Journal of Econometrics 13 (1980) 185–201. © North-Holland Publishing Company

ESTIMATING THE AUTOCORRELATED ERROR MODEL WITH TRENDED DATA

Rolla Edward PARK and Bridger M. MITCHELL*

The Rand Corporation, Santa Monica, CA 90406, USA

Received March 1978, final version received June 1979

A Monte Carlo study of the small sample properties of various estimators of the linear regression model with first-order autocorrelated errors. When independent variables are trended, estimators using T transformed observations (Prais–Winsten) are much more efficient than those using $T-1$ (Cochrane–Orcutt). The best of the feasible estimators is iterated Prais–Winsten using a sum-of-squared-error minimizing estimate of the autocorrelation coefficient ρ. None of the feasible estimators performs well in hypothesis testing; all seriously underestimate standard errors, making estimated coefficients appear to be much more significant than they actually are.

1. Introduction and summary

Several estimators are commonly used to estimate the linear regression model with first-order autocorrelated errors. This Monte Carlo study extends the investigation of the small-sample properties of such estimators, first undertaken by Rao and Griliches (1969), in two major respects: (1) We provide a systematic comparison of the estimation efficiency of *all principal estimators* with *trended data* and (2) we compare estimator performance in *hypothesis testing.*[1]

The major estimators can be classified according to (a) whether $T-1$ or T transformed observations are used, (b) whether the autocorrelation coefficient ρ is known or estimated, and, if estimated, (c) whether the estimate of ρ is iterated. With trended data we find that estimators using $T-1$ observations have very low efficiency, often less than ordinary least squares (OLS), regardless of whether ρ is known or estimated. For unknown ρ iterative estimators using all T observations dominate OLS and are somewhat more efficient than two-step estimators. We find that the iterated Prais–Winsten estimator using the sum-of-squares minimizing ρ estimate performs marginally better than the full maximum likelihood estimator.

*We are grateful to James MacKinnon and the referees for many helpful suggestions which have substantially improved this paper. Remaining shortcomings are definitely our own fault.

[1]There has been a good deal of recent work on particular aspects of estimating the autocorrelated error model, but none treats all of the principal estimators and none deals with hypothesis testing. See the discussion below (section 2).

For the empirical researcher reliable hypothesis testing procedures are as important as efficient coefficient estimates. Perhaps the most serious deficiency of OLS in the presence of autocorrelation is not inefficiency but the bias in its estimated standard errors — a bias that in many situations will make the estimated coefficients appear to be much more significant than they actually are. Unfortunately, our results show that in this regard the preferred estimators, though substantially better than OLS, can still be seriously misleading.

The estimation problem and the estimators that we consider are described in the next section. For short time series ($T=20$), section 3 compares the efficiency of the various estimators, and section 4 describes their performance in hypothesis testing. Results for longer time series ($T=50$) are presented in section 5. Section 6 is a concise list of recommendations based on our results.

2. The estimation problem

2.1. The model

The commonly encountered econometric model is

$$y_t = \boldsymbol{x}_t\beta + u_t,$$
$$u_t = \rho u_{t-1} + \varepsilon_t, \qquad t=1,2,\ldots,T, \tag{1}$$

where

$$|\rho|<1, \qquad E(\varepsilon)=0, \qquad E(\varepsilon\varepsilon')=\sigma_\varepsilon^2 \boldsymbol{I}.$$

In general, $\boldsymbol{x}_t$ will include a 1 for the constant term; that is,

$$\boldsymbol{x}_t = [1, x_{2,t}, \ldots, x_{K,t}]. \tag{2}$$

For this model the $T\times T$ covariance matrix of the error vector is

$$E(\boldsymbol{uu}')=\sigma_u^2 V=\sigma_u^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & . \\ \rho^2 & \rho & 1 & \cdots & . \\ . & . & . & \cdots & . \\ \rho^{T-1} & . & . & \cdots & 1 \end{bmatrix}, \tag{3}$$

where $\sigma_u^2 = \sigma_\varepsilon^2/(1-\rho^2)$.

If ρ is known, the Aitken estimator,

$$b=(X'V^{-1}X)^{-1}X'V^{-1}y, \tag{4}$$

is best linear unbiased. Computationally, it is convenient to decompose

$$V^{-1}=[1/(1-\rho^2)]R'R,$$

where

$$R=\begin{bmatrix} \sqrt{(1-\rho^2)} & 0 & 0 & \dots & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot & \cdot \\ 0 & 0 & 0 & \dots & -\rho & 1 \end{bmatrix}, \tag{5}$$

and calculate

$$b=(X'R'RX)^{-1}X'R'Ry \tag{6}$$

as an OLS regression of the transformed variables $y^*=Ry$ on $X^*=RX$.

