DigitalCommons @WayneState

Journal of Modern Applied Statistical Methods

Volume 8 | Issue 1 Article 9

5-1-2009

Email: A Note on Hypothesis Tests after Correction for Autocorrelation: Solace for the Cochrane-Orcutt Method?

Terry E. Dielman
Texas Christian University, t.dielman@tcu.edu

Journal of Modern Applied Statistical Methods
May 2009, Vol. 8, No. 1, 100-109

1538 – 9472/09/$95.00

A Note on Hypothesis Tests after Correction for Autocorrelation: Solace for the Cochrane-Orcutt Method?

Terry E. Dielman
Texas Christian University

The behavior of the *t* test in small samples for coefficient significance in time-series regressions is examined after using the Prais-Winsten (PW) and Cochrane-Orcutt (CO) corrections for autocorrelation. Results are compared to ordinary least squares and generalized least squares.

Key words: First-order autocorrelation generalized least squares, ordinary least squares, Prais-Winsten, time series regression.

Introduction

The Prais-Winsten (PW) and Cochrane-Orcutt (CO) methods are popular procedures for correcting for autocorrelation in time-series regression models. Both methods transform the data using a differencing transformation to remove autocorrelation. Ordinary least squares (OLS) applied to the transformed observations will yield estimators that are asymptotically more efficient than OLS applied to the original data.

The PW and CO methods are essentially equivalent except for the treatment of the first observation in the data set. The CO method simply omits the first observation, while the PW method transforms the observation and retains it. Asymptotically, there is no difference in the efficiency of estimators produced by the two methods. In previous studies of small sample behavior, however, the superior performance of the PW procedure has been documented. Using the CO procedure results in estimators that are less efficient in small samples. Under certain conditions, the CO estimator can even be less efficient than OLS applied to the original data.

Terry E. Dielman is a Professor in the M.J. Neeley School of Business, Department of Information Systems and Supply Chain Management. Email: t.dielman@tcu.edu.

Due to the inefficiency of the CO estimator, comparisons of hypothesis testing results from models estimated by PW and CO have not been considered. This article examines the behavior of the *t* test in small samples for coefficient significance in time-series regressions. Tests are compared using four estimation procedures: OLS, CO, PW and generalized least squares estimation (GLS) using the true value of the autocorrelation coefficient.

The results suggest that the PW and CO methods perform similarly when testing hypotheses, but in certain cases, CO outperforms PW. This does not, however, mean that either method performed particularly well. Both had levels of significance that were much higher than desirable in certain circumstances. The poor performance of these procedures in situations when they are intended to correct for autocorrelation suggests the need for either better estimates of the autocorrelation coefficient, better procedures for correcting for autocorrelation, or alternative approaches that will result in improved hypothesis tests.

Methodology

The following simple regression model is considered:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \text{ with } \varepsilon_t = \rho\, \varepsilon_{t-1} + \eta_t \quad (1)$$

for $t = 1,2,...,T$. In equation (1), y_t and x_t are the t^{th} observations on the dependent and explanatory variables, respectively, and ε_t is a random disturbance for the t[th] observation and may be subject to autocorrelation. The η_t represents disturbance components that are assumed to be independent and identically distributed. The parameters β_0 and β_1 are unknown and must be estimated. The parameter ρ is the autocorrelation coefficient, with $|\rho|<1$. Using matrix notation, the model can be written as:

$$\mathbf{Y} = \mathbf{X\beta} + \boldsymbol{\varepsilon} \tag{2}$$

where

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_T \end{bmatrix},\ \mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ . & . \\ 1 & x_T \end{bmatrix},\ \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_T \end{bmatrix},\ \text{and } \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \tag{3}$$

Two procedures to correct for autocorrelation are examined. These are the Prais-Winsten (1954) and Cochrane-Orcutt (1949) procedures. Both procedures transform the data using the autocorrelation coefficient, ρ, after which the transformed data are used in estimation. The procedures differ in their treatment of the first observation, (x_1, y_1). The PW transformation matrix is:

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

Pre-multiplying the model in (2) by $\mathbf{M}_{\text{PW}}$ yields

$$\mathbf{M_{PW}Y} = \mathbf{M_{PW}X\beta} + \mathbf{M_{PW}}\boldsymbol{\varepsilon} \tag{5}$$

or

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

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

$$\mathbf{Y}^* = \left\lfloor \sqrt{1-\rho^2}\, y_1, y_2 - \rho y_1, \ldots, y_T - \rho y_{T-1} \right\rfloor \tag{7}$$

and

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

In (6), $\boldsymbol{\eta}$ is the vector of serially uncorrelated η_t errors.

