tilde{\Gamma}_n = \sum_{i=1}^n \tilde{V}_{i,n}\tilde{V}'_{i,n}/n$, where $\tilde{V}_{i,n} = \sum_{j=1}^n M_{ij,n}V_{j,n}$.

Theorem 1 *Suppose Assumptions 1–3 hold and suppose* $\tilde{\Gamma}_n^{-1} = O_p(1)$. *Then*

$$\Omega_n^{-1/2}\sqrt{n}(\hat{\beta}_n - \beta) \to_d \mathcal{N}(0, I_d), \qquad \Omega_n = \Gamma_n^{-1}\Sigma_n\Gamma_n^{-1}, \tag{6}$$

where $\Gamma_n = \mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n]$ *and* $\Sigma_n = \sum_{i=1}^n \tilde{V}_{i,n}\tilde{V}_{i,n}'\mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n]/n$.

Achieving (2), the counterpart of (6) in which the unknown matrices (Γ_n, Σ_n) are replaced by the estimators $(\hat{\Gamma}_n, \hat{\Sigma}_n)$, requires additional assumptions. One possibility is to impose homoskedasticity.

Theorem 2 *Suppose the assumptions of Theorem 1 hold. If* $\mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n] = \sigma_n^2$ *and if* $\overline{\lim}_{n\to\infty} K_n/n < 1$, *then (2) holds with* $\hat{\Sigma}_n = \hat{\Sigma}_n^{\mathtt{HO}}$.

This result shows in quite some generality that homoskedastic inference in linear models remains valid even when K_n is proportional to n, provided the variance estimator incorporates a degrees-of-freedom correction, as $\hat{\Sigma}_n^{\mathtt{HO}}$ does.

Establishing (2) is also possible when K_n is assumed to be a vanishing fraction of n, as is of course the case in the usual fixed-K_n linear regression model setup. The following theorem establishes consistency of the conventional standard error estimator $\hat{\Sigma}_n^{\mathtt{EW}}$ under the assumption $K_n/n \to 0$, and also derives an asymptotic representation for estimators of the form $\hat{\Sigma}_n(\kappa_n)$ without imposing this assumption. Let $\|\kappa_n\|_\infty = \max_{1\le i\le n}\sum_{j=1}^n |\kappa_{ij,n}|$.

Theorem 3 *Suppose the assumptions of Theorem 1 hold.*
(a) If $K_n/n \to 0$, *then (2) holds with* $\hat{\Sigma}_n = \hat{\Sigma}_n^{\mathtt{EW}}$.
(b) If $\|\kappa_n\|_\infty = O_p(1)$, *then*

$$\hat{\Sigma}_n(\kappa_n) = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\left[\sum_{k=1}^n \kappa_{ik,n}M_{jk,n}^2\right]\tilde{V}_{i,n}\tilde{V}_{i,n}'\mathbb{E}[U_{j,n}^2|\mathcal{X}_n, \mathcal{W}_n] + o_p(1).$$

The conclusion of part (a) typically fails when the condition $K_n/n \to 0$ is dropped. For example, when specialized to $\kappa_n = I_n$ part (b) implies that in the homoskedastic case (i.e., when the assumptions of Theorem 2 are satisfied)

$$\hat{\Sigma}_n^{\mathtt{EW}} = \Sigma_n - \frac{\sigma_n^2}{n}\sum_{i=1}^n(1 - M_{ii,n})\tilde{V}_{i,n}\tilde{V}_{i,n}' + o_p(1),$$

where $\sum_{i=1}^n(1 - M_{ii,n})\tilde{V}_{i,n}\tilde{V}_{i,n}'/n \neq o_p(1)$ in general (unless $K_n/n \to 0$).[3] On the other hand, because $\sum_{1\le k\le n}\kappa_{ik,n}^{\mathtt{HC}}M_{jk,n}^2 = \mathbb{1}(i = j)$ by construction, part (b) implies that $\hat{\Sigma}_n^{\mathtt{HC}}$ is consistent provided $\|\kappa_n^{\mathtt{HC}}\|_\infty = O_p(1)$. To state a condition for this to occur, let

$$\mathcal{C}_n^M = \mathbb{1}\{\lambda_{\min}(\sum_{i=1}^n w_{i,n}w_{i,n}') > 0\}\min_{1\le i\le n} M_{ii,n}.$$

[3] Similar remarks apply to the variants of the HCk estimators mentioned above.

If $\mathcal{C}_n^M > 1/2$, then $\kappa_n^{\texttt{HC}}$ is diagonally dominant and it follows from Theorem 1 of Varah (1975) that

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

As a consequence, we obtain the following theorem, whose conditions can hold even if $K_n/n \nrightarrow 0$.

Theorem 4 *Suppose the assumptions of Theorem 1 hold.*
If $\mathbb{P}[\mathcal{C}_n^M > 1/2] \to 1$ *and if* $1/(\mathcal{C}_n^M - 1/2) = O_p(1)$, *then (2) holds with* $\hat{\Sigma}_n = \hat{\Sigma}_n^{\texttt{HC}}$.

Because $\sum_{i=1}^n M_{ii,n} = n - K_n$ when $\lambda_{\min}(\sum_{i=1}^n w_{i,n} w_{i,n}') > 0$, we have $\mathcal{C}_n^M \leq 1 - K_n/n$ and a necessary condition for Theorem 4 to be applicable is therefore that $\overline{\lim}_{n\to\infty} K_n/n < 1/2$. When the design is balanced, that is, when $M_{11,n} = \ldots = M_{nn,n}$ (as occurs in the panel data model (5)), the condition $\overline{\lim}_{n\to\infty} K_n/n < 1/2$ is also sufficient, but in general it seems difficult to formulate primitive sufficient conditions for the assumption made about $\mathcal{C}_n^M$ in Theorem 4. In practice, the fact that $\mathcal{C}_n^M$ is observed means that the condition $\mathcal{C}_n^M > 1/2$ is verifiable, and therefore unless $\mathcal{C}_n^M$ is found to be "close" to $1/2$ there is reason to expect $\hat{\Sigma}_n^{\texttt{HC}}$ to perform well.

4.3 Examples

4.3.1 Linear Regression Model with Increasing Dimension

Specializing Theorems 2-4 to the linear regression model, we obtain the following result.

Theorem LR Suppose Assumptions LR1–LR3 hold and suppose $\overline{\lim}_{n\to\infty} K_n/n < 1$.

(a) If $\mathbb{E}[u_{i,n}^2 | x_{i,n}, z_{i,n}] = \sigma_n^2$, then (2) holds with $\hat{\Sigma}_n = \hat{\Sigma}_n^{\texttt{HO}}$.

(b) If $K_n/n \to 0$, then (2) holds with $\hat{\Sigma}_n = \hat{\Sigma}_n^{\texttt{EW}}$.

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

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

Remark. Our main results for linear models concern large-sample approximations for the finite-sample distribution of the usual t-statistics. An alternative, equally automatic approach is to employ the bootstrap and closely related resampling procedures (see, among others, Freedman (1981), Mammen (1993), Gonçalvez and White (2005), Kline and Santos (2012)). Assuming $K_n/n \nrightarrow 0$, Bickel and Freedman (1983) demonstrated an invalidity result for the bootstrap. We conjecture that similar results can be obtained for other resampling procedures, but it is beyond the scope of this paper to do so.

4.3.2 Semiparametric Partially Linear Model

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

Theorem PL Suppose Assumptions PL1–PL3 hold and suppose $\overline{\lim}_{n\to\infty} K_n/n < 1$.

(a) If $\mathbb{E}[\varepsilon_i^2|x_i, z_i] = \sigma^2$, then (2) holds with $\hat{\Sigma}_n = \hat{\Sigma}_n^{\mathtt{HO}}$.

(b) If $K_n/n \to 0$, then (2) holds with $\hat{\Sigma}_n = \hat{\Sigma}_n^{\mathtt{EW}}$.

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

- DISCUSS CONNECTIONS TO Cattaneo, Jansson, and Newey (2014) here.

4.3.3 Fixed Effects Panel Data Regression Model

Finally, consider the panel data model (5). Because $K_n/n = 1/T$ is fixed this model does not admit an analog of Theorem 3. On the other hand, it does admit an analog of Theorems 2 and 4.

