View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Robust Standard Error Estimation In Fixed-Effects Panel Models

Gábor Kézdi*
Budapest University of Economics, IE/HAS† and CEU‡

October 13, 2003

Abstract

This paper focuses on standard error estimation in FE models if there is serial correlation in the error process. Applied researchers have often ignored the problem, probably because major statistical packages do not estimate robust standard errors in FE models. Not surprisingly, this can lead to severe bias in the standard error estimates, both in hypothetical and real-life situations. The paper gives a systematic overview of the differnt standard error estimators and the assumptions under which they are consistent (in the usual large N, small T asymptotics). One of the possible reasons why the robust estimators are not used often is a fear of their bad finite sample properties. The most important results of the paper, based on an extensive Monte Carlo study, show that those fears are in general unwarranted. I also present evidence that it is the abolute size of the cross-sectional sample that primarily affects the finite-sample behavior, not the relaitve size compared to the time-series dimension. That indicates good small-sample behavior even when $N \approx T$. I introduce a simple direct test analogous to that of White (1980) for the restrictive assumptions behind the estimators. Its finite sample properties are fine except for low power in very small samples.

This paper focuses on Fixed-Effects panel models (FE) with exogenous regressors on pooled cross sectional and time series data with relatively few within-individual observations. Empirical studies that estimate this kind of FE models are abundant, and they routinely estimate standard errors under the assumption of no serial error correlation within individual units. In the past three years, the top three economics journals with a focus on applied empirical research published 42 papers that estimated linear FE models with time series

*My first thanks go to John Bound and Gary Solon for their suggestions and support. Jinyong Hahn, Steven Levitt, Shinichi Sakata, and Douglas Staiger provided many helpful comments. All remaining errors are mine.
Correspondance: kezdi@econ.core.hu.

†Institute of Economics, Hungarian Academy of Sciences

‡Central European University, Budapest

within individual units.[1] Out of the 42, only 6 took serial correlation into account when estimated the standard errors.[2]

Serial correlation in the error process affects standard errors in FE models with more than two observations per individual unit, unless all right-hand side variables are serially uncorrelated. The effect is larger the stronger the correlation and the longer the time horizon. Serial correlation consistent standard error estimators for panel models without fixed effects are covered by most econometrics textbooks. Same is not true, however for FE. Similar estimators were developed explicitly for FE models by Kiefer (1980), Bhargava et al. (1982), and Arellano (1987), but they have been overlooked by practitioners. It seems that worries about finite sample properties are responsible for this fact. Major statistical computer packages do not allow for any robust standard error estimation in FE models. Stata^{TM}, for example, calculates standard errors that are robust to serial correlation for all linear models but FE (and random effects). It does so for an analogous model but it explicitly cautions against using robust methods in samples with long time-series within individual units.[3] As we will see, however, even this warning is unwarranted.

In this paper I give a systematic overview of standard error estimation in FE models, together with the assumptions under which the estimators are consistent. I also introduce a very simple test for the assumptions in question (it is analogous to White's 1980 direct test for heteroskedasticity). The asymptotic results consider the case when T is fixed and $N \to \infty$, and they are straightforward applications of White's (1984) general results. The novelty in this paper is a thorough examination of the finite-sample properties of the estimators and tests. The Monte-Carlo study considers various combinations of the time-series and cross-sectional sample size, and the degree of serial correlation and cross-sectional heteroskedasticity.

The most important result is that the general robust standard error estimator, known in other models as the "cluster" estimator (introduced to FE by Arellano, 1987) is not only consistent in general but it behaves well in finite samples. The Monte Carlo experiments reveal that the cluster estimator is unbiased in samples of usual size although it is slightly biased downward if the cross-sectional sample is very small. The results suggest that it is the cross-sectional dimension itself that matters, not its relative size to the time-series dimension (N and not N/T). The variance of the estimator naturally increases as the sample gets small but stays moderate at usual sample sizes. Kiefer's (1980) estimator is consistent under the assumption of conditional homoskedasticity across individuals. Quite naturally, when consistent, it is superior to the robust estimator in terms of both variance and small-sample bias. The bias of the estimators that assume no serial correlation is substantial when the

[1] The examined journal issues were the following: American Economic Review 88(4) to 91(3); Journal of Political Economy 106(4) to 109(3); and Quarterly Journal of Economics 103(3) to 106(2). Only papers that estimated linear FE models on panel data with time-series ($T > 2$) within the individual units were considered.

[2] Two did that by a parametric specification of the error process, one by using the cluster estimator (see later). The other three did not specify the standard error estimator they used.

[3] "Why is it dangerous to use the robust cluster() option on areg (areg estimates the same fixed-effects model as xtreg, fe)?" http://www.stata.com/support/faqs/stat/aregclust.html. I thank John Bound for this note.

assumption is not met, and it is larger than the finite-sample bias of the robust estimators at any sample size. The bias is a function of serial correlation both in the right-hand-side variables and the error term. The test that looks at the restrictive assumtpions delivers the desirable size and power properties in relatively large samples. Its power, however, is quite low in small samples unless the serial correlation is very strong.

Bertrand, Duflo, and Mullainathan (2001) have drawn attention to robust standard error estimation in the context of a special FE model, the "Difference-in-Differences" (DD) model. Typically, DD models estimate effects of binary treatments on different individual units by comparing before and after treatment outcomes. Serial correlation in the error process has especially large effect on standard errors in these models because the main right-hand-side variable is highly correlated through time (the binary treatment variable changes only once in most cases). The problem is irrelevant if only two points in time are compared but it can lead to a severe bias to conventional standard error estimates in longer series. Bertrand et al. report simulation results on frequently used data (yearly earnings for US states) that show 45 to 65 percent rejection rates of a t-test on "placebo" binary treatments instead of the nominal size of 5%. This size distortion is probably due to downward biased standard errors. Bertrand et al. suggest an intuitively appealing simulation-based method to overcome the problem. Apart from being a little complicated for applied research, their method is specific to binary treatment effects. The alternative solutions I present here are more conventional, easier to implement, and general to all FE models. They also behave well in finite samples.

The asymptotic results are stated in the main text. To keep things simple, I consider a data generating process that is *i.i.d.* in the individual units. This simplification is justified because our main concern is about the process within the individual units. The usual T fixed, $N \to \infty$ asymptotics is considered for the results. The proofs are straightforward applications of standard *i.i.d.* results (see White, 1984, for example). They are not presented in the paper for this reason. Exceptions are the simplified versions of the asymptotic covariance matrix of the FE estimator under the appropriate assumptions. They are derived in the main text because of their importance.

The remainder of the paper is organized the following way. The first section introduces the assumptions underlying the data generating process, the model, and the fixed-effects estimator. The second part presents the sampling covariance matrix of the FE estimator and its simplified versions under restrictive assumptions, and it introduces the estimators. The third part examines the finite sample properties of the four proposed estimators. The fourth part introduces a direct test for the restrictions and examines its finite sample properties, and the last part concludes.

1 Setup

1.1 Data generating process

Assume that a T dimensional random vector y_i and a $T \times K$ dimensional random matrix x_i are generated by an independent and identically distributed process. More formally, we assume that the $T \times (K+1)$ dimensional random process $\{y_i, x_i\}_{i \in \mathbb{N}}$ on $\{\Omega, \mathcal{F}, P\}$is *i.i.d.*, with finite fourth moments. Note that there is no restriction in the time series dimension. In particular, nonconstant variance, unit roots, an unequal spells are allowed. We can do so because of the T fixed assumption. All asymptotic results well be driven by the cross-sectional properties of the process.

The intuition behind the data generating process (DGP) assumption is that each i is an individual observation that is drawn from a population in a random fashion. The assumption implies that there is one $E[y_i]$ and one $E[x_i]$, which are the population means. The goal of the exercise is to reveal the relationship between y and x in the population.