The CO transformation matrix is the (T−1) x 1 matrix obtained by removing the first row of the $\mathbf{M}_{\text{PW}}$ transformation matrix. The use of the CO transformation means that (T−1) observations, rather than T, are used to estimate the model. In the CO transformation, the first observation is omitted, whereas it is transformed and included in the estimation in the PW transformation. Asymptotically, the loss of this single observation is probably of minimal concern. However, for small samples, omitting the first observation has been shown to result in an estimator inferior to that obtained when the first observation is retained and transformed. See Dielman & Pfaffenberger (1984), Maeshiro (1979), and Park & Mitchell (1980) for simulation studies demonstrating the efficiency gains of PW, and Doran (1981), Magee (1987), Taylor (1981), and Thornton (1987) for analytical results.

In practice the value of ρ will be unknown and it must be estimated from sample data. The estimators of ρ used will be as follows:

$$\hat{\rho}_{PW} = \frac{\sum_{t=2}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=2}^{T} \hat{\varepsilon}_t^2} \tag{9}$$

when all T observations are used, and

$$\hat{\rho}_{CO} = \frac{\sum_{t=2}^{T} \hat{\varepsilon}_t \hat{\varepsilon}_{t-1}}{\sum_{t=1}^{T-1} \hat{\varepsilon}_t^2} \quad (10)$$

when $T-1$ observations are used, where the $\hat{\varepsilon}_t$ represent OLS residuals. Park and Mitchell (1980) showed that these two estimators minimize the error sum of squares conditional on β when T and $T-1$ observations are used, respectively, in the estimation process.

The actual estimation procedures for both PW and CO are iterative procedures. OLS is run to obtain estimates of the regression coefficients and, subsequently, the $\hat{\varepsilon}_t$. The estimator of the autocorrelation coefficient, ρ, is computed, the data are transformed, and new estimates of the regression coefficients are obtained. The autocorrelation coefficient estimate is recomputed and compared to the previous estimate. In the results, if these estimates differ by less than 0.000001, the iterative procedure stops. The procedure also stops when it reaches 25 iterations. If boundary conditions are encountered the estimate of ρ is set at ±0.999999.

The model considered in this article is described in equation (1). The explanatory variable values are generated as follows:

1. $x_t = \lambda x_{t-1} + u_t$ for t = 1, 2, …, T with the u_t chosen from the N(0,2) distribution. The values of λ used were 0.0, 0.4 and 0.8.

2. A stochastic time trend is used. In this case $x_t = \lambda t + u_t$ for t = 1, 2, ..., T and u_t is chosen from the N(0,2) distribution for λ = 0.4 and 0.8.

Once generated, these values are held fixed throughout the experiment for each sample size. The disturbances, η_t, are chosen from the N(0,1) distribution. After generating the η_t, the ε_t values are created as $\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t$ where $\varepsilon_0 = \frac{\eta_0}{1-\rho^2}$ and η_0 is an initial draw from the disturbance distribution. The explanatory variable values were generated independently of the disturbances.

The parameter β_0 was set equal to zero (without loss of generality). The parameter β_1 was set equal to zero to examine the level of significance. For each factor combination in the experimental design, ten thousand Monte Carlo trials were used to assess levels of significance. A sample size of T = 20 was used. The values of ρ were 0.0, 0.2, 0.4, 0.6, 0.8, and 0.95. The null hypothesis H_0: $\beta_1 = 0$ is tested using the t test and the number of rejections of the null hypothesis was recorded to assess the level of significance.

The hypothesis tests were also conducted with β_1 = 0.2, 0.3, 0.4, 0.5, and 1.0. When H_0 is rejected, the proportion of correct rejections can be used to construct empirical power functions. Power comparisons based on the original simulations are complicated by the differences in the observed significance levels. A valid power comparison can be made only if the true significance levels of the tests are similar, which is clearly not the case based on our results.

