Psychological Methods
2000, Vol. 5, No. 1, 87–101

1082-989X/00/$5.00 DOI: 10.1037//1082-989X.5.1.87

A Double Bootstrap Method to Analyze Linear Models With Autoregressive Error Terms

Scott D. McKnight
Toledo Hospital

Joseph W. McKean and Bradley E. Huitema
Western Michigan University

A new method for the analysis of linear models that have autoregressive errors is proposed. The approach is not only relevant in the behavioral sciences for analyzing small-sample time-series intervention models, but it is also appropriate for a wide class of small-sample linear model problems in which there is interest in inferential statements regarding all regression parameters and autoregressive parameters in the model. The methodology includes a double application of bootstrap procedures. The 1st application is used to obtain bias-adjusted estimates of the autoregressive parameters. The 2nd application is used to estimate the standard errors of the parameter estimates. Theoretical and Monte Carlo results are presented to demonstrate asymptotic and small-sample properties of the method; examples that illustrate advantages of the new approach over established time-series methods are described.

Data collected under a time-series design are frequently encountered in many applied and theoretical areas. Although there are many types of time-series design, the interrupted time-series quasiexperiment is perhaps the most familiar to behavioral science researchers. This frequently encountered version of time-series design has two phases. The first phase is called the baseline phase and is followed with the introduction of the second phase, an intervention of some type. All observations obtained subsequent to the introduction of the intervention condition constitute the intervention data. Because the conditions present during the baseline phase are interrupted by the introduction of the intervention condition, the term *interrupted time-series quasiexperiment* is often used to label this type of study. An examination of recent methodological literature reveals many expository presentations and recommendations for the increased use of this design (e.g., Cook & Shadish, 1994; Glass, 1997; Marcantonio & Cook, 1994; Mohr, 1995; Rushe & Gottman, 1993).

The main purpose of the statistical analysis is to provide an evaluation of possible differential performance under the two conditions of the time series. A useful framework for this analysis is often a linear model. Although pure autoregressive integrated moving average (ARIMA) intervention models are frequently recommended for the analysis of interrupted time-series data, it has recently been demonstrated that appropriate forms of conventional regression models often fit such data adequately (Huitema & McKean, 1998). This result is especially important when sample size is small (i.e., $N < 50$) because conventional ARIMA modeling is not generally recommended for small samples (see, e.g., Box, Jenkins, & Reinsel, 1994, p. 17). Unfortunately, however, it is not unusual for the errors of a conventional regression model applied to time-series data to be autocorrelated; an alternative regression model that acknowledges the autocorrelated errors should be considered in this case.

In this article we are concerned with linear models

Scott D. McKnight, Department of Clinical Outcomes and Resource Management, Toledo Hospital, Toledo, Ohio; Joseph W. McKean, Department of Mathematics and Statistics, Western Michigan University; Bradley E. Huitema, Department of Psychology, Western Michigan University.

Correspondence concerning this article should be addressed to Bradley E. Huitema, Department of Psychology, Western Michigan University, Kalamazoo, Michigan 49008. Electronic mail may be sent to brad.huitema@wmich.edu.

whose errors follow an autoregressive time-series process. Models of this type are described in the statistical literature under various headings such as *combined transfer function-disturbance models, dynamic regression models,* and *models with regression terms and time-series errors* (Box et al., 1994; Fuller, 1996). These models have considerable potential in the behavioral sciences for the analysis of interrupted time-series quasiexperiments, but, as with ARIMA, they require a reasonably large sample size ($N \geq 50$). Hence, both ARIMA and dynamic regression models are likely to be poor choices if the sample size is small.

This is the situation in which the method of the present article may be of interest; it was specifically developed and evaluated for linear regression intervention applications having both small sample size and autocorrelated errors. Although the focus of the methodology presented here is on simple, interrupted time-series designs having the intervention(s) modeled in the design matrix, our development can be generalized for more complex linear models with time-series errors. Extensions can include polynomials used to model trends over the different phases of the design (each phase may have its own polynomial model) and covariates of an appropriate form. Because models of the type mentioned here require regression parameters as well as autoregressive parameters, the goal of our work is to provide valid inferential procedures for both types of parameter. Existing approaches are available to provide hypothesis tests and confidence intervals on the regression parameters, but they have not been found to be successful in the small sample size situation.

Popular approaches for time-series regression analysis include the Cochrane–Orcutt (Cochrane & Orcutt, 1949) and Prais–Winsten (Prais & Winsten, 1954) versions of generalized least squares (GLS), the Durbin two-stage procedure (Durbin, 1960), and, more recently, maximum-likelihood estimation. (Examples of the application of GLS to interrupted time-series data can be found in Berry & Lewis-Beck, 1986). Except for maximum-likelihood estimation, these methods first obtain estimates of the autoregressive parameters from the residuals of an initial fit of the linear model (first stage) and then, on the basis of these estimates, refit the linear model (second stage). Unfortunately, it has been shown in the econometrics literature that conventional GLS procedures produce unacceptably high Type I error rates when applied to small samples (e.g., Johnston, 1984). Although researchers sometimes suggest that the departure of empirical Type I error from the nominal value can be attributed completely to bias in the autoregressive coefficient estimators, we demonstrate that in the small-sample case the problem has two sources: (a) the bias of the initial estimates of the autoregressive parameters and (b) error variance estimation issues that are not solved through improved estimation of the autoregressive parameters.

Evidence of the first problem is presented in Table 1. This table contains partial results of a large Monte Carlo study we performed to investigate the performance of the well-known Prais–Winsten version of GLS in the context of a two-phase intervention model with first-order autoregressive errors (u_t; defined as $\rho u_{t-1} + e_t$, $t = 1, \ldots, N$, where the errors e_t are independent and normally distributed with a mean of 0 and a variance of 1). The design matrix associated with this model is of the form shown here as Equation 1:

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & n_1 & 0 & 0 \\ 1 & n_1+1 & 1 & 0 \\ 1 & n_1+2 & 1 & 1 \\ 1 & n_1+3 & 1 & 2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & n_1+n_2 & 1 & n_2-1 \end{bmatrix}. \quad (1)$$

An inspection of the empirical mean values in Table 1 (based on 5,000 simulations) reveals a large negative bias in the estimates of ρ associated with the

Table 1
Empirical Means and Variances of $\hat{\rho}_{LS}$ Under Equation 1 and First-Order Autoregressive Errors

	$\hat{\rho}_{LS}$	
Actual ρ	*M*	Variance
.90	.391	.029
.80	.360	.030
.70	.323	.032
.50	.211	.032
.30	.083	.030
.20	.015	.029
.00	–.135	.028
–.30	–.360	.024

Note. Mean values are based on 5,000 simulations. $N = 30$ ($n_1 = n_2 = 15$).

Prais–Winsten procedure when $N = n_1 + n_2 = 30$ for the entire sample. The Prais–Winsten estimator for ρ is

$$\hat{\rho}_{LS} = \frac{\sum_{t=2}^{n_1+n_2} r_t r_{t-1}}{\sum_{t=2}^{n_1+n_2} r_t^2},$$

where r_t is the least squares (LS) residual from the fit of the responses (y_t) using the design matrix in Equation 1.

Figure 1 illustrates the problem of inflated Type I error resulting partially from using the biased estimator $\hat{\rho}_{LS}$, in the Prais–Winsten transformation. It can be seen that Type I error greatly exceeds the nominal level (.05) for tests on each of the four parameters in the model. Parameters β_0, β_1, β_2, and β_3 are the intercept, first-phase slope, level change, and slope-change parameters, respectively. Because the degree of departure of the empirical value from the nominal value increases greatly with positive ρ, it is clear that the Prais–Winsten procedure is not effective in modeling autocorrelation among the errors when analyzing a short time series.

Bias adjustments to $\hat{\rho}_{LS}$ and other traditional estimates have been suggested (e.g., Beesley, Doran, & Griffiths, 1987; Griliches & Rao, 1969), but these adjusted estimators only modestly improve on the results displayed in Figure 1 (Judge, Griffiths, Hill, Lutkepohl, & Lee, 1988; McKnight, 1994). More effective methods are required for bias reduction in the estimation of both the autoregressive coefficients and the standard errors of the coefficients in the beta vector. Bootstrap methodology (Efron & Tibshirani, 1994) lends itself to problems of this nature.

