# Heteroskedasticity-Robust Standard Errors for Fixed Effects Panel Data Regression

May, 2006

This revision: July, 2007

James H. Stock

Department of Economics, Harvard University

and the NBER

Mark W. Watson[1]

Department of Economics and Woodrow Wilson School, Princeton University

and the NBER

**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 $\sqrt{nT}$ -consistent under any sequences ($n$, $T$) in which $n$ and/or $T$ increase to $\infty$. This estimator can be extended to handle serial correlation of fixed order.

Key words: White standard errors, longitudinal data, clustered standard errors

JEL codes: C23, C12

---

[1] We thank Alberto Abadie, Gary Chamberlain, Guido Imbens, Doug Staiger, Hal White, and the referees for helpful comments and/or discussions, Mitchell Peterson for providing the data in footnote 2, and Anna Mikusheva for research assistance. This research was supported in part by NSF grant SBR-0617811.

## 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 serially 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. $Q_{\tilde{X}\tilde{X}} \equiv E\,T^{-1}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'$ 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 $\Sigma$ 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 $T > 2$,

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

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

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

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

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

---

[2] For example, at the time of writing $\hat{\Sigma}^{HR-XS}$ is the estimator used in STATA and Eviews. Petersen (2007) reports a survey of 207 panel data papers published in the *Journal of Finance*, the *Journal of Financial Economics*, and the *Review of Financial Studies* between 2001 and 2004. Of these, 15% used $\hat{\Sigma}^{HR-XS}$, 23% used clustered standard errors, 26% used uncorrected OLS standard errors, and the remaining papers used other methods.

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),

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

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 $\beta$ 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 (cf. Lancaster [2000]).

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 $\hat{\Sigma}^{HR-XS}$ can be biased up or down.
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. Under $T$ fixed, $n \to \infty$ asymptotics, the assumptions of stationarity and twelve summable cumulants can be relaxed, and assumption 5 can be replaced by $EX_{it}^{12} < \infty$ and $Eu_{it}^{12} < \infty$, $t = 1,\ldots, T$. The assumption of twelve moments, 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 heteroskedasticity- and autocorrelation-consistent (HAC) variance estimation in time series data; it arises here because the number of nuisance parameters (entity means) increases with $n$.
8. 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 $\hat{\Sigma}^{HR-FE}_{psd} = Q'|\Lambda|Q$ is psd.

9. If the errors are serially correlated, then (3) holds with the modification that $\Sigma = ET^{-1}\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right)\left(\sum_{t=1}^{T}\tilde{X}_{it}\tilde{u}_{it}\right)' = ET^{-1}\left(\sum_{t=1}^{T}\tilde{X}_{it}u_{it}\right)\left(\sum_{t=1}^{T}\tilde{X}_{it}u_{it}\right)'$ (the second equality arises from the idempotent matrix identity). The first of these expressions leads to 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 $\beta$ 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. The problem of no consistent estimation of the entity means does not affect the consistency of $\hat{\Sigma}^{cluster}$, however it generally does introduce $O(T^{-1})$ bias into weighted sum-of-covariances estimators based on kernels other than the nontruncated rectangular kernel used for $\hat{\Sigma}^{cluster}$.

10. If $n$ and/or $T \to \infty$, then $\hat{\Sigma}^{cluster} = \Sigma + O_p(1/\sqrt{n})$ (see the appendix of the working paper version of Hansen [2007]). 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 $\hat{\Sigma}^{HR-FE}$ is more efficient than $\hat{\Sigma}^{cluster}$. The efficiency gain of $\hat{\Sigma}^{HR-FE}$ arises because imposing the condition that $u_{it}$ is conditionally serially uncorrelated permits averaging over both entities and time, whereas $\hat{\Sigma}^{cluster}$ averages only across entities.
11. Under $n$ fixed, $T \to \infty$ asymptotics and i.i.d. observations across entities, the asymptotic null distribution of the $t$-statistic computed using $\hat{\Sigma}^{cluster}$ testing one element of $\beta$ is $\sqrt{\frac{n}{n-1}}t_{n-1}$ (Hansen [2007]), and the $F$-statistic testing $p$ elements of $\beta$ is distributed as $\left(\frac{n}{n-p}\right)F_{p,n-p}$ (this follows from Hansen [2007, Corollary 4.1] and Rao [1973, Section 8b]). If the divisor used to compute the clustered variance estimator is

$(n-1)T$, not $nT$ as in (10), then the Wald chi-squared statistic using $\hat{\Sigma}^{cluster}$ and testing $p$ restrictions on $\beta$ has the Hotelling $T^2(p,n-1)$ distribution. In contrast, if $\hat{\Sigma}^{HR-FE}$ is used the $t$-statistic is distributed N(0,1) and the $F$-statistic testing $p$ restrictions is distributed $\chi_p^2 / p$ under any sequence with $n$ and/or $T \to \infty$. All this suggests that, when $n$ is small or moderate, the increased precision of $\hat{\Sigma}^{HR-FE}$ over $\hat{\Sigma}^{cluster}$ will translate into improved power and more accurate confidence intervals.

