SEASONALITY IN REGRESSION

by

Christopher A. Sims

Discussion Paper No. 23, September 1972
Revised July 1973

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

SEASONALITY IN REGRESSION

by

Christopher A. Sims*

Recent work by D. Grether and M. Nerlove (1970) and by H. M. Rosenblatt (1968) among others has addressed again the old question of how best to seasonally adjust economic time series. This work has not, however, considered the question of how seasonal adjustment or the failure to adjust series contaminated by seasonal noise might affect analysis of relations between series. Other recent work which has dealt with seasonality in regression models (J. J. Thomas and Kenneth F. Wallis (1971), D. Jorgenson (1964), M. C. Lovell (1963)) has given specific guidance only for narrower definitions of seasonality than that now commonly used in considering seasonal adjustment of individual series.[1/]

This paper develops useful results for regression in the presence of seasonal noise which evolves slowly, producing sharp but not infinitesimally narrow peaks at seasonal frequencies in the spectral densities of the variables.[2/]

The paper's analytical parts assume the reader is familiar with the theory of covariance-stationary stochastic processes, though at some points there is an attempt to recapitulate briefly in less technical language. A bivariate distributed lag model is the basis for discussion throughout. Extensions to multivariate distributed lag regressions are obvious, and time series regression models in which no lags appear are a special case of the paper's general distributed lag model.

Section I examines the nature of asymptotic bias in least-squares estimates of lag distributions when seasonal noise is present and seasonal adjustment is either not attempted or not complete.[3] It is shown that the bias is likely either to be small or, if large, to be clearly recognizable as "seasonal" if the regression is estimated as a long, two-sided lag distribution with no strong a priori smoothness constraints. Imposition of the usual sorts of a priori constraints on the length or smoothness of the lag distribution is likely to increase the asymptotic bias and to make it less obvious in form.

Section II develops a procedure which reverses the effect of a priori constraints. If independent and dependent variables are both seasonally "overadjusted" in a certain sense, using the same linear procedure to adjust both series, then even quite weak a priori constraints which are approximately valid for the true lag distribution will reduce bias due to seasonality to negligible proportions.

Section III considers the case where independent and dependent variables are adjusted by different procedures, as must frequently be the case if published adjusted series are used. In this case there is no presumption that use of the adjusted data will yield smaller biases than use of the original data.

Section IV takes up the case, probably rare in practice, where the seasonal noises in independent and dependent variables are independent of each other. In this case Grether-Nerlove optimal adjustment of the independent variable proves to be part of an unbiased estimation procedure.

I. Effects of Seasonal Noise

Suppose we have a true dependent variable y and a true independent variable x, contaminated by seasonal noises z and w, respectively. The variables are related by the following distributed lag regressions:

1) $y = x * b + u$

2) $z = w * c + v$,

where u is orthogonal[4/] to x, v is orthogonal to w, and all the stochastic processes in (1) and (2) (y, x, u, z, w, and v) are jointly covariance-stationary. The notation "x * b" is defined by

3) $x * b(t) = \sum_{s=-\infty}^{\infty} b(s)x(t-s)$.

Assume further that the noises z and w are orthogonal to x and y.

Now if we observe, instead of y and x, the contaminated variables $y' = y+z$ and $x' = x+w$, then there will be a distributed lag relation analogous to (1) relating y' and x', namely

4) $y' = x' * b' + u'$,[5/]

with u' orthogonal to x'. The question we address initially is: How is b' related to b?

From what we have assumed it follows that $\tilde{b} = S_{y'x'}/S_{x'}$, where $\tilde{b}'$ is the Fourier transform of b', $S_{y'x'}$ is the cross-spectral density of y' with x' and $S_{x'}$ is the spectral density of x'. Rewriting this expression for $\tilde{b}'$ in terms of the underlying variables of (1) and (2) gives us

5) $\tilde{b}' = \tilde{b}(S_x/(S_x+S_w)) + \tilde{c}(S_w/(S_x+S_w))$.

Hence <u>in the frequency domain</u> we have $\tilde{b}'$ a weighted average of $\tilde{c}$ and $\tilde{b}$, with weights which vary with frequency. We shall see that this is not at all the same thing as having b' a weighted average of b and c in the time domain.

If w is a seasonal noise, then S_w will be small relative to S_x except in narrow bands of frequencies near the seasonal frequencies. Thus the first term on the right-hand side of (5) is the Fourier transform of b multiplied by a function which is near one except near seasonal frequencies, where it is less than one. In other words, this first term is the Fourier transform of b after a partial "deseasonalization" through a symmetric linear filter. Of course where c is small relative to b, this first term dominates the expression; so it is worthwhile initially to ask how this deseasonalized b differs from the original. It might be expected that, since b is unlikely to have a periodic form in most applications, deseasonalization should not have any substantial effect on the form of b. As a rough first approximation this is true, as can be seen from the examples displayed in Figure 1.

Figure 1a shows the spectral density of a hypothetical x' process obtained by adding to an x which has $E[x(s)x(s-t)] = R_x(t) = .9^{|t|}$ (a first-order Markov x with parameter .9) a w satisfying

6) $$S_w = \left(\frac{\sin(10\omega)}{5\sin(2\omega)}\right)^2 \cdot \left(\frac{17}{16} - \frac{\sin(8.5\omega)}{\sin(.5\omega)}\right)^2 \cdot K,$$ [6/]

where K is a constant chosen to give w a variance one-fourth that of x. In Figures 1b and 1c we see the b' from a true lag distribution b which is an identity lag distribution ($b(0) = 1$, $b(t) = 0$ for $t \neq 0$), and a Koyck lag distribution ($b(t) = 0$, $t < 0$, $b(t) = .9^t$, $t \geq 0$) respectively, on the assumption that $c=0$ with the S_x and $S_{x'}$ displayed in Figure 1a.