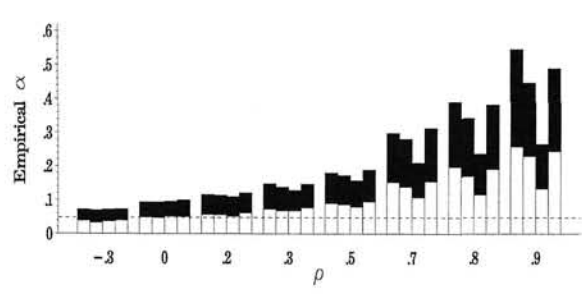

Figure 1. Prais–Winsten empirical Type I errors. Empirical αs are based on 5,000 simulations of $N = 30$. The four parameters plotted at each level of ρ are, in order, as follows: intercept (β_0), first-phase slope (β_1), level change (β_2), and slope change (β_3). Open boxes correspond to left-tailed tests and filled boxes to right-tailed tests. The horizontal line corresponds to a nominal α of .05.

Recent work has been reported using bootstrap methods to control the levels of the test for zero slope in linear-trend models with autoregressive errors. Nankervis and Savin (1996) and Woodward, Bottone, and Gray (1997) presented bootstrap procedures formulated under a simple linear model constrained by the null hypothesis of zero slope. Sun and Pantula (1996) have evaluated the Woodward et al. procedure.

The goal of our bootstrap procedure is somewhat more complex than that of the procedures mentioned above. We are concerned with inference (confidence intervals as well as tests of general linear hypotheses) for regression parameters as well as autoregressive parameters. The estimation combines a two-stage Durbin-type approach (Durbin, 1960; Fuller, 1996, chap. 9) and a bootstrap procedure similar to one developed by Freedman (1984). This approach leads to natural estimates of the degree of bias of the autoregressive parameters and, ultimately, to correction terms for these parameters. This process can be iterated until the incremental change in the autoregressive estimates is small. These new autoregressive estimates are then used to obtain Durbin-type estimates of the regression parameters and bootstrap estimates of their standard errors. Details of the method are presented in subsequent sections of this article.

Method

Model and Notation

Consider the general model

$$y_t = \mathbf{x}_t'\boldsymbol{\beta} + u_t,\ t = 1, \ldots, N \qquad (2)$$

where $\mathbf{x}_t$ is a $(p + 1) \times 1$ vector of prespecified design regressors, the error term u_t follows a stationary autoregressive (k) series,

$$u_t = \rho_1 u_{t-1} + \cdots + \rho_k u_{t-k} + e_t,$$

and the errors e_t are independent and identically distributed with a mean of zero and finite variance. Hence, we are assuming homoscedastic errors. We further assume that the model contains an intercept (i.e., let $x_{t1} = 1$ for all t). Our interest is in the estimation of the autoregressive and regression parameters and inference concerning these parameters.

The model in Equation 2 is the second model discussed by Durbin (1960) and is also discussed by

Fuller (1996, p. 530). Note that we can rewrite Equation 2 equivalently as Equation 3:

$$y_t = \rho_1 y_{t-1} + \cdots + \rho_k y_{t-k} + \mathbf{x}_t'\boldsymbol{\beta} - \mathbf{x}_{t-1}'\rho_1\boldsymbol{\beta} - \cdots - \mathbf{x}_{t-k}'\rho_k\boldsymbol{\beta} + e_t. \tag{3}$$

Suppose, for the moment, the vector of autoregressive parameters $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_k)'$ is known. Consider the transformation

$$v_t(\boldsymbol{\rho}) = y_t - \rho_1 y_{t-1} - \cdots - \rho_k y_{t-k}, \tag{4}$$

$$\mathbf{w}_t(\boldsymbol{\rho}) = \mathbf{x}_t - \rho_1\mathbf{x}_{t-1} - \cdots - \rho_k\mathbf{x}_{t-k} \tag{5}$$

for $t = k + 1, \ldots, N$. Under normal errors the LS estimate of $\boldsymbol{\beta}$ based on the regression of $v_t(\boldsymbol{\rho})$ on $\mathbf{w}_t(\boldsymbol{\rho})$ is optimal, except for end effects. Under regularity conditions, if the errors are not normally distributed, the LS estimate of $\boldsymbol{\beta}$ is asymptotically normal (Fuller, 1996).

Of course in practice, $\boldsymbol{\rho}$ is not known. The Cochrane–Orcutt and Prais–Winsten estimated GLS procedures make use of this transformation on the basis of initial estimates of $\boldsymbol{\rho}$ (see, e.g., chap. 12 of Neter, Kutner, Nachtsheim, & Wasserman, 1996). Our bootstrap procedure uses the Durbin two-stage procedure for estimation of the regression and autoregressive parameters; hence, we briefly describe this procedure.

Stage 1. Equation 3 is fit using ordinary LS to obtain estimates of both the autoregressive parameters and the regression coefficients simultaneously. Let $\hat{\boldsymbol{\rho}}_1$ denote the Durbin estimate of the vector of the autoregressive parameters.

Stage 2. Perform the transformations in Equations 4 and 5 using $\hat{\boldsymbol{\rho}}_1$ to form $v(\hat{\boldsymbol{\rho}}_1)$ and $\mathbf{w}_t(\hat{\boldsymbol{\rho}}_1)$. Note that $\mathbf{w}_t(\hat{\boldsymbol{\rho}}_1)$ is a $(p + 1) \times 1$ vector. Then obtain the LS fit of $v_t(\hat{\boldsymbol{\rho}}_1)$ on $\mathbf{w}_t(\hat{\boldsymbol{\rho}}_1)$. Denote the LS estimates by $\hat{\boldsymbol{\gamma}}$. Finally, let $\hat{\boldsymbol{\beta}}_1$ be the $(p + 1) \times 1$ dimensional vector with components

$$\hat{\beta}_{1j} = \begin{cases} \dfrac{\hat{\gamma}_0}{1 - \hat{\rho}_1 - \cdots - \hat{\rho}_k} & \text{if } j = 0 \\ \hat{\gamma}_j & \text{if } j = 1, \ldots, p\,. \end{cases}$$

The estimator $\hat{\boldsymbol{\beta}}_1$ is a consistent estimate of $\boldsymbol{\beta}$ (see Durbin, 1960). Durbin also showed that the asymptotic distribution of $\sqrt{n}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta})$ is the same as the LS estimate of $\boldsymbol{\beta}$ when the actual values of the autoregressive parameters $\rho_1, \ldots, \rho_k$ are known.

Bootstrap Procedure

For ease of discussion, we define our bootstrap procedure for Equation 2 with first-order autoregressive (AR [1]) errors; thus, $k = 1$ here. Our discussion, however, is general and holds for AR(k). The basic idea of the bootstrap is to generate replicate series of the original time series to estimate bias in the Durbin estimate of ρ and standard errors of the Durbin estimates of $\boldsymbol{\beta}$. To generate the replicate series, first obtain the Durbin two-stage estimators $\hat{\boldsymbol{\beta}}_1$ and $\hat{\rho}_1$ as discussed in the previous section. Next, form the residuals $\hat{e}_t$ as

$$\hat{e}_t = y_t - \hat{\rho}_1 y_{t-1} - (\mathbf{x}_t - \hat{\rho}_1\mathbf{x}_{t-1})'\hat{\boldsymbol{\beta}}_1,\ t = 2, \ldots, N. \tag{6}$$

The residuals $\hat{e}_t$ are centered and rescaled by the factor $\sqrt{(N - k - p)/[N - 2(k + p)]}$ to counter possible deflation caused by the fitting (see Stine, 1987). We shall denote these centered and rescaled residuals by $\hat{e}_{s2}, \ldots, \hat{e}_{sN}$, where s denotes scaled. Let $\hat{F}_{N-1}$ denote the empirical cumulative distribution function (cdf) of $\hat{e}_{s2}, \ldots, \hat{e}_{sN}$; $\hat{F}_{N-1}$ has mass $1/(N - 1)$ at each $\hat{e}_{st}$. For general autoregressive k, $N - 1$ is replaced by $N - k$.

For each bootstrap replicate, we obtain a random sample, with replacement, of size N from $\hat{F}_N$, say, $\hat{e}_{s1}, \ldots, \hat{e}_{sN}$. We then form the series

$$y_t^* = \hat{\rho}_1 y_{t-1}^* - (\mathbf{x}_t - \mathbf{x}_{t-1})'\hat{\boldsymbol{\beta}}_1 + \hat{e}_{st}^*,\ t = 2, \ldots, N, \tag{7}$$

where y_1 is used to start the series. Note that the empirical distribution with cdf $\hat{F}_N$ has a mean of zero and constant variance and that the residuals $\hat{e}_{st}^*$ are selected randomly from $\hat{F}_N$. Hence, this bootstrap replicated series satisfies the assumptions of the Durbin estimation process.

Let N_B denote the number of replicated series (Equation 7), that is, the size of the bootstrap. For the ith replicate, let $\hat{\rho}_i^*$ and $\hat{\boldsymbol{\beta}}_i^*$ denote the Durbin two-stage estimates of ρ and $\boldsymbol{\beta}$. Note that these estimates are asymptotically unbiased for $\hat{\rho}_1$ and $\hat{\boldsymbol{\beta}}_1$.