12. The estimator $\hat{\Sigma}^{HR-FE}$ can alternatively be derived as a method of moments estimator in which zero restrictions on the conditional autocovariances of $u_{it}$ are used to impose restrictions on the conditional autocovariances of $\tilde{u}_{it}$. Let $u_i = (u_{i1},\ldots, u_{iT})'$, $\tilde{u}_i = (\tilde{u}_{i1},\ldots,\tilde{u}_{iT})'$, $\tilde{X}_i = (\tilde{X}_{i1},\ldots,\tilde{X}_{iT})'$, $\Omega_i = E\left(u_i u_i' \mid \tilde{X}_i\right)$, and $\tilde{\Omega}_i = E\left(\tilde{u}_i \tilde{u}_i' \mid \tilde{X}_i\right)$. Then $\tilde{\Omega}_i = M_\iota \Omega M_\iota$, where $M_\iota = I_T - T^{-1}\iota\iota'$, where $\iota$ is the $T$-vector of 1's. Now $\Sigma = T^{-1}E\left(\tilde{X}_i' \Omega_i \tilde{X}_i\right)$, so $vec\Sigma = T^{-1}E\left[\left(\tilde{X}_i \otimes \tilde{X}_i\right)' vec\Omega_i\right]$. Let $S$ be a $T^2 \times r$ selection matrix with full column rank such that $S'vec\Omega_i$ is the $r\times 1$ vector of the $r$ nonzero elements of $\Omega_i$. If these zero restrictions are valid, then $M_S vec\Omega_i = 0$ (where $M_S = I_{T^2} - P_S$ and $P_S = S\left(S'S\right)^{-1}S'$), so $vec\Sigma = T^{-1}E\left[\left(\tilde{X}_i \otimes \tilde{X}_i\right)' P_S vec\Omega_i\right]$. Under these zero restrictions, if $S'\left(M_\iota \otimes M_\iota\right)S$ is invertible, then (as is shown in the appendix)

$$vec\Sigma = T^{-1}E\left[\left(\tilde{X}_i \otimes \tilde{X}_i\right)' H vec\tilde{\Omega}_i\right], \tag{11}$$

where $H = S\left[S'\left(M_\iota \otimes M_\iota\right)S\right]^{-1}S'$. This suggests the estimator,

$$vec\hat{\Sigma}^{MA(q)} = \frac{1}{nT}\sum_{i=1}^{n}\left(\tilde{X}_i \otimes \tilde{X}_i\right)' H\left(\hat{\tilde{u}}_i \otimes \hat{\tilde{u}}_i\right), \tag{12}$$

where the superscript $MA(q)$ indicates that this estimator imposes a conditional moving average structure for the errors. Under the assumption of no conditional autocorrelation (so $q = 0$), $S$ selects the diagonal elements of $\Omega_i$, and the resulting estimator $\hat{\Sigma}^{MA(0)}$ is the same as $\hat{\Sigma}^{HR-FE}$ in (6) except that $k$ is dropped in the degrees-of-freedom correction (see the appendix). If no zero restrictions are imposed, then $S = I_{T^2}$ and $S'(M_t \otimes M_t)S$ is not invertible but setting $H = I_{T^2}$ yields $\hat{\Sigma}^{MA(T-1)} = \hat{\Sigma}^{cluster}$. The estimator for the MA(1) case obtains by setting $S$ to select the diagonal and first off-diagonal elements of a vectorized $T \times T$ matrix.

13. If time fixed effects are estimated as well, the results of this section continue to hold under fixed $T$, $n \to \infty$ asymptotics, for then the time effects are $\sqrt{nT}$ -consistently estimated.
14. The theoretical results and remarks 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 quantify the foregoing theoretical results. The design has a single regressor and conditionally Gaussian errors:

$$y_{it} = x_{it}\beta + u_{it} \tag{13}$$

$$x_{it} = \zeta_{it} + \theta\zeta_{it-1},\ \zeta_{it} \sim \text{i.i.d. } N(0,1),\ t = 1,\ldots, T, \tag{14}$$

$$u_{it} = \varepsilon_{it} + \theta\varepsilon_{it-1},\ \varepsilon_{it}|x_i \sim \text{i.n.i.d. } N(0, \sigma_{it}^2),\ \sigma_{it}^2 = \lambda(0.1 + x_{it}^2)^{\kappa},\ t = 1,\ldots, T, \tag{15}$$

where $\zeta_{i0}$ and $\varepsilon_{i0}$ are drawn from their stationary distributions, $\kappa = \pm 1$, and $\lambda$ is chosen so that $\text{var}(\varepsilon_{it}) = 1$.