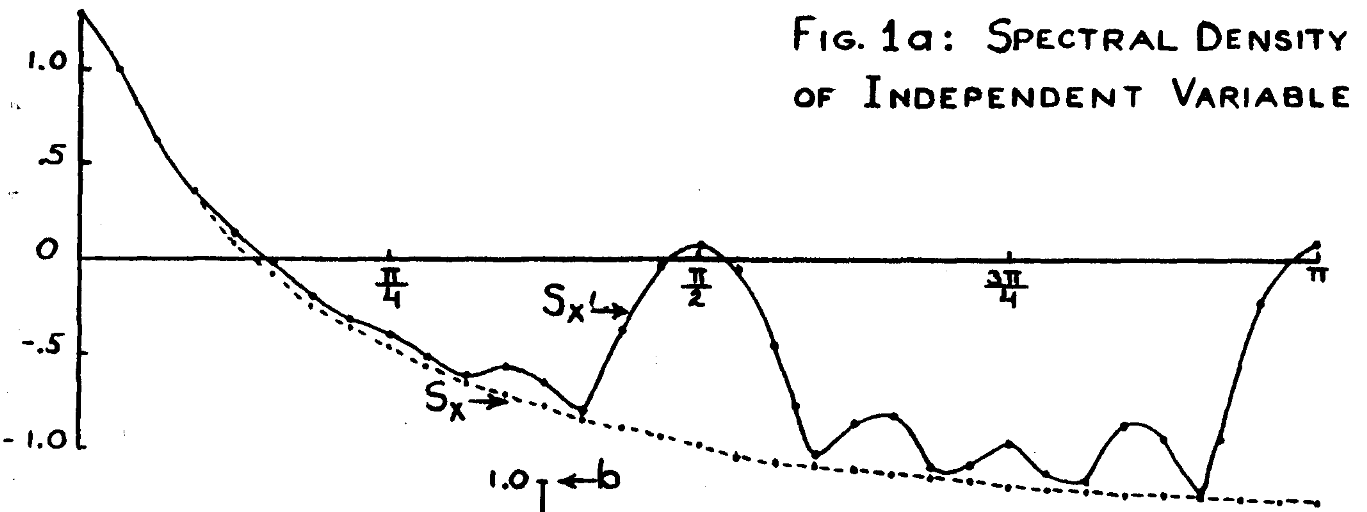

Fig. 1a: Spectral Density of Independent Variable

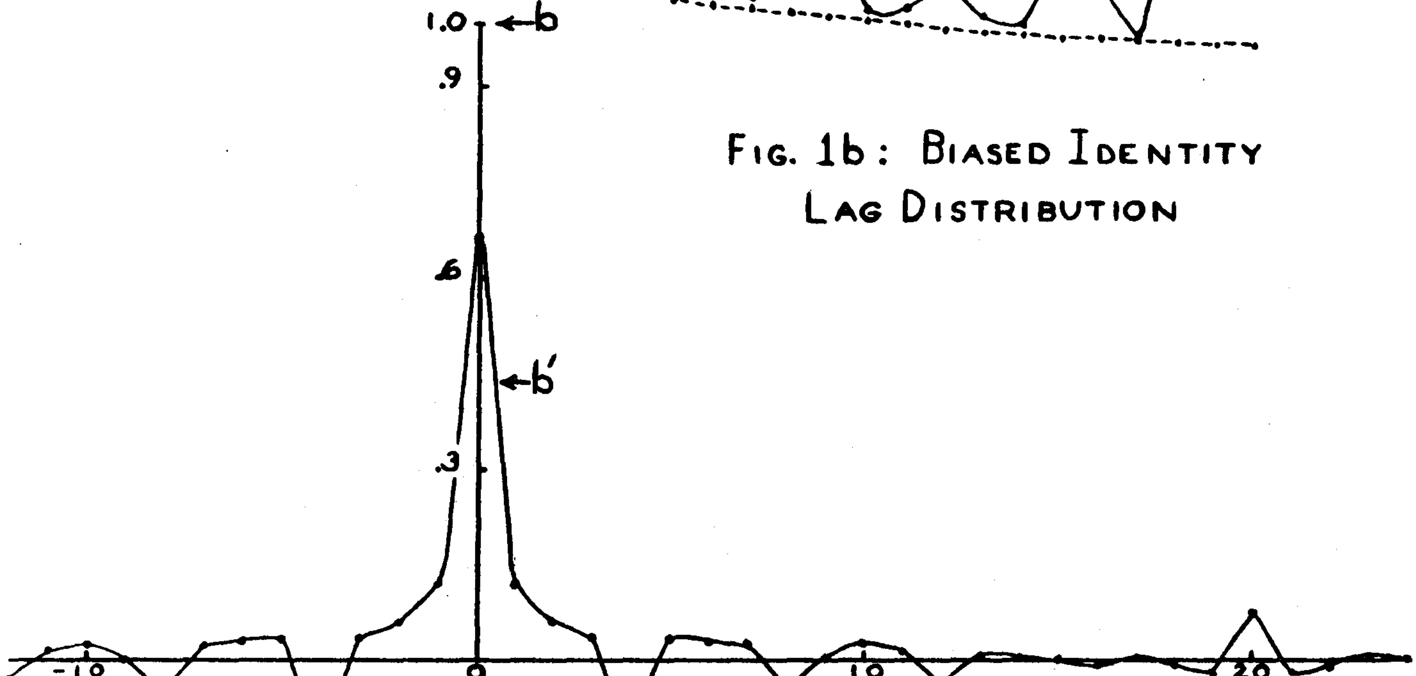

Fig. 1b: Biased Identity Lag Distribution

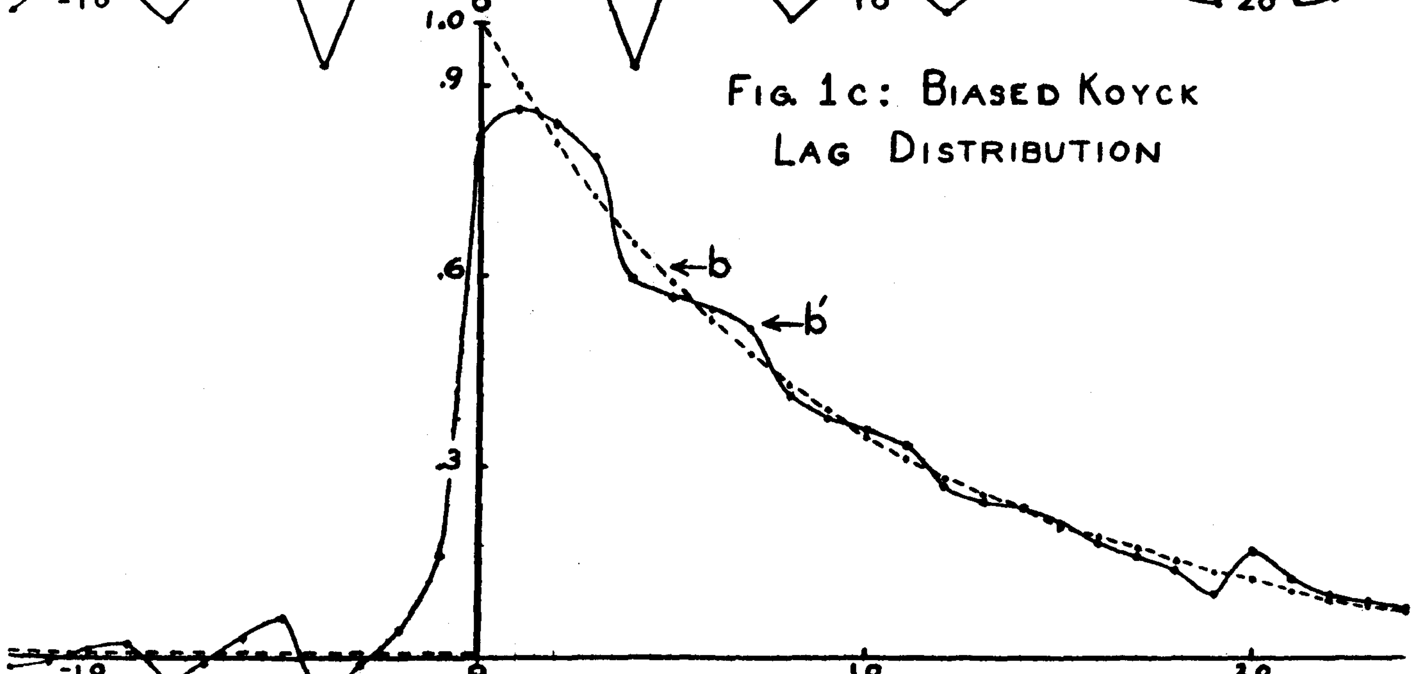

Fig. 1c: Biased Koyck Lag Distribution

In both the examples of Figure 1, because of the c=0 assumption, the basic shape of the true lag distribution is only mildly distorted. Because of the assumption that seasonal noises have negligible power at very low frequencies, the sums of coefficients in the lag distributions show little bias. Yet even for these relatively mild examples of bias, fallacious inferences might result. In the case of the identity distribution, the coefficients on future and lagging variables are non-negligible. Even in samples of moderate size, a test of the null-hypothesis that the full effect of the independent variable is felt in the current quarter might strongly reject the hypothesis. For both distributions, coefficients on future values are large enough that the hypothesis of one-sidedness might be rejected in samples of moderate size.

Of course this possibility of fallacious inference is mitigated by the fact that unconstrained estimates of the biased distributions would reveal the presence of bias due to the periodic patterns in the coefficients. Note that even in the examples of Figure 1, where seasonal power has been <u>removed</u> from the true distributions to give the biased distributions, the effect is to <u>introduce</u> seasonal fluctuations into the sequence of coefficients. The reason is that seasonal bias, because it is concentrated at seasonal frequencies, makes the difference between the true and biased distributions large in absolute value only in the neighborhood of seasonals. Hence this difference must itself show a pattern of slowly evolving seasonal fluctuations. Unless the true lag distribution itself shows seasonal fluctuations, the biased distribution must therefore show seasonal fluctuations around its true form.

