A Note on the Effect of Seasonal Dummies on the Periodogram Regression

Marius Ooms and Uwe Hassler

May 20, 1996

Econometric Institute, Erasmus University Rotterdam
and
Institute for Statistics and Econometrics, Free University Berlin

Correspondence to:
Marius Ooms
Econometric Institute, Erasmus University Rotterdam
PO Box 1738, NL- 3000 DR Rotterdam, The Netherlands
email: ooms@ect.few.eur.nl

Abstract

We discuss how prior regression on seasonal dummies leads to singularities in periodogram regression procedures for the detection of long memory. We suggest a modified procedure. We illustrate the problems using monthly inflation data from Hassler and Wolters (1995).

Keywords
Long Memory, Seasonal Adjustment

1 A Numerical Problem

In this section we propose a numerical problem which came up when we redid a regression analysis of simple fractionally integrated models for inflation, where two packages came up with different answers. In the next sections we propose solutions.

Periodogram regression is by now a standard procedure to start the examination of long memory in a time series. A popular model with long memory is the ARFIMA(p, d, q) model: $\Phi(L)(1-L)^d y_t = \Theta(L)\upsilon_t$, $t = 1, 2, \ldots$ with $\Phi(L)$ and $\Theta(L)$ polynomials of orders p and q in the lag operator $L : L^k y_t = y_{t-k}$ and $(1-L)^d = 1 - dL - \frac{d(1-d)}{2}L^2 - \ldots$, see e.g. Hosking (1981). In the first stage of the estimation of an ARFIMA(p, d, q) model one simply regresses the log periodogram on the logarithm of the spectrum of simple fractionally integrated process to obtain an estimate of the fractional order of integration d. One uses the following equation:

$$\ln I(\omega_j) = \ln f_u(0) + \delta R_j + \epsilon_j \quad j = m_1, m_1 + 1, \ldots, m \tag{1}$$

where the regressand $\ln I(\omega_j)$ is the log periodogram at frequency $\omega_j = 2\pi j/T$, with T the number of observations, where the constant $\ln f_u(0)$ is the log of spectrum at zero of $(1-L)^d y_t = u_t$, where the regressor R_j is defined by $R_j = -\ln 4\{\sin^2(\omega_j/2)\}$ and where the error term $\epsilon_j = \ln\{I(\omega_j)/f(\omega_j, d)\}$ measures differences between the periodogram and the model spectrum $f(\omega_j, d)$. Robinson (1995) showed that standard regression results can be employed to test hypotheses about the fractional integration parameter d using the OLS estimate δ in (1), provided m_1 and m are chosen appropriately. The only modification in the inference compared with OLS is that the error variance is fixed at $\pi^2/6$. For the simple fractionally integrated process $(p = q = 0)$ one should use as many independent periodogram points as possible: $m_1 = 1$, $m = T/2$, in finite samples, see e.g. Hurvich and Beltrao (1994)
.

Hassler and Wolters (1995) found the simple fractionally integrated process to provide a good description of 5 post war consumer price inflation series, and compared periodogram estimates of d with estimates from more efficient procedures.

First they regressed monthly changes in logs of price indices on seasonal dummies to get rid of seasonal variation. They used the sample period 1969.01-1992.09 $(T = 285)$. Both dummy regression and periodogram regression were done in MicroTSP. They fixed $m_1 = 1$ and presented results for a range of values of m. In the upper panel of Table 1 their results for U.K. inflation are extended with results for shorter sample periods, obtained by subsequently deleting observations from the beginning of the sample period.

Table 1 *around here*

We redid their regressions using a different computer package written in Borland Pascal. We obtained much larger estimates for the integration parameter for some sample periods for some choices of m, see the lower panel of Table 1 For sample size 283 and 281 the results of the different panels agree. For sample size 285, 284, 282 and 280 they agree only for low values of m. How come?

2 Singularities in the log periodogram

Numerical problems are usually do to (near-)singularities. This problem is no exception. A simple examination of the values of the regressand $\ln I(\omega_j)$ shows some large negative values, pointing to values close to zero for the periodogram ordinate. The singularity of the periodogram at the seasonal frequencies for seasonally adjusted data is a well known feature for data series containing full years of data. Depending on the number of observations this extreme singularity problem can pop up at one or more frequencies. For $T = 283$ and $T = 281$ it does not occur, but this is not to say that there is no need to worry in that case.

