A. Proofs of main results

A.1. Proofs of Theorems 3 and 4

The proofs of the main results rely on the following simple lemma.

Lemma 9. *Consider the $\boldsymbol{W}$-estimate as defined in Algorithm 1. Assume $\|\boldsymbol{x}_i\|_2^2 \leq C$. Then for any i,*

$$\|\mathrm{I}_p - \boldsymbol{W}_{i-1}\boldsymbol{X}_{i-1}\|_F^2 - \|\mathrm{I}_p - \boldsymbol{W}_i\boldsymbol{X}_i\|_F^2 \asymp 2\lambda\|\boldsymbol{w}_i\|_2^2$$

Proof. This follows directly from the fact that $\boldsymbol{W}_i\boldsymbol{X}_i = \boldsymbol{W}_{i-1}\boldsymbol{X}_{i-1} + \boldsymbol{w}_i\boldsymbol{x}_i^{\mathsf{T}}$ and the following formula for $\boldsymbol{w}_i$:

$$\boldsymbol{w}_i = \frac{(\mathrm{I}_p - \boldsymbol{W}_{i-1}\boldsymbol{X}_{i-1})\boldsymbol{x}_i}{\lambda + \|\boldsymbol{x}_i\|_2^2}$$

which implies:

$$\|\mathrm{I}_p - \boldsymbol{W}_{i-1}\boldsymbol{X}_{i-1}\|_F^2 - \|\mathrm{I}_p - \boldsymbol{W}_i\boldsymbol{X}_i\|_F^2 = (2\lambda + \|\boldsymbol{x}_i\|_2^2)\|\boldsymbol{w}_i\|_2^2$$

The result follows as $\|\boldsymbol{x}_i\|_2^2$ is bounded uniformly. □

We can now prove Theorems 3 and 4 in a straightforward fashion.

Proof of Theorem 3. We have:

$$\begin{aligned}\mathsf{Tr}\{\mathrm{Var}(\mathsf{v})\} &= \sigma^2\mathbb{E}\Big\{\sum_i \|\boldsymbol{w}_i\|_2^2\Big\}\\ &\asymp \frac{\sigma^2}{2\lambda}\Big(\|\mathrm{I}_p\|_F^2 - \mathbb{E}\{\|\mathrm{I}_p - \boldsymbol{W}_n\boldsymbol{X}_n\|_F^2\}\Big),\end{aligned}$$

where in the second line we use Lemma 9 and sum over the telescoping series in i. The result follows. □

Proof of Theorem 4. From Lemma 2, the definition of the spectral norm $\|\cdot\|_{op}$, and Cauchy-Schwarz we have that

$$\|\beta - \mathbb{E}\{\widehat{\beta}\}\|_2^2 \leq \mathbb{E}\{\|\mathrm{I}_p - \boldsymbol{W}_n\boldsymbol{X}_n\|_{op}^2\}\mathbb{E}\{\|\widehat{\beta}_{\mathsf{OLS}} - \beta\|_2^2\}.$$

Using Theorem 1, the second term is bounded by $p\sigma^2\mathbb{E}\{\log\lambda_{\max}/\lambda_{\min}\}$. We first show that this term is at most $p\sigma^2\log n/\lambda(n)$, under the conditions of Theorem 4. First, note that

$$\begin{aligned}\lambda_{\max} &\leq \mathsf{Tr}\{\boldsymbol{X}_n\boldsymbol{X}_n^{\mathsf{T}}\}\\ &\leq \sum_i \|\boldsymbol{x}_i^2\|_2\\ &\leq Cn.\end{aligned}$$

With this and condition (5), we have that:

$$\begin{aligned}\mathbb{E}\Big\{\frac{\log(\lambda_{\max})}{\lambda_{\min}}\Big\} &\leq \frac{\log n}{\lambda(n)} + O\Big(\frac{1}{n}\Big)\\ &= O\Big(\frac{\log n}{\lambda(n)}\Big).\end{aligned}$$

Combining with condition (4) and Theorem 3 gives the result. □

We split the proof of Proposition 5 for the different conditions independently in the following lemmas.

