



Asymptotic properties of a robust variance matrix estimator for panel data when T is large

Christian B. Hansen

University of Chicago, Graduate School of Business, 5807 South Woodlawn Ave., Chicago, IL 60637, USA

Available online 20 November 2006

Abstract

I consider the asymptotic properties of a commonly advocated covariance matrix estimator for panel data. Under asymptotics where the cross-section dimension, n , grows large with the time dimension, T , fixed, the estimator is consistent while allowing essentially arbitrary correlation within each individual. However, many panel data sets have a non-negligible time dimension. I extend the usual analysis to cases where n and T go to infinity jointly and where $T \rightarrow \infty$ with n fixed. I provide conditions under which t and F statistics based on the covariance matrix estimator provide valid inference and illustrate the properties of the estimator in a simulation study.

© 2007 Elsevier B.V. All rights reserved.

JEL classification: C12; C13; C23

Keywords: Panel; Heteroskedasticity; Autocorrelation; Robust; Covariance matrix

1. Introduction

The use of heteroskedasticity robust covariance matrix estimators, cf. [White \(1980\)](#), in cross-sectional settings and of heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimators, cf. [Andrews \(1991\)](#), in time series contexts is extremely common in applied econometrics. The popularity of these robust covariance matrix estimators is due to their consistency under weak functional form assumptions. In particular, their use allows the researcher to form valid confidence regions about a set of parameters from a model of interest without specifying an exact process for the disturbances in the model.

E-mail address: chansen1@chicagoGSB.edu.

With the increasing availability of panel data, it is natural that the use of robust covariance matrix estimators for panel data settings that allow for arbitrary within individual correlation are becoming more common. A recent paper by Bertrand et al. (2004) illustrated the pitfalls of ignoring serial correlation in panel data, finding through a simulation study that inference procedures which fail to account for within individual serial correlation may be severely size distorted. As a potential resolution of this problem, Bertrand et al. (2004) suggest the use of a robust covariance matrix estimator proposed by Arellano (1987) and explored in Kezdi (2002) which allows arbitrary within individual correlation and find in a simulation study that tests based on this estimator of the covariance parameters have correct size.

One drawback of the estimator of Arellano (1987), hereafter referred to as the “clustered” covariance matrix (CCM) estimator, is that its properties are only known in conventional panel asymptotics as the cross-section dimension, n , increases with the time dimension, T , fixed. While many panel data sets are indeed characterized by large n and relatively small T , this is not necessarily the case. For example, in many differences-in-differences and policy evaluation studies, the cross-section is composed of states and the time dimension of yearly or quarterly (or occasionally monthly) observations on each state for 20 or more years.

In this paper, I address this issue by exploring the theoretical properties of the CCM estimator in asymptotics that allow n and T to go to infinity jointly and in asymptotics where T goes to infinity with n fixed. I find that the CCM estimator, appropriately normalized, is consistent without imposing any conditions on the rate of growth of T relative to n even when the time series dependence between the observations within each individual is left unrestricted. In this case, both the OLS estimator and the CCM estimator converge at only the \sqrt{n} -rate, essentially because the only information is coming from cross-sectional variation. If the time series process is restricted to be strongly mixing, I show that the OLS estimator is \sqrt{nT} -consistent but that, because high lags are not down weighted, the robust covariance matrix estimator still converges at only the \sqrt{n} -rate. This behavior suggests, as indicated in the simulations found in Kezdi (2002), that it is the n dimension and not the size of n relative to T that matters for determining the properties of the CCM estimator.

It is interesting to note that the limiting behavior of $\hat{\beta}$ changes “discontinuously” as the amount of dependence is limited. In particular, the rate of convergence of $\hat{\beta}$ changes from \sqrt{n} in the “no-mixing case” to \sqrt{nT} when mixing is imposed. However, despite the difference in the limiting behavior of $\hat{\beta}$, there is no difference in the behavior of standard inference procedures based on the CCM estimator between the two cases. In particular, the same t and F statistics will be valid in either case (and in the $n \rightarrow \infty$ with T fixed case) without reference to the asymptotics or degree of dependence in the data.

I also derive the behavior of the CCM estimator as $T \rightarrow \infty$ with n fixed, where I find the estimator is not consistent but does have a limiting distribution. This result corresponds to asymptotic results for HAC estimators without truncation found in recent work by Kiefer and Vogelsang (2002, 2005), Phillips et al. (2003), and Vogelsang (2003). While the limiting distribution is not proportional to the true covariance matrix in general, it is proportional to the covariance matrix in the important special case of iid data across individuals,¹

¹Note that this still allows arbitrary correlation and heteroskedasticity within individuals, but restricts that the pattern is the same across individuals.

allowing construction of asymptotically pivotal statistics in this case. In fact, in this case, the standard t -statistic is not asymptotically normal but converges in distribution to a random variable which is exactly proportional to a t_{n-1} distribution. This behavior suggests the use of the t_{n-1} for constructing confidence intervals and tests when the CCM estimator is used as a general rule, as this will provide asymptotically correct critical values under any asymptotic sequence.

I then explore the finite sample behavior of the CCM estimator and tests based upon it through a short simulation study. The simulation results indicate that tests based on the robust standard error estimates generally have approximately correct size in serially correlated panel data even in small samples. However, the standard error estimates themselves are considerably more variable than their counterparts based on simple parametric models. The bias of the simple parametric estimators is also typically smaller in the cases where the parametric model is correct, suggesting that these standard error estimates are likely preferable when the researcher is confident in the form of the error process. In the simulation, I also explore the behavior of an analog of White's (1980) direct test for heteroskedasticity proposed by Kezdi (2002).² The results indicate the performance of the test is fairly good for moderate n , though it is quite poor when n is small. This simulation behavior suggests that this test may be useful for choosing between the use of robust standard error estimates and standard errors estimated from a more parsimonious model when n is reasonably large.

The remainder of this paper is organized as follows. In Section 2, I present the basic framework and the estimator and test statistics that will be considered. The asymptotic properties of these estimators are collected in Section 3, and Section 4 contains a discussion of a Monte Carlo study assessing the finite sample performance of the estimators in simple models. Section 5 concludes.

2. A heteroskedasticity–autocorrelation consistent covariance matrix estimator for panel data

Consider a regression model defined by

$$y_{it} = x_{it}'\beta + \varepsilon_{it}, \quad (1)$$

where $i = 1, \dots, n$ indexes individuals, $t = 1, \dots, T$ indexes time, x_{it} is a $k \times 1$ vector of observable covariates, and ε_{it} is an unobservable error component. Note that this formulation incorporates the standard fixed effects model as well as models which include other covariates that enter the model with individual specific coefficients, such as individual specific time trends, where these covariates have been partialled out. In these cases, the variables x_{it} , y_{it} , and ε_{it} should be interpreted as residuals from regressions of x_{it}^* , y_{it}^* , and ε_{it}^* on an auxiliary set of covariates z_{it}^* from the underlying model $y_{it}^* = x_{it}^*\beta + z_{it}^*\gamma + \varepsilon_{it}^*$. For example, in the fixed effects model, Z^* is a matrix of dummy variables for each individual and γ is a vector of individual specific fixed effects. In this case, $x_{it} = x_{it}^* - (1/T)\sum_{t=1}^T x_{it}^*$, and y_{it} and ε_{it} are defined similarly. Alternatively, x_{it} , y_{it} , and ε_{it} could be interpreted as variables resulting from other transformations which

²Solon and Inoue (2004) offers a different testing procedure for detecting serial correlation in fixed effects panel models. See also Bhargava et al. (1982), Baltagi and Wu (1999), Wooldridge (2002, pp. 275, 282–283), and Drukker (2003).

remove the nuisance parameters from the equation, such as first-differencing to remove the fixed effects. In what follows, all properties are given in terms of the transformed variables for convenience. Alternatively, conditions could be imposed on the underlying variables and the properties derived as $T \rightarrow \infty$ as in Hansen (2006).³

Within each individual, the equations defined by (1) may be stacked and represented in matrix form as

$$y_i = x_i\beta + \varepsilon_i, \tag{2}$$

where y_i is a $T \times 1$ vector of individual outcomes, x_i is a $T \times k$ vector of observed covariates, and ε_i is a $T \times 1$ vector of unobservables affecting the outcomes y_i with $E[\varepsilon_i\varepsilon_i'|x_i] = \Omega_i$. The OLS estimator of β from Eq. (2) may then be defined as $\hat{\beta} = (\sum_{i=1}^n x_i'x_i)^{-1} \sum_{i=1}^n x_i'y_i$. The properties of $\hat{\beta}$ as $n \rightarrow \infty$ with T fixed are well known. In particular, under regularity conditions, $\sqrt{n}(\hat{\beta} - \beta)$ is asymptotically normal with covariance matrix $Q^{-1}WQ^{-1}$ where $Q = \lim_n (1/n)\sum_{i=1}^n E[x_i'x_i]$ and $W = \lim_n (1/n)\sum_{i=1}^n E[x_i'\Omega_i x_i]$.

The problem of robust covariance matrix estimation is then estimating W without imposing a parametric structure on the Ω_i . In this paper, I consider the estimator suggested by Arellano (1987) which may be defined as

$$\hat{W} = \frac{1}{nT} \sum_{i=1}^n x_i' \hat{\varepsilon}_i \hat{\varepsilon}_i' x_i, \tag{3}$$

where $\hat{\varepsilon}_i = y_i - x_i\hat{\beta}$ are OLS residuals from Eq. (2). This estimator is an appealing generalization of White’s (1980) heteroskedasticity consistent covariance matrix estimator that allows for arbitrary intertemporal correlation patterns and heteroskedasticity across individuals.⁴ The estimator is also appealing in that, unlike HAC estimators for time series data, its implementation does not require the selection of a kernel or bandwidth parameter. The properties of \hat{W} under conventional panel asymptotics where $n \rightarrow \infty$ with T fixed are well-established. In the remainder of this paper, I extend this analysis by considering the properties of \hat{W} under asymptotic sequences where $T \rightarrow \infty$ as well.

