View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

TECHNICAL WORKING PAPER SERIES

HETEROSKEDASTICITY-ROBUST STANDARD ERRORS FOR FIXED EFFECTS PANEL DATA REGRESSION

James H. Stock
Mark W. Watson

Technical Working Paper 323
http://www.nber.org/papers/T0323

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2006

We thank Alberto Abadie, Gary Chamberlain, Doug Staiger, and Hal White for helpful discussions and/or comments and Anna Mikusheva for research assistance. This research was supported in part by NSF grant SBR-0214131. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

©2006 by James H. Stock and Mark W. Watson. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression
James H. Stock and Mark W. Watson
NBER Technical Working Paper No. 323
June 2006
JEL No. C23, C12

ABSTRACT

The conventional heteroskedasticity-robust (HR) variance matrix estimator for cross-sectional regression (with or without a degrees of freedom adjustment), applied to the fixed effects estimator for panel data with serially uncorrelated errors, is inconsistent if the number of time periods T is fixed (and greater than two) as the number of entities n increases. We provide a bias-adjusted HR estimator that is $(nT)^{1/2}$ -consistent under any sequences (n, T) in which n and/or T increase to ∞.

James H. Stock
Department of Economics
Harvard University
Littauer Center M4
Cambridge, MA 02138
and NBER
james_stock@harvard.edu

Mark W. Watson
Department of Economics
Princeton University
Princeton, NJ 08544-1013
and NBER
mwatson@princeton.edu

1. Model and Theoretical Results

Consider the fixed effects regression model,

$$Y_{it} = \alpha_i + \beta' X_{it} + u_{it},\ i = 1,\ldots, n,\ t = 1,\ldots, T \tag{1}$$

where X_{it} is a $k\times 1$ vector of regressors and where (X_{it}, u_{it}) satisfy:

Heteroskedastic panel data model with conditionally uncorrelated errors

1. $(X_{i1},\ldots, X_{iT}, u_{i1},\ldots, u_{iT})$ are i.i.d. over $i = 1,\ldots, n$ (i.i.d. over entities),
2. $E(u_{it}|X_{i1},\ldots, X_{iT}) = 0$ (strict exogeneity)
3. $Q_{\tilde{X}\tilde{X}} \equiv ET^{-1}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'$ is nonsingular (no perfect multicollinearity), and
4. $E(u_{it}u_{is}|\ X_{i1},\ldots, X_{iT}) = 0$ for $t \neq s$ (conditionally serially uncorrelated errors).

For the asymptotic results we will further assume:

Stationarity and moment condition

5. (X_{it}, u_{it}) is stationary and has absolutely summable cumulants up to order twelve.

The fixed effects estimator is,

$$\hat{\beta}_{FE} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{Y}_{it} \tag{2}$$

where the superscript "~" over variables denotes deviations from entity means ($\tilde{X}_{it} = X_{it} - T^{-1}\sum_{s=1}^{T} X_{is}$, etc.). The asymptotic distribution of $\hat{\beta}_{FE}$ is [e.g. Arrelano (2003)]

$$\sqrt{nT}\,(\hat{\beta}_{FE} - \beta) \xrightarrow{d} \mathrm{N}(0,\ Q_{\tilde{X}\tilde{X}}^{-1}\,\Sigma\, Q_{\tilde{X}\tilde{X}}^{-1}), \text{ where } \Sigma = \frac{1}{T}\sum_{t=1}^{T} E\left(\tilde{X}_{it}\tilde{X}_{it}' u_{it}^{2}\right). \tag{3}$$

The variance of the asymptotic distribution in (3) is estimated by $\hat{Q}_{\tilde{X}\tilde{X}}^{-1}\hat{\Sigma}\hat{Q}_{\tilde{X}\tilde{X}}^{-1}$, where $\hat{Q}_{\tilde{X}\tilde{X}} = (nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'$ and $\hat{\Sigma}$ is a heteroskedasticity-robust (HR) covariance matrix estimator.

A frequently used HR estimator of Σ is

$$\hat{\Sigma}^{HR-XS} = \frac{1}{nT - n - k}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\hat{\tilde{u}}_{it}^{2} \tag{4}$$

where $\{\hat{\tilde{u}}_{it}\}$ are the fixed-effects regression residuals, $\hat{\tilde{u}}_{it} = \tilde{u}_{it} - (\hat{\beta}_{FE} - \beta)'\tilde{X}_{it}$.[2]

Although $\hat{\Sigma}^{HR-XS}$ is consistent in cross-section regression [White (1980)], it turns out to be inconsistent in panel data regression with fixed *T*. Specifically, an implication of the results in the appendix is that, under fixed-*T* asymptotics with *T* > 2,

[2] For example, $\hat{\Sigma}^{HR-XS}$ is the estimator used in STATA and Eviews.

$$\hat{\Sigma}^{HR-XS} \xrightarrow{p}_{(n\to\infty,\ T\ fixed)} \Sigma + \frac{1}{T-1}(B-\Sigma)\text{, where } B = E\left[\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)\left(\frac{1}{T}\sum_{s=1}^{T}u_{is}^{2}\right)\right]. \quad (5)$$

The expression for B in (5) suggests the bias-adjusted estimator,

$$\hat{\Sigma}^{HR-FE} = \left(\frac{T-1}{T-2}\right)\left(\hat{\Sigma}^{HR-XS} - \frac{1}{T-1}\hat{B}\right),$$

$$\text{where } \hat{B} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}\hat{\tilde{u}}_{is}^{2}\right) \quad (6)$$

where the estimator is defined for $T > 2$.

It is shown in the appendix that, if assumptions 1-5 hold, then under any sequence (n, T) in which $n \to \infty$ and/or $T \to \infty$ (which includes the cases of n fixed or T fixed),

$$\hat{\Sigma}^{HR-FE} = \Sigma + O_p(1/\sqrt{nT}) \quad (7)$$

so the problematic bias term of order T^{-1} is eliminated if $\hat{\Sigma}^{HR-FE}$ is used.