The power comparison was accomplished using a procedure suggested by Zhang and Boos (1994). From the original 10,000 simulations with $\beta_1 = 0$, the test statistics were sorted and the critical values producing a 5% level of significance were chosen for each design point. These values represent estimates of the critical values under the null hypothesis that produce an exact 5% level of significance. The simulation was repeated with the non-zero values of β_1, using the empirically determined critical values. The test statistics from the second set of simulations will have similar levels of significance, making their powers comparable. Zhang and Boos (1994) suggested using a larger number of Monte Carlo trials to estimate the correct critical value under the null hypothesis if possible. In this experiment 10,000 trials under the null and 5,000 under the alternative hypotheses were used.

Results are reported for four estimation procedures: OLS (assuming ρ = 0), PW and CO and GLS (which is the PW procedure using the true value of ρ). All random numbers were generated using IMSL subroutines and the simulation was written in FORTRAN.

Results

Consider Tables 1 and 2. These tables show the number of rejections of the true null hypothesis that the slope is zero for all factor combinations in the Monte Carlo simulation. Table 1 shows the results for the autoregressive independent variable; Table 2 for the stochastic trend variable. The most striking results are for the autoregressive case when λ is 0.8 and the stochastic trend case for λ equal to both 0.4 and 0.8. As the level of autocorrelation increases, the observed levels of significance become very high for OLS, but this is not unexpected. OLS is not expected to perform well when disturbances are autocorrelated.

However, the two methods that correct for autocorrelation do not perform well either. PW has very high rejection rates with some cases approaching 50%. The rejection rates for CO are high as well, but often not as high as PW. This is particularly evident when the independent variable is autoregressive. These results suggest that correcting for autocorrelation does not guarantee reliable inferences about the slope coefficient.

Selected power comparisons using 5,000 Monte Carlo trials are shown in Table 3 for the autoregressive independent variable with $\lambda = 0.8$ and in Table 4 for the stochastic trend variable with $\lambda = 0.8$. When the independent variable is autoregressive, CO generally has power equal to or slightly higher than PW. Figures 1 and 2 plot the empirical power curves for $\rho = 0.0$ and $\rho = 0.95$ from Table 3. When $\rho = 0.0$ there is little difference in adjusted power; when $\rho = 0.95$ CO has higher power than PW.

When the independent variable is a stochastic trend, there is little difference between PW and CO as evidenced in the empirical power curves in Figures 3 and 4 for $\rho = 0.0$ and $\rho = 0.95$, respectively. In this case, when $\rho = 0.95$, it is especially troublesome that OLS has higher adjusted power than either PW or CO, which supposedly adjust for autocorrelation. This result is driven by the very high levels of significance for OLS of course.

Conclusion

Previous studies have shown that the PW method is superior to CO as a correction for autocorrelation in terms of estimator efficiency. However, these results do not hold up in an examination of inference results. CO generally performs as well and in many cases better than PW in terms of observed level of significance and adjusted power. This should not be taken as a suggestion that the PW method should be abandoned and CO resorted to, however. Perhaps both methods should be abandoned and a better approach sought for handling autocorrelation in regression models. In terms of inference, a bootstrap approach as Rayner (1991) suggested might be preferred to either the PW or CO method. Alternatively, as suggested by Mizon (1995), perhaps another approach to correcting for autocorrelation should be considered. Bayesian estimators (see Ohtani, 1990, and Kennedy & Simons, 1991) also hold promise for improvements.

References

Cochrane, D., & Orcutt, G. (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. (1984). Small sample properties of estimators in the autocorrelated error model: A review and some additional simulations. *Statistical Papers/Statistiche Hefte*, *30*, 163-183.

Doran, H. (1981). Omission of an observation from a regression analysis: A discussion on efficiency loss with applications. *Journal of Econometrics*, *16*, 367-374.

Kennedy, P., & Simons, D. (1991). Fighting the teflon factor: Comparing classical and Bayesian estimators for autocorrelated errors. *Journal of Econometrics*, *48*, 15-27.

Maeshiro, A. (1979). On the retention of the first observation in serial correlation adjustment of regression models. *International Economic Review*, *20*, 259-265.

Magee, L. (1987). A note on Cochrane-Orcutt estimation. *Journal of Econometrics*, *35*, 211-218.