The following theorem states that the periodogram ordinates of a seasonally adjusted series are zero at frequencies $2\pi i/s$, $i = 0, \ldots, s$ where s is the number of observations per year.

Theorem

Let $y_{t,}$ $t = 1, \ldots, T$, be a time series contained in the $T \times 1$ vector y. Let $x = M_D y$ be the seasonally adjusted time series vector obtained by regression on a complete set of s seasonal dummy variables with period s contained in the $T \times s$ matrix D, $M_D = I_T - D(D'D)^{-1}D'$. Then the periodogram $I_x(\lambda_i)$ for x equals zero at frequencies $\lambda_i = 2\pi i/s$, $i = 0, \ldots, s$.

The appendix contains the proof of the theorem, which is based on the regression interpretation of the Discrete Fourier Transform. Consider the case of Hassler and Wolters, $s = 12$, $T = 285$, where the periodogram is computed at frequencies $2\pi j/285, j = 1, \ldots, 142$. The first frequency with a singularity appears for $j = 95$, since $95/285 = 4/12$. Theoretically $\ln I(\omega_{95})$ would be minus infinity and the estimate of d using this ordinate would be ill defined, but in practice using finite precision in the computation finite negative values will be obtained. The larger the precision, the more negative the log periodogram ordinate, the larger the estimate of d. For $T = 284$ one obtains singularities at $j = 71$, for $T = 282$ at $j = 47$, 94 and for $T = 280$ at $j = 70, 140$. Suppose we would have full years of data *e.g.* 24 years: $T = 288$. Then we get the familiar case with singularities at all the seasonal frequencies: $2\pi i/12$, $i = 0, 1, \ldots, 6$, i.e. at $2\pi j/288$, $j = 0, 24, \ldots, 144$. The periodogram regression results in Hassler and Wolters ($T = 285$) for $m > 95$ are artificial, since they include a singular frequency. This explains why Package 1 and Package 2 differ for $m = 100, 120, 140$. Please note that no singularities arise for $T = 281$ and $T = 283$; that is why the corresponding columns show no differences in Table 1.

3 Proper Estimation

How do we avoid using the spurious estimates involving the singularities? One way would be to omit only the singular ordinates from the regression. This could lead to the situation of using a decreasing number of periodogram points with an increasing number of observations, e.g. 143 points for $T = 287$ and 138 points for $T = 288$. A preferable way is to extend the original data set to full years by adding zeros at the end, i.e. by "zero padding" as it is called in the popular econometric software package RATS. The periodogram of this extended series will contain ordinates for all the seasonal frequencies. These ordinates are then omitted in the subsequent periodogram regression. This has the additional advantage that the subsequent estimator of d does no longer depend on the regression estimates for the seasonal means. These means are hard to estimate in models with long memory. See e.g. Samarov and Taqqu (1988), who discussed the efficiency of regression estimation of the mean for fractionally integrated processes in detail. Note that our procedure makes prior regression on seasonal dummies obsolete. In sum: instead of seasonal adjustment in the time domain by prior regression, we suggest seasonal adjustment in the frequency domain by omitting all periodogram ordinates at the seasonal frequencies.

In Table 2 we present the results of this estimation procedure for the four "seasonal" inflation series analyzed in Hassler and Wolters (1995), which are now also reliable for $m > 95$ as well. We also show outcomes for the asymptotically efficient approximate frequency domain ML estimator for the simple fractionally integrated process applied in Boes et al. (1989), which minimizes

$$\sum_j \ln g(\omega_j, \delta) + T \ln \left(\frac{2\pi}{T} \sum_j \frac{I(\omega_j)}{g(\omega_j, \delta)} \right) \tag{2}$$

over δ, where $g(\omega_j, \delta) = \frac{2\pi}{\sigma_v^2} f(\omega_j, \delta) = \{4 \sin^2(\omega_j/2)\}^{-\delta}$. Here periodogram ordinates with zero values do not lead to numerical problems. Seasonal adjustment is again done by zero padding and omission of the contribution of the seasonal frequencies in the objective function (2). Note that minimizing only the second term of (2) leads to the "simple" Whittle estimator suggested by Fox and Taqqu (1986), which Hassler and Wolters applied to check their periodogram regression results. Robinson (1994) provided an overview of different frequency domain estimators of the fractional integration parameter. The results of the periodogram regression and the two approximate ML estimators are now close, see the last rows of Table 2.