2.2. The estimators

The estimators that we consider may all be thought of as variants on the Aitken estimator. As shown in table 1, some use T transformed observations, and some use $T-1$; they also use different values for ρ.

Table 1
Estimators considered in this paper.

Estimated autocorrelation coefficient ($\hat{\rho}$)	Number of observations	
	$T-1$	T
Zero	—	OLS
True ρ	TRUECO	AITKEN
Sum-of-squared-error minimizing	2SCO, ITERCO	2SPW, ITERPW
Likelihood maximizing	—	BM

It is common to omit the first row in the transformation matrix $\boldsymbol{R}$.[2] We denote the reduced matrix by $\boldsymbol{S}$. Then the transformed variables $\boldsymbol{y}^* = \boldsymbol{S}\boldsymbol{y}$ and $\boldsymbol{X}^* = \boldsymbol{S}\boldsymbol{X}$ are the $T-1$ weighted first differences,

$$y_t^* = y_t - \rho y_{t-1},$$
$$\boldsymbol{x}_t^* = [1-\rho, x_{2,t} - \rho x_{2,t-1}, \ldots, x_{K,t} - \rho x_{K,t-1}]. \tag{7}$$

This is the transformation first proposed by Cochrane and Orcutt (1949).

Alternatively, Prais and Winsten (1954) recommend retaining the first row of $\boldsymbol{R}$, in which case one has T transformed observations, including in addition to (7) the transformed first observation

$$y_1^* = \surd(1-\rho^2)\, y_1,$$
$$\boldsymbol{x}_1^* = [\surd(1-\rho^2), \surd(1-\rho^2)\, x_{2,1}, \ldots, \surd(1-\rho^2)\, x_{K,1}]. \tag{8}$$

If the true value of ρ were known, its use in $\boldsymbol{R}$ with all T observations would yield the Aitken estimator. Using true ρ and $T-1$ observations would give what we call the TRUECO (true ρ, Cochrane–Orcutt) estimator.

In practice, ρ is almost never known. It is common to substitute a consistent estimate $\hat{\rho}$ based on the residuals $\hat{u}_t$ from a first-stage OLS regression using untransformed variables. The estimators we use minimize the sum of squared errors for the transformed regression, conditional on given estimates of β.[3] For the Cochrane–Orcutt (CO) transformation, the estimator is

$$\hat{\rho}_{\mathrm{CO}} = \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} \Big/ \sum_{t=1}^{T-1} \hat{u}_t^2, \tag{9a}$$

and for the Prais–Winsten (PW) transformation it is

$$\hat{\rho}_{\mathrm{PW}} = \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} \Big/ \sum_{t=2}^{T-1} \hat{u}_t^2. \tag{9b}$$

Using $\hat{\rho}_{\mathrm{CO}}$ in $\boldsymbol{S}$ produces what we call 2SCO (two-stage Cochrane–Orcutt) estimates; using $\hat{\rho}_{\mathrm{PW}}$ in $\boldsymbol{R}$ produces 2SPW (two-stage Prais–Winsten) estimates.

[2] This is done, for example, in the widely used regression package TSP; the TSP procedure CORC is the same as what we call ITERCO.

[3] We are grateful to James MacKinnon for suggesting that we use the sum-of-squared-errors minimizing estimate of ρ. See subsection 2.3 for a discussion of alternative estimators.

Iterative estimates based on estimated ρ are obtained as follows: (a) Use the second-stage estimate of β to calculate new residuals $\hat{\boldsymbol{u}}=\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b}$. (b) Use these to calculate a new estimate of ρ. (c) Use the new $\hat{\rho}$ in $\boldsymbol{S}$ (or $\boldsymbol{R}$) to calculate a new estimate of β. (d) Repeat these steps until successive estimates of ρ differ by less than $\pm\delta$. We set $\delta=0.00001$ and call the resulting estimators ITERCO (iterated Cochrane–Orcutt) using $\boldsymbol{S}$ and ITERPW (iterated Prais–Winsten) using $\boldsymbol{R}$.

There is some chance that $\hat{\rho}$ estimated according to (9a) or (9b) will take on inadmissible values ($|\hat{\rho}|\geqq 1$). When $\hat{\rho}\geqq 1$, we reset it to 0.99999 ($=1-\delta$); $\hat{\rho}\leqq -1$ becomes -0.99999.