The chief reason for interest in the CCM estimator is for performing inference about $\hat{\beta}$. Suppose $\sqrt{d_{nT}}(\hat{\beta} - \beta) \xrightarrow{d} N(0, B)$ and define an estimator of the asymptotic variance of $\hat{\beta}$ as $(1/d_{nT})\hat{B}$ where $\hat{B} \xrightarrow{p} B$. The following estimator of the asymptotic variance of $\hat{\beta}$ based on \hat{W} is used throughout the remainder of the paper:

$$\begin{aligned} \widehat{Avar}(\hat{\beta}) &= \left(\sum_{i=1}^n x_i'x_i \right)^{-1} (nT\hat{W}) \left(\sum_{i=1}^n x_i'x_i \right)^{-1} \\ &= \left(\sum_{i=1}^n x_i'x_i \right)^{-1} \left(\sum_{i=1}^n x_i' \hat{\varepsilon}_i \hat{\varepsilon}_i' x_i \right) \left(\sum_{i=1}^n x_i'x_i \right)^{-1}. \end{aligned} \tag{4}$$

³This is especially relevant in Theorem 3 where the mixing conditions will not hold for the transformed variables if, for example, the transformation is to remove fixed effects by differencing out the individual means. Hansen (2006) provides conditions on the untransformed variables which will cover this case in a different but related context. This approach complicates the proof and notation and is not pursued here.

⁴It does, however, ignore the possibility of cross-sectional correlation, and it will be assumed that there is no cross-sectional correlation for the remainder of the paper.

In addition, for testing the hypothesis $R\beta = r$ for a $q \times k$ matrix R with rank q , the usual t (for R a $1 \times k$ vector) and Wald statistics can be defined as

$$t^* = \frac{\sqrt{nT}(R\hat{\beta} - r)}{\sqrt{R\hat{Q}^{-1}\hat{W}\hat{Q}^{-1}R'}} \quad (5)$$

and

$$F^* = nT(R\hat{\beta} - r)'[R\hat{Q}^{-1}\hat{W}\hat{Q}^{-1}R']^{-1}(R\hat{\beta} - r), \quad (6)$$

respectively, where \hat{W} is defined above and $\hat{Q} = (1/nT)\sum_{i=1}^n x_i'x_i$. In Section 3, I verify that, despite differences in the limiting behavior of $\hat{\beta}$, $t^* \xrightarrow{d} N(0, 1)$, $F^* \xrightarrow{d} \chi_q^2$, and $\widehat{Avar}(\hat{\beta})$ is valid for estimating the asymptotic variance of $\hat{\beta}$ as $n \rightarrow \infty$ regardless of the behavior of T . I also consider the behavior of t^* and F^* as $T \rightarrow \infty$ with n fixed. In this case, \hat{W} is not consistent for W but does have a limiting distribution; and when the data are iid across i ,⁵ I show that $t^* \xrightarrow{d} (n/(n-1))^{1/2}t_{n-1}$ and that F^* is asymptotically pivotal and so can be used to construct valid tests. This behavior suggests that inference using $(n/(n-1))\hat{W}$ and forming critical values using a t_{n-1} distribution will be valid regardless of the asymptotic sequence considered.

It is worth noting that the estimator \hat{W} has also been used extensively in multilevel models to account for the presence of correlation between individuals within cells; cf. [Liang and Zeger \(1986\)](#) and [Bell and McCaffrey \(2002\)](#). For example, in a schooling study, one might have data on individual outcomes where the individuals are grouped into classes. In this case, the cross-sectional unit of observation could be defined as the class, and arbitrary correlation between all individuals within each class could be allowed. In this case, one would expect the presence of a classroom specific random effect resulting in equicorrelation between all individuals within a class. While this would clearly violate the mixing assumptions imposed in obtaining the asymptotic behavior as $T \rightarrow \infty$ with n fixed, it would not invalidate the use of \hat{W} for inference about β in cases where n and T go to infinity jointly.

In addition to being useful for performing inference about $\hat{\beta}$, \hat{W} may also be used to test the specification of simple parametric models of the error process.⁶ Such a test may be useful for a number of reasons. If a parametric model is correct, the estimates of the variance of $\hat{\beta}$ based on this model will tend to behave better than the estimates obtained from \hat{W} . In particular, parametric estimates of the variance of $\hat{\beta}$ will often be considerably less variable and will typically converge faster than estimates made using \hat{W} ; and if the parametric model is deemed to be adequate, this model may be used to perform FGLS estimation. The FGLS estimator is asymptotically more efficient than the OLS estimator, and simulation evidence in [Hansen \(2006\)](#) suggests that the efficiency gain to using FGLS over OLS in serially correlated panel data may be substantial.

⁵Note that this still allows arbitrary correlation and heteroskedasticity within individuals but restricts that the pattern is the same across individuals.

⁶The test considered is a straightforward generalization of the test proposed by [White \(1980\)](#) for heteroskedasticity and was suggested in the panel context by [Kezdi \(2002\)](#).

To define the specification test, called hereafter the heteroskedasticity–autocorrelation (HA) test, let $\widehat{W}(\widehat{\theta}) = (1/nT)\sum_{i=1}^n x_i' \Omega_i(\widehat{\theta}) x_i$ where $\widehat{\theta}$ are estimates of a finite set of parameters describing the disturbance process and $\Omega_i(\widehat{\theta})$ is the implied covariance matrix for individual i .⁷ Define a test statistic

$$S^* = (nT)[\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))' \widehat{D}^- \text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))], \quad (7)$$

where \widehat{D} is a positive semi-definite weighting matrix that estimates the variance of $\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))$ and A^- is the generalized inverse of a matrix A .⁸ In the following section, it will be shown that $S^* \xrightarrow{d} \chi_{k(k+1)/2}^2$ for \widehat{D} defined below.

A natural choice for \widehat{D} is

$$\widehat{D} = \frac{1}{nT} \sum_{i=1}^n [(\text{vec}(x_i' \widehat{\varepsilon}_i \widehat{\varepsilon}_i' x_i - x_i' \Omega_i(\widehat{\theta}) x_i))(\text{vec}(x_i' \widehat{\varepsilon}_i \widehat{\varepsilon}_i' x_i - x_i' \Omega_i(\widehat{\theta}) x_i))']. \quad (8)$$

Under asymptotics where $\{n, T\} \rightarrow \infty$ jointly, another potential choice for \widehat{D} is an estimate of the asymptotic variance of \widehat{W} :

$$\widehat{V} = \frac{1}{nT} \sum_{i=1}^n [(\text{vec}(x_i' \widehat{\varepsilon}_i \widehat{\varepsilon}_i' x_i - \widehat{W}))(\text{vec}(x_i' \widehat{\varepsilon}_i \widehat{\varepsilon}_i' x_i - \widehat{W}))']. \quad (9)$$

That \widehat{V} provides an estimator of the variance of $\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))$ follows from the fact that as $\{n, T\} \rightarrow \infty$, $\text{vec}(\widehat{W})$ is \sqrt{n} -consistent while $\text{vec}(\widehat{W}(\widehat{\theta}))$ will be \sqrt{nT} -consistent in many cases, so $\text{vec}(\widehat{W}(\widehat{\theta}))$ may be taken as a constant relative to $\text{vec}(\widehat{W})$. The difference in rates of convergence would arise, for example, in a fixed effects panel model where the errors follow an AR process with common AR coefficients across individuals. However, it is important to note that this will not always be the case. In particular, in random effects models, the estimator of the variance of the individual specific shock will converge at only a \sqrt{n} rate, implying the same rate of convergence for both the robust and parametric estimators of the variance. In the following section, I outline the asymptotic properties of $\widehat{\beta}$, \widehat{W} , and \widehat{V} from which the behavior of t^* , F^* , and S^* will follow. The properties of \widehat{D} , though not discussed, will generally be the same as those of \widehat{V} under the different asymptotic sequences considered.

3. Asymptotic properties of the robust covariance matrix estimator

To develop the asymptotic inference results, I impose the following conditions.

⁷Consistency and asymptotic normality of $\widehat{W}(\widehat{\theta})$ will generally follow from consistency and asymptotic normality of $\widehat{\theta}$. In particular, defining $W_i(\theta)$ as the derivative of W with respect to θ_i and letting θ be a $p \times 1$ vector, a Taylor series expansion of $W(\widehat{\theta})$ yields $W(\widehat{\theta}) = W(\theta) + \sum_{i=1}^p W_i(\widehat{\theta})(\widehat{\theta} - \theta)$ where $\widehat{\theta}$ is an intermediate value. As long as a uniform law of large numbers applies to $W_i(\theta)$, $W(\widehat{\theta}) - W(\theta)$ will inherit the properties of $\widehat{\theta} - \theta$. The problem is then reduced to finding an estimator of θ that is consistent and asymptotically normal with a mean zero asymptotic distribution. Finding such an estimator in fixed effects panel models with serial correlation and/or heteroskedasticity when $n \rightarrow \infty$ and $T/n \rightarrow \rho$ where $\rho < \infty$ is complicated, though there are estimators which exist. See, for example, Nickell (1981), MaCurdy (1982), Solon (1984), Lancaster (2002), Hahn and Kuersteiner (2002), Hahn and Newey (2004), and Hansen (2006).

⁸The test could alternatively be defined by only considering the $(k(k+1))/2$ unique elements of $\widehat{W} - \widehat{W}(\widehat{\theta})$ and using the inverse of the implied covariance matrix. This test will be equivalent to the test outlined above.

Assumption 1. $\{x_i, \varepsilon_i\}$ are independent across i , and $E[\varepsilon_i \varepsilon_i' | x_i] = \Omega_i$.