Lemma 10. *Suppose that the data collection process satisfies (6) and (7). Then for any $\lambda \geq 1$ we have that:*

$$\mathbb{E}\{\|\mathrm{I}_p - \boldsymbol{W}_n\boldsymbol{X}_n\|_F^2\} \leq p\exp\Big(-\frac{n\bar{\mu}_n}{\lambda}\Big)$$

Proof. Define $\boldsymbol{M}_i = \mathrm{I}_p - \boldsymbol{W}_i \boldsymbol{X}_i$. Then, from Lemma 9 and the closed form for $\boldsymbol{w}_i$ we have that:

$$\begin{aligned}\|\boldsymbol{M}_{i-1}\|_F^2 - \|\boldsymbol{M}_i\|_F^2 &= \frac{2\lambda + \|\boldsymbol{x}_i\|_2^2}{(\lambda + \|\boldsymbol{x}_i\|_2^2)^2} \mathsf{Tr}\{\boldsymbol{M}_{i-1}\boldsymbol{x}_i\boldsymbol{x}_i^\mathsf{T}\boldsymbol{M}_{i-1}^\mathsf{T}\} \\ &\geq \frac{1}{\lambda + \|\boldsymbol{x}_i\|_2^2} \mathsf{Tr}\{\boldsymbol{M}_{i-1}\boldsymbol{x}_i\boldsymbol{x}_i^\mathsf{T}\boldsymbol{M}_{i-1}^\mathsf{T}\}.\end{aligned}$$

We now take expectations conditional on $\mathcal{G}_{i-1}$ on both sides. Observing that (i) $\boldsymbol{W}_n$, $\boldsymbol{X}_n$ and, therefore, $\boldsymbol{M}_n$ are well-adapted and (ii) using condition (6), we have

$$\begin{aligned}\mathbb{E}\{\|\boldsymbol{M}_{i-1}\|_F^2 \,|\mathcal{G}_{i-1}\} - \mathbb{E}\{\|\boldsymbol{M}_i\|_F^2 \,|\mathcal{G}_{i-1}\} &\geq \frac{\mu_i(n)}{\lambda}\mathbb{E}\{\|\boldsymbol{M}_i\|_F^2 \,|\mathcal{G}_{i-1}\}, \\ \text{or } \ \mathbb{E}\{\|\boldsymbol{M}_i\|_F^2 \,|\mathcal{G}_{i-1}\} &\leq \exp\Big(\frac{-\mu_i(n)}{\lambda}\Big)\mathbb{E}\{\|\boldsymbol{M}_{i-1}\|_F^2 \,|\mathcal{G}_{i-1}\}.\end{aligned}$$

Removing the conditioning on $\mathcal{G}_{i-1}$ and iterating over $i = 1, 2, \dots, n$ gives the claim. □

Lemma 11. *If the matrices $\{\boldsymbol{x}_i\boldsymbol{x}_i^\mathsf{T}\}_{i\leq n}$ commute, we have that*

$$\|\mathrm{I}_p - \boldsymbol{W}_n\boldsymbol{X}_n\|_{op} \leq \exp\Big(-\frac{\lambda_{\min}}{\lambda}\Big)$$

Proof. From the closed form in Lemma 9 and induction, we get that:

$$\mathrm{I}_p - \boldsymbol{W}_n\boldsymbol{X}_n = \prod_{i\leq n}\Big(\mathrm{I}_p - \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\mathsf{T}}{\lambda + \|\boldsymbol{x}_i\|_2^2}\Big).$$

