ELSEVIER

Economics Letters 48 (1995) 249–256

economics letters

Spurious deterministic seasonality☆

Philip Hans Franses[a,*], Svend Hylleberg[b], Hahn S. Lee[c]

[a]*Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, Rotterdam, NL-3000 DR, Netherlands*
[b]*Institute of Economics, Aarhus University, Aarhus, Denmark*
[c]*Department of Economics, Tulane University, New Orleans, LA 70118-5698, USA*

Received 20 July 1994; accepted 18 August 1994

Abstract

It is sometimes assumed that the R^2 of a regression of a first-order differenced time series on seasonal dummy variables reflects the amount of seasonal fluctuations that can be explained by deterministic variation in the series. In this paper we show that neglecting the presence of seasonal unit roots may yield spuriously high values of this coefficient.

Keywords: Seasonal dummies; Stochastic seasonality

JEL classification: C22

1. Introduction and motivation

The model specification of seasonally observed macroeconomic time series has gained quite some attention recently. A stylized fact of most macroeconomic variables is that they contain at least one stochastic trend and that they display seasonal fluctuations that often account for the major part of the variation. As the construction of multivariate models is often based on the univariate analysis of such series, issues involved with model specification in the multivariate framework; such as, common stochastic trends, cointegration, and common seasonal movements, are also interesting topics for further analysis. One may therefore expect that the outcomes of univariate analysis can suggest possible routes for investigation. Further, one can readily imagine that a sound initial analysis of the individual series is a prerequisite for successful empirical modeling. In the case of seasonally observed univariate time series, the

☆ The first author thanks the Royal Netherlands Academy of Arts and Sciences for its financial support. The third author acknowledges the financial support of Tulane University through a short-term research grant. Thanks are also due to Bart Hobijn for research assistance.

* Corresponding author.

0165-1765/95/$09.50 © 1995 Elsevier Science B.V. All rights reserved
SSDI 0165-1765(94)00638-5

focus is mostly on describing the stochastic trend and the characteristics of the seasonally recurring patterns.

In some applications it is assumed that y_t, defined as $\Delta_1 x_t$ where x_t is the level (or log level) of a certain macroeconomic time series, and Δ_k is defined by $\Delta_k z_t = (1 - B^k) z_t = z_t - z_{t-k}$, which is observed s times per year, where s usually is 4 or 12, can be described by

$$y_t = \sum_{i=1}^{s} \delta_i D_{it} + \nu_t \,, \tag{1}$$

where D_{it} are seasonal dummy variables and ν_t is some error process. This model assumes that the stochastic trend in x_t is removed by applying the first-order differencing filter Δ_1, and that the seasonal fluctuations are stationary around a deterministic seasonal pattern. Since it may sometimes be interesting to assign part of the variation in a univariate series to seasonality, the model in (1) can be useful. The coefficient of determination of the auxiliary regression (1) may then be interpreted as the amount of variation that can be explained by deterministic seasonality, see, for example, Beaulieu et al. (1992) and Miron (1990). In fact, these studies report on high R^2 values like 0.7 for several macroeconomic time-series. As discussed in Beaulieu and Miron (1991), a comparison of the R^2 values across several auxiliary regressions, and of the successive values of the estimated δ_i parameters in (1), may then be used to yield insights into common aspects of various macroeconomic time series.

Alternative to the approach reflected by (1), several other methods have been proposed in the recent literature on the analysis of seasonal time series. Examples are the models that allow for the presence of so-called seasonal unit roots, see Hylleberg et al. (1990), the models that allow the parameters to be periodic, i.e. to vary with the seasons, originating from the seminal work of Gladyshev (1961), and periodic models that allow for seasonal unit roots, see, for example, Franses (1994). As an example of the class of models with seasonal unit roots, consider the fourth-order differencing filter, which is sometimes applied to quarterly time series. It can be decomposed as

$$\Delta_4 = (1 - B^4) = (1 - B)(1 + B)(1 - iB)(1 + iB) \,, \tag{2}$$

and the roots corresponding to the polynomials $(1 + B)$, $(1 - iB)$ and $(1 + iB)$ are called seasonal unit roots. Hylleberg et al. (1990) propose a test for such roots in quarterly data, and Hylleberg et al. (1993) document that many quarterly macroeconomic time series, which are similar to those considered in Beaulieu et al. (1992) and Miron (1990), possess one or more seasonal unit roots.