Table 1 presents results for $\hat{\Sigma}^{HR-XS}$ (given in (4)), $\hat{\Sigma}^{HR-FE}$ (given in (6)), and $\hat{\Sigma}^{cluster}$ (given in (10)) for $\kappa = 1$ (panel (a)) and $\kappa = -1$ (panel (b)), for conditionally serially uncorrelated errors ($\theta = 0$). The first three columns of results report the bias of the three estimators, relative to the true value of $\Sigma$ [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 were the entity means and $\beta$ known. The final three columns report the size of the 10% two-sided tests of $\beta = \beta_0$ based on the *t*-statistic using the indicated variance estimator and asymptotic critical value (standard normal for $\hat{\Sigma}^{HR-XS}$ and $\hat{\Sigma}^{HR-FE}$, $\sqrt{\frac{n}{n-1}}t_{n-1}$ for $\hat{\Sigma}^{cluster}$). 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 $n = 500$, the relative bias of $\hat{\Sigma}^{HR-XS}$ is –11.5% when $\kappa = 1$ and is 32% when $\kappa = -1$. This large bias of $\hat{\Sigma}^{HR-XS}$ can produce 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 of the infeasible estimator, apparently reflecting a bias-variance tradeoff.

Second, the bias correction in $\hat{\Sigma}^{HR-FE}$ does its job: the relative bias of $\hat{\Sigma}^{HR-FE}$ is less than 2% 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.

Third, consistent with remark 12, 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 $\hat{\Sigma}^{cluster}$ considerably exceeds the MSE of $\hat{\Sigma}^{HR-FE}$ when *T* is moderate or large, regardless of *n*.

Fourth, although the focus of this note has been bias and MSE, in practice variance estimators are used mainly for inference on $\beta$, and one would suspect that the variance estimators with less bias would produce tests of $\beta = \beta_0$ 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 $T = 5$ and $n = 500$, the size of the nominal 10% test is 12.2% for $\kappa = 1$ and is 5.8% when $\kappa = -1$. In contrast, in all cases with $n = 500$, tests based on $\hat{\Sigma}^{HR-FE}$ and $\hat{\Sigma}^{cluster}$ have sizes within Monte Carlo error of 10%. In unreported designs with more heteroskedasticity, the size distortions of tests based on $\hat{\Sigma}^{HR-XS}$ are even larger than reported in Table 1.

Table 2 also considers inference on $\beta$ by comparing the size-adjusted power of two-sided *t*-tests of $\beta = \beta_0$ using $\hat{\Sigma}^{HR-FE}$ or $\hat{\Sigma}^{cluster}$ when the errors are conditionally serially uncorrelated ($\theta = 0$). Monte Carlo critical values are used to correct for finite-sample distortions in distribution of the *t*-ratio under the null. Consistent with remark 11, when *n* is small the power of Wald tests based on the more precise estimator $\hat{\Sigma}^{HR-FE}$ can considerably exceed the power of the test based on $\hat{\Sigma}^{cluster}$.

As discussed in remark 12, the approach used to obtain $\hat{\Sigma}^{HR-FE}$ can be extended to conditionally moving average errors. Table 3 considers the MA(1) case ($\theta = \pm 0.8$) and compares the performance of $\hat{\Sigma}^{MA(1)}$, defined in and subsequent to (12), to $\hat{\Sigma}^{cluster}$. As expected, both estimators show little bias and produce Wald tests with small or negligible size distortions. Because $\hat{\Sigma}^{MA(1)}$ in effect estimates fewer covariances than $\hat{\Sigma}^{cluster}$, however, $\hat{\Sigma}^{MA(1)}$ has a lower MSE than $\hat{\Sigma}^{cluster}$, with its relative precision increasing as *T* increases.

## 3. Conclusions

Our theoretical results and Monte Carlo simulations, combined with the results in Hansen (2007), 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 $\hat{\Sigma}^{cluster}$ and tests based on $\hat{\Sigma}^{HR-FE}$ will be more powerful than tests based on $\hat{\Sigma}^{cluster}$, so $\hat{\Sigma}^{HR-FE}$ should be used, especially if *T* is moderate or large. If the errors are well modeled as a low-order moving average and *T* is moderate or large, then $\hat{\Sigma}^{MA(q)}$ is an appropriate choice and is more efficient than $\hat{\Sigma}^{cluster}$. If, however, no restrictions can be placed on the serial correlation structure of the errors, then $\hat{\Sigma}^{cluster}$ should be used in conjunction with $\sqrt{\frac{n}{n-1}}t_{n-1}$ or $\left(\frac{n}{n-p}\right)F_{p,n-p}$ critical values for hypothesis tests on $\beta$.

## 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. (2007). "Asymptotic Properties of a Robust Variance Matrix Estimator for Panel Data with *T* is Large," manuscript, Graduate School of Business, University of Chicago; forthcoming, *Journal of Econometrics*.

Lancaster, T. (2000), "The Incidental Parameter Problem Since 1948." *Journal of Econometrics*, 95, 391-413.

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

Petersen, M. (2007). "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." Forthcoming, *Review of Financial Studies*.

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 $\text{var}\left(\sum_{j=1}^{m} a_j\right) \leq \left(\sum_{j=1}^{m} \sqrt{\text{var}(a_j)}\right)^2$.