Remarks

1. The bias arises because the entity means are not consistently estimated when T is fixed, so the usual step of replacing estimated regression coefficients with their probability limits is inapplicable. This can be seen by considering

$$\tilde{\Sigma}^{HR-XS} \equiv \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\tilde{u}_{it}^2, \tag{8}$$

which is the infeasible version of $\hat{\Sigma}^{HR-XS}$ in which β is treated as known and the degrees-of-freedom correction k is omitted. The bias calculation is short:

$$\begin{aligned}
E\tilde{\Sigma}^{HR-XS} &= E\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\left(u_{it}-\frac{1}{T}\sum_{s=1}^{T}u_{is}\right)^2 \\
&= \frac{1}{T-1}E\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'u_{it}^2 - \frac{2}{T(T-1)}E\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'u_{it}u_{is} + \frac{1}{T^2(T-1)}E\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{r=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'u_{is}u_{ir} \\
&= \left(\frac{T-2}{T-1}\right)\frac{1}{T}\sum_{t=1}^{T}E\left(\tilde{X}_{it}\tilde{X}_{it}'u_{it}^2\right) + \frac{1}{T^2(T-1)}E\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'u_{is}^2 \\
&= \left(\frac{T-2}{T-1}\right)\Sigma + \frac{1}{T-1}B,
\end{aligned} \tag{9}$$

where the third equality uses the assumption $E(u_{it}u_{is}|\, X_{i1},\ldots,X_{iT}) = 0$ for $t \neq s$; rearranging the final expression in (9) yields the plim in (5). The source of the bias is the final two terms in the second line of (9), both of which appear because of estimating the entity means. The problems created by the entity means is an example of the general problem of having increasingly many incidental parameters.

2. The asymptotic bias in $\hat{\Sigma}^{HR-XS}$ is $O(1/T)$. An implication of the calculations in the appendix is that var($\hat{\Sigma}^{HR-XS}$) = $O(1/nT)$, so MSE($\hat{\Sigma}^{HR-XS}$) = $O(1/T^2) + O(1/nT)$.
3. In general, $B - \Sigma$ is neither positive nor negative semidefinite, so standard errors computed using $\hat{\Sigma}^{HR-XS}$ can in general either be too large or too small.

4. If (X_{it}, u_{it}) are i.i.d. over t as well as over i, then the asymptotic bias in $\hat{\Sigma}^{HR-XS}$ is proportional to the asymptotic bias in the homoskedasticity-only estimator, $\hat{\Sigma}^{homosk} = \hat{Q}_{\tilde{X}\tilde{X}}\hat{\sigma}_u^2$, where $\hat{\sigma}_u^2 = (nT-n-k)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\hat{\tilde{u}}_{it}^2$. Specifically, plim$(\hat{\Sigma}^{HR-XS} - \Sigma) = b_T$plim$(\hat{\Sigma}^{homosk} - \Sigma)$, where $b_T = (T-2)/(T-1)^2$. In this sense, $\hat{\Sigma}^{HR-XS}$ undercorrects for heteroskedasticity.

5. One case in which $\hat{\Sigma}^{HR-XS} \xrightarrow{p} \Sigma$ is when $T = 2$, in which case the fixed effects estimator and $\hat{\Sigma}^{HR-XS}$ are equivalent to the estimator and HR variance matrix computed using first-differences of the data (suppressing the intercept).

6. Another case in which $\hat{\Sigma}^{HR-XS}$ is consistent is when the errors are homoskedastic: if $E(u_{it}^2|X_{i1},\ldots,X_{iT}) = \sigma_u^2$, then $B = \Sigma = Q_{\tilde{X}\tilde{X}}\ \sigma_u^2$.

7. Another estimator of Σ is the clustered (over entities) variance estimator,

$$\hat{\Sigma}^{cluster} = \frac{1}{nT}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\tilde{X}_{it}\hat{\tilde{u}}_{it}\right)\left(\sum_{s=1}^{T}\tilde{X}_{is}\hat{\tilde{u}}_{is}\right)' \qquad (10)$$

If $T = 3$, then the infeasible version of $\hat{\Sigma}^{HR-FE}$ (in which β is known) equals the infeasible version of $\hat{\Sigma}^{cluster}$, and $\hat{\Sigma}^{HR-FE}$ is asymptotically equivalent to $\hat{\Sigma}^{cluster}$ to order $1/\sqrt{n}$; but for $T > 3$, $\hat{\Sigma}^{cluster}$ and $\hat{\Sigma}^{HR-FE}$ differ. Interestingly, the problem of no consistent estimation of the entity means does not affect the clustered variance estimator for any value of T because of the (idempotent matrix) identity $\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it} = \sum_{t=1}^{T}\tilde{X}_{it}u_{it}$. This identity does not hold in general for heteroskedasticity- and autocorrelation-consistent (HAC) kernel estimators of Σ, rather it arises as a special case for the untruncated rectangular kernel used in the cluster variance estimator. Thus the means-estimation problem discussed above for $\hat{\Sigma}^{HR-XS}$ seems likely to arise for HAC panel data estimators other than $\hat{\Sigma}^{cluster}$.

8. Under general (n, T) sequences (n and/or $T \to \infty$), $\hat{\Sigma}^{cluster} = \Sigma + O_p(1/\sqrt{n})$ [Hansen (2005)]. Because $\hat{\Sigma}^{HR-FE} = \Sigma + O_p(1/\sqrt{nT})$, if the errors are conditionally serially uncorrelated and T is moderate or large then $\hat{\Sigma}^{HR-FE}$ will be more efficient than $\hat{\Sigma}^{cluster}$.

9. The assumption of 12 absolutely summable cumulants, which is used in the proof of the $\sqrt{nT}$-consistency of $\hat{\Sigma}^{HR-FE}$, is stronger than needed to justify HR variance estimation in cross-sectional data or HAC estimation in time series data. In the proof in the appendix, this stronger assumption arises because the number of nuisance parameters (entity means) is increasing when $n \to \infty$. Under T fixed, $n \to \infty$ asymptotics, stationarity and summable cumulants are unnecessary and assumption 5 can be replaced by $EX_{it}^{12} < \infty$ and $Eu_{it}^{12} < \infty$, $t = 1,\ldots,$ T.

10. As written, $\hat{\Sigma}^{HR-FE}$ is not guaranteed to be positive semi-definite (psd). Asymptotically equivalent psd estimators can be constructed in a number of standard ways. For example if the spectral decomposition of $\hat{\Sigma}^{HR-FE}$ is $Q'\Lambda Q$, then $\hat{\Sigma}_{psd}^{HR-FE} = Q'|\Lambda|Q$ is psd.

11. These results should extend to IV panel data regression with heteroskedasticity, albeit with different formulas.

2. Monte Carlo Results

A small Monte Carlo study was performed to assess the quantitative importance of the bias in $\hat{\Sigma}^{HR-XS}$ and the relative MSEs of the variance estimators. The design has a single regressor and Gaussian errors:

$$y_{it} = x_{it}\beta + u_{it} \tag{11}$$

$$x_{it} \sim \text{i.i.d. N}(0,1) \tag{12}$$

$$u_{it}|x_i \sim \text{i.n.i.d. N}(0, \sigma_{it}^2),\ \sigma_{it}^2 = \lambda(0.1 + x_{it}^2)^{\kappa}, \tag{13}$$

where $\kappa = \pm 1$ and λ is chosen so that the unconditional variance of u_{it} is 1. The variance estimators considered are $\hat{\Sigma}^{HR-XS}$ (given in (4)), $\hat{\Sigma}^{HR-FE}$ (given in (6)), and $\hat{\Sigma}^{cluster}$ (given in (10)).

