NBER TECHNICAL PAPER SERIES

PITFALLS IN THE USE OF TIME AS AN EXPLANATORY VARIABLE IN REGRESSION

Charles R. Nelson

Heejoon Kang

Technical Working Paper No. 30

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

November 1983

The research reported here is part of the NBER's research program in Financial Markets and Monetary Economics. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Technical Working Paper #30
November 1983

Pitfalls in the use of Time as an Explanatory Variable in Regression

ABSTRACT

Regression of a trendless random walk on time produces R-squared values around .44 regardless of sample length. The residuals from the regression exhibit only about 14 percent as much variation as the original series even though the underlying process has no functional dependence on time. The autocorrelation structure of these "detrended" random walks is pseudo-cyclical and purely artifactual. Conventional tests for trend are strongly biased towards finding a trend when none is present, and this effect is only partially mitigated by Cochrane-Orcutt correction for autocorrelation. The results are extended to show that pairs of detrended random walks exhibit spurious correlation.

KEY WORDS: Detrending, Regression, Random Walk, Autocorrelation, Spurious Correlation.

Charles R. Nelson
Department of Economics
University of Washington
Seattle, WA 98105

Heejoon Kang
Deapartment of Business
Economics and Public Policy
School of Business
Indiana University
Bloomington, IN 47405

1. Introduction

A stationary time series is one which fluctuates around a mean value. More technically, it has a mean value, variance and autocovariances which are finite and constant through time. Many economic time series, however, are clearly nonstationary in the sense that they tend to depart ever farther from any given value as time goes on. When movement in a series appears to be predominantly in one direction it is often said to exhibit "trend." In applied work trend is often attributed to a functional dependence on time. Accordingly, nonstationary time series are frequently "detrended" by regressing the series on time or a function of time. The residuals are then treated as a stationary time series with well defined variance, autocovariances and covariances with other detrended variables. The model or representation implicit in these procedures is

$$(1.1) \qquad Y_t = f(t) + u_t$$

where $\{Y_t\}$ is an observed nonstationary time series and $\{u_t\}$ is the stationary series of deviations around the trend function f(t). In the case of a linear trend, f(t) has the form $\alpha + \beta t$.

An alternative hypothesis consistent with nonstationarity is the one popularized by Box and Jenkins (1970), namely that the observed series represents the accumulation (integration) of changes or first differences which are a stationary time series. A series of this type evolves according to the relation

(1.2) $$Y_t = Y_{t-1} + \beta + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a stationary series with mean zero and constant variance σ_ε^2 and β is the (fixed) mean of the first differences often called the drift. The simplest member of this class is the random walk in which the steps, ε_t, are serially random. Accumulating changes in Y from any initial value, say Y_0, we have

(1.3) $$Y_t = Y_0 + \beta t + \sum_{i=1}^{t} \varepsilon_i$$

which has the same form as a linear version of (1.1). It is fundamentally different, however. The intercept is not a fixed parameter but rather depends on the initial value Y_0. The disturbance is not stationary but rather the variance and autocovariances depend on time. In the random case the variance is just $t\sigma_\varepsilon^2$. The width of a confidence interval for future values of Y in (1.1) is limited in width by the finite dispersion of u, while in (1.3) it increases without bound (in proportion to $\sqrt{t}$ for a random walk with normally distributed errors). We follow Nelson and Plosser (1982) in referring to the first class of models (1.1) as trend stationary processes (TSP) and the second class (1.2) as difference stationary processes (DSP).

A test for the hypothesis that a time series belongs to the DSP class against the alternative that it belongs to the

TSP class has been developed by Dickey and Fuller (1979). In the simplest case one estimates by least squares the coefficients in the model.

$$(1.4) \qquad Y_t = \alpha + \rho Y_{t-1} + \beta t + \varepsilon_t,$$

which belongs to the DSP class if $\rho = 1$ and $\beta = 0$ and belongs to the TSP class if $|\rho| < 1$. Dickey and Fuller show that the least squares estimate of ρ is not distributed around unity under the DSP hypothesis but rather around a value less than one. The negative bias diminishes with the number of observations. Nelson and Plosser applied this testing procedure to a wide range of historical time series for the U.S. economy and found that the DSP hypothesis was accepted in all cases with the exception of the unemployment rate which, not surprisingly, appears to be stationary and population which was not consistent with linear versions of either hypothesis. Further, many variables such as real GNP and employment were found to be reasonably characterized as random walks with drift.

In this paper we explore the pitfalls inherent in regression models in which time is included as an explanatory variable under the TSP hypothesis when in fact the time series we are interested in explaining belongs to the DSP class. We take as our benchmark the case where the series is generated by a random walk. In Section 2 of the paper we investigate the properties of standard regression statistics including R^2, t-ratios and sample autocorrelations of residuals when time is the only explanatory variable as it is in "detrending" regressions. In Section 3 we extend the investigation to regressions which include other explanatory variables. Particular attention is focused on the danger of spurious regression relationships among variables which are unrelated.

2. Goodness-of-Fit and Residual Variance in Linear Time Trend Regressions

The simplest example of regression on time and the one most often encountered in practice is the linear "detrending" regression

$$Y_t = \alpha + \beta t + u_t \tag{2.1}$$

in which the nonstationarity in Y is assumed to be explained by a linear dependence on time with the remaining variation in Y being due to a stationary "cyclical" component $\{u_t\}$. If the series is one which grows exponentially it is often transformed to natural logarithms prior to detrending. Regression (2.1) is properly specified if the series belongs to the TSP class and account is taken of autocorrelation in the disturbances prior to drawing inferences about β. However, if Y belongs to the DSP class then the disturbance in (2.1) is not stationary and the appropriate transformation to produce a stationary series would be differencing. Suppose one nevertheless fits the linear trend to a DSP series. What can we say about the properties of the usual regression statistics and the resulting "detrended" data?

We begin by noting that a DSP has the form of a linear regression with cumulative errors, as in (1.3). The initial value Y_0 may be regarded as an unknown parameter, say α, so we have

$$(2.2) \qquad Y_t = \alpha + \beta t + \sum_{i=1}^{t} \varepsilon_i; \; t=1,\ldots,N.$$

In this context, the realization $(Y_1,\ldots,Y_N)$ is thought of as resulting from a random drawing from the joint distribution of disturbances $(\varepsilon_1,\ldots,\varepsilon_N)$. This joint distribution implies a joint distribution for the cumulative errors in (2.2). These regression errors have mean zero and a N x N covariance matrix, say Ω. The properties of the ordinary least squares (OLS) coefficients $\hat{\alpha}$ and $\hat{\beta}$ and the resulting residuals are of considerable practical interest since it is these residuals which are often interpreted as "detrended" data in applied work.

