October 27, 1976

DATA ERRORS AND DATA DIFFERENCES

Rodney L. Jacobs*
University of California, Los Angeles

Discussion Paper Number 82

*I would like to thank Edward Leamer for comments on an earlier draft. Needless to say, he is absolved from all guilt by association.

I. Introduction

Data differencing has become common practice in many econometric studies. For instance, differencing plays an important role in time series analysis as exemplified by the work of Box and Jenkins. If we are concerned with a system in which some input X_t produces the output Y_t, the time series approach is to determine the cross correlation function between the input and output. That is, we regard X_t and Y_t as being generated by a bivariate stationary stochastic process, and our goal is to identify the correlation function of this process. The identification involves considerable prefiltering of the data because most time series are dominated by some type of trend such as exponential growth and are clearly nonstationary. It is customary to assume that stationarity can be induced by differencing the data a suitable number of times.[1] For example, if the series are dominated by a linear trend then the first differences $\Delta X_t = X_t - X_{t-1}$ and $\Delta Y_t = Y_t - Y_{t-1}$ will approximate stationary processes.

Granger and Newbold have recently demonstrated that differencing the data to obtain a stationary series is desirable for standard regression models. They show that two independent random walks X_t and Y_t will in most instances falsely indicate some regression relationship between X and Y. These regressions are usually typified by low values of the Durbin-Watson statistic and relatively high values of R^2. If the random walks are differenced to obtain stationary processes no regression relationship is found to exist. Granger and Newbold thus argue that regressions with a low value of the Durbin-Watson statistic may well be spurious no matter how high the value of R^2. To guard against spurious regressions, they recommend working with stationary series. Since many economic time series have high serial correlation between adjacent values, the stationarity can usually be induced by taking first differences of the data.

This paper illustrates that inducing stationarity by differencing the data may not be desirable because differencing tends to amplify the influence of observational errors.[2] Such errors may be small relative to levels or systematic (trend) changes in the time series but may not be small relative to the differenced series. Observation errors may, thus, seriously bias empirical results obtained from the differenced data. Since the Cochrane-Orcutt correction for serial correlation involves some degree of data differencing, it also amplifies the biases due to data errors.

II. Theoretical Model

In this section we compare theoretical regression results for levels and first differences of the linear model

(1) $$Y = \beta X + u$$

where Y is a T vector of observations Y_t, X is a T vector of the independent variable X_t, and u is a T vector of errors u_t. The error term is assumed to be

(2) $$u_t = \rho_u u_{t-1} + \varepsilon_t,$$

where ε_t is a white noise process with mean zero and variance σ_ε^2. The series X_t is generated from

(3) $$X_t = \alpha t + v_t,$$

where α is equal to a constant and

(4) $$v_t = \rho_v v_{t-1} + \eta_t,$$

where η_t is distributed independently of ε_t and is white noise with mean zero and variance σ_η^2. We further assume that X_t is measured with some error so that the observations $\hat{X}_t$ become

(5) $$\hat{X}_t = X_t + w_t.$$

The error term w_t is

(6) $$w_t = \rho_w w_{t-1} + \gamma_t,$$

where γ_t is distributed independently of ε_t and η_t and is white noise with mean zero and variance σ_γ^2. Any observation errors on Y_t are, of course, incorporated in the error term u_t.

The model is intended to be illustrative of a large number of economic time series which are composed of trend and movements about that trend. The structure of the model insures that first differences will induce stationarity in both series. In addition, if the values of ρ are close to unity these differenced series will approximate a white noise process. An investigator who wishes to estimate the model of equation (1) may not have a clear idea of the process which generates X or its observation errors. Should he use least squares on data levels or on data differences? In the remainder of this section we illustrate that differencing the data before estimation will significantly amplify the bias due to observation errors.

Least squares applied to data levels gives the following estimate of β

$$(7) \qquad b = \beta - \frac{\beta\hat{X}'w}{\hat{X}'\hat{X}} + \frac{\hat{X}'u}{\hat{X}'\hat{X}} .$$

We are interested in the behavior of b for large samples so we consider

$$(8) \qquad \text{Plim}(b) = \beta - \text{Plim}(\frac{1}{T^2}\hat{X}'\hat{X})^{-1}\{\beta\text{Plim}(\frac{1}{T^2}\hat{X}'w) - \text{Plim}(\frac{1}{T^2}\hat{X}'u)\}.$$