Given that seasonal unit roots seem to be present in many economic time series, it may be of interest to investigate the implications of neglecting seasonal unit roots in the regressions as (1). In this paper, we study the impact on the R^2 values to be obtained from the regressions in (1) when the seasonal time series under consideration contain unit roots at the seasonal frequencies, i.e. they contain factors like $(1 + B^2)$ and $(1 + B)$.

In the next section we discuss the distribution theory for the regression model in (1) when the DGP contains unit roots at some seasonal frequencies.

2. Distribution theory for R^2

Consider time-series processes generated by

$$\phi(B)y_t = \varepsilon_t \,, \tag{3}$$

where $\phi(B)$ contains (combinations of) polynomials corresponding to seasonal unit roots. For quarterly time series, the polynomial $\phi(B)$ in (3) can contain some or all of the elements in (2) so that $\phi(B)$ can be equal to $(1+B)$, $(1+B^2)$ or $(1+B)(1+B^2)$. As a simple example, suppose that $\phi(B) = (1+B)$. This implies that y_t is generated by

$$(1+B)y_t = \varepsilon_t \,. \tag{4}$$

Further, we assume that $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean zero and variance σ^2. While the discussion here is based on the i.i.d. assumption, the result can be shown to hold for a general innovation process satisfying the conditions for the functional central limit theorem in Phillips (1987).

In the case of (4), the regression model in (1) can be written as

$$y_t = \sum_{i=1}^{s} \delta_i D_{it} + \nu_t \,. \tag{5}$$

When we assume that a sample of N years, each containing s observations, is available, i.e. that the sample size is $T = Ns$, the OLS estimate of $\delta = (\delta_1, \delta_2, \ldots, \delta_s)'$ can be obtained as

$$\hat{\delta}_i = (1/N) \sum_{i=1}^{T} D_{it} y_t = \bar{y}_i \quad (i = 1, 2, \ldots, s) \,, \tag{6}$$

so that the estimated coefficients for the dummy variables are simply the mean values of y_t in each season. According to the usual results of regression theory for stationary series, the coefficients $\hat{\delta}_i$ $(i = 1, \ldots, s)$ converge in probability to constants as T (and hence N) grows. For non-stationary series, however, it is well known that the convergence to constants does not hold, see for example, Phillips (1986). The contrast with the standard regression theory has also been pointed out by Lee and Siklos (1992) in the context of seasonal time series. That is, the estimated coefficients in regression (1) are inconsistent and, in fact, diverge as the sample size T increases. In particular, it can be shown that

$$T^{-1/2}\hat{\delta}_i \rightarrow \sigma \int_0^1 W_i(r)\,\mathrm{d}r \quad (i = 1, 2, \ldots, s) \,, \tag{7}$$

where $W_i(r)$ is a standard Brownian motion such that $W_i(r) \equiv -W_{i-1}(r)$, $i = 1, 2, \ldots, s$, while the symbol '$\equiv$' signifies equality in distribution.

This result also extends to the distribution of the coefficient of multiple determination R^2 from the regression. This is stated in the following theorem.

Theorem. Suppose that Eq. (1) *is estimated by least squares regression, where* y_t *is generated by* (4). *Then, as* $T \uparrow \infty$,

$$R^2 \to \frac{\sum_{i=1}^{s}\left[\int_0^1 W_i(r)\,\mathrm{d}r\right]^2}{s\int_0^1 W_1(r)^2\,\mathrm{d}r} \equiv \frac{\left[\int_0^1 W_1(r)\,\mathrm{d}r\right]^2}{\int_0^1 W_1(r)^2\,\mathrm{d}r}. \tag{8}$$

Proof. The coefficient of determination is

$$R^2 = \frac{\sum_{t=1}^{T}(\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{T}(y_t - \bar{y})^2} = \frac{\sum_{t=1}^{T}\hat{y}_t^2 - T\bar{y}^2}{\sum_{t=1}^{T} y_t^2 - T\bar{y}^2} = \frac{T^{-2}\sum_{t=1}^{T}\hat{y}_t^2 - T^{-1}\bar{y}^2}{T^{-2}\sum_{t=1}^{T} y_t^2 - T^{-1}\bar{y}^2}.$$