The main purpose of the bootstrap is to gather information about the bias in the estimate of ρ. The bootstrap resamples from a population with autocorrelation $\hat{\rho}_1$. Hence, for each bootstrap replication $\hat{\rho}_i^* - \hat{\rho}_1$ is a random variable whose expectation approximates the true bias. As our measure of bias then, we take the average of these terms, that is,

$$\hat{b}_{\text{bias}}^{(1)} = \sum_{i=1}^{N_B} \frac{\hat{\rho}_i^* - \hat{\rho}_1}{N_B} = \frac{1}{N_B}\sum_{i=1}^{N_B}\hat{\rho}_i^* - \hat{\rho}_1.$$

This bootstrap procedure is similar to a method proposed by Freedman (1984) for bootstrapping two-stage LS estimates in stationary linear models. One of Freedman's results is that his bootstrap estimates are asymptotically equivalent to the LS estimates based on the transformations in Equations 4 and 5 using the true ρ. Although Freedman's main model is similar to our Equation 2, the independent variables in our model are lagged. For the subsequent discussion we refer to Freedman's model as a Type 1 model and to our Equation 2 as a Type 2 model; Durbin (1960) discussed both types. Durbin's first model is a Type 1 and his second model is a Type 2. Durbin showed that his two-stage estimation scheme for Type 2 models is asymptotically equivalent to LS estimation for Type I models. Estimates based on the two-stage scheme are asymptotically equivalent to optimal LS estimates based on the transformations in Equations 4 and 5 with known $\boldsymbol{\rho}$ under the assumption of normality. As we noted above, the bootstrap model satisfies the assumptions of the Durbin estimation process; hence, a similar equivalence can be made between the bootstrap for models of Type 2 and for those of Type 1. Estimates based on Freedman's bootstrap, though, are asymptotically equivalent to LS estimates of models of Type 1. Hence, it follows from Durbin's equivalence and Freedman's theory that the bootstrap estimates described have the same asymptotic properties as the LS estimates based on a known $\boldsymbol{\rho}$. This conclusion does not, however, speak to the small-sample properties of the bootstrap estimates.

We have discovered that one iteration of the bootstrap does not suffice to correct the bias of the estimates of $\boldsymbol{\rho}$. Therefore, we propose the following iterative procedure:

1. Start with the Durbin estimates $\hat{\rho}_1$ and $\hat{\boldsymbol{\beta}}_1$ of ρ and $\boldsymbol{\beta}$, respectively.
2. Use the bootstrap to calculate $\hat{b}^{(1)}_{\text{bias}}$. Let $\hat{\rho}_2 = \hat{\rho}_1 - \hat{b}^{(1)}_{\text{bias}}$, and use $\hat{\rho}_2$ in Durbin's second step to calculate $\hat{\boldsymbol{\beta}}_2$.
3. Perform the bootstrap again using $\hat{\rho}_2$ and $\hat{\boldsymbol{\beta}}_2$ in Equation 7 to obtain replications, as we did with $\hat{\rho}_1$ and $\hat{\boldsymbol{\beta}}_1$ in Equation 6. Calculate $\hat{b}^{(2)}_{\text{bias}}$ as
$$\sum_{i=1}^{N_B} \frac{\hat{\rho}_i^* - \hat{\rho}_2}{N_B}.$$
Let $\hat{\rho}_3 = \hat{\rho}_1 - \hat{b}^{(2)}_{\text{bias}}$, and use $\hat{\rho}_3$ to calculate the Durbin estimate $\hat{\boldsymbol{\beta}}_3$.
4. Repeat Step 3 up to m times or until $|\hat{\rho}_j - \hat{\rho}_{j-1}| < \epsilon$ where $j = 2, \ldots, m$ for some specified $\epsilon > 0$. We recommend setting ϵ at .01. We denote the bootstrap bias-adjusted estimate of ρ by $\hat{\rho}_F = \hat{\rho}_1 - \hat{b}^{(j-1)}_{\text{bias}}$ and the final Durbin estimate of $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_F$, where the subscript F stands for final. Because of the stationary assumption made earlier, if $\hat{\rho}_F \geq 1$ or $\hat{\rho}_F \leq -1$, we set $\hat{\rho}_F$ equal to .99 or $\hat{\rho}_F$ equal to $-.99$.

After obtaining the final estimates, a residual analysis should be performed on $\hat{e}_{Ft}$ to evaluate the adequacy of the fit:

$$\hat{e}_{Ft} = y_t - \hat{\rho}_F y_{t-1} - (\mathbf{x}_t - \hat{\rho}_F \mathbf{x}_{t-1})\hat{\boldsymbol{\beta}}_F, t = 2, \ldots, N. \quad (8)$$

Inference

After fitting the model and conducting a residual analysis, we proceed with inference on $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$. For relatively short series ($N < 100$), the bias-adjusted estimate of $\boldsymbol{\rho}$ may nevertheless lead to underestimation of the standard errors [$SE(\hat{\beta}_{Fk})$, where $k = 1, \ldots, p$] and thus to inflation of the test statistics. Thus, when the sample size is small we propose to produce $SE(\hat{\beta}_{Fk})$ by mimicking the variation in $\hat{\boldsymbol{\beta}}_F$ by using the Durbin estimates, $\hat{\boldsymbol{\beta}}^*$, obtained from N_B replicate series based on bootstrap methodology. Note that this is the second application of the bootstrap. That is, the fitting procedure described above is our first bootstrap for this problem. The inference we discuss next is based on our second bootstrap. As in the presentation of the first bootstrap, our discussion is for the case of the AR(1) model, that is, where $k = 1$.

To this end, N_B bootstrap replicated series are created using $\hat{\rho}_F$ and $\hat{\boldsymbol{\beta}}_F$ in Equation 7. That is, the bootstrap replicated series is

$$y_t^* = \hat{\rho}_F y_{t-1}^* - (\mathbf{x}_t - \mathbf{x}_{t-1})'\hat{\boldsymbol{\beta}}_F + \hat{e}_{st}^*, t = 2, \ldots, N. \quad (9)$$

We have found that using y_1 to start this series, as we did for the series in Equation 7, results in underestimation of the variance of the intercept. Instead, we recommend a starting value chosen at random from the original time series; this alleviates the underestimation of the variance of the intercept without changing the other regression parameters' bootstrap standard errors. Stine (1987) proceeded the same way in choosing a starting value for his bootstrap method for time series. The bootstrap errors $\hat{e}_{st}^*$ are randomly selected with replacement from the residuals in Equation 8. We calculate the Durbin estimates, $\hat{\boldsymbol{\beta}}_i^*$, for $i = 1, \ldots, N_B$. Because they are Durbin two-stage estimates, the random variables $\hat{\boldsymbol{\beta}}_i^*$ are consistent for $\hat{\boldsymbol{\beta}}_F$. These bootstrap estimates can then be used to estimate the variance–covariance matrix of $\hat{\boldsymbol{\beta}}_F$ in the usual way, that is

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}_F) = \frac{1}{N_B} \sum_{i=1}^{N_B} (\hat{\boldsymbol{\beta}}_i^* - \hat{\boldsymbol{\beta}}_F)(\hat{\boldsymbol{\beta}}_i^* - \hat{\boldsymbol{\beta}}_F)'. \quad (10)$$

There are other bootstrap estimates of this variance–covariance matrix. A common method is to use the average of the $\hat{\boldsymbol{\beta}}_i^*$ in place of $\hat{\boldsymbol{\beta}}_F$. In our empirical studies, we have found that using the average led to more liberal confidence procedures than using $\hat{\boldsymbol{\beta}}_F$. As our empirical results below show, the empirical coverages using Equation 10 are quite close to nominal confidence levels over the situations that we examined.

