Almost unbiased variance estimation in linear regressions with many covariates

Stanislav Anatolyev*

CERGE-EI, Czech Republic and New Economic School, Russia

May 2018

Abstract

We propose an adjustment to the variance estimator of Cattaneo, Jansson, and Newey (2018) when there are many covariates. The finite-sample correction in the spirit of Horn, Horn and Duncan (1975) makes the estimator exactly unbiased under homoskedasticity. Simulations show that the adjustment reduces test size distortions, especially with skewed regressors. We also verify whether further degrees-of-freedom adjustments in the spirit of Bell and McCaffrey (2002) bring improvements to the control over test size.

Keywords: linear regression, ordinary least squares, variance estimation, many regressor asymptotics

JEL codes: C2, C13

*Address: Stanislav Anatolyev, CERGE-EI, Politických vězňů 7, 11121 Prague 1, Czech Republic. E-mail: stanislav.anatolyev@cerge-ei.cz.

This research was supported by the grant 17-26535S from the Czech Science Foundation. I thank a referee for constructive suggestions.

1 Introduction and setup

Consider the standard linear regression model under random sampling

$$Y = X\beta + W\gamma + U,$$

where we are interested in inference about the $d \times 1$ parameter vector β when facing many covariates in W. That is, while d is asymptotically fixed, the dimensionality p of the nuisance parameter vector γ asymptotically grows, possibly with the same rate, as the sample size n grows. The regression assumption is supposed to hold: $E[U|X, W] = 0$, the data are IID.

Denote $M = I_n - W(W'W)^{-1}W'$. Let $\hat{V} = MX$ be the 'partialled out' regressors of interest, $\hat{\mathbf{v}}_i$ being the i^{th} column of $\hat{V}$, and let

$$\hat{P} = \hat{V}(\hat{V}'\hat{V})^{-1}\hat{V}',$$

be the associated projection matrix. Denote

$$\hat{\Gamma} = \frac{\hat{V}'\hat{V}}{n}.$$

The OLS asymptotic variance estimate is computed as

$$\hat{\Omega} = \hat{\Gamma}^{-1}\hat{\Sigma}\hat{\Gamma}^{-1},$$

where $\hat{\Sigma}$ is an estimate of

$$\Sigma = var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\hat{\mathbf{v}}_i u_i\right).$$

Cattaneo, Jansson, and Newey (2018) (sometimes abbreviated as CJN) show that the traditional Eicker–White variance estimator is inconsistent under the many covariate asymptotics, and propose a variance estimator HC that takes into account numerosity of covariates. We construct an adjustment to this variance estimator, in the spirit of almost unbiased variance estimator of Horn, Horn and Duncan (1975) and termed $HC2$ in the subsequent literature (see, e.g., MacKinnon, 2012 and Imbens and Kolesár, 2016).[1] We too call the adjusted estimator *almost unbiased.*

Furthermore, we verify whether further degrees-of-freedom adjustments in the spirit of Bell and McCaffrey (2002) bring improvements to the control over test size. The idea of the adjustment is to use critical values from the Student's distribution with the degrees of freedom parameter such that, under homoskedasticity, the first two moments of the variance estimator are matched to ones of the associated chi-squared distribution.

[1] The adjusted estimator is sharply different from the $HC2$ estimator considered and deemed inappropriate in Cattaneo, Jansson, and Newey (2018), the one that does not take into account numerosity of covariates.

2 Almost unbiased variance estimator

The general variance estimator is

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

where $\hat{u}_j$ are OLS residuals, elements of the vector $(M-\hat{P})U$. The conditional expectation of $\hat{\Sigma}$ is

$$\begin{aligned} E\left[\hat{\Sigma}|X,W\right] &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_{ij}\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'E\left[\hat{u}_j^2|X,W\right] \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\kappa_{ik}M_{kj}^2\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'E\left[u_j^2|X,W\right]+o(1) \qquad (1)\end{aligned}$$

almost surely. Cattaneo, Jansson, and Newey (2018) set

$$\kappa^{HC} = (M\odot M)^{-1} \qquad (2)$$

to match the leading term with the target conditional expectation

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'E\left[u_i^2|X,W\right]. \qquad (3)$$

Cattaneo, Jansson, and Newey (2018) show that the sufficient condition for positive definiteness of $M\odot M$ is that $\min_{1\le i\le n}\{2M_{ii}-1\}>0$.