As

$$\sum_{t=1}^{T}\hat{y}_t^2 = N\sum_{i=1}^{s}\hat{\delta}_i^2,$$

it follows from (7) that

$$T^{-2}\sum_{t=1}^{T}\hat{y}_t^2 = s^{-1}\sum_{i=1}^{s}(T^{-1/2}\hat{\delta}_i)^2 \to s^{-1}\sum_{i=1}^{s}\sigma^2\left[\int_0^1 W_i(r)\,\mathrm{d}r\right]^2. \tag{9}$$

Using the results in Chan and Wei (1988) and Lee (1992), it can be shown that

$$T^{-2}\sum_{t=1}^{T} y_t^2 \to \sigma^2\int_0^1 W_1(r)^2\,\mathrm{d}r, \tag{10}$$

$$T^{-1}\bar{y}^2 = (T^{-1/2}\bar{y})^2 = \left(T^{-1/2}T^{-1}\sum_{t=1}^{T} y_t\right)^2 \to 0. \tag{11}$$

The relation (8) then follows from the results (9), (10) and (11). □

The above theorem shows that, in contrast to the standard regression theory for stationary series, the R^2 has a non-degenerating asymptotic distribution. Thus, moderate and spuriously high values of R^2 are to be expected from the regression in (1). Note that the values of R^2 are, however, bounded by 1, since

$$\int_0^1\left[W(r) - \int_0^1 W(r)\,\mathrm{d}r\right]^2\mathrm{d}r = \int_0^1 W(r)^2\,\mathrm{d}r - \left[\int_0^1 W(r)\,\mathrm{d}r\right]^2 \geq 0.$$

Obviously, the R^2 values are also bounded from below by 0.

The results in the theorem can be extended to a more general case when $\phi(B) = \Delta_s/(1-B)$. The coefficient of determination R^2 in this case can be shown to converge in distribution to functions of Brownian motions similar to the expression given in (8). Hence, neglecting

seasonal unit roots may yield spuriously high R^2 values in the regression (1). Whether this asymptotic result has any bearing in small samples is investigated in the next section using Monte Carlo experiments.

3. Monte Carlo results

The design of the Monte Carlo experiments is as follows. We consider time series processes generated by (3), i.e. $\phi(B)y_t = \varepsilon_t$, where the ε_t process is assumed to be i.i.d. N(0, 1). The standard normal variables are generated using the Gauss programming language for the sample of size $T = 120$. Different sample sizes do not give dramatically different outcomes, and hence we only report on the sample size of 120. One could include ARMA-type error processes, but this would unnecessarily complicate the Monte Carlo exercise, as appropriate inference on the values of R^2 would require a proper model specification for the error process ν_t.

For quarterly time series, the $\phi(B)$ polynomial can be of the form $(1+B)$, $(1+B^2)$ or $(1+B)(1+B^2)$. From Table 1, which reports the quantiles of the distribution of the R^2 values from the regression in (1), it can be observed that the R^2 value exceeds, for example, 0.7 in more than 50% of the 10,000 replications and that it can be as high as 0.9 in more than 10% of the cases. Note that such frequencies already emerge in the case when there is only one neglected seasonal root, i.e. when $\phi(B)$ equals $(1+B)$.

In the case of monthly time series, the twelfth-order differencing filter $(1-B^{12})$, can be decomposed as

$$\begin{aligned}&(1-B^{12})\\&\quad=(1-B)(1+B)(1+B^2)(1+B\sqrt{3}+B^2)(1-B\sqrt{3}+B^2)(1+B+B^2)(1-B+B^2)\,,\end{aligned} \tag{12}$$

where all polynomials, except the $(1-B)$, correspond to seasonal unit roots. An expression for the $\phi(B)$ polynomial for the data-generating process (4) can now be found by combining parts of the polynomial in (13). This yields 63 data-generating processes. To save space, we

Table 1
Distribution of the R^2 values from the regression $y_t = \sum_{i=1}^{4} \delta_1 D_{it} + \eta_t$, where the DGP is $\phi(B)y_t = \varepsilon_t$, where ε_t is N(0, 1) based on 10,000 Monte Carlo replications (sample size is 120)