Inference for $\boldsymbol{\beta}$ can be based on the bootstrap variance–covariance matrix in Equation 10. But in our simulation studies, presented below, we have found that a standardized version motivated by the percentile t method of Hall (1988) has better properties. For the ith replicated model $i = 1, \ldots, N_B$ (Equation 9), let $\hat{\mathbf{e}}_{si}^*$ denote the vector of residuals. Denote the mean square error of these residuals by

$$MSE(\hat{\mathbf{e}}_{si}^*) = \frac{1}{N-1}\sum_{t=1}^{N-1}(\hat{e}_{sit}^* - \overline{\hat{e}_{si}^*})^2,\ i = 1, \ldots, N_B.$$

Let $MSE(\hat{\mathbf{e}}_F)$ denote the mean square error based on the residuals in Equation 8. Then, our modified estimate of the variance–covariance matrix of $\hat{\boldsymbol{\beta}}_F$ is $\hat{\mathbf{V}}_M$, which is defined as

$$\frac{MSE(\hat{\mathbf{e}}_F)}{N_B}\sum_{i=1}^{N_B}\frac{(\hat{\boldsymbol{\beta}}_i^* - \hat{\boldsymbol{\beta}}_F)(\hat{\boldsymbol{\beta}}_i^* - \hat{\boldsymbol{\beta}}_F)'}{MSE(\hat{\mathbf{e}}_{si}^*)}. \qquad (11)$$

Using Equation 11 as our variance–covariance estimator of $\hat{\boldsymbol{\beta}}_F$, tests of general linear hypotheses, H_0: $\mathbf{M}\boldsymbol{\beta} = \mathbf{0}$ versus H_A: $\mathbf{M}\boldsymbol{\beta} \neq \mathbf{0}$, for a specified matrix **M**, can be conducted. In particular, a test of H_0: $\beta_j = 0$ can be based on the t statistic

$$t_j = \frac{\hat{\beta}_{Fj}}{\sqrt{\hat{\mathbf{V}}_{Mjj}}}. \qquad (12)$$

As the simulation study in the subsequent section shows, comparing t_j with t-critical values based on $N - p$ degrees of freedom gives reasonable empirical levels. Likewise, confidence intervals for β_j can be formulated as

$$\hat{\beta}_{Fj} \pm t_{\alpha/2,n-p}\sqrt{\hat{\mathbf{V}}_{Mjj}}.$$

Turning to inference for the autoregressive parameters, we proceed as we did with $\boldsymbol{\beta}$. Note that the estimate of ρ based on a replicate series will be biased; hence, we can estimate this bias by running the four-step bootstrap procedure for each replicate series. Thus, approximate confidence intervals for ρ are constructed on the basis of the bootstrap bias-adjusted estimate $\hat{\rho}_F$ and a bootstrap standard error estimate, $s_{\hat{\rho}F}$. To calculate $s_{\hat{\rho}F}$, $\hat{\rho}_F$ and $\hat{\boldsymbol{\beta}}_F$ are used to construct N_B replicate series as described above. Then the bootstrap procedure (Steps 1–4) is used to get bias-adjusted estimates, $\hat{\rho}_{iF}$ for each replicate series, where $i = 1, \ldots, N_B$. As with standardizing the regression parameters (Equation 11), we have found it best to standardize each replicate by the mean square error of the replicate residuals. Hence, our estimate of the standard error of $\hat{\rho}_F$ is given by

$$s_{\hat{\rho}F} = \sqrt{MSE(\hat{\mathbf{e}}_F)}\left[\frac{1}{N_B}\sum_{i=1}^{N_B}\frac{(\hat{\rho}_{iF} - \hat{\rho}_F)^2}{MSE(\hat{\mathbf{e}}_{si}^*)}\right]^{1/2}.$$

The confidence interval for ρ is then

$$\hat{\rho}_F \pm t_{\alpha/2,n-p}s_{\hat{\rho}F}. \qquad (13)$$

Simulation results presented in the next section reveal that these confidence intervals provide good empirical coverage for ρ.

Monte Carlo Studies

We carried out two Monte Carlo studies to compare our bootstrap procedure with alternative approaches, using the two-phase interrupted time-series model discussed earlier (see Equation 1 for the form of the design matrix) in both studies. In our first study we compared the bootstrap with the Durbin estimation procedure and with a GLS procedure for which the values of ρ were assumed to be known. In the second study we compared the bootstrap with the maximum-likelihood ARIMA procedure. The particular implementation we chose was the SAS (SAS Institute, 1987) ARIMA procedure. We refer to this as the *ARIMA procedure* in the subsequent text. Methodological details for both studies are described below.

Study 1. The model $y_t = \mathbf{x}'\boldsymbol{\beta} + u_t$ (Equation 2), where $\mathbf{x}_t$ is the tth row of the design matrix given by Equation 1, defines the deterministic parameters of interest. The error term u_t was modeled as a first-order autoregressive process, $u_t = \rho u_{t-1} + e_t$, where the e_t represents the independent and identically distributed normal variates with $\mu = 0, \sigma^2 = 1$, and $|\rho| < 1.0$ (σ^2 = variance). For each specified value of ρ and series length N, 5,000 interrupted series were generated. The phase lengths were equal ($n_1 = n_2$). The e_ts were generated as independent normal deviates from a pair of independent uniform variates as proposed by Marsaglia and Bray (1964). The uniform variates were generated by the portable Fortran generator UNI developed by Kahner, Moler, and Nash (1988). The first simulated autoregressive series $\{u_t\}_i$ was started after the generation of 300 variates of the form $v_j = \rho v_{j-1} + e_j$, where $j = 1, \ldots, 300$. Letting $u_0 = v_{300}$, the first simulated autoregressive series $\{u_t\}_i$ was generated as $u_i = \rho u_{i-1} + e_i$, where $i = 1, \ldots, n$. Then, this entire process was repeated for each of the 5,000 series. Hence, between each generated series 300 ignored variates were generated. Thus, each simulated series began with a random normal deviate with μ = zero, $\sigma^2 = 1/(1 - \rho^2)$, and there was assurance that the last observation of one simulated series was virtually uncorrelated with the first observation of the next simulated series.

The centered model was used in the second stage of Durbin's estimation procedure. The centered observations and the centered design matrix are $y_{ct} = y_t - \bar{y}$ and $\mathbf{x}'_{ct} = (x_{t,1} - \bar{x}_1, x_{t,2} - \bar{x}_2, x_{t,3} - \bar{x}_3)$, respectively. The Phase 1 intercept parameter, β_0, was then estimated by $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}_1 - \hat{\beta}_2\bar{x}_2 - \hat{\beta}_3\bar{x}_3$.

Study 2. The second study involved a comparison of the properties of our bootstrap procedure with those of the ARIMA procedure. We used the latter to obtain maximum-likelihood estimates of the parameters of Equation 2 with Equation 1 under both the null model (i.e., $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 0$) and the alternative model (i.e., $\beta_0 = \beta_1 = 0, \beta_2 = 1.0, \beta_3 = .5$) in order to evaluate Type I error and power, respectively. Each simulated series had 15 preintervention observations (n_1) and 15 postintervention observations (n_2). The simulation size was 5,000. The values of the autoregressive parameter ρ associated with the error terms of the simulated series (defined as $u_t = \rho u_{t-1} + e_t$) were randomly chosen from the interval (0, 1); the normal deviates e_t were generated from the SAS function NORMAL to have zero mean and unit variance. The SAS model, ARIMA, was fit using the following SAS code:

```
identify var = y crosscor = (x1 x2 x3 x4);
estimate p = 1 input = (x1 x2 x3 x4) method =
ML noconstant;
```

where x1–x4 denote the columns of the design matrix for the two-phase intervention design.

Results

Study 1

Table 2 shows the results of Study 1 with respect to the confidence interval coverages of ρ based on the bootstrap confidence interval in Equation 13. Because this is a nested bootstrap procedure (i.e., a bootstrap within a bootstrap) and therefore is extremely computer intensive, the results in Table 2 are based on only 1,000 simulations for $N = 30$. The empirical coverages are quite close to the nominal confidence coefficients for both the 90% and 95% intervals.

Table 2
Empirical Confidence Interval Coverages of the Bootstrap Confidence Interval in Equation 13 for ρ

ρ	Nominal confidence 90%	95%
.90	.903	.940
.80	.926	.954
.70	.941	.967
.50	.890	.957
.30	.911	.948
.20	.901	.960
.10	.913	.948
.00	.926	.962
−.30	.931	.966

Note. Confidence interval coverages are based on 1,000 simulations of $N = 30$.

Tables 3 and 4 display the empirical means and variances of the initial Durbin estimator of ρ ($\hat{\rho}_1$) and the bootstrap bias-adjusted estimator ($\hat{\rho}_F$) for series of $N = 20$, 30, 50, and 100. As shown in these tables, the bootstrap bias-adjusted estimates are considerably less biased than the initial estimates of ρ, and for $N = 100$ they are nearly unbiased.

Figures 2 through 5 present, for $N = 20$, 30, 50, and 100, empirical results of the hypothesis tests H_0: $\beta_k = 0$, where $k = 0, \ldots, p$, at nominal $\alpha = .05$, using our bootstrap procedure. The true β_j here are all zero and hence, Figures 2 through 5 give the empirical Type I errors. The standardized test statistics of Equation 12, t_j, were used for these results. In all of these figures we see that Type I error rates for β_2 are approximately nominal. The Type I error rates for the hypothesis tests for the parameters β_0, β_1, and β_3 are close to nominal, with some inflation when ρ is near .90.

