# ESTIMATION AFTER PRE-TESTING IN LEAST ABSOLUTE VALUE REGRESSION WITH AUTOCORRELATED ERRORS

Terry E. Dielman*
Elizabeth L. Rose**

*We study least absolute value (LAV) estimation and inference in the context of simple time series regression when the disturbances are autocorrelated. Several different estimation techniques are compared: uncorrected LAV, LAV after a Cochrane-Orcutt (CO)-type transformation to correct for autocorrelation, LAV after a Prais-Winsten (PW)-type transformation to correct for autocorrelation, and two pre-test estimators that transform (by CO and by PW, respectively) when a pre-test suggests that autocorrelation is present. Monte Carlo simulation methods are used to compare the small-sample performances of the different estimators. The Prais-Winsten approach to correction for autocorrelation is preferable to the Cochrane-Orcutt approach, and there appears to be minimal loss associated with always correcting.*

The most commonly-employed approach to estimating the parameters of regression models is that of ordinary least squares (OLS). The OLS approach provides estimates that are unbiased and have minimum variance, as long as the errors are independent and identically distributed normal random variables. However, when the distribution of the errors is nonnormal and subject to outlying values, OLS estimators may perform

* *Terry E. Dielman is a Professor of Decision Sciences at Texas Christian University, Fort Worth, Texas.*

** *Elizabeth L. Rose is an Assistant Professor of Information and Operations Management in the School of Business at University of Southern California, Los Angeles, CA.*

quite poorly. In that situation, the use of an estimation technique that is robust to such violations of the assumptions of OLS is advisable. Least absolute value (LAV) regression is one of the most commonly-employed robust regression techniques.

The use of regression models for time series data often involves the violation of another assumption crucial to the optimality of OLS. With time series applications, the error terms often are not independent, displaying a correlation structure through time. This autocorrelation can arise for many reasons, including the existence of random shocks whose effects extend for more than one period. For example, a natural disaster, such as an earthquake or hurricane, can be expected to affect the local economy (e.g., construction activity) for months or years. On a more positive note, a corporation hopes that the effect of a large advertising campaign on its sales is not a single-period phenomenon, but will endure for some time. In this paper, we investigate the use of LAV estimation approaches in the presence of the first-order autocorrelation of the disturbances that is so frequently an issue when dealing with time series regression.

The problem of first-order autocorrelation has been investigated extensively in the context of OLS (see e.g., Dielman & Pfaffenberger [1989]), and some work has also been done in the context of least absolute value (LAV) regression. Weiss (1990) investigated the asymptotic properties of the LAV estimator when the disturbances are autocorrelated, and suggested a correction for first-order autocorrelation. Weiss' method is analogous to the approach proposed by Cochrane & Orcutt (1949), which is used frequently in least squares regression. Dielman & Rose (1994a) examined the performances of several estimators in the regression model with first-order autocorrelation of the disturbances, including uncorrected OLS and LAV, and OLS- and LAV-based versions of the Cochrane-Orcutt and Prais-Winsten (1954) corrections for first-order serial correlation.

**"The use of regression models for time series data often involves the violation of another assumption crucial to the optimality of OLS. With time series applications, the error terms often are not independent, displaying a correlation structure through time."**

The results obtained by both Weiss (1990) and Dielman & Rose (1994a, 1994b) suggest that correction for autocorrelation is important for moderate to large values of the autocorrelation coefficient, whether estimation is by least squares or least absolute value. For LAV estimation, Dielman & Rose (1994a) found that the efficiency of the uncorrected estimator is considerably worse than that of the corrected estimator, particularly as the severity of the autocorrelation increases. In another study, Dielman & Rose (1994b) observed that the out-of-sample forecasting performance of the models suffers when the autocorrelation correction is not employed. As with many problems of mismatching between the data and the model assumptions, the problems associated with uncorrected autocorrelation tend to be magnified when the sample sizes are small. Dielman & Rose (1994a) found that the inclusion of the first observation in the correction process is important for LAV estimation, particularly for small-sample situations. This result has also been observed with least squares-based methods, and leads to a preference for Prais-Winsten (PW)-type correction over Cochrane-Orcutt (CO)-type correction.

Previous studies have examined the use of a pre-test procedure when the estimation is conducted using least squares-based techniques; see, for example, King & Giles (1984) and Griffiths & Beesley (1984). In this research, we extend the literature by considering the use of a pre-test procedure with LAV regression. Five estimation techniques are compared: LAV estimation with no correction, LAV estimation with CO-type and PW-type transformations to correct for autocorrelation, and two pre-test estimators that transform (by CO in one case and by PW in the other) only if a pre-test suggests that autocorrelation is present. Monte Carlo simulation methods are used to compare the small-sample properties of the estimators and the observed significance levels of the tests.