Standard least squares regression theory provides us with some immediate and useful results, namely (1) $\hat{\alpha}$ and $\hat{\beta}$ are unbiased (2) sampling errors $(\hat{\alpha}-\alpha)$ and $(\hat{\beta}-\beta)$ in a given sample depend only on the ε's underlying that sample and not on α or β (3) the particular residuals, or detrended data, obtained from a given sample depend only on the ε's and not on α or β. These properties imply that we can investigate the distributions of $(\hat{\alpha}-\alpha)$, $(\hat{\beta}-\beta)$, the detrended data, and derived statistics such as standard errors and t-ratios by Monte Carlo methods without being concerned that our results depend on arbitrarily chosen values of α and β.

To make our investigation operational we need to choose a particular DSP process as an archtype and our choice is the random walk for which the ε are independent and identically distributed. The random walk is the simplest member of the DSP class and also provides a reasonable characterization of many economic time series. For this case the elements of Ω are just $\Omega_{j,k} = \min(j,k)\ \sigma_\varepsilon^2$ so that Ω is completely determined by the number of observations and σ_ε^2.

The covariances between errors in this model are all positive and standard analysis of regression with positively correlated disturbances suggests that conventional standard errors and t-statistics will be misleading by overstating significance of coefficient estimates. Correspondingly, R^2 will exaggerate the extent to which movement of the data is actually accounted for by time. This is analogous to the spurious regression phenomenon discussed by Granger and Newbold (1974) in the context of pairs of stochastic time series. To see how the spurious regression phenomenon works in our situation, consider the case where $\beta = 0$ so that the level of Y does not have any functional dependence on time. The population value of R^2 is zero since time does not in fact account for any of the variation in Y. If we nevertheless run the regression of Y on time we will obtain a sample value of R^2 which is given by

(2.3) $$R^2 = 1 - (SSE/SST)$$

where SSE is the conventional error sum of squares and SST the total sum of squares. Using the <u>true</u> parameters we would have SSE = SST and $R^2 = 0$. As we shall see, sample R^2 will be wrong on both counts, that is both SSE and SST are distorted in the sample.

Taking the SST part of R^2 first, we note that by definition the sample variance of N observations is

$$\text{(2.4)} \qquad SST/(N-1) \equiv (N-1)^{-1} (\textstyle\sum Y^2) - [N(N-1)]^{-1} (\textstyle\sum Y)^2.$$

For model (2.2) with $\beta = 0$ the expected value of the first term is readily shown to be

$$\text{(2.5)} \qquad E[(N-1)^{-1}(\textstyle\sum Y^2) = [N(N+1)/2(N-1)]\sigma_\varepsilon^{\ 2} \approx (N/2)\sigma_\varepsilon^{\ 2}$$

which is the correct average variance of Y over the sample since the variances of Y_1, Y_2, Y_N are $\sigma_\varepsilon^{\ 2}$, $2\sigma_\varepsilon^{\ 2}, \ldots, N\sigma_\varepsilon^{\ 2}$ respectively. The distortion of SST comes with the second term which arises from using the sample mean of Y in calculating SST since

$$\text{(2.6)} \qquad E(\textstyle\sum Y)^2 = (1 + 2^2 + \ldots + N^2)\sigma_\varepsilon^{\ 2}$$

$$= [N(N+1)(2N+1)/6]\sigma_\varepsilon^{\ 2}$$

which is of order N^3 instead of order N as would be the case in a random sampling situation. Thus we have

$$\text{(2.7)} \qquad E[(\textstyle\sum Y)^2/(N-1)N] = [(N+1)(2N+1)/6(N-1)]\sigma_\varepsilon^{\ 2}$$

which combined with (2.5) gives us

(2.8) $E[SST/(N-1)] = [(N+1)/6]\sigma_{\varepsilon}^{2}.$

Comparing (2.8) with (2.5) we see that the effect of using the sample mean of Y instead of the true mean, zero, in calculating the sample variance of Y is to cut the measured variation in the data on average by two-thirds.

In studying the behavior of SSE we are able to make use of a formula given in Nelson and Kang (1981), based on work by Chan, Hayya, and Ord (1977) for the expected values of the autocovariances of residuals from regression of a random walk on time. This result does not depend on whether the mean of the changes, β, is zero or not. Recognizing that the variance of the residuals is their autocovariance at lag zero, equation (2.1) of Nelson and Kang (1981) yields the approximation

(2.9) $E[SSE/(N-1)] \approx [(N+1)/15]\sigma_{\varepsilon}^{2}.$

Comparing this result with (2.5) which gives the true variance of Y for a realization of length N, we see that the variance of "detrended" data is only 2/15 as large. _Thus, even when no trend or drift is present, regression detrending will remove about 86 percent of the variation from the original data_.

_If_ it were true that $E(R^2) = 1 - E(SSE)/E(SST)$ instead of $1 - E(SSE/SST)$ we could combine results (2.8) and (2.9) to obtain $E(R^2) = 3/5 = .6$ for regression of a random walk with zero drift on time. As we shall see from the results of sampling experiments, $E(R^2)$ is not that large, but rather around .44 due to the difference between E(SSE)/E(SST) and E(SSE/SST). The fact that E(SSE) and E(SST) are both of order N <u>does</u> suggest that $E(R^2)$ does not depend importantly on sample size, a result which is confirmed by sampling experiments. To explore the sampling characteristics of sample R^2, t-ratios, residual variance, and residual autocorrelation we have performed a set of experiments in which the trend model was fitted by OLS to 1000 independent replications of a random walk with $\alpha = 0$, $\beta = 0$ and ε_t i.i.d. N(0,1). The basic results for N = 100 observations are presented in Table 1. Only the results for R^2 and the mean values of $\hat{\alpha}$ and $\hat{\beta}$ depend on α and β.

Note first that the mean sample R^2 is .443 although Y does not in fact depend on t. The R^2 obtained in each realization is a lower bound on the R^2 which would have been obtained in the multiple OLS regression which included other independent variables. The minimum observed value of R^2 was .000+ and the maximum .978. This occurs in spite of the fact that $\hat{\beta}$ is unbiased.