We should compare power only for procedures whose Type I error rates are close to nominal values. Thus, we do not use the Prais–Winsten GLS procedure based on the Durbin estimate of ρ. Instead we choose the Prais–Winsten GLS procedure based on the true value of ρ in the transformation. We denote this procedure by TV, for true value, so as not to confuse it with the Prais–Winsten GLS procedure

Table 3
Empirical Estimates and Variances for the Durbin Estimator $\hat{\rho}_1$ and the Bootstrap Bias-Adjusted Estimator $\hat{\rho}_F$ for N = 20 and N = 30

	N = 20				N = 30			
ρ	$\hat{\rho}_1$	Var.	$\hat{\rho}_F$	Var.	$\hat{\rho}_1$	Var.	$\hat{\rho}_F$	Var.
.90	.239	.068	.696	.118	.456	.040	.769	.057
.80	.221	.066	.677	.126	.410	.038	.747	.064
.70	.180	.066	.630	.139	.361	.041	.689	.075
.60	.131	.062	.564	.142	.294	.038	.600	.075
.50	.083	.060	.495	.147	.229	.038	.511	.077
.40	.030	.059	.414	.148	.165	.038	.422	.074
.30	−.030	.056	.321	.145	.087	.035	.315	.066
.20	−.090	.055	.223	.140	.015	.034	.219	.061
.10	−.159	.051	.117	.127	−.065	.035	.112	.060
.00	−.223	.049	.018	.119	−.147	.032	.007	.054
−.10	−.287	.047	−.081	.107	−.217	.030	−.084	.050
−.20	−.357	.044	−.184	.097	−.302	.030	−.192	.049
−.30	−.431	.041	−.292	.086	−.380	.026	−.291	.045
−.40	−.498	.038	−.389	.077	−.462	.026	−.394	.042
−.50	−.563	.036	−.479	.070	−.536	.023	−.485	.036
−.60	−.636	.031	−.578	.057	−.620	.020	−.588	.030
−.70	−.716	.026	−.682	.043	−.706	.017	−.688	.023
−.80	−.789	.021	−.771	.021	−.784	.014	−.776	.017
−.90	−.863	.014	−.856	.019	−.869	.009	−.867	.010

Note. Estimates are based on 5,000 simulations. Var. = variance.

Table 4
Empirical Estimates and Variances for the Durbin Estimator $\hat{\rho}_1$ and the Bootstrap Bias-Adjusted Estimator $\hat{\rho}_F$ for N = 50 and N = 100

	N = 50				N = 100			
ρ	$\hat{\rho}_1$	Var.	$\hat{\rho}_F$	Var.	$\hat{\rho}_1$	Var.	$\hat{\rho}_F$	Var.
.90	.637	.018	.860	.023	.780	.006	.898	.007
.80	.558	.018	.797	.029	.696	.007	.804	.009
.70	.506	.019	.709	.031	.610	.007	.707	.009
.60	.426	.019	.609	.031	.519	.008	.607	.010
.50	.344	.021	.508	.030	.427	.009	.508	.010
.40	.262	.021	.410	.029	.333	.009	..407	.011
.30	.175	.021	.308	.029	.240	.011	.308	.011
.20	.090	.020	.210	.028	.148	.010	.209	.012
.10	.002	.021	.108	.028	.055	.010	.110	.012
.00	−.085	.020	.007	.027	−.041	.010	.007	.012
−.10	−.169	.018	−.089	.026	−.134	.010	−.093	.012
−.20	−.259	.019	−.192	.025	−.228	.009	−.194	.011
−.30	−.343	.017	−.289	.023	−.322	.009	−.296	.011
−.40	−.435	.016	−.395	.021	−.416	.008	−.396	.009
−.50	−.522	.014	−.493	.019	−.510	.007	−.497	.008
−.60	−.612	.013	−.594	.016	−.603	.006	−.595	.007
−.70	−.699	.010	−.690	.012	−.698	.005	−.694	.006
−.80	−.789	.008	−.786	.009	−.791	.004	−.790	.004
−.90	−.879	.005	−.878	.005	−.887	.002	−.886	.002

Note. Estimates are based on 5,000 simulations. Var. = variance.

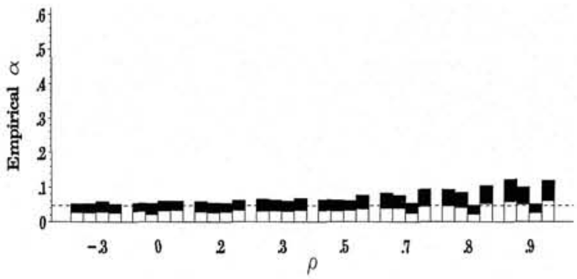

Figure 2. Bootstrap empirical Type I errors. Empirical αs are based on 5,000 simulations of $N = 20$. The four parameters plotted at each level of ρ are, in order, as follows: intercept (β_0), first-phase slope (β_1), level change (β_2), and slope change (β_3). Open boxes correspond to left-tailed tests and filled boxes to right-tailed tests. The horizontal line corresponds to a nominal α of .05.

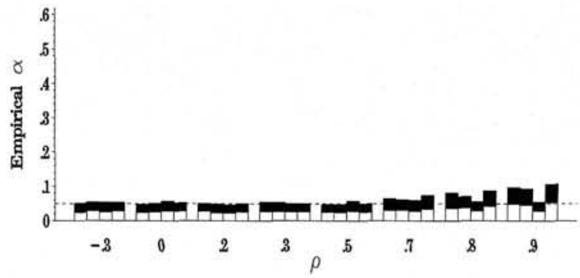

Figure 4. Bootstrap empirical Type I errors. Empirical αs are based on 5,000 simulations of $N = 50$. The four parameters plotted at each level of ρ are, in order, as follows: intercept (β_0), first-phase slope (β_1), level change (β_2), and slope change (β_3). Open boxes correspond to left-tailed tests and filled boxes to right-tailed tests. The horizontal line corresponds to a nominal α of .05.

based on the Durbin estimate of ρ. The TV procedure, being based on the true value of a parameter, is of course not a statistical procedure, but it does provide a benchmark comparison of power with our bootstrap procedure.

Figures 6 and 7 compare the simulated power results of the bootstrap method with that of the TV procedure. Power results are based on the hypothesis H_0: $\beta_2 = 0$, when in fact the true $\beta_2 = 1$. Results are shown for $N = 30$ and 50, using nominal $\alpha = .05$. We observe from Figures 6 and 7 that the bootstrap procedure has comparable power to the results of the TV procedure.

Study 2

Table 5 compares the empirical means of the bootstrap and ARIMA procedures. For both procedures the empirical means of the regression coefficients are quite close to each other and quite close to their true values. Note, however, that the procedures differ on the estimate of the autoregression coefficient, ρ. As the empirical means show, the bootstrap's estimate of ρ is much closer to the actual value than is the estimate based on the ARIMA procedure. Figures 8 and 9 present the Type I error rates for tests of the regression model parameters using the bootstrap method and output produced from ARIMA. The Type I error rate for the bootstrap procedure is much closer to the nominal .05 level than that for the ARIMA procedure. Notice that extremely liberal values are obtained using the ARIMA approach.

Figure 10 displays the empirical power of the bootstrap procedure for the test on β_3. Power results for the ARIMA procedure are not reported because of its

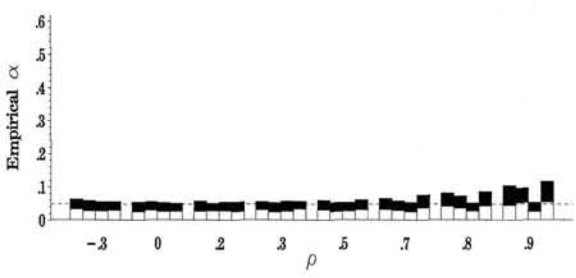

Figure 3. Bootstrap empirical Type I errors. Empirical αs are based on 5,000 simulations of $N = 30$. The four parameters plotted at each level of ρ are, in order, as follows: intercept (β_0), first-phase slope (β_1), level change (β_2), and slope change (β_3). Open boxes correspond to left-tailed tests and filled boxes to right-tailed tests. The horizontal line corresponds to a nominal α of .05.

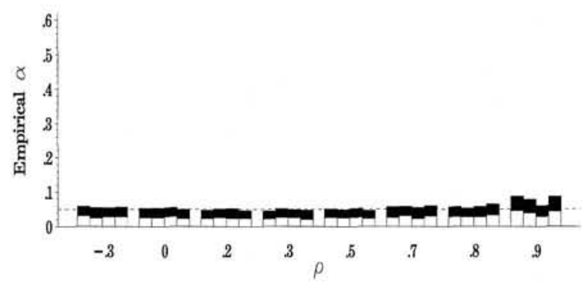

Figure 5. Bootstrap empirical Type I errors. Empirical αs are based on 5,000 simulations of $N = 100$. The four parameters plotted at each level of ρ are, in order, as follows: intercept (β_0), first-phase slope (β_1), level change (β_2), and slope change (β_3). Open boxes correspond to left-tailed tests and filled boxes to right-tailed tests. The horizontal line corresponds to a nominal α of .05.