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

To see the connection between our results and those in Stock and Watson (2008), observe that $M_n = I_N \otimes [I_T - \iota_T\iota_T'/T]$ for $\iota_T \in \mathbb{R}^T$ a $T \times 1$ vector of ones. We then obtain $M_{ii,n} = 1 - 1/T$ (for $i = 1,\ldots,n$) and therefore $\mathcal{C}_n^M \geq 2/3$ because $T \geq 3$. More importantly, perhaps, we obtain a closed-form expression for $\kappa_n^{\mathtt{HC}}$ given by

$$\kappa_n^{\mathtt{HC}} = I_N \otimes \frac{T}{T-2}\left[I_T - \frac{1}{(T-1)^2}\iota_T\iota_T'\right].$$

As a consequence,

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

where $\hat{U}_{it} = Y_{it} - T^{-1}\sum_{1\leq s\leq T} Y_{is} - \hat{\beta}_n'\tilde{X}_{it}$. 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).

Remark. The result above not only highlights a tight connection between our general standard error estimator and the one in Stock and Watson (2008), but also suggests that our general formula $\hat{\Sigma}_n^{\mathtt{HC}}$ could be used to derive explicit, simple expressions in other contexts where multi-way fixed effects or similar discrete regressors are included.

5 Simulations

We report the results from a small Monte Carlo experiment aimed to capture the extent to which our main theoretical findings are present in samples of moderate size. To facilate comparability with other studies, we employ a data generating process (DGP) that is as similar as possible to those employed in the literature before. In particular, we consider the following model:

$$\begin{aligned} y_i &= \beta x_i + \gamma' w_i + u_i, & u_i|(x_i, w_i) &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_{ui}^2), & \sigma_{ui}^2 &= \varkappa_u(1 + (x_i + \iota' w_i)^2)^{\vartheta}, \\ x_i &= v_i, & v_i|w_i &\sim \text{i.i.d. } \mathcal{N}(0, \sigma_{vi}^2), & \sigma_{vi}^2 &= \varkappa_v(1 + (\iota' w_i)^2)^{\vartheta}, \end{aligned}$$

where $\iota = (1, 1, \cdots, 1)'$, $\beta = 0$ and $\gamma = 0$, and the constants $\varkappa_u$ and $\varkappa_v$ are chosen so that $\mathbb{V}[u_i] = \mathbb{V}[v_i] = 1$. In the absence of the additional covariates w_i, this design coincides with the one in Stock and Watson (2008), and is very similar to the one considered in MacKinnon (2012).

The simulation study employs $5,000$ replications, sets the sample size to $n = 500$, and consideres models with $K_n/n \in \{0.1, 0.2, 0.3, 0.4\}$. The two main parameters varying in the Monte Carlo experiments are: the constant ϑ and the distribution of the covariates w_i. The first parameter controls the degree of heteroskedasticity: $\vartheta = 0$ corresponds to homoskedasticity, $\vartheta = 1$ corresponds to moderate heteroskedasticity, and $\vartheta = 3/2$ corresponds to pronounced heteroskedasticity. For the distribution of the covariates we consider the following cases: independent standard $\mathcal{N}(0, 1)$ (Model 1), independent $U(-1, 1)$ (Model 2), independent discrete covariates constructed as $\mathbf{1}(\mathcal{N}(0, 1) \geq 2.33)$.

The results are given in Table 2.

6 Conclusion

We investigate the properties of several popular heteroskedasticity-robust standard error estimators in linear regression models with many nuisance covariates, and showed that none of the usual formulas deliver consistent standard errors when the number of covariates is not a vanishing proportion of the sample size. We also proposed a new standard error formula that is consistent under (conditional) heteroskedasticity and many covariates, which is fully automatic and does not assume any special structure on the regressors.

Our results concern high-dimensional models where the number of covariates is at most a non-vanishing fraction of the sample size. A quite recent related literature concerns ultra-high-dimensional models where the number of covariates is much larger than the sample size, but some form of (approximate) sparsity is imposed in the model (e.g., Belloni, Chernozhukov, and Hansen (2014b), Belloni, Chernozhukov, and Hansen (2014a), Belloni, Chernozhukov, Hansen, and Fernandez-Val (2014) and Farrell (2014)). In that setting, inference is conducted after covariate selection, where the resulting number of selected covariates is at most a vanishing fraction of the sample size (usually much smaller). Thus, it would be of interest to investigate whether the methods proposed herein can be applied also for inference post covariate selection in ultra-high-dimensional settings, which would allow for weaker forms of sparsity because more covariates could be selected for inference.

References

Belloni, A., V. Chernozhukov, D. Chetverikov, and K. Kato (2014): "On the Asymptotic Theory for Least Squares Series: Pointwise and Uniform Results," *Journal of Econometrics*, forthcoming.

Belloni, A., V. Chernozhukov, and C. Hansen (2014a): "High-Dimensional Methods and Inference on Structural and Treatment Effects," *Journal of Economic Perspectives*, 28(2), 29–50.

——— (2014b): "Inference on Treatment Effects after Selection among High-Dimensional Controls," *Review of Economic Studies*, 81(2), 608–650.

Belloni, A., V. Chernozhukov, C. Hansen, and I. Fernandez-Val (2014): "Program Evaluation with High-Dimensional Data," Working paper, MIT.

Bera, A. K., T. Suprayitno, and G. Premaratne (2002): "On Some Heteroskedasticity-robust Estimators of Variance-Covariance Matrix of the Least-squares Estimators," *Journal of Statistical Planning and Inference*, 108, 121–136.

Bickel, P. J., and D. A. Freedman (1983): "Bootstrapping Regression Models with Many Parameters," in *A Festschrift for Erich L. Lehmann*, ed. by P. Bickel, K. Doksum, and J. Hodges. Chapman and Hall.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2014): "Alternative Asymptotics and the Partially Linear Model with Many Regressors," Working paper, University of Michigan.

Chesher, A. (1989): "Hájek Inequalities, Measures of Leverage and the Size of Heteroskedasticity Robust Wald Tests," *Econometrica*, 57(4), 971–977.

Chesher, A., and I. Jewitt (1987): "The Bias of a heteroskedasticity consistent covariance matrix estimator," *Econometrica*, 55(5), 1217–1222.

Cribari-Neto, F., and M. da Gloria A. Lima (2011): "A sequence of improved standard errors under heteroskedasticity of unknown form," *Journal of Statistical Planning and Inference*, 141(11), 3617–3627.

Cribari-Neto, F., S. L. P. Ferrari, and G. M. Cordeiro (2000): "Improved Heteroscedasticity-Consistent Covariance Matrix Estimators," *Biometrika*, 87(4), 907–918.

Donald, S. G., and W. K. Newey (1994): "Series Estimation of Semilinear Models," *Journal of Multivariate Analysis*, 50(1), 30–40.

Farrell, M. H. (2014): "Robust Inference on Average Treatment Effects with Possibly More Covariates than Observations," working paper, University of Chicago.

Freedman, D. A. (1981): "Bootstrapping Regression Models," *Annals of Statistics*, 9(6), 1218–1228.

Gonçalvez, S., and H. White (2005): "Bootstrap Standard Error Estimates for Linear Regression," *Journal of the American Statistical Association*, 100(471), 970–979.

Kline, P., and A. Santos (2012): "Higher order properties of the wild bootstrap under misspecification," *Journal of Econometrics*, 171(1), 54–70.

Koenker, R. (1988): "Asymptotic Theory and Econometric Practice," *Journal of Applied Econometrics*, 3(2), 139–147.

MacKinnon, J., and H. White (1985): "Some Heteroskedasticity-consistent Covariance Matrix Estimators with Improved Finite Sample Properties," *Journal of Econometrics*, 29, 305–325.

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.

Mammen, E. (1993): "Bootstrap and Wild Bootstrap for High Dimensional Linear Models," *Annals of Statistics*, 21(1), 255–285.

Müller, U. K. (2013): "Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix," *Econometrica*, 81(5), 1805–1849.

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

Shao, J., and C. F. J. Wu (1987): "Heteroscedasticity-Robustness of Jackknife Variance Estimators in Linear Models," *Annals of Statistics*, 15(4), 1563–1579.

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.

White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48(4), 817–838.

Wu, C. F. J. (1986): "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *Annals of Statistics*, 14(4), 1261–1295.

Table 1: Empirical Coverage of 95% Confidence Intervals ($n = 500$).