We take into consideration the higher-order term in (1) that includes estimation noise in $\hat{\beta}$. Let $e_{m,j}$ stand for the j^{th} column of the $m\times m$ identity matrix. Under the homoskedasticity restriction, for observation j the expectation $E\left[\hat{u}_j^2|X,W\right]$ accounting for this noise is equal to

$$\begin{aligned} E\left[((M-\hat{P})U)_j^2|X,W\right]_{HO} &= E[e_{n,j}'(M-\hat{P})UU'(M-\hat{P})e_{n,j}|X,W]_{HO} \\ &= e_{n,j}'(M-\hat{P})E\left[UU'|X,W\right]_{HO}(M-\hat{P})e_{n,j} \\ &= \sigma^2 e_{n,j}'(M-\hat{P})e_{n,j} \\ &= \sigma^2(M_{jj}-\hat{P}_{jj}),\end{aligned}$$

where $\sigma^2 = E\left[u_i^2\right]$. Hence, the leading term in (1) evaluated under the homoskedasticity restriction yields that

$$\begin{aligned} E\left[\hat{\Sigma}|X,W\right]_{HO} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n}\kappa_{ik}(M_{kk}-\hat{P}_{kk})\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'\sigma^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\kappa_{ik}(M_{kj}^2-\hat{P}_{kj}^2)\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'\sigma^2,\end{aligned}$$

because $M_{kk} = \sum_{j=1}^{n} M_{kj}^2$ and $\hat{P}_{kk} = \sum_{j=1}^{n} \hat{P}_{kj}^2$. Hence, the expectation will exactly match the target conditional expectation (3) under homoskedasticity if, instead of (2), one sets[2]

$$\kappa^{AU} = \left(M \odot M - \hat{P} \odot \hat{P}\right)^{-1}. \tag{4}$$

Note that this κ^{AU} is also symmetric by construction. The sufficient condition for positive definiteness is strengthened to $\min_{1\le i\le n}\{M_{ii}(2M_{ii}-1) - \hat{P}_{ii}\} > 0$. The Appendix demonstrates that the estimator $\hat{\Sigma}^{AU}$ based on κ^{AU} is consistent for Σ under the same conditions when the estimator $\hat{\Sigma}^{HC}$ based on κ^{HC} is consistent when an additional condition is placed on diagonal elements of M and $\hat{P}$.

3 Degrees of freedom adjustment

We also verify whether the use of the Student's distribution instead of normal can further improve size control. The idea of the Bell and McCaffrey (2002) adjustment is to use critical values from the Student's distribution with the degrees of freedom parameter such that, under homoskedasticity, the first two moments of the variance estimator are matched to ones of the associated chi-squared distribution.

We are interested in the decomposition of the p^{th} diagonal element of $\hat{\Omega}^{AU}$:

$$\hat{\Omega}_{pp}^{AU} = \left(\hat{\Gamma}^{-1}\hat{\Sigma}^{AU}\hat{\Gamma}^{-1}\right)_{pp} = e_{d,p}'\hat{\Gamma}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\kappa_{ij}^{AU}\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'\hat{u}_j^2\right)\hat{\Gamma}^{-1}e_{d,p} = \sum_{j=1}^{n}\mu_{p,j}^{AU}\left((M-\hat{P})U\right)_j^2,$$

where

$$\mu_{p,j}^{AU} = e_{d,p}'\hat{\Gamma}^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\kappa_{ij}^{AU}\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'\right)\hat{\Gamma}^{-1}e_{d,p}.$$

Under homoskedastic normal errors U,

$$\hat{\Omega}_{pp}^{AU} = \sigma^2\sum_{j=1}^{n}\lambda_{p,j}^{AU}Z_j,$$

[2] An alternative and more straightforward estimator would result upon noticing that

$$E\left[\hat{\Sigma}|X,W\right] = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\kappa_{ik}(M_{kj}-\hat{P}_{kj})^2\hat{\mathbf{v}}_i\hat{\mathbf{v}}_i'E\left[u_j^2|X,W\right],$$

which, when matched to (3), leads to the solution

$$\kappa = \left((M-\hat{P})\odot(M-\hat{P})\right)^{-1}.$$

However, as simulations show, this results in *worsening* the performance in small samples: the rejection rates experience even higher distortions.