The mean sample variance of residuals, [SSE/(N-1)], is 6.79 which agrees closely with the value given by approximation (2.9) for N = 100 and $\sigma_\varepsilon^2 = 1$, namely 6.73. Since the true variance of Y averaged over the sample is 50, the detrending procedure has indeed removed 86 percent of the variation in the data. This attenuation effect does not depend on σ_ε^2.

The spurious regression phenomenon is further reflected in the large dispersion of t-ratios for $\hat{\beta}$ and $\hat{\alpha}$, their standard deviations being about 15 and 7 times respectively what they would be in a properly specified regression model. As a result, the true null hypotheses $\beta = 0$ and $\alpha = 0$ are rejected with frequencies of 87 percent and 80 percent respectively at a nominal 5 percent significance level.

Earlier papers by Chan, Hayya and Ord (1977) and Nelson and Kang (1981) discuss the behavior of sample autocorrelations of residuals from regression of a random walk on time. An approximation is developed in which the expected sample autocorrelation for a given lag is a function only of the ratio of the lag to the length of the series except for terms of order N^{-1} and smaller. For example, at lag one the expected autocorrelation is roughly (1-10/N). Thus the shape of the sample autocorrelation function is effectively an artifact of detrending with the value of the autocorrelation for given lag depending on the particular sample length. Further, the function is shown by Nelson and Kang to resemble a damped sine wave as can be seen in Figure 1 where the function is plotted for N = 101. This implies pseudo-periodic behavior in the detrended series where none is present in the underlying data. At lags which are small relative to sample size, however, the function declines roughly exponentially. These results do not depend on the true value of β or of σ_{ε}^2.

In our experimental case with N = 100 the empirical means of the first three sample autocorrelations of residuals, denoted r_1, r_2 and r_3 in Table 1 are .88, .77, and .68 respectively. The values predicted by the approximation formula, .91, .82 and .74,

are somewhat too high. The empirical mean values do decline roughly exponentially. This suggests that an investigator versed in time series model identification would typically specify a stationary first order AR process for this "detrended" series even though the original data are not stationary around a trend. This conjecture is supported by values of the partial autocorrelations at lags two and three, denoted r_{22} and r_{33} in Table 1. The means are -.04 and -.02 respectively which should be compared with a standard error of $1/\sqrt{N}$ or .100 in this case. We conclude that an investigator who takes the time trend model seriously is likely to specify an AR(1) process for the residuals with a coefficient typically around .88 in the case of 100 observations or in general about (1-10/N).

To check the sensitivity of our results to sample size we also ran the experiment for N = 20 observations, again with 1000 replications. Mean R^2 was nearly the same, .435, as expected. The mean sample variance of residuals was 1.40, so again detrending removed about 86 percent of the variation in the data. In accordance with the smaller number of observations, t-ratios are reduced in absolute size with standard deviations falling to about 6 and 3 for

$\hat{\beta}$ and $\hat{\alpha}$ respectively, the true null hypotheses being rejected 73 percent and 54 percent of the time respectively at a 5 percent nominal significance level. The means of r_1, r_2 and r_3 were .50 and .17 and -.04 compared with corresponding values from the approximation formula of .59, .28 and .05. Thus with only 20 observations the oscillatory character of the autocorrelation function becomes more evident at low lags. This is further evident in the partial autocorrelations r_{22} and r_{33} which have mean values -.18 and -.13 though the appropriate standard error is now .22. If a second order autoregression is specified for the residuals the coefficients implied by sample autocorrelations indicate complex roots (oscillatory behavior) in 69 percent of the cases. By comparison, complex roots were indicated in less than 1 percent of the cases for N = 100.

An alert investigator would presumably recognize the presence of autocorrelation in the trend regression residuals and would rerun the regression after correction for autocorrelation if the objective was to test for the presence of trend rather than simply detrending. Under the hypothesis of AR(1) errors, the appropriately transformed regression would be

$$(2.10) \qquad (Y_t - \rho Y_{t-1}) = \alpha(1-\rho)+\beta[t-\rho(t-1)]+(u_t-\rho u_{t-1}),$$

where ρ is the AR coefficient for the OLS residuals. Indeed, (2.10) would be a correctly specified regression if ρ were set at unity which corresponds to first differencing.

However, an investigator who believed the errors were stationary would presumably use the value of r_1 from the OLS residuals. Estimation of (2.10) with r_1 in place of ρ amounts to a two-step Cochrane-Orcutt (1949) procedure. The transformation of course under-corrects for autocorrelation since r_1 is typically around .88 for N = 100. As a result the spurious regression phenomenon by no means disappears. In the same set of experiments with N = 100 the standard deviation of the t-ratio for the trend coefficient $\hat{\beta}$ dropped only to 4.57 after transformation, with the true null hypothesis being rejected in 58 percent of the realizations at a nominal 5 percent level. Similarly, the standard deviation of $t(\hat{\alpha})$ fell to 2.67 and the rejection frequency to 45 percent. The mean R^2 in these regressions was .122. The residuals in the transformed regressions give no hint of trouble: the mean Durbin-Watson statistic was 1.88 corresponding to a mean autocorrelation at lag one of only .05.[1]

The Cochrane-Orcutt procedure is often iterated using successive estimates of (α,β) and ρ until convergence occurs. In the special case of time trend regression, convergence effectively occurs at the first iteration. The algebraic reasons for this are explored in the Appendix. This result is entirely a result of the algebra of least squares and the special nature of the time trend variable and is not dependent on the data being generated by a TSP, DSP, or whatever.

To sum up, an investigator following standard methodology will tend to find evidence of a trend in random walk data when none is in fact present. This is essentially because testing for trend in the regression framework takes as a maintained hypothesis the assumption that the time series is stationary, apart from a deterministic trend (if any). Since trend lines fit random walk data well in an *ex post* sense, the detrending procedure will tend to remove much variation in the data which is in fact stochastic rather than deterministic. As Nelson and Kang (1981) have demonstrated, the autocorrelation properties of the detrended data will be dominated by the effects of detrending, thereby obscuring evidence of any actual autocorrelation in first differences. We conclude that the appropriate test for trend developed by Dickey and Fuller (1979) should be conducted prior to analysis of nonstationary series. At a minimum, the alternative that the series are stationary in differences should be considered.