When c is zero, that is when the seasonal noises are uncorrelated, the magnitude of the seasonal bias is limited. At worst $\tilde{b}' = \tilde{b}\, S_x/S_{x'}$ will fall near to zero in absolute value near the seasonals. But when $c \neq 0$, bias can be much larger if $|\tilde{c}|$ is greater than $|\tilde{b}|$ near seasonals. For as can be seen from (5), in that case $|\tilde{b}'|$ will tend toward $|\tilde{c}|$ near seasonals, then return to $|\tilde{b}|$ away from seasonals.

Seasonal bias is probably widespread in econometric work. Since becoming aware of its possible existence and likely form, I have found it in several pieces of applied work, both by myself and others. An example of such bias, resulting from a distributed lag regression of GNP on money supply using only 3 standard quarterly dummies with otherwise unadjusted data, appears in Figure 2. But such bias reveals itself directly only when a lag distribution is not constrained a priori to be short and/or smooth. Most econometric work has imposed such a priori constraints; and as we shall now see, the use of such constraints with seasonally noisy data is likely to make bias worse as well as conceal it.

To get an intuitive feel for why a priori constraints on the estimate make matters worse, consider the plots in Figure 1. What would happen if we estimated these biased lag distributions by least squares, subject only to the constraints that for the estimated lag distribution b, $b(s) = 0$ for $s < 0$ and, say, for $s > 8$? In doing this we would be omitting a large number of variables. In the presence of a strong quarterly seasonal,