with $\{Z_j\}_{j=1}^n \sim \text{IID}\,\chi^2(1)$ and $\left\{\lambda_{p,j}^{AU}\right\}_{j=1}^n$ eigenvalues of the matrix

$$\sum_{j=1}^n \mu_{p,j}^{AU}(M_j - \hat{P}_j)(M_j - \hat{P}_j)'.$$

The degrees of freedom parameter is computed as (Imbens and Kolesár, 2016)

$$\nu^{AU} = \frac{\left(\sum_{j=1}^n \lambda_{p,j}^{AU}\right)^2}{\sum_{j=1}^n \left(\lambda_{p,j}^{AU}\right)^2}.$$

4 Simulation evidence

We borrow the simulation setup from MacKinnon (2012) and adapt it to the case of many covariates. The model is

$$y_i = \beta_1 + \sum_{k=2}^d \beta_k x_{ik} + \sum_{k=1}^p \gamma_k w_{ik} + u_i,$$

where $\beta_k = 1$ for $1 \leq k \leq d-1$, $\beta_d = 0$, and $\gamma_k = 0$ for $1 \leq k \leq p$; the regular regressors x_{ik} are IID standard lognormal, the nuisance covariates w_{ik} are IID uniform on $[-1, 1]$; the errors are potentially heteroskedastic: $u_i = \sigma_i \varepsilon_i$, with ε_i IID standard normal and

$$\sigma_i = z_\zeta \left(1 + \left|\sum_{k=2}^d \log x_{ik} + \sum_{k=1}^p \left(|w_{ik}| - \tfrac{1}{2}\right)\right|^\zeta\right),$$

where the multiplier z_ζ ensures that the variance of u_i is unity. The parameter ζ indexes the strength of heteroskedasticity: $\zeta = 0$ (homoskedasticity), $\zeta = 1$ (moderate), or $\zeta = 2$ (severe). As long as $\zeta \neq 0$, the conditional variance depends on all regressors and exhibits high variability across i.[3] Note that the presence of skewed and thus of high leverage regressors lends support to both the correction for leverage (MacKinnon, 2012) and the Bell–McCaffrey degrees of freedom adjustment (Imbens and Kolesár, 2016). The coefficient of interest is β_d, and we track rejection rates of the null $H_0 : \beta_d = 0$. We set $d = 5$ as in MacKinnon (2012).

The sample size is $n = 500$ throughout. We set $p = 25 \cdot 2^m$, where $m \in \{0, 1, 2, 3\}$. In Table 1 we present rejection rates for 10%, 5% and 1% size two-sided tests based on 5,000 Monte–Carlo replications. The first test uses weights κ^{HC} in the variance estimate

[3] For example, with $p = 200$, the cross-sectional mean and standard deviation of σ_i are equal to 0.86 and 0.53 when $\zeta = 1$ and to 0.65 and 0.79 when $\zeta = 2$.

and normal critical values, the second test uses weights κ^{AU} and normal critical values, and the third uses weights κ^{AU} and critical values of Student's with ν^{AU} degrees of freedom.[4]

Clearly, the HC estimator in all cases experiences overrejection which is pretty stable for different degrees of covariate numerosity. The distortions go slightly down or slightly up with severity of heteroskedasticity depending on the nominal size. The AU estimator reduces the size distortions in all cases, sometimes a fewfold, more so under strong heteroskedasticity. The tests based on the AU variance estimator and Student's distribution exhibits slight underrejection, but usually of a smaller degree than that of overrejection by the AU estimator. The underrejection is a bit stronger when heteroskedasticity is stronger, but not significantly.

Finally, we take a closer look at possible sources of distortions and compare the cases of skewed (lognormal) and symmetrically distributed (normal) regressors. Within the same simulation setup, we change the distribution of x_{ik} to IID standard normal, set $p = 1$, and shut heteroskedasticity down. Table 2 reports means and standard deviations of the two estimates of Σ, as well as actual rejection rates. One can see that with the skewed regressors the HC estimator is significantly downward biased while the AU is almost unbiased, though its standard deviation is smaller.[5] The situation straightens in the case of normal regressors – the HC is almost unbiased, the AU is unbiased, and both have equal variability. This results in actual rejection rates being close to the nominal.

5 Conclusion

We have proposed and analyzed a finite-sample adjustment to the variance estimator of Cattaneo, Jansson, and Newey (2018) in the spirit of Horn, Horn and Duncan (1975). The almost unbiased estimator is constructed to be exactly unbiased under homoskedasticity. Simulations show that overrejection is reduced by the adjustment when regressors are skewed. We also find out that a degrees-of-freedom adjustment in the spirit of Bell and McCaffrey (2002) can bring further improvements to the control over test size, though it may lead to slight underrejection.