It is doubtless true that many economic time series such as output, sales, and prices do in fact exhibit a positive rate of drift which would mean $\beta > 0$ in our simple model. If this is the case, does it mitigate any of our objections to time trend regression? Not at all. As noted previously, the properties of estimation errors $(\hat{\alpha}-\alpha)$ and $(\hat{\beta}-\beta)$, the residuals, and statistics constructed from them do not depend on β; residual variance will understate the true stochastic variation in the data to the same degree, residual autocorrelations will display the same pseudo-periodicity, and conventional measures of significance will be overstated. R^2 will still reflect the spurious regression phenomenon. The behavior of R^2 in the case $\beta > 0$ is illuminated by considering the effect of β on SSE or SST. While SSE does not depend on β at all, E(SST) is augmented by a term which is a function of β^2 and N and is of order N^3. The ratio E(SSE)/E(SST) is then of order N^{-1} implying that R^2 will tend toward one as N increases in the case $\beta \neq 0$, instead of toward a constant. Denoting by $\tilde{R}^2$ the expression [1-E(SSE)/E(SST)], it is easy to show that

$$(2.11) \qquad \tilde{R}^2 = [\tilde{R}_0{}^2 + (N/2)(\beta^2/\sigma_\varepsilon{}^2)]/[1+(N/2)(\beta^2/\sigma_\varepsilon{}^2)]$$

where $\tilde{R}_0{}^2$ is the previously derived value of $\tilde{R}^2$ for $\beta = 0$. Note that $\tilde{R}^2$ exceeds $\tilde{R}_0{}^2$ if $\beta \neq 0$, the difference increasing with the ratio $(\beta^2/\sigma_\varepsilon{}^2)$ and N. It is the case for postwar annual U.S. real GNP in logs that β and σ_ε are roughly equal. For $\beta^2 = \sigma_\varepsilon{}^2$ we have

$$(2.12) \qquad \tilde{R}^2 = [\tilde{R}_0{}^2 + (N/2)]/(1+N/2)$$

which is always greater than N/(N+2). This suggests that for economic time series R^2 will be driven close to unity very rapidly as N increases. The high values of R^2 reported in the literature for regressions which include time are therefore not surprising. Obviously, R^2 gives us little useful information about the explanatory power of the independent variables in such regressions.

To check on the validity of (2.11) as an approximation we reran our Monte Carlo experiments setting $\beta^2 = \sigma_\varepsilon^2$ and N at 20 and 100. The mean sample R^2 in the case N = 20 was .953 compared with the .949 implied by equation (2.12) with $\tilde{R}_0^2$ replaced by the Monte Carlo mean for $\beta = 0$. At N = 100 the mean R^2 rose to .992 which compares with .989 predicted by (2.12).

An example from the literature may help to emphasize some of the points made in this section. Perloff and Wachter (1979, Table 3) report a series of regressions of aggregate output in the U.S. on time, time squared, time cubed and an index of labor and capital inputs constructed under various assumptions about the form of the aggregate production function. The polynomial trend function is intended to capture the effects of technological change. The data cover the 92 quarters from 1955 through 1977. The reported R^2 values are about .995. Using (2.12), we would expect an R^2 of about .988 in a regression of output on time <u>alone</u> if output were a random walk and did not depend on the index of inputs, assuming $(\beta/\sigma_\varepsilon) = 1$ and taking $\tilde{R}_0^2 = .44$. Thus the reported R^2 in itself provides no real information about the relationship. The t-ratios for time squared and time cubed are not significant but for time they are in the range of 8 to 10 which is within one standard deviation of zero for the results reported in our Table 1 for the case of a random walk with zero drift. The t-ratios for the input index are ten or larger, however. This suggests that the upward drift in output may be largely due to the upward drift in the index of inputs. However, Perloff and Wachter warn that "because of autocorrelation, these statistical tests should be viewed with caution" (p. 130, fn. 25). Indeed, the Durbin-Watson statistic for the preferred model is .231 corresponding to an r_1 of .884, very close to the mean values recorded in our Table 1 for a slightly larger number of observations. Therefore, we interpret the Perloff and Wachter results as being consistent with the hypothesis that the contribution of technology (inputs other than labor and capital) to output is not stationary around a deterministic trend but rather is a nonstationary stochastic process akin to a random walk. If this hypothesis is correct, then technological change occurs in an irregular stochastic fashion rather than in a smooth deterministic one.

3. Spurious Regression Relationships Between Nonstationary Variables Resulting From Inappropriate Use of Time as an Independent Variable

In situations where it is the relationship between nonstationary variables which is of primary interest one often encounters regression equations of the form

$$Y_t = \alpha + \beta t + \gamma X_t + u_t \tag{3.1}$$

where $\{Y_t\}$ is a nonstationary variable, such as output, $\{X_t\}$ is a nonstationary independent variable (or set of such variables), such as a production input, and $\{u_t\}$ is a sequence of disturbances. The role of time is to account for growth in Y not attributable to X, for example the impact of technological change on output. The parameters of (3.1) are estimated by ordinary or generalized least squares procedures which make the assumption that the errors $\{u_t\}$ are stationary. Equivalently, another way to state this assumption that the part of Y is not explained by X, $(Y_t - \gamma X_t)$ is a TSP variable. If this quantity is instead a DSP variable then first differencing of the relationship, that is

$$(3.2) \qquad \Delta Y_t = \beta + \gamma(\Delta X_t) + \varepsilon_t,$$

would put it in the form suitable for estimation by standard procedures. What we are interested in here is the consequence of estimating the relationship in levels (3.1) when in fact the differenced relationship (3.2) is the one which has stationary disturbances.

We begin by noting as in Section 2 that a relationship between levels of the variables including time is obtained by accumulating changes given by (3.2) from an arbitrary time period zero which gives

$$(3.3) \qquad Y_t = Y_0 + \beta t + \gamma X_t + \sum_{i=1}^{t} \varepsilon_i .$$

This equation has the same form as (3.1) except that the disturbance is cumulative rather than stationary and the intercept is given by the arbitrary initial value Y_0 rather than being a fixed parameter. Standard results in regression theory tell us that OLS estimates of β and γ in

(3.3) will be unbiased but inefficient since the disturbances in (3.3) will be correlated across time periods. In the case that the ε_t are serially random then the disturbance in (3.3) is a random walk. The covariance matrix Ω for these N disturbances would then have elements $\Omega_{j,k} = \min(j,k)\sigma_\varepsilon^2$; $j,k = 1,\ldots,N$, and an appropriate GLS estimator could be constructed for Y_0, β, and γ which would recognize this covariance structure. Alternatively, the equation could be estimated efficiently in first differences by OLS, which is much simpler.