1.2 Model

For estimating this relationship, consider a linear panel model with exogenous regressors and individual-specific constants ("fixed effects"). The panel has a cross-sectional dimension i and a time-series dimension t.

$$y_{it} = x_{it}'\beta + \varepsilon_{it} = \alpha_i + x_{it}'\beta + u_{it}, \tag{1}$$

or, in vector notation,

$$y_i = \alpha_i \ell + x_i \beta + u_i \tag{2}$$

where $y_i = [y_{i1,}..., y_{iT}]'$ is $T \times 1$, $x_i = \left[x_{i1,}'..., x_{iT}'\right]'$ is $T \times K$, $\varepsilon_{it} = \alpha_i + u_{it}$, α_i is a scalar, $\ell = [1, .., 1]'$ is $T \times 1$, and $u_i = [u_{i1,}..., u_{iT}]'$ is $T \times 1$. $i = 1....N$, and $t = 1...T$. For future reference, let x_{ik} be the $T \times 1$ vector of the k-th right-hand side variable so that $x_i = [x_{i1}, ..., x_{iK}]$

The intuition behind the model is the following. We would like to uncover something about the conditional mean of y given x, which may be different across individuals. (2) models the conditional mean of y given x in a linear fashion. There is an i-specific intercept denoted by α_i. It is interpreted as the conditional mean of y_i given $x_i = 0$. The model is restrictive in that apart from the intercept this conditional mean is the same across both the i and the t dimension. One interpretation of β is that it is a population average of the relationship after countering for the i-specific intercept. The model does not put any restriction on the covariance of x_i and α_i, the latter treated as a random variable itself. Formally, we assume that all relevant moments exist and that $E[x_{ik}u_i'] = 0$ for $k = 1...K$. On the other hand, we allow for $E[\alpha_i x_i] \neq 0$.

We want a consistent estimator for β and its asymptotic covariance matrix. We can take the limit in both the cross-sectional and the time-series dimension, so it is important to be explicit what we mean by consistency and an asymptotic distribution. In this paper, the $N \to \infty$, T fixed asymptotics will be considered. In that case, it is the limiting distribution of $\sqrt{N}\left(\hat{\beta} - \beta\right)$ that we are interested in.

The $N \to \infty$, T fixed asymptotics is a natural setup for household or individual panels like the PSID (the Panel Study of Income Dynamics of the University of Michigan). It is also a natural approximation for country or regional panels if the time series is relatively short

$(N > T)$. The simulation results suggest, however, that the proposed estimators behave well also in the finite $N < T$ setup.

1.3 The fixed-effects estimator

OLS with N constants for capturing each of the α_i is a natural candidate for estimation. This estimator is often called the "least-squares dummy-variables" estimator or LSDV in order to distinguish it from OLS with only one constant. For computational reasons, however, it is common to use the fixed-effects (FE, also known as Within-) estimator instead. FE is OLS on mean-differenced variables, which are defined as $\underset{T\times 1}{\tilde{y}_i} \equiv [y_{i1} - \bar{y}_i, ..., y_{iT} - \bar{y}_i]'$, $\underset{T\times K}{\tilde{x}_i} \equiv [x'_{i1} - \bar{x}'_i, ..., x'_{iT} - \bar{x}'_i]'$, and $\underset{T\times 1}{\tilde{u}_i} \equiv [u_{i1} - \bar{u}_i, ..., u_{iT} - \bar{u}_i]'$, where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$ etc. To simplify notation, let $\underset{T\times T}{M} \equiv I_T - \frac{1}{T}\ell_T\ell_T'$. Note that M is idempotent. Then, $\tilde{y}_i = My_i, \tilde{x}_i = Mx_i$,and $\tilde{u}_i = Mu_i$. The mean-differenced equation to estimate is $\tilde{y}_i = \tilde{x}_i\beta + \tilde{u}_i$, and the fixed-effect estimator for β is defined as

$$\hat{\beta}_{FE} \equiv \left(\sum_{i=1}^{N} \tilde{x}_i'\tilde{x}_i\right)^{-1} \left(\sum_{i=1}^{N} \tilde{x}_i'\tilde{y}_i\right) = \tilde{S}_{xx}^{-1}\tilde{S}_{xy}. \tag{3}$$

$\tilde{S}_{xx} \equiv \frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\tilde{x}_i$, and $\tilde{S}_{xy} \equiv \frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\tilde{y}_i$.A standard result is that FE and the LSDV estimator for β on levels are computationally equivalent.

Recall that we assume that the data generating process is *i.i.d.* in the cross-sectional dimension, and therefore the $(\tilde{y}_i, \tilde{x}_i)$ are *i.i.d.*, too. $\hat{\beta}_{FE}$ is consistent for β in the $N \to \infty$, T fixed asymptotics without further assumptions about the time-series dimension. The conditional covariance matrix of $\tilde{u}_i$ affects the asymptotic covariance of $\hat{\beta}_{FE}$. Serial correlation and heteroskedasticity of any kind would also make $\hat{\beta}_{FE}$ inefficient. The rest of the paper focuses on consistent estimation of the sampling covariance of $\hat{\beta}_{FE}$. Efficient estimation of β is not addressed here.[4]

2 Asymptotic Distribution of the Fixed-Effects Estimator

The covariance matrix of $\widehat{\beta}_{FE}$ is easy to derive because of cross-sectional independence and the linearity of the model.

[4] Some of the introduced covariance matrix estimators could be used in efficient estimation (feasible GLS) of the parameters. Although that seems like a natural extension of my analysis, it would introduce other problems that should be dealt with. It could aggravate bias from measurement error or misspecification of the timing of binary variables or lagged effects.

$$
\begin{aligned}
\hat{\beta}_{FE} &= \left(\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i\right)^{-1}\left(\sum_{i=1}^{N}\tilde{x}_i'\tilde{y}_i\right) = \left(\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i\right)^{-1}\left(\sum_{i=1}^{N}\tilde{x}_i'\left(\tilde{x}_i\beta+\tilde{u}_i\right)\right) \\
&= \beta + \tilde{S}_{xx}^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{u}_i\right).
\end{aligned}
$$

Proposition 1 *Suppose that* $\{y_i, x_i\}_{i\in\mathbb{N}}$ *is i.i.d. with finite second moments. Consider the fixed-effect (FE) panel model (1-**??**) and assume that* $E[\tilde{x}_i'\tilde{x}_i]$ *and* $\tilde{S}_{xx} \equiv \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i$ *are positive definite. The FE estimator defined by (3) is consistent and asymptotically normal with covariance matrix* D *defined below (5-6):*

$$
\begin{aligned}
\hat{\beta}_{FE} &= \tilde{S}_{xx}^{-1}\tilde{S}_{xy} \to \beta \quad prob-P \ \ as\ N\to\infty,\ and \\
& \quad D^{-1/2}\sqrt{N}\left(\hat{\beta}_{FE}-\beta\right)\overset{A}{\sim}N(0,I),\ \ where \qquad (4)\\
D &\equiv E[\tilde{x}_i'\tilde{x}_i]^{-1}VE[\tilde{x}_i'\tilde{x}_i]^{-1}\ \ and \qquad (5)\\
V &\equiv E[\tilde{x}_i'\tilde{u}_i\tilde{u}_i'\tilde{x}_i]. \qquad (6)
\end{aligned}
$$

The standard errors of the elements in $\hat{\beta}_{FE}$ *are therefore the square root of the diagonal elements of* D *divided by* N*, or with some abuse of notation,*

$$
\hat{\beta}_{FE}\overset{A}{\sim}N\left(\beta,\frac{1}{N}D\right)
$$

The proof is a straightforward application of Theorems 3.5 and 5.3 in White (1984). Note that the time-series properties of $\{\tilde{u}_{it}\}$ or $\{\tilde{x}_{it}\}$ are not restricted in any way. Among other things, serial correlation and time-series heteroskedasticity of any kind are allowed, and so are unit roots and unequal spacing. All asymptotic results follow from the fixed length of the time series and the cross-sectional *i.i.d.* assumption.

The next few subsections will consider simplified versions of $V = E[\tilde{x}_i'\tilde{u}_i\tilde{u}_i'\tilde{x}_i]$ under restrictive assumptions.

2.1 Cross-sectional homoskedasticity

Under conditional homoskedasticity in the cross-sectional dimension but no restriction in the time series dimension, we have that $E[u_iu_i'|x_i] = E[u_iu_i'] \equiv \Omega$. Since M is nonstochastic, $E[\tilde{u}_i\tilde{u}_i'|\tilde{x}_i] = ME[u_iu_i'|x_i]M$, and so

$$
\tilde{\Omega} \equiv E[\tilde{u}_i\tilde{u}_i'] = E[\tilde{u}_i\tilde{u}_i'|\tilde{x}_i] = M\Omega M.
$$

This implies that

$$
V = E[\tilde{x}_i'\tilde{u}_i\tilde{u}_i'\tilde{x}_i] = E[\tilde{x}_i'E[\tilde{u}_i\tilde{u}_i'|\tilde{x}_i]\tilde{x}_i] = E\left[\tilde{x}_i'\tilde{\Omega}\tilde{x}_i\right].
$$

Here, again, no time-series restrictions are used.[5] Notice that $E[\tilde{x}_i'\tilde{u}_i\tilde{u}_i'\tilde{x}_i] = E[x_i'MMu_iu_i'MMx_i] = E[x_i'Mu_iu_i'Mx_i]$. Using this fact, we can simplify V further to get $V = E[x_i'M\Omega Mx_i] = E[\tilde{x}_i'\Omega\tilde{x}_i']$. This result is not used for the present estimator because we naturally want everything to be a function of the mean-differenced variables. The result is important in itself nevertheless. It means that using the levels error covariance matrix or mean-differenced error covariance matrix are equivalent.