The results, which are based on 20,000 Monte Carlo draws, are summarized in Table 1(a) (for $\kappa = 1$) and 1(b) (for $\kappa = -1$). The first three columns of results report the bias of the three estimators, relative to the true value of Σ (e.g., $E[\hat{\Sigma}^{HR-XS} - \Sigma]/\Sigma$). The next three columns report their MSEs, relative to the MSE of the infeasible HR estimator $\hat{\Sigma}^{inf} = (nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'u_{it}^2$ that could be constructed if the true errors were observed. The final three columns report the size of the 10% two-sided test of $\beta = \beta_0$ based on the t-statistic using the indicated variance estimator and the asymptotic normal critical value. Several results are noteworthy.

First, the bias in $\hat{\Sigma}^{HR-XS}$ can be large, it persists as n increases with T fixed, and it can be positive or negative depending on the design. For example, with $T = 5$, and $n = 1000$, the relative bias of $\hat{\Sigma}^{HR-XS}$ is –11.2% when $\kappa = 1$ and is 31% when $\kappa = -1$.

Second, a large bias in $\hat{\Sigma}^{HR-XS}$ can result in a very large relative MSE. Interestingly, in some cases with small n and T and $\kappa = 1$, the MSE of $\hat{\Sigma}^{HR-XS}$ is less than the MSE the infeasible estimator, apparently reflecting a bias-variance tradeoff.

Third, the bias correction in $\hat{\Sigma}^{HR-FE}$ does its job: the relative bias of $\hat{\Sigma}^{HR-FE}$ is less than 3% in all cases with $n \geq 100$, and in most cases the MSE of $\hat{\Sigma}^{HR-FE}$ is very close to the MSE of the infeasible HR estimator.

Fourth, consistent with remark 8, the ratio of the MSE of the cluster variance estimator to the infeasible estimator depends on T and does not converge to 1 as n gets large for fixed T. The MSE of the cluster estimator considerably exceeds the MSE of $\hat{\Sigma}^{HR-FE}$ when T is moderate or large, regardless of n.

Fifth, although the focus of this note has been bias and MSE, one would suspect that the variance estimators with less bias would produce tests with better size. Table 1 is consistent with this conjecture: When $\hat{\Sigma}^{HR-XS}$ is biased up, the t-tests reject too infrequently, and when $\hat{\Sigma}^{HR-XS}$ is biased down, the t-tests reject too often. When T is small, the magnitudes of these size distortions can be considerable: for $T = 3$ and $n = 1000$, the size of the nominal 10% test is 13.0% for $\kappa = 1$ and is 6.2% when $\kappa = -1$. In contrast, in all cases with $n \geq 500$, the other two variance estimators produce tests with sizes that are within Monte Carlo error of 10%. In more complicated designs, the size distortions of tests based on $\hat{\Sigma}^{HR-XS}$ are even larger than reported in Table 1.

3. Conclusions

Our theoretical results and Monte Carlo simulations, combined with the results in Hansen (2005), suggest the following advice for empirical practice. The usual estimator $\hat{\Sigma}^{HR-XS}$ can be used if $T = 2$ but it should not be used if $T > 2$. If $T = 3$, $\hat{\Sigma}^{HR-FE}$ and $\hat{\Sigma}^{cluster}$ are asymptotically equivalent and either can be used. If $T > 3$ and there are good reasons to believe that u_{it} is conditionally serially uncorrelated, then $\hat{\Sigma}^{HR-FE}$ will be more efficient than $\hat{\Sigma}^{cluster}$, so $\hat{\Sigma}^{HR-FE}$ should be used. If, however, serially correlated errors are a possibility – as they are in many applications – then $\hat{\Sigma}^{cluster}$ should be used in conjunction with t_n or $F_{\cdot,n}$ critical values for hypothesis tests on β [see Hansen (2005)].

References

Arrelano, M. (2003). *Panel Data Econometrics*, Oxford: Oxford University Press.

Brillinger, D. (1981). *Time Series Data Analysis and Theory*. San Francisco: Holden-Day.

Hansen, C. (2005). "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data with *T* is Large," manuscript, Graduate School of Business, University of Chicago.

Leonov, V.P. and Shiryaev, A.N. (1959). "On a Method of Calculation of Semi-Invariants." *Theoretical Probability and its Applications*, 4, 319-329.

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

Appendix: Proof of (7)

All limits in this appendix hold for any nondecreasing sequence (n, T) in which $n \to \infty$ and/or $T \to \infty$. To simplify the calculations, we consider the special case that X_{it} is a scalar. Without loss of generality, let $EX_{it} = 0$. Adopt the notation $\bar{u}_i = T^{-1}\sum_{t=1}^{T} u_{it}$ and $\bar{X}_i = T^{-1}\sum_{t=1}^{T} X_{it}$. The proof repeatedly uses the inequality $\text{var}\left(\sum_{j=1}^{m} a_j\right) \le \left(\sum_{j=1}^{m}\sqrt{\text{var}(a_j)}\right)^2$.

Begin by writing $\sqrt{nT}\,(\hat{\Sigma}^{HR-FE} - \Sigma)$ as the sum of four terms using (6) and (9):

$$\sqrt{nT}\,(\hat{\Sigma}^{HR-FE} - \Sigma) = \sqrt{nT}\left[\left(\frac{T-1}{T-2}\right)\left(\hat{\Sigma}^{HR-XS} - \frac{1}{T-1}\hat{B}\right) - \left(\frac{T-1}{T-2}\right)\left(E\tilde{\Sigma}^{HR-XS} - \frac{1}{T-1}B\right)\right]$$

$$= \left(\frac{T-1}{T-2}\right)\sqrt{nT}\left(\hat{\Sigma}^{HR-XS} - E\tilde{\Sigma}^{HR-XS}\right) - \frac{\sqrt{nT}}{T-2}\left(\hat{B} - B\right)$$

$$= \left(\frac{T-1}{T-2}\right)\left[\sqrt{nT}\left(\hat{\Sigma}^{HR-XS} - \tilde{\Sigma}^{HR-XS}\right) + \sqrt{nT}\left(\tilde{\Sigma}^{HR-XS} - E\tilde{\Sigma}^{HR-XS}\right)\right]$$

$$-\left(\frac{T}{T-2}\right)\left[\sqrt{\frac{n}{T}}\left(\hat{B} - \tilde{B}\right) + \sqrt{\frac{n}{T}}\left(\tilde{B} - B\right)\right] \qquad (14)$$

where $\tilde{\Sigma}^{HR-XS}$ is given in (8) and $\tilde{B}$ is $\hat{B}$ given in (6) with $\hat{\tilde{u}}_{it}$ replaced by $\tilde{u}_{it}$.