Estimation of γ by OLS in levels will be subject to the spurious regression phenomenon discussed by Granger and Newbold (1974). The classical formula will understate the sampling variance of $\hat{\gamma}$ and therefore overstate its significance. A heuristic explanation of this is as follows. Lovell (1963) showed that the OLS coefficient $\hat{\gamma}$ obtained in a regression of Y on time and X is numerically identical to the OLS coefficient obtained in the regression of detrended Y on detrended X. The detrended values of any time series may be expressed as the product of an NxN matrix, say T, times the vector of observations for the series. Note that T depends only on N and is the same for all series of length N. Multiplying (3.3) in vector form through by T we have

$$TY = \gamma TX + Tu \tag{3.4}$$

where Y, X and u denote column vectors of the respective time series and we use the fact that the detrended values of the intercept variable and time are identically zero. Using tildes to denote the detrended variables, we have

$$(3.5) \qquad \tilde{Y}_t = \gamma\tilde{X}_t + \tilde{u}_t.$$

If $\{X_t\}$ and $\{u_t\}$ are both random walks, then $\{Y_t\}$ is also a random walk and (3.5) is the regression of a detrended random walk on another detrended random walk with a disturbance which is also a detrended random walk. In short, estimating γ by least squares in (3.3) is equivalent to a regression where the independent variable and the error term have the same autocorrelation function. A standard textbook result, originally due to Wold (1953), is that the effect of autocorrelation in regression errors is to inflate the variance of the OLS coefficient by a factor of the form $(1 + \sum\rho_i r_i)$ where ρ_i is the autocorrelation of the disturbance at lag i and r_i is the sample autocorrelation of the independent variable at lag i. Since ρ_i and r_i will tend to be the same in our situation, the variance of $\hat{\gamma}$ will be larger than the classical formula would indicate. Further, since these autocorrelation coefficients become larger with sample size (recall that it is roughly (1-10/N) at lag 1), the spurious regression effect will be <u>more</u> pronounced in larger samples.

To get an idea of the magnitude of the spurious regression phenomenon in this situation for sample sizes typically encountered in economics we have conducted a Monte Carlo experiment corresponding to that reported in Table 1 but with the addition of an independent variable which is a random walk. Since the spurious regression phenomenon depends on the random walk nature of the variables and not on specific parameter values we have set β and γ equal to zero so that the R^2 obtained provides a lower bound for the general case. $\{X_t\}$ and $\{u_t\}$ are zero-drift random walks with unit variance Normal disturbances. The primary results are given in Table 2. As in the detrending regressions, only the means of $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ and values of R^2 depend on α, β and γ.

Note that the empirical mean of the conventional standard error for $\hat{\gamma}$ is about one fourth as large as the empirical standard deviation of $\hat{\gamma}$, implying that the precision of the estimate of γ will be greatly overstated if serial correlation in the regression errors is ignored. Correspondingly, the empirical standard deviation of the conventional t-ratio is too large by a factor of about 4. If a t-table is used to assess significance, $\hat{\gamma}$ is significant at the 5 percent level or better 64 percent of the time and at the 1 percent level or better 55 percent of the time. The spurious relationship with time is somewhat reduced by the inclusion of X, as one would expect, but the rejection frequency for the true null hypothesis is nevertheless 83 percent at a nominal 5 percent level, compared with the 87 percent frequency we obtained when time was the only regressor.

The values of R^2 are of course higher than in the simple time trend regressions of Section 2 since inclusion of an additional variable cannot lower R^2 in a given sample. The empirical mean rises to .501 from .443. Thus, time and a random walk variable will typically explain about 50 percent of the variation in a random walk which is in fact unrelated to either. This further reinforces our conclusion in Section 2 that R^2 values are highly misleading or at least uninformative in regressions involving time as a variable. The increase in R^2 due to inclusion of X will of course typically appear "significant" by conventional criteria since the F-test used to assess its significance under classical regression assumptions will correspond to the t-test for $\hat{\gamma}$.

Finally, we note that the empirical means of sample autocorrelations of the residuals reported in Table 2 are somewhat smaller than those reported in Table 1 in the absence of X, for example at lag one we obtain .852 instead of .883. The mean Durbin-Watson statistic is correspondingly somewhat larger, .260 compared with .198. Nevertheless, an alert investigator would again typically reject the hypothesis of serially random errors. Believing the regression disturbances to be stationary, and noting the roughly exponential decline of the sample autocorrelations in a typical realization, our imaginary investigator would presumably follow the popular procedure of assuming a first order AR process for the errors and use r_1 as an estimate of the AR coefficient. The transformed regression equation would be

$$(3.6) \qquad (Y_t - r_1 Y_{t-1}) = \alpha(1-r_1) + \beta(t - r_1(t-1)) + \gamma(X_t - r_1 X_{t-1}) + (u_t - r_1 u_{t-1}).$$

Now (3.6) would be a properly specified classical regression if r_1 were set at unity (corresponding to first differences), since $(u_t - u_{t-1})$ is indeed random in our situation. Sample values of r_1 are, however, rarely close to unity since the empirical standard deviation is only .064 around the mean of .852. The transformed regression would still suffer from the problem of nonrandom, indeed nonstationary, disturbances.

The corresponding results for the transformed regressions (3.6) using a first round estimate of r_1 are reported in Table 3. Note that the empirical standard deviation of the t-ratio for $\hat{\gamma}$ is reduced by the transformation from 4.490 to 1.287, although the standard deviation for a t-distributed variable with 96 degrees of freedom is 1.010. Thus, the frequency of rejection of the true null hypothesis on γ is still too large, in particular 11.3 percent at a nominal 5 percent level and 5.4 percent at a nominal 1 percent level. This overstatement of significance is reflected in the comparison of the empirical standard deviation of $\hat{\gamma}$, .126, with the mean standard error,

.100. The spurious relationship of Y to time is also diminished by the transformation although it is still very strong. The empirical standard deviation of the t-ratio for $\hat{\beta}$ is 4.69. Rejection frequencies are 60 percent at a nominal 5 percent level and 51.5 percent at a nominal 1 percent level.

The autocorrelation coefficients of the residuals in the transformed regressions are of course considerably diminished relative to those obtained for the original regressions, averaging only .079 at lag one with a standard deviation of .097. An investigator using the standard error $1/\sqrt{N}$ = .100 would rarely reject the hypothesis of serially random errors. The mean Durbin-Watson statistic of 1.820 is similarly well inside the acceptance region.[2] We conclude then that an investigator who believed the disturbances in the levels regression to be stationary would typically find the results of the transformed regression to be satisfactory in the sense of passing the usual tests of random errors and therefore run a substantial risk of finding a significant relationship between Y and X where none exists.