(a) Model 1: Gaussian $w_{i,n}$ Regressors

ϑ	K_n/n	HO_0	HO_1	HC_0	HC_1	HC_2	HC_3	HC_4	HC_K
0	0.1	0.933	0.947	0.929	0.943	0.943	0.954	0.975	0.943
0	0.2	0.915	0.946	0.911	0.945	0.945	0.968	0.993	0.945
0	0.3	0.894	0.948	0.892	0.946	0.946	0.978	0.985	0.943
0	0.4	0.870	0.947	0.869	0.948	0.947	0.987	0.971	0.945
1	0.1	0.431	0.451	0.886	0.906	0.908	0.925	0.956	0.922
1	0.2	0.443	0.486	0.829	0.882	0.884	0.921	0.976	0.915
1	0.3	0.432	0.509	0.774	0.859	0.859	0.923	0.938	0.909
1	0.4	0.454	0.561	0.723	0.843	0.843	0.931	0.889	0.908

(b) Model 2: Uniform $w_{i,n}$ Regressors

ϑ	K_n/n	HO_0	HO_1	HC_0	HC_1	HC_2	HC_3	HC_4	HC_K
0	0.1	0.939	0.951	0.934	0.949	0.949	0.961	0.977	0.948
0	0.2	0.912	0.950	0.913	0.948	0.947	0.970	0.992	0.946
0	0.3	0.892	0.948	0.888	0.949	0.949	0.978	0.984	0.946
0	0.4	0.857	0.949	0.856	0.947	0.946	0.989	0.973	0.943
1	0.1	0.375	0.394	0.885	0.907	0.907	0.926	0.956	0.924
1	0.2	0.410	0.452	0.839	0.889	0.889	0.930	0.977	0.922
1	0.3	0.427	0.497	0.775	0.865	0.865	0.930	0.948	0.919
1	0.4	0.429	0.538	0.731	0.850	0.850	0.934	0.896	0.915

(c) Model 3: Discrete $w_{i,n}$ Regressors

ϑ	K_n/n	HO_0	HO_1	HC_0	HC_1	HC_2	HC_3	HC_4	HC_K
0	0.1	0.936	0.951	0.935	0.948	0.947	0.958	0.979	0.945
0	0.2	0.922	0.952	0.924	0.952	0.949	0.971	0.991	0.949
0	0.3	0.905	0.955	0.910	0.956	0.953	0.979	0.982	0.951
0	0.4	0.880	0.952	0.885	0.955	0.950	0.986	0.968	0.948
1	0.1	0.298	0.313	0.747	0.778	0.853	0.930	0.991	0.914
1	0.2	0.390	0.432	0.701	0.764	0.847	0.949	0.988	0.933
1	0.3	0.480	0.553	0.683	0.773	0.848	0.962	0.929	0.929
1	0.4	0.531	0.658	0.654	0.783	0.847	0.976	0.858	0.936

Notes: (i) .

Table 2: Empirical Coverage of 95% Confidence Intervals ($n = 1000$).

(a) Model 1: Gaussian $w_{i,n}$ Regressors

ϑ	K_n/n	HO_0	HO_1	HC_0	HC_1	HC_2	HC_3	HC_4	HC_K
0	0.1	0.939	0.952	0.940	0.950	0.950	0.963	0.977	0.950
0	0.2	0.922	0.952	0.920	0.952	0.951	0.969	0.993	0.950
0	0.3	0.899	0.948	0.897	0.949	0.949	0.981	0.987	0.947
0	0.4	0.867	0.952	0.866	0.951	0.951	0.987	0.974	0.950
1	0.1	0.421	0.442	0.899	0.917	0.918	0.932	0.962	0.929
1	0.2	0.436	0.479	0.852	0.896	0.896	0.934	0.976	0.929
1	0.3	0.446	0.516	0.809	0.878	0.881	0.941	0.955	0.928
1	0.4	0.442	0.554	0.742	0.858	0.858	0.937	0.902	0.922

(b) Model 2: Uniform $w_{i,n}$ Regressors

ϑ	K_n/n	HO_0	HO_1	HC_0	HC_1	HC_2	HC_3	HC_4	HC_K
0	0.1	0.937	0.950	0.938	0.950	0.950	0.962	0.980	0.950
0	0.2	0.930	0.959	0.929	0.960	0.960	0.975	0.993	0.959
0	0.3	0.905	0.955	0.904	0.953	0.952	0.982	0.988	0.953
0	0.4	0.872	0.951	0.870	0.952	0.951	0.989	0.973	0.950
1	0.1	0.387	0.406	0.901	0.922	0.922	0.939	0.964	0.936
1	0.2	0.420	0.463	0.862	0.905	0.906	0.939	0.981	0.932
1	0.3	0.427	0.499	0.807	0.886	0.885	0.943	0.959	0.931
1	0.4	0.419	0.521	0.736	0.853	0.853	0.942	0.908	0.927

(c) Model 3: Discrete $w_{i,n}$ Regressors

ϑ	K_n/n	HO_0	HO_1	HC_0	HC_1	HC_2	HC_3	HC_4	HC_K
0	0.1	0.936	0.949	0.935	0.948	0.947	0.960	0.976	0.946
0	0.2	0.919	0.948	0.920	0.948	0.947	0.967	0.993	0.946
0	0.3	0.896	0.945	0.897	0.947	0.946	0.978	0.982	0.945
0	0.4	0.868	0.945	0.871	0.947	0.943	0.988	0.971	0.943
1	0.1	0.346	0.366	0.834	0.861	0.900	0.949	0.991	0.942
1	0.2	0.516	0.569	0.802	0.856	0.893	0.957	0.992	0.940
1	0.3	0.616	0.703	0.777	0.858	0.892	0.970	0.964	0.943
1	0.4	0.670	0.790	0.751	0.867	0.899	0.982	0.927	0.950

Notes: (i) .

Heteroskedasticity Consistent Standard Errors for Linear Models with Many Covariates: Supplemental Appendix

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

January 6, 2015

Abstract

This supplemental appendix contains proofs of the results reported in the paper.

Contents

[*]Department of Economics, University of Michigan.
[†]Department of Economics, UC Berkeley and *CREATES*.
[‡]Department of Economics, MIT.

1 Setup

The model is

$$y_{i,n} = \beta' x_{i,n} + \gamma_n' w_{i,n} + u_{i,n}, \qquad i = 1, \ldots, n, \tag{SA-1}$$

where $y_{i,n} \in \mathbb{R}$, $x_{i,n} \in \mathbb{R}^d$, $w_{i,n} \in \mathbb{R}^{K_n}$, and $u_{i,n} \in \mathbb{R}$.

Let $\{\mathcal{T}_{i,n} : 1 \leq i \leq N_n\}$ be a partition of $\{1, \ldots, n\}$ and define $\mathcal{C}_n^{\mathcal{T}} = \max_{1 \leq i \leq N_n}(\#\mathcal{T}_{i,n})$, where $\#\mathcal{T}_{i,n}$ is the cardinality of $\mathcal{T}_{i,n}$. Let $\mathcal{X}_n = (x_{1,n}, \ldots, x_{n,n})$ and for a set $\mathcal{W}_n$ of random variables satisfying $\mathbb{E}[w_{i,n}|\mathcal{W}_n] = w_{i,n}$, define the constants

$$\begin{aligned}
\varrho_n &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[|R_{i,n}^u|^2], &\qquad R_{i,n}^u &= \mathbb{E}[u_{i,n}|\mathcal{X}_n, \mathcal{W}_n],\\
\bar{\varrho}_n &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[|Q_{i,n}^u|^2], &\qquad Q_{i,n}^u &= \mathbb{E}[u_{i,n}|\mathcal{W}_n],\\
\rho_n &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[\|R_{i,n}^v\|^2], &\qquad R_{i,n}^v &= \mathbb{E}[v_{i,n}|\mathcal{W}_n],
\end{aligned}$$

where $v_{i,n} = x_{i,n} - \delta_n' w_{i,n}$ with $\delta_n = (\sum_{i=1}^n \mathbb{E}[w_{i,n}w_{i,n}'])^{-1} \sum_{i=1}^n \mathbb{E}[w_{i,n}x_{i,n}']$.

We impose the following three assumptions, where

$$\overline{\mathcal{C}}_{U,n} = 1 + \max_{1\leq i\leq n} \mathbb{E}[U_{i,n}^4|\mathcal{X}_n, \mathcal{W}_n], \quad \overline{\mathcal{C}}_{V,n} = 1 + \max_{1\leq i\leq n} \mathbb{E}[\|V_{i,n}\|^4|\mathcal{W}_n], \quad \underline{\mathcal{C}}_{U,n} = \min_{1\leq i\leq n} \mathbb{E}[U_{i,n}^2|\mathcal{X}_n, \mathcal{W}_n],$$

with $U_{i,n} = u_{i,n} - R_{i,n}^u = y_{i,n} - \mathbb{E}[y_{i,n}|\mathcal{X}_n, \mathcal{W}_n]$ and $V_{i,n} = v_{i,n} - R_{i,n}^v = x_{i,n} - \mathbb{E}[x_{i,n}|\mathcal{W}_n]$.

Assumption 1 $\{(U_{t,n}, V_{t,n}) : t \in \mathcal{T}_{i,n}\}$ *are independent over* i *conditional on* $\mathcal{W}_n$*,* $\mathbb{C}[U_{i,n}, U_{j,n}|\mathcal{X}_n, \mathcal{W}_n] = 0$ *for* $i \neq j$*, and* $\mathcal{C}_n^{\mathcal{T}} = O(1)$.