Assumption 2. $Q_{nT} = E[\sum_{i=1}^n (x_i' x_i / nT)]$ is uniformly positive definite with constant limit Q where limits are taken as $n \rightarrow \infty$ with T fixed in Theorem 1, as $\{n, T\} \rightarrow \infty$ in Theorems 2 and 3, and as $T \rightarrow \infty$ with n fixed in Theorem 4.

In addition, I impose either Assumption 3(a) or Assumption 3(b) depending on the context.

Assumption 3.

- (a) $E[\varepsilon_i | x_i] = 0$.
- (b) $E[x_{it} \varepsilon_{it}] = 0$.

Assumptions 1–3 are quite standard for panel data models. Assumption 1 imposes independence across individuals, ruling out cross-sectional correlation, but leaves the time series correlation unconstrained and allows general heterogeneity across individuals. Assumption 2 is a standard full rank condition, and the restriction that Q_{nT} has a constant limit could be relaxed at the cost of more complicated notation. Assumption 3 imposes that one of two orthogonality conditions is satisfied. Assumption 3(b) imposes that x_{it} and ε_{it} are uncorrelated and is weaker than the strict exogeneity imposed in Assumption 3(a). Assumption 3(a) is stronger than necessary, but it simplifies the proof of asymptotic normality of \widehat{W} and consistency of \widehat{V} . In addition, Assumption 3(a) would typically be imposed in fixed effects models.⁹

The first theorem, which is stated here for completeness, collects the properties of $\widehat{\beta}$ and \widehat{W} in asymptotics where $n \rightarrow \infty$ with T fixed.

Theorem 1. *Suppose the data are generated by model (1), that Assumptions 1 and 2 are satisfied, and that $n \rightarrow \infty$ with T fixed.*

- (i) *If Assumption 3(b) holds and $E|x_{it}|^{4+\delta} < \Delta < \infty$ and $E|\varepsilon_{it}|^{4+\delta} < \Delta < \infty$ for some $\delta > 0$, then*

$$\sqrt{nT}(\widehat{\beta} - \beta) \xrightarrow{d} Q^{-1}N\left(0, W = \lim_n \frac{1}{nT} \sum_{i=1}^n E[x_i' \Omega_i x_i]\right),$$

and

$$\widehat{W} \xrightarrow{p} W.$$

- (ii) *In addition, if Assumption 3(a) holds and $E|x_{it}|^{8+\delta} < \Delta < \infty$ and $E|\varepsilon_{it}|^{8+\delta} < \Delta < \infty$ for some $\delta > 0$, then*

$$\begin{aligned} &\sqrt{nT}[\text{vec}(\widehat{W} - W)] \\ &\xrightarrow{d} N\left(0, V = \lim_n \frac{1}{nT} \sum_{i=1}^n E[(\text{vec}(x_i' \varepsilon_i \varepsilon_i' x_i - W))(\text{vec}(x_i' \varepsilon_i \varepsilon_i' x_i - W))']\right), \end{aligned}$$

⁹Note that a balanced panel has also implicitly been assumed. All of the results with the exception of Corollary 4.1 could be extended to accommodate unbalanced panels at the cost of more complicated notation.

and

$$\widehat{V} \xrightarrow{p} V.$$

Remark 3.1. It follows from Theorem 1(i) that the asymptotic variance of $\widehat{\beta}$ can be estimated using (4) since

$$\begin{aligned} \widehat{Avar}(\widehat{\beta}) &= \left(\sum_{i=1}^n x_i' x_i \right)^{-1} \sum_{i=1}^n x_i' \widehat{\varepsilon}_i \widehat{\varepsilon}_i' x_i \left(\sum_{i=1}^n x_i' x_i \right)^{-1} \\ &= \frac{1}{nT} \left(\frac{1}{nT} \sum_{i=1}^n x_i' x_i \right)^{-1} \widehat{W} \left(\frac{1}{nT} \sum_{i=1}^n x_i' x_i \right)^{-1} = \frac{1}{nT} \widehat{Q}^{-1} \widehat{W} \widehat{Q}^{-1}, \end{aligned}$$

where $\widehat{Q}^{-1} \widehat{W} \widehat{Q}^{-1} \xrightarrow{p} Q^{-1} W Q^{-1}$. It also follows immediately from the definitions of t^* and F^* in Eqs. (5) and (6) and Theorem 1(i) that, under the null hypothesis, $t^* \xrightarrow{d} N(0, 1)$ and $F^* \xrightarrow{d} \chi_q^2$. Similarly, using Theorem 1(ii) and assuming $\widehat{W}(\widehat{\theta})$ has properties similar to those of \widehat{W} , it will follow that the HA test statistic, S^* , formed using \widehat{D} defined above converges in distribution to a $\chi_{k(k+1)/2}^2$ under the null hypothesis.

Theorem 1 verifies that $\widehat{\beta}$ and \widehat{W} are consistent and asymptotically normal as $n \rightarrow \infty$ with T fixed without imposing any restrictions on the time series dimension. In the following results, I consider alternate asymptotic approximations under the assumption that both n and T are going to infinity.¹⁰ In these cases, consistency and asymptotic normality of suitably normalized versions of \widehat{W} are established under weak conditions.

Theorem 2, given immediately below, covers the case where n and T are going to infinity and there is not weak dependence in the time series. In particular, the results of Theorem 2 are only interesting in the case where $W = \lim_{n,T} (1/nT^2) \sum_{i=1}^n E[x_i' \Omega_i x_i] > 0$. Perhaps the leading case where this behavior would occur is in a model where ε_{it} includes an individual specific random effect that is uncorrelated to x_{it} and the estimated model does not include an individual specific effect. In this case, all observations for a given individual will be equicorrelated, and the condition given above will hold. Theorem 3, given following Theorem 2, covers the case where there is mixing in the time series.

Theorem 2. Suppose the data are generated by model (1), that Assumptions 1 and 2 are satisfied, and that $\{n, T\} \rightarrow \infty$ jointly.

- (i) If Assumption 3(b) holds and $E|x_{it}|^{4+\delta} < \Delta < \infty$ and $E|\varepsilon_{it}|^{4+\delta} < \Delta < \infty$ for some $\delta > 0$, then

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} Q^{-1} N(0, W) = \lim_{n,T} \frac{1}{nT^2} \sum_{i=1}^n E[x_i' \Omega_i x_i],$$

¹⁰One could also consider sequential limits in which one takes limits as n or T goes to infinity with the other dimension fixed and then lets the other dimension go to infinity. It could be shown that under the conditions of Theorem 2 and appropriate normalizations sequential limits taken first with respect to either n or T would yield the same results as the joint limit. Similarly, under the conditions of Theorem 3, the sequential limits taken first with respect to either n or T would produce the same results as the joint limit.

and

$$\widehat{W}/T \xrightarrow{P} W.$$

(ii) In addition, if Assumption 3(a) holds and $E|x_{iit}|^{8+\delta} < \Delta < \infty$ and $E|\varepsilon_{it}|^{8+\delta} < \Delta < \infty$ for some $\delta > 0$, then

$$\begin{aligned} &\sqrt{n}[\text{vec}(\widehat{W}/T - W)] \\ &\xrightarrow{d} N\left(0, V = \lim_{n,T} \frac{1}{nT^4} \sum_{i=1}^n E[(\text{vec}(x'_i \varepsilon_i \varepsilon'_i x_i - W))(\text{vec}(x'_i \varepsilon_i \varepsilon'_i x_i - W))']\right), \end{aligned}$$

and

$$\widehat{V}/T^3 \xrightarrow{P} V.$$

Remark 3.2. It is important to note that the results presented in Theorem 2 are not interesting in the setting where the $\{j, k\}$ element of Ω_i becomes small when $|j - k|$ is large since in these circumstances $(1/nT^2) \sum_{i=1}^n E[x'_i \Omega_i x_i] \rightarrow 0$. Theorem 3 presents results which are relevant in this case.

Remark 3.3. Theorem 2 verifies consistency and asymptotic normality of both $\widehat{\beta}$ and \widehat{W} while imposing essentially no constraints on the time series dependence in the data. The large cross-section effectively allows the time series dimension to be ignored even when T is large. However, without constraints on the time series, $\widehat{\beta}$ is \sqrt{n} -consistent, not \sqrt{nT} -consistent. Intuitively, the slower rate of convergence is due to the fact that there may be little information contained in the time series since it is allowed to be arbitrarily dependent.

Remark 3.4. The fact that $\widehat{\beta}$ and \widehat{W} are not \sqrt{nT} -consistent will not affect practical implementation of inference about $\widehat{\beta}$. In particular, the estimate of the asymptotic variance of $\widehat{\beta}$ based on Eq. (4) is

$$\begin{aligned} \widehat{Avar}(\widehat{\beta}) &= \left(\sum_{i=1}^n x'_i x_i \right)^{-1} \sum_{i=1}^n x'_i \widehat{\varepsilon}_i \widehat{\varepsilon}'_i x_i \left(\sum_{i=1}^n x'_i x_i \right)^{-1} \\ &= \frac{1}{n} \left(\frac{1}{nT} \sum_{i=1}^n x'_i x_i \right)^{-1} (\widehat{W}/T) \left(\frac{1}{nT} \sum_{i=1}^n x'_i x_i \right)^{-1} = \frac{1}{n} \widehat{Q}^{-1} (\widehat{W}/T) \widehat{Q}^{-1}, \end{aligned}$$

where $\widehat{Q}^{-1} (\widehat{W}/T) \widehat{Q}^{-1} \xrightarrow{P} Q^{-1} W Q^{-1}$. The t -statistic defined in Eq. (5) may also be expressed as

$$\begin{aligned} t^* &= \frac{\sqrt{nT}(R\widehat{\beta} - r)}{\sqrt{R\widehat{Q}^{-1} \widehat{W} \widehat{Q}^{-1} R'}} \\ &= \frac{\sqrt{n}(R\widehat{\beta} - r)}{\sqrt{R\widehat{Q}^{-1} (\widehat{W}/T) \widehat{Q}^{-1} R'}} \end{aligned}$$

which converges in distribution to a $N(0, 1)$ random variable under the null hypothesis, $R\beta = r$, by Theorem 2(i). Similarly, it follows that $F^* \xrightarrow{d} \chi^2_q$ under the null. Finally, the H_A

test statistic, S^* , defined above also satisfies

$$S^* = (nT)[\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))' \widehat{D}^- \text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))] \\ = n[\text{vec}(\widehat{W}/T - \widehat{W}(\widehat{\theta})/T)' (\widehat{D}/T^3)^- \text{vec}(\widehat{W}/T - \widehat{W}(\widehat{\theta})/T)],$$

which converges in distribution to a $\chi^2_{k(k+1)/2}$ under the conditions of the theorem and the additional assumption that $\widehat{W}(\widehat{\theta})$ behaves similarly to \widehat{V} .