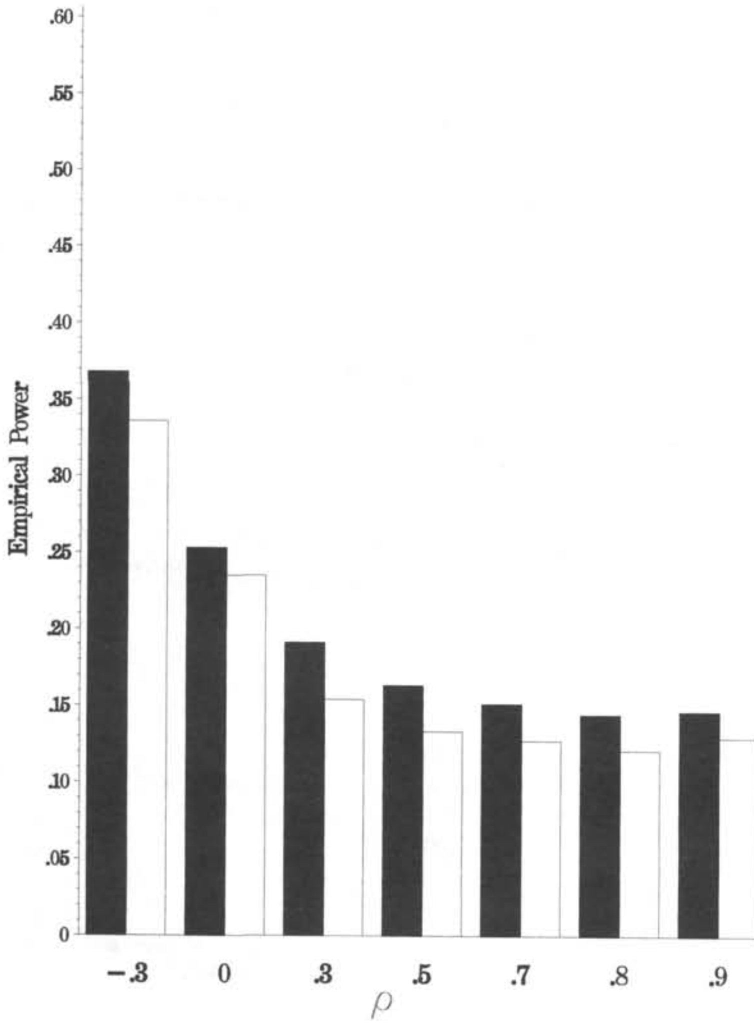

Figure 6. Empirical power comparisons, based on 5,000 simulations of $N = 30$. Nominal $\alpha = .05$. Filled boxes correspond to the Prais–Winsten generalized least squares method based on the true value of ρ, and open boxes correspond to the bootstrap method.

poor Type I behavior in the above study. Table 6 shows the empirical means of the bootstrap estimates of the regression coefficients and the autoregressive parameters. Once again the estimates are close to their actual values.

Example

An example of the application of the new double bootstrap method to a published study is presented below. We first show that there are conditions under which the outcome of ordinary least squares (OLS), various versions of GLS, ARIMA (1, 0, 0), and the new method are very similar. Next, we show that the new method can provide results that differ considerably from those provided by conventional solutions when autocorrelation of the errors appears to be high. In both instances the results are consistent with simulation results and statistical theory cited earlier in this article.

Analysis of Original Data

Dyer, Schwartz, and Luce (1984) investigated the potential intervention effects of staff training on the behavior of severely handicapped students. A portion of the data they reported can be viewed as having been collected under a two-phase time-series design. We used the four-parameter design matrix presented earlier in Equation 1 with the following methods: (a) OLS, (b) Cochrane–Orcutt, (c) Prais–Winsten, (d)

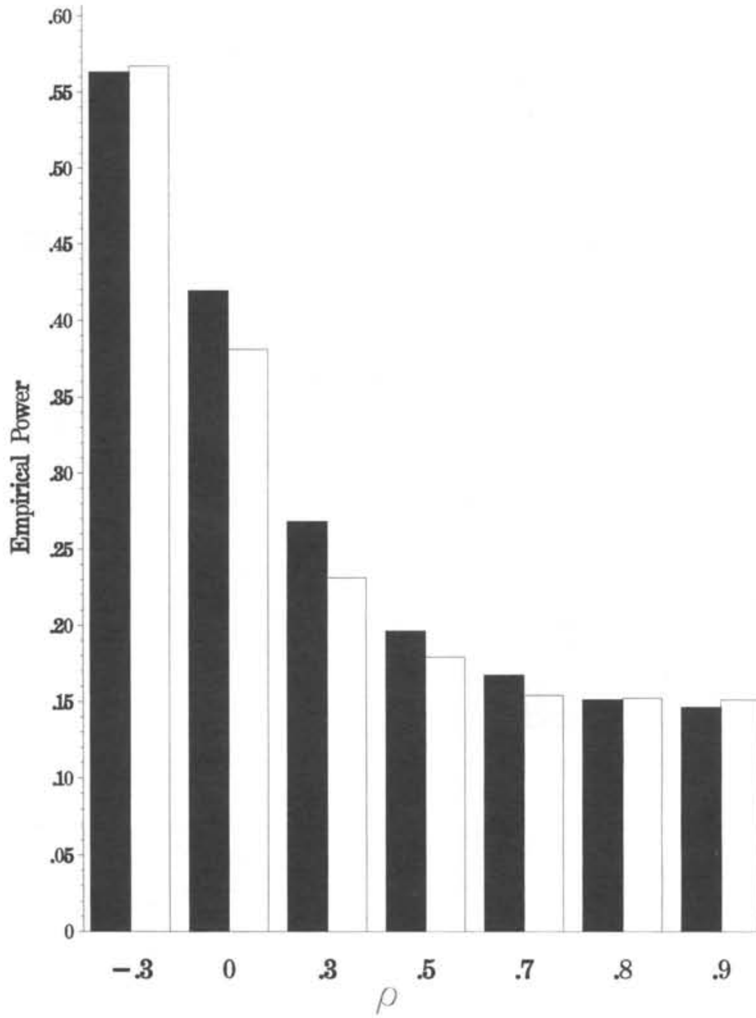

Figure 7. Empirical power comparisons, based on 5,000 simulations of $N = 50$. Nominal $\alpha = .05$. Filled boxes correspond to the Prais–Winsten generalized least squares method based on the true value of ρ, and open boxes correspond to the bootstrap method.

maximum-likelihood, (e) ARIMA (1, 0, 0), and (f) the new double bootstrap method.

The results are presented in the top of Table 7, which shows that the estimates of the four regression coefficients as well as the *t* values associated with these coefficients are very similar. The estimates of autocorrelation among the errors are almost identical using the various analyses.

Because parameters β_1 and β_3 in the tentative four-parameter model do not appear to be necessary (note the *t* values), we estimated a more parsimonious model containing only two parameters (viz., the intercept β_0 and the level-change coefficient β_2). Once again, we used all methods of estimation mentioned above; the results are presented in the bottom of Table 7. The parameter estimates are almost identical for all methods, and the *t* statistics are quite similar. We conclude from all methods that there is an intervention effect of about nine points and that it is not reasonable to interpret this difference as attributable to sampling error.