The proof of (7) proceeds by showing that, under the stated moment conditions,

(a) $\sqrt{nT}\left(\tilde{\Sigma}^{HR-XS} - E\tilde{\Sigma}^{HR-XS}\right) = O_p(1)$,

(b) $\sqrt{n/T}\left(\tilde{B}-B\right)$ = $O_p(1/\sqrt{T}\ )$,

(c) $\sqrt{nT}\left(\hat{\Sigma}^{HR-XS}-\tilde{\Sigma}^{HR-XS}\right)\xrightarrow{p}0$,

(d) $\sqrt{n/T}\left(\hat{B}-\tilde{B}\right)\xrightarrow{p}0$.

Substitution of (a) – (d) into (14) yields $\sqrt{nT}\ (\hat{\Sigma}^{HR-FE}\ -\Sigma) = O_p(1)$ and thus the result (7)

.

(a) From (8), we have that

$$\text{var}\left[\sqrt{nT}\left(\tilde{\Sigma}^{HR-XS}-E\tilde{\Sigma}^{HR-XS}\right)\right] = \text{var}\left[\left(\frac{T}{T-1}\right)\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\tilde{X}_{it}^{2}\tilde{u}_{it}^{2}-E\tilde{X}_{it}^{2}\tilde{u}_{it}^{2}\right)\right]$$

$$=\left(\frac{T}{T-1}\right)^{2}\text{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{X}_{it}^{2}\tilde{u}_{it}^{2}\right)$$

so (a) follows if it can be shown that $\text{var}\left(T^{-1/2}\sum_{t=1}^{T}\tilde{X}_{it}^{2}\tilde{u}_{it}^{2}\right)$ = $O(1)$. Expanding

$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{X}_{it}^{2}\tilde{u}_{it}^{2}$ yields:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{X}_{it}^{2}\tilde{u}_{it}^{2} = A_0-2A_1D_3+\frac{1}{\sqrt{T}}\left(A_1^2D_2+A_2^2D_1-2A_2A_4\right)+\frac{4}{T}A_1A_2A_3-\frac{3}{T^{3/2}}A_1^2A_2^2$$

where $A_0 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{it}^2 u_{it}^2$, $A_1 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{it}$, $A_2 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} u_{it}$, $A_3 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{it}u_{it}$, $A_4 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{it}^2 u_{it}$, $D_1 = \frac{1}{T}\sum_{t=1}^{T} X_{it}^2$, $D_2 = \frac{1}{T}\sum_{t=1}^{T} u_{it}^2$, and $D_3 = \frac{1}{T}\sum_{t=1}^{T} X_{it}u_{it}^2$. Thus

$$\text{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\tilde{X}_{it}^2\tilde{u}_{it}^2\right) \le \{\text{var}(A_0)^{1/2} + 2\text{var}(A_1D_3)^{1/2} + T^{-1/2}\text{var}(A_1^2D_2)^{1/2}$$

$$+ T^{-1/2}\text{var}(A_2^2D_1)^{1/2} + 2T^{-1/2}\text{var}(A_2A_4)^{1/2}$$

$$+ 4T^{-1}\text{var}(A_1A_2A_3)^{1/2} + 3T^{-3/2}\text{var}(A_1^2A_2^2)^{1/2}\}^2$$

$$\le \left\{\text{var}(A_0)^{1/2} + 2\left(EA_1^4ED_3^4\right)^{1/4} + T^{-1/2}\left(EA_1^8ED_2^4\right)^{1/4} + T^{-1/2}\left(EA_2^8ED_1^4\right)^{1/4}\right.$$

$$\left. + 2T^{-1/2}\left(EA_2^4EA_4^4\right)^{1/4} + 4T^{-1}\left(EA_1^8EA_2^8\right)^{1/8}\left(EA_3^4\right)^{1/4} + 3T^{-3/2}\left(EA_1^8EA_2^8\right)^{1/4}\right\}^2 \quad (15)$$

where the second inequality uses term-by-term inequalities, for example the second term in the final expression obtains using $\text{var}(A_1D_3) \le EA_1^2D_3^2 \le \left(EA_1^4ED_3^4\right)^{1/2}$. Thus a sufficient condition for $\text{var}\left(T^{-1/2}\sum_{t=1}^{T}\tilde{X}_{it}^2\tilde{u}_{it}^2\right)$ to be $O(1)$ is that $\text{var}(A_0)$, EA_1^8, EA_2^8, EA_3^4, EA_4^4, ED_1^4, ED_2^4, and ED_3^4 all are $O(1)$.

First consider the D terms. Because $ED_1^4 \le EX_{it}^8$, $ED_2^4 \le Eu_{it}^8$, and (by Hölder's inequality) $ED_3^4 \le EX_{it}^4u_{it}^8 \le \left(EX_{it}^{12}\right)^{1/3}\left(Eu_{it}^{12}\right)^{2/3}$, under assumption 5 all the D moments in (15) are $O(1)$.

For the remainder of the proof of (a), drop the subscript i. Now turn to the A terms, starting with A_1. Because X_t (X_{it}) has mean zero and absolutely summable eighth cumulants,

$$EA_1^8 = E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_t\right)^8 \le h_8\left(\sum_{j=-\infty}^{\infty}\left|\mathrm{cov}(X_t, X_{t-j})\right|\right)^4 + O(T^{-1}) = O(1)$$

where h_8 is the eighth moment of a standard normal random variable.[3] The same argument applied to u_t yields $EA_2^8 = O(1)$.