The previous theorem establishes the properties of $\widehat{\beta}$ and the robust variance matrix estimator as n and T go to infinity jointly without imposing restrictions on the time series dependence. While the result is interesting, there are many cases in which one might expect the time series dependence to diminish over time. In the following theorem, the properties of $\widehat{\beta}$ and \widehat{W} are established under the assumption that the data are strong mixing in the time series dimension.

Theorem 3. *Suppose the data are generated by model (1), that Assumptions 1 and 2 are satisfied, and that $\{n, T\} \rightarrow \infty$ jointly.*

- (i) *If Assumption 3(b) is satisfied, $E|x_{it}|^{r+\delta} < \Delta$ and $E|\varepsilon_{it}|^{r+\delta} < \Delta$ for some $\delta > 0$, and $\{x_{it}, \varepsilon_{it}\}$ is a strong mixing sequence in t with α of size $-3r/(r-4)$ for $r > 4$,*

$$\sqrt{nT}(\widehat{\beta} - \beta) \xrightarrow{d} Q^{-1}N\left(0, W = \lim_{n,T} \frac{1}{nT} \sum_{i=1}^n E[x_i' \Omega_i x_i]\right)$$

and

$$\widehat{W} - W \xrightarrow{p} 0.$$

- (ii) *In addition, if Assumption 3(a) is satisfied, $E|x_{it}|^{r+\delta} < \Delta$ and $E|\varepsilon_{it}|^{r+\delta} < \Delta$ for some $\delta > 0$, and $\{x_{it}, \varepsilon_{it}\}$ is a strong mixing sequence in t with α of size $-7r/(r-8)$ for $r > 8$,*

$$\sqrt{n}[\text{vec}(\widehat{W} - W)] \\ \xrightarrow{d} N\left(0, V = \lim_{n,T} \frac{1}{nT^2} \sum_{i=1}^n E[(\text{vec}(x_i' \varepsilon_i \varepsilon_i' x_i - W))(\text{vec}(x_i' \varepsilon_i \varepsilon_i' x_i - W))']\right),$$

and

$$\widehat{V}/T \xrightarrow{p} V.$$

Remark 3.5. Theorem 3 verifies consistency and asymptotic normality of both $\widehat{\beta}$ and \widehat{W} under fairly conventional conditions on the time series dependence of the variables. The added restriction on the time series dependence allows estimation of β at the \sqrt{nT} -rate, which differs from the case above where $\widehat{\beta}$ is only \sqrt{n} -consistent. Intuitively, the increase in the rate of convergence is due to the fact that under the mixing conditions, the time series is more informative than in the case analyzed in Theorem 2.

Remark 3.6. It follows immediately from the conclusions of Theorem 3 and the definitions of $\widehat{Avar}(\widehat{\beta})$, t^* , and F^* in Eqs. (4)–(6) that $\widehat{Avar}(\widehat{\beta})$ is valid for estimating the asymptotic variance of $\widehat{\beta}$ and that $t^* \xrightarrow{d} N(0, 1)$ and $F^* \xrightarrow{d} \chi^2_q$ under the null hypothesis. The HA test

statistic, S^* , also satisfies

$$\begin{aligned} S^* &= (nT)[\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))' \widehat{D}^- \text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))] \\ &= n[\text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))' (\widehat{D}/T)^- \text{vec}(\widehat{W} - \widehat{W}(\widehat{\theta}))], \end{aligned}$$

which converges in distribution to a $\chi^2_{k(k+1)/2}$ under the conditions of the theorem and the assumption that \widehat{D} behaves similarly to \widehat{V} . In this case, \widehat{V} could also typically be used as the weighting matrix in forming S^* since it will often be the case that $\widehat{W}(\widehat{\theta})$ will be \sqrt{nT} -consistent while \widehat{W} is \sqrt{n} -consistent.

Theorems 1–3 establish that conventional estimators of the asymptotic variance of $\widehat{\beta}$ and t and F statistics formed using \widehat{W} have their usual properties as long as $n \rightarrow \infty$ regardless of the behavior of T . In addition, the results indicate that it is essentially only the size of n that matters for the asymptotic behavior of the estimators under these sequences. To complete the theoretical analysis, I present the asymptotic properties of \widehat{W} as $T \rightarrow \infty$ with n fixed below. The results are interesting in providing a justification for a commonly used procedure and in unifying the results and the different asymptotics considered.

Theorem 4. *Suppose the data are generated by model (1), that Assumptions 1, 2, and 3(b) are satisfied, and that $T \rightarrow \infty$ with n fixed. If $E|x_{it}|^{r+\delta} < \Delta$, $E|\varepsilon_{it}|^{r+\delta} < \Delta$, and $\{x_{it}, \varepsilon_{it}\}$ is a strong mixing sequence in t with α of size $-3r/(r-4)$ for $r > 4$, then*

$$\sqrt{nT}(\widehat{\beta} - \beta) \xrightarrow{d} Q^{-1}N(0, W), \quad x'_i x_i/nT - Q_i/n \xrightarrow{p} 0, \quad x'_i \varepsilon_i/\sqrt{nT} \xrightarrow{d} N(0, W_i/n),$$

and

$$\begin{aligned} \widehat{W} \xrightarrow{d} U &= \frac{1}{n} \sum_{i=1}^n (\Lambda_i B_i B'_i \Lambda_i - \Lambda_i B_i \left(\sum_{j=1}^n B'_j \Lambda_j \right) \left(\sum_{j=1}^n Q_j \right)^{-1} Q_i \\ &\quad - Q_i \left(\sum_{j=1}^n Q_j \right)^{-1} \left(\sum_{j=1}^n \Lambda_j B_j \right) B'_i \Lambda_i \\ &\quad + Q_i \left(\sum_{j=1}^n Q_j \right)^{-1} \left(\sum_{j=1}^n \Lambda_j B_j \right) \left(\sum_{j=1}^n B'_j \Lambda_j \right) \left(\sum_{j=1}^n Q_j \right)^{-1} Q_i, \end{aligned}$$

where $W_i = \lim_T (1/T)E[x'_i \Omega_i x_i]$, $W = \lim_T (1/nT) \sum_i E[x'_i \Omega_i x_i]$, $B_i \sim N(0, I_k)$ is a k -dimensional normal vector with $E[B_i B'_i] = 0$ and $\Lambda_i = W_i^{1/2}$.

Remark 3.7. Theorem 4 verifies that \widehat{W} is not consistent but does have a limiting distribution as $T \rightarrow \infty$ with n fixed. Unfortunately, the result here differs from results obtained in Phillips et al. (2003), Kiefer and Vogelsang (2002, 2005), and Vogelsang (2003) who consider HAC estimation in time series data without truncation in that how to construct asymptotically pivotal statistics from U is not immediately obvious. However, in one important special case, U is proportional to the true covariance matrix allowing construction of asymptotically pivotal tests.

Corollary 4.1. *Suppose the conditions of Theorem 4 are satisfied and that $Q_i = Q$ and $W_i = W$ for all i . Then*

$$\widehat{W} \xrightarrow{d} U = \frac{1}{n} \Lambda \left(\sum_{i=1}^n B_i B'_i - \frac{1}{n} \sum_{i=1}^n B_i \sum_{i=1}^n B'_i \right) \Lambda$$

for B_i defined in Theorem 4 and $\Lambda = W^{1/2}$. Then, for testing the null hypothesis $H_0 : R\beta = r$ against the alternative $H_1 : R\beta \neq r$ for a $q \times k$ matrix R with rank q , the limiting distributions of the conventional Wald (F^*) and t -type (t^*) tests under H_0 are

$$F^* = (nT)(R\hat{\beta} - r)'[R\hat{Q}^{-1}\hat{W}\hat{Q}^{-1}R']^{-1}(R\hat{\beta} - r) \xrightarrow{d} \tilde{B}_{q,n}' \left[\frac{1}{n} \left(\sum_i B_{q,i} B_{q,i}' - \tilde{B}_{q,n} \tilde{B}_{q,n}' \right) \right]^{-1} \tilde{B}_{q,n} = \frac{nq}{n-q} F_{q,n-q}, \tag{10}$$

and

$$t^* = \frac{\sqrt{nT}(R\hat{\beta} - r)}{\sqrt{R\hat{Q}^{-1}\hat{W}\hat{Q}^{-1}R'}} \xrightarrow{d} \frac{\tilde{B}_{1,n}}{\sqrt{(1/n)(\sum_i B_{1,i}^2 - \tilde{B}_{1,n}^2)}} = \sqrt{\frac{n}{n-1}} t_{n-1}, \tag{11}$$

where $B_{q,i} \sim N(0, I_q)$, $\tilde{B}_{q,n} = (1/\sqrt{n})\sum_{i=1}^n B_{q,i}$, t_{n-1} is a t distribution with $n - 1$ degrees of freedom, and $F_{q,n-q}$ is an F distribution with q numerator and $n - q$ denominator degrees of freedom.