## MODEL, ESTIMATION, AND INFERENCE

### Model

We consider a simple regression model of the form

$$\begin{aligned} y_t &= \beta_0 + \beta_1 x_t + \epsilon_t \\ \epsilon_t &= \rho \epsilon_{t-1} + \eta_t, \end{aligned} \tag{1}$$

for t=1,2,...,T. In (1), $y_t$ and $x_t$ are the t$^{th}$ observations on the dependent and explanatory variables, respectively, and $\varepsilon_t$ is a random disturbance for the t$^{th}$ observation. The $\eta_t$, are assumed to be independent and identically distributed, but not necessarily normal. The parameters $\beta_0$ and $\beta_1$ are unknown and must be estimated. The parameter $\rho$ is the autocorrelation coefficient, with $|\rho|<1$.

Using matrix notation, the model can be written as

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \tag{2}$$

where

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_T \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ . \\ . \\ X_T \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ . \\ . \\ \epsilon_T \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}. \tag{3}$$

The LAV criterion selects estimates of $\beta_0$ and $\beta_1$ to minimize the sum of the absolute residuals. This problem can be stated and solved in a linear programming context. See Dielman & Pfaffenberger (1982) for a general survey of LAV regression.

We consider two different two-stage procedures to correct for autocorrelation in the LAV regression model. These procedures are analogous to the CO and PW procedures commonly employed in least squares regression, and differ in the treatment of the first observation. Both procedures transform the data using the autocorrelation coefficient, $\rho$. LAV estimation can then be applied to the transformed data. The PW transformation matrix is:

$$\mathbf{M}_1 = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & . & . & . & 0 & 0 \\ -\rho & 1 & . & . & . & 0 & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & . & . & . & -\rho & 1 \end{bmatrix}. \tag{4}$$

Premultiplying the model in (2) by $\mathbf{M}_1$ yields

$$\mathbf{M}_1\mathbf{Y} = \mathbf{M}_1\mathbf{X}\beta + \mathbf{M}_1\epsilon, \tag{5}$$

or

$$\mathbf{Y}^* = \mathbf{X}^*\beta + \eta, \tag{6}$$

where $\mathbf{Y}^*$ contains the transformed dependent variable values and $\mathbf{X}^*$ is the matrix of transformed independent variable values, so

$$\mathbf{Y}^{*\prime} = \left[\sqrt{1-\rho^2}\, y_1 \quad y_2 - \rho y_1 \quad \cdots \quad y_T - \rho y_{T-1}\right] \tag{7}$$

and

$$\mathbf{X}^* = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}\, x_1 \\ 1-\rho & x_2 - \rho x_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1-\rho & x_T - \rho x_{t-1} \end{bmatrix}. \tag{8}$$

In (6), $\eta$ is the vector of serially uncorrelated $\eta_t$ errors.

The CO transformation matrix is the (T-1)×1 matrix obtained by removing the first row of the $\mathbf{M}_1$ transformation matrix. Note that the use of the CO transformation means that (T-1) observations, rather than T, are used in the model estimation. The first observation is omitted, rather than transformed and included in the estimation. Asymptotically, the loss of this single observation is relatively inconsequential. However, for small samples, omitting the first observation has been shown to result in a least squares estimator inferior to that obtained when the first observation is retained and transformed; see Dielman & Pfaffenberger (1984), Maeshiro (1979), and Park & Mitchell (1980). Dielman & Rose (1994a) observed the same phenomenon in the context of LAV-estimated models.

Operationalization of the model requires that the correlation coefficient, ρ, be estimated from sample data. We estimate ρ by applying LAV estimation to the following equation:

$$\hat{\varepsilon}_t = \rho \hat{\varepsilon}_{t-1} + \eta_t, \tag{9}$$

where the $\hat{\varepsilon}_t$ are the residuals from the LAV fit to the model in (1).

In this paper, we consider the following estimators for the coefficients $\beta_0$ and $\beta_1$ in (1):

1. LAV estimation with no correction for autocorrelation: the "LAV" (never-correct) estimator.

2. LAV estimation after using the CO correction for autocorrelation: the "CO" (always-correct) estimator.

3. LAV estimation after using the PW correction for autocorrelation: the "PW" (always-correct) estimator.

4. LAV estimation after using the CO correction if a test for autocorrelation indicates that correction is required; otherwise LAV estimation with no correction: the "PRE-CO" (pre-test) estimator.

5. LAV estimation after using the PW correction if a test for autocorrelation indicates that correction is required; otherwise LAV estimation with no correction: the "PRE-PW" (pre-test) estimator.

The test for autocorrelation used in the pre-test estimators is the Durbin-Watson test applied to the residuals from the LAV fitted model. Weiss (1990) examined tests for autocorrelation in models estimated using LAV regression, and found that the Durbin-Watson test applied to the LAV residuals performed well. The test concludes that correction is necessary if the test statistic falls below the Durbin-Watson lower bound. Otherwise, no correction is performed and uncorrected LAV estimation is used.