Table 1: Empirical Significance Level: Number of Rejections of True Null Hypothesis H_0: $\beta_1 = 0$ Using Autoregressive Independent Variable (10,000 Trials)

Lambda = 0.0	Rho = 0.00	0.20	0.40	0.60	0.80	0.95
OLS	508	601	623	531	342	226
PW	754	757	716	637	523	462
GLS	508	500	516	514	526	530
CO	699	716	697	617	471	424
Lambda = 0.4	**Rho = 0.00**	**0.20**	**0.40**	**0.60**	**0.80**	**0.95**
OLS	516	712	870	852	700	491
PW	808	819	826	755	640	546
GLS	516	516	501	503	508	493
CO	744	771	764	694	566	447
Lambda = 0.8	**Rho = 0.00**	**0.20**	**0.40**	**0.60**	**0.80**	**0.95**
OLS	521	865	1307	1848	2658	3613
PW	806	942	1048	1232	1442	1679
GLS	521	512	512	506	505	496
CO	726	877	995	1057	949	822

Table 2: Empirical Significance Level: Number of Rejections of True Null Hypothesis H_0: $\beta_1 = 0$ Using Stochastic Trend Independent Variable (10,000 Trials)

Lambda = 0.4	Rho = 0.00	0.20	0.40	0.60	0.80	0.95
OLS	453	936	1685	2848	4537	6111
PW	842	1026	1291	1815	2965	4496
GLS	453	458	475	482	477	494
CO	805	1008	1307	1860	2955	4419
Lambda = 0.8	**Rho = 0.00**	**0.20**	**0.40**	**0.60**	**0.80**	**0.95**
OLS	467	1013	1822	3087	4812	6352
PW	831	1039	1347	1941	3181	4711
GLS	467	456	453	459	484	487
CO	819	1036	1399	2038	3288	4706

Table 3: Adjusted Power Comparisons Using Autoregressive Independent Variable with Lambda = 0.8 (5,000 trials)

Rho = 0.0	Beta =					
	0.0	0.2	0.3	0.4	0.5	1.0
OLS	250	3006	4572	4961	4998	5000
PW	250	3107	4535	4922	4988	5000
GLS	250	3006	4572	4961	4998	5000
CO	250	3108	4550	4920	4992	5000
Rho = 0.2	Beta =					
	0.0	0.2	0.3	0.4	0.5	1.0
OLS	250	2476	4079	4819	4976	5000
PW	250	2332	3856	4695	4944	5000
GLS	250	2361	4043	4817	4978	5000
CO	250	2457	3991	4731	4953	5000
Rho = 0.4	Beta =					
	0.0	0.2	0.3	0.4	0.5	1.0
OLS	250	2039	3407	4429	4845	5000
PW	250	1737	3003	4175	4749	5000
GLS	250	1860	3420	4507	4907	5000
CO	250	1932	3254	4346	4795	5000
Rho = 0.6	Beta =					
	0.0	0.2	0.3	0.4	0.5	1.0
OLS	250	1684	2711	3659	4364	5000
PW	250	1286	2313	3359	4252	5000
GLS	250	1526	2925	4120	4729	5000
CO	250	1655	2771	3749	4490	5000
Rho = 0.8	Beta =					
	0.0	0.2	0.3	0.4	0.5	1.0
OLS	250	1478	2153	2849	3475	4906
PW	250	1089	1808	2656	3497	4967
GLS	250	1424	2771	3935	4637	5000
CO	250	1651	2430	3287	3970	4977
Rho = 0.95	Beta =					
	0.0	0.2	0.3	0.4	0.5	1.0
OLS	250	1430	1874	2361	2852	4473
PW	250	973	1526	2235	3048	4821
GLS	250	1532	2911	4072	4686	5000
CO	250	1726	2307	3025	3626	4835

Figure 1: Power Curve for Testing Slope Equal Zero:
Autoregressive With Lambda = 0.8; rho = 0.0