Finally, Beach and MacKinnon (1978) proposed a full maximum likelihood estimator. Because the log likelihood function includes the term $0.5\log(1-\rho^2)$, the estimated ρ is bounded away from ± 1, so that inadmissible values of $\hat{\rho}$ do not occur. Computationally, the Beach–MacKinnon (BM) procedure is the same as ITERPW, except that a different estimate of ρ is used; the BM estimate of ρ maximizes the likelihood function conditional on estimated β.

2.3. *Alternative estimators of* ρ

Many previous Monte Carlo studies have used the following estimate of ρ:

$$\hat{\rho}=\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}\Big/\sum_{t=2}^{T}\hat{u}_t^2. \tag{10}$$

This estimator is consistent but unlike (9a) or (9b) it does not minimize the sum-of-squared-errors for either CO or PW.

In preparing this paper we found (10) to be inferior. Using (9b) rather than (10) in 2SPW and ITERPW reduces the well-known downward bias in estimated ρ, and results in slightly smaller root mean squared errors for both $\hat{\rho}$ and $\boldsymbol{b}$ in almost all cases. Using (9a) rather than (10) makes little difference in 2SCO, but has a large effect in ITERCO on the number of times $\hat{\rho}$ sticks at the boundary value 0.99999. With (10), $\hat{\rho}=0.99999$ in a large fraction of the experiments in which true $\rho\geqq 0.8$ — over 50 percent in one case. With (9a), the fraction never exceeded 4 percent, and was usually much smaller than that.[4] Since boundary estimates of ρ result in very bad estimates of the intercept coefficient β_1 (for reasons discussed in section 3 below), (9a) is decidedly preferable to (10) in ITERCO.

[4]See appendix table A.8, which is available from the authors on request, together with the other tables listed at the end of the paper. The reason for the difference is that $\hat{u}_1^2$ tends to be larger than $\hat{u}_T^2$ in ITERCO, because of the relatively small weight given the first observation in the CO transformed regression.

Other consistent estimators of ρ have been proposed. Theil (1971, p. 254) suggests incorporating a degrees-of-freedom correction that would yield estimates smaller in absolute value than (9a) and (9b). In light of the downward bias in (9a) and (9b) as they stand, this correction does not seem desirable.

Durbin (1960) proposes running an auxiliary regression to estimate ρ. Using (9a) in ITERCO and (9b) in either 2SPW or ITERPW almost always resulted in ρ estimates that were less biased and had smaller mean squared errors than the Durbin estimates.[5] On balance, the sum-of-squared-errors minimizing ρ estimators (9a) and (9b) appear to be better than any of the commonly used alternatives.

2.4. Independent variables

We analyze the relative performance of the estimators in table 1 using three independent variables. They are (a) one artificial trended series:

$$x_t = [1, t],$$

(b) one real trended series:

$$x_t = [1, GNP_t],$$

the annual U.S. gross national product in constant dollars beginning in 1950, and, for comparison, (c) one real untrended series:

$$x_t = [1, CAP_t],$$

the annual U.S. manufacturing capacity utilization rate, also beginning in 1950.[6]

We work chiefly with 20 observations ($T = 20$), a sample size representative of many time series studies. In section 5 we discuss the differences found when 50 (quarterly) observations are available.

Strictly speaking, our results are all conditional on the particular X matrices we have used. But we believe our findings are generally applicable for trended independent variables because the results are very much the same for the artificial and the real trended series. Each type of series answers

[5]Compare table A.6 from the appendix to this paper with table 2 in Park and Mitchell (1978, p. 12).

[6]Maeshiro (1976) also used (a) and (b). Instead of (c), he used quarterly capacity utilization starting with 1948.

questions left open by the other. The artificial series clearly establishes the effect of pure trends, but leaves open the question of whether the results would hold for the quirks present in real-world data – a question that the *GNP* results answer in the affirmative.[7]

2.5. Other recent work

Although several recent papers have discussed aspects of estimating the autocorrelated error model, this is the first to provide a unified investigation of all of the major estimators with trended data. Furthermore, with the exception of Park and Mitchell (1978), none of the previous work has taken up the question of hypothesis testing.

In a pair of articles, Maeshiro compares the efficiency of estimators using *known* ρ. In (1976), he shows that TRUECO is less efficient than OLS with trended data, and in (1978) he demonstrates that the Aitken estimator is often substantially more efficient than TRUECO. Beach and MacKinnon (1978) compare their proposed full maximum likelihood estimator, BM, with ITERCO, and find BM to be more efficient, especially when the data are trended. Harvey and McAvinchey, in an unpublished paper (1978), make efficiency comparisons of most of the major estimators applied to both trended and untrended data. They do not, however, consider ITERPW, the estimator that we find to be the best performer. Park and Mitchell (1978) do not consider any iterative procedures. Using untrended data, Spitzer (1979) revisits the estimators investigated by Rao and Griliches (1969) — these do not include ITERCO and ITERPW — and adds BM.