Assumption 2 $\varrho_n + n(\varrho_n - \bar{\varrho}_n) \to 0$, $\rho_n \to 0$, *and* $n\varrho_n\rho_n \to 0$.

Assumption 3 $\mathbb{P}[\lambda_{\min}(\sum_{i=1}^n w_{i,n}w_{i,n}') > 0] \to 1$, $\overline{\mathcal{C}}_{U,n} + \overline{\mathcal{C}}_{V,n} = O_p(1)$*, and* $1/\underline{\mathcal{C}}_{U,n} = O_p(1)$.

2 Useful Lemmas

Our main results are based on six lemmas and are obtained by working with the representation

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

where $\hat{\Gamma}_n = \sum_{1\leq i\leq n} \hat{v}_{i,n}\hat{v}_{i,n}'/n$ and $S_n = \sum_{1\leq i\leq n} \hat{v}_{i,n}u_{i,n}/\sqrt{n}$. and $\hat{v}_{i,n} = \sum_{1\leq j\leq n} M_{ij,n}x_{j,n}$. Strictly speaking, the displayed representation is valid only when $\lambda_{\min}(\sum_{i=1}^n w_{i,n}w_{i,n}') > 0$ and $\lambda_{\min}(\hat{\Gamma}_n) > 0$. 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 $\omega_n = \mathbb{1}\{\lambda_{\min}(\sum_{k=1}^{n} w_{k,n}w_{k,n}') > 0\}$, and, in a slight abuse of notation, we define

$$\hat{\Gamma}_n = \frac{1}{n}\sum_{1\leq i\leq n}\hat{v}_{i,n}\hat{v}_{i,n}', \qquad S_n = \frac{1}{\sqrt{n}}\sum_{1\leq i\leq n}\hat{v}_{i,n}u_{i,n}, \qquad \hat{v}_{i,n} = \sum_{1\leq j\leq n}\bar{M}_{ij,n}x_{j,n},$$

and

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

The first lemma can be used to approximate $\hat{\Gamma}_n$. Let $\tilde{V}_{i,n} = \sum_{1\leq j\leq n}\bar{M}_{ij,n}V_{j,n}$.

Lemma SA-1 *Suppose Assumptions 1 and 3 hold. If $\rho_n \to 0$, then*

$$\hat{\Gamma}_n = \tilde{\Gamma}_n + o_p(1) = \mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n] + o_p(1), \qquad \tilde{\Gamma}_n = \frac{1}{n}\sum_{1\leq i\leq n}\tilde{V}_{i,n}\tilde{V}_{i,n}'.$$

The second lemma can be used to show asymptotic normality of S_n. This result employs the Berry-Esseen inequality, the Cramér-Wold device, and some simple bounding arguments.

Lemma SA-2 *Suppose Assumptions 1–3 hold. If $\tilde{\Gamma}_n^{-1} = O_p(1)$, then $\mathbb{V}[\tilde{S}_n|\mathcal{X}_n,\mathcal{W}_n]^{-1} = O_p(1)$ and*

$$\mathbb{V}[\tilde{S}_n|\mathcal{X}_n,\mathcal{W}_n]^{-1/2}S_n \to_d \mathcal{N}(0, I_d), \qquad \tilde{S}_n = \frac{1}{\sqrt{n}}\sum_{1\leq i\leq n}\tilde{V}_{i,n}U_{i,n}.$$

The third lemma can be used to approximate

$$\hat{\sigma}_n^2 = \frac{1}{n-d-K_n}\sum_{1\leq i\leq n}\hat{u}_{i,n}^2, \qquad \hat{u}_{i,n} = \sum_{1\leq j\leq n}\bar{M}_{ij,n}(y_{j,n} - \hat{\beta}_n' x_{j,n}).$$

Let $\tilde{U}_{i,n} = \sum_{1\leq j\leq n}\bar{M}_{ij,n}U_{j,n}$.

Lemma SA-3 *Suppose Assumptions 1–3 hold. If $\tilde{\Gamma}_n^{-1} = O_p(1)$ and if $\overline{\lim}_{n\to\infty}K_n/n < 1$, then*

$$\hat{\sigma}_n^2 = \tilde{\sigma}_n^2 + o_p(1) = \mathbb{E}[\tilde{\sigma}_n^2|\mathcal{X}_n,\mathcal{W}_n] + o_p(1), \qquad \tilde{\sigma}_n^2 = \frac{1}{n-K_n}\sum_{1\leq i\leq n}\tilde{U}_{i,n}^2.$$

The fourth lemma can be used to approximate

$$\hat{\Sigma}_n(\kappa_n) = \frac{1}{n}\sum_{1\leq i,j\leq n}\kappa_{ij,n}\hat{v}_i\hat{v}_i'\hat{u}_j^2.$$

Let $\mathcal{C}_n^\kappa = \max_{1\leq i\leq n}\sum_{1\leq j\leq n}|\kappa_{ij,n}|$.

Lemma SA-4 *Suppose Assumptions 1–3 hold. If* $\tilde{\Gamma}_n^{-1} = O_p(1)$ *and if* $\mathcal{C}_n^\kappa = O_p(1)$, *then*

$$\hat{\Sigma}_n(\kappa_n) = \tilde{\Sigma}_n(\kappa_n) + o_p(1) = \mathbb{E}[\tilde{\Sigma}_n(\kappa_n)|\mathcal{X}_n, \mathcal{W}_n] + o_p(1), \qquad \tilde{\Sigma}_n(\kappa_n) = \frac{1}{n}\sum_{1\le i,j\le n} \kappa_{ij,n}\tilde{V}_{i,n}\tilde{V}_{i,n}'\tilde{U}_{j,n}^2.$$

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

Lemma SA-5 *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* $\mathbb{V}[\tilde{S}_n|\mathcal{X}_n, \mathcal{W}_n] = \sigma_n^2\tilde{\Gamma}_n$.

Finally, the sixth lemma can be combined with the fourth lemma to show consistency of $\hat{\Sigma}_n(\kappa_n)$. Part (a) is a general result stated under a high-level condition. Part (b) gives sufficient conditions for the assumption of part (a) for estimators of HCk type and part (c) does likewise for $\hat{\Sigma}_n^{\mathtt{HC}}$. Let $\mathcal{C}_n^M = \min_{1\le i\le n} \bar{M}_{ii,n}$.

Lemma SA-6 *Suppose Assumption 3 is satisfied.*

(a) If

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

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

(b) If $K_n/n \to 0$ *and if* $\kappa_{ij,n} = \omega_n \mathbb{1}(i=j)d_{i,n}M_{ii,n}^{-\xi_{i,n}}$, *where* $d_{i,n} \ge 0$, $0 \le \xi_{i,n} \le 4$, *and*

$$\frac{1}{n}\sum_{1\le i\le n}|1 - d_{i,n}| = o_p(1), \qquad \max_{1\le i\le n} d_{i,n} = O_p(1),$$

$$\max_{1\le i\le n}\omega_n\mathbb{1}(\xi_{i,n} > 0)M_{ii,n}^{-1} = O_p(1),$$

then $\mathcal{C}_n^\kappa = O_p(1)$ *and*

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

(c) If $\mathbb{P}[\mathcal{C}_n^M > 1/2] \to 1$, $1/(\mathcal{C}_n^M - 1/2) = O_p(1)$, *and if* $\kappa_n = \omega_n\kappa_n^{HC}$, *then* $\mathcal{C}_n^\kappa = O_p(1)$ *and*

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

2.1 Proof of Lemma SA-1

Assuming without loss of generality that $d = 1$, we have

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

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

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

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

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

In particular, $\tilde{\Gamma}_n - \mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n] = o_p(1)$ and $\tilde{\Gamma}_n = O_p(1)$ because

$$\mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n] \leq \frac{1}{n}\sum_{1\leq i\leq n}\mathbb{E}[V_{i,n}^2|\mathcal{W}_n] \leq \overline{\mathcal{C}}_{V,n} = O_p(1).$$

To complete the proof it suffices to show that $\hat{\Gamma}_n - \tilde{\Gamma}_n = o_p(1)$. Defining $\tilde{R}_{i,n}^v = \sum_{1\leq j\leq n}\bar{M}_{ij,n}R_{j,n}^v$ and using $\rho_n \to 0$, we have

$$\frac{1}{n}\sum_{1\leq i,j\leq n}|\tilde{R}_{i,n}^v|^2 \leq \frac{1}{n}\sum_{1\leq i\leq n}|R_{i,n}^v|^2 = O_p(\rho_n) = o_p(1)$$

and, by implication,

$$|\frac{1}{n}\sum_{1\leq i\leq n}\tilde{V}_{i,n}\tilde{R}_{i,n}^v|^2 \leq (\frac{1}{n}\sum_{1\leq i\leq n}\tilde{V}_{i,n}^2)(\frac{1}{n}\sum_{1\leq i,j\leq n}|\tilde{R}_{i,n}^v|^2) = \tilde{\Gamma}_n o_p(1) = o_p(1),$$

where the inequality uses the Cauchy-Schwarz inequality. As a consequence,

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

as was to be shown.