Now consider A_3 and let $\xi_t = X_t u_t$. Then

$$\begin{aligned}
EA_3^4 &= E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_t\right)^4 = \frac{1}{T^2}\sum_{t_1,\ldots,t_4=1}^{T} E\xi_{t_1}\xi_{t_2}\xi_{t_3}\xi_{t_4} \\
&= 3\left[\frac{1}{T}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\mathrm{cov}(\xi_{t_1},\xi_{t_2})\right]^2 + \frac{1}{T^2}\sum_{t_1,\ldots,t_4=1}^{T}\mathrm{cum}(\xi_{t_1},\xi_{t_2},\xi_{t_3},\xi_{t_4}) \\
&= 3\mathrm{var}(\xi_t)^2 + \frac{1}{T}\sum_{t_1,t_2,t_3=1}^{T}\mathrm{cum}(\xi_0,\xi_{t_1},\xi_{t_2},\xi_{t_3}) \\
&\le 3\sqrt{EX_t^4 Eu_t^4} + \frac{1}{T}\sum_{t_1,t_2,t_3=1}^{T}\left|\mathrm{cum}(X_0u_0, X_{t_1}u_{t_1}, X_{t_2}u_{t_2}, X_{t_3}u_{t_3})\right| \qquad (16)
\end{aligned}$$

where cum(.) denotes the cumulant, the third equality follows from assumption 1 and the definition of the fourth cumulant (see definition 2.3.1 of Brillinger (1981)), the fourth

[3] If a_t is stationary with mean zero, autocovariances γ_j, and absolutely summable cumulants up to order $2k$, then $E(T^{-1/2}\sum_{t=1}^{T}a_t)^{2k} \le h_{2k}\left(\sum_j|\gamma_j|\right)^k + O(T^{-1})$.

equality follows by the stationarity of (X_t, u_t) and because $\text{cov}(\xi_t,\xi_s) = 0$ for $t \neq s$ by assumption 4, and the inequality follows by Cauchy-Schwartz (first term).

It remains to show that the final term in (16) is finite. We do so by using a result of Leonov and Shiryaev (1959), stated as Theorem 2.3.2 in Brillinger (1981), to express the cumulant of products as the product of cumulants. Let $z_{s1} = X_s$ and $z_{s2} = u_s$, and let $\nu = \bigcup_{j=1}^{m} \nu_j$ denote a partition of the set of index pairs $\mathcal{S}_{A_3} = \{(0,1), (0,2), (t_1,1), (t_1,2), (t_2,1), (t_2,2), (t_3,1), (t_3,2)\}$. Theorem 2.3.2 implies that $\text{cum}(X_0u_0, X_{t_1}u_{t_1}, X_{t_2}u_{t_2}, X_{t_3}u_{t_3}) = \text{cum}(z_{01}z_{02}, z_{t_1 1}z_{t_1 2}, z_{t_2 1}z_{t_2 2}, z_{t_3 1}z_{t_3 2}) = \sum_{\nu} \text{cum}(z_{ij}, ij \in \nu_1)\cdots\text{cum}(z_{ij}, ij \in \nu_m)$, where the summation extends over all indecomposable partitions of $\mathcal{S}_{A_3}$. Because (X_t, u_t) has mean zero, $\text{cum}(X_0) = \text{cum}(u_0) = 0$ so all partitions with some ν_k having a single element make a contribution of zero to the sum. Thus nontrivial partitions must have $m \leq 4$. Separating out the partition with $m = 1$, we therefore have that

$$\sum_{t_1,t_2,t_3=1}^{T} \left|\text{cum}(X_0u_0, X_{t_1}u_{t_1}, X_{t_2}u_{t_2}, X_{t_3}u_{t_3})\right| \leq \sum_{t_1,t_2,t_3=-\infty}^{\infty} \left|\text{cum}(X_0, u_0, X_{t_1}, u_{t_1}, X_{t_2}, u_{t_2}, X_{t_3}, u_{t_3})\right|$$

$$+ \sum_{\nu: m=2,3,4} \sum_{t_1,t_2,t_3=-\infty}^{\infty} \left|\text{cum}(z_{ij}, ij \in \nu_1)\cdots\text{cum}(z_{ij}, ij \in \nu_m)\right|. \tag{17}$$

The first term on the right hand side of (17) satisfies

$$\sum_{t_1,t_2,t_3=1}^{T} \left|\text{cum}(X_0, u_0, X_{t_1}, u_{t_1}, X_{t_2}, u_{t_2}, X_{t_3}, u_{t_3})\right| \leq \sum_{t_1,t_2,\ldots,t_7=-\infty}^{\infty} \left|\text{cum}(X_0, u_{t_1}, X_{t_2}, u_{t_3}, X_{t_4}, u_{t_5}, X_{t_6}, u_{t_7})\right|$$

which is finite by assumption 5.

It remains to show that the second term in (17) is finite. Consider cumulants of the form $\mathrm{cum}(X_{t_1},\dots,X_{t_r},u_{s_1},\dots,u_{s_p})$ (including the case of no X's). When $p = 1$, by assumption 1 this cumulant is zero. When $p = 2$, by assumption 4 this cumulant is zero if $s_1 \neq s_2$. Thus the only nontrivial partitions of $\mathcal{S}_{A_3}$ either (i) place two occurrences of u in one set and two in a second set, or (ii) place all four occurrences of u in a single set.

In case (i), the three-fold summation reduces to a single summation which can be handled by bounding one or more cumulants and invoking summability. For example, one such term is

$$\sum_{t_1,t_2,t_3=-\infty}^{\infty} \left|\mathrm{cum}(X_0,X_{t_3})\mathrm{cum}(X_{t_1},u_0,u_{t_2})\mathrm{cum}(X_{t_2},u_{t_1},u_{t_3})\right|$$

$$= \sum_{t=-\infty}^{\infty} \left|\mathrm{cum}(X_0,X_t)\mathrm{cum}(X_t,u_0,u_0)\mathrm{cum}(X_0,u_t,u_t)\right|$$

$$\leq \mathrm{var}(X_0)\sqrt{EX_0^2Eu_0^4}\sum_{t_1,t_2=-\infty}^{\infty}\left|\mathrm{cum}(X_0,u_{t_1},u_{t_2})\right| < \infty \tag{18}$$

where the inequality uses $\left|\mathrm{cum}(X_0,X_t)\right| \leq \mathrm{var}(X_0)$, $\left|\mathrm{cum}(X_t,u_0,u_0)\right| \leq \left|EX_tu_0^2\right| \leq \sqrt{EX_0^2Eu_0^4}$, and $\sum_{t=-\infty}^{\infty}\left|\mathrm{cum}(X_0,u_t,u_t)\right| \leq \sum_{t_1,t_2=-\infty}^{\infty}\left|\mathrm{cum}(X_0,u_{t_1},u_{t_2})\right|$; all terms in the final line of (18) are finite by assumption 5. For a partition to be indecomposable, it must be that at least one cumulant under the single summation contains both time indexes 0 and t (if not, the partition satisfies Equation (2.3.5) in Brillinger (1981) and thus violates the row equivalency necessary and sufficient condition for indecomposability). Thus all terms in case (i) can be handled in the same way (bounding and applying summability to a cumulant with indexes of both 0 and t) as the term handled in (18). Thus all terms in case (i) are finite.