3. Efficiencies of estimators

3.1. Exact theoretical efficiencies

We can make two of our efficiency comparisons using exact formulas. For the case of known values of ρ, the exact variances of the OLS, TRUECO and Aitken estimators are given by the formulas

$$\operatorname{var}(\boldsymbol{b}_{\mathrm{OLS}})=\sigma_u^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}, \tag{11}$$

$$\operatorname{var}(\boldsymbol{b}_{\mathrm{TRUECO}})=\sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{S}'\boldsymbol{S}\boldsymbol{X})^{-1}, \tag{12}$$

$$\operatorname{var}(\boldsymbol{b}_{\mathrm{Aitken}})=\sigma_u^2(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}=\sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{R}'\boldsymbol{R}\boldsymbol{X})^{-1}. \tag{13}$$

[7]Furthermore, Monte Carlo experiments with quite different artificial time series yielded similar results; see Park and Mitchell (1979).

For these estimators we define relative efficiency as the *square root* of the ratio of the variances of the estimators being compared, for example,

$$EFF(b_{1,\mathrm{TRUECO}}) = [\mathrm{var}(b_{1,\mathrm{OLS}})/\mathrm{var}(b_{1,\mathrm{TRUECO}})]^{1/2}.$$

This definition is in accord with comparisons of standard errors or t-ratios commonly used by applied researchers; to use the ratio itself would make the differences between estimators appear larger than they 'really' are.

3.2. Experimental efficiencies

We used Monte Carlo simulation to assess the relative efficiencies for the other five estimators — 2SCO, ITERCO, 2SPW, ITERPW, and BM. For each of the independent variables $\boldsymbol{x}$ and for each value of $\rho = -0.8$, 0.0, 0.4, 0.8, 0.9, 0.98, we generated 1000 samples using model (1) with $\beta = [1, 1]$. A value u_0 was generated by drawing a random ε_0 from $N(0,1)$ and dividing by $\sqrt{(1-\rho^2)}$. Successive values of ε_t drawn from $N(0,1)$ were used to calculate $u_t = \rho u_{t-1} + \varepsilon_t$, and hence $y_t = \boldsymbol{x}_t\beta + u_t$. We then applied each estimation method and averaged the squared errors of the estimated coefficients over the 1000 samples.[8] For these estimators we define relative efficiency as the ratio of the root mean squared errors of the estimators being compared, for example

$$EFF(b_{1,2\mathrm{SCO}}) = RMSE(b_{1,\mathrm{OLS}})/RMSE(b_{1,2\mathrm{SCO}}),$$

where

$$RMSE(b_1) = \left[\sum_{1}^{1000} (b_1 - \beta_1)^2/1000\right]^{1/2}.$$

3.3. Efficiency of estimators that use $T-1$ transformed observations

Table 2 shows the efficiency, relative to OLS, of the three estimators that use $T-1$ observations. Here we focus on the results for positive ρ.[9] For trended variables all three estimators are less efficient than OLS in almost all of the cases tabulated.

For $\boldsymbol{x}_t = [1, t]$, TRUECO has extremely low efficiency as ρ approaches 1. This poor performance is the result of collinearity; because the transformed

[8]The calculations were done in double precision on an IBM 370-158 using regression analysis subroutines from the STATLIB statistical package.

[9]Results for $\rho = -0.8$ and $\rho = 0.0$ are included in the appendix tables.

Table 2

Efficiency, relative to OLS, of estimators that use $T-1$ transformed observations ($T=20$).