2.2 No serial correlation

In the absence of serial correlation in the error process $\{u_{it}\}_t$, we have that $E[u_{it}u_{is}] = 0$ $\forall s \neq t$, and therefore $\Omega_i \equiv E[u_iu_i'|x_i] = \langle\omega_{it}\rangle_{T\times T}$, a diagonal matrix, with elements $\omega_{it} = E[u_{it}^2|x_i]$. Therefore,

$$\begin{aligned} V &= E[x_i'Mu_iu_i'Mx_i] = E[\tilde{x}_i'\Omega_i\tilde{x}_i] \\ &= E\left[\sum_{t=1}^{T}\omega_{it}\tilde{x}_{it}\tilde{x}_{it}'\right] = E\left[\sum_{t=1}^{T}u_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right], \end{aligned}$$

We would like to express this in terms of the conditional variance of the mean-differenced errors, because we estimate the model on mean-differenced data. One can show that $E[\tilde{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'] = \frac{T-1}{T}E[u_{it}^2\tilde{x}_{it}\tilde{x}_{it}']$, and therefore

$$V = E\left[\sum_{t=1}^{T}u_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right] = \frac{T}{T-1}E\left[\sum_{t=1}^{T}\tilde{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right].$$

The same result is implied by zero serial correlation in the right-hand-side variables, that is if $E[x_{it}x_{is}'] = 0$ $\forall t \neq s$. Let $\tilde{\Omega}_i \equiv E[\tilde{u}_i\tilde{u}_i'|x_i]$ and write

$$\begin{aligned} V &= E[x_i'Mu_iu_i'Mx_i] = E[x_i'E[\tilde{u}_i'\tilde{u}_i|x_i]\,x_i] = E\left[x_i'\tilde{\Omega}_ix_i\right] \\ &= E\left[\sum_{t=1}^{T}\tilde{\omega}_{it}x_{it}x_{it}'\right] = E\left[\sum_{t=1}^{T}\tilde{u}_{it}^2x_{it}x_{it}'\right] = \frac{T}{T-1}E\left[\sum_{t=1}^{T}\tilde{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right], \end{aligned}$$

where we used the fact that $E[x_{it}'x_{is}] = 0$ $\forall s \neq t$ and $E\left[\sum_{t=1}^{T}x_{it}x_{it}'\right] = E[x_i'x_i]$, both implied by $E[x_{it}x_{is}'] = 0$. The last equality makes use the fact that $E[\tilde{x}_i'\tilde{x}_i] = \frac{T-1}{T}E[x_i'x_i]$.

The assumption we use is zero serial correlation in the error process or in (and across) the right-hand-side variables. The error process may be heteroskedastic in any dimension. This sampling covariance matrix is in fact a $\frac{T}{T-1}$-scaled version of the one that is behind the original White heteroskedasticity-consistent estimator, applied to the mean-differenced data.

[5] V is basically a seemengly unrelated regressions (SUR) covariance matrix, with T equations and the β constrained to be the same. Kiefer (1980) has introduced this estimator in the FE context.

Note that it is the error terms or the right-hand-side variables in levels (as opposed to mean-differences) that are assumed to be serially uncorrelated. In the fixed T setup we focus on, mean-differencing induces serial correlation in the first-differenced errors, because all $\tilde{u}_{it}$ are correlated with $\bar{u}_{it}$. Assuming no serial correlation in the mean-differenced error terms would deliver a similar result without the $\frac{T}{T-1}$ factor. We think that that assumption has no intuitive appeal. The model is set up in levels, while mean-differencing is only a way to get around the correlation of α_i and x_i.

We can already see that the unscaled White estimator is going to be inconsistent in the fixed-T framework. This is an example if the incidental parameter problem (Lancaster, 2000), since the White estimator is an ML estimator in the *i.i.d.* setup (Huber, 1967). The adjustment is analogous to "degrees of freedom" corrections for the α_i parameters when the model is estimated in levels.

2.3 Homoskedasticity and no serially correlation

If there is no serial correlation and the conditional variance of u_{it} is the same at every t, that is $E\left[u_{it}^2|x_{it}\right] = \Omega_i = \Omega = \sigma^2 I_T$, we get back the appropriately scaled *i.i.d.* OLS estimator for V.

$$\begin{aligned} V &= E\left[\tilde{x}_i'\Omega\tilde{x}_i\right] = \sigma^2 E\left[\tilde{x}_i'\tilde{x}_i\right], \text{ where} \\ \sigma^2 &= E\left[u_{it}^2\right]. \end{aligned}$$

D simplifies in this case to $D = \sigma^2 E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1}$. We would like to have an expression in terms of the mean-differenced error term. Analogously to the relationship of the conditional level and mean-differenced variances, we have that $\sigma^2 = E\left[u_{it}^2\right] = \frac{T}{T-1} E\left[\tilde{u}_{it}^2\right]$.

Homoskedastic errors and serially independent right-hand side variables imply the same covariance of $\hat{\beta}_{FE}$ Assume that $E\left[u_i u_i'|x_i\right] = \Omega$ with $\omega_{tt} = \sigma^2$, and $E\left[x_{it}x_{is}'\right] = 0\ \forall s \neq t$. Recall that no serial correlation across and within right-hand side variables implies that $E\left[\widetilde{x}_i'\widetilde{x}_i\right] = \frac{T-1}{T} E\left[x_i'x_i\right]$. Therefore,

$$V = E\left[x_i'M\Omega M x_i\right] = E\left[x_i'\tilde{\Omega}x_i\right] = E\left[\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{\omega}_{st}x_{it}'x_{is}\right] = \sum_{t=1}^{T}\tilde{\omega}_{tt}E\left[x_{it}'x_{it}\right],$$

where the last equality holds because $E\left[\sum_{t=1}^{T}\sum_{s=1}^{T}x_{it}'x_{is}\right] = 0$ if $s \neq t$. Using $\tilde{\omega}_{tt} = E\left[\tilde{u}_{it}\right] = \frac{T-1}{T}\sigma^2$, we get the same result as before.

The asymptotic variance of the fixed-effect estimator is the $\frac{T-1}{T}$-scaled asymptotic variance of the OLS estimator on the mean-differenced data. Just like before, the zero serial correlation is assumed about u_{it} or x_{it} and not their mean-differenced counterparts. And again, conventional OLS standard errors based on the FE residuals are going to be inconsistent because of the incidental parameter problem, with the same bias as in the White estimator.

2.4 Estimation

We have considered four cases for V. (0) the general case, (1) cross-sectional conditional homoskedasticity but no restriction in the time dimension, (2) no serial correlation, and (3) cross-sectional and time-series conditional homoskedasticity and no serial correlation. The four asymptotic covariance matrices are, respectively,

$$D_0 \equiv E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} E\left[\tilde{x}_i\tilde{u}_i\tilde{u}_i'\tilde{x}_i\right] E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} \quad (7)$$

$$D_1 \equiv E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} E\left[\tilde{x}_i'\Omega\tilde{x}_i\right] E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} \quad (8)$$

$$D_2 \equiv \frac{T}{T-1} E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} E\left[\sum_{t=1}^{T} \tilde{u}_{it}^2 \tilde{x}_{it}\tilde{x}_{it}'\right] E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} \quad (9)$$

$$D_3 \equiv \sigma^2 E\left[\tilde{x}_i'\tilde{x}_i\right]^{-1} \quad (10)$$

Let $\check{u}$ denote the FE residuals. By the analogy principle, the proposed estimators for D_0 through D_3 are, respectively,

$$\hat{D}_0 \equiv \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\tilde{x}_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\check{u}_i\check{u}_i'\tilde{x}_i\right) \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\tilde{x}_i\right)^{-1}, \quad \text{where} \quad (11)$$

$$\check{u}_i \equiv \tilde{y}_i - \tilde{x}_i\hat{\beta}_{FE} \quad (12)$$

$$\hat{D}_1 \equiv \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\tilde{x}_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\check{\Omega}\tilde{x}_i\right) \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\tilde{x}_i\right)^{-1}, \quad \text{where} \quad (13)$$

$$\check{\Omega} \equiv \frac{1}{N}\sum_{i=1}^{N} \check{u}_i\check{u}_i', \quad (14)$$

$$\hat{D}_2 \equiv \frac{T}{T-1}\left(\frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\tilde{x}_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1it}^{T} \check{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right) \left(\frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\tilde{x}_i\right)^{-1}, \quad (15)$$

$$\hat{D}_3 \equiv \hat{\sigma}^2\left(\frac{1}{N}\sum_{i=1}^{N} \tilde{x}_i'\tilde{x}_i\right)^{-1}, \quad \text{where} \quad (16)$$

$$\hat{\sigma}^2 \equiv \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T} \check{u}_{it}^2. \quad (17)$$