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

$$\begin{aligned}
\sqrt{nT}\,(\hat{\Sigma}^{HR-FE} - \Sigma) &= \sqrt{nT}\left[\left(\frac{T-1}{T-2}\right)\left(\hat{\Sigma}^{HR-XS} - \frac{1}{T-1}\hat{B}\right) - \left(\frac{T-1}{T-2}\right)\left(E\tilde{\Sigma}^{HR-XS} - \frac{1}{T-1}B\right)\right] \\
&= \left(\frac{T-1}{T-2}\right)\sqrt{nT}\left(\hat{\Sigma}^{HR-XS} - E\tilde{\Sigma}^{HR-XS}\right) - \frac{\sqrt{nT}}{T-2}\left(\hat{B} - B\right) \\
&= \left(\frac{T-1}{T-2}\right)\left[\sqrt{nT}\left(\hat{\Sigma}^{HR-XS} - \tilde{\Sigma}^{HR-XS}\right) + \sqrt{nT}\left(\tilde{\Sigma}^{HR-XS} - E\tilde{\Sigma}^{HR-XS}\right)\right] \\
&\quad - \left(\frac{T}{T-2}\right)\left[\sqrt{\frac{n}{T}}\left(\hat{B} - \tilde{B}\right) + \sqrt{\frac{n}{T}}\left(\tilde{B} - B\right)\right] \qquad (16)
\end{aligned}$$

where $\tilde{\Sigma}^{HR-XS}$ is given in (8) and $\tilde{B}$ is $\hat{B}$ given in (6) with $\hat{\hat{u}}_{it}$ 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 (16) yields $\sqrt{nT}\,(\hat{\Sigma}^{HR-FE} - \Sigma) = O_p(1)$ and thus the result (7)

.

(a) From (8), we have that

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

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

so (a) follows if it can be shown that $\operatorname{var}\left(T^{-1/2}\sum_{t=1}^{T}\tilde{X}_{it}^2\tilde{u}_{it}^2\right) = O(1)$. 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 $A_0 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_{it}^2u_{it}^2$, $A_1 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_{it}$, $A_2 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}u_{it}$, $A_3 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_{it}u_{it}$, $A_4 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}X_{it}^2u_{it}$, $D_1 = \frac{1}{T}\sum_{t=1}^{T}X_{it}^2$, $D_2 = \frac{1}{T}\sum_{t=1}^{T}u_{it}^2$, and $D_3 = \frac{1}{T}\sum_{t=1}^{T}X_{it}u_{it}^2$. Thus

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

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

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

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

$$\left. + 2T^{-1/2}\left(EA_2^4EA_4^4\right)^{1/4} + 4T^{-1}\left(EA_1^8EA_2^8\right)^{1/8}\left(EA_3^4\right)^{1/4} + 3T^{-3/2}\left(EA_1^8EA_2^8\right)^{1/4}\right\}^2 \quad (17)$$

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 (17) 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,

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

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 $\xi_t = X_tu_t$. Then

$$\begin{aligned} EA_3^4 &= E\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\xi_t\right)^4 = \frac{1}{T^2}\sum_{t_1,\ldots,t_4=1}^{T}E\xi_{t_1}\xi_{t_2}\xi_{t_3}\xi_{t_4} \\ &= 3\left[\frac{1}{T}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\text{cov}(\xi_{t_1},\xi_{t_2})\right]^2 + \frac{1}{T^2}\sum_{t_1,\ldots,t_4=1}^{T}\text{cum}(\xi_{t_1},\xi_{t_2},\xi_{t_3},\xi_{t_4}) \\ &= 3\text{var}(\xi_t)^2 + \frac{1}{T}\sum_{t_1,t_2,t_3=1}^{T}\text{cum}(\xi_0,\xi_{t_1},\xi_{t_2},\xi_{t_3}) \end{aligned}$$

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

$$\leq 3\sqrt{EX_t^4 Eu_t^4} + \frac{1}{T}\sum_{t_1,t_2,t_3=1}^{T}\left|\text{cum}(X_0u_0, X_{t_1}u_{t_1}, X_{t_2}u_{t_2}, X_{t_3}u_{t_3})\right| \qquad (18)$$

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 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 (18) 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 $\nu_k$ 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

$$\sum_{t_1,t_2,t_3=1}^{T}\left|\text{cum}(X_0u_0, X_{t_1}u_{t_1}, X_{t_2}u_{t_2}, X_{t_3}u_{t_3})\right| \leq \sum_{t_1,t_2,t_3=-\infty}^{\infty}\left|\text{cum}(X_0, u_0, X_{t_1}, u_{t_1}, X_{t_2}, u_{t_2}, X_{t_3}, u_{t_3})\right|$$
$$+ \sum_{\nu:m=2,3,4}\sum_{t_1,t_2,t_3=-\infty}^{\infty}\left|\text{cum}(z_{ij}, ij \in \nu_1)\cdots\text{cum}(z_{ij}, ij \in \nu_m)\right|. \qquad (19)$$