2.2 Proof of Lemma SA-2

Assuming without loss of generality that $d = 1$, we have

$$S_n - \tilde{S}_n = \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \tilde{R}^v_{i,n} U_{i,n} + \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \tilde{Q}^u_{i,n} V_{i,n} + \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \tilde{R}^v_{i,n} \tilde{R}^u_{i,n} + \frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \tilde{V}_{i,n} (\tilde{R}^u_{i,n} - \tilde{Q}^u_{i,n}),$$

where $\tilde{R}^u_{i,n} = \sum_{1 \le j \le n} \bar{M}_{ij,n} R^u_{j,n}$ and $\tilde{Q}^u_{i,n} = \sum_{1 \le j \le n} \bar{M}_{ij,n} Q^u_{j,n}$.

Under Assumptions 1 and 3, $n^{-1/2} \sum_{1 \le i \le n} \tilde{R}^v_{i,n} U_{i,n} = o_p(1)$ because $\mathbb{E}[\tilde{R}^v_{i,n} U_{i,n} | \mathcal{X}_n, \mathcal{W}_n] = 0$ and

$$\begin{aligned}
\mathbb{V}[\frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \tilde{R}^v_{i,n} U_{i,n} | \mathcal{X}_n, \mathcal{W}_n] &= \frac{1}{n} \sum_{1 \le i \le N_n} \mathbb{V}[\sum_{t \in \mathcal{T}_i} \tilde{R}^v_{t,n} U_{t,n} | \mathcal{X}_n, \mathcal{W}_n] \\
&\le \frac{1}{n} \sum_{1 \le i \le N_n} (\# \mathcal{T}_i) \sum_{t \in \mathcal{T}_i} |\tilde{R}^v_{t,n}|^2 \mathbb{V}[U_{t,n} | \mathcal{X}_n, \mathcal{W}_n] \\
&\le \mathcal{C}^{\mathcal{T}}_n \overline{\mathcal{C}}_{U,n} \frac{1}{n} \sum_{1 \le i \le N_n} \sum_{t \in \mathcal{T}_i} |\tilde{R}^v_{t,n}|^2 = \mathcal{C}^{\mathcal{T}}_n \overline{\mathcal{C}}_{U,n} \frac{1}{n} \sum_{1 \le i \le n} |\tilde{R}^v_{i,n}|^2 = O_p(\rho_n) = o_p(1).
\end{aligned}$$

Similarly, $n^{-1/2} \sum_{1 \le i \le n} \tilde{Q}^u_{i,n} V_{i,n} = o_p(1)$ because $\mathbb{E}[\tilde{Q}^u_{i,n} V_{i,n} | \mathcal{W}_n] = 0$ and

$$\mathbb{V}[\frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \tilde{Q}^u_{i,n} V_{i,n} | \mathcal{W}_n] \le \mathcal{C}^{\mathcal{T}}_n \overline{\mathcal{C}}_{V,n} \frac{1}{n} \sum_{1 \le i \le n} |Q^u_{i,n}|^2 = O_p(\bar{\varrho}_n) = o_p(1),$$

where the last equality uses $\bar{\varrho}_n \le \varrho_n = o(1)$. Finally, using the the Cauchy-Schwarz inequality,

$$|\frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \tilde{R}^v_{i,n} \tilde{R}^u_{i,n}|^2 \le n(\frac{1}{n} \sum_{1 \le i \le n} |R^v_{i,n}|^2)(\frac{1}{n} \sum_{1 \le i \le n} |R^u_{i,n}|^2) = O_p(n \rho_n \varrho_n) = o_p(1)$$

and

$$|\frac{1}{\sqrt{n}} \sum_{1 \le i \le n} \tilde{V}_{i,n} (\tilde{R}^u_{i,n} - \tilde{Q}^u_{i,n})|^2 \le n(\frac{1}{n} \sum_{1 \le i \le n} V^2_{i,n})(\frac{1}{n} \sum_{1 \le i \le n} |R^u_{i,n} - Q^u_{i,n}|^2) = O_p[n(\varrho_n - \bar{\varrho}_n)] = o_p(1),$$

where the penultimate equality uses $n^{-1} \sum_{1 \le i \le n} V^2_{i,n} = O_p(1)$ and $\mathbb{E}[|R^u_{i,n} - Q^u_{i,n}|^2] = \mathbb{E}[|R^u_{i,n}|^2] - \mathbb{E}[|Q^u_{i,n}|^2]$. As a consequence, $S_n = \tilde{S}_n + o_p(1)$.

Next, using Assumption 1,

$$\begin{aligned}
\mathbb{V}[\tilde{S}_n | \mathcal{X}_n, \mathcal{W}_n] &= \frac{1}{n} \sum_{1 \le i \le N_n} \sum_{t \in \mathcal{T}_i} \tilde{V}_{t,n} \tilde{V}'_{t,n} \mathbb{V}[U_{t,n} | \mathcal{X}_n, \mathcal{W}_n] = \frac{1}{n} \sum_{1 \le i \le n} \tilde{V}_{i,n} \tilde{V}'_{i,n} \mathbb{V}[U_{i,n} | \mathcal{X}_n, \mathcal{W}_n] \\
&\ge \underline{\mathcal{C}}_{U,n} \frac{1}{n} \sum_{1 \le i \le n} \tilde{V}_{i,n} \tilde{V}'_{i,n} = \underline{\mathcal{C}}_{U,n} \tilde{\Gamma}_n,
\end{aligned}$$

so $\mathbb{V}[\tilde{S}_n|,\mathcal{X}_n,\mathcal{W}_n]^{-1} = O_p(1)$. The proof can therefore be completed by showing that

$$\mathbb{V}[\tilde{S}_n|\mathcal{X}_n,\mathcal{W}_n]^{-1/2}\tilde{S}_n \to_d \mathcal{N}(0, I_d).$$

We shall do so assuming without loss of generality $\lambda_{\min}(\mathbb{V}[\tilde{S}_n|\mathcal{X}_n,\mathcal{W}_n]) > 0$ (a.s.) and that $d = 1$. (The case where $d > 1$ can be handled by means of the Cramér-Wold device and simple bounding arguments.) Because

$$\mathbb{V}[\tilde{S}_n|\mathcal{X}_n,\mathcal{W}_n]^{-1/2}\tilde{S}_n = \frac{1}{\sqrt{n}}\sum_{1\le i\le N_n}\sum_{t\in\mathcal{T}_i}\breve{V}_{t,n}U_{t,n}, \qquad \breve{V}_{t,n} = \mathbb{V}[\tilde{S}_n|\mathcal{X}_n,\mathcal{W}_n]^{-1/2}\tilde{V}_{t,n},$$

where, conditional on $(\mathcal{X}_n,\mathcal{W}_n)$, $\sum_{t\in\mathcal{T}_i}\breve{V}_{t,n}U_{t,n}$ are mean zero independent random variables with

$$\frac{1}{n}\sum_{1\le i\le N_n}\mathbb{V}[\sum_{t\in\mathcal{T}_i}\breve{V}_{t,n}U_{t,n}|\mathcal{X}_n,\mathcal{W}_n] = 1,$$

the result will follow from the Berry-Esseen inequality provided

$$\frac{1}{n}\sum_{1\le i\le N_n}\mathbb{E}[|\sum_{t\in\mathcal{T}_i}\breve{V}_{t,n}U_{t,n}|^3|\mathcal{X}_n,\mathcal{W}_n] = o_p(\sqrt{n}).$$