## THE MONTE CARLO DESIGN

The experimental design for the Monte Carlo simulation consists of the following factors.

**Sample size:** The sample size is $T$=20, chosen to represent a data history that is short enough so that asymptotic results cannot reasonably be assumed to be valid. Many applications of practical interest in business and economics involve samples of approximately this size (e.g., five years of quarterly data). In addition, previous LAV-related studies (e.g., Dielman & Pfaffenberger [1990, 1992]) have indicated few differences in model behavior for $T \geq 40$, which would imply that asymptotic results may be applicable above that threshold. These studies also showed that results for $T$=20 and $T$=30 were similar to each other, while results for $T$=14 were noticeably different. Therefore, the selection of the sample size for this study is motivated by a desire to investigate small-sample results, while considering a sufficiently large number of observations to maintain the viability of the model.

**Coefficient values:** The parameters $\beta_0$ and $\beta_1$ are both set equal to zero. This selection causes no loss of generality. Refer to Andrews (1986), who showed that the choices of the parameter values are irrelevant in Monte Carlo efficiency comparisons using iterative generalized least squares or LAV estimators of the type considered in this study.

**Degree of autocorrelation:** $\rho = 0, 0.2, 0.4, 0.6, 0.8, 0.95$. We do not include values for $\rho < 0$ in the study, since negative autocorrelation is not encountered frequently in practical business and economic applications.

**Disturbance distributions:** Four different distributional forms for the $\eta_t$ disturbances are considered, to include a variety of outlier-producing distributions:

1. Normal, with mean 0 and variance 1; i.e., N(0,1).

2. Contaminated normal, where the $\eta_t$ are drawn from a N(0,1) distribution with probability 0.85 and from a N(0,25) distribution with probability 0.15.

3. Laplace (double exponential), with mean 0 and variance 2. Uniform (0,1) random deviates are generated and transformed to Laplace deviates using the inverse c.d.f. transformation.

4. Cauchy, with median 0 and scale parameter 1.

After generating the $\eta_t$ values, the $\varepsilon_t$ values are created as $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$, where $\varepsilon_0 = \eta_0/(1-\rho^2)$, and $\eta_0$ is an initial draw from the disturbance distribution.

**Explanatory variable:** The explanatory variables are generated in the following ways, to enhance the generalizability of the results:

1. Normally distributed: $x_t = u_t$, with $u_t \sim N(0,2)$.
2. Autoregression: $x_t = \lambda x_{t-1} + u_t$, for $\lambda = 0.4, 0.8$, with $u_t \sim N(0,2)$.
3. Stochastic time trend: $x_t = \lambda t + u_t$, for $\lambda = 0.4, 0.8$, with $u_t \sim N(0,1)$.

The patterns in the explanatory variables generated in these ways should be representative of patterns encountered in practical time series applications. Once generated, these values are held fixed throughout the experiment.

For each factor combination in the experimental design, 5000 Monte Carlo trials are used, and the resulting parameter estimates are recorded. All random numbers are generated using IMSL subroutines, and the explanatory variable values are generated independently of the disturbances. The simulation software is written in FORTRAN and run on an IBM 4341, Model 12.

## RESULTS

The performances of various estimators can be compared using a variety of criteria. In this study, we consider the mean absolute deviations (MAD) of the slope and the intercept estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$, for each method, computed as

$$\text{MAD} = \frac{\sum_{i=1}^{5000} |\hat{\beta}_i - \beta_i|}{5000},$$

where $\beta_i$ represents the true value of either $\beta_0$ or $\beta_1$, as appropriate, and $\hat{\beta}_i$ is the estimated value of $\beta_i$. Note that the MAD value can be viewed as a measure of an estimator's efficiency. Since a smaller MAD value is better, pairs of estimation methods can be compared by considering the ratio of their MAD values. These ratios are then measures of relative efficiency. Clearly, if the MAD ratio has a value that is less than 1, the estimation approach represented in the numerator is preferable to that represented in the denominator, based on this criterion.

**"Is pre-testing better than always correcting? The answer is a resounding 'no.'"**

Various MAD ratios for the 5000 trials have been tabulated for testing $\beta_0$ and $\beta_1$. Figures 1-6 provide graphical displays of some of the most important results, showing median MAD ratios (across explanatory variable types) for estimating $\beta_0$ and $\beta_1$, for the values of $\rho$ and the four different error distribution types.

As expected, the performances of the PW-type and the CO-type corrections are quite similar with respect to the estimation of $\beta_1$, with some preference for the PW-type approach; these results parallel those observed by Dielman & Rose (1994a). The comparison between the two approaches is much more dramatic when estimating $\beta_0$. The CO-type correction performs extremely poorly when estimating the intercept term. Therefore, our analysis of the results focuses more heavily on the PW-type corrections. Three questions are of primary interest and are addressed below.