Corollary 4.1 gives the limiting distribution of \hat{W} as $T \rightarrow \infty$ under the additional restriction that $Q_i = Q$ and $W_i = W$ for all i . These restrictions would be satisfied when the data vectors for each individual $\{x_i, y_i\}$ are iid across i . While this is more restrictive than the condition imposed in Assumption 1, it still allows for quite general forms of conditional heteroskedasticity and does not impose any structure on the time series process within individuals.

The most interesting feature about the result in Corollary 4.1 is that under the conditions imposed, the limiting distribution of \hat{W} is proportional to the actual covariance matrix in the data. This allows construction of asymptotically pivotal statistics based on standard t and Wald tests as in Phillips et al. (2003), Kiefer and Vogelsang (2002, 2005), and Vogelsang (2003). This is particularly convenient in the panel case since the limiting distribution of the t -statistic is exactly $\sqrt{(n/(n-1))} t_{n-1}$ where t_{n-1} denotes the t distribution with $n - 1$ degrees of freedom.¹¹ It is also interesting that $EU = (1 - (1/n))W$. This suggests normalizing the estimator \hat{W} by $n/(n-1)$ will result in an asymptotically unbiased estimator in asymptotics where $T \rightarrow \infty$ with n fixed and will likely reduce the finite-sample bias under asymptotics where $n \rightarrow \infty$. In addition, the t -statistic constructed based on the estimator defined by $(n/(n-1))\hat{W}$ will be asymptotically distributed as a t_{n-1} for which critical values are readily available.¹²

The conclusions of Corollary 4.1 suggest a simple procedure for testing hypotheses regarding regression coefficients which will be valid under any of the asymptotics considered. Using $(n/(n-1))\hat{W}$ and obtaining critical values from a t_{n-1} distribution will yield tests which are asymptotically valid regardless of the asymptotic sequence since the

¹¹If $n = 1$, \hat{W} is identically equal to 0. In this case, it is easy to verify that U equals 0, though the results of Theorem 4 and Corollary 4.1 are obviously uninteresting in this case.

¹²This is essentially the normalization used in Stata's *cluster* command, which normalizes \hat{W} by $[(nT - 1)/(nT - k)]/[n/(n - 1)]$, where the normalization is motivated as a finite-sample adjustment under the usual $n \rightarrow \infty, T$ fixed asymptotics; see *Stata User's Guide Release 8*, p. 275 (Stata Corporation, 2003).

$t_{n-1} \rightarrow N(0, 1)$ and $n/(n-1) \rightarrow 1$ as $n \rightarrow \infty$. Thus, this approach will yield valid tests under any of the asymptotics considered in the presence of quite general heteroskedasticity and serial correlation.¹³

In addition, it is important to note that in the cases where there is weak dependence in the time series and T is large, more efficient estimators of the covariance matrix which make use of this information are available. In particular, standard time series HAC estimators which downweight the correlation between observations that are far apart will have faster rates of convergence than the CCM estimator.

Finally, it is worth noting that the maximum rank of \widehat{W} will generally be $n-1$, which suggests that \widehat{W} will be rank deficient when $k > n-1$. Since \widehat{W} is supposed to estimate a full rank matrix, it seems likely that inference based on \widehat{W} will perform poorly in these cases. Also, the above development ignores time effects, which will often be included in panel data models. Under T fixed, $n \rightarrow \infty$ asymptotics, the time effects can be included in the covariate vector x_{it} and pose no additional complications. However, as $T \rightarrow \infty$, they also need to be considered separately from x and partialled out with the individual fixed effects. This partialing out will generally result in the presence of an $O(1/n)$ correlation between individuals. When n is large, this correlation should not matter, but in the fixed n , $T \rightarrow \infty$ case, it will invalidate the results. The effect of the presence of time effects was explored in a simulation study with the same design as that reported in the following section where each model was estimated including a full set of time fixed effects. The results, which are not reported below but are available upon request, show that tests based on \widehat{W} are somewhat more size distorted than when no time effects are included for small n , but that this size distortion diminishes quickly as n increases.

4. Monte Carlo evidence

The asymptotic results presented above suggest that tests based on the robust standard error estimates should have good properties regardless of the relative sizes of n and T . I report results from a simple simulation study used to assess the finite sample effectiveness of the robust covariance matrix estimator and tests based upon it below. Specifically, the simulation focuses on t -tests for regression coefficients and the HA test discussed above.

The Monte Carlo simulations are based on two different specifications: a “fixed effect” specification and a “random effects” specification. The terminology refers to the fact that in the “fixed effect” specification, the models will be estimated including individual specific fixed effects with the goal of focusing on the case where the underlying disturbances exhibit weak dependence. In the “random effects” specification individual specific effects are not estimated and the goal is to examine the behavior of the CCM estimator and tests based upon it in an equicorrelated model.

The fixed effect specification is

$$y_{it} = x'_{it}\beta + \alpha_i + \varepsilon_{it},$$

where x_{it} is a scalar and α_i is an individual specific effect. The data generating process for the fixed effect specification allows for serial correlation in both x_{it} and ε_{it} and

¹³This argument also applies to testing multiple parameters using F^* .

heteroskedasticity:

$$\begin{aligned}x_{it} &= .5x_{it-1} + v_{it}, \quad v_{it} \sim \mathbf{N}(0, .75), \\ \varepsilon_{it} &= \rho\varepsilon_{it-1} + \sqrt{a_0 + a_1x_{it}^2}u_{it}, \quad u_{it} \sim \mathbf{N}(0, 1 - \rho^2), \\ \alpha_i &\sim \mathbf{N}(0, .5).\end{aligned}$$

Data are simulated using four different values of ρ , $\rho \in \{0, .3, .6, .9\}$, in both the homoskedastic ($a_0 = 1$, $a_1 = 0$) and heteroskedastic ($a_0 = a_1 = .5$) cases, resulting in a total of eight distinct parameter settings. The models are estimated including x_{it} and a full set of individual specific fixed effects.¹⁴

The random effects specifications is

$$y_{it} = x'_{it}\beta + \varepsilon_{it},$$

where x_{it} is a normally distributed scalar with $E[x_{it}^2] = 1$ and $E[x_{it_1}x_{it_2}] = .8$ for all $t_1 \neq t_2$. ε_{it} contains an individual specific random component and a random error term:

$$\begin{aligned}\varepsilon_{it} &= \alpha_i + u_{it}, \\ \alpha_i &\sim \mathbf{N}(0, \rho), \\ u_{it} &\sim \mathbf{N}(0, 1 - \rho).\end{aligned}$$

Note that the random effects data generating process implies that $E[\varepsilon_{it_1}\varepsilon_{it_2}] = \rho$ for $t_1 \neq t_2$. Three values of ρ are employed for the random effects specification: .3, .6, and .9. The model is estimated by regressing y_{it} on x_{it} and a constant.

The fixed effects model is commonly used in empirical work when panel data are available. The random effects specification is also widely used in the policy evaluation literature. In many policy evaluation studies, the covariate of interest is a policy variable that is highly correlated within aggregate cells, often with a correlation of one, which has led to the dominance of the random effects estimator in this context. For example, a researcher may desire to estimate the effect of classroom level policies on student-level micro data containing observations from multiple classrooms. In this setting, T indexes the number of students within each class, n indexes the number of classrooms, and α_i is a classroom specific random effect. The CCM estimator has been widely utilized in such situations in order to consistently estimate standard errors.¹⁵

Simulation results for various values of the cross-sectional (n) and time (T) dimensions are reported. For each $\{n, T\}$ combination, reported results for each of the 11 parameter settings (eight for the fixed effects specification and three for the random effects specification) are based on 1,000 simulation repetitions. Each simulation estimates three types of standard errors for $\hat{\beta}$: unadjusted OLS standard errors, $\hat{\sigma}_{OLS}$, CCM standard errors, $\hat{\sigma}_{CLUS}$, and standard errors consistent with an AR(1) process, $\hat{\sigma}_{AR(1)}$.¹⁶ For the

¹⁴Since α_i is uncorrelated with x_i , this model could be estimated using random effects. I chose to consider a different specification for the random effects estimates where the x_{it} were generated to more closely resemble covariates which appear in policy analysis studies.

¹⁵This is, in fact, one of the original motivations for the development of the CCM estimator, cf. Liang and Zeger (1986).

¹⁶ $\hat{\sigma}_{AR(1)}$ imposes the parametric structure implied by an AR(1) process. The ρ parameter is estimated from the OLS residuals using the procedure described in Hansen (2006) which consistently estimates AR parameters in fixed effects panel models. The standard errors are then computed as $(X'X)^{-1}X'\Omega(\hat{\rho})X(X'X)^{-1}$ where $\Omega(\hat{\rho})$ is the covariance matrix implied by an AR(1) process.

random effects specification, standard errors consistent with random effects, \widehat{s}_{RE} , are substituted for $\widehat{s}_{AR(1)}$.¹⁷ \widehat{s}_{CLUS} is consistent for all parameter settings. \widehat{s}_{OLS} is consistent only in the iid case (the homoskedastic data generating process with $\rho = 0$). $\widehat{s}_{AR(1)}$ is consistent in all homoskedastic data generating processes, and \widehat{s}_{RE} is consistent in all models for which it is reported. In all cases, the CCM estimator is computed using the normalization implied by $T \rightarrow \infty$ with n fixed asymptotics; that is, the CCM estimator is computed as $(n/(n-1))\widehat{W}$ for \widehat{W} defined in Eq. (3).

Tables 1–4 present the results of the Monte Carlo study, where each table corresponds to a different $\{n, T\}$ combination.¹⁸ In each table, Panel A presents the fixed effects results for the homoskedastic and heteroskedastic cases, while Panel B presents the random effects results. Column (1) presents t -test rejection rates for 5% level tests based on OLS, CCM, and AR(1) standard errors. The critical values for tests based on OLS and AR(1) errors are taken from a t_{nT-n-1} distribution, and the critical values for tests based on clustered standard errors are taken from a t_{n-1} distribution. Columns (2) and (3) present the mean and standard deviation of the estimated standard errors respectively. Column (4) presents the standard deviation of the $\widehat{\beta}$'s. The difference between columns (2) and (4) is therefore the bias of the estimated standard errors. Finally, column (5) presents the rejection rates for the HA test described above which tests the null hypothesis that both the CCM estimator and the parametric estimator are consistent.