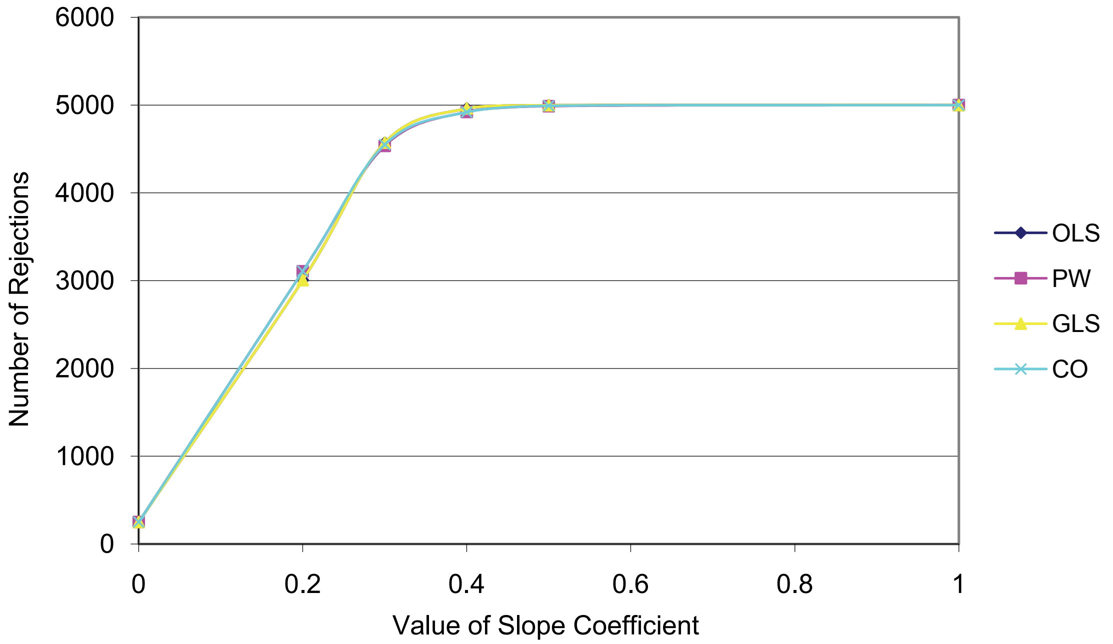


Figure 2: Power Curve for Testing Slope Equal Zero:
Autoregressive With Lambda=0.8, rho=0.95

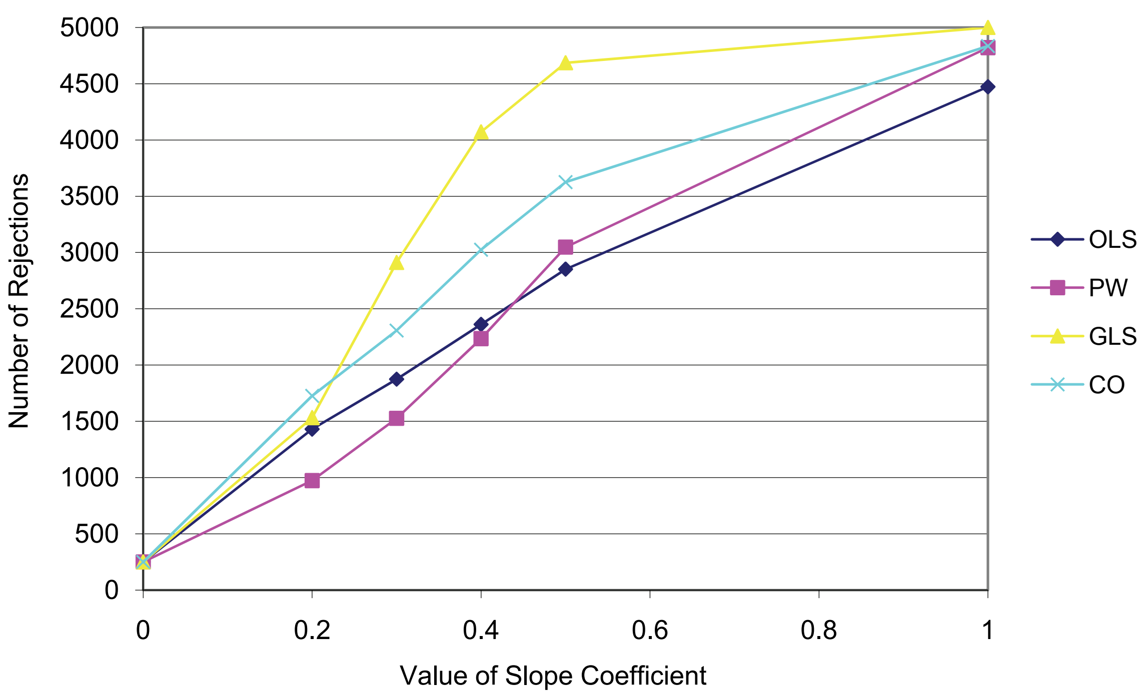


Table 4: Adjusted Power Comparisons Using Stochastic Trend Independent Variable with Lambda = 0.8 (5,000 trials)