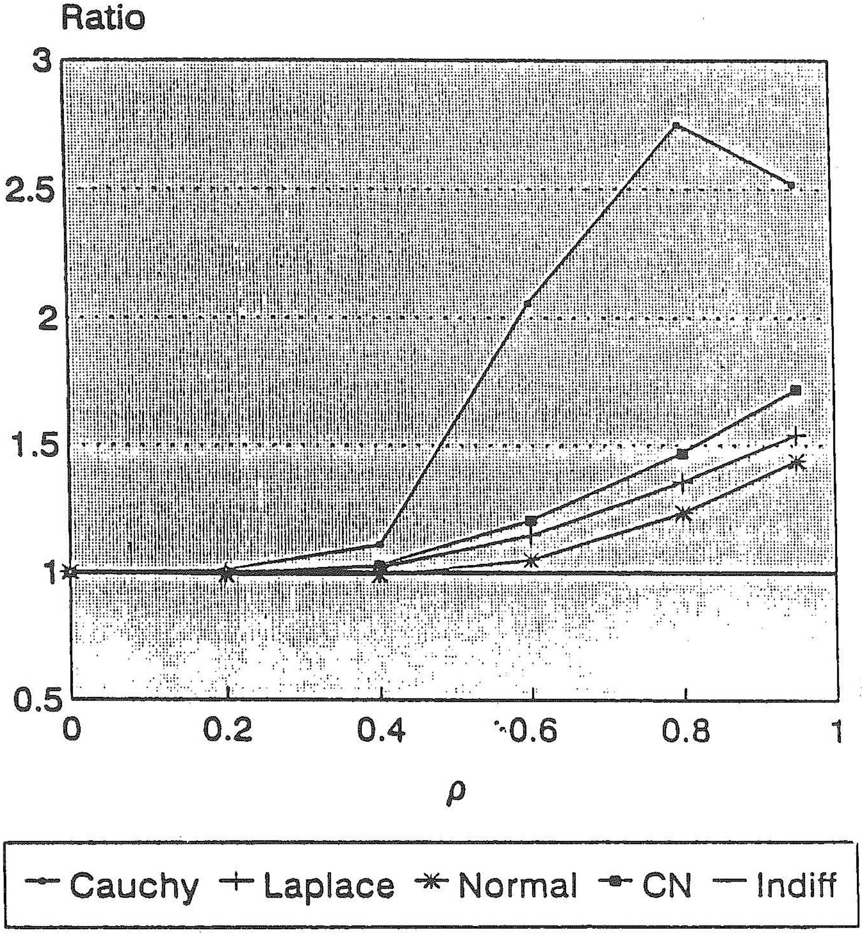
**Is pre-testing better than never correcting?** This comparison is made by studying the MAD ratios for LAV:PRE-PW and LAV:PRE-CO. (We report only PW results; complete results can be obtained from the authors.) Consider, first, the slope parameter (Figure 1). Not surprisingly, the preference for the pre-test correction increases with $\rho$. For Cauchy-distributed errors, correction is never worse, and is always preferred when $\rho \geq .2$. When the errors follow a Laplace distribution, correction is preferred when $\rho \geq .4$. For normally-distributed errors, correction is not uniformly preferred until $\rho \geq .8$, and correction is generally preferred for errors that follow a contaminated normal distribution. The loss associated with correction is minimal (e.g., relative efficiency=.96), whereas the loss associated with not correcting can be quite large (e.g., relative efficiency=2.65). For the intercept term (Figure 2), correction is generally preferred for Cauchy, Laplace, and contaminated normal errors. When the errors are normally-distributed, the preference is generally for uncorrected LAV estimation.

**"The pre-test estimator exhibits its best performance when the errors follow a normal distribution."**

**Is pre-testing better than always correcting?** The answer is a resounding "no." Given the preference for the PW-type approach to correction, this comparison is made primarily by considering the MAD ratio for PRE-PW:PW. For the slope parameter (Figure 3), there is little loss associated with always correcting, even when $\rho=0$. The pre-test correction performs poorly for Cauchy errors and $.2\leq\rho\leq.8$. The loss associated with pre-testing is reduced for $\rho=.95$ (the pre-testing results in correction for virtually all of the 5000 trials). While Laplace-distributed errors yield a slight preference for pre-testing when $\rho=0$, the loss associated with pre-testing with relatively low autocorrelation far exceeds this small gain. The pre-test estimator exhibits its best performance when the errors follow a normal distribution, which is not surprising, since the Durbin-Watson pre-test is based on an assumption of normality. The pre-test approach is preferred for contaminated normal errors only when $\rho=0$ and the explanatory variables follow a time trend. The results are fairly similar for the intercept term (Figure 4). Again, the pre-test correction is particularly bad for Cauchy-distributed errors and performs best for normally-distributed errors. For all error distributions, the pre-test approach is most appropriate when the explanatory variables follow a time trend.