If we substitute equation (5) for $\hat{X}$ and consider the case where $\rho < 1$ we obtain for each term[3]

$$\beta\text{Plim}(\frac{1}{T^2}\hat{X}'w) = 0,$$

$$\text{Plim}(\frac{1}{T^2}\hat{X}'u) = 0,$$

and

$$\text{Plim}(\frac{1}{T^2}\hat{X}'\hat{X})^{-1} = 0$$

so that

$$(9) \qquad \text{Plim}(b) = \beta.$$

With all $\rho = 1$ it is still the case that Plim(b) = beta.[4] For the regression based on levels, the fact that $\frac{1}{T^2}\hat{X}'\hat{X}$ grows at the rate T insures that in the limit Var (X)/Var (w) $\to \infty$ so that observational errors will have little influence on the estimate of b. We shall see later that this result also holds for small samples.

Taking first differences leaves each series with a non-zero mean. To avoid including a constant in the regression we assume that the regression equations apply to variations about the mean. If we use the notation Δ for first differences, the least squares estimate of β becomes

$$(10) \qquad b = \beta - \frac{\beta\Delta\hat{X}'\Delta w}{\Delta\hat{X}'\Delta\hat{X}} + \frac{\Delta\hat{X}'\Delta u}{\Delta\hat{X}'\Delta\hat{X}} .$$

To obtain large sample results we examine

$$(11) \qquad \text{Plim}(b) = \beta - \text{Plim}(\tfrac{1}{T}\Delta\hat{X}'\Delta\hat{X})^{-1}\{\beta\text{Plim}(\tfrac{1}{T}\Delta\hat{X}'\Delta w) - \text{Plim}(\tfrac{1}{T}\Delta\hat{X}'\Delta u)\}.$$

Using equation (5) to eliminate $\Delta\hat{X}$ we obtain for the individual terms

$$\beta\text{Plim}(\tfrac{1}{T}\Delta\hat{X}'\Delta w) = \beta\,\frac{2\sigma_\gamma^2}{1+\rho_w},$$

$$\text{Plim}(\tfrac{1}{T}\Delta\hat{X}'\Delta u) = 0,$$

and

$$\text{Plim}(\tfrac{1}{T}\Delta\hat{X}'\Delta\hat{X})^{-1} = \left(\frac{2\sigma_\gamma^2}{1+\rho_v} + \frac{2\sigma_\gamma^2}{1+\rho_w}\right)^{-1}.$$

Therefore, Plim(b) becomes

$$(12) \qquad \text{Plim}(b) = \frac{\beta\sigma_\eta^2(1+\rho_w)}{\sigma_\eta^2(1+\rho_w) + \sigma_\gamma^2(1+\rho_v)} .$$

The observation errors on X_t can be small relative to the trend changes in X_t and still produce a serious bias in b. What is important is their size relative to the stochastic variations in X_t. If the observation errors are of the same order of magnitude as the stochastic variations the estimate of β can be biased to half the actual value of β. An even more serious problem occurs if Y_t depends on smoothed values of X_t (equivalent to $\eta_t = 0$). In that case, differencing would destroy the relationship between X and Y and a regression of ΔY on ΔX would give a value of zero for b.

III. Simulation Results

In the previous section we derived theoretical results for the asymptotic effect of observation errors. This section considers small sample results obtained from two Monte Carlo studies. In the studies, theoretical $\hat{X}$ and Y data were generated from equations (5) and (1) respectively with the error term selected from tables of normal random deviates. The least squares estimate b of β was then computed for data levels and data differences. In each case the data were defined as the difference from the mean so no constant term was necessary. For each study the parameters were $\alpha = 1.0$, $\beta = 1.0$, $\sigma_\varepsilon^2 = 0.5$, $\sigma_\eta^2 = 0.5$, and $\sigma_\gamma^2 = 0.25$.

The first Monte Carlo study assumed that each error term was a random walk so that all values of ρ are unity. The sample size was varied from five to fifty observations and seventy-five separate simulations were performed for each sample size. Table 1 summarizes the results for each sample size. It lists the average value and the standard deviation of b computed from the seventy-five coefficients obtained for each sample size. The asymptotic result of b equal to 1.0 for levels and b equal to 2/3 for differences is seen to hold for very small samples.[5] The standard deviation of b declines as the sample size increases. Table 2 summarizes the regression results for a sample size of 35 observations. For data levels, the average standard error obtained from the regressions seriously understates the true standard deviation computed from the seventy-five estimates of β. This result is not surprisidg in view of the average Durbin-Watson statistic obtained for these regressions. The regression on differences gives an average standard error equal to the true standard deviation and an average Durbin-Watson statistic of 2.099. Unfortunately the estimate of β is seriously biased. The regression on differences is equivalent to using the Cochrane-Orcutt technique to correct the regression on levels for first order serial correlation. It has become standard practice in econometric studies to report results after correcting for serial correlation. That such a standard technique could bias the regression results is disturbing and calls for further study.