Now,

$$\begin{aligned}
\frac{1}{n}\sum_{1\le i\le N_n}\mathbb{E}[|\sum_{t\in\mathcal{T}_i}\breve{V}_{t,n}U_{t,n}|^3|\mathcal{X}_n,\mathcal{W}_n] &\le \frac{N_n}{n} + \frac{1}{n}\sum_{1\le i\le N_n}\mathbb{E}[|\sum_{t\in\mathcal{T}_i}\breve{V}_{t,n}U_{t,n}|^4|\mathcal{X}_n,\mathcal{W}_n] \\
&\le \frac{N_n}{n} + \frac{1}{n}\sum_{1\le i\le N_n}(\#\mathcal{T}_i)^3\sum_{t\in\mathcal{T}_i}\breve{V}_{t,n}^4\mathbb{E}[U_{t,n}^4|\mathcal{X}_n,\mathcal{W}_n] \\
&\le \frac{N_n}{n} + (\mathcal{C}_n^{\mathcal{T}})^3\overline{\mathcal{C}}_{U,n}\lambda_{\min}(\mathbb{V}[\tilde{S}_n|\mathcal{X}_n,\mathcal{W}_n])^{-2}\frac{1}{n}\sum_{1\le i\le N_n}\sum_{t\in\mathcal{T}_i}\tilde{V}_{t,n}^4 \\
&= O_p(1),
\end{aligned}$$

the last equality using the fact that

$$\frac{1}{n}\sum_{1\le i\le N_n}\sum_{t\in\mathcal{T}_i}\tilde{V}_{t,n}^4 = \frac{1}{n}\sum_{1\le i\le n}\tilde{V}_{i,n}^4 = O_p(1)$$

because

$$\begin{aligned}
\mathbb{E}[\tilde{V}_{i,n}^4|\mathcal{W}_n] &= \mathbb{E}[|\sum_{1\leq j\leq n}\bar{M}_{ij,n}V_{j,n}|^4|\mathcal{W}_n] \\
&= \frac{1}{n}\sum_{1\leq i,j,k,l,m\leq n}\mathbb{E}[\bar{M}_{ij,n}\bar{M}_{ik,n}\bar{M}_{il,n}\bar{M}_{im,n}V_{j,n}V_{k,n}V_{l,n}V_{m,n}|\mathcal{W}_n] \\
&= \frac{1}{n}(\sum_{1\leq i,j\leq n}\bar{M}_{ij,n}^4\mathbb{E}[V_{j,n}^4|\mathcal{W}_n] + 3\sum_{1\leq i,j,k\leq n,k\neq j}\bar{M}_{ij,n}^2\bar{M}_{ik,n}^2\mathbb{E}[V_{j,n}^2V_{k,n}^2|\mathcal{W}_n]) \\
&\leq 3\overline{\mathcal{C}}_{V,n}\frac{1}{n}(\sum_{1\leq i,j\leq n}\bar{M}_{ij,n}^4 + \sum_{1\leq i,j,k\leq n,k\neq j}\bar{M}_{ij,n}^2\bar{M}_{ik,n}^2) \\
&= 3\overline{\mathcal{C}}_{V,n}\frac{1}{n}\sum_{1\leq i,j,k\leq n}\bar{M}_{ij,n}^2\bar{M}_{ik,n}^2 = 3\overline{\mathcal{C}}_{V,n}\frac{1}{n}\sum_{1\leq i\leq n}\bar{M}_{ii,n}^2 \leq 3\overline{\mathcal{C}}_{V,n} = O_p(1).
\end{aligned}$$

2.3 Proof of Lemma SA-3

Defining $\tilde{n} = n - K_n$, we have

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

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

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

and $\sum_{1\leq i,j\leq N_n}\sum_{s\in\mathcal{T}_{i,n},t\in\mathcal{T}_{j,n}}\bar{M}_{st,n}^2 = \tilde{n}$. As a consequence,

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

and

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

In particular, $\tilde{\sigma}_n^2 - \mathbb{E}[\tilde{\sigma}_n^2|\mathcal{X}_n,\mathcal{W}_n] = o_p(1)$ and $\tilde{\sigma}_n^2 = O_p(1)$ because

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

To complete the proof it suffices to show that $\hat{\sigma}_n^2 - \tilde{\sigma}_n^2/(1-d/\tilde{n}) = o_p(1)$. Using $\varrho_n \to 0$, we have

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

Also, by Lemmas SA-1 and SA-2, $(\hat{\beta}_n-\beta)'\tilde{\Gamma}_n(\hat{\beta}_n-\beta) = o_p(1)$. As a consequence, using $\hat{u}_{i,n}-\tilde{U}_{i,n} = \tilde{R}^u_{i,n}+\tilde{V}'_{i,n}(\hat{\beta}_n-\beta)$,

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

implying that

$$|\hat{\sigma}^2_n-\tilde{\sigma}^2_n\frac{1}{1-d/\tilde{n}}| \leq \frac{1}{\tilde{n}-d}\sum_{1\leq i\leq n}(\hat{u}_{i,n}-\tilde{U}_{i,n})^2+\sqrt{\tilde{\sigma}^2_n\frac{1}{1-d/\tilde{n}}}\sqrt{\frac{1}{\tilde{n}-d}\sum_{1\leq i\leq n}(\hat{u}_{i,n}-\tilde{U}_{i,n})^2} = o_p(1),$$

where the inequality uses the Cauchy-Schwarz inequality and the equality uses $\tilde{\sigma}^2_n = O_p(1)$.

2.4 Proof of Lemma SA-4

Assuming without loss of generality that $d = 1$, we have

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

$$A_{ij,n} = \sum_{s\in\mathcal{T}_{i,n},t\in\mathcal{T}_{j,n}}(\sum_{1\leq k,l\leq n}\kappa_{kl,n}\tilde{V}^2_{k,n}\bar{M}_{sl,n}\bar{M}_{tl,n})U_{s,n}U_{t,n},$$

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

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

$$\begin{aligned}
&\sum_{1\leq i,j\leq N_n}\sum_{s\in\mathcal{T}_{i,n},t\in\mathcal{T}_{j,n}}\sum_{1\leq k,l,K,L\leq n}\kappa_{kl,n}\kappa_{KL,n}\tilde{V}^2_{k,n}\tilde{V}^2_{K,n}\bar{M}_{sl,n}\bar{M}_{tl,n}\bar{M}_{sL,n}\bar{M}_{tL,n} \\
&= \sum_{1\leq i,j\leq n}\sum_{1\leq k,l,K,L\leq n}\kappa_{kl,n}\kappa_{KL,n}\tilde{V}^2_{k,n}\tilde{V}^2_{K,n}\bar{M}_{il,n}\bar{M}_{jl,n}\bar{M}_{iL,n}\bar{M}_{jL,n} \\
&= \sum_{1\leq k,l,K,L\leq n}\kappa_{kl,n}\kappa_{KL,n}\tilde{V}^2_{k,n}\tilde{V}^2_{K,n}\bar{M}^2_{lL,n} \leq \sum_{1\leq k,l,K,L\leq n}|\kappa_{kl,n}||\kappa_{KL,n}|\tilde{V}^2_{k,n}\tilde{V}^2_{K,n}\bar{M}^2_{lL,n},
\end{aligned}$$

and, using $\mathbb{E}[\tilde{V}_{k,n}^2 \tilde{V}_{K,n}^2 | \mathcal{W}_n] \leq 3\overline{\mathcal{C}}_{V,n}$,

$$\begin{aligned}\mathbb{E}[\sum_{1\leq k,l,K,L\leq n} |\kappa_{kl,n}||\kappa_{KL,n}|\tilde{V}_{k,n}^2 \tilde{V}_{K,n}^2 M_{lL,n}^2 | \mathcal{W}_n] &\leq 3\overline{\mathcal{C}}_{V,n} \sum_{1\leq k,l,K,L\leq n} |\kappa_{kl,n}||\kappa_{KL,n}|\bar{M}_{lL,n}^2 \\ &\leq 3(\mathcal{C}_n^\kappa)^2 \overline{\mathcal{C}}_{V,n} \sum_{1\leq l,L\leq n} \bar{M}_{lL,n}^2 \leq 3(\mathcal{C}_n^\kappa)^2 \overline{\mathcal{C}}_{V,n} n = O_p(n).\end{aligned}$$

As a consequence,

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

and

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

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

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

To complete the proof it suffices to show that

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

To do so, it suffices (by the Cauchy-Schwarz inequality and $\tilde{\Sigma}_n(\kappa_n) = O_p(1)$) to show that

$$\frac{1}{n}\sum_{1\leq i,j\leq n} |\kappa_{ij,n}|\tilde{V}_{j,n}^2|\tilde{Q}_{i,n}^u|^2 = o_p(1), \qquad \frac{1}{n}\sum_{1\leq i,j\leq n} |\kappa_{ij,n}||\tilde{R}_{j,n}^v|^2\tilde{U}_{i,n}^2 = o_p(1),$$

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

$$\frac{1}{n}\sum_{1\leq i,j\leq n} |\kappa_{ij,n}||\tilde{R}_{j,n}^v|^2|\tilde{Q}_{i,n}^u|^2 = o_p(1), \qquad \frac{1}{n}\sum_{1\leq i,j\leq n} |\kappa_{ij,n}||\tilde{R}_{j,n}^v|^2|\tilde{R}_{i,n}^u - \tilde{Q}_{i,n}^u|^2 = o_p(1),$$

