Seasonality in Revisions of Macroeconomic Data

Philip Hans Franses*
Rene Segers

Econometric Institute and Tinbergen Institute
Erasmus University Rotterdam

ECONOMETRIC INSTITUTE REPORT EI 2008-09

This version: April 14th, 2008

Abstract

We analyze five vintages of eighteen quarterly macroeconomic variables for the Netherlands and we focus on the degree of deterministic seasonality in these series. We document that the data show most such deterministic seasonality for their first release vintage and for the last available vintage. In between vintages show a variety of seasonal patterns. We show that seasonal patterns in later vintages can hardly be predicted by those in earlier vintages. The consequences of these findings for the interpretation and modeling of macroeconomic data are discussed.

Keywords: Seasonality, real-time data
JEL Classification Codes: C32, C52, C82, E20

* Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, the Netherlands. Tel.: +31-10-4081273, fax: +31-10-4089162. E-mail addresses: franses@few.eur.nl, rsegers@few.eur.nl.
We thank Dick van Dijk for his helpful comments.

1. Introduction

With the advent of real time databases, carefully compiled by academics and statistics institutes alike, one can observe a growing interest in analyzing the properties of various vintages of data. It is of course tremendously relevant to understand what first-release data actually tell us about the economy, and also what later vintages can add to that. There are many recent studies on the properties of real time data, and recent summaries are given in Croushore (2006) and in Corradi *et al*. (2008). So far, the literature did not address the issue of seasonality in vintages of data, nor on whether such seasonality is constant across vintages or not. It is this issue that we address in the current paper.

One reason why seasonality is rarely considered is that quite often only seasonally adjusted data are available. Indeed for the US, data on various vintages are available but only after seasonal adjustment. In contrast, Statistics Netherlands has compiled a real-time database which gives only seasonally unadjusted data. Even though these data are reported in terms of annual growth rates (in time series jargon, after applying the fourth-differencing filter to quarterly data), we can still analyze seasonal patterns when we put forward the proper tool for analysis, as we will do in Section 2 below. We conjecture that with an airline type of model, and drawing on the results in Bell (1987), it is possible to estimate the degree of deterministic seasonality in the data. Section 2 describes the model used.

In Section 3 we apply our methodology to five vintages of eighteen important quarterly macroeconomic quarterly observed variables, such as Gross Domestic Product, Consumption by Households and Exports. We find that there is much variation in seasonality across the vintages for about all eighteen variables. We also document that there is strong variation across the eighteen different parameters for each of the five vintages. Next, we find that there is not much correlation between these parameters, which implies that each vintage's deterministic-seasonality parameter can only mildly be predicted by the recent vintage parameter. Moreover, seasonal patterns in the last vintage data cannot be inferred from seasonality in the first release data. Section 4 discusses the implication of these findings.

2. Estimating the Degree of Deterministic Seasonality

We aim to estimate the degree of deterministic seasonality, as this parameter can be retrieved from a time series model for the annual growth rates of otherwise seasonally unadjusted data. Indeed, the data that we have are approximately of the format $\Delta_4 y_t$, where y_t denotes the natural logarithm of a quarterly macroeconomic variable as measured in quarter t. The way we can find the degree of deterministic seasonality in y_t follows from the results in Bell (1987). In this paper, it is shown that when $\theta_4 = -1$ in the moving average model of order 4 [MA(4)], written as

$$(1) \qquad \Delta_4 y_t = \mu + (1 + \theta_4 L^4)\varepsilon_t ,$$

where L denotes the familiar lag operator and where ε_t is a standard white noise variable, that then the model

$$(2) \qquad y_t = \delta_1 D_{1,t} + \delta_2 D_{2,t} + \delta_3 D_{3,t} + \delta_4 D_{4,t} + \varepsilon_t$$

appears. Hence when $\theta_4 = 0$ in (1), y_t is a seasonal random walk with seasonal fluctuations that can vary widely over the sample period. At the other end, when $\theta_4 = -1$ the seasonal fluctuation in y_t are fully deterministic, see Franses and Paap (2004) and Ghysels and Osborn (2001) for recent surveys on models for seasonality. In sum, the deterministic seasonality parameter of interest is θ_4 in a model as in (1).