Rho = 0.00	Beta = 0.0	0.2	0.3	0.4	0.5	1.0
OLS	250	4866	4999	5000	5000	5000
PW	250	4785	4983	4998	5000	5000
GLS	250	4866	4999	5000	5000	5000
CO	250	4666	4959	4993	4996	5000
Rho = 0.20	**Beta = 0.0**	**0.2**	**0.3**	**0.4**	**0.5**	**1.0**
OLS	250	4475	4990	5000	5000	5000
PW	250	4124	4873	4988	5000	5000
GLS	250	4494	4994	5000	5000	5000
CO	250	3900	4788	4959	4989	5000
Rho = 0.40	**Beta = 0.0**	**0.2**	**0.3**	**0.4**	**0.5**	**1.0**
OLS	250	3591	4817	4992	5000	5000
PW	250	2860	4389	4869	4984	5000
GLS	250	3768	4882	4998	5000	5000
CO	250	2656	4207	4752	4931	5000
Rho = 0.60	**Beta = 0.0**	**0.2**	**0.3**	**0.4**	**0.5**	**1.0**
OLS	250	2268	3856	4712	4959	5000
PW	250	1520	2963	4176	4714	5000
GLS	250	2673	4294	4916	4996	5000
CO	250	1440	2840	3964	4540	4997
Rho = 0.80	**Beta = 0.0**	**0.2**	**0.3**	**0.4**	**0.5**	**1.0**
OLS	250	1133	2124	3209	4022	5000
PW	250	715	1406	2295	3193	4962
GLS	250	1713	3143	4291	4841	5000
CO	250	701	1371	2232	3049	4880
Rho = 0.95	**Beta = 0.0**	**0.2**	**0.3**	**0.4**	**0.5**	**1.0**
OLS	250	685	1161	1766	2446	4661
PW	250	415	713	1112	1608	4074
GLS	250	1294	2457	3634	4450	5000
CO	250	442	733	1164	1660	4024

Figure 3: Power Curve for Testing Slope Equal Zero:
Stochastic Trend With Lambda = 0.8; rho = 0.0