Under our cross-sectional *i.i.d.* assumption it is straightforward to show that the $\hat{D}_j$ are consistent for the corresponding D_j $(j = 0, 1, 2, 3)$ if T is fixed and $N \to \infty$. The proofs are straightforward application of Theorem 5.3 (v) in White (1984). One should note that the estimators don't correct for degrees of freedom decreased by the dimension of $\tilde{x}_i$. That is only for keeping things as simple as possible. Not surprisingly, the simulation results presented in the next section suggest that such corrections would slightly improve upon the finite-sample bias of the consistent estimators.

$\hat{D}_0$ is known as the "clustered" covariance estimator, and was introduced by Arellano (1987). It is always consistent in our setup. $\hat{D}_1$, introduced by Kiefer (1980), makes use of the covariance matrix of the FE residuals, $\check{\Omega}$. It is consistent under any time-series behavior as long as the error term is homoskedastic in the cross-sectional dimension. $\hat{D}_2$ is the original heteroskedasticity-consistent estimator of White (1980) scaled by $\frac{T}{T-1}$. It is consistent if the error term or the right-hand-side variables are serially uncorrelated. $\hat{D}_3$ is the scaled version of the homoskedasticity-consistent OLS estimator. It is the conventional sampling covariance estimator of $\hat{\beta}_{FE}$, calculated as the default by all software packages. It is consistent only under cross-sectional and time-series homoskedasticity and if either the error term or the right-hand-side variables are serially uncorrelated and have the same variance.

3 Finite-sample properties

In this section Monte Carlo simulation results are presented. To keep things simple, the analysis was restricted to a one-dimensional x variable. The data generating process involved the possibility of serial correlation in both x_{it} and u_{it}. In particular, stationary AR(1) processes were considered with autoregressive parameters 0, 0.1, 0.3, 0.5, 0.7, and 0.9 for each process (all 36 combinations were analyzed). Two separate sets were examined. In one, u was homoskedastic in the cross-sectional dimension, in the other it was heteroskedastic conditional on x. The two data generating processes were the following.

DGP(1)

$$\begin{aligned} x_{it} &= \rho_x x_{i(t-1)} + v_{xit}, \quad x_{it}\tilde{\ } N(0,1), \\ u_{it} &= \rho u_{i(t-1)} + v_{uit}, \quad u_{it}\tilde{\ } N(0,1), \end{aligned}$$

DGP(2)

$$\begin{aligned} &\text{same as DGP(1), plus} \\ v_{uit} &= \sqrt{h_{it}} w_{it}, \quad w_{it}\tilde{\ } iidN(0,1) \\ h_{it} &= a_0 + a_1 x_{it}^2, \quad a_0 = a_1 = 0.5. \end{aligned}$$

10,000 Monte Carlo simulations were conducted for each of the 2×36 parameter settings. I have estimated the sampling distribution of the $\hat{\beta}_{FE}$ and compared its standard deviation to the mean of the 10,000 estimated standard error estimates ($SE_j = \sqrt{\frac{1}{M}\sum_{m=1}^{M} SE_{mj}}$, $j = 0,1,2,3$). These means were then used to calculate the relative bias $\left(\frac{SE_j - std(\hat{\beta}_{SE})}{std(\hat{\beta}_{SE})}\right)$. In addition to the relative bias, I also present the standard deviation of the SE_j. Several combinations of (N,T) were considered. The $(500,10)$ case establishes large-sample properties while the $(50,10)$ case looks at what happens in relatively small-N samples. The $(50,50)$, case illustrates what happens when $N = T$ in relatively small samples, and the $(10,50)$ case is an illustration of what happens in a small-sample $N < T$ setup. Finally, a $(10,10)$ example illustrates extreme small sample behavior. The results are contained in Tables 1 and 2.

In order to assess the results, note that in the first set (Table 1), SE_0 and SE_1 are always consistent for the true SE, and SE_2 and SE_3 are consistent if $\rho_u \rho_x = 0$ (either of the two is zero). In the second set (Table 2), SE_0 is always consistent for the true SE, SE_2 is consistent if $\rho_u \rho_x = 0$, but SE_1 and SE_3 are never consistent because of cross-sectional heteroskedasticity.

Tables 1.1 and 2.1 present the large-sample results. Bias of the consistent estimators is virtually zero. The bias of the inconsistent estimators increases as ρ_u and ρ_x increase. Unbiasedness of SE_2 and SE_3 when they are consistent indicates that the unscaled White and OLS estimators are biased in small samples as well. In the heteroskedastic setup, the bias of SE_1 and SE_3 is dominated by heteroskedasticity in the small-ρ setups, and serial correlation takes over as ρ_u and ρ_x increase. The variance of the estimators behave the predictable way, with the more restrictive ones having smaller variation. These differences, however, are very small for practical purposes.

Smaller-N samples (Table 1.2 and 2.2) basically deliver the large-sample results in terms of the bias. SE_0 shows a small-sample bias that is larger than other consistent estimators, but it is still negligible. Differences in the variance are magnified, as expected, but they are not extremely large, either. Both the small-sample bias and the variance of the consistent estimators increases as ρ_u and ρ_x increase. This reflects the fact that higher serial correlation decreases the variation in the variables if the overall error and RHS variance is fixed, as were throughout the simulation.

Bias due to serial correlation is greater as T increases. Small-sample bias of SE_0 stays small in the $(50, 50)$ setup but becomes a significant negative 8-16 percent when in the $N = 10$ setups. The results indicate that it is the overall sample size, and especially N, the size of the cross-sectional sample that determines the small-sample bias. Reluctance of using the cluster estimator (SE_0) when T is large is unjustified. The variance disadvantage of SE_0 is larger in the $(50, 50)$ case than the $(50, 10)$ case, as expected, but the differences remain modest. The small-sample properties of SE_1, the more restrictive serial correlation consistent estimator, are significantly better when it is consistent. Its small sample bias stays close to zero even in the $(10, 10)$ sample, and its standard deviation is below 25 percent larger than that of SE_3 when $N = 10, T = 10$ and 50 per cent when $N = 10, T = 50$.

The good small-sample behavior may seem somewhat surprising. But they may simply reflect that the standard error estimators take an average over all NT observations. In this light, even the (10.10) sample is not small: it consists of 100 observations altogether.

The results have the following practical implications. In large samples, SE_0 is just as good for applied work as the restricted estimators even when the latter are also consistent. In smaller samples, there is some advantage SE_1 (the Omega-estimator) if that is consistent for the true SE. The conventional estimator (SE_3) has no substantial advantage over SE_1 other than computational simplicity. The simulation results suggest that properties of the estimators don't depend much on the relative size of T and N but rather on the total sample size NT and especially N itself. At the same time, an increasing T increases the bias due to serial correlation. Cautioning against using the "clustered" estimator (SE_0) when the time-series is long is therefore not simply unnecessary but quite misleading.

4 A Direct Test for Homoskedasticity and No Serial Correlation

In this section I propose a direct test for the restrictive assumptions under which the alternative (less robust) estimators are consistent. $\hat{D}_3$ is easier to compute and has the best properties if consistent. $\hat{D}_1$ performs significantly better in terms of variance than $\hat{D}_0$ when both are consistent, especially in smaller samples. Moreover, the properties of $\hat{D}_1$ match closely those of $\hat{D}_3$ if both are consistent. If we can test for the restrictions that make $\hat{D}_1$, $\hat{D}_2$, or $\hat{D}_3$ consistent we can always choose the best consistent estimator. In this section I develop such a test.