The scalar equality $\exp(a+b) = \exp(a)\exp(b)$ extends to commuting matrices $\boldsymbol{A}, \boldsymbol{B}$. Applying this to the terms in the product above, which commute by assumption:

$$\begin{aligned}\mathrm{I}_p - \boldsymbol{W}_n\boldsymbol{X}_n &= \exp\Big[\sum_i \log\Big(\mathrm{I}_p - \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\mathsf{T}}{\lambda + \|\boldsymbol{x}_i\|_2^2}\Big)\Big] \\ &\preccurlyeq \exp\Big(-\sum_i \frac{\boldsymbol{x}_i\boldsymbol{x}_i^\mathsf{T}}{\lambda}\Big),\end{aligned}$$

using the fact that $\exp(\log(1-a)) \leq -a$. Finally, employing commutativity the fact that $\lambda_{\min}$ is the minimum eigenvalue of $\boldsymbol{X}_n^\mathsf{T}\boldsymbol{X}_n = \sum_i \boldsymbol{x}_i\boldsymbol{x}_i^\mathsf{T}$, the desired result follows. □

We can now prove Proposition 5.

Proof of Proposition 5. We need to satisfy conditions (4) and (5) for both the cases. Using either Lemma 10 or 11, with the appropriate choice of $\lambda(n)$ we have that

$$\mathbb{E}\{\|\mathrm{I}_p - \boldsymbol{W}_n\boldsymbol{X}_n\|_{op}^2\} = o(1/\log n),$$

thus obtaining condition (4). In fact, this can be made polynomially small with a constant factor smaller choice of $\lambda(n)$. Condition (5) only needs to be verified for the case of Lemma 10 or condition (6). It follows from a standard application of the matrix Azuma inequality (Tropp, 2012), the fact that $n\bar{\mu}_n \geq K\sqrt{n}$ and the fact that $\|\boldsymbol{x}_i\|_2^2$ are bounded. □

A.2. Proof of Theorem 10: Central limit theorem

It suffices to show the case for $p = 1$ of the theorem, and the case when $\sigma^2 = 1$ by rescaling. In this case, it is not hard to show that the moment stability assumption of the main article subsumes the condition of the following:

Theorem 12. *Let $(w_i(n), \varepsilon_i(n), \mathcal{F}_i(n))_{i\leq m(n)}$ be a triangular martingale difference array. Here for each $n \geq 1$, $\mathcal{F}_i(n)$ is a non-decreasing sequence of sub-sigma-algebras, $\varepsilon_i(n)$ are i.i.d. (uniformly) bounded random variables with $\mathbb{E}\{\varepsilon_i|\mathcal{F}_{i-1}\} = 0, \mathbb{E}\{\varepsilon_i^2|\mathcal{F}_{i-1}\} = 1$ and $w_i \in \mathrm{m}\mathcal{F}_{i-1}$ is predictable and bounded by 1 almost surely. Further, define $S_\ell = \sum_{i\leq \ell} w_i\varepsilon_i$ and $\sigma_i^2 = \sum_{\ell\leq i} w_i^2$ for $i \leq m(n)$.*

Suppose that for all $t \in \mathbb{R}$:

$$\sum_{i=1}^m \mathbb{E}\big|\mathbb{E}\{w_i\varepsilon_i \exp(-\sigma_m^2 t^2/2)|\mathcal{F}_{i-1}\}\big| = o_n(1)$$
$$\sum_{i=1}^m \mathbb{E}\big|\mathbb{E}\{w_i^2(\varepsilon_i^2-1) \exp(-\sigma_m^2 t^2/2)|\mathcal{F}_{i-1}\}\big| = o_n(1),$$
$$\sum_{i=1}^m \mathbb{E}\{w_i^3\} = o_n(1).$$

Then, for any function $\varphi : \mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}$, bounded and continuous, and $\xi \sim \mathsf{N}(0,1)$ independent of σ_m^2, we have that:

$$\limsup_{n\to\infty} \mathbb{E}\{\varphi(S_m;\sigma_m^2)\} - \mathbb{E}\{\varphi(\sigma_m\xi;\sigma_m^2)\} = 0.$$

Proof. By Levy's continuity theorem it suffices to consider the subclass of functions $\varphi(x;y) = \varphi_t(x;y) = \exp(\mathrm{i}txy)$ for any (fixed) $t \in \mathbb{R}$. Note that, since ξ is independent of $\bar{\sigma}_n$, $\mathbb{E}\{\exp(\mathrm{i}t\sigma_n\xi)|\mathcal{F}_i\} = \mathbb{E}\{\exp(-\sigma_n^2t^2/2)|\mathcal{F}_i\}$ a.s. for all t and i. The following proof is mostly standard, and uses the Fourier analytic technique.