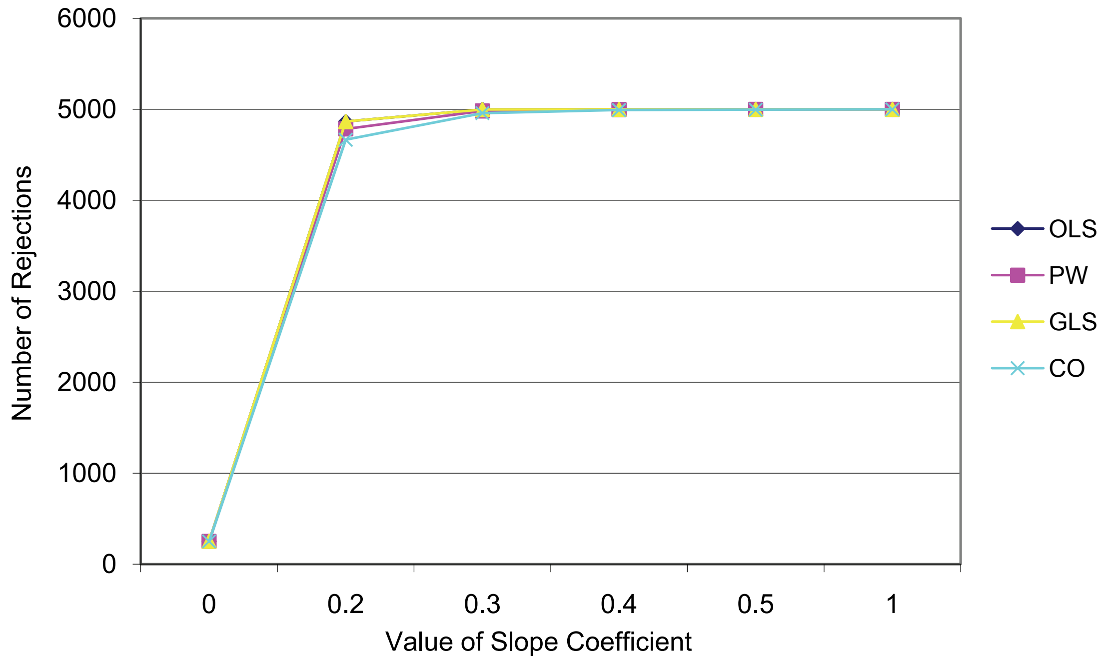
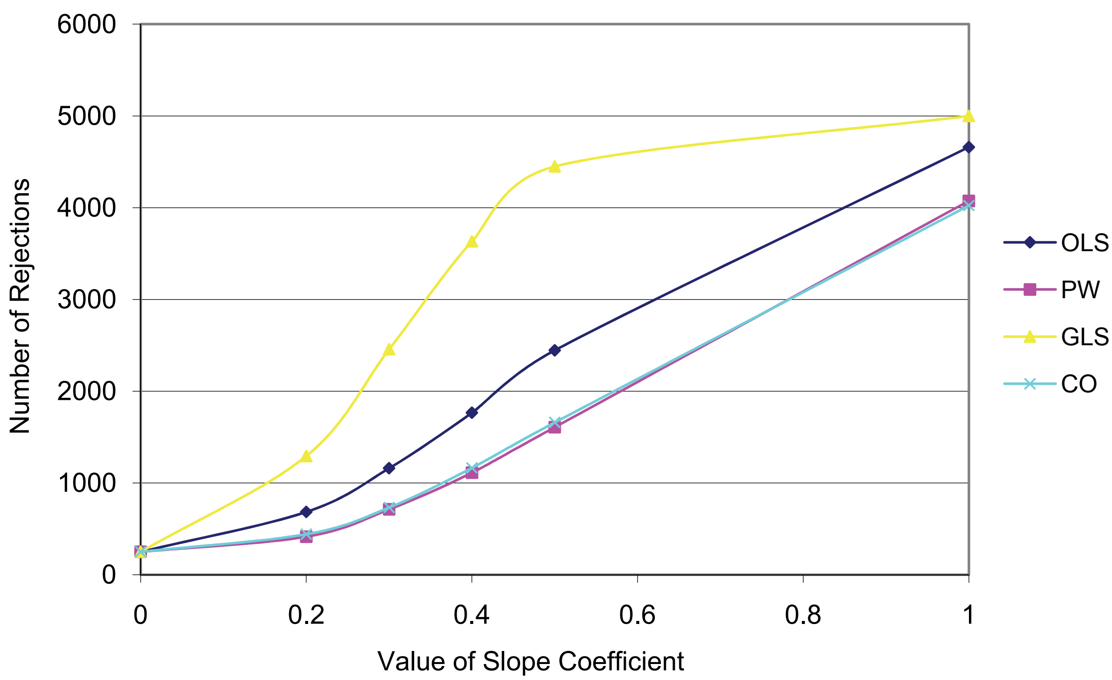


Figure 4: Power Curve for Testing Slope Equal Zero:
Stochastic Trend With Lambda = 0.8; rho = 0.95



Mizon, G. (1995). A simple message for autocorrelation correctors: Don't. *Journal of Econometrics*, *69*, 267-288.

Ohtani, K. (1990). On estimating and testing in a linear regression model with autocorrelated errors. *Journal of Econometrics*, *44*, 333-346.

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, IL.

Rayner, R. (1991). Resampling methods for tests in regression models with autocorrelated errors. *Economics Letters*, *36*, 281-284.

Taylor, W. (1981). On the efficiency of the Cochrane-Orcutt estimator. *Journal of Econometrics*, *17*, 67-82.

Thornton, D. (1987). A note on the efficiency of the Cochrane-Orcutt estimator of the AR(1) regression model. *Journal of Econometrics*, *36*, 369-376.

Zhang, J., & Boos, D. (1994). Adjusted power estimates in Monte Carlo experiments. *Communication in Statistics - Simulation and Computation*, *23*, 165-173.