In case (ii), the summation remains three-dimensional and all cases can be handled by bounding the cumulants not containing the u's and invoking absolute summability for the cumulant containing the u's. A typical term is

$$\sum_{t_1,t_2,t_3=-\infty}^{\infty} \left|\mathrm{cum}(X_0,u_0,u_{t_1},u_{t_2},u_{t_3})\mathrm{cum}(X_{t_1},X_{t_2},X_{t_3})\right| \leq E\left|X_0\right|^3 \sum_{t_1,t_2,t_3=-\infty}^{\infty} \left|\mathrm{cum}(X_0,u_0,u_{t_1},u_{t_2},u_{t_3})\right|$$

$$\leq E\left|X_0\right|^3 \sum_{t_1,\ldots,t_4=-\infty}^{\infty} \left|\mathrm{cum}(X_0,u_{t_1},u_{t_2},u_{t_3},u_{t_4})\right| < \infty.$$

Because the number of partitions is finite, the final term in (17) is finite, and it follows from (16) that $EA_3^4 = O(1)$.

Next consider A_4. The argument that $EA_4^4 = O(1)$ closely follows the argument for A_3. The counterpart of the final line of (16) is

$$EA_4^4 \leq 3\sqrt{EX_t^8 Eu_t^4} + \frac{1}{T}\sum_{t_1,t_2,t_3=1}^{T} \left|\mathrm{cum}(X_0X_0u_0, X_{t_1}X_{t_1}u_{t_1}, X_{t_2}X_{t_2}u_{t_2}, X_{t_3}X_{t_3}u_{t_3})\right|$$

so the leading term in the counterpart of (17) is a twelfth cumulant, which is absolutely summable by assumption 5. Following the remaining steps shows that $EA_4^4 < \infty$.

Now turn to A_0. The logic of (17) implies that

$$\begin{aligned}\text{var}(A_0) &= \text{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{it}^2 u_{it}^2\right) \\ &\le \sum_{t=-\infty}^{\infty}\left|\text{cov}(X_0^2u_0^2, X_t^2u_t^2)\right| \\ &\le \sum_{t=-\infty}^{\infty}\left|\text{cum}(X_0, X_0, u_0, u_0, X_t, X_t, u_t, u_t)\right| \\ &\quad + \sum_{\nu:m=2,3,4}\sum_{t=-\infty}^{\infty}\left|\text{cum}(z_{ij}, ij\in\nu_1)\cdots\text{cum}(z_{ij}, ij\in\nu_m)\right| \qquad (19)\end{aligned}$$

where the summation over ν extends over indecomposable partitions of $\mathcal{S}_{A_0}$ = {(0,1), (0,1), (0,2), (0,2), (t,1) , (t,1) , (t,2) , (t,2)} with $2 \le m \le 4$. The first term in the final line of (19) is finite by assumption 5. For a partition of $\mathcal{S}_{A_0}$ to be indecomposable, at least one cumulant must have indexes of both 0 and t (otherwise Brillinger's (1981) Equation (2.3.5) is satisfied). Thus the bounding and summability steps of (18) can be applied to all partitions in (19), so var(A_0) = O(1). This proves (a).

(b) First note that $E\tilde{B} = B$:

$$\begin{aligned}E\tilde{B} &= E\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^2\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}\tilde{u}_{is}^2\right) \\ &= E\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^2\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}u_{is}^2 - \frac{T}{T^2(T-1)}\sum_{s=1}^{T}\sum_{r=1}^{T}u_{is}u_{ir}\right)\end{aligned}$$

$$= E\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^2\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}u_{is}^2 - \frac{1}{T(T-1)}\sum_{s=1}^{T}u_{is}^2\right) = B$$

where the penultimate equality obtains because u_{it} is conditionally serially uncorrelated. Thus

$$E\left[\sqrt{\frac{n}{T}}\left(\tilde{B}-B\right)\right]^2 = \frac{1}{T}\text{var}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^2\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}\tilde{u}_{is}^2\right)\right]$$

$$\leq \frac{1}{T}E\left(\frac{1}{T}\sum_{t=1}^{T}X_{it}^2\right)^2\left(\frac{1}{T-1}\sum_{s=1}^{T}u_{is}^2\right)^2$$

$$\leq \frac{1}{T}\sqrt{EX_{it}^8 Eu_{is}^8} \tag{20}$$

where the first inequality uses $\sum_{t=1}^{T}\tilde{X}_{it}^2 \leq \sum_{t=1}^{T}X_{it}^2$ and $\sum_{t=1}^{T}\tilde{u}_{it}^2 \leq \sum_{t=1}^{T}u_{it}^2$. The result (b) follows from (20). Inspection of the right hand side of the first line in (20) reveals that this variance is positive for finite *T*, so that under fixed-*T* asymptotics the estimation of *B* makes a $1/nT$ contribution to the variance of $\hat{\Sigma}^{HR-FE}$.

(c) $$\sqrt{nT}\left(\hat{\Sigma}^{HR-XS} - \tilde{\Sigma}^{HR-XS}\right) = \frac{\sqrt{nT}}{nT-n-k}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\hat{\tilde{u}}_{it}^2 - \frac{\sqrt{nT}}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\tilde{u}_{it}^2$$

$$= \left(\frac{nT}{n(T-1)-k}\right)\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\left(\hat{\tilde{u}}_{it}^2 - \tilde{u}_{it}^2\right) - \left(\frac{k\sqrt{nT}}{n(T-1)-k}\right)\tilde{\Sigma}^{HR-XS}. \tag{21}$$

An implication of (a) is that $\tilde{\Sigma}^{HR-XS} \xrightarrow{p} E\tilde{\Sigma}^{HR-XS}$, so the second term in (21) is $O_p(1/\sqrt{nT}\,)$. To show that the first term in (21) is $o_p(1)$ it suffices to show that $\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}{}'\left(\hat{\tilde{u}}_{it}^2-\tilde{u}_{it}^2\right) \xrightarrow{p} 0$. Because $\hat{\tilde{u}}_{it} = \tilde{u}_{it} - (\hat{\beta}-\beta)\tilde{X}_{it}$,

$$\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}{}'\left(\hat{\tilde{u}}_{it}^2-\tilde{u}_{it}^2\right) = \sqrt{nT}\left(\hat{\beta}-\beta\right)^2\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^4$$

$$- 2\sqrt{nT}\left(\hat{\beta}-\beta\right)\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^3\tilde{u}_{it}$$

$$= \left[\sqrt{nT}\left(\hat{\beta}-\beta\right)\right]^2\frac{1}{(nT)^{3/2}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^4 - 2\sqrt{nT}\left(\hat{\beta}-\beta\right)\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^3u_{it}$$

$$+ 2\sqrt{nT}\left(\hat{\beta}-\beta\right)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^3\right)\bar{u}_i \,. \qquad (22)$$