**Figure 1**

**MAD Ratios For Estimating $\beta_1$, LAV:PRE-PW (Ratio greater than 1 implies preference for pre-testing, rather than never correcting)**




Medians across explanatory variable types

**Figure 2**

**MAD Ratios For Estimating $\beta_0$, LAV:PRE-PW (Ratio greater than 1 implies preference for pre-testing, rather than never correcting)**

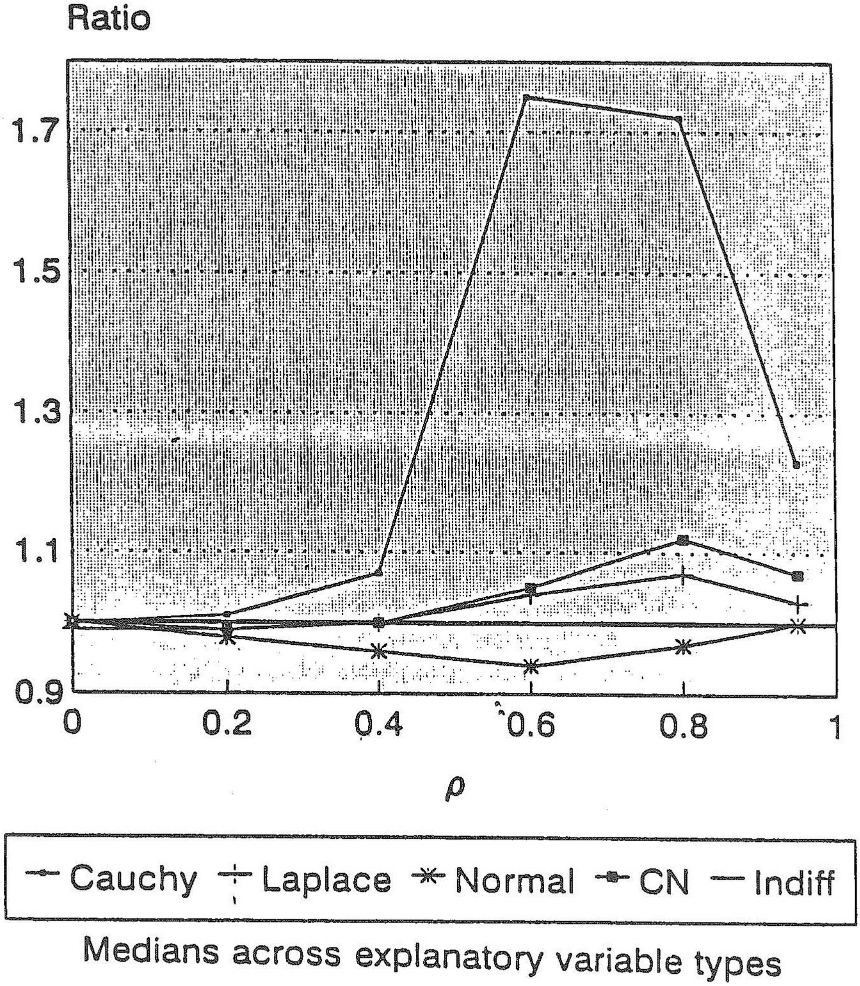


**Figure 3**

**MAD Ratios For Estimating $\beta_1$, PRE-PW:PW (Ratio greater than 1 implies preference for always correcting, rather than pre-testing)**