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

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

$$\begin{aligned}\mathbb{E}[\frac{1}{n}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}|\tilde{V}_{j,n}^2|\tilde{Q}_{i,n}^u|^2|\mathcal{W}_n] &= \frac{1}{n}\sum_{1\leq i\leq n}|\tilde{Q}_{i,n}^u|^2\sum_{1\leq j\leq n}|\kappa_{ij,n}|\mathbb{E}[\tilde{V}_{j,n}^2|\mathcal{W}_n] \\ &\leq \overline{\mathcal{C}}_{V,n}\mathcal{C}_n^\kappa\frac{1}{n}\sum_{1\leq i\leq n}|\tilde{Q}_{i,n}^u|^2 = O_p(\bar{\varrho}_n) = o_p(1)\end{aligned}$$

and $n^{-1}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}||\tilde{R}_{j,n}^v|^2\tilde{U}_{i,n}^2 = o_p(1)$ because

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

Next,

$$\frac{1}{n}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}|\tilde{V}_{j,n}^2|\tilde{R}_{i,n}^u-\tilde{Q}_{i,n}^u|^2 = n\mathcal{C}_n^\kappa(\frac{1}{n}\sum_{1\leq i\leq n}\tilde{V}_{i,n}^2)(\frac{1}{n}\sum_{1\leq i\leq n}|\tilde{R}_{i,n}^u-\tilde{Q}_{i,n}^u|^2) = o_p(1)$$

because $n^{-1}\sum_{1\leq i\leq n}|\tilde{R}_{i,n}^u-\tilde{Q}_{i,n}^u|^2 = O_p(\varrho_n-\bar{\varrho}_n) = o_p(n^{-1})$ and $n^{-1}\sum_{1\leq i\leq n}\tilde{V}_{i,n}^2 = O_p(1)$.

Similarly,

$$\frac{1}{n}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}||\tilde{R}_{j,n}^v|^2|\tilde{Q}_{i,n}^u|^2 \leq n\mathcal{C}_n^\kappa(\frac{1}{n}\sum_{1\leq i\leq n}|\tilde{R}_{i,n}^v|^2)(\frac{1}{n}\sum_{1\leq i\leq n}|\tilde{Q}_{i,n}^u|^2) = O_p(n\rho_n\bar{\varrho}_n) = o_p(1)$$

and

$$\begin{aligned}\frac{1}{n}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}||\tilde{R}_{j,n}^v|^2|\tilde{R}_{i,n}^u-\tilde{Q}_{i,n}^u|^2 &\leq n\mathcal{C}_n^\kappa(\frac{1}{n}\sum_{1\leq i\leq n}|\tilde{R}_{i,n}^v|^2)(\frac{1}{n}\sum_{1\leq i\leq n}|\tilde{R}_{i,n}^u-\tilde{Q}_{i,n}^u|^2) \\ &\leq O_p[n\rho_n(\varrho_n-\bar{\varrho}_n)] = o_p(1).\end{aligned}$$

Finally,

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

and

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

because $\hat{\beta}_n-\beta = o_p(1)$ and

$$\mathbb{E}[\frac{1}{n}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}|\tilde{V}_{j,n}^2\tilde{V}_{i,n}^2|\mathcal{W}_n] \leq 3\overline{\mathcal{C}}_{V,n}\frac{1}{n}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}| \leq 3\overline{\mathcal{C}}_{V,n}\mathcal{C}_n^\kappa = O_p(1),$$

$$\mathbb{E}[\frac{1}{n}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}||\tilde{R}^v_{j,n}|^2\tilde{V}^2_{i,n}|\mathcal{W}_n] \leq \overline{\mathcal{C}}_{V,n}\frac{1}{n}\sum_{1\leq i,j\leq n}|\kappa_{ij,n}||\tilde{R}^v_{j,n}|^2$$

$$\leq \overline{\mathcal{C}}_{V,n}\mathcal{C}^{\kappa}_n\frac{1}{n}\sum_{1\leq i\leq n}|\tilde{R}^v_{i,n}|^2 = O_p(\rho_n) = o_p(1).$$

2.5 Proof of Lemma SA-5

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

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

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

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

and

$$\mathbb{V}[\tilde{S}_n|\mathcal{X}_n,\mathcal{W}_n] = \frac{1}{n}\sum_{1\leq i\leq n}\tilde{V}_{i,n}\tilde{V}'_{i,n}\mathbb{E}[U^2_{i,n}|\mathcal{X}_n,\mathcal{W}_n] = \sigma^2_n\frac{1}{n}\sum_{1\leq i\leq n}\tilde{V}_{i,n}\tilde{V}'_{i,n} = \sigma^2_n\tilde{\Gamma}_n.$$

2.6 Proof of Lemma SA-6

Assuming without loss of generality that $d=1$, we have

$$\mathbb{E}[\tilde{\Sigma}_n(\kappa_n)|\mathcal{X}_n,\mathcal{W}_n] - \mathbb{V}[\tilde{S}_n|\mathcal{X}_n,\mathcal{W}_n] = \frac{1}{n}\sum_{1\leq i,j\leq n}A_{ij,n}\tilde{V}^2_{i,n}\mathbb{E}[U^2_{j,n}|\mathcal{X}_n,\mathcal{W}_n],$$

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

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

$$|\mathbb{E}[\tilde{\Sigma}^{EW}_n|\mathcal{X}_n,\mathcal{W}_n] - \mathbb{V}[\tilde{S}_n|\mathcal{X}_n,\mathcal{W}_n]| \leq \frac{1}{n}\sum_{1\leq i,j\leq n}|A_{ij,n}|\tilde{V}^2_{i,n}\mathbb{E}[U^2_{j,n}|\mathcal{X}_n,\mathcal{W}_n]$$

$$\leq \overline{\mathcal{C}}_{U,n}\frac{1}{n}\sum_{1\leq i,j\leq n}|A_{ij,n}|\tilde{V}^2_{i,n} = o_p(1)$$

because

$$\mathbb{E}[\frac{1}{n}\sum_{1\leq i,j\leq n}|A_{ij,n}|\tilde{V}^2_{i,n}|\mathcal{W}_n] = \frac{1}{n}\sum_{1\leq i,j\leq n}|A_{ij,n}|\mathbb{E}[\tilde{V}^2_{i,n}|\mathcal{W}_n] \leq \overline{\mathcal{C}}_{V,n}\frac{1}{n}\sum_{1\leq i,j\leq n}|A_{ij,n}| = o_p(1).$$

This establishes part (a).

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

$$\begin{aligned}\frac{1}{n}\sum_{1\le i,j\le n}|A_{ij,n}| &= \frac{1}{n}\sum_{1\le i\le n}|d_{i,n}M_{ii,n}^{-\xi_{i,n}}M_{ii,n}^2 - 1| + \frac{1}{n}\sum_{1\le i,j\le n, j\ne i} d_{i,n}M_{ii,n}^{-\xi_{i,n}}M_{ji,n}^2 \\ &\le \frac{1}{n}\sum_{1\le i\le n}|d_{i,n}-1| + \frac{1}{n}\sum_{1\le i\le n} d_{i,n}|M_{ii,n}^{2-\xi_{i,n}} - 1| + \frac{1}{n}\sum_{1\le i,j\le n, j\ne i} d_{i,n}M_{ii,n}^{-\xi_{i,n}}M_{ji,n}^2,\end{aligned}$$

where, using $\sum_{1\le i\le n}(1-M_{ii,n}^2) \le 2K_n$ and $\sum_{1\le i,j\le n, j\ne i} M_{ji,n}^2 \le K_n$,

$$\begin{aligned}&\frac{1}{n}\sum_{1\le i\le n} d_{i,n}|M_{ii,n}^{2-\xi_{i,n}} - 1| \\ &\le (\max_{1\le i\le n} d_{i,n})\left[\frac{1}{n}\sum_{1\le i\le n}\mathbb{1}(\xi_{i,n}=0)(1-M_{ii,n}^2) + \frac{1}{n}\sum_{1\le i\le n}\mathbb{1}(\xi_{i,n}>0)M_{ii,n}^{-2}|M_{ii,n}^{4-\xi_{i,n}} - M_{ii,n}^2|\right] \\ &\le (\max_{1\le i\le n} d_{i,n})(1 + [\max_{1\le i\le n}\mathbb{1}(\xi_{i,n}>0)M_{ii,n}^{-1}]^2)\frac{1}{n}\sum_{1\le i\le n}(1-M_{ii,n}^2) \\ &\le 2\frac{K_n}{n}(\max_{1\le i\le n} d_{i,n})(1 + [\max_{1\le i\le n}\mathbb{1}(\xi_{i,n}>0)M_{ii,n}^{-1}]^2)\end{aligned}$$