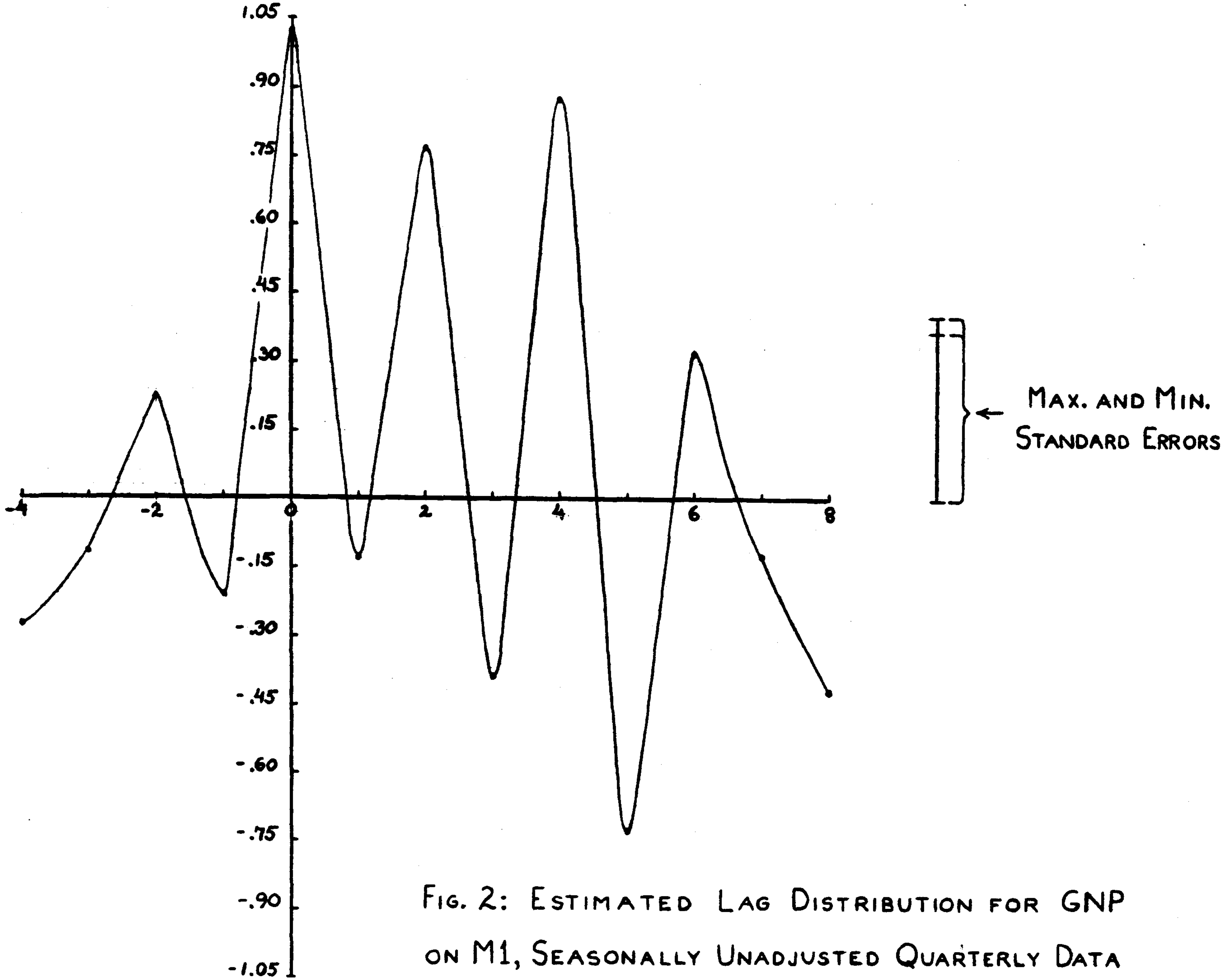

Fig. 2: Estimated Lag Distribution for GNP on M1, Seasonally Unadjusted Quarterly Data (1951.1 - 1965.4)

Notes to Figure 2:

The regression equation which produced this lag distribution was:

$$y(t) = \sum_{s=1}^{3} a_s D_s(t) + c + b*x(t) + u(t),$$

where $y(t) = g*Y(t)$, $x(t) = g*X(t)$, Y is the log of nominal GNP, and X is the log of nominal M1, currency plus demand deposits. The prewhitening filter g has the values 1, -1.5, .5625, for t=0, 1, 2, and the value zero for other values of t. T is a linear time trend and D_j is the j'th quarter seasonal dummy. The estimated equation thus corresponds exactly to those used in my paper "Money, Income, and Causality", in the American Economic Review in September, 1972. The reader may verify for himself that the severe seasonal bias apparent here does not appear in the regressions with properly adjusted data.

$x(t+4s)$ will tend to show strong positive correlation with $x(t)$ for integer s of moderate size. If $x(t)$ is an included variable in our restricted regression, the excluded variables of the $x(t+4s)$ type will tend to have coefficients whose sign matches the sign of the seasonal bias in the coefficient of $x(t)$, as is clear from inspection of Figure 1. Hence by excluding these variables we amplify the bias.

A tight but not obviously unreasonable prior restriction would, in the case of the example in Figure 1b, require $\hat{b}(s) = 0$ for all $s \neq 0$. Using the same x' employed in computing the unconstrained lag distribution of Figure 1c, we find that the constrained estimate would in large samples give a coefficient on $x(t)$ of .80. The lag distribution in Figure 1c has a greater bias in the zero-order coefficient than this constrained estimate, but yields a better estimate of the sum of coefficients. If we were to relax the constraint slightly and add $x(t-1)$ as well as $x(t)$ to the regression, the coefficients would tend to .58, .31, on current and lagged x respectively. If these were estimates of, say, a short-run demand function for labor, the evidence for "short-run increasing returns to scale" would appear very strong, and by keeping the lag distribution short evidence of the presence of seasonal bias would be suppressed.

A clearer idea of the reasons for the bad effect of imposing a priori restrictions when using unadjusted data can be obtained by looking at the problem in the frequency domain, using the apparatus of an earlier paper of mine (1972). That paper shows that when we estimate b' in (4) by least squares subject to prior constraints, the estimate $\hat{b}$ will minimize, subject to the constraints, the quantity

$$7)\quad \int_{-\pi}^{\pi} S_{x'}(\omega)\left|\tilde{\hat{b}}(\omega) - \tilde{b}'(\omega)\right|^2 d\omega .$$

This is a weighted average of the squared differences between the Fourier transforms $\tilde{\hat{b}}$ and $\tilde{b}'$ of $\hat{b}$ and b', with weights given by the spectral density of x', $S_{x'}$. Constraints of the type usually imposed on estimated lag distributions, requiring that $\hat{b}(t)$ vanish for large t or that $\hat{b}(t)$ decay rapidly (say exponentially), amount in the frequency domain to smoothness constraints on $\tilde{\hat{b}}$. They will, in other words, prevent $\tilde{\hat{b}}$ from displaying sharp peaks or dips of the type $\tilde{b}'$ has at the seasonal frequencies. Since we do not expect the true $\tilde{b}$ to display such sharp peaks and dips, preventing $\tilde{\hat{b}}$ from displaying them would be all to the good, were it not that $S_{x'}$ itself has sharp peaks at the seasonals: Because $|\tilde{\hat{b}} - \tilde{b}'|^2$ at seasonal frequencies is given especially strong weight in the criterion function (7), constraints preventing <u>sharp</u> peaks or dips in $\tilde{\hat{b}}$ at seasonals only result in estimates which have <u>wide</u> peaks or dips at seasonals, trading relatively more error at values away from the seasonal for a good fit very close to the seasonal.

The reader may now see that one useful way to proceed would be to impose the prior restrictions but at the same time be sure that the spectral density of the independent variable is small near the seasonals instead of large. We consider such a procedure in Section II.

From the analysis up to this point it might appear that if w is zero, the seasonal noise z in y will produce no bad effects.

The only sort of effect considered to this point, however, has been asymptotic bias. Seasonal noise in y does not by itself produce asymptotic bias, but it does produce bad effects in small samples by introducing a seasonal pattern of serial correlation in residuals. Seasonal serial correlation is only a special case of the general serial correlation problem and raises no special theoretical problems. In practice, however, most econometricians use estimation and testing methods which allow for only low-order autoregressive patterns of serial correlation. Thus seasonal serial correlation is likely to go undetected by the usual tests and uncorrected by the usual methods of modifying least squares to allow for serial correlation. Some consideration of how to proceed when seasonal noise appears in y only can be found in Thomas and Wallis (1971).

II. Correcting for Seasonal Bias

If we choose some sequence $a(s)$, $s=-\infty, \ldots, \infty$, and "filter" both y' and x' through it to obtain $y'' = a*y'$ and $x'' = a*x'$, then we can still write, analogous to (4),

8) $y'' = x''*b' + u''$,

where x'' and u'' are orthogonal and b' is the same lag distribution that appears in (4). If $|\tilde{a}|$ is chosen to have sharp dips at the seasonal frequencies, then x'' and y'' will be deseasonalized versions of x' and y'. By deseasonalizing the data, we have left unchanged the asymptotic bias in the lag distribution.

But this result depends on the assumption that no constraints are placed on b' in the estimation process. If the estimate $\hat{b}$ is constrained, then estimation will again minize the expression (7), but with the difference that now the weighting function $S_{x'}$ is replaced by the weighting function $S_{x''} = |\tilde{a}|^2 S_{x'}$. Since the common sorts of prior constraints on $\hat{\tilde{b}}$ will prevent it from showing the sharp peaks or dips at seasonals we expect in $\tilde{b}'$, we can, by making $S_{x''}$ small near the seasonals, insure that our estimate $\hat{\tilde{b}}$ will fit closely to the true $\tilde{b}$ over frequencies away from the seasonals and simply cut across the bases of the seasonal peaks and dips in $\tilde{b}'$.