Let me introduce the following notation. Recall that the alternative standard error estimators differ only in how they estimate $V = E\left[\tilde{x}_i'\tilde{u}_i\tilde{u}_i'\tilde{x}_i\right]$. The assumptions behind the restricted estimators can therefore be tested by comparing the corresponding V estimates to one that is always consistent. Define

$$\hat{V}_0 \equiv \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\breve{u}_i\breve{u}_i'\tilde{x}_i \tag{18}$$

$$\hat{V}_1 \equiv \frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i'\breve{\Omega}\tilde{x}_i, \quad \breve{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\breve{u}_i\breve{u}_i' \tag{19}$$

$$\hat{V}_2 \equiv \frac{1}{N}\frac{T}{T-1}\sum_{i=1}^{N}\sum_{t=1it}^{T}\breve{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}' \tag{20}$$

$$\hat{V}_3 \equiv \frac{\hat{\sigma}^2}{N}\sum_{i=1}^{N}\tilde{x}_i'\tilde{x}_i, \quad \hat{\sigma}^2 = \frac{1}{N\left(T-1\right)}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{u}_{it}^2 \tag{21}$$

$\hat{V}_0$ is always consistent for V. $\hat{V}_1$ is consistent under cross-sectional homoskedasticity. $\hat{V}_2$ is consistent under no serial correlation in the (levels) error or the (levels) right-hand-side variables. $\hat{V}_3$ is consistent if both $\hat{V}_1$ and $\hat{V}_2$ is consistent and time-series homoskedasticity also holds. A direct way to test whether the more restrictive assumptions hold is to check whether $V_1 = V$, $V_2 = V$, or $V_3 = V$. In order to formulate the linear hypotheses, let's use the *vech* operator that stacks columnwise the diagonal and sub-diagonal elements of a symmetric matrix.[6]

[6]Suppose that $K = 3$, and A is symmetric:

$$A = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{21} & a_{22} & a_{32} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Then, $vech\left(A\right) \triangleq \left(a_{11}, a_{21}, a_{31}, a_{22}, a_{32}, a_{33}\right)'$. See Magnus and Neudecker (1988), e.g., for more discussion.

$$
\begin{aligned}
v_j &\equiv vech\left(V_j\right), && (22)\\
\hat{v}_j &\equiv vech\left(\hat{V}_j\right) \quad j=0,1,2,3 && (23)
\end{aligned}
$$

The hypotheses are

$$
\begin{aligned}
H_0 &: v_j - v_0 = 0, \quad j=1,2,3\\
H_1 &: v_j - v_0 \neq 0 \quad\; j=1,2,3.
\end{aligned}
$$

The test I propose is analogous to White's (1980) test for heteroskedasticity. Since $\hat{v}_0$ is always consistent and the $\hat{v}_j$ are consistent only under the appropriate H_0, their distance is an intuitive test statistic. If they are close enough, the restrictions probably hold. If they are very far, they probably do not hold.

Proposition 2 *Suppose that* $\{y_i, x_i\}_{i\in\mathbb{N}}$ *is i.i.d. with finite fourth moments. Consider the fixed-effect (FE) panel model (1-??) and assume that* $E\left[\tilde{x}_i'\tilde{x}_i\right]$ *and* $\tilde{S}_{xx}\equiv\frac{1}{N}\sum_{i=1}^N\tilde{x}_i'\tilde{x}_i$ *are positive definite. The test-statistic* h_j *defined below using (18-23) are distributed chi-squared under* H_0*. Their asymptotic power is 1. That is,*

$$
\begin{aligned}
h_j &\equiv N\left(\hat{v}_j-\hat{v}_0\right)'\hat{C}_j^{-1}\left(\hat{v}_j-\hat{v}_0\right)\sim\chi^2\left(\frac{K\left(K+1\right)}{2}+1\right)\\
&\quad \textit{under } H_0 \ \ (j=1,2,3)\,, \ \textit{and}\\
\lim_{N\to\infty}\Pr\left(h_j>c\right) &= 1 \quad (j=1,2,3) \ \textit{for any } c\in\mathbb{R} \ \textit{otherwise.}\\
\hat{C}_1 &\equiv \frac{1}{N}\sum_{i=1}^N\left(\tilde{x}_i'\check{u}_i\check{u}_i'\tilde{x}_i-\tilde{x}_i'\check{\Omega}\tilde{x}_i\right)\left(\tilde{x}_i'\check{u}_i\check{u}_i'\tilde{x}_i-\tilde{x}_i'\check{\Omega}\tilde{x}_i\right)',\\
\hat{C}_2 &\equiv \frac{1}{N}\sum_{i=1}^N\left(\tilde{x}_i'\check{u}_i\check{u}_i'\tilde{x}_i-\sum_{t=1}^T\check{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right)\left(\tilde{x}_i'\check{u}_i\check{u}_i'\tilde{x}_i-\sum_{t=1}^T\check{u}_{it}^2\tilde{x}_{it}\tilde{x}_{it}'\right)',\\
\hat{C}_3 &\equiv \frac{1}{N}\sum_{i=1}^N\left(\tilde{x}_i'\check{u}_i\check{u}_i'\tilde{x}_i-\hat{\sigma}^2\tilde{x}_i'\tilde{x}_i\right)\left(\tilde{x}_i'\check{u}_i\check{u}_i'\tilde{x}_i-\hat{\sigma}^2\tilde{x}_i'\tilde{x}_i\right)', \ \textit{where}\\
\check{\Omega} &\equiv \frac{1}{N}\sum_{i=1}^N\check{u}_i\check{u}_i' \ \textit{and}\\
\hat{\sigma}^2 &\equiv \frac{1}{N\left(T-1\right)}\sum_{i=1}^N\sum_{t=1}^T\check{u}_{it}^2
\end{aligned}
$$

The proof is straightforward provided the simplifications to V derived earlier and the consistency of the estimators $\hat{V}_j$ for the appropriate V_j. It is therefore skipped here and is available upon request.

4.1 Finite-sample properties

Tables 3 and 4 report simulated rejection rates for the three tests in the data generating processes identical to Tables 1-2, respectively, based on 10,000 Monte Carlo trials. Results for the $(500, 10)$ and $(50, 10)$ setups are presented only but all setups from Tables 1-2 were examined. The unpublished results indicate that given N, the size does not change but the power increases with T, and the test loses almost all of its power extremely small-N samples.

The results in general reflect the finite-sample properties of the estimators. The tests deliver their asymptotic properties in the $N = 500$, $T = 10$ setup. The notable exceptions are h_2 and h_3 under conditional homoskedasticity and very weak serial correlation ($\rho_u < 0.3$, $\rho_x < 0.3$, Table 3.1), and h_1 under conditional heteroskedasticity and very strong serial correlation (Table 4.1). The former are quite natural while the latter reflects that strong serial correlation dominates heteroskedasticity in the conditional variance (see Table 2.1).

The size is about right in moderate size samples. It is slightly biased upward, which makes the test a little too conservative (the actual size varies between 0.06 and 0.09 compared to a nominal size of 0.05). The power varies considerably with the alternatives. In the homoskedastic setup, the power, quite naturally, is a positive function of the serial correlation in u and x. The heteroskedastic setup yields the same result except against V_1, the heteroskedasticity-inconsistent but serial correlation consistent estimator.

5 Conclusion

The paper examined linear FE models with short time series within individual units. Serial correlation in the error process and the right-hand-side variables was shown to induce severe bias in the conventional standard error estimates. At the same time, the paper has shown that well-known robust ("clustered") estimator applied to the mean-differenced data is not only consistent but also behaves well in finite samples. Applied researchers should, therefore, routinely estimate the robust estimator in moderate-sized and large samples, the same way they already routinely estimate the heteroskedasticity-consistent estimator in cross-sectional models. The robust estimator does not get biased or significantly more disperse as the time-series dimension increases. At the same time, however, the serial correlation bias of the inconsistent estimators increases with the time-series dimension. Therefore, contrary to the intuition of many applied researchers, the advantages of the robust estimator increase as the time series get longer. It is the cross-sectional size of the sample that primarily affects the finite-sample behavior of the estimator.

In small samples and under cross-sectional homoskedasticity, there is some advantage of using the alternative serial correlation consistent estimator, the "Omega"-estimator. The conventional FE standard error estimator (the scaled version of the conventional OLS estimator on the mean-differenced data) has no significant advantage over the Omega-estimator even if both are consistent. In small samples, therefore, the Omega-estimator should be used unless there is evidence for cross-sectional heteroskedasticity.

The paper has also introduced a simple direct test for the assumptions under which the restrictive estimators are consistent. The test delivers the appropriate size properties. Its power is quite small in small samples but good enough to detect strong serial correlation.