Table 2 *around here*

One might be tempted to use seasonal adjustment for "stochastic seasonality" like Census X-11 or ARIMA model based methods as an alternative way to avoid the singularities. That is not an option. These seasonal adjustment methods introduce "seasonal moving average unit roots" in the adjusted series. See Maravall (1993, p.23) for a theoretical account. This leads to singularities in the log of the model spectrum for the adjusted series. The use of sample spectrum ordinates around the seasonal frequencies in the periodogram regression for the seasonally adjusted series will therefore lead to artificial results as well. See Ooms (1994, p. 274) for an empirical illustration of this phenomenon using a seasonal extension of the periodogram regression.

4 Conclusions

Prior regression on seasonal dummies can lead to artifacts in subsequent periodogram regressions for the detection of long memory. We suggest a combination of zero padding and seasonal adjustment in the frequency domain to avoid this problem. This method can also be applied to approximate frequency domain ML estimators. The problems are illustrated using data from Hassler and Wolters (1995). Our modified periodogram regression confirms their fractional specification and provides an even closer agreement between periodogram regression and frequency domain ML estimation.

This example clearly shows the benefit of checking empirical results across computer programs to reveal hidden numerical problems. It also shows the benefit of influence analysis as a standard procedure in empirical regressions, even in auxiliary regressions.

References

[1] Boes, D.C., Davis, R.A., and Gupta, S.N. (1989). Parameter estimation in low order fractionally differenced ARMA processes. *Stochastic Hydrology and Hydraulics*, 3:97–110.

[2] Fox, R., and Taqqu, M.S. (1986). Large-Sample properties of Parameter estimates for Strongly Dependent Stationary Gaussian Time Series. *The Annals of Statistics*, 14:517–532.

[3] Harvey, A.C. (1993). *Time Series Models, second edition.* Harvest Wheatsheaf, New York, London.

[4] Hassler, U. and Wolters, J. (1995). Long Memory in Inflation Rates: International Evidence. *Journal of Business and Economic Statistics*, 13:37–46.

[5] Hosking, J.R.M. (1981). Fractional Differencing. *Biometrika*, 68:165–176.

[6] Hurvich, C.M., and Beltrao, K.I. (1994). Automatic Semiparametric Estimation of the Memory Parameter of a Long-Memory Process. *Journal of Time Series Analysis*, 15:285–302.

[7] Maravall, A. (1993). Stochastic Linear Trends. *Journal of Econometrics*, 56:5–37.

[8] Ooms, M. (1994). *Empirical Vector Autoregressive Modeling*, volume 407 of *Lecture Notes in Economics and Mathematical Systems.* Springer-Verlag, Berlin, Germany.

[9] Robinson, P. (1994). Time Series with Strong Dependence. In Sims, C.A., editor, *Advances in Econometrics, sixth World Congress*, pages 47–95. Cambridge University Press, Cambridge.

[10] Robinson, P.M. (1995). Log-Periodogram Regression of Time Series with Long Range Dependence. *Annals of Statistics*, 23:1048–1072.

[11] Samarov, A., and Taqqu, M.S. (1988). On the Efficiency of the Sample Mean in Long-Memory Noise. *Journal of Time Series Analysis*, 9:191–210.

Appendix: Proof of Theorem

Proof of Theorem 1

We prove the theorem for even T and even s. The proof for odd T and/or odd s proceeds along similar lines. We use the regression interpretation of the Fourier Transform. Define the Discrete Fourier representation for y_t as, see e.g. Harvey (1993, sect. 6.2)

$$y_t = T^{-\frac{1}{2}} a_0 + (T/2)^{-\frac{1}{2}} \sum_{j=1}^{n-1} (a_j \cos \omega_j t + b_j \sin \omega_j t) + T^{-\frac{1}{2}} a_n (-1)^t, \tag{3}$$

$t = 1, \ldots T$, T even, n=T/2. The last term in the summation drops out if T is odd. Define the periodogram ordinates for y_t at the standard frequencies as:

$$I_y(\omega_j) = \frac{1}{2\pi T} \left| \sum_{t=1}^{T} y_t e^{it\omega_j} \right|^2, \ \omega_j = 2\pi j/T, \ j = 0, \ldots, n, \ n = T/2.$$

These can also be written, see Harvey (1993, sect. 6.2), as:

$$I_y(\omega_j) = \frac{1}{4\pi}\left(a_j^2 + b_j^2\right), \; j = 0, \ldots, n$$

with $b_0 = b_n = 0$. For odd T one has $n = (T-1)/2$.