The first term on the right hand side of (19) 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 (19) is finite. Consider cumulants of the form $\text{cum}(X_{t_1},...,X_{t_r},u_{s_1},...,u_{s_p})$ (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

$$\begin{aligned}
&\sum_{t_1,t_2,t_3=-\infty}^{\infty} \left|\text{cum}(X_0,X_{t_3})\text{cum}(X_{t_1},u_0,u_{t_2})\text{cum}(X_{t_2},u_{t_1},u_{t_3})\right| \\
&\qquad = \sum_{t=-\infty}^{\infty} \left|\text{cum}(X_0,X_t)\text{cum}(X_t,u_0,u_0)\text{cum}(X_0,u_t,u_t)\right| \\
&\qquad \leq \text{var}(X_0)\sqrt{EX_0^2Eu_0^4} \sum_{t_1,t_2=-\infty}^{\infty} \left|\text{cum}(X_0,u_{t_1},u_{t_2})\right| < \infty \qquad (20)
\end{aligned}$$

where the inequality uses $\left|\text{cum}(X_0,X_t)\right| \leq \text{var}(X_0)$, $\left|\text{cum}(X_t,u_0,u_0)\right| \leq \left|EX_tu_0^2\right| \leq \sqrt{EX_0^2Eu_0^4}$, and $\sum_{t=-\infty}^{\infty}\left|\text{cum}(X_0,u_t,u_t)\right| \leq \sum_{t_1,t_2=-\infty}^{\infty}\left|\text{cum}(X_0,u_{t_1},u_{t_2})\right|$; all terms in the final line of (20) 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 (20). 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

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

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

Because the number of partitions is finite, the final term in (19) is finite, and it follows from (18) 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 (18) is

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

so the leading term in the counterpart of (19) 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 (19) implies that

$$\begin{aligned}
\text{var}(A_0) &= \text{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} X_{it}^2 u_{it}^2\right) \\
&\le \sum_{t=-\infty}^{\infty} \left|\text{cov}(X_0^2u_0^2, X_t^2u_t^2)\right| \\
&\le \sum_{t=-\infty}^{\infty} \left|\text{cum}(X_0, X_0, u_0, u_0, X_t, X_t, u_t, u_t)\right| \\
&\quad + \sum_{\nu:m=2,3,4} \sum_{t=-\infty}^{\infty} \left|\text{cum}(z_{ij}, ij \in \nu_1)\cdots\text{cum}(z_{ij}, ij \in \nu_m)\right| \qquad (21)
\end{aligned}$$

where the summation over $\nu$ 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 (21) 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 (20) can be applied to all partitions in (21), so var($A_0$) = $O(1)$. This proves (a).

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

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

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

$$\begin{aligned} E\left[\sqrt{\frac{n}{T}}\left(\tilde{B}-B\right)\right]^2 &= \frac{1}{T}\mathrm{var}\left[\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^2\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}\tilde{u}_{is}^2\right)\right] \\ &\le \frac{1}{T}E\left(\frac{1}{T}\sum_{t=1}^{T}X_{it}^2\right)^2\left(\frac{1}{T-1}\sum_{s=1}^{T}u_{is}^2\right)^2 \\ &\le \frac{1}{T}\sqrt{EX_{it}^8 Eu_{is}^8} \end{aligned} \qquad (22)$$

where the first inequality uses $\sum_{t=1}^{T}\tilde{X}_{it}^2 \le \sum_{t=1}^{T}X_{it}^2$ and $\sum_{t=1}^{T}\tilde{u}_{it}^2 \le \sum_{t=1}^{T}u_{it}^2$. The result (b) follows from (22). Inspection of the right hand side of the first line in (22) 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) $$\sqrt{nT}\left(\hat{\Sigma}^{HR-XS} - \tilde{\Sigma}^{HR-XS}\right) = \frac{\sqrt{nT}}{nT-n-k}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\hat{\tilde{u}}_{it}^2 - \frac{\sqrt{nT}}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\tilde{u}_{it}^2$$

$$= \left(\frac{nT}{n(T-1)-k}\right)\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\left(\hat{\tilde{u}}_{it}^2-\tilde{u}_{it}^2\right) - \left(\frac{k\sqrt{nT}}{n(T-1)-k}\right)\tilde{\Sigma}^{HR-XS}. \qquad (23)$$

An implication of (a) is that $\tilde{\Sigma}^{HR-XS} \xrightarrow{p} E\tilde{\Sigma}^{HR-XS}$, so the second term in (23) is $O_p(1/\sqrt{nT})$. To show that the first term in (23) 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 $\hat{\tilde{u}}_{it} = \tilde{u}_{it} - (\hat{\beta}_{FE}-\beta)\tilde{X}_{it}$,

$$\begin{aligned}\frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\left(\hat{\tilde{u}}_{it}^2-\tilde{u}_{it}^2\right) &= \sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right)^2\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^4 \\ &\quad - 2\sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right)\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^3\tilde{u}_{it} \\ &= \left[\sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right)\right]^2\frac{1}{(nT)^{3/2}}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^4 - 2\sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right)\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{X}_{it}^3u_{it} \\ &\quad + 2\sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right)\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^3\right)\overline{u}_i. \qquad (24)\end{aligned}$$

Consider the first term in (24). Now $\sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right) = O_p(1)$ and

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

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 (24) converges in probability to zero. Next consider the second term in (24). 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),

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

This result and $\sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right) = O_p(1)$ imply that the second term in (24) converges in probability to zero. Turning to the final term in (24), because $u_{it}$ is conditionally serially uncorrelated, $\tilde{X}_{it}$ has 12 moments, $u_{it}$ has 4 moments,

$$\begin{aligned}\text{var}\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{3}\right)\overline{u}_i\right) &= \frac{1}{nT}E\left(\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{3}\right)^2\left(\frac{1}{T}\sum_{t=1}^{T}u_{it}^{2}\right)\right)\\ &\le \frac{1}{nT}\sqrt{E\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{3}\right)^4}\sqrt{Eu_{it}^4} \to 0\end{aligned}$$

This result and $\sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right) = O_p(1)$ imply that the final term in (24) converges in probability to zero, and (c) follows.