The second Monte Carlo study examines small sample results for different values of ρ (for simplicity we assume that $\rho_u = \rho_v = \rho_w$). For each value of ρ, seventy-five separate simulations were performed for sample sizes of 15 and 35 observations. Each simulation involved regressions on data levels, data differences, and data transformed by the Cochrane-Orcutt procedure. The Cochrane-Orcutt results are summarized in Figure 1. The bias in the estimate of β decreases as ρ decreases and decreases as the sample size increases. As T becomes large the bias goes to zero except when ρ is equal to unity. This occurs because the regression on levels is dominated by the behavior of the term $\frac{\alpha^2}{T^2}\Sigma t^2$ in $\frac{1}{T^2}\hat{X}'\hat{X}$. This term is approximately $\alpha^2 T/3$ and drives $(\frac{1}{T^2}X'X)^{-1}$ to zero for large T. If the Cochrane-Orcutt transformation is applied to data levels, the corresponding term is approximately $\alpha^2(1-\rho)^2 T/3$. As long as ρ is less than unity this term will drive $(\frac{1}{T^2}\hat{X}'\hat{X})^{-1}$ to zero for large T and eliminate the bias in the estimate of β. For small samples, however, multiplying the term $\alpha^2 T/3$ by $(1-\rho)^2$ greatly diminishes its impact if ρ is close to unity and we get some bias in the estimate of β.

Table 3 summarizes regression results on levels, differences, and transformed data for $\rho = 0.8$ and $T = 35$. The regression on levels gives an unbiased estimate of β but understates the true standard deviation of b.

The regression on differences gives a biased estimate of β, but the average standard error is approximately equal to the true standard deviation computed from the seventy-five estimates of β. The Cochrane-Orcutt regression produces a mixed bag consisting of a biased estimate (but not as biased as the regression with difference) which understates the true standard deviation (but not as much as the regression with levels). The mixed result occurs because X still contains the nonstationary trend $\alpha(1-\rho)t$.

IV. Conclusions

Our purpose has been to illustrate that data differencing may be more hazardous to healthy econometrics than one would expect judging from its widespread use. Differencing tends to amplify the biases due to observation errors. These biases can be large even if the data errors are small compared to the trend variations of the input variable. What is important is the observation error relative to the stochastic variation which remains after the trend is removed by differencing. When the output depends on smoothed values of the input, differencing may completely destroy the relationship between the variables. Correcting data levels for serial correlation of the residual is a form of data differencing in which some nonstationarity remains. As a result, it inherits some of the undesirable aspects of the regression on levels (understates the true variance) and the regression on differences (biased coefficients).

Our results, however, should not be taken as a recommendation to ignore the problem of spurious regressions. With most time series dominated by trend it is, unfortunately, all too easy to obtain high values of R^2 by estimating one variable as a distributed lag on another. Data differencing combined with estimation procedures which take explicit account of observation errors would alleviate the problem of spurious regressions. This implies that we need to be as concerned with modeling the error structure of our data as we are with modeling the error structure of our equations. Unfortunately, data differencing will not help if the output depends on smoothed values of the input. It is my own feeling that this situation describes the underlying structure of most economic models. If that is the case, it indicates the need for a more careful specification of the dynamic properties (the transfer function) of economic models. Such a specification would impose theoretical constraints on the lag structure between the input and output and make it harder to obtain high values of R^2 from spurious regressions.

REFERENCES

Box, George E.P. and Jenkins, Givelym M., <u>Time Series Analysis: Forecasting and Control</u>, Holden-Day, San Francisco, 1970.

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

Zellner, Arnold and Palm, Franz, "Time Series Analysis and Simultaneous Equation Econometric Models," <u>Journal of Econometrics</u>, 2 (1974), 17-54.

Footnotes

[1]Identification of the cross correlation function is greatly simplified if the stationary input is white noise. In time series analysis, the stationary input is filtered a second time by representing it as an autoregressive moving average process with a white noise residual.

[2]This point has been made by others such as Zellner and Palm. Our purpose is to develop the idea in the context of a simple theoretical model and provide explicit analytical results. We also want to illustrate the impact of data errors on the Cochrane-Orcutt correction for serial correlation.