For matrix notation we define the orthogonal $T \times T$ matrix Z and the $T \times 1$ vector $\gamma_y = (a_0, a_1, b_1, a_2, \ldots, a_n)'$, according to the definition of α_j and β_j in (3), so that we can rewrite (3) as $y = Z\gamma_y$. Consequently one has $\gamma_y = Z'y$.

Look then at the corresponding representation for the seasonally adjusted vector x.

Define $\gamma_x = (c_0, c_1, d_1, c_2, \ldots, c_n)' = Z'x$, so that

$$I_x(\omega_j) = \frac{1}{4\pi}\left(c_j^2 + d_j^2\right), \; j = 0, \ldots, n \tag{4}$$

with $d_0 = d_n = 0$. Let Z_j be the $T \times 2$ sub matrix of Z that corresponds to the parameters c_j and d_j , $j = 1, \ldots, n-1$ so that

$$\begin{pmatrix} c_j \\ d_j \end{pmatrix} = Z_j'x \, .$$

Let $S(D)$ be the space spanned by seasonal dummy columns in the matrix D. Let Z_i correspond to parameters c_i and d_i for the frequencies $\lambda_i = 2\pi i/s$, $i = 1, \ldots, s/2-1$. Let $Z_{tj}, t = 1, ..., T$, denote the rows of Z_j. Z_i lies in $S(D)$, since $Z_{ti} = Z_{t'i}$ for $\left|t - t'\right| = ks$, $k = 0, 1, \ldots, [T/s]$. The vector x lies in the orthogonal complement of $S(D)$ by construction. Thus $Z_i'x = 0$ for $i = 1, \ldots, s/2$. Therefore $c_j = d_j = 0$ for $j = i/s$, $i = 1, 2, \ldots, s-1$. Analogously we have $c_0 = Z_0'x = 0$ and $c_n = Z_n'x = 0,$ corresponding to λ_0 and $\lambda_{s/2}$, respectively. Consequently $I_x(\lambda_i) = 0$ for $\lambda_i = 2\pi i/s$, $i = 0, 1, \ldots, [s/2]$. Finally one has $I_x(\lambda_i) = I_x(2\pi - \lambda_i)$, which completes the proof.

Table 1: *Results of Periodogram Regressions for Seasonally Adjusted U.K. Inflation Rate With Increasing Range m for varying sample sizes T and two computer packages. Estimates of d.*

Package 1 (Hassler and Wolters (1995))							
m	SE	$T = 285$	284	283	282	281	280
20	.182	.59	.60	.63	.65	.65	.64
40	.119	.57	.56	.54	.52	.53	.54
60	.095	.39	.39	.43	.64*	.41	.42
80	.081	.45	.69*	.44	.53*	.45	.65*
100	.073	.55*	.56*	.42	.60*	.41	.52*
120	.068	.54*	.53*	.46	.57*	.46	.50*
140	.064	.51*	.50*	.44	.51*	.43	.59*
Package 2							
m	SE	$T = 285$	284	283	282	281	280
20	.182	.59	.60	.63	.65	.65	.64
40	.119	.57	.56	.54	.52	.53	.54
60	.095	.39	.39	.43	.83*	.41	.42
80	.081	.45	.84*	.44	.61*	.45	.80*
100	.073	.69*	.65*	.42	.76*	.41	.61*
120	.068	.63*	.59*	.46	.67*	.46	.56*
140	.064	.58*	.54*	.44	.57*	.43	.69*

NOTE: Estimation period for $1969.01 + k$, $k = 0, 1, 2, 3, 4, 5$ until 1992.09. Changes in log consumer price index, from the OECD Main Economic Indicators. SE denotes approximate standard error for $T = 285$. Asterisks indicate differences between packages.

Table 2: *Adjusted Periodogram Regression for four countries. Results for increasing range, with zero padding and frequency domain seasonal adjustment.* Estimates of d.

m	U.K.	France	Germany	Italy
20	.57	.74	.71	.55
40	.53	.60	.36	.52
60	.42	.49	.34	.44
80	.47	.48	.33	.51
100	.42	.46	.30	.51
120	.45	.46	.30	.52
140	.45	.47	.33	.51
FDML	.40	.50	.36	.51
SE	(.045)	(.042)	(.045)	(.045)
Whittle	.39	.48	.35	.50

NOTE: FDML: approximate Frequency Domain Maximum Likelihood, see objective function (2). SE: corresponding standard error. Whittle: Whittle Estimates from Hassler and Wolters (1995), computed using only the second term of objective function (2).