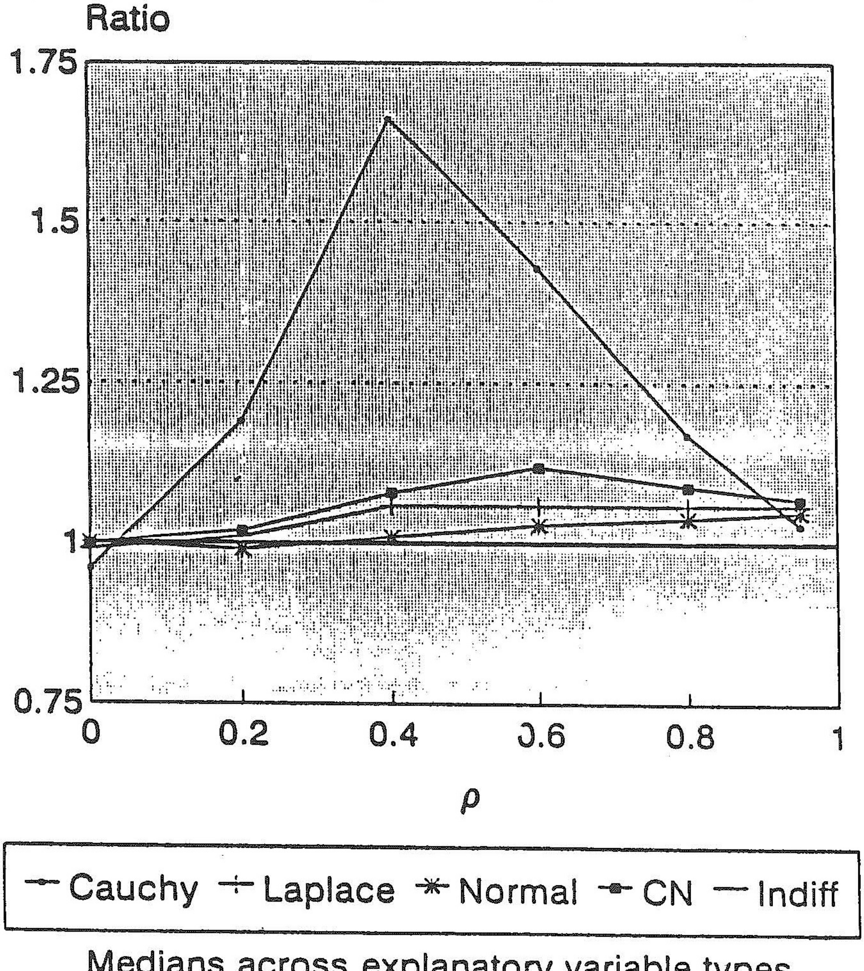
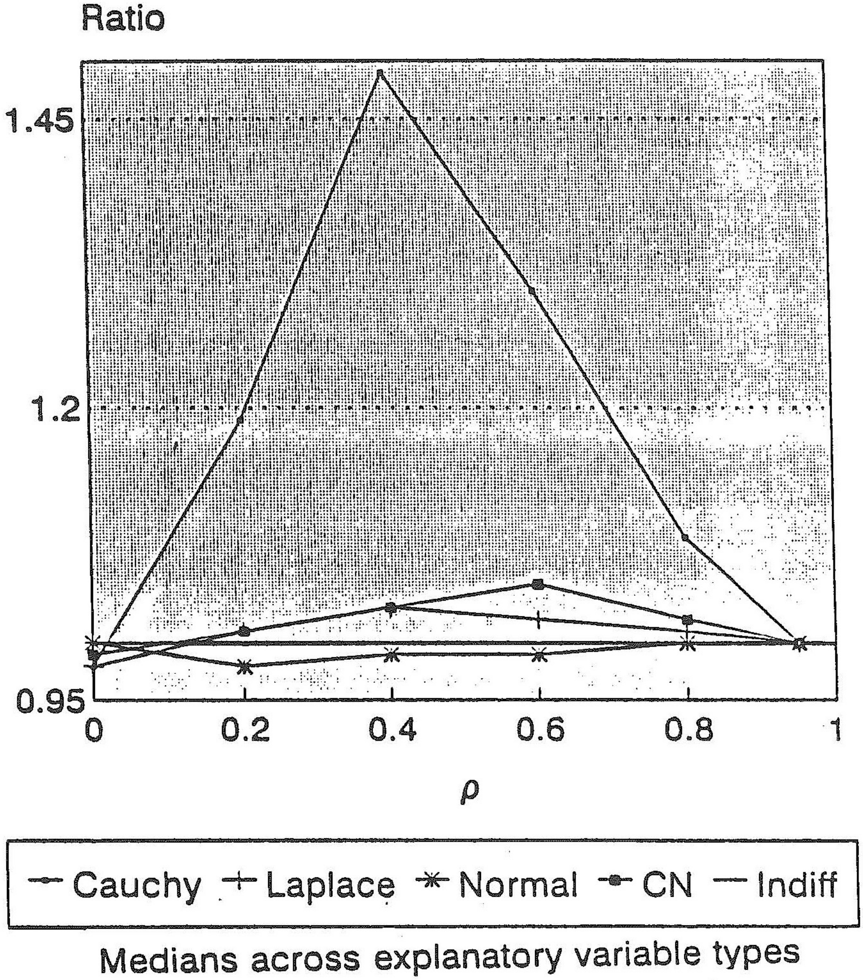


**Figure 4**

**MAD Ratios For Estimating $\beta_0$, PRE-PW:PW (Ratio greater than 1 implies preference for always correcting, rather than pre-testing)**



**"There seems to be little loss associated with adopting the always-correct approach, in terms of estimation performance."**

**When pre-testing, is the PW approach better than the CO approach?** Comparing the MAD ratios for PRE-PW:PRE-CO, the two approaches seem relatively similar for the slope coefficient (Figure 5), although the results indicate a preference for the PW approach when the autocorrelation is stronger, and a strong preference for the PW-type correction when the errors follow a Cauchy distribution. PRE-CO is never preferred to PRE-PW, except for normally-distributed explanatory variables and $\rho=.95$. However, in estimating the intercept term (Figure 6), PRE-CO exhibits remarkably poor performance when $\rho\geq.4$. Therefore, the PW approach is preferred to the CO approach.