As expected, tests based on \widehat{s}_{OLS} and $\widehat{s}_{AR(1)}$ perform well in the cases where the assumed model is consistent with the data across the full range of n and T combinations. The results are also consistent with the asymptotic theory, clearly illustrating the \sqrt{nT} -consistency of $\widehat{\beta}$ and \widehat{W} with the bias of \widehat{W} and the variance of both $\widehat{\beta}$ and \widehat{W} decreasing as either n or T increases. Of course, when the assumed parametric model is inconsistent with the data, tests based on parametric standard errors suffer from size distortions and the standard error estimates are biased. The RE tests have the correct size for moderate and large n , but not for small n (i.e. $n = 10$); and as indicated by the asymptotic theory, the T dimension has no apparent impact on the size of RE based tests or the overall performance of the RE estimates.

Tests based on the CCM estimator have approximately correct size across all combinations of n and T and all models of the disturbances considered in the fixed effect specification. The estimator does, however, display a moderate bias in the small n case; it seems likely that this bias does not translate into a large size distortion due to the fact that the bias is small relative to the standard error of the estimator and the use of the t_{n-1} distribution to obtain the critical values. While the clustered standard errors perform well in terms of size of tests and reasonably well in terms of bias, the simulations reveal that a potential weakness of the clustered estimator is a relatively high variance. The CCM estimates have a substantially higher standard deviation than the other estimators and this difference, in percentage terms, increases with T . This behavior is consistent with the

¹⁷ \widehat{s}_{RE} is estimated in a manner analogous to $\widehat{s}_{AR(1)}$ where the covariance parameters are estimated in the usual manner from the OLS and within residuals.

¹⁸Tables 1–4 correspond to $\{n, T\} = \{10, 10\}$, $\{n, T\} = \{10, 50\}$, $\{n, T\} = \{50, 10\}$, $\{n, T\} = \{50, 50\}$, respectively. Additional results for $\{n, T\} = \{10, 200\}$, $\{n, T\} = \{50, 20\}$, $\{n, T\} = \{50, 200\}$, $\{n, T\} = \{200, 10\}$, and $\{n, T\} = \{200, 50\}$ are available from the author upon request. The results are consistent with the asymptotic theory with the performance of the CCM estimator improving as either n or T increases in the fixed effects specification and as n increases in the random effects specification. In the random effects case, the performance does not appear to be greatly influenced by the size of T relative to n .

Table 1

Data generating process	<i>t</i> -test rejection rate (1)	Mean (s.e.) (2)	Std (s.e.) (3)	Std (β) (4)	HA test rejection rate (5)
<i>N</i> = 10, <i>T</i> = 10					
<i>A. Fixed effects</i>					
Homoskedastic, $\rho = 0$					
OLS	0.038	0.1180	0.0133	0.1152	0.152
Cluster	0.043	0.1149	0.0330	0.1152	
AR1	0.041	0.1170	0.0141	0.1152	0.135
Homoskedastic, $\rho = .3$					
OLS	0.082	0.1130	0.0136	0.1269	0.095
Cluster	0.054	0.1212	0.0357	0.1269	
AR1	0.055	0.1240	0.0161	0.1269	0.133
Homoskedastic, $\rho = .6$					
OLS	0.093	0.1005	0.0133	0.1231	0.074
Cluster	0.060	0.1167	0.0352	0.1231	
AR1	0.051	0.1219	0.0181	0.1231	0.123
Homoskedastic, $\rho = .9$					
OLS	0.145	0.0609	0.0090	0.0818	0.038
Cluster	0.053	0.0772	0.0249	0.0818	
AR1	0.054	0.0795	0.0136	0.0818	0.085
Heteroskedastic, $\rho = 0$					
OLS	0.126	0.1150	0.0126	0.1502	0.051
Cluster	0.057	0.1410	0.0458	0.1502	
AR1	0.126	0.1140	0.0137	0.1502	0.042
Heteroskedastic, $\rho = .3$					
OLS	0.171	0.1165	0.0137	0.1708	0.036
Cluster	0.068	0.1538	0.0500	0.1708	
AR1	0.143	0.1284	0.0172	0.1708	0.044
Heteroskedastic, $\rho = .6$					
OLS	0.187	0.1238	0.0153	0.1853	0.027
Cluster	0.074	0.1717	0.0572	0.1853	
AR1	0.117	0.1503	0.0219	0.1853	0.049
Heteroskedastic, $\rho = .9$					
OLS	0.198	0.1406	0.0209	0.2181	0.031
Cluster	0.087	0.1872	0.0641	0.2181	
AR1	0.097	0.1830	0.0336	0.2181	0.074
<i>B. Random effects</i>					
$\rho = .3$					
OLS	0.295	0.1063	0.0231	0.1926	0.017
Cluster	0.115	0.1561	0.0609	0.1926	
RE	0.097	0.1693	0.0460	0.1926	0.027
$\rho = .6$					
OLS	0.399	0.1030	0.0248	0.2438	0.054
Cluster	0.118	0.2024	0.0788	0.2438	
RE	0.094	0.2180	0.0600	0.2438	0.023
$\rho = .9$					
OLS	0.482	0.0987	0.0293	0.2925	0.093
Cluster	0.108	0.2346	0.0909	0.2925	
RE	0.095	0.2546	0.0723	0.2925	0.018

Table 2

Data generating process	<i>t</i> -test rejection rate (1)	Mean (s.e.) (2)	Std (s.e.) (3)	Std (β) (4)	HA test rejection rate (5)
<i>N</i> = 10, <i>T</i> = 50					
<i>A. Fixed effects</i>					
Homoskedastic, $\rho = 0$					
OLS	0.054	0.0462	0.0024	0.0472	0.184
Cluster	0.050	0.0449	0.0117	0.0472	
AR1	0.057	0.0460	0.0026	0.0472	0.185
Homoskedastic, $\rho = .3$					
OLS	0.088	0.0459	0.0024	0.0519	0.077
Cluster	0.043	0.0520	0.0133	0.0519	
AR1	0.050	0.0529	0.0031	0.0519	0.159
Homoskedastic, $\rho = .6$					
OLS	0.155	0.0447	0.0028	0.0590	0.049
Cluster	0.042	0.0574	0.0150	0.0590	
AR1	0.047	0.0598	0.0044	0.0590	0.184
Homoskedastic, $\rho = .9$					
OLS	0.225	0.0372	0.0034	0.0600	0.046
Cluster	0.046	0.0562	0.0159	0.0600	
AR1	0.049	0.0583	0.0072	0.0600	0.150
Heteroskedastic, $\rho = 0$					
OLS	0.158	0.0459	0.0021	0.0637	0.052
Cluster	0.051	0.0606	0.0169	0.0637	
AR1	0.162	0.0458	0.0023	0.0637	0.057
Heteroskedastic, $\rho = .3$					
OLS	0.199	0.0479	0.0022	0.0724	0.046
Cluster	0.041	0.0735	0.0198	0.0724	
AR1	0.142	0.0553	0.0032	0.0724	0.047
Heteroskedastic, $\rho = .6$					
OLS	0.229	0.0558	0.0031	0.0934	0.067
Cluster	0.043	0.0928	0.0260	0.0934	
AR1	0.112	0.0748	0.0054	0.0934	0.059
Heteroskedastic, $\rho = .9$					
OLS	0.239	0.0857	0.0079	0.1490	0.059
Cluster	0.046	0.1428	0.0451	0.1490	
AR1	0.076	0.1338	0.0163	0.1490	0.099
<i>B. Random effects</i>					
$\rho = .3$					
OLS	0.568	0.0471	0.0092	0.1636	0.147
Cluster	0.104	0.1356	0.0547	0.1636	
RE	0.097	0.1475	0.0413	0.1626	0.014
$\rho = .6$					
OLS	0.703	0.0466	0.0105	0.2331	0.212
Cluster	0.104	0.1897	0.0727	0.2331	
RE	0.095	0.2079	0.0567	0.2331	0.007
$\rho = .9$					
OLS	0.744	0.0450	0.0130	0.2785	0.245
Cluster	0.106	0.2310	0.0920	0.2785	
RE	0.103	0.2539	0.0701	0.2785	0.014