A prior analysis of the data to be analyzed in the next section indicated that

$$(3) \qquad \Delta_4 y_t = \mu + \rho \Delta_4 y_{t-1} + (1 + \theta_1 L)(1 + \theta_4 L^4)\varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2)$$

is an appropriate model for all data, that is, the estimated residuals do not show strong signs of residual autocorrelation, nor of autoregressive conditional heteroskedasticity (ARCH), nor of extreme non-normality. In the next section we will estimate the parameter θ_4 for eighteen series across five vintages. In a next round of analysis, we will examine the correlations across these estimates.

In order to draw potentially stronger conclusions we will also estimate model (3) when restricting the MA parameters to be the same across the eighteen variables. That is, we then consider

$$(4) \qquad \Delta_4 y_{i,t} = \mu_i + \rho_i \Delta_4 y_{i,t-1} + (1+\theta_1 L)(1+\theta_4 L^4)\varepsilon_{i,t}, \qquad \varepsilon_{i,t} \sim N(0,\sigma_i^2),$$

where the different variables are indexed by *i*. This yields an estimate of θ_4 with a potentially smaller standard error (at least, if the pooling assumption is valid).

The applied econometrics literature contains many studies that examine the type of seasonality in macroeconomic data. Tests for seasonal unit roots –1 and ±i, using one of the many variants of the Hylleberg *et al*. (1990) method, usually reveal that not all of these roots are present in the data, see for example Osborn (1990) and Franses (1996) for surveys. Hence a seasonal random walk model is unlikely to characterize macroeconomic variables, but some stochastic seasonality is present. Harvey's structural time series models are also frequently considered for such data, and then it is typically found that the variance in the seasonal component equation is small, see Harvey (1989) and others. Upon writing such structural time series models into an ARMA representation as in (3), this small variance leads to values of θ_4 close to –1. Finally, Franses and Paap (2004) argue that the so-called periodic integration model is best to describe macroeconomic data. When this model is approximated by a non-periodic ARMA model, it is found again that the θ_4 parameter in a model as in (3) is close to –1. Taking all this evidence in the available literature together, we are inclined to put forward a first hypothesis, which is

H_1: The final vintage of quarterly seasonally unadjusted macroeconomic data shows seasonality that is close to deterministic and hence a model as in (3) has a θ_4 parameter with a value close to –1.

We will examine this hypothesis by looking at the parameter estimates for θ_4 in the final wave of data.

Now, how about the seasonality properties of earlier releases? A substantial part of the first vintage necessarily concerns forecasted values. This means that the nature of the seasonality in the first vintage depends on the forecasting model. With respect to the latter, we hypothesize that there are two most likely scenarios. The first is

H_{2a}: The first vintage of quarterly seasonally unadjusted macroeconomic data shows close to deterministic seasonality as the forecasting models used for the components are models with deterministic seasonality, that is, as in (2).

Aggregation of all components gives a variable that is close to deterministic, and hence θ_4 is then also close to –1. The second plausible scenario is that

H_{2b}: The first vintage of quarterly seasonally unadjusted macroeconomic data shows strong stochastic seasonality, as the components are forecasted using models as in (1) with $\theta_4 = 0$, that is, by simply using forecasts for annual growth rates.

Aggregation then leads to close to fully stochastic seasonality. For the data releases in between the first and final vintage, that is, the fifth vintage, Statistics Netherlands computes the quarterly data from updated forecasts for annual time series. Hence, in between the first and last vintage, the data are re-allocated across quarters. Therefore one may expect more randomness in these vintages.

If we would estimate and arrange the five values of θ_4 using model (3) for actual data, then H_1 with H_{2a} should show (over the five waves) an inverted U shape (like: ∩), while H_1 with H_{2b} would show a downward sloping trend (like: \). In the next section, we will report on the empirical results.

3. Results for Eighteen Dutch Macroeconomic Variables

Statistics Netherlands compiles a real-time database concerning eighteen macroeconomic variables and makes it available to the general public. We have quarterly data for the period 1990Q1 – 2007Q2, however, in order to harmonize the estimation samples we have the samples end in 2004Q4. A data summary is given in Table 1.