		ρ							
		0.4		0.8		0.9		0.98	
Independent variable	Estimator	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
t	TRUECO	0.81	0.86	0.50	0.62	0.29	0.42	0.04	0.11
	2SCO	0.81	0.86	0.64	0.77	0.31	0.62	0.66	0.74
	ITERCO	0.80	0.85	0.51	0.69	0.27	0.56	0.54	0.64
GNP_t	TRUECO	0.88	0.91	0.71	0.81	0.57	0.75	0.29	0.71
	2SCO	0.91	0.93	0.84	0.91	0.87	0.95	0.95	1.03
	ITERCO	0.73	0.85	0.59	0.80	0.51	0.83	0.52	0.88
CAP_t	TRUECO	1.10	1.10	1.85	1.83	2.10	2.19	1.04	2.51
	2SCO	1.05	1.04	0.01	1.41	0.00	1.65	0.00	1.83
	ITERCO	1.03	1.03	0.01	1.65	0.00	2.03	0.00	2.27

Note: Exact theoretical relative efficiency for TRUECO; experimental relative efficiency for 2SCO and ITERCO.

variable $x^*_{2,t}=x_{2,t}-\rho x_{2,t-1}$ approaches the same value for all t, the $T-1$ vector $\boldsymbol{x}^*_2$ becomes collinear with the transformed vector for the constant term $x^*_{1,t}=1-\rho$, a situation well known for producing inefficient estimates. More generally, the CO transformation of any linearly trended variable (using $\rho>0$) produces observations that are more nearly constant than are the untransformed values, and hence more nearly collinear with the constant vector.[10]

When ρ is large, 2SCO using trended data performs better than TRUECO.[11] However, it is less efficient than OLS in almost all cases. Iterating on $\hat{\rho}$ makes matters worse. For trended variables, ITERCO is less efficient than 2SCO in all of the cases tabulated.

For the untrended variable CAP the picture is mixed. All three CO slope estimators are more efficient than OLS, but 2SCO and ITERCO produce

[10]Maeshiro (1976) explains the poor performance of the TRUECO estimator as a result of reduction of variance of the transformed independent variables. In one sense this is equivalent to our 'collinearity' explanation, because a low-variance independent variable is represented by a nearly constant vector, which is nearly collinear with the vector for the constant term. But in another sense it is somewhat misleading. In a model without a constant term, an independent variable with low (or even zero) variance causes no problems. For example, the estimator of β in $y_t=\beta x_t+u_t$, $E(u_t)=0$, when $x_t=k$ for all t, has variance $\text{var}(b)=\sigma^2/Tk^2$.

[11]This refutes Maeshiro's (1976) conjecture that 'an estimator utilizing relevant extraneous information (e.g., the true value of ρ) cannot be less efficient than an estimator not utilizing it'. The reason, paradoxically, is the downward bias in $\hat{\rho}$ (see appendix table A.7). The higher the value of ρ used in the transformation, the more collinear are the transformed variables.

very bad intercept estimates when $\hat{\rho}$ takes on the boundary value 0.99999. This happened between 3 and 35 times out of 1000 trials in the cases tabulated,[12] causing very large root mean squared errors and resulting in near-zero relative efficiency.

Since a few very bad estimates can dominate the experimental relative efficiencies, table 2 might conceivably hide a good performance in most of the trials. This is *not* the case for trended variables, although it is true for CAP. Table 3 shows the number of times out of 1000 that each estimator came closer to the true value of the coefficient than did OLS. For trended variables, none of the CO estimators was closer than OLS as much as half of the time in any case tabulated.[13]

Table 3

Number of times in 1000 trials that estimators that use $T-1$ transformed observations beat OLS ($T=20$).

		ρ							
		0.4		0.8		0.9		0.98	
Independent variable	Estimator	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
t	TRUECO	371	394	236	316	151	222	24	74
	2SCO	382	413	325	378	324	349	352	397
	ITERCO	381	410	321	372	318	340	347	383
GNP_t	TRUECO	443	461	382	441	324	398	192	374
	2SCO	443	457	421	468	433	454	438	495
	ITERCO	441	453	405	451	395	423	395	450
CAP_t	TRUECO	568	563	726	731	728	713	523	781
	2SCO	547	534	737	737	742	741	721	793
	ITERCO	532	521	722	725	720	732	703	786

Note: Counts greater than 531 or smaller than 469 are significantly different from 500 at the 0.05 level.

The lesson is clear: Avoid using any form of the Cochrane–Orcutt estimator.

3.4. *Efficiency of estimators that use T transformed observations*

Turning now to estimators that use all T transformed observations, we see in table 4 that they all provide more efficient estimates than does OLS. By

[12] See appendix table A.8.

[13] The proportion is significantly less than one-half on a binomial test at the 0.05 significance level in all but two cases.

retaining the differentially weighted first observation, these estimators break the collinearity that plagued the CO estimators.

For trended data, the Aitken estimator provides respectable, although not spectacular, efficiency improvements over OLS, ranging up to 26 percent in the cases tabulated. The efficiency gain is highest in just those cases where CO performs worst, that is, for $\rho \geqq 0.8$. The other methods, using estimated rather than true ρ, preserve about half of the Aitken efficiency improvement. For untrended data, the efficiency gains are larger, and more of the gain is retained when ρ must be estimated.