Consider the first term in (22). Now $\sqrt{nT}\left(\hat{\beta}-\beta\right) = O_p(1)$ and

$$E\left|\frac{1}{(nT)^{3/2}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^4\right| = \frac{1}{\sqrt{nT}}E(\tilde{X}_{it}^4)\rightarrow 0$$

where convergence follows because $E(\tilde{X}_{it}^4) < \infty$ is implied by $E(X_{it}^4) < \infty$. Thus, by Markov's inequality the first term in (22) converges in probability to zero. Next consider the second term in (22). Because u_{it} is conditionally serially uncorrelated, u_{it} has (respectively) 4 moments, and $\tilde{X}_{it}$ has 12 moments (because X_{it} has 12 moments),

$$\text{var}\left(\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^{3}u_{it}\right) = \frac{1}{nT}E\left(\tilde{X}_{it}^{6}u_{it}^{2}\right) \le \frac{1}{nT}\sqrt{\left(E\tilde{X}_{it}^{12}\right)\left(Eu_{it}^{4}\right)} \to 0.$$

This result and $\sqrt{nT}\left(\hat{\beta}-\beta\right) = O_p(1)$ imply that the second term in (22) converges in probability to zero. Turning to the final term in (22), because u_{it} is conditionally serially uncorrelated, $\tilde{X}_{it}$ has 12 moments, u_{it} has 4 moments,

$$\text{var}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{3}\right)\bar{u}_i\right) = \frac{1}{nT}E\left(\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{3}\right)^2\left(\frac{1}{T}\sum_{t=1}^{T}u_{it}^{2}\right)\right)$$

$$\le \frac{1}{nT}\sqrt{E\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{3}\right)^4}\sqrt{Eu_{it}^{4}} \to 0$$

This result and $\sqrt{nT}\left(\hat{\beta}-\beta\right) = O_p(1)$ imply that the final term in (22) converges in probability to zero, and (c) follows.

(d) Use $\hat{\tilde{u}}_{it} = \tilde{u}_{it} - (\hat{\beta}-\beta)\tilde{X}_{it}$ and collect terms to obtain

$$\sqrt{n/T}\left(\hat{B}-\tilde{B}\right) = \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}\left(\hat{\tilde{u}}_{is}^{2}-\tilde{u}_{is}^{2}\right)\right)$$

$$= \left(\frac{T}{T-1}\right)\left[\sqrt{nT}\left(\hat{\beta}-\beta\right)\right]^2\frac{1}{(nT)^{3/2}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{2}\right)^2$$

$$- 2\sqrt{nT}\left(\hat{\beta}-\beta\right)\frac{1}{nT}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{2}\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}\tilde{X}_{is}\tilde{u}_{is}\right). \qquad (23)$$

Because $\sqrt{nT}\left(\hat{\beta}-\beta\right)$ = $O_p(1)$ and X_{it} has four moments, by Markov's inequality the first term in (23) converges in probability to zero (the argument is like that used for the first term in (22)). Turning to the second term in (23),

$$\text{var}\left[\frac{1}{nT}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{2}\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}\tilde{X}_{is}\tilde{u}_{is}\right)\right] = \frac{1}{n(T-1)^2}\text{var}\left(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{X}_{it}^{2}\tilde{X}_{is}\tilde{u}_{is}\right)$$

$$\leq \frac{1}{n(T-1)^2}\sqrt{E\tilde{X}_{it}^{12}Eu_{it}^{4}} \to 0$$

so the second term in (23) converges in probability to zero, and (d) follows.

Details of remark 9. The only place in this proof that the summable cumulant condition is used is to bound the A moments in part (a). If T is fixed, a sufficient condition for the moments of A to be bounded is that X_{it} and u_{it} have 12 moments. Stationarity of (X_{it}, u_{it}) is used repeatedly but, if T is fixed, stationarity could be relaxed by replacing moments such as EX_{it}^{4} with $\max_t EX_{it}^{4}$. Thus, under T-fixed, $n \to \infty$ asymptotics, assumption 5 could be replaced by the assumption that $EX_{it}^{12} < \infty$ and $Eu_{it}^{12} < \infty$ for $t = 1,\ldots, T$.

Details of remark 4. If (X_{it}, u_{it}) is i.i.d., t = 1,…, T, i = 1,…, n, then $\Sigma = E\tilde{X}_{it}\tilde{X}_{it}'u_{it}^2 = Q_{\tilde{X}\tilde{X}}\sigma_u^2 + \Omega$, where $\Omega_{jk} = \text{cov}(\tilde{X}_{jit}\tilde{X}_{kit}, u_{it}^2)$, where $\tilde{X}_{jit}$ is the j^{th} element of $\tilde{X}_{it}$. Also, the (j,k) element of B is

$$B_{jk} = E\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{X}_{jit}\tilde{X}_{kit}u_{is}^2 = Q_{\tilde{X}\tilde{X},jk}\sigma_u^2 + \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\text{cov}(\tilde{X}_{jit}\tilde{X}_{kit}, u_{is}^2)$$

$$= Q_{\tilde{X}\tilde{X},jk}\sigma_u^2 + \frac{1}{T-1}\Omega_{jk},$$

where the final equality uses, for $t \neq s$, $\text{cov}(\tilde{X}_{jit}\tilde{X}_{kit}, u_{is}^2) = T^{-2}\text{cov}(X_{jit}X_{kit}, u_{it}^2) = (T-1)^{-2}\Omega_{jk}$ (because (X_{it}, u_{it}) is i.i.d. over t). Thus $B = Q_{\tilde{X}\tilde{X}}\sigma_u^2 + (T-1)^{-1}\Omega = Q_{\tilde{X}\tilde{X}}\sigma_u^2 + (T-1)^{-1}(\Sigma - Q_{\tilde{X}\tilde{X}}\sigma_u^2)$. The result stated in the remark follows by substituting this final expression for B into (5), noting that $\hat{\Sigma}^{homosk} \xrightarrow{p} Q_{\tilde{X}\tilde{X}}\sigma_u^2$, and collecting terms.