The two major questions which arise in implementing the proposed procedure are how to choose a and how to decide whether strong enough prior constraints have been imposed. Consider first the choice of a. Ideally, one would choose a so that $\tilde{a}$ is zero in some band of width δ about the seasonal frequencies and one elsewhere. The number δ would be chosen larger than the width of the largest seasonal peak in the spectral density of x'. With δ chosen equal to π/k, the implication is that correlation between annual seasonal patterns may become small within as little as 2k/S years, where S is the number of observations per year (4 in the case of quarterly data). A δ of $\pi/8$ is probably adequate for most quarterly data. One might start with this choice of δ, then examine the estimated spectral density of the resulting x" to be sure that the desired deep dips in $S_{x''}$ at seasonals have been achieved.[7/]

No filter of finite length, unfortunately, can have $\tilde{a} = 0$ over any interval of non-zero length. Hence the best we can hope for is to make $\tilde{a} = 0$ at the seasonal frequencies and to keep $|\tilde{a}|$ small over the intervals of length δ (henceforth called the "seasonal bands"). There is a doubly-infinite a which has $\tilde{a} = 1$ except in intervals of length of δ about the seasonals. We will call this ideal a, a_δ. For quarterly data its form is

$$9)\quad a_\delta(t) = \begin{cases} -[2\sin(\delta t/2)]\,\alpha(t)/t, & t \neq 0 \\ 1-3\delta, & t=0 \end{cases}$$

$\alpha(t)=3$, when $t=4s$ for integer s,
$\alpha(t)=-1$, all other t.

One reasonable way to proceed would be to truncate a_δ at some length which avoids loss of too high a proportion of the data at the ends of the sample and apply the truncated a_δ as the deseasonalizing filter.

Other deseasonalizing filters, easier to compute, may give practical results as good as truncated versions of (9). In order that the filter have Fourier transform near zero at the seasonals, it is necessary only to be sure that the same-month or same-quarter means, $(S/T)\sum_{j=0}^{T/S} x''(k+Sj)$, $j=1, \ldots, S$, where S is observations per year and T is sample size, are all close to the overall sample mean. For example, choosing $a(0)=1$, $a(4)=-1$, $a(t)=0$ for other values of t, gives a filter with zeroes at the quarterly seasonals. The trouble with choice of such simple deseasonalizing filters is that it is hard to relate them to the choice to δ.[8/] The fourth-order difference filter just described creates much broader dips in the spectral density than will be required for most econometric work, while at the same time making the dips so gradual that δ is very ill-defined. Without any value for δ, we have no firm basis for making the degrees of freedom corrections discussed below.

An alternative to explicity filtering the data is to remove seasonality by regression. A procedure which will, as sample size approaches infinity, remove all power in x' and y' in a band of width δ about each seasonal, is to estimate the seasonal in, e.g., x', by regressing $x'(t)$ on the set of functions $\cos(\omega_j t)$ and $\sin(\omega_j t)$ for all ω_j of the form $\omega_j = 2\pi j/T$ and for which ω_j lies in the seasonal bands of width δ. If both y' and x' have seasonal power removed this way, the result is equivalent to including the set of cosine and sine functions explicitly in the regression. Hence the implied loss in degrees of freedom is easily computed. Note that the loss in degrees of freedom is proportional to sample size, so that, e.g., with 80 quarters of data and a δ of $\pi/8$, 15 degrees of freedom are lost. In some applications a δ of $\pi/12$ might suffice, allowing a loss of only 9 degrees of freedom. To get back to a loss of only 3 degrees of freedom, the situation equivalent to use of standard seasonal dummies, requires assuming a δ smaller than $\pi/20$, or that seasonal patterns remain strongly correlated over more than half the full sample of 20 years.

This suggested procedure of using cosine and sine functions to deseasonalize by regression is equivalent to estimating the regression in the frequency domain, suppressing entirely the observations occurring in the seasonal bands.[9/]

Whenever estimation is carried out with deseasonalized data, it would seem reasonable to correct the degrees of freedom of the regression where possible by determining δ and reducing the degrees of freedom by a factor of $[1-(3\delta/2\pi)]$. This procedure is justified theoretically by the assumption that use of the adjusted data corresponds to frequency-domain regression with observations in the seasonal bands omitted.

It may be more convenient in some instances to use instead of the suggested set of cosine and sine functions, a set of equivalent size of interactions of polynomial terms with seasonal dummy variables. These would be variables of the form

10) $D_{pj}(t) = 3t^p$, $t=j+4s$

$D_{pj}(t) = -t^p$, $t \neq j+4s$ for integer j and s,

$j=1, 2, 3, p=1, \ldots, P$

Note that in order to maintain a constant δ, P must be allowed to increase in proportion to sample size. Otherwise we implicitly impose a slower rate of change on the seasonal in large samples than in small samples. If one has chosen δ, a reasonable choice of P is probably $\delta T/2\pi$. Thus in a sample of size 80 and $\delta=\pi/8$, a p of 5 is appropriate. In practice it may sometimes be easier to choose P by inspection of the raw data.

There are some drawbacks to use of polynomial-seasonal interactions rather than a corresponding number of trigonometric terms for removing seasonal power. If p is large, the polynomial terms may be highly collinear if used in the form given in (10). Collinearity can be avoided by using orthogonal polynomials in place of t^p, but this is at the cost of the algebraic simplicity which is the main appeal of the method. Furthermore, to use the trigonometric deseasonalization method, one can draw on packaged Fourier analysis programs, within which the method amounts to taking the Fourier transform of $x(t)$, $t=1, \ldots, T$, setting the components of the transform in the seasonal band to zero, then taking the inverse transform. This is much more efficient computationally than using packaged regression programs to remove seasonality with the polynomial-seasonal interactions. Finally, it is possibly not true, though this remains an open question, that if p is allowed to increase linearly with T the limiting effect is to remove all power in a set of seasonal bands and no power elsewhere.

The regression methods for deseasonalization are preferable to time domain filtering mainly on grounds of neatness. It may appear that because the interpretation of the former of the two regression methods as a frequency-domain regression involves completely ignoring observations in the seasonal bands, this method avoids the difficulty that $\tilde{a}$ cannot be zero over intervals of non-zero length. But this is a fallacy, since eliminating the finite number of observations in the seasonal bands in frequency domain regressions cannot prevent leakage of a powerful seasonal noise into the remaining observations. It may also appear that the regression methods allow an "exact small sample theory", since they involve no ambiguity in the degrees of freedom correction. But this too is an illusion. Even if the seasonal noise has power only in the seasonal bands, not just primarily in the seasonal bands (which is unlikely), the seasonal pattern will not be exactly a linear combination of the regression variables we use to estimate it. Hence even with the inclusion of seasonal variables in the regression, some error remains in the independent variables and the statistical theory for regression with independent errors can apply only approximately.