Iteration helps. ITERPW is a little more efficient than 2SPW using trended data, and substantially more efficient with untrended data. The BM estimator has virtually the same efficiency as ITERPW.

Using ITERPW as a standard of comparison, we show in table 5 the number of times in 1000 trials that the other estimators using T transformed observations come closer to the true coefficient value. In almost all cases, ITERPW outperformed the other estimators that use estimated ρ.

4. Hypothesis testing

In our opinion, the most troublesome characteristic of OLS when $V \neq I$ is not the loss in efficiency, but the bias in the statistic $\hat{\sigma}_u^2(X'X)^{-1}$, which is

Table 4

Efficiency, relative to OLS, of estimators that use T transformed observations ($T=20$)

		ρ							
		0.4		0.8		0.9		0.98	
Independent variable	Estimator	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
t	AITKEN	1.02	1.02	1.08	1.09	1.08	1.10	1.03	1.08
	2SPW	1.01	1.01	1.05	1.06	1.05	1.08	1.02	1.05
	ITERPW	1.01	1.01	1.06	1.07	1.05	1.08	1.02	1.05
	BM	1.01	1.01	1.05	1.06	1.05	1.07	1.02	1.04
GNP_t	AITKEN	1.02	1.02	1.13	1.14	1.16	1.20	1.13	1.26
	2SPW	1.01	1.01	1.08	1.09	1.10	1.13	1.06	1.12
	ITERPW	1.00	1.01	1.08	1.09	1.11	1.14	1.07	1.14
	BM	1.01	1.01	1.08	1.09	1.10	1.13	1.06	1.12
CAP_t	AITKEN	1.14	1.14	1.85	1.86	2.15	2.21	1.95	2.52
	2SPW	1.06	1,06	1.38	1.38	1.57	1.60	1.54	1.75
	ITERPW	1.04	1.05	1.61	1.61	1.94	2.00	1.77	2.15
	BM	1.05	1.06	1.58	1.58	1.92	1.97	1.75	2.12

Note: Exact theoretical relative efficiency for AITKEN; experimental relative efficiency for 2SPW, ITERPW, and BM.

Table 5

Number of times in 1000 trials that estimators that use T transformed observations beat ITERPW ($T=20$).

		ρ							
		0.4		0.8		0.9		0.98	
Independent variable	Estimator	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
t	AITKEN	517	503	534	537	517	546	563	542
	2SPW	493	514	449	445	430	420	458	439
	BM	489	510	446	445	450	406	433	432
GNP_t	AITKEN	532	532	527	527	564	591	568	550
	2SPW	502	499	433	428	391	383	419	377
	BM	499	494	438	433	382	380	409	364
CAP_t	AITKEN	561	565	561	564	523	531	539	538
	2SPW	500	500	398	390	399	423	434	417
	BM	512	514	430	419	436	449	427	423

Note: Counts greater than 531 or smaller than 469 are significantly different from 500 at the 0.05 level.

conventionally used as an estimator of the covariance matrix of the estimated coefficients. How serious this bias can be is illustrated by the experimental results shown in table 6. We focus on the methods that use all T transformed observations, since they are always more efficient than methods using $T-1$ observations. OLS seriously underestimates standard errors for all cases shown. For example, for trended data and $\rho \geqq 0.8$, when one applies a two-tailed test at the 0.05 level, the underestimate is large enough to lead one to judge an estimated coefficient to be significantly different from its true value 45 to 85 percent of the time.

The Aitken estimator provides unbiased variance estimates, of course.[14] Unfortunately, the procedures using estimated ρ do not do as well, although they do improve on OLS. ITERPW is the best of the lot. Still, for $\rho \geqq 0.8$ and trended data, it would result in rejection of a correct null hypothesis at least 25 percent of the time.

For the untrended variable *CAP*, the results are qualitatively similar, but the biases are less severe than in the case of trended variables.

On hypothesis testing grounds, ITERPW appears to be superior to both 2SPW and BM.[15]

[14]For all of the estimation methods that use transformed variables, the covariance matrix of the estimated coefficients is estimated directly in the transformed regressions as $\hat{\sigma}_\varepsilon^2(X^{*\prime}X^*)^{-1}$, where $\hat{\sigma}_\varepsilon^2 = \hat{\varepsilon}'\hat{\varepsilon}/(T-K)$.