[3]The term $\hat{X}'w/T^2$ consists of $\alpha\Sigma tw/T^2$, $\Sigma vw/T^2$ and $\Sigma w^2/T^2$. The last two become cov(v,w)/T and var(w)/T for large T and are clearly zero in the limit. The first term is $\alpha(w_T + w_{T-1}(T-1)/T + w_{T-2}(T-2)/T + \cdots)/T$ and is zero in the limit because of the assumption that $E(w) = 0$. The term $\hat{X}'u/T^2$ is equal to zero by a similar argument. The behavior of $\hat{X}'\hat{X}/T^2$ is dominated by the term $\alpha^2\Sigma t^2/T^2$. The summation is approximately $T^3/3$ so that $\hat{X}'\hat{X}/T^2$ grows at a rate proportional to T. In the limit, therefore, $(\hat{X}'\hat{X}/T^2)^{-1}$ goes to zero.

[4]With ρ equal to unity we need to be concerned with terms of the form $\Sigma w^2/T^2$. Each w_k is equal to $w_o + \sum_{j=1}^{k}\gamma_j$ so that

$$\frac{1}{T^2}\sum_{k=1}^{T} w_k^2 = \frac{1}{T^2}\sum_{k=1}^{T}\{w_o^2 + 2w_o\sum_{j=1}^{k}\gamma_j + (\sum_{j=1}^{k}\gamma_j)^2\}.$$

The first term is w_o^2/T and goes to zero for large T. The second term is $2w_o(\gamma_1 + \gamma_2(T-1)/T + \gamma_3(T-2)/T + \cdots)/T$ and goes to zero because of the assumption that $E(\gamma) = 0$. The third term is $(\gamma_1^2 + \gamma_2^2(T-1)/T + \gamma_3^2(T-2)/T + \cdots)/T$ plus various cross products which are zero in the limit because of the assumption that $\mathrm{cov}(\gamma_i,\gamma_j) = 0$. The third term is less than $\Sigma\gamma^2/T$ which in the limit becomes σ_γ^2. Since $\hat{X}'\hat{X}/T^2$ grows at the rate T driving $(\hat{X}'\hat{X}/T^2)^{-1}$ to zero, we still obtain $E(b) = \beta$.

[5]The small sample results for data differences are independent of α since differencing removes any influence of the trend. As equation (12) indicates, the bias merely depends on the relative size of v and w. The results for data levels depend on the trend dominating the effect of observation errors. For large samples the trend will dominate if $\alpha > 0$. However, if α is small relative to w we would expect some bias in the small sample estimates of β. For the results of Table 1, the changes in $\hat{X}$ due to trend are twice the one sigma changes due to observation errors. If we reduce α to 1/2 so these changes are equal, we obtain $\bar{b} = 1.006$ and $\sigma_b = 0.394$ for $T = 20$. For α equal to 1/4 we obtain $\bar{b} = 0.833$ and $\sigma_b = 0.520$. The unbiased small sample results for data levels, thus, require that changes in $\hat{X}$ due to observation errors not swamp the trend changes.

Table 1

Simulations on Sample Size (All $\rho = 1$)

Sample Size	Levels		Differences	
	$\bar{b}$	σ_b	$\bar{b}$	σ_b
5	1.005	0.541	0.578	0.577
10	0.990	0.313	0.687	0.355
15	0.982	0.220	0.663	0.241
20	1.021	0.215	0.713	0.234
25	1.029	0.237	0.677	0.213
35	1.016	0.158	0.633	0.163
50	1.002	0.121	0.662	0.139

Note: The average b is $\bar{b} = \sum_{j=1}^{75} b_j/75$ and the standard deviation is $\sigma_b = (\sum_{j=1}^{75} (b_j - \bar{b})^2/75)^{1/2}$.

Table 2

Simulation Results (All $\rho = 1$)
Sample Size = 35

	$\bar{b}$	Average S.E. b	σ_b	Average R^2	Average D-W
Levels	1.016	0.022	0.158	0.982	0.588
Differences	0.633	0.163	0.163	0.317	2.099

Table 3

Simulation Results (All $\rho = 0.8$)
Sample Size = 35

	$\bar{b}$	Average S.E. b	σ_b	Average R^2	Average D-W
Levels	0.995	0.019	0.055	0.987	0.783
Differences	0.628	0.163	0.166	0.313	2.232
C.O. Procedure	0.917	0.066	0.108	0.824	2.022

Figure 1

Cochrane-Orcutt Simulation