Recall that the asymptotic bias in the unrestricted lag distribution is unaffected by linear seasonal adjustment of the forms suggested here. If the a priori restrictions we place on the lag distribution are too weak, the seasonal adjustment will not prevent the asymptotic bias from showing itself. The question is, what criteria are there for determining when restrictions are "too weak"? Where our restrictions are truncations, restricting b(s) to be zero for $s > M_1$, $s < M_2$, it is not hard to show that the width of peaks in $\tilde{\hat{b}}$ is implicitly restricted to be no smaller than about $4\pi/(M_1-M_2)$. Thus in our example of quarterly data, $\delta = \pi/8$, truncation restrictions which hold M_1-M_2 below 32 (8 years) suffice to prevent asymptotic bias from showing itself if seasonal power in the independent variable has been removed. Lag distributions covering 8 years are certainly long enough for most econometric work; where a longer lag distribution is needed, imposition of probabilistic smoothness restrictions, using, e.g., Robert Shiller's (1971) convenient procedure, would suffice to prevent asymptotic bias from appearing.[10/] Alternatively, one might follow Shiller in using the Theil mixed estimation procedure but use it to assert that second differences of the Fourier transform of the estimated lag distribution across the seasonal bands are known a priori to be small. This uses directly a priori information that there should not be large seasonal components in the lag distribution, rather than relying on smoothness constraints to achieve the same effect.

Deterministic smoothness constraints of the common Almon polynomial or "rational distribution" type will prevent sharp seasonal peaks in $\tilde{\hat{b}}$ if the number of parameters estimated is kept small. Obviously a polynomial lag distribution spread over M quarters can produce no sharper peaks in $\tilde{\hat{b}}$ than a distribution constrained to cover the same quarters without the polynomial constraints. Furthermore, when we consider that a polynomial of degree p can have only p-1 local maxima, it is clear that if a peak of width as narrow as δ can be produced only with a lag distribution k years long, a peak that narrow must require a polynomial lag distribution of degree at least k+1.

Rational lag distributions are somewhat trickier in this respect. If the degree of the denominator of the generating function for the lag distribution is as high as S, the number of observations per year, then the parametric family of lag distributions is capable of generating arbitrarily narrow peaks at the seasonals. These peaks are of a special type, of course, and will not appear in the estimates, unless the peaks or dips in $\tilde{b}'$ happen to be of the same special type. Nonetheless, even fairly low order rational lag distributions may allow asymptotic bias to creep in.

III. Use of Published Adjusted Data

When a lag distribution estimated from seasonally unadjusted data shows obvious seasonal bias, more often than not in my experience re-estimation with adjusted data removes the seasonal pattern in the estimates. This, however, is a result which cannot be relied on to hold generally.

Occasionally it may happen that official adjustment procedures are close approximations to the kind of "a" filter we suggested in the preceding section. However, we must expect that official adjustment procedures will be deficient for regression work in two respects: They will not in general eliminate all or even most of the power in x' at the seasonal frequency; and the width of the band of frequencies in which they reduce variance may differ between y' and x'. Evidence that official procedures do not produce the required deep dips in the spectral density of the adjusted series appears, e.g., in Rosenblatt's paper (1968).[11/] Grether and Nerlove's (1971) criteria for optimal seasonal adjustment of an individual series imply that in long data series we will have approximately $\tilde{a}_x = S_{x'}/(S_x+S_w)$ for adjustment of x', and $\tilde{a}_y = S_{y'}/(S_y+S_z)$ for adjustment of y'. Application of a_x to x' would yield an $S_{x''}$ which, on a log scale, would show dips at the seasonals of the same size and shape as the peaks at the seasonals in $S_{x'}$. Thus optimal adjustment of a single series never produces a drop to zero in $S_{x''}$.

The Grether-Nerlove criterion also implies that the width of the dips at seasonals in $|\tilde{a}_x|$ and $|\tilde{a}_y|$ will differ if the widths of the seasonal peaks in $S_{x'}$ and $S_{y'}$ differ. Whether or not agencies publishing adjusted data do in fact commonly allow the width of the dips in their adjustment filters to vary in this way is unclear, at least to me. Posed in the time domain, the question is, when the actual seasonal pattern appears to change only very slowly over time, are more years used to estimate it? In any case, I have encountered examples where data from different agencies clearly show differences in the widths of the seasonal dips.[12/]

If for whatever reason different adjustment procedures are used on y' and x', the adjustment procedure itself may introduce or amplify seasonal bias in b'. Suppose both y' and x' are deseasonalized by linear filters, g and h respectively. Then the distributed lag regression of $y'' = g*y'$ on $x'' = h*x'$ is given by

11) $y'' = g*h^{-1}*b'*x'' + u''$

assuming that h^{-1} has an inverse under convolution.[13/] Comparing (11) with (8) one can see that the use of distinct deseasonalizing filters for the two variables has resulted in a change in the asymptotic bias in the lag distribution. If we call the lag distribution relating y'' and x'', b'', then we can write $\tilde{b}'' = \tilde{g}\,\tilde{b}'/\tilde{h}$. If, say, $|\tilde{b}'|$ has dips at the seasonals and $|\tilde{g}/\tilde{h}|$ has peaks of the right size and shape at the seasonals, b'' may be less biased than b'. But if on the other hand $|\tilde{b}'|$ has peaks at the seasonals and $|\tilde{h}|$ approaches nearer to zero than $|\tilde{g}|$ or remains small in a wider seasonal band, the bias in b' may be made many times worse. Of course the expression (7), showing the least squares criterion for estimation subject to restrictions, now becomes

12) $$\int_{-\pi}^{\pi} S_{x''}(\omega)\,|\tilde{\hat{b}}(\omega) - \tilde{g}(\omega)\,\tilde{b}'(\omega)/\tilde{h}(\omega)|^2\,d\omega$$

$$= \int_{-\pi}^{\pi} S_{x'}|\tilde{h}|^2|\tilde{\hat{b}} - \tilde{b}''|^2\,d\omega\ ,$$

so that besides the change in bias resulting from replacing b' by b'', there is a reduction in the weight given the seasonal bands due to replacement of $S_{x'}$ in (7) by $S_{x''} = S_{x'}|\tilde{h}|^2$ in (12). The change in bias may not always be for the worse, and the change in weights is always for the better, so it may be that more often than not use of adjusted data will reduce bias in the estimates, even when independent and dependent variables have been adjusted by different methods.