Unlike the case of the detrending regression, however, continued iteration of the Cochrane-Orcutt procedure does alter the estimates and improve their properties. The frequency of rejection of the true null hypothesis on γ drops to 6.7 percent at a nominal 5 percent level and to 2.5 percent at a nominal one percent level, reflecting the fact that estimated standard errors are closer to the actual standard deviation. The effect on inference about β is less dramatic. The frequencies of rejection fall only to 51.5 percent and 39.8 percent for the 5 percent and 1 percent tests respectively.

4. Summary and Conclusions

It is common practice in applied regression to attribute nonstationarity or "trend" in a time series to a functional dependence on time with the remaining variation in the series assumed to be stationary. In this paper we have considered the consequences of such an assumption when the time series is not stationary around a function of time but rather is stationary in first differences. The results reported by Nelson and Plosser (1982) are consistent with this hypothesis for a wide range of economic variables. The prototypical example of such nonstationary stochastic processes is the random walk process which forms the basis for our exploratory analysis.

The primary findings are as follows.

1) Regression of a random walk on time by least squares will produce R^2 values of around .44 regardless of sample size when in fact the variable has no dependence on time whatever (zero drift). For random walks with drift the R^2 will be higher and will increase with sample size, reaching one in the limit regardless of the actual rate of drift of the series or its variability.

2) Residuals from regression of a random walk on time will have a variance that on average is only about 14 percent of the true stochastic variance of the series . This result holds regardless of sample length or the true rate of drift (mean rate of change) of the series. If these residuals are mistakenly interpreted as a "detrended" series, then their variance will greatly understate the actual variance of the series. Equivalently, stochastic variation is mistakenly attributed to dependence on time which is present in only an ex post sense, not an ex ante one.

3) The mean values of sample autocorrelations of a "detrended" random walk are a function of sample length, being roughly (1-10/N) at lag one for example, and therefore are purely artifactual. Since the function oscillates with a period of roughly (2/3)N the detrended data will appear to exhibit a long cycle which is spurious. Nelson and Kang (1981) showed that this result is quite robust with respect to serial correlation in the first differences of the series.

4) A conventional t-statistic for the least squares coefficient of time is a very poor test for the presence of trend in the sense of a dependence on time. Such tests lead to rejection of the null hypothesis of no dependence in 87 percent of the cases for a sample

length of 100 at a nominal 5 percent level when in fact there is no dependence on time. Attempts to correct for serial correlation in the residuals only partially corrects this effect. An investigator applying a first order AR correction based on sample autocorrelations of the residuals would still reject the true hypothesis at a nominal 5 percent level with 73 percent probability. The correct procedure would be to take first differences in which case the size of the test would be correct.

5) Regression of one random walk on another, with time included to account for trend, is strongly subject to the spurious regression phenomenon. That is, a conventional t-test will tend to indicate a relationship between the variables when none is present. Since such regressions can be thought of as regression of one "detrended" random walk on another, the phenomenon can be viewed in the framework developed by Wold (1953) for stationary autocorrelated series. For the case of zero drift, unrelated random walks, the true null hypothesis of no relationship is rejected with a frequency of 64 percent at a nominal 5 percent level using 100 observations. The true null hypothesis of no dependence on time is correspondingly rejected with an 83 percent frequency. Attempts to correct for serial correlation on the assumption

that the disturbances are stationary and first order autoregressive only partially alleviate the problem; the rejection frequencies drop to 11.3 percent and 60 percent respectively. Continued iteration reduces these to 6.7 percent and 51.5 percent respectively. First differencing would of course correct the size of the tests.

Our advice to practitioners based on this investigation is to regard stationarity around a function of time as a tentative rather than a maintained hypothesis. It is certainly not a harmless assumption, but rather one fraught with potential pitfalls. Regression models involving non-stationary time series should be estimated in differenced form and the results carefully compared with those for the regression in levels with time as an explanatory variable. Two recent papers by Plosser and Schwert (1977, 1978) suggest that the consequences of differencing when it is not needed to achieve stationarity are much less costly in the present context than those of failing to difference when it is appropriate. In addition, the tests for stationarity in differences as opposed to stationary around a trend line developed by Dickey and Fuller (1979) are strongly recommended as a guide to the appropriate transformation of time series data.

ACKNOWLEDGEMENTS

The authors wish to thank Dennis F. Kraft for thoughtful comments and useful suggestions for improvement. This research was supported by the National Science Foundation under grants SES-7906948 and SES-8112687.