Insert Table 1 about here

There are five vintages of data. The first (*i*) is the so-called Flash estimate, which is released 45 days after the end of each quarter. The second (*ii*) is the regular quarterly estimate, released 90 days after the end of each quarter. The third (*iii*) concern the preliminary annual estimates for each quarter, released 6 months after the end of the fourth quarter, from which new quarterly data are constructed. The fourth (*iv*) concerns the second preliminary annual estimates for each quarter, released 18 months after the end of the fourth quarter. Finally, the fifth vintage (*v*) involves the final annual estimates for each quarter, released 30 months after the end of the fourth quarter

Insert Tables 2 and 3 about here

As said, for each of the eighteen variables and for each of the five vintages we estimate the parameters of model (3). We report the estimates of θ_4 in Table 2, and we give a summary of these values in Table 3. Looking at the minimum and maximum values of these estimates in Table 3 we see a strong variation. On the other hand, some of these extreme values are due to one or a few variables, and hence the median estimate seems quite reliable. The relevant numbers in Table 3, which are –0.893, –0.266, –0.384, –0.398 and –0.526 for the five vintages, respectively, seem to give strong support for H_1 with H_{2a}. When we consider the pooled model in (4), with parameter estimates given in Table 4, we find similar support for the same hypotheses.

Insert Table 4 about here

Insert Figure 1 about here

See also Figure 1 which summarizes these results graphically. In sum, seasonal patterns in these macroeconomic data are mostly of a deterministic nature for the first and final, that is, fifth release data.

Insert Tables 5 and 6 about here

That the second to fourth release data have seasonal patterns that differ from those at the start and at the end, can also be observed from the estimated correlations between the eighteen estimates, as displayed in Table 5. Vintage (*v*) shows largest correlation with vintage (*iv*) and a little less with vintage (*iii*), but much less with, for example, vintage (*ii*). This is even further substantiated by the regression results in Table 6, which shows that the degree of deterministic seasonality in the final release data can be predicted with a fit of 72% from al previous measures of that degree. The parameters in vintages (*ii*) and (*iii*) can hardly be foreseen.

4. Discussion and Conclusion

In this paper we analyzed five vintages of eighteen Dutch quarterly macroeconomic variables, focusing on the degree of deterministic seasonality in these series. The data show most such deterministic seasonality for their first and final vintages.

What are the potential consequences of such changing nature of seasonality? First, quarter-to-quarter changes will become much more difficult to interpret for the intermediate vintages of the data. Second, as seasonal patterns become confounded at the end of the sample, we would recommend excluding the last few years for modelling. Third, potentially we need to rethink the seasonal adjustment methods of sequential vintages of data as these methods cannot be the same across vintages. Fourth, and related, is the question whether our findings can help to explain why rationality across adjusted and unadjusted data seems to differ, as is documented in Kavajecz and Collins (1995) and Swanson and van Dijk (2006).

A. Tables and Figures

Table 1: The effective samples

Variable	Estimation sample	
	First release	Later releases
Gross Domestic Product	1990Q2-2004Q4	1990Q2-2004Q4
Gross Investments Industry	2001Q2-2004Q4	1990Q2-2004Q4
Gross Investments Government	2001Q2-2004Q4	1990Q2-2004Q4
Consumption Households	1990Q4-2004Q4	1990Q2-2004Q4
Consumption Government	1999Q3-2004Q4	1990Q2-2004Q4
Imports	1999Q3-2004Q4	1990Q2-2004Q4
Value Added, Construction	2001Q2-2004Q4	1990Q2-2004Q4
Value Added, Energy	2001Q2-2004Q4	1990Q2-2004Q4
Value Added, Commerce	2001Q2-2004Q4	1990Q2-2004Q4
Value Added, Industry	2001Q2-2004Q4	1990Q2-2004Q4
Value Added, Agriculture	2001Q2-2004Q4	1990Q2-2004Q4
Value Added, Government	2001Q2-2004Q4	1990Q2-2004Q4
Value Added, Transport	2001Q2-2004Q4	1990Q2-2004Q4
Value Added, Health Care	2001Q2-2004Q4	1990Q2-2004Q4
Exports	1999Q3-2004Q4	1990Q2-2004Q4
Value Added, Financial Sector	2001Q2-2004Q4	1990Q2-2004Q4
Value Added, Mining	2001Q2-2004Q4	1990Q2-2004Q4
Total Value Added	2001Q2-2004Q4	1990Q2-2004Q4