**Figure 5**

**MAD Ratios For Estimating $\beta_1$, PRE-PW:PW-CO (Ratio less than 1 implies preference pre-testing using PW, rather than pre-testing using CO)**

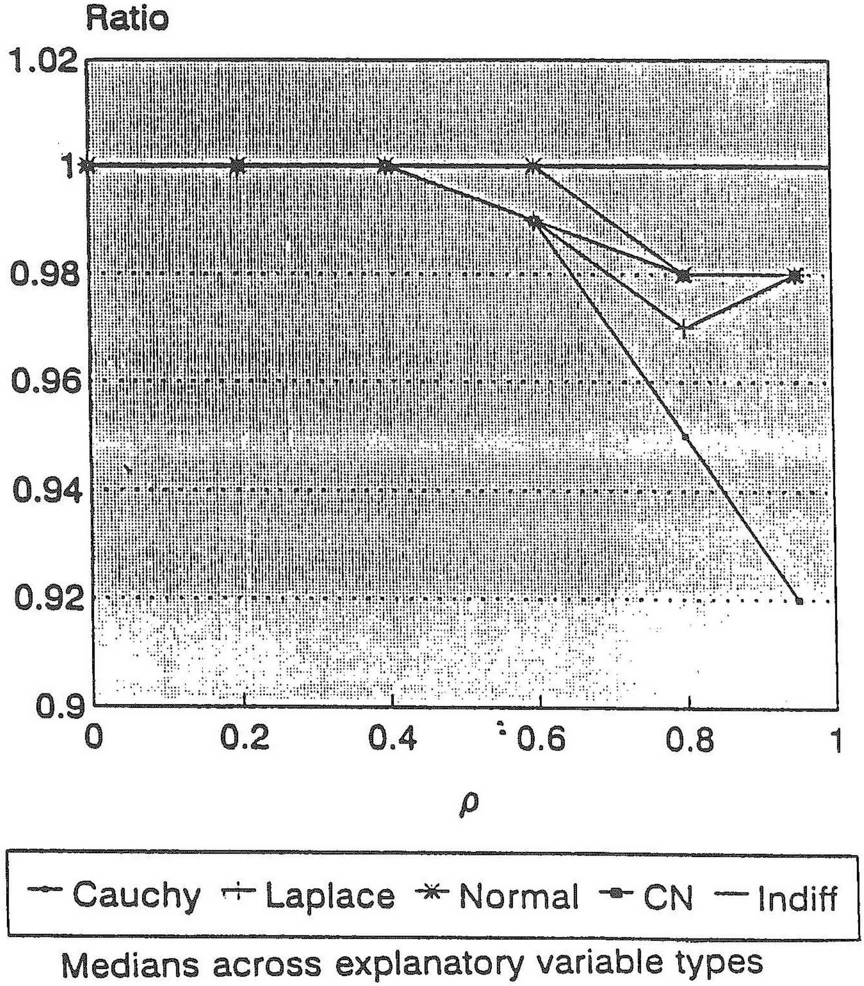
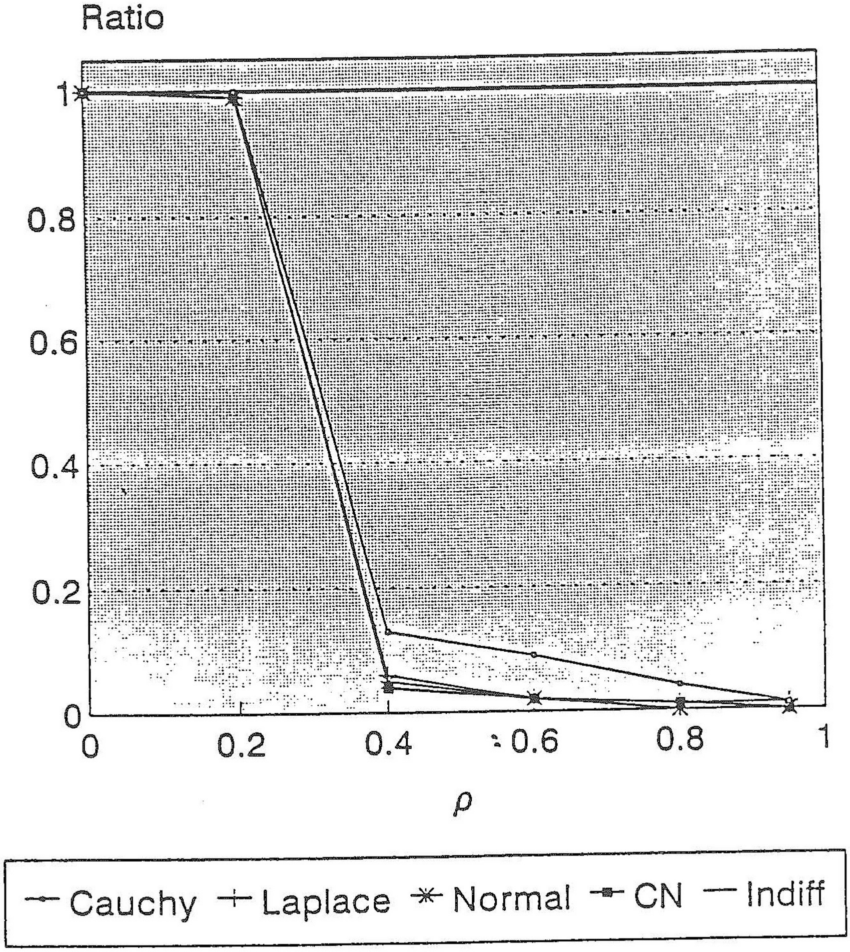