References

[1] Arellano, M. (1987), "Computing Robust Standard Errors for Within-Groups Estimators" *Oxford Bulletin of Economics and Statistics*; 49(4), 431-434.

[2] Bertrand, M., E. Duflo, and S. Mullainathan (2001), "How Much Should We Trust Differences-in-Differences Estimates?" *Working Paper presented at the UCLA/RAND Labor and Population Workshop*, Feb. 20, 2001.

[3] Bhargava, A., L. Franzini, and W. Narendranathan (1982), "Serial Correlation and the Fixed Effects Model," *The Review of Economic Studies* 49(4) 553-549.

[4] Kiefer, N. M. (1980), " Estimation of Fixed Effect Models for Time Series of Cross-Sections with Arbitrary Intertemporal Covariance," *Journal of Econometrics* 14(2), 195-202.

[5] Magnus, J. R. and H. Neudecker (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: Wiley.

[6] White, Halbert (1980), "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48:4, 817-838.

[7] White, Halbert (1984), *Asymptotic Theory for Econometricians*. Orlando: Academic Press.

Table 1. Relative Bias ("bias": mean estimated SE over the standard deviation of the simulated distribution of β_{FE}) and **Coefficient of Variation** ("CV": standard error of the estimated SE distribution over its mean) of the four different SE estimators. **Homoskedastic errors.** In each cell, the first row corresponds to the general estimator (SE_0), the second row to the Omega-estimator (SE_1 consistent under cros-sectional homoskedasticity), the third row to the scaled version of the original White estimator (SE_2, consistent under no serial correlation) , and the fourht row to the scaled version of conventional estimator (SE_3 consistent under homoskedasticity and no serial correlation). Results from 10,000 Monte Carlo experiments.

Table 1.1. $N = 500$, $T = 10$. Homoskedastic errors.

		rho(x)							
rho(u)		**0.0**		**0.3**		**0.5**		**0.9**	
		bias	**cv**	**bias**	**cv**	**bias**	**cv**	**bias**	**cv**
0.0	**SE0**	-0.01	0.03	0.00	0.04	-0.01	0.04	-0.01	0.04
	SE1	-0.01	0.01	0.00	0.02	-0.01	0.02	0.00	0.03
	SE2	-0.02	0.02	0.00	0.02	0.00	0.02	0.00	0.03
	SE3	-0.01	0.01	0.00	0.02	-0.01	0.02	0.00	0.02
0.3	**SE0**	0.00	0.03	0.00	0.04	0.00	0.04	0.00	0.04
	SE1	0.00	0.02	0.01	0.02	0.00	0.02	0.00	0.03
	SE2	0.00	0.02	-0.06	0.02	-0.10	0.02	-0.16	0.03
	SE3	0.00	0.02	-0.06	0.02	-0.10	0.02	-0.17	0.02
0.5	**SE0**	0.01	0.04	-0.01	0.04	0.01	0.04	-0.01	0.05
	SE1	0.01	0.02	-0.01	0.02	0.01	0.02	-0.01	0.03
	SE2	0.01	0.02	-0.11	0.02	-0.16	0.02	-0.26	0.03
	SE3	0.01	0.02	-0.11	0.02	-0.16	0.02	-0.27	0.02
0.9	**SE0**	0.00	0.049	-0.01	0.049	0.00	0.049	0.00	0.051
	SE1	0.00	0.026	-0.01	0.027	0.00	0.027	0.00	0.030
	SE2	0.00	0.026	-0.17	0.027	-0.25	0.028	-0.39	0.034
	SE3	0.00	0.021	-0.17	0.021	-0.27	0.022	-0.42	0.026

Table 1.2. $N = 50$, $T = 10$. Homoskedastic errors.

rho(u)		rho(x) 0.0 bias	cv	0.3 bias	cv	0.5 bias	cv	0.9 bias	cv
0.0	**SE0**	-0.02	0.12	-0.02	0.12	-0.02	0.12	-0.04	0.14
	SE1	0.00	0.05	0.00	0.05	0.00	0.06	-0.02	0.08
	SE2	0.00	0.07	0.00	0.07	-0.01	0.07	-0.02	0.08
	SE3	0.00	0.05	0.00	0.05	0.00	0.05	-0.01	0.07
0.3	**SE0**	-0.02	0.12	-0.02	0.12	-0.02	0.12	-0.01	0.14
	SE1	0.00	0.05	0.00	0.06	0.00	0.06	0.01	0.08
	SE2	-0.01	0.07	-0.07	0.07	-0.11	0.07	-0.16	0.08
	SE3	0.00	0.05	-0.06	0.05	-0.10	0.05	-0.16	0.07
0.5	**SE0**	-0.01	0.12	-0.03	0.13	-0.02	0.13	-0.04	0.15
	SE1	0.01	0.06	-0.01	0.06	0.00	0.07	-0.02	0.09
	SE2	0.00	0.07	-0.12	0.07	-0.17	0.07	-0.27	0.09
	SE3	0.01	0.05	-0.11	0.05	-0.17	0.06	-0.27	0.07
0.9	**SE0**	-0.02	0.141	-0.02	0.145	0.00	0.147	-0.04	0.159
	SE1	0.00	0.085	0.00	0.083	0.02	0.085	-0.02	0.097
	SE2	0.00	0.082	-0.17	0.086	-0.24	0.088	-0.40	0.107
	SE3	0.00	0.066	-0.17	0.067	-0.25	0.070	-0.43	0.082

Table 1.3. $N = 50$, $T = 50$. Homoskedastic errors.

rho(u)		rho(x) 0.0 bias	cv	0.3 bias	cv	0.5 bias	cv	0.9 bias	cv
0.0	**SE0**	-0.02	0.11	0.00	0.10	-0.01	0.11	-0.01	0.12
	SE1	-0.01	0.02	0.01	0.02	0.00	0.03	0.00	0.06
	SE2	-0.01	0.03	0.01	0.03	0.00	0.03	0.01	0.05
	SE3	-0.01	0.02	0.01	0.02	0.00	0.02	0.01	0.04
0.3	**SE0**	-0.01	0.11	0.00	0.11	0.00	0.11	-0.02	0.12
	SE1	0.01	0.03	0.01	0.03	0.01	0.03	-0.01	0.06
	SE2	0.01	0.03	-0.07	0.03	-0.13	0.03	-0.23	0.05
	SE3	0.01	0.02	-0.07	0.02	-0.12	0.02	-0.23	0.04
0.5	**SE0**	-0.02	0.11	-0.01	0.11	-0.02	0.11	-0.01	0.12
	SE1	-0.01	0.03	0.00	0.03	-0.01	0.03	0.00	0.06
	SE2	-0.01	0.03	-0.13	0.03	-0.22	0.04	-0.36	0.05
	SE3	-0.01	0.03	-0.13	0.03	-0.22	0.02	-0.36	0.04
0.9	**SE0**	-0.02	0.120	-0.01	0.123	-0.02	0.123	-0.03	0.135
	SE1	-0.01	0.059	0.00	0.059	0.00	0.059	-0.01	0.071
	SE2	0.00	0.048	-0.23	0.047	-0.37	0.047	-0.63	0.065
	SE3	0.00	0.041	-0.23	0.041	-0.37	0.040	-0.64	0.054

Table 1.4. $N = 10$, $T = 50$. Homoskedastic errors.

		rho(x)							
rho(u)		**0.0**		**0.3**		**0.5**		**0.9**	
		bias	**cv**	**bias**	**cv**	**bias**	**cv**	**bias**	**cv**
0.0	**SE0**	-0.08	0.24	-0.08	0.25	-0.07	0.25	-0.09	0.27
	SE1	-0.01	0.08	-0.01	0.09	0.00	0.09	-0.02	0.14
	SE2	-0.01	0.06	-0.01	0.06	0.00	0.07	0.00	0.10
	SE3	0.00	0.04	0.00	0.05	0.01	0.05	0.00	0.09
0.3	**SE0**	-0.08	0.25	-0.09	0.25	-0.08	0.25	-0.09	0.28
	SE1	-0.02	0.09	-0.01	0.09	-0.02	0.09	-0.02	0.14
	SE2	-0.01	0.07	-0.09	0.07	-0.14	0.07	-0.23	0.10
	SE3	-0.01	0.05	-0.09	0.05	-0.14	0.05	-0.23	0.09
0.5	**SE0**	-0.08	0.25	-0.08	0.25	-0.08	0.25	-0.10	0.28
	SE1	-0.01	0.09	-0.01	0.09	-0.01	0.10	-0.02	0.14
	SE2	-0.01	0.07	-0.14	0.07	-0.22	0.07	-0.37	0.11
	SE3	0.00	0.05	-0.13	0.05	-0.22	0.06	-0.37	0.10
0.9	**SE0**	-0.10	0.275	-0.09	0.273	-0.09	0.273	-0.11	0.296
	SE1	-0.02	0.141	-0.02	0.143	-0.02	0.145	-0.03	0.167
	SE2	-0.02	0.104	-0.24	0.104	-0.37	0.108	-0.64	0.147
	SE3	-0.01	0.092	-0.23	0.094	-0.37	0.096	-0.63	0.124