Table 1. Monte Carlo Results: Bias, Relative MSE, and Size for Three Variance Estimators

Design: $y_{it} = x_{it}\beta + u_{it}$, $i = 1,\ldots, n$, $t = 1,\ldots, T$

$x_{it} \sim$ i.i.d. N(0,1)

$u_{it}|x_i \sim$ i.n.i.d. $N(0, \sigma_{it}^2)$; $\sigma_{it}^2 = (0.1 + x_{it}^2)^{\kappa}/E[(0.1 + x_{it}^2)^{\kappa}]$

(a) $\kappa = 1$

		Bias relative to true			MSE relative to infeasible			Size (nominal level 10%)		
T	***n***	$\hat{\Sigma}^{HR-XS}$	$\hat{\Sigma}^{HR-FE}$	$\hat{\Sigma}^{cluster}$	$\hat{\Sigma}^{HR-XS}$	$\hat{\Sigma}^{HR-FE}$	$\hat{\Sigma}^{cluster}$	$\hat{\Sigma}^{HR-XS}$	$\hat{\Sigma}^{HR-FE}$	$\hat{\Sigma}^{cluster}$
3	50	-0.180	-0.052	-0.068	0.78	1.05	1.02	0.147	0.125	0.128
5	50	-0.135	-0.029	-0.046	0.84	0.98	1.14	0.132	0.113	0.122
10	50	-0.073	-0.013	-0.034	0.92	0.99	1.47	0.119	0.108	0.119
25	50	-0.030	-0.005	-0.026	0.96	0.99	2.42	0.107	0.102	0.113
50	50	-0.015	-0.002	-0.021	0.98	0.99	3.82	0.103	0.102	0.110
100	50	-0.008	-0.001	-0.020	0.99	1.00	6.95	0.099	0.098	0.107
3	100	-0.160	-0.027	-0.035	0.89	1.11	1.10	0.144	0.118	0.120
5	100	-0.123	-0.015	-0.023	0.95	1.02	1.20	0.127	0.106	0.110
10	100	-0.067	-0.006	-0.016	0.99	1.01	1.54	0.116	0.105	0.108
25	100	-0.028	-0.002	-0.012	1.00	1.00	2.43	0.103	0.099	0.104
50	100	-0.014	-0.001	-0.012	1.00	1.00	3.95	0.102	0.100	0.104
100	100	-0.007	-0.001	-0.012	1.00	1.00	6.94	0.101	0.100	0.106
3	500	-0.142	-0.006	-0.008	1.60	1.21	1.20	0.123	0.097	0.097
5	500	-0.113	-0.003	-0.004	1.70	1.07	1.30	0.123	0.101	0.102
10	500	-0.062	-0.001	-0.003	1.45	1.03	1.55	0.114	0.103	0.104
25	500	-0.026	0.000	-0.003	1.19	1.01	2.48	0.104	0.100	0.101
50	500	-0.013	0.000	-0.002	1.10	1.00	4.06	0.102	0.100	0.101
100	500	-0.007	0.000	-0.002	1.05	1.00	7.24	0.101	0.100	0.101
3	1000	-0.139	-0.002	-0.003	2.35	1.22	1.22	0.130	0.104	0.104
5	1000	-0.112	-0.001	-0.002	2.59	1.08	1.29	0.122	0.099	0.100
10	1000	-0.062	-0.001	-0.002	2.00	1.02	1.56	0.109	0.098	0.099
25	1000	-0.026	0.000	-0.002	1.43	1.01	2.46	0.105	0.101	0.101
50	1000	-0.013	0.000	-0.001	1.23	1.00	3.93	0.102	0.100	0.100
100	1000	-0.006	0.000	0.000	1.11	1.00	7.22	0.103	0.102	0.102

Table 1, ctd.

(b) $\kappa = -1$

		Bias relative to true			MSE relative to infeasible			Size (nominal level 10%)		
T	***n***	$\hat{\Sigma}^{HR-XS}$	$\hat{\Sigma}^{HR-FE}$	$\hat{\Sigma}^{cluster}$	$\hat{\Sigma}^{HR-XS}$	$\hat{\Sigma}^{HR-FE}$	$\hat{\Sigma}^{cluster}$	$\hat{\Sigma}^{HR-XS}$	$\hat{\Sigma}^{HR-FE}$	$\hat{\Sigma}^{cluster}$
3	50	0.274	0.013	-0.012	2.72	1.32	1.28	0.067	0.105	0.110
5	50	0.313	0.007	-0.014	5.20	1.68	2.02	0.060	0.104	0.107
10	50	0.233	0.003	-0.017	6.96	1.51	4.57	0.068	0.101	0.110
25	50	0.119	0.001	-0.017	6.36	1.33	14.20	0.083	0.101	0.108
50	50	0.065	0.000	-0.018	4.62	1.19	32.51	0.091	0.101	0.111
100	50	0.034	0.000	-0.020	3.14	1.11	69.91	0.094	0.100	0.110
3	100	0.270	0.006	-0.007	3.78	1.30	1.28	0.064	0.099	0.101
5	100	0.312	0.003	-0.006	8.65	1.66	2.10	0.059	0.099	0.101
10	100	0.233	0.001	-0.009	12.68	1.51	4.68	0.065	0.098	0.102
25	100	0.119	0.001	-0.008	11.09	1.33	14.22	0.082	0.102	0.106
50	100	0.065	0.000	-0.009	7.93	1.19	32.62	0.090	0.101	0.107
100	100	0.034	0.000	-0.010	5.19	1.12	70.98	0.094	0.100	0.105
3	500	0.271	0.001	-0.002	13.59	1.31	1.30	0.063	0.098	0.098
5	500	0.309	0.000	-0.001	35.28	1.66	2.04	0.059	0.099	0.099
10	500	0.231	0.001	-0.001	55.72	1.50	4.81	0.066	0.099	0.099
25	500	0.118	0.000	-0.002	49.32	1.31	14.35	0.081	0.098	0.100
50	500	0.064	0.000	-0.002	34.61	1.19	32.99	0.090	0.100	0.101
100	500	0.034	0.000	-0.001	21.26	1.12	71.91	0.093	0.098	0.099
3	1000	0.269	0.001	0.000	25.27	1.31	1.31	0.062	0.099	0.099
5	1000	0.310	0.000	-0.001	70.65	1.66	2.09	0.059	0.099	0.099
10	1000	0.231	0.000	-0.001	108.60	1.50	4.66	0.069	0.099	0.099
25	1000	0.118	0.000	-0.001	97.76	1.32	14.49	0.084	0.103	0.103
50	1000	0.064	0.000	-0.001	68.18	1.19	33.12	0.088	0.098	0.099
100	1000	0.034	0.000	-0.001	40.87	1.10	70.28	0.093	0.098	0.100

Notes to Table 1: The first three columns of results report the bias of the indicated estimator as a fraction of the true variance. The next three columns report the MSE of the indicated estimator, relative to the MSE of the infeasible estimator $\hat{\Sigma}^{inf} = (nT)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^2 u_{it}^2$. The final three columns report rejection rate under the null hypothesis of the 2-sided test of $\beta = \beta_0$ based on the *t*-statistic computed using the indicated variance estimator and the asymptotic normal critical value, where the test has a nominal level of 10%. All results are based on 20,000 Monte Carlo draws.