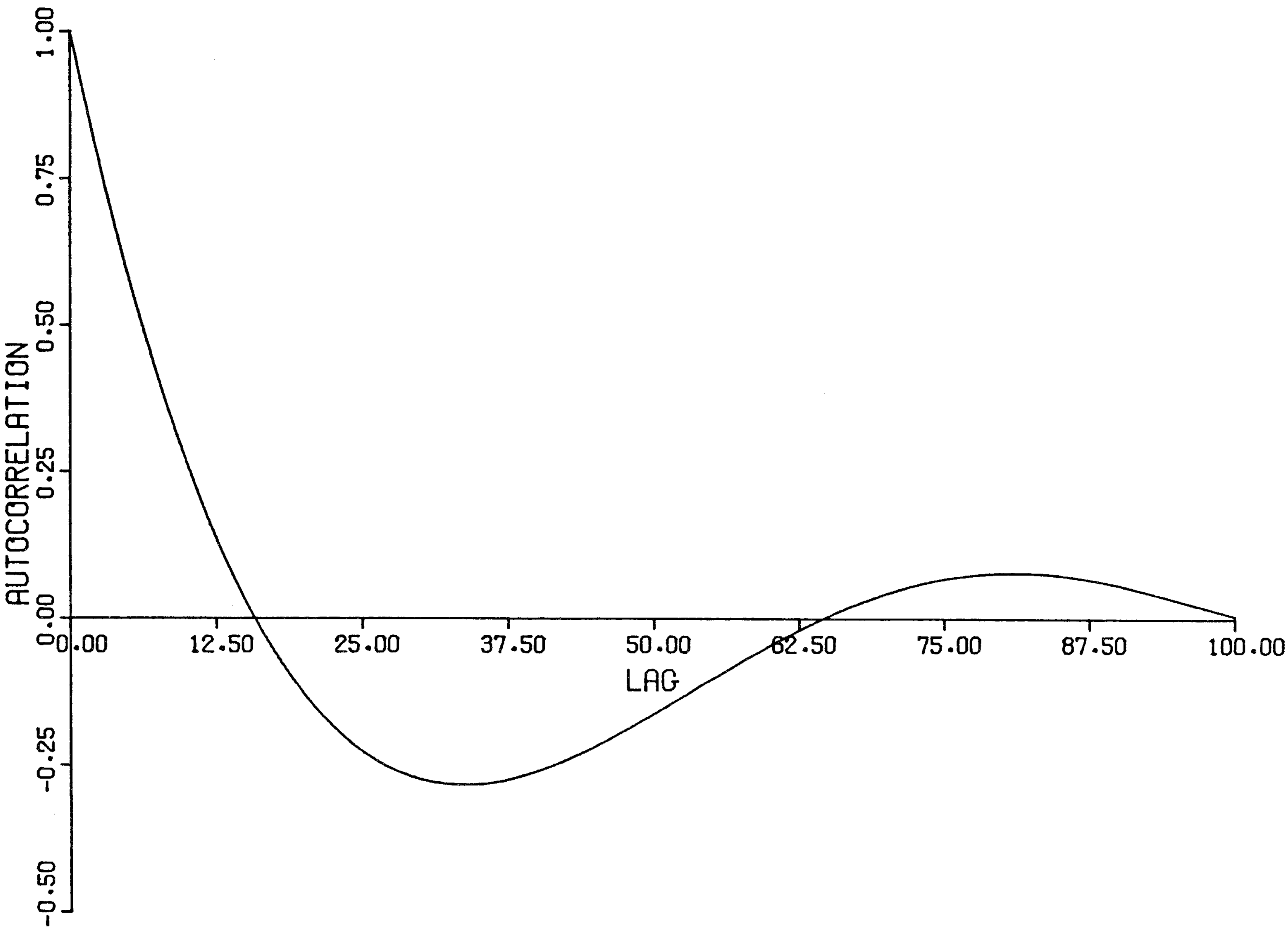

FIG. 1 THEORETICAL AUTOCORRELATIONS FOR DETRENDED RANDOM WALK

TABLE 1. Empirical Moments of Summary Statistics for OLS Regression of a Zero-Drift Random Walk on Time With One Hundred Observations, Based on One Thousand Replications: $Y_t = \hat{\alpha} + \hat{\beta}t + \hat{u}_t$.

Statistic	Mean	Variance	Standard Deviation
R^2	.443	.092	.304
SSE/(N-1)	6.7	18.32	4.28
$\hat{\beta}$	-.002	.013	.115
Est. Var. ($\hat{\beta}$)	.81 E-4	.26 E-8	.51 E-4
t($\hat{\beta}$)	-.16	220.2	14.84
$\hat{\alpha}$	.02	13.81	3.72
Est, Var, ($\hat{\alpha}$)	.28	.03	.17
t($\hat{\alpha}$)	-.005	50.41	7.10
D.W.	.198	.011	.104
r_1	.88	.28 E-2	.532 E-1
r_2	.77	.91 E-2	.956 E-1
r_3	.68	.17 E-1	.129
r_{22}	-.04	.01	.09
r_{23}	-.02	.01	.09

NOTES: $\hat{\alpha}$ and $\hat{\beta}$ are computed by OLS and their estimated variances and t-ratios under the (inappropriate) assumptions of the classical linear regression model. In this situation $\hat{\alpha}$ and $\hat{\beta}$ are unbiased (have expected value zero).

TABLE 2: Empirical Moments of Summary Statistics for OLS Regression of a Zero-Drift Random Walk on an Unrelated Zero-Drift Random Walk and Time With One Hundred Observations. Based on One Thousand Replications: $Y_t = \hat{\alpha} + \hat{\beta}t + \hat{\gamma}X_t + \hat{u}_t$.

Statistic	Mean	Variance	Standard Deviation
R^2	.501	.077	.278
SSE/(N-3)	5.830	14.131	3.759
$\hat{\gamma}$	-.011	.177	.421
Std. Error ($\hat{\gamma}$)	.102	.002	.045
t($\hat{\gamma}$)	-.054	20.164	4.490
$\hat{\beta}$	.005	.014	.117
Std. Error ($\hat{\beta}$)	.013	.000+	.007
t($\hat{\beta}$)	.324	133.556	11.557
D.W.	.260	.016	.128
r_1	.852	.004	.064
r_2	.721	.013	.112
r_3	.604	.021	.146

(See Notes to Table 1)

TABLE 3: Empirical Moments of Summary Statistics For Transformed Regressions (3.6) Based on the Same One Thousand Samples Used in Table 2.

Statistic	Mean	Variance	Standard Deviation
R^2	.162	.032	.178
SSE/(N-3)	.970	.024	.153
$\hat{\gamma}$	.003	.016	.126
Std. Error ($\hat{\gamma}$)	.100	.000+	.012
t($\hat{\gamma}$)	.044	1.656	1.287
$\hat{\beta}$	.005	.014	.118
Std. Error ($\hat{\beta}$)	.031	.000+	.014
t($\hat{\beta}$)	.142	22.010	4.692
D.W.	1.820	.038	.195
r_1	.079	.009	.097
r_2	.064	.010	.099
r_3	.052	.010	.098

(See Notes to Table 1).

FOOTNOTES

1. The alert reader will recognize that the Durbin-Watson statistic is not strictly appropriate for testing for random disturbances in regression (2.10) since ρ is estimated from the data. In effect, the model includes a lagged dependent variable, Y_{t-1}. Following Durbin (1970), one could estimate ρ with α and β and compute the h-statistic which takes into account the estimated variance of the estimate of ρ. However, this variance will tend to be small for large values of ρ such as we have in this situation. See the discussion in Durbin (1970, p. 419).
2. See footnote 1.

REFERENCES

Box, George E.P. and Jenkins, Gwilym M. (1970), Time Series Analysis and Control, Forecasting, San Francisco: Holden-Day.

Chan, K. Hung; Hayya, Jack C. and Ord, J. Keith (1977), "A Note on Trend Removal Methods: The Case of Polynomial Regression Versus Variate Differencing," Econometrica 45: 737-744.

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.

Dickey, David A. and Fuller, Wayne A. (1979), "Distribution for the Estimators for Autoregressive Time Series With a Unit Root," Journal of the American Statistical Association 74: 427-431.