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

$$\begin{aligned}\sqrt{n/T}\left(\hat{B}-\tilde{B}\right) &= \frac{1}{\sqrt{nT}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}\tilde{X}_{it}'\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}\left(\hat{\tilde{u}}_{is}^{2}-\tilde{u}_{is}^{2}\right)\right)\\ &= \left(\frac{T}{T-1}\right)\left[\sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right)\right]^2\frac{1}{(nT)^{3/2}}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{2}\right)^2\\ &\quad - 2\sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right)\frac{1}{nT}\sum_{i=1}^{n}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{X}_{it}^{2}\right)\left(\frac{1}{T-1}\sum_{s=1}^{T}\tilde{X}_{is}\tilde{u}_{is}\right). \qquad (25)\end{aligned}$$

Because $\sqrt{nT}\left(\hat{\beta}_{FE}-\beta\right) = O_p(1)$ and $X_{it}$ has four moments, by Markov's inequality the first term in (25) converges in probability to zero (the argument is like that used for the first term in (24)). Turning to the second term in (25),

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

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

so the second term in (25) converges in probability to zero, and (d) follows.

***Details of remark 7.*** The only place in the foregoing 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., $t = 1,\ldots, T$, $i = 1,\ldots, n$, 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 $j^{\text{th}}$ 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.

***Details of remark 12.*** To derive (11), first note that $vec\tilde{\Omega}_i = \left(M_\iota \otimes M_\iota\right) vec\Omega_i = \left(M_\iota \otimes M_\iota\right)\left(P_S + M_S\right) vec\Omega_i = \left(M_\iota \otimes M_\iota\right) SS' vec\Omega_i$, where the final equality imposes the zero restrictions $M_S vec\Omega_i = 0$ and uses the fact that $S'S = I$. The system of equations $vec\,\tilde{\Omega}_i = (M_\iota \otimes M_\iota) SS' vec\Omega_i$, is overdetermined, but the system $S' vec\,\tilde{\Omega}_i = S'(M_\iota \otimes M_\iota) SS' vec\Omega_i$ is exactly determined, so if $S'\left(M_\iota \otimes M_\iota\right) S$ is invertible, then $S' vec\Omega_i = \left[S'\left(M_\iota \otimes M_\iota\right) S\right]^{-1} S' vec\tilde{\Omega}_i$. (This final expression also can be obtained as the least squares solution to the overdetermined system.) One obtains (11) by substituting this final expression into $vec\Sigma = T^{-1} E\left[\left(\tilde{X}_i \otimes \tilde{X}_i\right) P_S vec\Omega_i\right]$ and using $S'S = I$.

We now show that $\hat{\Sigma}^{MA(0)}$, given by (12) for the MA(0) case, is the same as $\hat{\Sigma}^{HR-FE}$ up to the degree of freedom correction involving $k$. For the MA(0) case, $S$ is $T^2 \times T$ with nonzero elements $\{S_{(j-1)T+j,j} = 1, j = 1,\ldots, T\}$. Direct calculations show that $S'\left(M_\iota \otimes M_\iota\right) S = T^{-1}\left(T-2\right)\left[I + T^{-1}\left(T-2\right)^{-1} \iota\iota'\right]$ and $\left[S'\left(M_\iota \otimes M_\iota\right) S\right]^{-1} = \left(T-2\right)^{-1} T\left[I - T^{-1}\left(T-1\right)^{-1} \iota\iota'\right]$. Now $S'\left(\tilde{u}_i \otimes \tilde{u}_i\right) = (\tilde{u}_{i1}^2,\ldots,\tilde{u}_{iT}^2)' \equiv \tilde{u}_i^2$ and $S'\left(\tilde{X}_i \otimes \tilde{X}_i\right) = (\tilde{X}_{i1} \otimes \tilde{X}_{i1},\ldots,\ \tilde{X}_{iT} \otimes \tilde{X}_{iT})' \equiv \tilde{X}_i^2$. Thus, starting with (12) and the definition of $H$,