Table 1.5. *N* = 10, *T* = 10. Homoskedastic errors.

		rho(x)							
rho(u)		**0.0**		**0.3**		**0.5**		**0.9**	
		bias	**cv**	**bias**	**cv**	**bias**	**cv**	**bias**	**cv**
0.0	**SE0**	-0.10	0.27	-0.09	0.27	-0.11	0.28	-0.13	0.31
	SE1	-0.03	0.13	-0.02	0.13	-0.02	0.15	-0.04	0.20
	SE2	-0.03	0.14	-0.03	0.15	-0.02	0.15	-0.04	0.18
	SE3	-0.01	0.11	-0.01	0.11	-0.01	0.12	-0.02	0.15
0.3	**SE0**	-0.08	0.27	-0.11	0.27	-0.10	0.28	-0.13	0.31
	SE1	-0.01	0.14	-0.03	0.14	-0.03	0.15	-0.04	0.19
	SE2	-0.02	0.14	-0.10	0.15	-0.13	0.15	-0.19	0.19
	SE3	0.00	0.11	-0.08	0.11	-0.11	0.12	-0.17	0.15
0.5	**SE0**	-0.10	0.27	-0.09	0.28	-0.11	0.29	-0.13	0.32
	SE1	-0.02	0.14	-0.01	0.15	-0.03	0.16	-0.04	0.20
	SE2	-0.03	0.15	-0.13	0.16	-0.18	0.16	-0.27	0.19
	SE3	-0.01	0.12	-0.10	0.12	-0.17	0.13	-0.26	0.16
0.9	**SE0**	-0.09	0.311	-0.11	0.310	-0.11	0.314	-0.14	0.331
	SE1	-0.02	0.193	-0.04	0.193	-0.03	0.195	-0.05	0.217
	SE2	-0.02	0.180	-0.19	0.185	-0.28	0.188	-0.42	0.222
	SE3	0.00	0.150	-0.18	0.154	-0.26	0.155	-0.42	0.185

Table 2. Relative Bias ("bias": mean estimated SE over the standard deviation of the simulated distribution of β_{FE}) and **Coefficient of Variation** ("CV": standard error of the estimated SE distribution over its mean) of the four different SE estimators. **Conditoinal heteroskedasticity in the cross-sectional dimension.** In each cell, the first row corresponds to the general estimator (SE_0), the second row to the Omega-estimator (SE_1 consistent under cros-sectional homoskedasticity), the third row to the scaled version of the original White estimator (SE_2, consistent under no serial correlation) , and the fourht row to the scaled version of conventional estimator (SE_3 consistent under homoskedasticity and no serial correlation). Results from 10,000 Monte Carlo experiments.

Table 2.1. $N = 500$, $T = 10$. Cross-sectional conditional heteroskedasticity.

		rho(x)							
rho(u)		**0.0**		**0.3**		**0.5**		**0.9**	
		bias	**cv**	**bias**	**cv**	**bias**	**cv**	**bias**	**cv**
0.0	**SE0**	-0.01	0.04	0.01	0.04	0.00	0.05	-0.02	0.05
	SE1	-0.28	0.01	-0.25	0.01	-0.24	0.02	-0.13	0.03
	SE2	-0.03	0.02	-0.01	0.02	-0.02	0.02	-0.02	0.02
	SE3	-0.28	0.01	-0.25	0.01	-0.24	0.02	-0.13	0.02
0.3	**SE0**	0.00	0.04	0.00	0.05	0.00	0.05	-0.01	0.06
	SE1	-0.26	0.01	-0.24	0.02	-0.23	0.02	-0.12	0.03
	SE2	-0.02	0.02	-0.08	0.02	-0.12	0.02	-0.18	0.02
	SE3	-0.26	0.01	-0.29	0.02	-0.31	0.02	-0.27	0.02
0.5	**SE0**	0.00	0.05	0.00	0.04	0.00	0.05	-0.01	0.06
	SE1	-0.23	0.01	-0.22	0.02	-0.20	0.02	-0.11	0.03
	SE2	-0.02	0.02	-0.12	0.02	-0.18	0.02	-0.27	0.03
	SE3	-0.23	0.02	-0.31	0.02	-0.34	0.02	-0.35	0.02
0.9	**SE0**	-0.02	0.05	-0.02	0.05	0.00	0.06	0.00	0.06
	SE1	-0.14	0.02	-0.12	0.03	-0.10	0.02	-0.05	0.03
	SE2	-0.02	0.02	-0.17	0.02	-0.25	0.03	-0.38	0.03
	SE3	-0.14	0.02	-0.26	0.02	-0.34	0.02	-0.45	0.03

Table 2.2. $N = 50$, $T = 10$. Cross-sectional conditional heteroskedasticity.

		rho(x)							
rho(u)		0.0		0.3		0.5		0.9	
		bias	cv	bias	cv	bias	cv	bias	cv
0.0	SE0	-0.02	0.13	-0.03	0.14	-0.03	0.14	-0.03	0.16
	SE1	-0.27	0.05	-0.26	0.05	-0.24	0.06	-0.11	0.09
	SE2	-0.03	0.09	-0.04	0.09	-0.03	0.09	-0.01	0.10
	SE3	-0.27	0.04	-0.26	0.05	-0.24	0.05	-0.11	0.07
0.3	SE0	-0.02	0.14	-0.02	0.14	-0.02	0.15	-0.03	0.16
	SE1	-0.25	0.05	-0.24	0.06	-0.22	0.06	-0.12	0.09
	SE2	-0.03	0.09	-0.09	0.09	-0.12	0.09	-0.18	0.10
	SE3	-0.26	0.05	-0.29	0.05	-0.31	0.05	-0.26	0.07
0.5	SE0	-0.02	0.14	-0.03	0.14	-0.02	0.15	-0.03	0.17
	SE1	-0.23	0.06	-0.22	0.06	-0.20	0.07	-0.10	0.10
	SE2	-0.03	0.09	-0.13	0.09	-0.18	0.09	-0.26	0.10
	SE3	-0.23	0.05	-0.31	0.05	-0.34	0.06	-0.34	0.08
0.9	SE0	-0.02	0.15	-0.03	0.16	-0.03	0.17	-0.05	0.19
	SE1	-0.12	0.08	-0.11	0.08	-0.10	0.09	-0.07	0.11
	SE2	-0.01	0.09	-0.17	0.09	-0.26	0.10	-0.40	0.12
	SE3	-0.12	0.07	-0.26	0.07	-0.34	0.07	-0.46	0.09

Table 2.3. $N = 50$, $T = 50$. Cross-sectional conditional heteroskedasticity.

rho(u)		rho(x) 0.0 bias	cv	0.3 bias	cv	0.5 bias	cv	0.9 bias	cv
0.0	**SE0**	-0.03	0.11	-0.01	0.11	-0.02	0.11	-0.03	0.13
	SE1	-0.29	0.02	-0.28	0.02	-0.28	0.03	-0.23	0.05
	SE2	-0.02	0.04	0.00	0.04	-0.01	0.04	-0.01	0.05
	SE3	-0.30	0.02	-0.28	0.02	-0.28	0.02	-0.23	0.03
0.3	**SE0**	-0.01	0.11	-0.01	0.11	-0.01	0.11	-0.03	0.14
	SE1	-0.27	0.02	-0.26	0.03	-0.26	0.03	-0.23	0.05
	SE2	0.00	0.04	-0.08	0.04	-0.13	0.04	-0.24	0.05
	SE3	-0.27	0.02	-0.32	0.02	-0.36	0.02	-0.41	0.03
0.5	**SE0**	-0.01	0.11	-0.02	0.11	-0.03	0.12	-0.03	0.14
	SE1	-0.23	0.03	-0.24	0.03	-0.24	0.03	-0.22	0.06
	SE2	0.00	0.04	-0.14	0.04	-0.23	0.05	-0.38	0.05
	SE3	-0.24	0.02	-0.34	0.03	-0.41	0.02	-0.51	0.04
0.9	**SE0**	-0.02	0.13	-0.02	0.13	-0.02	0.13	-0.03	0.16
	SE1	-0.09	0.06	-0.09	0.06	-0.10	0.06	-0.13	0.07
	SE2	-0.01	0.05	-0.23	0.05	-0.36	0.05	-0.63	0.07
	SE3	-0.09	0.04	-0.30	0.04	-0.43	0.04	-0.68	0.05