and

$$\begin{aligned}&\frac{1}{n}\sum_{1\le i,j\le n, j\ne i} d_{i,n}M_{ii,n}^{-\xi_{i,n}}M_{ji,n}^2 \\ &\le (\max_{1\le i\le n} d_{i,n})\left[\frac{1}{n}\sum_{1\le i,j\le n, j\ne i}\mathbb{1}(\xi_{i,n}=0)M_{ji,n}^2 + \frac{1}{n}\sum_{1\le i,j\le n, j\ne i}\mathbb{1}(\xi_{i,n}>0)M_{ii,n}^{-\xi_{i,n}}M_{ji,n}^2\right] \\ &\le (\max_{1\le i\le n} d_{i,n})(1 + [\max_{1\le i\le n}\mathbb{1}(\xi_{i,n}>0)M_{ii,n}^{-1}]^4)\frac{1}{n}\sum_{1\le i,j\le n, j\ne i} M_{ji,n}^2 \\ &\le \frac{K_n}{n}(\max_{1\le i\le n} d_{i,n})(1 + [\max_{1\le i\le n}\mathbb{1}(\xi_{i,n}>0)M_{ii,n}^{-1}]^4)\frac{1}{n}\sum_{1\le i,j\le n, j\ne i} M_{ji,n}^2.\end{aligned}$$

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

Finally, if $\lambda_{\min}(\sum_{k=1}^n w_{k,n}w_{k,n}') > 0$ and if $\mathcal{C}_n^M > 1/2$, then

$$\frac{1}{n}\sum_{1\le i\le n}\left|\sum_{1\le k\le n}\kappa_{ik,n}^{HC}M_{ik,n}^2 - 1\right| + \frac{1}{n}\sum_{1\le i,j\le n, j\ne i}\left|\sum_{1\le k\le n}\kappa_{ik,n}^{HC}M_{jk,n}^2\right| = 0$$

and, by Theorem 1 of Varah (1975),

$$\mathcal{C}_n^\kappa \le \frac{1}{\mathcal{C}_n^M - 1/2}.$$

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

3 Proofs of Main Results

3.1 General Results

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

3.2 Linear Regression Model with Increasing Dimension

If Assumption LR1 holds, then Assumption 1 holds with $\mathcal{W}_n = (w_{1,n}, \ldots, w_{n,n})$, $N_n = n$, $\mathcal{T}_{i,n} = \{i\}$, and $\mathcal{C}_n^{\mathcal{T}} = 1$. Moreover, in this case we have

$$\rho_n = \min_{\delta \in \mathbb{R}^{K_n \times d}} \mathbb{E}[\|\mathbb{E}(x_{i,n}|w_{i,n}) - \delta' w_{i,n}\|^2] = \rho_n^{\mathtt{LR}}.$$

If also Assumption LR2 holds, then $R_{i,n}^u = Q_{i,n}^u = 0$, so Assumption 2 holds (with $\varrho_n + n(\varrho_n - \bar{\varrho}_n) = 0$ and $n\varrho_n\rho_n = 0$). Finally, Assumption 3 is implied by Assumption LR3 under Assumptions LR1 and LR2.

Theorem LR follows from Theorems 2-4 provided $\tilde{\Gamma}_n^{-1} = O_p(1)$. By Lemma SA-1, a sufficient condition for this to hold is that $\mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n]^{-1} = O_p(1)$. Now, $\mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n] = n^{-1}\sum_{1\leq i\leq n}\mathbb{E}[\tilde{V}_{i,n}\tilde{V}_{i,n}'|w_{i,n}]$ and therefore

$$\begin{aligned}
\lambda_{\min}(\mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n]) &= \omega_n\lambda_{\min}(\frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n}\mathbb{E}[V_{i,n}V_{i,n}'|w_{i,n}]) \\
&\geq \omega_n\frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n}\lambda_{\min}(\mathbb{E}[V_{i,n}V_{i,n}'|w_{i,n}]) \\
&\geq \omega_n(\frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n})\min_{1\leq i\leq n}\lambda_{\min}(\mathbb{E}[V_{i,n}V_{i,n}'|w_{i,n}]) \\
&= \omega_n(1 - K_n/n)\min_{1\leq i\leq n}\lambda_{\min}(\mathbb{E}[V_{i,n}V_{i,n}'|w_{i,n}]),
\end{aligned}$$

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

3.3 Semiparametric Partially Linear Model

If Assumption PL1 holds, then Assumption 1 holds with $\mathcal{W}_n = (z_1, \ldots, z_n)$, $N_n = n$, $\mathcal{T}_{i,n} = \{i\}$, and $\mathcal{C}_n^{\mathcal{T}} = 1$. Moreover, in this case we have

$$\rho_n = \min_{\delta \in \mathbb{R}^{K_n \times d}} \mathbb{E}[\|\mathbb{E}(x_i|z_i) - \delta' p_n(z_i)\|^2] = \rho_n^{\mathtt{PL}}$$

and, using $\mathbb{E}(y_i - \beta' x_i|x_i, z_i) = g(z_i) = \mathbb{E}(y_i - \beta' x_i|z_i)$,

$$\bar{\varrho}_n = \min_{\gamma \in \mathbb{R}^{K_n}} \mathbb{E}[|\mathbb{E}(y_i - \beta' x_i|z_i) - \gamma' p_n(z_i)|^2] = \min_{\gamma \in \mathbb{R}^{K_n}} \mathbb{E}[|\mathbb{E}(y_i - \beta' x_i|x_i, z_i) - \gamma' p_n(z_i)|^2] = \varrho_n = \varrho_n^{\mathtt{PL}},$$

so Assumption 2 holds (with $n(\varrho_n - \bar{\varrho}_n) = 0$) when Assumption PL2 holds. Finally, Assumption 3 is implied by Assumption PL3 under Assumptions PL1 and PL2.

Theorem PL follows from Theorems 2-4 provided $\tilde{\Gamma}_n^{-1} = O_p(1)$. By Lemma SA-1, a sufficient condition for this to hold is that $\mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n]^{-1} = O_p(1)$. Now, $\mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n] = \omega_n n^{-1} \sum_{1\leq i\leq n} M_{ii,n}\mathbb{E}[\nu_i\nu_i'|z_i]$ and therefore

$$\begin{aligned}
\lambda_{\min}(\mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n]) &\geq \omega_n \frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n}\lambda_{\min}(\mathbb{E}[\nu_i\nu_i'|z_i]) \\
&\geq \omega_n(\frac{1}{n}\sum_{1\leq i\leq n} M_{ii,n}) \min_{1\leq i\leq n} \lambda_{\min}(\mathbb{E}[\nu_i\nu_i'|z_i]) \\
&= \omega_n(1 - K_n/n) \min_{1\leq i\leq n} \lambda_{\min}(\mathbb{E}[\nu_i\nu_i'|z_i]),
\end{aligned}$$

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

3.4 Fixed Effects Panel Data Regression Model

If Assumption FE1 holds, then Assumption 1 holds with $\mathcal{W}_n = (w_{1,n}, \ldots, w_{n,n})$, $N_n = N = n/T$, $\mathcal{T}_{i,n} = \{T(i-1)+1, \ldots, Ti\}$, and $\mathcal{C}_n^{\mathcal{T}} = T$. Moreover, in this case we have $\rho_n = \bar{\varrho}_n = \varrho_n = 0$ under Assumption FE2, so Assumption 2 holds when Assumption FE2 holds. Finally, Assumption 3 is implied by Assumption FE3 under Assumptions FE1 and FE2.

Theorem FE follows from Theorems 2 and 4 provided $\tilde{\Gamma}_n^{-1} = O_p(1)$. By Lemma SA-1, a sufficient condition for this to hold is that $\mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n]^{-1} = O_p(1)$. Now, $\mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n] = T^{-1}\sum_{1\leq t\leq T} \mathbb{E}[\tilde{X}_{it}\tilde{X}_{it}']$ and therefore $\mathbb{E}[\tilde{\Gamma}_n|\mathcal{W}_n]^{-1} = O_p(1)$ under Assumption FE3.

4 Properties of $M_n \odot M_n$

Because M_n is symmetric, so is $M_n \odot M_n$ and it follows from Gerschgorin's Theorem (see, e.g., Barnes and Hoffman (1981) for an interesting discussion) that

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

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

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

Thus, $\lambda_{\min}(M_n \odot M_n) > 0$ (i.e., $M_n \odot M_n$ is positive definite) whenever $\mathcal{C}_n^M = \min_{1\leq i\leq n} M_{ii,n} > 1/2$.

Under the same condition, $M_n \odot M_n$ is diagonally dominant and it follows from Theorem 1 of

Varah (1975) that

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

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.

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