$$
\begin{aligned}
vec\,\hat{\Sigma}^{MA(0)} &= \frac{1}{nT}\sum_{i=1}^{n}\left(\tilde{X}_i \otimes \tilde{X}_i\right)' S\left[S'\left(M_\iota \otimes M_\iota\right) S\right]^{-1} S'\left(\tilde{u}_i \otimes \tilde{u}_i\right) \\
&= \frac{1}{nT}\sum_{i=1}^{n} \tilde{X}_i^{2\prime}\left(\frac{T}{T-2}\right)\left[I - \frac{1}{T\left(T-1\right)}\iota\iota'\right]\tilde{u}_i^2 \\
&= \left(\frac{T-1}{T-2}\right)\left[\frac{1}{n(T-1)}\sum_{i=1}^{n}\tilde{X}_i^{2\prime}\tilde{u}_i^2 - \frac{1}{T-1}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\tilde{X}_i^{2\prime}\iota}{T}\right)\left(\frac{\iota'\tilde{u}_i^2}{T-1}\right)\right] \\
&= \left(\frac{T-1}{T-2}\right)\left[\left(\frac{nT-n-k}{nT-n}\right) vec\hat{\Sigma}^{HR-XS} - \frac{1}{T-1} vec\hat{B}\right]. \qquad (26)
\end{aligned}
$$

The only difference between $\hat{\Sigma}^{MA(0)}$ in (26) and $\hat{\Sigma}^{HR-FE}$ in (6) is that $k$ in the degrees-of-freedom adjustment in $\hat{\Sigma}^{HR-XS}$ is eliminated in (26).

Table 1. Monte Carlo results for $\hat{\Sigma}^{HR-XS}$, $\hat{\Sigma}^{HR-FE}$, and $\hat{\Sigma}^{cluster}$: bias, MSE relative to the infeasible variance estimator with known entity means and $\beta$, and size of two-sided Wald tests of $\beta = 0$

Monte Carlo design: Equations (13) – (15) with $\theta = 0$ (uncorrelated errors) and $\beta = 0$

| *T* | *n* | Bias relative to true | | | MSE relative to infeasible | | | Size (nominal level 10%) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | $\hat{\Sigma}^{HR-XS}$ | $\hat{\Sigma}^{HR-FE}$ | $\hat{\Sigma}^{cluster}$ | $\hat{\Sigma}^{HR-XS}$ | $\hat{\Sigma}^{HR-FE}$ | $\hat{\Sigma}^{cluster}$ | $\hat{\Sigma}^{HR-XS}$ | $\hat{\Sigma}^{HR-FE}$ | $\hat{\Sigma}^{cluster}$ |
| **(a) $\kappa = 1$** | | | | | | | | | | |
| 5 | 20 | -0.170 | -0.069 | -0.111 | 0.72 | 0.87 | 1.07 | 0.152 | 0.132 | 0.124 |
| 10 | 20 | -0.085 | -0.025 | -0.072 | 0.83 | 0.92 | 1.52 | 0.125 | 0.113 | 0.111 |
| 20 | 20 | -0.045 | -0.014 | -0.062 | 0.91 | 0.96 | 2.29 | 0.113 | 0.108 | 0.108 |
| 50 | 20 | -0.016 | -0.003 | -0.050 | 0.96 | 0.98 | 4.73 | 0.105 | 0.102 | 0.101 |
| | | | | | | | | | | |
| 5 | 100 | -0.126 | -0.018 | -0.027 | 1.05 | 1.04 | 1.32 | 0.128 | 0.107 | 0.107 |
| 10 | 100 | -0.064 | -0.004 | -0.013 | 1.01 | 1.01 | 1.69 | 0.114 | 0.103 | 0.102 |
| 20 | 100 | -0.038 | -0.006 | -0.017 | 1.03 | 1.01 | 2.44 | 0.107 | 0.102 | 0.103 |
| 50 | 100 | -0.016 | -0.003 | -0.014 | 1.02 | 1.00 | 4.87 | 0.103 | 0.101 | 0.101 |
| | | | | | | | | | | |
| 5 | 500 | -0.115 | -0.004 | -0.006 | 2.06 | 1.09 | 1.38 | 0.122 | 0.103 | 0.103 |
| 10 | 500 | -0.060 | 0.001 | -0.001 | 1.57 | 1.03 | 1.72 | 0.110 | 0.099 | 0.099 |
| 20 | 500 | -0.013 | 0.019 | 0.017 | 0.87 | 1.01 | 2.29 | 0.104 | 0.098 | 0.097 |
| 50 | 500 | -0.014 | -0.001 | -0.003 | 1.16 | 1.01 | 4.95 | 0.102 | 0.100 | 0.100 |
| **(b) $\kappa = -1$** | | | | | | | | | | |
| 5 | 20 | 0.317 | 0.022 | -0.032 | 2.05 | 1.32 | 1.41 | 0.060 | 0.106 | 0.094 |
| 10 | 20 | 0.235 | 0.006 | -0.044 | 2.25 | 1.21 | 2.37 | 0.068 | 0.101 | 0.096 |
| 20 | 20 | 0.149 | 0.007 | -0.044 | 2.08 | 1.14 | 4.39 | 0.077 | 0.099 | 0.098 |
| 50 | 20 | 0.058 | -0.007 | -0.055 | 1.47 | 1.06 | 10.58 | 0.091 | 0.101 | 0.100 |
| | | | | | | | | | | |
| 5 | 100 | 0.320 | 0.009 | -0.001 | 4.72 | 1.30 | 1.46 | 0.060 | 0.102 | 0.100 |
| 10 | 100 | 0.228 | -0.004 | -0.013 | 5.52 | 1.21 | 2.42 | 0.069 | 0.102 | 0.100 |
| 20 | 100 | 0.134 | -0.008 | -0.018 | 4.49 | 1.12 | 4.46 | 0.081 | 0.101 | 0.101 |
| 50 | 100 | 0.052 | -0.013 | -0.022 | 2.30 | 1.05 | 10.09 | 0.091 | 0.101 | 0.101 |
| | | | | | | | | | | |
| 5 | 500 | 0.320 | 0.007 | 0.005 | 18.16 | 1.31 | 1.50 | 0.058 | 0.099 | 0.099 |
| 10 | 500 | 0.230 | -0.003 | -0.005 | 22.13 | 1.20 | 2.43 | 0.067 | 0.099 | 0.099 |
| 20 | 500 | 0.150 | 0.006 | 0.004 | 20.83 | 1.12 | 4.37 | 0.078 | 0.099 | 0.099 |
| 50 | 500 | 0.056 | -0.009 | -0.010 | 7.73 | 1.05 | 9.37 | 0.091 | 0.102 | 0.103 |