Table 2.4. $N = 10$, $T = 50$. Cross-sectional conditional heteroskedasticity.

		rho(x)							
rho(u)		**0.0**		**0.3**		**0.5**		**0.9**	
		bias	**cv**	**bias**	**cv**	**bias**	**cv**	**bias**	**cv**
0.0	**SE0**	-0.09	0.25	-0.09	0.25	-0.08	0.26	-0.12	0.29
	SE1	-0.27	0.09	-0.27	0.09	-0.26	0.09	-0.23	0.14
	SE2	-0.02	0.09	-0.02	0.09	-0.01	0.09	-0.02	0.10
	SE3	-0.29	0.04	-0.29	0.04	-0.28	0.05	-0.23	0.08
0.3	**SE0**	-0.08	0.25	-0.09	0.26	-0.09	0.26	-0.10	0.29
	SE1	-0.25	0.09	-0.25	0.09	-0.25	0.10	-0.21	0.14
	SE2	-0.02	0.09	-0.10	0.09	-0.15	0.09	-0.23	0.10
	SE3	-0.27	0.04	-0.33	0.04	-0.37	0.05	-0.39	0.08
0.5	**SE0**	-0.07	0.26	-0.09	0.26	-0.09	0.26	-0.12	0.30
	SE1	-0.22	0.09	-0.22	0.10	-0.23	0.10	-0.21	0.15
	SE2	0.00	0.09	-0.14	0.09	-0.23	0.09	-0.38	0.11
	SE3	-0.23	0.05	-0.34	0.05	-0.41	0.05	-0.50	0.08
0.9	**SE0**	-0.10	0.28	-0.10	0.28	-0.11	0.29	-0.12	0.33
	SE1	-0.11	0.14	-0.10	0.14	-0.11	0.15	-0.13	0.18
	SE2	-0.02	0.11	-0.24	0.11	-0.37	0.11	-0.63	0.15
	SE3	-0.10	0.09	-0.30	0.09	-0.43	0.09	-0.67	0.12

Table 2.5. $N = 10$, $T = 10$. Cross-sectional conditional heteroskedasticity.

		rho(x)							
rho(u)		0.0		0.3		0.5		0.9	
		bias	cv	bias	cv	bias	cv	bias	cv
0.0	SE0	-0.11	0.29	-0.12	0.29	-0.13	0.30	-0.14	0.33
	SE1	-0.26	0.13	-0.25	0.14	-0.24	0.14	-0.14	0.21
	SE2	-0.07	0.19	-0.07	0.19	-0.07	0.18	-0.05	0.20
	SE3	-0.27	0.10	-0.26	0.10	-0.25	0.11	-0.12	0.17
0.3	SE0	-0.11	0.29	-0.13	0.30	-0.13	0.30	-0.14	0.34
	SE1	-0.24	0.13	-0.25	0.14	-0.23	0.15	-0.13	0.21
	SE2	-0.06	0.18	-0.14	0.18	-0.16	0.18	-0.20	0.21
	SE3	-0.26	0.10	-0.30	0.11	-0.31	0.11	-0.25	0.17
0.5	SE0	-0.12	0.30	-0.12	0.30	-0.12	0.31	-0.14	0.34
	SE1	-0.24	0.14	-0.22	0.15	-0.20	0.16	-0.13	0.22
	SE2	-0.07	0.18	-0.16	0.19	-0.21	0.19	-0.29	0.21
	SE3	-0.24	0.11	-0.30	0.11	-0.33	0.12	-0.33	0.17
0.9	SE0	-0.12	0.32	-0.13	0.33	-0.13	0.34	-0.16	0.36
	SE1	-0.14	0.19	-0.13	0.19	-0.13	0.20	-0.11	0.24
	SE2	-0.05	0.19	-0.20	0.20	-0.28	0.20	-0.43	0.24
	SE3	-0.13	0.14	-0.26	0.15	-0.34	0.15	-0.45	0.20

Table 3. Rejection rates of h_1 (H_0: $V_1 = V_0$), h_2 (H_0: $V_2 = V_0$) and h_3 (H_0: $V_3 = V_0$). Nominal size=0.05. **Homoskedastic errors.** V_1 and V_0 are asymptotically equivalent always; V_2 and V_0, and V_3 and V_0 are asymptotically equivalent if rho(x)=0 or rho(u)=0. Results from 10,000 Monte Carlo experiments.

Table 3.1. $N = 500$, $T = 10$. Homoskedastic errors.

		rho(x)			
rho(u)		**0.0**	**0.3**	**0.5**	**0.9**
0.0	**h1**	0.05	0.05	0.04	0.04
	h2	0.04	0.06	0.04	0.06
	h3	0.05	0.05	0.05	0.05
0.3	**h1**	0.04	0.04	0.04	0.04
	h2	0.05	0.56	0.92	1.00
	h3	0.05	0.36	0.79	0.99
0.5	**h1**	0.04	0.04	0.04	0.04
	h2	0.06	0.90	1.00	1.00
	h3	0.05	0.79	0.99	1.00
0.9	**h1**	0.04	0.04	0.04	0.03
	h2	0.06	1.00	1.00	1.00
	h3	0.05	0.99	1.00	1.00

Table 3.2. $N = 50$, $T = 10$. Homoskedastic errors.

rho(u)		rho(x) 0.0	0.3	0.5	0.9
0.0	**h1**	0.09	0.08	0.08	0.08
	h2	0.08	0.08	0.08	0.09
	h3	0.09	0.09	0.09	0.10
0.3	**h1**	0.09	0.08	0.08	0.07
	h2	0.08	0.05	0.07	0.14
	h3	0.09	0.04	0.04	0.06
0.5	**h1**	0.08	0.08	0.08	0.07
	h2	0.08	0.06	0.18	0.46
	h3	0.09	0.04	0.09	0.26
0.9	**h1**	0.08	0.07	0.07	0.06
	h2	0.09	0.14	0.44	0.86
	h3	0.10	0.08	0.29	0.74

Table 4. Rejection rates of h_1 (H_0: $V_1 = V_0$), h_2 (H_0: $V_2 = V_0$) and h_3 (H_0: $V_3 = V_0$). Nominal size=0.05. **Conditional cross-sectional heteroskedasticity in the errors.** V_1 and V_0, and V_3 and V_0 are never asymptotically equivalent; V_2 and V_0 are asymptotically equivalent if rho(x)=0 or rho(u)=0. Results from 10,000 Monte Carlo experiments.

Table 4.1. $N = 500$, $T = 10$. Cross-sectional conditional heteroskedasticity.

		rho(x)			
rho(u)		**0.0**	**0.3**	**0.5**	**0.9**
0.0	**h1**	1.00	1.00	1.00	0.77
	h2	0.10	0.07	0.06	0.06
	h3	1.00	1.00	1.00	0.71
0.3	**h1**	1.00	1.00	1.00	0.53
	h2	0.08	0.73	0.96	0.99
	h3	1.00	1.00	1.00	1.00
0.5	**h1**	1.00	1.00	1.00	0.37
	h2	0.05	0.96	1.00	1.00
	h3	1.00	1.00	1.00	1.00
0.9	**h1**	0.60	0.48	0.34	0.04
	h2	0.04	0.99	1.00	1.00
	h3	0.58	1.00	1.00	1.00

Table 4.2. $N = 50$, $T = 10$. Cross-sectional conditional heteroskedasticity.

		rho(x)			
rho(u)		**0.0**	**0.3**	**0.5**	**0.9**
0.0	**h1**	0.36	0.30	0.19	0.03
	h2	0.05	0.05	0.05	0.08
	h3	0.36	0.29	0.18	0.03
0.3	**h1**	0.28	0.21	0.12	0.03
	h2	0.05	0.05	0.07	0.11
	h3	0.28	0.39	0.38	0.15
0.5	**h1**	0.19	0.14	0.08	0.03
	h2	0.06	0.07	0.18	0.32
	h3	0.20	0.40	0.48	0.36
0.9	**h1**	0.03	0.03	0.03	0.04
	h2	0.08	0.11	0.35	0.74
	h3	0.04	0.16	0.36	0.65