For simplicity, define the following errors that will ultimately be controlled by the conditions of the theorem:

$$\nu_i^1(t) \equiv \mathbb{E}\{w_i\varepsilon_i\exp(-\sigma_n^2t^2/2)|\mathcal{F}_{i-1}\},$$
$$\nu_i^2(t) \equiv \mathbb{E}\{w_i^2(\varepsilon_i^2-1)\exp(-\sigma_n^2t^2/2)|\mathcal{F}_{i-1}\}.$$

We will show that the following main estimate:

$$\begin{aligned}|\mathbb{E}\{\exp(\mathrm{i}tS_i + (\sigma_i^2-\sigma_n^2)t^2/2)|\mathcal{F}_{i-1}\} - \exp(\mathrm{i}tS_{i-1} + \sigma_{i-1}^2t^2/2)\mathbb{E}\{\exp(-\sigma_n^2t^2/2)|\mathcal{F}_{i-1}\}| \\ = O\big(t|\nu_i^1(t)| + t^2|\nu_i^2(t)| + t^3w_i^3\big). \qquad (8)\end{aligned}$$

Assuming this estimate, the result is straightforward. Taking expectations above and summing the above estimate from $i=1$ to $m(n)$ yields

$$|\mathbb{E}\{\exp(\mathrm{i}tS_m)\} - \mathbb{E}\{\exp(-\sigma_m^2t^2/2\}| = O\bigg(t\sum_i \mathbb{E}\{|\nu_i^1(t)|\} + \frac{t^2}{2}\sum_{i=1}\mathbb{E}\{|\nu_i^2(t)|\} + Ct^3\sum_i\mathbb{E}\{w_i^3\}\bigg).$$

By assumptions of the theorem, this implies the claim for complex exponential test functions as required.

It remains to show the main estimate (8). By Taylor expansion and boundedness of ε_i

$$\begin{aligned}\exp(\mathrm{i}tS_i) &= \exp(\mathrm{i}tS_{i-1})\exp(\mathrm{i}tw_i\varepsilon_i)\\ &= \exp(\mathrm{i}tS_{i-1})\bigg(1 + \mathrm{i}tw_i\varepsilon_i - \frac{t^2w_i^2\varepsilon_i^2}{2}\bigg) + O\big(t^3|w_i|^3\varepsilon_i^3\big).\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{E}\{\exp(\mathrm{i}tS_i + (\sigma_i^2-\sigma_m^2)t^2/2)|\mathcal{F}_{i-1}\} &= \exp(\mathrm{i}tS_{i-1} + t^2\sigma_i^2/2)\times\\ &\quad \mathbb{E}\bigg\{\bigg(1 + \mathrm{i}tw_i\varepsilon_i - \frac{t^2w_i^2\varepsilon_i^2}{2}\bigg)\exp(-t^2\sigma_m^2/2)\bigg|\mathcal{F}_{i-1}\bigg\} + Ct^3w_i^3\\ &= \exp(\mathrm{i}tS_{i-1} + t^2\sigma_i^2/2)\mathbb{E}\big\{\exp(-t^2(\sigma_m^2+w_i^2)/2)|\mathcal{F}_{i-1}\big\}\\ &\quad + O\bigg(t|\nu_i^1(t)| + \frac{t^2|\nu_i^2(t)|}{2} + t^3w_i^3\bigg),\end{aligned}$$

where we used the fact that $|w_i| \leq 1$ a.s. and that $\exp(-z^2/2) = 1 - z^2/2 + O(z^4)$ for $|z| \leq 1$. Rearranging, we obtain (8).