Table 3

Data generating process	<i>t</i> -test rejection rate (1)	Mean (s.e.) (2)	Std (s.e.) (3)	Std (β) (4)	HA test rejection rate (5)
<i>N</i> = 50, <i>T</i> = 10					
<i>A. Fixed effects</i>					
Homoskedastic, $\rho = 0$					
OLS	0.049	0.0522	0.0026	0.0526	0.106
Cluster	0.057	0.0515	0.0062	0.0526	
AR1	0.047	0.0522	0.0028	0.0526	0.099
Homoskedastic, $\rho = .3$					
OLS	0.080	0.0500	0.0027	0.0569	0.053
Cluster	0.059	0.0552	0.0072	0.0569	
AR1	0.055	0.0556	0.0033	0.0569	0.092
Homoskedastic, $\rho = .6$					
OLS	0.102	0.0447	0.0026	0.0539	0.132
Cluster	0.048	0.0549	0.0071	0.0539	
AR1	0.049	0.0553	0.0037	0.0539	0.072
Homoskedastic, $\rho = .9$					
OLS	0.156	0.0273	0.0273	0.0387	0.220
Cluster	0.075	0.0364	0.0367	0.0387	
AR1	0.067	0.0367	0.0367	0.0387	0.078
Heteroskedastic, $\rho = 0$					
OLS	0.119	0.0517	0.0025	0.0659	0.213
Cluster	0.047	0.0673	0.0093	0.0659	
AR1	0.116	0.0516	0.0028	0.0659	0.210
Heteroskedastic, $\rho = .3$					
OLS	0.197	0.0521	0.0026	0.0768	0.369
Cluster	0.062	0.0741	0.0114	0.0768	
AR1	0.139	0.0581	0.0033	0.0768	0.140
Heteroskedastic, $\rho = .6$					
OLS	0.214	0.0558	0.0031	0.0840	0.451
Cluster	0.048	0.0820	0.0126	0.0840	
AR1	0.108	0.0688	0.0045	0.0840	0.056
Heteroskedastic, $\rho = .9$					
OLS	0.152	0.0623	0.0043	0.0883	0.324
Cluster	0.038	0.0899	0.0144	0.0883	
AR1	0.057	0.0834	0.0070	0.0883	0.023
<i>B. Random effects</i>					
$\rho = .3$					
OLS	0.291	0.0451	0.0041	0.0822	0.673
Cluster	0.062	0.0776	0.0135	0.0822	
RE	0.059	0.0788	0.0091	0.0822	0.058
$\rho = .6$					
OLS	0.357	0.0452	0.0049	0.1034	0.892
Cluster	0.073	0.1004	0.0183	0.1034	
RE	0.068	0.1028	0.0127	0.1034	0.054
$\rho = .9$					
OLS	0.497	0.0447	0.0056	0.1246	0.943
Cluster	0.062	0.1192	0.0212	0.1246	
RE	0.063	0.1210	0.0147	0.1246	0.048

Table 4

Data generating process	<i>t</i> -test rejection rate (1)	Mean (s.e.) (2)	Std (s.e.) (3)	Std (β) (4)	HA test rejection rate (5)
<i>N</i> = 50, <i>T</i> = 20					
<i>A. Fixed effects</i>					
Homoskedastic, $\rho = 0$					
OLS	0.050	0.0342	0.0013	0.0341	0.097
Cluster	0.049	0.0341	0.0040	0.0341	
AR1	0.052	0.0342	0.0014	0.0341	0.088
Homoskedastic, $\rho = .3$					
OLS	0.094	0.0334	0.0013	0.0393	0.077
Cluster	0.051	0.0379	0.0045	0.0393	
AR1	0.056	0.0382	0.0016	0.0393	0.086
Homoskedastic, $\rho = .6$					
OLS	0.120	0.0315	0.0014	0.0414	0.300
Cluster	0.059	0.0407	0.0052	0.0414	
AR1	0.050	0.0412	0.0021	0.0414	0.092
Homoskedastic, $\rho = .9$					
OLS	0.200	0.0222	0.0013	0.0336	0.580
Cluster	0.059	0.0327	0.0047	0.0336	
AR1	0.060	0.0329	0.0024	0.0336	0.094
Heteroskedastic, $\rho = 0$					
OLS	0.168	0.0340	0.0011	0.0479	0.408
Cluster	0.063	0.0458	0.0056	0.0479	
AR1	0.171	0.0340	0.0012	0.0479	0.406
Heteroskedastic, $\rho = .3$					
OLS	0.209	0.0350	0.0012	0.0536	0.675
Cluster	0.051	0.0527	0.0068	0.0536	
AR1	0.145	0.0399	0.0016	0.0536	0.294
Heteroskedastic, $\rho = .6$					
OLS	0.228	0.0394	0.0017	0.0653	0.802
Cluster	0.050	0.0636	0.0084	0.0653	
AR1	0.119	0.0514	0.0027	0.0653	0.123
Heteroskedastic, $\rho = .9$					
OLS	0.196	0.0507	0.0028	0.0775	0.681
Cluster	0.036	0.0809	0.0131	0.0775	
AR1	0.058	0.0751	0.0056	0.0775	0.034
<i>B. Random effects</i>					
$\rho = .3$					
OLS	0.405	0.0320	0.0029	0.0756	0.915
Cluster	0.069	0.0726	0.0131	0.0756	
RE	0.063	0.0738	0.0085	0.0756	0.064
$\rho = .6$					
OLS	0.515	0.0318	0.0033	0.1012	0.944
Cluster	0.066	0.0976	0.0169	0.1012	
RE	0.055	0.0996	0.0118	0.1012	0.055
$\rho = .9$					
OLS	0.614	0.0314	0.0038	0.1203	0.948
Cluster	0.054	0.1166	0.0204	0.1203	
RE	0.051	0.1194	0.0140	0.1203	0.053

\sqrt{n} -consistency of the estimator and does suggest that if a parametric estimator is available, it may have better properties for estimating the variance of $\hat{\beta}$.

The clustered estimator performs less well in the random effects specification. For small n , tests based on the CCM estimator suffer from a substantial size distortion for all values of T . For moderate to large values of n , the tests have the correct size, and the overall performance does not appear to depend on T . In addition, the variance of $\hat{\beta}$ does not appear to decrease as T increases. These results are consistent with the lack of \sqrt{nT} -consistency in this case.¹⁹

The performance of the HA test is much less robust than that of t -tests based on clustered standard errors. For small n , the tests are badly size distorted and have essentially no power against any alternative hypotheses. As n and T grow, the test performance improves. With $n = 50$, the test remains size distorted, but it does have some power against alternatives that increases as T increases. The HA test also performs poorly for the random effects specification for small n . However, for moderate or large n , the test has both the correct size and good power.

Overall, the simulation results support the use of clustered standard errors for performing inference on regression coefficient estimates in serially correlated panel data, though they also suggest care should be taken if n is small and one suspects a “random effects” structure. The poor performance of \hat{W} in “random effects” models with small n is already well-known; see for example Bell and McCaffrey (2002) who also suggest a bias reduction for \hat{W} in this case. However, that the estimator does quite well even for small n in the serially correlated case where the errors are mixing is somewhat surprising and is a new result which is suggested by the asymptotic analysis of the previous section. The simulation results confirm the asymptotic results, suggesting that the clustered standard errors are consistent as long as $n \rightarrow \infty$ and that they are not sensitive to the size of n relative to T . The chief drawback of the CCM estimator is that the robustness comes at the cost of increasing the variance of the standard error estimate relative to that of standard errors estimated through more parsimonious models.

The HA test offers one simple information based criterion for choosing between the CCM estimator and a simple parametric model of the error process. However, the simulation evidence regarding its usefulness is mixed. In particular, the properties of the test are poor in small sample settings where there is likely to be the largest gain to using a parsimonious model. However, in moderate sized samples, the test performs reasonably well, and there still may be gains to using a simple parametric model in these cases.

5. Conclusion

This paper explores the asymptotic behavior of the robust covariance matrix estimator of Arellano (1987). It extends the usual analysis performed under asymptotics where $n \rightarrow \infty$ with T fixed to cases where n and T go to infinity jointly, considering both non-mixing and mixing cases, and to the case where $T \rightarrow \infty$ with n fixed. The limiting behavior of the OLS estimator, $\hat{\beta}$, in each case is different. However, the analysis shows that the conventional estimator of the asymptotic variance and the usual t and F statistics have the same properties regardless of the behavior of the time series as long as $n \rightarrow \infty$. In addition,

¹⁹The inconsistency of $\hat{\beta}$ when T increases with n fixed in differences-in-differences and policy evaluation studies has also been discussed in Donald and Lang (2001).

when $T \rightarrow \infty$ with n fixed and the data satisfy mixing conditions and an iid assumption across individuals, the usual t and F statistics can be used for inference despite the fact that the robust covariance matrix estimator is not consistent but converges in distribution to a limiting random variable. In this case, it is shown that the t statistic constructed using $n/(n-1)$ times the estimator of Arellano (1987) is asymptotically t_{n-1} , suggesting the use of $n/(n-1)$ times the estimator of Arellano (1987) and critical values obtained from a t_{n-1} in all cases. The use of this procedure is also supported in a short simulation experiment, which verifies that it produces tests with approximately correct size regardless of the relative size of n and T in cases where the time series correlation between observations diminishes as the distance between observations increases. The simulations also verify that tests based on the robust standard errors are consistent as n increases regardless of the relative size of n and T even in cases when the data are equicorrelated.

Acknowledgments

The research reported in this paper was motivated through conversations with Byron Lutz, to whom I am very grateful for input in developing this paper. I would like to thank Whitney Newey and Victor Chernozhukov as well as anonymous referees and a coeditor for helpful comments and suggestions. This work was partially supported by the William S. Fishman Faculty Research Fund at the Graduate School of Business, the University of Chicago. All remaining errors are mine.

Appendix

For brevity, sketches of the proofs are provided below. More detailed versions are available in an additional Technical Appendix from the author upon request and in Hansen (2004).

Proof of Theorem 1. $\hat{\beta} - \beta \xrightarrow{p} 0$ and $\sqrt{nT}(\hat{\beta} - \beta) \xrightarrow{d} Q^{-1}N(0, W = \lim_n(1/nT)\sum_{i=1}^n E[x_i' \Omega_i x_i])$ follow immediately under the conditions of Theorem 1 from the Markov LLN and the Liapounov CLT. The remaining conclusions follow from repeated use of the Cauchy–Schwarz inequality, Minkowski’s inequality, the Markov LLN, and the Liapounov CLT. \square

The proofs of Theorems 2 and 3 make use of the following lemmas which provide a LLN and CLT for inid data as $\{n, T\} \rightarrow \infty$ jointly.

Lemma 1. Suppose $\{Z_{i,T}\}$ are independent across i for all T with $E[Z_{i,T}] = \mu_{i,T}$ and $E|Z_{i,T}|^{1+\delta} < \Delta < \infty$ for some $\delta > 0$ and all i, T . Then $(1/n)\sum_{i=1}^n (Z_{i,T} - \mu_{i,T}) \xrightarrow{p} 0$ as $\{n, T\} \rightarrow \infty$ jointly.