[15]Had we used the maximum likelihood estimate $\tilde{\sigma}_\varepsilon^2 = \hat{\varepsilon}'\hat{\varepsilon}/T$ in the BM procedure, the margin of superiority of ITERPW over BM would have been slightly larger.

Table 6

Number of type 1 errors in 1000 trials at 0.05 significance level ($T=20$).

Independent variable	Estimator	ρ = 0.4 b_1	0.4 b_2	0.8 b_1	0.8 b_2	0.9 b_1	0.9 b_2	0.98 b_1	0.98 b_2
t	OLS	193	197	502	490	645	571	848	709
	2SPW	125	132	302	293	411	340	690	473
	ITERPW	124	131	293	285	401	336	700	474
	BM	126	133	312	305	433	360	731	503
GNP_t	OLS	186	185	457	449	601	596	730	666
	2SPW	138	138	258	251	354	343	509	413
	ITERPW	136	136	254	246	343	322	486	397
	BM	139	138	261	258	375	352	534	432
CAP_t	OLS	147	143	304	294	341	323	407	322
	2SPW	110	106	153	149	154	144	211	137
	ITERPW	115	113	102	101	90	86	144	86
	BM	112	109	107	107	98	92	176	86

Note: Counts greater than 63 or smaller than 37 are significantly different from 50 at the 0.05 level.

5. Longer time series

In this section we investigate how the results change when 50, rather than 20, observations are available for one of our independent variables, $x_t = [1, GNP_t]$. In order to increase the length of the *GNP* time series to get 50 observations, we must shift from annual to quarterly observations. Because the quarterly series exhibits short-term fluctuations that are averaged out in the annual series, our $T=50$ series is less trended than the annual *GNP* series used above. It is, however, typical of longer economic time series.

A larger sample markedly improves the estimators of ρ. For example, when true $\rho=0.9$, the mean value of $\hat{\rho}_{\text{ITERPW}}$ increased from 0.59 for $T=20$ to 0.80 for $T=50$.[16] The bias, although still clearly apparent, is greatly reduced.

Tables 7, 8 and 9 repeat the information in tables 2 through 6 for $T=50$. For the most part, the conclusions for $T=20$ also apply for the larger sample size. Estimators using $T-1$ transformed observations are usually less efficient than OLS. Those using T observations always improve on OLS, and the margin is wider for the larger sample size. Also, methods using estimated ρ retain more of the Aitken estimator's margin of improvement (reflecting the

[16]See appendix table A.7.

Table 7

Efficiency comparisons for estimators that use $T-1$ transformed observations ($T=50$).

Independent variable	Estimator	ρ 0.4 b_1	b_2	0.8 b_1	b_2	0.9 b_1	b_2	0.98 b_1	b_2
		Efficiency relative to OLS[a]							
GNP_t	TRUECO	0.90	0.91	0.84	0.87	0.87	0.94	0.92	1.39
	2SCO	0.91	0.92	0.85	0.88	0.91	0.95	0.02	1.13
	ITERCO	0.91	0.92	0.70	0.77	0.72	0.84	0.02	1.12
		Number of times in 1000 trials that other estimators beat OLS[b]							
GNP_t	TRUECO	427	428	406	430	433	461	445	599
	2SCO	425	432	401	416	449	477	509	553
	ITERCO	425	430	387	400	399	429	423	512

[a]Exact theoretical relative efficiency for TRUECO; experimental relative efficiency for 2SCO and ITERCO.

[b]Counts greater than 531 or smaller than 469 are significantly different from 500 at the 0.05 level.

Table 8

Efficiency comparisons for estimators that use T transformed observations ($T=50$).

Independent variable	Estimator	ρ 0.4 b_1	b_2	0.8 b_1	b_2	0.9 b_1	b_2	0.98 b_1	b_2
		Efficiency relative to OLS[a]							
GNP_t	AITKEN	1.02	1.02	1.19	1.19	1.40	1.41	1.78	1.89
	2SPW	1.02	1.02	1.15	1.14	1.28	1.28	1.41	1.44
	ITERPW	1.02	1.02	1.12	1.11	1.26	1.26	1.52	1.57
	BM	1.02	1.02	1.14	1.13	1.28	1.29	1.48	1.52
		Number of times in 1000 trials that other estimators beat ITERPW[b]							
GNP_t	AITKEN	490	494	564	556	553	561	586	597
	2SPW	500	491	440	432	426	415	383	354
	BM	493	484	441	445	427	415	384	353

[a]Exact theoretical relative efficiency for AITKEN; experimental relative efficiency for 2SPW, ITERPW, and BM.

[b]Counts greater than 531 or smaller than 469 are significantly different from 500 at the 0.05 level.