Durbin, J. (1970), "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," Econometrica 38: 410-421.

Granger, C.W.J. and Newbold, P. (1974), "Spurious Regressions in Econometrics," Journal of Econometrics 2: 111-120.

Lovell, Michael C. (1963),"Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis," Journal of the American Statistical Association 58: 993-1010.

Nelson, Charles R. and Kang, Heejoon (1981), "Spurious Periodicity in Inappropriately Detrended Time Series," Econometrica 49: 741-751.

Nelson, Charles R. and Plosser, Charles I. (1982), "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," Journal of Monetary Economics 10: 139-162.

Perloff, Jeffrey M. and Wachter, Michael L. (1979), "A Production Function-Nonaccelerating Inflation Approach to Potential Output: Is Measured Potential Output Too High?" Carnegie-Rochester Conference Series on Public Policy 10: 113-164.

Plosser, Charles I. and Schwert, G. William (1977), "Estimation of a Non-Invertible Moving Average Process: The Case of Overdifferencing," Journal of Econometrics 6: 199-224.

___, (1978), "Money, Income, and Sunspots: Measuring Economic Relationships and the Effects of Differencing," Journal of Monetary Economics 4: 637-660.

Wold, Herman (1953), Demand Analysis, New York: John Wiley and Sons.

APPENDIX

RAPID CONVERGENCE OF SLOPE AND AUTOREGRESSIVE PARAMETERS IN LINEAR TIME TREND REGRESSION

The familiar regression model with AR(1) errors is

(A1) $$Y_t = \beta X_t + \varepsilon_t$$

(A2) $$\varepsilon_t = \rho\varepsilon_{t-1} + v_t$$

where Y and X are deviations from means. Following Cochrane and Orcutt (1949), one obtains a first round estimate of β by OLS, say $\hat{\beta}$, and then runs an autoregression on the residuals to get a first round estimate of ρ, say $\hat{\rho}$. The data are then transformed to $(Y_t - \rho Y_{t-1})$ and $(X_t - \rho X_{t-1})$ from which a second round estimate of β is obtained, and so forth. Convergence is effectively immediate in the special case of time trend regression for which $X_t = t$; $t = -n,\ldots,+n$.

Consider first the least squares estimate of β given a value of ρ, say β^*. It can be shown to be

(A3) $$\beta^* = \frac{\sum Y_t X_t + \rho^2 \sum Y_{t-1} X_{t-1} - \rho \sum Y_t X_{t-1} - \rho \sum X_t Y_{t-1}}{\sum X_t^2 + \rho^2 \sum X_{t-1}^2 - 2\rho \sum X_t X_{t-1}}$$

where the summations run from t = - n+1 to +n. Consider dividing each term by $\sum X_t^2$, in which case we have

(A4) $$\beta^* \doteq \frac{\hat{\beta} + (\rho^2 \sum Y_{t-1} X_{t-1} - \rho \sum Y_t X_{t-1} - \rho \sum X_t Y_{t-1}) / \sum X_t^2}{1 + \rho^2 - 2\rho (\sum X_t X_{t-1}) / \sum X_t^2}$$

with the approximation being due to leaving the first observation out of $\hat{\beta}$. Obviously, we can say little in general about the difference between β^* and $\hat{\beta}$, the OLS estimate with $\rho = 0$. This difference depends on autocorrelation in X and cross correlation between Y and X at lags (+1) and (-1). However, in the time trend case we have $X_t = t = X_{t-1} + 1$ which implies

$$\text{(A5)} \qquad \beta^* \doteq \frac{\hat{\beta}+\rho^2\hat{\beta}-2\rho\hat{\beta}}{1+\rho^2-2\rho} = \hat{\beta}.$$

The degree of approximation is the order of difference between $\sum t(t-1)/\sum t^2$ and unity, which is 1/N. Thus the nature of the independent variable in the time trend case implies that the slope estimate is insensitive to the value of ρ.

Similarly, consider estimating ρ on the basis of a given value for β, say $(\hat{\beta}+\Delta\beta)$. In the general case we have

$$\text{(A6)} \qquad \rho^* = \frac{\sum\hat{e}_t\hat{e}_{t-1}-(\Delta\beta)\sum\hat{e}_tX_{t-1}+(\Delta\beta)\sum X_t\hat{e}_{t-1}}{\sum\hat{e}_{t-1}^2+(\Delta\beta)^2\sum X_{t-1}^2-2(\Delta\beta)\sum\hat{e}_{t-1}X_{t-1}}$$

where the summations run from $t = -n+1$ to $+n$ and $\hat{e}$ denotes OLS residuals associated with $\hat{\beta}$. Now dividing each term by $\sum\hat{e}_{t-1}^2$ and noting that the third term in the denominator becomes negligible

$$\text{(A7)} \qquad \rho^* \doteq \frac{\hat{\rho}+(\Delta\beta)(\sum X_t\hat{e}_{t-1}-\sum\hat{e}_tX_{t-1})/\sum\hat{e}_{t-1}^2}{1+(\Delta\beta)^2\sum X_{t-1}^2/\sum\hat{e}_{t-1}^2}$$

which will differ from $\hat{\rho}$ depending on $(\Delta\beta)$, cross correlation between X and $\hat{e}$ at lags (+1) and (-1) and the variance ratio of X and e. In the special case $X_t = t = X_{t-1} + 1$ we see that the approximation reduces to

$$\text{(A8)} \qquad \rho^* = \hat{\rho}/(1+(\Delta\beta)^2 \textstyle\sum X_{t-1}{}^2/\sum e_{t-1}{}^2)$$

since $\sum X_t \hat{e}_{t-1}$ becomes $(\sum X_{t-1}\hat{e}_{t-1} + \sum \hat{e}_{t-1})$ which differs from zero only by end terms (by the algebra of least squares), and similarly for $\sum e_t X_{t-1}$. Since $(\Delta\beta)$ tends to be small in the time trend case the adjustments to $\hat{\rho}$ will be small.

To sum up, iteration of trend line coefficients and autoregressive coefficients for detrended data will be rapid due to the special nature of the independent variable. In our experience, widely-used computer programs for Cochrane-Orcutt have never gone beyond one iteration. This result is entirely due to the algebra of least squares. The true nature of the underlying data is irrelevant, whether it be a TSP, a DSP, or something else. Further the result implies that rapid convergence will occur whenever the independent variables are closely approximated by linear trends. The result also generalizes to higher order autoregressive schemes for the residuals.