Analysis of Autocorrelation-Contaminated Data

To demonstrate the properties of the new method in the context of autocorrelated data, the original Dyer et al. (1984) data (analyzed above) were contaminated by adding a high level of autocorrelation ($\rho_1 = .80$) to the residuals of the OLS fit. An adequate analysis of these contaminated data should recover most of the

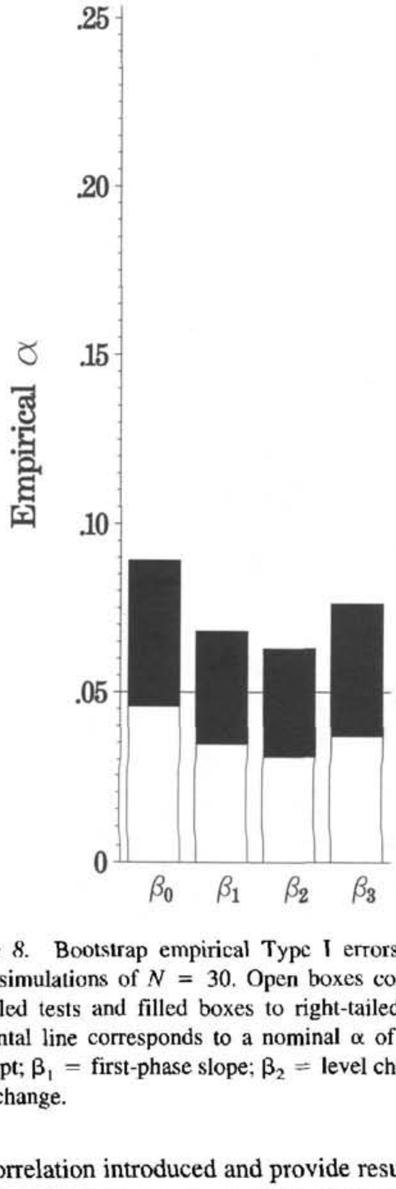

Figure 8. Bootstrap empirical Type I errors, based on 5,000 simulations of $N = 30$. Open boxes correspond to left-tailed tests and filled boxes to right-tailed tests. The horizontal line corresponds to a nominal α of .05. β_0 = intercept; β_1 = first-phase slope; β_2 = level change; β_3 = slope change.

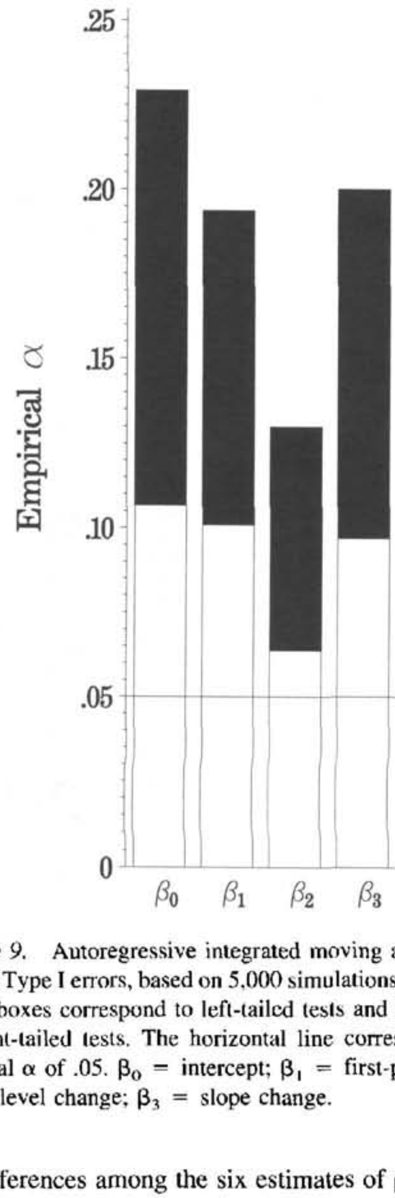

Figure 9. Autoregressive integrated moving average empirical Type I errors, based on 5,000 simulations of $N = 30$. Open boxes correspond to left-tailed tests and filled boxes to right-tailed tests. The horizontal line corresponds to a nominal α of .05. β_0 = intercept; β_1 = first-phase slope; β_2 = level change; β_3 = slope change.

autocorrelation introduced and provide results similar to those obtained on the uncontaminated data. The results obtained using the six methods described above can be seen in Table 8.

Table 5
Empirical Means of Parameter Estimates

Parameter	Actual	Bootstrap	ARIMA
ρ	.498	.491	.221
β_0	.000	.020	.026
β_1	.000	−.002	−.003
β_2	.000	.002	.003
β_3	.000	.004	.004

Note. Estimates are based on 5,000 simulations of $N = 30$. ARIMA = autoregressive integrated moving average.

Differences among the six estimates of ρ_1 are substantial. Notice that the bootstrap estimate of ρ_1 is .74, whereas the values produced by the other methods are much lower. The autocorrelation coefficient along with the test of significance on this coefficient (not shown) indicates that (a) the errors are highly autocorrelated, (b) OLS is inappropriate in this application, and (c) correction for autocorrelated errors should be undertaken.

Notice that the values for regression coefficients β_1 and β_3 (first-phase slope and slope change, respectively) and the corresponding t values associated with the bootstrap differ greatly from those produced by the other methods. One outcome of the bootstrap analysis is that neither β_1 nor β_3 are needed in the

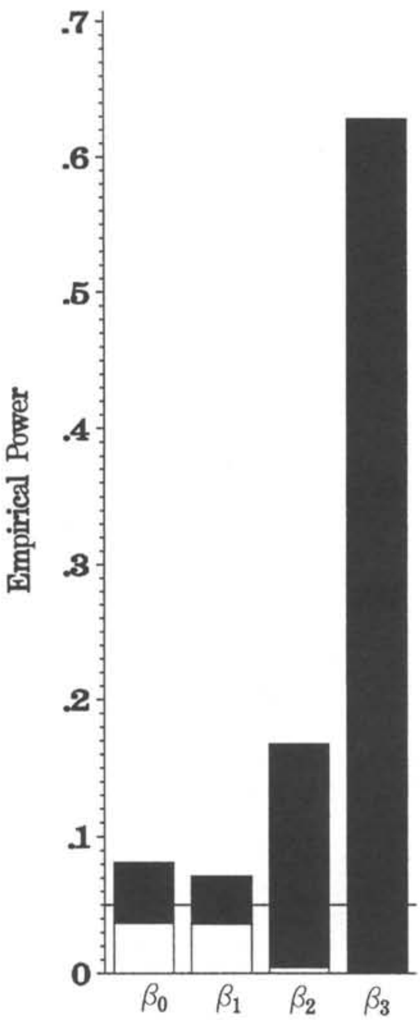

Figure 10. Bootstrap empirical power results, based on 5,000 simulations of $N = 30$. Open boxes correspond to left-tailed tests and filled boxes to right-tailed tests. The horizontal line corresponds to a nominal α of .05. β_0 = intercept; β_1 = first-phase slope; β_2 = level change; β_3 = slope change.

model; most of the other methods lead to the conclusion that these parameters should be retained in the model. Recall from the analysis of the original data (Table 7) that β_1 and β_3 are not necessary in the model. Hence, the model identified by the bootstrap is consistent with the analysis of the original data, whereas the conventional procedures argue for a more complex (and incorrect) model. Note also in the bottom of Table 8 that the level-change estimate (9.12) provided by the bootstrap analysis of the contaminated data is consistent with the level-change estimate

Table 6

Empirical Means of the Bootstrap Parameter Estimates for the Power Study

Parameter	Actual	Bootstrap
ρ	.487	.479
β_0	.000	−.016
β_1	.000	−.001
β_2	1.000	.999
β_3	.500	.501

Note. Estimates based on 5,000 simulations of $N = 30$.

(9.05) provided by the analysis of the original noncontaminated data.

Discussion

We have presented a theoretically based bootstrap procedure to analyze linear models with autoregressive error terms. Our focus was on a particular two-phase intervention model with first-order autoregressive error terms. However, the theoretical framework is general and can be applied to other linear models with higher order autoregressive error terms. Simulation results show the superiority of the bootstrap over other popular methods (including maximum-likelihood procedures) in controlling Type I errors. In

Table 7

Parameter Estimates and t Values Associated With Six Estimation Methods Applied to Linearly Transformed Time-Series Data From Dyer, Schwartz, and Luce (1984)

	Method					
Coef.	BS	OLS	C-O	P-W	ML	ARIMA
	Four-parameter design matrix					
$\hat{\rho}_1$	−.24	−.25	−.25	−.25	−.25	−.25
$\hat{\beta}_0$	4.28	4.40	4.26	4.35	4.35	4.35
$\hat{\beta}_1$	0.15	0.14	0.15	0.14	0.14	0.14
$\hat{\beta}_2$	7.31	7.39	7.31	7.36	7.36	7.36
$\hat{\beta}_3$	−0.12	−0.10	−0.12	−0.11	−0.11	−0.11
t_0	3.69	3.44	3.78	4.22	4.22	4.22
t_1	1.27	1.03	1.31	1.33	1.33	1.33
t_2	5.61	4.80	5.71	5.90	5.90	5.90
t_3	−0.94	−0.72	−0.97	−0.96	−0.96	−0.96
	Two-parameter design matrix					
$\hat{\rho}_1$	−.21	−.21	−.21	−.21	−.21	−.21
$\hat{\beta}_0$	5.58	5.56	5.58	5.56	5.56	5.56
$\hat{\beta}_2$	9.05	9.08	9.05	9.08	9.08	9.08
t_0	10.58	9.17	10.84	11.14	11.14	11.14
t_2	14.92	12.16	14.35	14.76	14.76	14.76

Note. Coef. = coefficient; BS = double bootstrap; OLS = ordinary least squares; C-O = Cochrane–Orcutt; P-W = Prais–Winsten; ML = maximum-likelihood; ARIMA = autoregressive integrated moving average.