$\phi(B)$	Percentiles								
	1%	5%	10%	25%	50%	75%	90%	95%	99%
$(1+B)$	0.002	0.028	0.087	0.378	0.730	0.886	0.943	0.961	0.981
$(1+B^2)$	0.054	0.214	0.357	0.600	0.780	0.877	0.923	0.942	0.966
$(1+B)(1+B^2)$	0.162	0.363	0.485	0.660	0.793	0.870	0.915	0.934	0.957

Table 2
The distribution of the R^2 values from the regression $y_t = \sum_{i=1}^{12} \delta_i D_{it} + \eta_t$, where the data-generating processes are $\phi(B)y_t = \varepsilon_t$, with $\phi(B)$ equal to $(1+B)$, $(1+B^2)$, $(1+B\sqrt{3}+B^2)$, $(1-B\sqrt{3}+B^2)$, $(1+B+B^2)$, $(1-B+B^2)$, or combinations of these polynomials. The cells in the table are summary statistics for the data-generating processes corresponding to a certain number of seasonal unit roots. Based on 10,000 Monte Carlo replications (sample size is 120)

No. seasonal unit roots	Percentiles								
	1%	5%	10%	25%	50%	75%	90%	95%	99%
1	0.009	0.035	0.092	0.379	0.710	0.889	0.946	0.963	0.981
2	0.061	0.221	0.365	0.603	0.780	0.877	0.926	0.944	0.966
3	0.152	0.352	0.474	0.654	0.790	0.872	0.917	0.936	0.960
4	0.245	0.438	0.541	0.685	0.796	0.867	0.909	0.928	0.952
5	0.247	0.446	0.545	0.686	0.796	0.866	0.909	0.927	0.953
6	0.346	0.515	0.597	0.708	0.800	0.862	0.901	0.918	0.944
7	0.336	0.503	0.586	0.703	0.797	0.862	0.902	0.920	0.946
8	0.427	0.568	0.634	0.725	0.801	0.855	0.891	0.908	0.933
9	0.414	0.556	0.624	0.720	0.799	0.855	0.893	0.911	0.936
10	0.491	0.616	0.668	0.740	0.802	0.851	0.883	0.899	0.923
11	0.487	0.601	0.658	0.736	0.800	0.849	0.883	0.900	0.923

summarize the results for the various R^2 values obtained from regressions in (1) so that we only report the distribution of the R^2 values according to the number of seasonal unit roots in Table 2. We find high values of R^2 when the presence of seasonal unit roots is neglected. This tendency increases with the number of roots on the unit circle. Notice also that the figures at 75 percentile and higher appear to decrease as the number of seasonal unit roots grows, indicating that the distribution of the R^2 values tends to be more concentrated around the mode as the number of seasonal unit roots increases. As discussed below, however, our results suggest that the more seasonal unit roots in the DGP, the higher will be the R^2 in the regression (1).

Generally, the power of unit root tests can be low, and it is likely that tests for seasonal unit roots suffer from the same drawback. It may, therefore, be worthwhile to investigate the distribution of the R^2 from (1) when the DGP is, for example, $y_t = -\rho y_{t-1} + \varepsilon_t$ with ρ equal to, say, 0.9. Asymptotically, the R^2 should be zero, but given the results in Table 3, even in the near unit root case spuriously high R^2 may emerge in limited samples. In Table 3, we report on the empirical R^2 distribution in the case where this ρ takes values like 0.5 through 0.99. It is clear from the results in this table that our conjecture seems to be verified. In the case where $\rho = 0.95$, one may find R^2 values as high as about 0.4 in 20% of the cases. Furthermore, F tests for the joint significance of seasonal dummy variables in (1) can spuriously exceed 5% critical values.