Unless we have some prior information about the nature of the bias in b' (as it is assumed we do in the next section) it is clearly better to hold bias constant and force the estimation procedure to ignore seasonal frequencies, adjusting y' and x' by the same filter, than to risk making matters worse by using separately adjusted series.

IV. The Case of Independent Seasonal Noises

Seasonal noises in economic time series are probably only rarely independent of one another. Certainly in aggregate time series certain identifiablefactors affect most of the seasonal fluctuations: weather, seasonal habits of retail customers, influence of annual budget cycles on spending patterns of large institutions. The reason for excluding seasonal "noises" in estimating a regression, then, is not that the "true" components of independent and dependent variables are related and the "noise" components are unrelated, but rather that the two components are related in different ways. In the notation of Section I, c in (2) is not zero, it is just different from b in (1).

Nonetheless there may be situations where seasonal factors in the variables of a regression are unrelated. In such a situation it is possible to make the asymptotic bias small by appropriate seasonal adjustment. From (5) we can see that if c=0, $\tilde{b}' = [S_x/(S_x+S_w)]\tilde{b}$. From (11) we know that use of different adjustment filters on x' and y' alters the bias to give us $\tilde{b}'' = \tilde{g}\,\tilde{b}'/\tilde{h}$. Suppose we take $\tilde{h} = S_x/(S_x+S_w)$ and $\tilde{g} \equiv 1$. Then b"=b and asymptotic bias has been eliminated.

This particular form for $\tilde{h}$ is, as Grether and Nerlove (1971) pointed out, the optimum seasonal adjustment filter when a doubly infinite sample is available. In long but finite time series optimal seasonal adjustment will be approximately in the form of this h except at the beginnings and the ends of the series. Hence in this case of uncorrelated seasonal noises, asymptotically unbiased regression estimates can be obtained by adjusting the independent variable only, using the Grether-Nerlove optimal procedure. Of course if there is a seasonal component in y', then least-squares regression estimates will be inefficient, but this now is only a standard problem in correcting for serial correlation.[14/]

Grether and Nerlove propose that the ratio $S_x/(S_x+S_w)$ can be estimated if we have finite parameterizations for both S_x and S_w, making them both rational functions. But choosing such finite parameterizations is an extremely hazardous business. Having estimated rational forms for both S_x and S_w, it would always be prudent to compare the implied $S_x{}' = S_x+S_w$ with a less tightly parameterized estimate of $S_{x'}$. A better way to carry out the estimation in my opinion would be to estimate $S_{x'}$ directly (taking care not to let a broad spectral window give misleadingly wide spectral peaks), then to estimate S_x by interpolating $S_{x'}$ across the bases of spectral peaks, finally taking $\tilde{h}$ to be $S_x/S_{x'}$.[15/]

V. Seasonality in Frequency-Domain Estimation Techniques

As has already been pointed out, if we simply Fourier transform all variables to arrive at a frequency-domain regression, the natural interpretation of the suggestions of Section II is that observations falling in the seasonal bands simply be ignored. The "Hannan efficient" estimation prcoedure is just weighted regression in the frequency domain, so we already have a recommendation for handling seasonality in this estimation method.

The Hannan inefficient procedure, which takes $\tilde{b} = S_{y'x'}/S_{x'}$, with unadjusted data, has to be handled somewhat differently. In my own work with the Hannan inefficient procedure, I have always found it possible to eliminate seasonality while ta ing the seasonal bands to be smaller than the width of the window used in forming spectral estimates. In this case, there is little danger in simply applying the Hannan inefficient method directly to data adjusted by any of the methods proposed in Section II. There will be moderate dips at the seasonals in both $S_{y''x''}$ and $S_{x''}$, but so long as the same window is used in forming both estimates, the dips should be parallel and have the effect of interpolating the estimate $\tilde{\hat{b}}$ across the seasonal bands. Where the seasonal bands are too wide relative to the spectral window, one can instead use the estimated $\tilde{\hat{b}} = S_{y''x''}/S_{x''}$ only at the non-seasonal frequencies, interpolating values for $\tilde{\hat{b}}$ in the seasonal bands before taking the inverse transform to get $\hat{b}$ itself.

VI. Practical Suggestions

The deseasonalization methods of Section II, while certainly quite feasible, are not trivial computations. Time series regression work in the past has proceeded without the use of these methods, and certainly not all such previous work is seriously distorted by seasonal bias. Is there any way to decide when these procedures are necessary?

I would argue that whenever the seasonally unadjusted data contain noticeable seasonal variation, the methods of Section II should be applied at least as a check in the final stages of research. But when large numbers of time series regressions are being computed on an exploratory basis, this elaborate treatment of seasonality is not always necessary. The following classification of types of regression equations may help to indicate when use of unadjusted or officially adjusted data may be relatively safe. This is meant as a guide not only to research practice, but to critical evaluation of existing time series regression estimates where seasonality has been treated more casually.

There are two dimensions of variation to the classification--sample size and nature of prior restrictions on the lag distribution. In a large sample (exceeding k years when the seasonal pattern is known to remain stable only over periods of k years) use of the usual S-1 seasonal dummies will not suffice to remove seasonality. If seasonal bias is an important possibility, there is no choice but to determine an appropriate δ for the seasonal bands and apply one of the Section II methods. In a small sample, on the other hand, the standard seasonal dummies will adequately protect against seasonal bias in the regression.

But when the lag distribution is only very weakly restricted a priori (as is more likely in larger samples), in particular when the only restriction is a truncation that still leaves the lag distribution more than 2 years long, strong seasonal bias can often be discovered by inspection of the estimated lag distribution. The bias will show up as a pattern of seasonal oscillations in the estimate. Hence with such a loosely restricted estimation procedure, there is little danger in proceeding with a casual treatment of seasonality, using unadjusted or officially adjusted data, so long as results are checked carefully for seasonal patterns.

Detecting seasonal bias by eye from time domain plots or listings of coefficients is an acquired skill, however. In some cases, e.g. when the estimated lag distribution contains large oscillations at a non-seasonal frequency, detecting oscillations at a seasonal frequency from time domain statistics may be nearly impossible. A safer procedure is to routinely compute the absolute value of the Fourier transforms of estimated lag distributions; seasonal bias will then show up as sharp peaks or dips at the seasonal frequencies.[16/]

The place where more elaborate methods are unavoidable is in a heavily restricted distributed lag regression estimated from a reasonably long sample. By the criterion given above, the more than 20 years available in postwar data is certainly a long sample, and the commonly employed rational or Almon polynomial types of prior restrictions are certainly strong enough to prevent seasonal bias from showing itself plainly. Thus much existing econometric research with postwar quarterly or monthly time series is potentially affected by hidden seasonal bias. If we are lucky, it will prove to be true that official adjustment procedures have been such as to keep the bias small in most instances.

FOOTNOTES

*/ This research was supported primarily by the National Bureau of Economic Research, with some support also from NSF Grant GS 36838.