□

B. Supplementary experiments

In Figure 5 we show the results of applying $\boldsymbol{W}$-decorrelation to the AR(1) process:

$$y_i = \beta_0 y_{i-1} + \varepsilon_i,$$

where $\beta_0 = 0.9$. The process is nearly non-stationary. As evidenced by Figure 5, $\boldsymbol{W}$-decorrelation provides estimates with an order of magnitude lower excess kurtosis and valid empirical coverage. Simulataneously, it provides smaller widths than concentration inequalities, particularly in the moderate confidence regime.

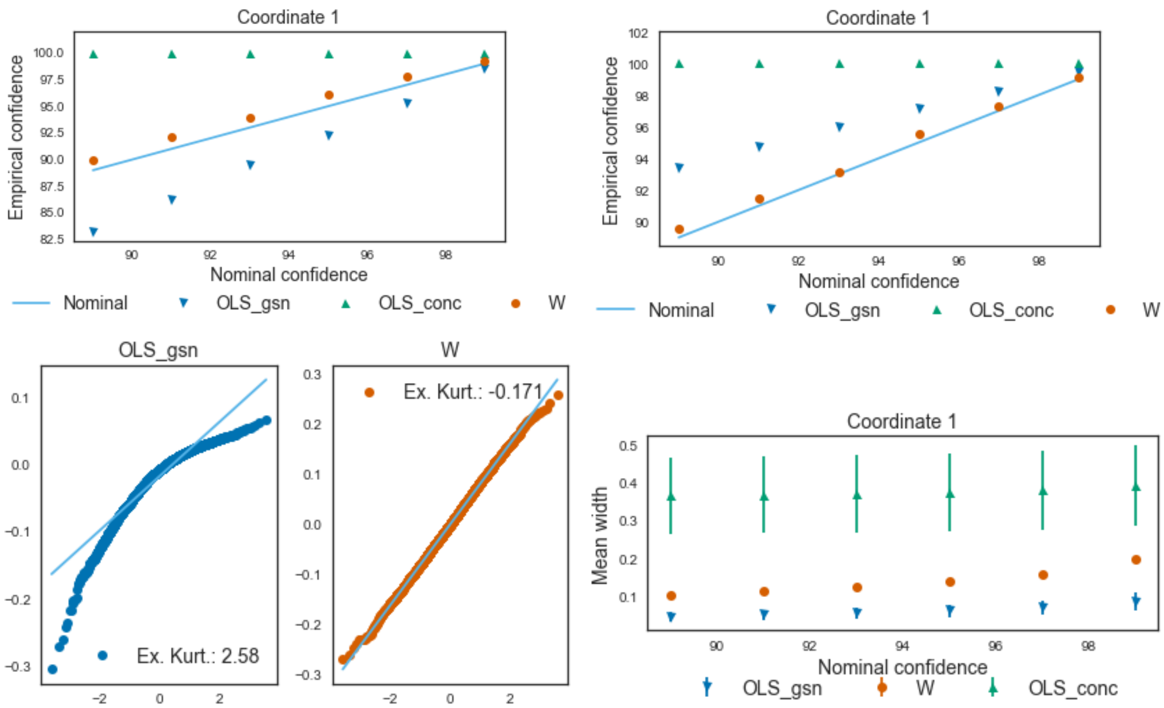

Figure 5. Lower (Top left) and upper (Top right) coverage probabilities for OLS with Gaussian intervals, OLS with concentration inequality intervals, and decorrelated $\boldsymbol{W}$-decorrelated estimate intervals. Note that 'Conc' has always 100% coverage. Bottom right: QQ plot for the distribution of errors of standard OLS estimate and the W-decorrelated estimate. Excess kurtosis is included inset. Right: Mean confidence widths for various estimators. The error bars show one (empirical) standard deviation of the confidence width.