Table 3
Distribution of the R^2 values from the regression $y_t = \Sigma_{i=1}^{4} \delta_i D_{it} + \eta_t$, where the DGP is $(1 + \rho B)y_t = \varepsilon_t$, where ε_t is N(0, 1) based on 10,000 Monte Carlo replications (sample size is 120)

ρ	Percentiles								
	1%	5%	10%	25%	50%	75%	90%	95%	99%
0.5	0.001	0.003	0.005	0.011	0.024	0.045	0.078	0.104	0.174
0.8	0.001	0.003	0.005	0.013	0.039	0.100	0.187	0.249	0.362
0.9	0.001	0.002	0.005	0.020	0.077	0.198	0.346	0.443	0.590
0.95	0.001	0.003	0.008	0.039	0.153	0.365	0.549	0.651	0.784
0.99	0.001	0.011	0.037	0.186	0.524	0.773	0.885	0.921	0.960

4. Concluding remarks

The distribution theory and the Monte Carlo experiments in the present paper show that neglecting seasonal unit roots can yield spuriously high R^2 values in the regression of first-differenced time series on some seasonal dummy variables, and hence suggest several implications for analyzing seasonally observed time series. The first and most obvious is that an interpretation of an R^2 value obtained from an auxiliary regression like (1) should be performed with great care. Its value can be spuriously high in the case when seasonal unit roots have been neglected, and hence it may then not be useful as a measure of seasonal variation.

Similar caution should be exercised when interpreting the parameter estimates for the seasonal dummies from regressions as (1). Time-series processes with seasonal unit roots can yield any kind of estimates for such dummy parameters. A comparison of the estimated δ_i values across several time series may then be hazardous. Given the result in Section 2, one can think of the case, for example, when two time series x_t and y_t, which are generated by $x_t = -x_{t-1} - x_{t-2} - x_{t-3} + \eta_t$ and $y_t = -y_{t-2} + \varepsilon_t$, where ε_t and η_t are independently drawn, can yield spuriously similar δ_i estimates in regressions like (1). Hence, although the two processes are completely independent, they may seem to be related with respect to their seasonally fluctuating patterns.

Since seasonal fluctuations in many macroeconomic time series seem to change over time, the presence of seasonal (near-) unit roots seems likely notwithstanding the possibly low power of currently applied test procedures. Several economic variables have been affected by exogenous shocks caused by oil crises and world-wide recessions, and one can observe that seasonal patterns for most of these series have changed throughout the years. This implies that it is unlikely that constant parameter models such as (1) yield the most accurate descriptions of economic time series. A useful description, which allows for changing seasonal patterns, is then a model that assumes the presence of seasonal unit roots in the autoregressive polynomial. An extension to include more than one series, and which facilitates the analysis of common aspects, is discussed in Lee (1992) and Engle et al. (1993).

References

Beaulieu, J.J., J.K. MacKie-Mason and J.A. Miron, 1992, Why do countries and industries with large seasonal cycles also have large business cycles? Quarterly Journal of Economics 107, 621–656.

Beaulieu, J.J. and J.A. Miron, 1991, The seasonal cycle in U.S. manufacturing, Economic Letters 37, 115–118.

Chan, N.H. and C.Z. Wei, 1988, Limiting distributions of least squares estimates of unstable autoregressive processes, Annals of Statistics, 16, 367–401.

Engle, R.F., C.W.J. Granger, S. Hylleberg and H.S. Lee, 1993, Seasonal cointegration: The Japanese consumption function, Journal of Econometrics 55, 275–298.

Franses, P.H., 1994, A multivariate approach to modeling univariate seasonal time series, Journal of Econometrics 63, forthcoming.

Gladyshev, E.G., 1961, Periodically correlated random sequences, Soviet Mathematics, 2, 385–388.

Hylleberg, S., R.F. Engle, C.W.J. Granger and B.S. Yoo, 1990, Seasonal integration and cointegration, Journal of Econometrics, 44, 215–238.

Hylleberg, S., C. Jørgensen and N.K. Sørensen, 1993, Seasonality in macroeconomic time series, Empirical Economics 18, 321–335.

Lee, H.S., 1992, Maximum likelihood inference on cointegration and seasonal cointegration, Journal of Econometrics 54, 1–47.

Lee, H.S. and P.L. Siklos, 1992, Seasonality in economic time series: Money–income causality in US data revisited, unpublished manuscript, Tulane University.

Miron, J.A., 1990, The economics of seasonal cycles, Paper presented at the Sixth World Meeting of the Econometric Society, Barcelona.

Phillips, P.C.B., 1986, Understanding spurious regressions in econometrics, Journal of Econometrics 33, 311–340.

Phillips, P.C.B., 1987, Time series regression with a unit root, Econometrica 55, 277–301.