1/ In particular, none of the latter three papers cited considers the effects of seasonal components in the independent variables which are not purely deterministic. Lovell anticipates some of the practical conclusions of this paper by suggesting that interactions of seasonal variables with polynomial or trigonometric trends be used to allow for evolving seasonal patterns, and suggests degrees of freedom corrections similar to those suggested here. But the ideas that the number of seasonal-interaction-with-trend variables required depends on sample size and that the theory of the normal linear regression model can be only approximate when seasonality is present were not explicit in Lovell's paper and are incompatible with the more precise statistical model of Jorgenson.

2/ Contemporaneous work by Wallis (1973) arrived independently at some of the theoretical conclusions of this paper, particularly expression (5) and an analogue of (11) below. Wallis's theoretical explorations do not extend into the area of interactions between adjustment methods and approximate a priori restrictions, which take up a major part of this paper. Wallis's simulation studies have verified that at least in a certain class of applications the Census X-11 method of seasonal adjustment behaves like a linear filter. Hence the relevance to practical work of theory based on treating seasonally adjusted data as linearly filtered is given indirect support.

3/ "Incomplete" seasonal adjustment results, e.g., when data are adjusted by regression on seasonal dummy variables but the seasonal noises actually evolve slowly over time rather than holding the fixed form implied by the use of seasonal dummies.

4/ A stochastic process u is orthogonal to a stochastic process x if and only if $\mathrm{Cov}(u(t), x(s)) = 0$, all t, s.

5/ That we can write (4) with u' and x' orthogonal is not a unique characteristic of the problem addressed in this paper. The random variable $x'{*}b'(t)$ is the projection of $y'(t)$ on the space spanned by $x'(s)$, $s=-\infty, \ldots, \infty$ under the covariance inner product. We can form such a projection for any covariance-stationary pair of processes to arrive at a relation like (4) or (1). Hence, any errors-in-variables problem for distributed lag relations (not just seasonal errors in variables) can be approached by comparing b' for a relation like (4) with the true b. I have applied essentially this same approach to a problem similar to an errors-in-variable problem in an earlier paper (1971).

6/ This formula was arrived at by taking the weights of a symmetric moving average representation of the process to be the convolution of $(1/5)\left(\sum_{s=-2}^{2} L^{4s}\right)$ with $\left(1 - (1/16) \sum_{\substack{s=-8 \\ s\neq 0}}^{8} L^{s}\right)$. The first term of this convolution gives peaks at the seasonals of width about $\pi/5$, while the second term eliminates the peak at 0 in the first term.

7/ Note that the reverse procedure, examining an estimate of $S_{x'}$ to determine δ, may not be reliable. If there is a narrow seasonal peak in $S_{x'}$, the smoothing inherent in spectral windowing may make the peak appear much broader than it is in fact.

8/ A second problem with the fourth-order difference filter given as an example is that it removes power near the frequency 0 in exactly the same way as at the seasonal frequencies. In much applied work it makes sense to remove trend from the data for the same reasons that it makes sense to remove seasonal variation. But where long-run effects are of central importance, a filter which removes power only at the seasonal frequencies is preferable.

9/ Such "band-limited regression" has been suggested by R.F. Engle in unpublished work as a way to handle other types of errors in variables as well.

10/ Shiller's procedure introduces prior information that fluctuations in b(t) are unlikely. Seasonal fluctuations in the estimated lag distribution can, if the data has been adjusted in one of the ways suggested above, be eliminated with only very small effect on the sum of squared errors in the estimated regression. Hence even very weak prior information that the lag distribution is likely to be smooth will flatten out seasonal oscillations.

11/ Though it should be noted that Rosenblatt used a spectral window whose width at the base exceeded $\pi/6$, so that if, e.g., seasonal patterns are stable over spans of five or six years and official procedures are Grether-Nerlove optimal, Rosenblatt would have seriously underestimated the depth and sharpness of seasonal dips.

12/ Wallis (1973) shows that the Census X-11 method behaves like a linear filter of fixed bandwidth.

13/ If $\tilde{h}$ is everywhere non-zero, h does have an inverse under convolution, given by $\tilde{h}^{-1} = 1/\tilde{h}$.

14/ To be precise, if $\tilde{h} = S_x/(S_x+S_w)$ is known exactly and applied to x' to yield x", the regression of y' on x" will be $y' = b*x'' + u'$, with b the same as in the relation (1) which applies to uncontaminated data. But of course h cannot be applied to x' unless an infinite sample on x' is available. If truncated versions of h are applied to finite samples on x', with the truncation points for h allowed to grow with sample size, least-squares regression estimates are undoubtedly asymptotically unbiased under weak regularity conditions on x, y, z, and w, but a rigorous proof of this seems more trouble than it's worth. Especially since, as we note below, h itself will in general be known only up to a good approximation.

15/ Note that while this procedure is adequate for the application proposed here, a good low-order rational parameterization of S_x and S_w, if it can be found, is certainly more convenient for some of the more refined applications Grether and Nerlove have in mind--e.g., finding the best one-sided deseasonalizing filters to apply at the ends of the series.

16/ As an example of the difficulty of detecting seasonal bias by eye in the time domain, consider the estimated lag distribution including future coefficients for GNP on M1 in Sims (1972b). Though no seasonal pattern leaps to the eye, seasonal effects probably are strongly biasing downward the 0-order coefficient.

References

Grether, D.M. and M. Nerlove (1970), "Some Properties of 'Optimal' Seasonal Adjustment", Econometrica, 38, pp. 682-703.

Jorgenson, D.W. (1964), "Minimum Variance, Linear, Unbiased, Seasonal Adjustment of Economic Time Series", J. Amer. Stat. Assoc., 59, pp. 681-724.

Lovell, M.C. (1963), "Seasonal Adjustment of Economic Time Series and Multiple Regression Analysis", J. Amer. Stat. Assoc., 58, pp. 993-1010.

Rosenblatt, Harry M. (1968), "Spectral Evaluation of SBLS and Census Revised Seasonal Adjustment Procedures", J. Amer. Stat. Assoc., 63, pp. 472-501.

Sims, C.A. (1971), "Discrete Approximations to Continuous Time Distributed Lags in Econometrics", Econometrics, 39, pp. 545-563.

__________ (1972a), "The Role of Approximate Prior Restrictions in Distributed Lag Estimation", J. Amer. Stat. Assoc., 67, pp. 169-175.

__________ (1972b), "Money, Income and Causality", American Economic Review, 62, pp. 540-52.

Thomas, J.J. and Kenneth F. Wallis (1971), "Seasonal Variation in Regression Analysis", J. Royal Stat. Soc., Series A (General), 134, pp. 57-72.

Wallis, Kenneth F. (1973), "Seasonal Adjustment and Relations Between Variables", mimeo.