[4] In case any variance estimate turns out negative, we set it equal to 10^{-5}, essentially forcing rejection of the null; this is a reasonable adjustment as a negative variance estimate signals of a very small positive population variance. This never happens with $n = 500$, while in additional experiments with $n = 100$ such occurrences are extremely rare tending to be higher for HC than for AU.

[5] In this design, the straightforward estimator described in footnote 2 has mean 7.43 and very big variance.

References

Bell, R. M. and D. F. McCaffrey (2002): "Bias reduction in standard errors for linear regression with multi-stage samples," *Survey Methodology*, 28(2):169–181.

Cattaneo, M. D., M. Jansson, and W. K. Newey (2018): "Inference in linear regression models with many covariates and heteroskedasticity," *Journal of American Statistical Association*, forthcoming.

Horn, S. D., R. A. Horn and D. B. Duncan (1975): "Estimating heteroskedastic variances in linear models," *Journal of the American Statistical Association*, 70, 380–385.

Imbens, G. W. and M. Kolesár (2016): "Robust standard errors in small samples: Some practical advice," *Review of Economics and Statistics*, 98, 701–712.

MacKinnon, J. G. (2012): "Thirty years of heteroskedasticity-robust inference." In: Chen, X. and N. R. Swanson, eds., *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, 437–461, Springer, New York.

Varah, J. M. (1975): "A lower bound for the smallest singular value of a matrix," *Linear Algebra and its Applications*, 11(1), 3–5.

Appendix

We show that under the conditions set forth in CJN, $||\hat{\Sigma}^{AU} - \hat{\Sigma}^{HC}|| \leq o_p(1)$, which implies that $\hat{\Sigma}^{AU}$ is consistent for Σ if $\hat{\Sigma}^{HC}$ is consistent. In particular, we too impose that $\max_{1\leq i\leq n} \|\hat{\mathbf{v}}_i\| / \sqrt{n} = o_p(1)$, see Section 4.2 of the supplemental appendix to CJN for its discussion. In addition, we impose that $\mathcal{P} \equiv \min_{1\leq i\leq n}\{M_{ii}(2M_{ii}-1) - \hat{P}_{ii}\} > 0$ and $\mathcal{P}^{-1} = O_p(1)$. Then, along the lines of the supplemental appendix to CJN (section 3),

$$\begin{aligned}
\lambda_{\min}(M \odot M - \hat{P} \odot \hat{P}) &\geq \min_{1\leq i\leq n}\{M_{ii}^2 - \hat{P}_{ii}^2 - \sum_{1\leq j\leq n, j\neq i} |M_{ij}^2 - \hat{P}_{ij}^2|\} \\
&\geq \min_{1\leq i\leq n}\{M_{ii}^2 - \hat{P}_{ii}^2 - \sum_{1\leq j\leq n, j\neq i} M_{ij}^2 - \sum_{1\leq j\leq n, j\neq i} \hat{P}_{ij}^2\} \\
&= \min_{1\leq i\leq n}\{M_{ii}^2 - \hat{P}_{ii}^2 - (M_{ii} - M_{ii}^2) - (\hat{P}_{ii} - \hat{P}_{ij}^2)\} \\
&= \min_{1\leq i\leq n}\{M_{ii}(2M_{ii}-1) - \hat{P}_{ii}\} > 0,
\end{aligned}$$

and, also using Theorem 1 of Varah (1975),

$$\left\|\kappa^{AU}\right\|_\infty < \left(\min_{1\leq i\leq n}\left\{M_{ii}^2 - \hat{P}_{ii}^2 - \sum_{1\leq j\leq n, j\neq i} |M_{ij}^2 - \hat{P}_{ij}^2|\right\}\right)^{-1} \leq \mathcal{P}^{-1} = O_p(1).$$

Under the conditions specified in CJN, we have that (CJN, Theorem 3)

$$\hat{\Sigma}^{HC} = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \kappa_{ik}^{HC} M_{kj}^2 \hat{\mathbf{v}}_i\hat{\mathbf{v}}_i' E\left[u_j^2|X,W\right] + o_p(1)$$

and

$$\hat{\Sigma}^{AU} = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \kappa_{ik}^{AU} M_{kj}^2 \hat{\mathbf{v}}_i\hat{\mathbf{v}}_i' E\left[u_j^2|X,W\right] + o_p(1).$$