Notes: 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 rates 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 10% asymptotic critical value (using $\hat{\Sigma}^{HR-XS}$ and $\hat{\Sigma}^{HR-FE}$, the critical value is from the standard normal distribution, using $\hat{\Sigma}^{cluster}$ it is from the $\sqrt{\frac{n}{n-1}}t_{n-1}$ distribution). All results are based on 50,000 Monte Carlo draws.

Table 2. Size-adjusted power of level-10% two-sided Wald tests of $\beta = 0$ based on $\hat{\Sigma}^{HR-FE}$ and $\hat{\Sigma}^{cluster}$ against the local alternative $\beta = b/\sqrt{nT}$

Monte Carlo design: (13) – (15) with $\kappa = 1$, $\theta = 0$ (uncorrelated errors), and $T = 20$

| $n$ | Size-adjusted power of z-test based on: | |
|---|---|---|
| | $\hat{\Sigma}^{HR-FE}$ | $\hat{\Sigma}^{cluster}$ |
| (a) $b = 2$ | | |
| 3 | 0.338 | 0.227 |
| 5 | 0.324 | 0.269 |
| 10 | 0.332 | 0.305 |
| 20 | 0.322 | 0.307 |
| 100 | 0.321 | 0.316 |
| (b) $b = 4$ | | |
| 3 | 0.758 | 0.504 |
| 5 | 0.750 | 0.629 |
| 10 | 0.760 | 0.710 |
| 20 | 0.760 | 0.731 |
| 100 | 0.761 | 0.756 |

Notes: Entries are Monte Carlo rejection rates of two-sided *t*-tests using the indicated variance estimator, with a critical value computed by Monte Carlo for $\beta = 0$, $\kappa = 1$, $\theta = 0$, $T = 20$, and $n$ given in the first column. Based on 50,000 Monte Carlo draws.

Table 3. Performance of $\hat{\Sigma}^{MA(1)}$ and $\hat{\Sigma}^{cluster}$ with MA(1) errors: Bias, MSE of $\hat{\Sigma}^{MA(1)}$ relative to $\hat{\Sigma}^{cluster}$, and size of Wald tests of $\beta = 0$

Monte Carlo design: (13) – (15) with $\kappa = 1$, $\theta = \pm 0.8$, $\beta = 0$, and $n = 100$

| $T$ | Bias relative to true | | $\frac{mse\left(\hat{\Sigma}^{MA(1)}\right)}{mse\left(\hat{\Sigma}^{cluster}\right)}$ | Size (nominal level 10%) | |
|---|---|---|---|---|---|
| | $\hat{\Sigma}^{MA(1)}$ | $\hat{\Sigma}^{cluster}$ | | $\hat{\Sigma}^{MA(1)}$ | $\hat{\Sigma}^{cluster}$ |
| **(a) $\theta = 0.8$** | | | | | |
| 5 | -0.022 | -0.023 | 0.99 | 0.113 | 0.108 |
| 10 | -0.013 | -0.019 | 0.73 | 0.107 | 0.105 |
| 20 | -0.006 | -0.015 | 0.52 | 0.103 | 0.102 |
| **(b) $\theta = -0.8$** | | | | | |
| 5 | -0.032 | -0.035 | 0.93 | 0.112 | 0.109 |
| 10 | -0.018 | -0.025 | 0.72 | 0.107 | 0.106 |
| 20 | -0.007 | -0.015 | 0.52 | 0.103 | 0.102 |

Notes: $\hat{\Sigma}^{MA(1)}$ is defined in (12), where $S$ selects the diagonal and first upper and lower off-diagonal elements of a vectorized $T \times T$ matrix. Size was computed using asymptotic critical values (standard normal for $\hat{\Sigma}^{MA(1)}$, $\sqrt{\frac{n}{n-1}}t_{n-1}$ for $\hat{\Sigma}^{cluster}$) for two-sided Wald tests using the indicated variance estimator. Based on 50,000 Monte Carlo draws.