Proof. The proof follows from standard arguments, cf. Chung (2001) Chapter 5. Details are given in Hansen (2004). \square

Lemma 2. For $k \times 1$ vectors $Z_{i,T}$, suppose $\{Z_{i,T}\}$ are independent across i for all T with $E[Z_{i,T}] = 0$, $E[Z_{i,T}Z_{i,T}'] = \Omega_{i,T}$, and $E\|Z_{i,T}\|^{2+\delta} < \Delta < \infty$ for some $\delta > 0$. Assume $\Omega = \lim_{n,T}(1/n)\sum_{i=1}^n \Omega_{i,T}$ is positive definite with minimum eigenvalue $\lambda_{\min} > 0$. Then $(1/\sqrt{n})\sum_{i=1}^n Z_{i,T} \xrightarrow{d} N(0, W)$ as $\{n, T\} \rightarrow \infty$ jointly.

Proof. The result follows from verifying the Lindeberg condition of Theorem 2 in Phillips and Moon (1999) using an argument similar to that used in the proof of Theorem 3 in Phillips and Moon (1999). Details are given in Hansen (2004). □

Proof of Theorem 2. The conclusions follow from conventional arguments making repeated use of the Cauchy–Schwarz inequality, Minkowski’s inequality, and Lemmas 1 and 2. □

In addition to using Lemmas 1 and 2, I make use of the following mixing inequality, restated from Doukhan (1994) Theorem 2 with a slight change of notation, to establish the properties of the estimators as $\{n, T\} \rightarrow \infty$ when mixing conditions are imposed. Its proof may be found in Doukhan (1994, p. 25–30).

Lemma 3. Let $\{z_t\}$ be a strong mixing sequence with $E[z_t] = 0$, $E\|z_t\|^{\tau+\varepsilon} < \Delta < \infty$, and mixing coefficient $\alpha(m)$ of size $(1 - c)r/(r - c)$ where $c \in 2\mathbb{N}$, $c \geq \tau$, and $r > c$. Then there is a constant C depending only on τ and $\alpha(m)$ such that $E|\sum_{t=1}^T y_t|^\tau \leq CD(\tau, \varepsilon, T)$ with $D(\tau, \varepsilon, T)$ defined in Doukhan (1994) and satisfying $D(\tau, \varepsilon, T) = O(T)$ if $\tau \leq 2$ and $D(\tau, \varepsilon, T) = O(T^{\tau/2})$ if $\tau > 2$.

Proof of Theorem 3. The conclusions follow under the conditions of the theorem by making use of the Cauchy–Schwarz inequality, Minkowski’s inequality, and Lemma 3 to verify the conditions of Lemmas 1 and 2. □

Proof of Theorem 4. Under the hypotheses of the theorem, $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} Q^{-1}N(0, W)$, $x'_i x_i / T - Q_i \xrightarrow{p} 0$, and $x'_i \varepsilon_i / \sqrt{T} \xrightarrow{d} N(0, W_i)$ are immediate from a LLN and CLT for mixing sequences, cf. White (2001, Theorems 3.47 and 5.20). The conclusion then follows from the definition of \hat{W} and $\hat{\varepsilon}_i$. □

Proof of Corollary 4.1. Consider $t^* = \sqrt{nT}(R\hat{\beta} - r) / \sqrt{R\hat{Q}^{-1}\hat{W}\hat{Q}^{-1}R}$. Under the null hypothesis, $R\beta = r$, so the numerator of t^* is $\sqrt{nT}R(\hat{\beta} - \beta) = R((1/nT)\sum_i x'_i x_i)^{-1} ((1/\sqrt{nT})\sum_i x'_i \varepsilon_i) \xrightarrow{d} RQ^{-1}A\sum_i B_i / \sqrt{n}$. From Theorem 4 and the hypotheses of the Corollary, the denominator of t^* converges in distribution to

$$\sqrt{RQ^{-1} \frac{1}{n} A \left(\sum_{i=1}^n B_i B'_i - \frac{1}{n} \sum_{i=1}^n B_i \sum_{i=1}^n B'_i \right) A Q^{-1} R}$$

It follows from the Continuous Mapping Theorem that

$$t^* \xrightarrow{d} \frac{RQ^{-1}A\sum_i B_i / \sqrt{n}}{\sqrt{(1/n)RQ^{-1}A(\sum_{i=1}^n B_i B'_i - (1/n)\sum_{i=1}^n B_i \sum_{i=1}^n B'_i)AQ^{-1}R}}$$

Define $\delta = (RQ^{-1}AAQ^{-1}R)^{1/2}$, so

$$\begin{aligned} t^* \xrightarrow{d} U &= \frac{\delta \sum_i B_{1,i} / \sqrt{n}}{\sqrt{(\delta^2/n)(\sum_{i=1}^n B_{1,i} B'_{1,i} - (1/n)\sum_{i=1}^n B_{1,i} \sum_{i=1}^n B'_{1,i})}} \\ &= \frac{\tilde{B}_{1,n}}{\sqrt{(1/n)(\sum_i B^2_{1,i} - \tilde{B}^2_{1,n})}} \end{aligned}$$

It is straightforward to show that $\tilde{B}_{1,n} \sim N(0, 1)$, that $\sum_i B_{1,i}^2 - \tilde{B}_{1,n}^2 \sim \chi_{n-1}^2$, and that $\sum_i B_{1,i}^2 - \tilde{B}_{1,n}^2$ and $\tilde{B}_{1,n}$ are independent, from which it follows that

$$U = \left(\frac{n}{n-1}\right)^{1/2} \frac{\tilde{B}_{1,n}}{\sqrt{(\sum_i B_{1,i}^2 - \tilde{B}_{1,n}^2)/(n-1)}} \sim \left(\frac{n}{n-1}\right)^{1/2} t_{n-1}.$$

The result for F^* is obtained through a similar argument, and using a result from Rao (2002) Chapter 8b to verify that the resulting quantity follows an F distribution. \square

References

- Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59 (3), 817–858.
- Arellano, M., 1987. Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics and Statistics* 49 (4), 431–434.
- Baltagi, B.H., Wu, P.X., 1999. Unequally spaced panel data regressions with AR(1) disturbances. *Econometric Theory* 15, 814–823.
- Bell, R.M., McCaffrey, D.F., 2002. Bias reduction in standard errors for linear regression with multi-stage samples. Mimeo RAND.
- Bertrand, M., Duflo, E., Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics* 119 (1), 249–275.
- Bhargava, A., Franzini, L., Narendranathan, W., 1982. Serial correlation and the fixed effects model. *Review of Economic Studies* 49, 533–549.
- Chung, K.L., 2001. *A Course in Probability Theory*, third ed. Academic Press, San Diego.
- Donald, S., Lang, K., 2001. Inference with differences in differences and other panel data. Mimeo.
- Doukhan, P., 1994. Mixing: properties and examples. In: Fienberg, S., Gani, J., Krickeberg, K., Olkin, I., Wermuth, N. (Eds.), *Lecture Notes in Statistics*, vol. 85. Springer, New York.
- Drukker, D.M., 2003. Testing for serial correlation in linear panel-data models. *Stata Journal* 3, 168–177.
- Hahn, J., Kuersteiner, G.M., 2002. Asymptotically unbiased inference for a dynamic panel model with fixed effects when both N and T are large. *Econometrica* 70 (4), 1639–1657.
- Hahn, J., Newey, W.K., 2004. Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica* 72 (4), 1295–1319.
- Hansen, C.B., 2004. Inference in linear panel data models with serial correlation and an essay on the impact of 401(k) participation on the wealth distribution. Ph.D. Dissertation, Massachusetts Institute of Technology.
- Hansen, C.B., 2006. Generalized least squares inference in multilevel models with serial correlation and fixed effects. *Journal of Econometrics*, doi:10.1016/j.jeconom.2006.07.011.
- Kezdi, G., 2002. Robust standard error estimation in fixed-effects panel models. Mimeo.
- Kiefer, N.M., Vogelsang, T.J., 2002. Heteroskedasticity–autocorrelation robust testing using bandwidth equal to sample size. *Econometric Theory* 18, 1350–1366.
- Kiefer, N.M., Vogelsang, T.J., 2005. A new asymptotic theory for heteroskedasticity–autocorrelation robust tests. *Econometric Theory* 21, 1130–1164.
- Lancaster, T., 2002. Orthogonal parameters and panel data. *Review of Economic Studies* 69, 647–666.
- Liang, K.-Y., Zeger, S., 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73 (1), 13–22.
- MaCurdy, T.E., 1982. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics* 18 (1), 83–114.
- Nickell, S., 1981. Biases in dynamic models with fixed effects. *Econometrica* 49 (6), 1417–1426.
- Phillips, P.C.B., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. *Econometrica* 67 (5), 1057–1111.
- Phillips, P.C.B., Sun, Y., Jin, S., 2003. Consistent HAC estimation and robust regression testing using sharp origin kernels with no truncation. Cowles Foundation Discussion Paper 1407.
- Rao, C.R., 2002. *Linear Statistical Inference and Its Application*. Wiley-Interscience.

- Solon, G., 1984. Estimating autocorrelations in fixed effects models. NBER Technical Working Paper no. 32.
- Solon, G., Inoue, A., 2004. A portmanteau test for serially correlated errors in fixed effects models. Mimeo.
- Stata Corporation, 2003. Stata User's Guide Release 8. Stata Press, College Station, Texas.
- Vogelsang, T.J., 2003. Testing in GMM models without truncation. In: Fomby, T.B., Hill, R.C. (Eds.), *Advances in Econometrics*, volume 17, *Maximum Likelihood Estimation of Misspecified Models: Twenty Years Later*. Elsevier, Amsterdam, pp. 192–233.
- White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48 (4), 817–838.
- White, H., 2001. *Asymptotic Theory for Econometricians*, revised edition. Academic Press, San Diego.
- Wooldridge, J.M., 2002. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, MA.