Table 8
Parameter Estimates and t Values Associated With Six Methods Applied to Time-Series Data From Dyer, Schwartz, and Luce (1984) That Have Been Contaminated by Adding Autocorrelation (ρ_1 = .80) to the Residuals

	Method					
Coef.	BS	OLS	C-O	P-W	ML	ARIMA
Four-parameter design matrix						
$\hat{\rho}_1$	.74	.46	.46	.46	.51	.51
$\hat{\beta}_0$	6.65	4.05	5.25	4.20	4.22	4.22
$\hat{\beta}_1$	0.03	0.32	0.19	0.28	0.27	0.27
$\hat{\beta}_2$	9.33	7.15	8.30	7.93	8.08	8.08
$\hat{\beta}_3$	−0.19	−0.49	−0.36	−0.45	−0.45	−0.45
t_0	2.03	5.08	3.25	3.64	3.44	3.44
t_1	0.11	3.91	1.32	2.46	2.27	2.27
t_2	6.06	7.44	6.38	6.41	6.33	6.33
t_3	−0.60	−5.60	−2.27	−3.52	−3.26	−3.26
Bootstrap analysis based on two-parameter design matrix						
$\hat{\rho}_1$	.86					
$\hat{\beta}_0$	5.63					
$\hat{\beta}_2$	9.12					
t_0	3.30					
t_2	6.61					

Note. Coef. = coefficient; BS = double bootstrap; OLS = ordinary least squares; C-O = Cochrane–Orcutt; P-W = Prais–Winsten; ML = maximum-likelihood; ARIMA = autoregressive integrated moving average.

contrast, traditional GLS and maximum-likelihood methods yield liberal Type I error rates. Our bootstrap procedure thereby offers a more exact method for testing hypotheses about the model parameters. Simulation results also show that the bootstrap procedure is nearly as powerful as the Prais–Winsten GLS procedure when the latter is provided (unrealistically) with the known (rather than an estimated) value of ρ in the TV procedure. It seems that there are no competing procedures that are as satisfactory in the small-sample case as the proposed bootstrap method.

The simulation results show that the bootstrap bias-adjusted estimates of the autoregressive parameters are much less biased than the original (Durbin) estimates. It can be concluded that the bootstrap autocorrelation estimation method is also far less biased than all competing estimation methods that are mathematically similar to the Durbin procedure. These methods include the popular Prais–Winston, Cochrane–Orcutt, and all related approaches. The bootstrap confidence intervals for these autoregressive parameters show good coverage properties over the situations studied.

The example discussed is illustrative of the theoretical discussion of our bootstrap procedure and our simulation study. The data were drawn from a two-phase study that we modeled using the four-parameter design matrix presented earlier. On the original data all the methods (OLS, various versions of GLS, ARIMA [1, 0, 0], and our double bootstrap procedure) are very similar. All the methods indicated little autocorrelation effect and that the two-slope parameters of the four-parameter design were not necessary. We then altered the data by inserting an autocorrelation effect into the errors. The double bootstrap analysis of these contaminated data recovered most of the autocorrelation introduced and provided results similar to those obtained on the uncontaminated data. The other procedures, though, had severely biased estimates of the autocorrelation parameter. As our simulation results predicted, this caused the other analyses to have deflated standard errors of the regression coefficients, leading to the finding that the slope effects are significant. Hence, the model identified by the bootstrap is consistent with the analysis of the original data, whereas the conventional procedures argued for a more complex (and incorrect) model.

Future work is needed to explore the usefulness of the bootstrap in analyzing more complicated models with higher order autoregressive and moving average error terms. We also intend to explore other bootstrap procedures and jackknife procedures (see Wu, 1986) for these types of problems. Preliminary simulation results for our double bootstrap (McKnight, 1994) based on a two-phase intervention model with second-order autoregressive error terms indicate that our bootstrap can be successfully applied to higher order autoregressive models. Because first- and second-order autoregressive models adequately capture the error structure for a large proportion of time-series regression models, the new procedure has many potential applications.

References

Beesley, P. A., Doran, H. E., & Griffiths, W. E. (1987). *Further attempts at improving the finite sample accuracy of interval estimation in a linear model with first-order autoregressive error and trended explanatory variable.* Unpublished manuscript, University of New England, Armidale, Australia.

Berry, W. D., & Lewis-Beck, M. S. (1986). Interrupted time series. In W. D. Berry & M. S. Lewis-Beck (Eds.), *New tools for social scientists: Advances and applications in research methods* (pp. 209–240). Newbury Park, CA: Sage.

Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. 1994. *Time* *series analysis, forecasting and control.* San Francisco: Holden-Day.

Cochrane, D., & Orcutt, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *Journal of the American Statistical Association, 44,* 32–61.

Cook, T. D., & Shadish, W. R. (1994). Social experiments: Some developments over the past fifteen years. *Annual Review of Psychology, 45,* 545–580.

Durbin, J. (1960). Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society: Series B (Methodological), 22,* 139–153.

Dyer, K., Schwartz, I. S., & Luce, S. C. (1984). A supervision program for increasing functional activities for severely handicapped students in a residential setting. *Journal of Applied Behavior Analysis, 17,* 249–259.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap.* New York: Chapman and Hall.

Freedman, D. (1984). On bootstrapping two-stage least squares estimates in regression problems. *Annals of Statistics, 12,* 827–842.

Fuller, W. A. (1996). *Introduction to statistical time series* (2nd ed.), New York: Wiley.

Glass, G. V. (1997). Interrupted time-series quasi-experiments. In R. M. Jaeger (Ed.), *Methods for research in education* (2nd ed., pp. 589–609). Washington, DC: American Educational Research Association.

Griliches, Z., & Rao, P. (1969). Small-sample properties of several two-stage regression methods in the context of auto-correlated errors. *Journal of the American Statistical Association, 64,* 253–272.

Hall, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Annals of Statistics, 16,* 927–953.

Huitema, B. E., & McKean, J. W. (1998). Irrelevant autocorrelation in least-squares intervention models. *Psychological Methods, 3,* 1–12.

Johnston, J. (1984). *Econometric methods* (3rd ed). New York: McGraw-Hill.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., & Lee, T. C. (1988). *The theory and practice of econometrics* (2nd ed.). New York: Wiley.

Kahner, D., Moler, C., & Nash, S. (1988). *Numerical methods and software.* Englewood Cliffs, NJ: Prentice Hall.

Marcantonio, R. J., & Cook, T. D. (1994). Convincing quasi-experiments: The interrupted time-series and regression-discontinuity designs. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 133–154). San Francisco: Jossey-Bass.

Marsaglia, G., & Bray, T. A. (1964). A convenient method for generating normal variables. *SIAM Review, 6,* 260–264.

McKnight, S. (1994). A bootstrap method to analyze an intervention model with autoregressive error terms (Doctoral dissertation, Western Michigan University, 1995). *Dissertation Abstracts International, 55,* 5416.

Mohr, L. B. (1995). *Impact analysis for program evaluation* (2nd ed.). Thousand Oaks, CA: Sage.

Nankervis, J. C., & Savin, N. E. (1996). The level and the power of the bootstrap t test in the AR(1) model with trend. *Journal of Business & Economic Statistics, 14,* 161–168.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). *Applied linear regression models* (3rd ed.). Homewood, IL: Irwin.

Prais, S. J., & Winsten, C. B. (1954). *Trend estimators and serial correlation.* Unpublished discussion paper for the Cowles Commission, Chicago (Stat. No. 383). (Available from the Department of Business, University of Chicago, Chicago, IL 60637)

Rushe, R. H., & Gottman, J. M. (1993). Essentials in the design and analysis of time-series experiments. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Statistical issues* (pp. 493–528). Hillsdale, NJ: Erlbaum.

SAS Institute. (1987). *SAS/graph guide for personal computers* (Version 6). Cary, NC: Author.

Stine, R. (1987). Estimating properties of autoregressive forecasts. *Journal of the American Statistical Association, 82,* 1072–1078.

Sun, H., & Pantula, S. G. (1996). *Testing for trends in correlated data.* Unpublished manuscript.

Woodward, W. A., Bottone, S., & Gray, H. L. (1997). Improved tests for trend in time series data. *Journal of Agricultural, Biological, and Environmental Statistics, 2,* 403–416.

Wu, C. F. J. (1986). Jackknife, bootstrap, and other resampling methods in regression analysis. *Annals of Statistics, 14,* 1261–1294.

Received March 16, 1998
Revision received October 19, 1999
Accepted October 19, 1999 ■