Because κ^{AU} defined by (4) solves $\sum_{k=1}^n \kappa_{ik}^{AU}(M_{kj}^2 - \hat{P}_{kj}^2) = \mathbb{I}_{\{i=j\}}$, and κ^{HC} defined by (2) solves $\sum_{k=1}^n \kappa_{ik}^{HC} M_{kj}^2 = \mathbb{I}_{\{i=j\}}$, we have $\sum_{k=1}^n \left(\kappa_{ik}^{AU} - \kappa_{ik}^{HC}\right) M_{kj}^2 = \sum_{k=1}^n \kappa_{ik}^{AU}\hat{P}_{kj}^2$. Then,

$$\hat{\Sigma}^{AU} - \hat{\Sigma}^{HC} = \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n \kappa_{ik}^{AU} \hat{P}_{kj}^2 \hat{\mathbf{v}}_i\hat{\mathbf{v}}_i' E\left[u_j^2|X,W\right] + o_p(1).$$

Denote $C_{\sigma^2} = \max_{1\leq i\leq n} E\left[u_i^2|X,W\right]$ which is $O_p(1)$ under the conditions specified in CJN. Without loss of generality, following CJN we let $d=1$. Using that $\sum_{k=1}^n \hat{P}_{kk} = d$, we obtain

$$\begin{aligned}
\left|\hat{\Sigma}^{AU} - \hat{\Sigma}^{HC}\right| &\leq \frac{C_{\sigma^2}}{n}\left(\max_{1\leq i\leq n} \hat{v}_i^2\right)\sum_{i=1}^n\sum_{k=1}^n \left|\kappa_{ik}^{AU}\right| \sum_{j=1}^n \hat{P}_{kj}^2 + o_p(1) \\
&\leq C_{\sigma^2}\left(\frac{\max_{1\leq i\leq n}|\hat{v}_i|}{\sqrt{n}}\right)^2 \left\|\kappa^{AU}\right\|_\infty \sum_{k=1}^n \hat{P}_{kk} + o_p(1) \\
&= O_p(1)o_p(1)^2 O_p(1) + o_p(1) = o_p(1).
\end{aligned}$$

	based on κ^{HC}			based on κ^{AU}			based on κ^{AU} & $t_{\nu^{AU}}$		
index ζ	10%	5%	1%	10%	5%	1%	10%	5%	1%
				$p = 25$					
0	13.4%	8.1%	2.6%	11.9%	6.6%	1.9%	9.7%	4.8%	0.9%
1	13.9%	7.9%	2.6%	12.3%	6.7%	1.7%	10.4%	5.1%	0.9%
2	13.1%	7.5%	2.2%	11.1%	5.9%	1.5%	9.5%	4.3%	0.8%
				$p = 50$					
0	13.0%	7.3%	2.2%	11.4%	6.1%	1.5%	9.5%	4.5%	0.5%
1	13.7%	8.0%	2.1%	12.2%	6.7%	1.5%	10.1%	4.8%	0.6%
2	13.4%	7.3%	1.7%	11.7%	5.8%	1.1%	9.9%	4.1%	0.5%
				$p = 100$					
0	12.9%	7.2%	2.1%	11.1%	5.7%	1.5%	8.9%	4.2%	0.9%
1	13.3%	7.2%	1.7%	11.2%	5.7%	1.2%	9.1%	4.1%	0.6%
2	12.5%	6.7%	1.8%	10.7%	5.3%	1.2%	8.8%	3.9%	0.5%
				$p = 200$					
0	13.2%	8.1%	2.6%	11.7%	6.4%	1.9%	9.9%	4.6%	0.9%
1	13.4%	7.7%	2.7%	11.5%	6.3%	1.9%	9.5%	4.5%	0.8%
2	13.6%	7.8%	2.0%	11.5%	6.0%	1.6%	9.6%	4.5%	0.5%

Table 1. Actual rejection rates, from simulations.

	based on κ^{HC}	based on κ^{AU}
Lognormal regressors, true variance equals 7.39		
mean	6.77	7.36
standard deviation	2.43	3.35
10% rejection rate	13.5%	12.0%
5% rejection rate	7.5%	6.4%
1% rejection rate	2.1%	1.5%
Normal regressors, true variance equals 1.00		
mean	0.98	1.00
standard deviation	0.12	0.12
10% rejection rate	10.2%	9.9%
5% rejection rate	5.2%	5.1%
1% rejection rate	1.2%	1.1%

Table 2. Actual rejection rates and other performance measures, from simulations.