Table 9

Number of type 1 errors in 1000 trials at 0.05 significance level ($T = 50$).

		ρ							
		0.4		0.8		0.9		0.98	
Independent variable	Estimator	b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
GNP_t	OLS	209	209	501	505	636	631	754	781
	2SPW	90	92	143	143	198	202	377	366
	ITERPW	87	92	138	137	184	191	307	296
	BM	90	92	151	151	200	208	357	338

Note: Counts greater than 63 or smaller than 37 are significantly different from 50 at the 0.05 level.

improved estimate of ρ). ITERPW appears to be slightly better than either 2SPW or BM.

Increased sample size does nothing to reduce the bias in the OLS estimated standard errors, but it does improve the ITERPW estimates. Nevertheless, ITERPW would still lead to rejection of a correct null hypothesis up to 30 percent of the time in the cases tabulated.

6. Recommendations

Our results lead us to offer the following guidelines to practicing econometricians working with trended data in the presence of autocorrelation:

(1) Avoid the Cochrane–Orcutt estimator (using $T-1$ transformed observations); it is more complicated than OLS and often less efficient.

(2) Use the iterative version of the Prais–Winsten estimator (using T transformed observations). It offers efficiency gains over OLS that range from modest to substantial. It is slightly but clearly superior to two-stage Prais–Winsten. For trended data and a large autocorrelation coefficient, it also appears to have a slight edge in small samples over the full maximum likelihood method proposed by Beach and MacKinnon.

(3) Distrust the conventional t-statistics. The OLS standard errors are vastly underestimated. The iterative Prais–Winsten standard errors are a substantial improvement, but still highly misleading. Because estimated coefficients seem much more significant than they really are, apply a more stringent confidence level for hypothesis testing.

Appendix tables

The following appendix tables are available from the authors upon request. While the text tables show results for selected estimators for positive values of ρ only, the appendix tables include results for all relevant estimators for $\rho = -0.8$, 0.0, 0.4, 0.8, 0.9, and 0.98.

Table	*Title*
A.1	Exact theoretical efficiency, relative to OLS, of estimators that use true ρ
A.2	Experimental efficiency of various estimators relative to OLS
A.3	Number of times in 1000 trials that various estimators beat OLS
A.4	Number of times in 1000 trials that various estimators beat ITERPW
A.5	Number of type 1 errors in 1000 trials at 0.05 significance level
A.6	Performance of estimators of ρ
A.7	Number of times in 1000 trials that various estimators of ρ beat the ITERPW estimator of ρ
A.8	Number of times in 1000 trials that estimated ρ equals boundary value
A.9	Number of iterations and failures to converge

References

Beach, Charles M. and James G. MacKinnon, 1978, A maximum likelihood procedure for regression with autocorrelated errors, Econometrica, Jan.

Cochrane, D. and G.H. Orcutt, 1949, Application of least squares regression to relationships containing autocorrelated error terms, American Statistical Association Journal, March.

Durbin, J., 1960, The fitting of time series models, International Statistical Institute Review 28, no. 3.

Harvey, A.C. and I.D. McAvinchey, 1978, The small sample efficiency of two-step estimators in regression models with autoregressive disturbances. Discussion Paper no. 78–10, April (University of British Columbia, Vancouver, BC).

Maeshiro, Asatoshi, 1976, Autoregressive transformations, trended independent variables and autocorrelated disturbance terms, Review of Economics and Statistics, Nov.

Maeshiro, Asatoshi, 1979, On the retention of the first observations in serial correlation adjustment of regression models, International Economic Review, Feb.

Park, R.E. and B.M. Mitchell, 1978, Estimating the autocorrelated error model with trended data, R-2273-NIE/RC, March (Rand Corporation, Santa Monica, CA).

Park, R.E. and B.M. Mitchell, 1979, Maximum likelihood vs. minimum sum-of-squares estimation of the autocorrelated error model, N-1325, Nov. (Rand Corporation, Santa Monica, CA).

Prais, G.J. and C.B. Winsten, 1954, Trend estimators and serial correlation, Discussion Paper no. 383 (Cowles Commission, Chicago, IL).

Rao, Potluri and Zvi Griliches, 1969, Small-sample properties of several two-stage regression methods in the context of auto-correlated errors, American Statistical Association Journal, March.

Spitzer, John J., 1979, Small-sample properties of nonlinear least squares and maximum likelihood estimators in the context of autocorrelated errors, Journal of the American Statistical Association, March.

Theil, Henri, 1971, Principles of econometrics (Wiley, New York).