Figure 6

**MAD Ratios For Estimating $\beta_0$, PRE-PW:PW-CO (Ratio less than 1 implies preference pre-testing using PW, rather than pre-testing using CO)**




Medians across explanatory variable types

**"In the context of LAV regression . . . this study does provide additional evidence that Prais-Winsten-type correction should be used, rather than Cochrane-Orcutt-type correction."**

## CONCLUSIONS

In comparing the pre-test approach against the strategy of always correcting for autocorrelation, this study indicates that there seems to be little loss associated with adopting the always-correct approach, in terms of estimation performance. Additional research is needed to develop a more complete understanding of the circumstances under which the use of a pre-test for serial correlation is useful in the context of LAV regression. This study does provide additional evidence that Prais-Winsten-type correction should be used, rather than Cochrane-Orcutt-type correction. The choice of correction is particularly crucial when inference is made regarding the intercept term, due to the remarkably poor performance of the Cochrane-Orcutt correction in that context.

## REFERENCES

Andrews, D.W.K. (1986). "A Note on the Unbiasedness of Feasible GLS, Quasi-maximum Likelihood, Robust, Adaptive, and Spectral Estimators of the Linear Model." ***Econometrica***, 54, 687-698.

Cochrane, D. and Orcutt, G.H. (1949). "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms." ***Journal of the American Statistical Association***, 44, 32-61.

Dielman, T. & Pfaffenberger, R. (1992). "A Further Comparison of Tests of Hypotheses in LAV Regression." ***Computational Statistics and Data Analysis***, 14, 375-384.

Dielman, T. & Pfaffenberger, R. (1990). "Tests of Linear Hypotheses in LAV Regression." ***Communications in Statistics - Simulation and Computation***, 19, 1179-1199.

Dielman, T. & Pfaffenberger, R. (1989). "Small Sample Properties of Estimators in the Autocorrelated Error Model: A Review and Some Additional Simulations." ***Statistical Papers/Statistische Hefte***, 30, 163-183.

Dielman, T. & Pfaffenberger, R. (1984). "Computational Algorithms for Calculating Least Absolute Value and Chebyshev Estimates in Multiple Regression." ***American Journal of Mathematical and Management Sciences***, 4, 169-197.

Dielman, T. & Pfaffenberger, R. (1982). "LAV (Least Absolute Value) Estimation in Linear Regression: A Review." In: S. Zanakis and J. Rustagi, eds., ***TIMS Studies in Management Science***. Amsterdam: North-Holland, 31-52.

Dielman, T. & Rose, E.L. (1994a). "Estimation in Least Absolute Value Regression with Autocorrelated Errors." ***Journal of Statistical Computation and Simulation***, 50, 29-43.

Dielman, T. & Rose, E.L. (1994b). "Forecasting in Least Absolute Value Regression with Autocorrelated Errors: A Small-sample Study." ***International Journal of Forecasting***, 10, 539-547.

Griffiths, W.E. & Beesley, P.A.A. (1984). "The Small-sample Properties of Some Preliminary Test Estimators in a Linear Model with Autocorrelated Errors." ***Journal of Econometrics***, 25, 49-62.

King, M.L. & Giles, D.E.A. (1984). "Autocorrelation Pre-testing in the Linear Model: Estimation, Testing and Prediction." ***Journal of Econometrics***, 25, 35-48.

Maeshiro, A. (1979). "On the Retention of the First Observation in Serial Correlation Adjustment of Regression Models." ***International Economic Review***, 20 , 259-265.

Park, R. & Mitchell, G. (1980). "Estimating the Autocorrelated Error Model with Trended Data." ***Journal of Econometrics***, 13, 185-201.

Prais, S. & Winsten, C. (1954). "Trend Estimators and Serial Correlation."Cowles Commission Discussion Paper: Stat. No. 383. Chicago.

Weiss, A. (1990). "Least Absolute Error Estimation in the Presence of Serial Correlation." ***Journal of Econometrics***, 44, 127-158.