Table 2: "Deterministic seasonality" parameters for each of the five vintages of data (estimated standard errors are in parentheses) estimated separately for each of the 18 series

Variable	Estimated θ_4 parameters for vintage				
	(*i*)	(*ii*)	(*iii*)	(*iv*)	(*v*)
Gross Domestic Product	-0.777	-0.255	-0.680	-0.624	-0.933
	(0.105)	(0.156)	(0.123)	(0.126)	(0.031)
Gross Investments Industry	-0.897	-0.557	-0.912	-0.586	-0.523
	(0.058)	(0.122)	(0.049)	(0.124)	(0.135)
Gross Investments Government	-0.908	-0.325	-0.396	-0.416	-0.438
	(0.064)	(0.143)	(0.141)	(0.143)	(0.139)
Consumption Households	-0.360	-0.185	-0.132	-0.315	-0.260
	(0.152)	(0.160)	(0.160)	(0.154)	(0.155)
Consumption Government	0.373	0.030	0.061	-0.053	-0.413
	(0.303)	(0.159)	(0.162)	(0.182)	(0.199)
Imports	-0.884	-0.969	-0.853	-0.844	-0.885
	(0.054)	(0.035)	(0.069)	(0.068)	(0.049)
Value Added, Construction	-0.947	-0.791	-0.344	-0.369	-0.407
	(0.068)	(0.092)	(0.145)	(0.142)	(0.139)
Value Added, Energy	-0.931	-0.859	-0.354	-0.379	-0.498
	(0.052)	(0.056)	(0.190)	(0.159)	(0.133)
Value Added, Commerce	-0.932	-0.228	-0.966	-0.916	-0.930
	(0.049)	(0.150)	(0.025)	(0.042)	(0.038)
Value Added, Industry	0.893	-0.084	-0.696	-0.921	-0.922
	(0.075)	(0.169)	(0.106)	(0.029)	(0.023)
Value Added, Agriculture	-0.931	-0.217	0.142	0.325	-0.079
	(0.049)	(0.152)	(0.142)	(0.133)	(0.142)
Value Added, Government	0.372	0.168	0.112	0.156	0.562
	(0.335)	(0.160)	(0.173)	(0.165)	(0.123)
Value Added, Transport	-0.889	-0.301	-0.371	0.024	-0.873
	(0.110)	(0.144)	(0.151)	(0.163)	(0.075)
Value Added, Health Care	-0.908	0.052	0.181	0.916	0.909
	(0.054)	(0.173)	(0.164)	(0.024)	(0.026)
Exports	-0.866	-0.450	-0.640	-0.911	-0.738
	(0.051)	(0.130)	(0.110)	(0.035)	(0.101)
Value Added, Financial Sector	0.909	-0.049	0.004	-0.003	-0.528
	(0.042)	(0.160)	(0.169)	(0.172)	(0.148)
Value Added, Mining	-0.990	-0.906	-0.909	-0.903	-0.902
	(0.091)	(0.023)	(0.024)	(0.027)	(0.027)
Total Value Added	-0.969	-0.276	-0.908	-0.929	-0.911
	(0.051)	(0.151)	(0.042)	(0.031)	(0.026)

Table 3: Statistics of estimated parameters θ_4 (sample size is 18)

Statistics	Vintage				
	(*i*)	(*ii*)	(*iii*)	(*iv*)	(*v*)
Mean	-0.536	-0.345	-0.434	-0.375	-0.487
Median	-0.893	-0.266	-0.384	-0.398	-0.526
Minimum	-0.990	-0.969	-0.990	-0.929	-0.933
Maximum	0.909	0.168	0.181	0.916	0.909
Standard deviation	0.671	0.344	0.421	0.522	0.520

Table 4: "Deterministic seasonality" parameters for each of the five vintages of data (estimated standard errors are in parentheses) estimated using the panel data model

	Estimated θ_4 parameters for vintage				
All variables	(*i*)	(*ii*)	(*iii*)	(*iv*)	(*v*)
	-0.391	-0.364	-0.278	-0.259	-0.329
	(0.047)	(0.029)	(0.030)	(0.030)	(0.029)

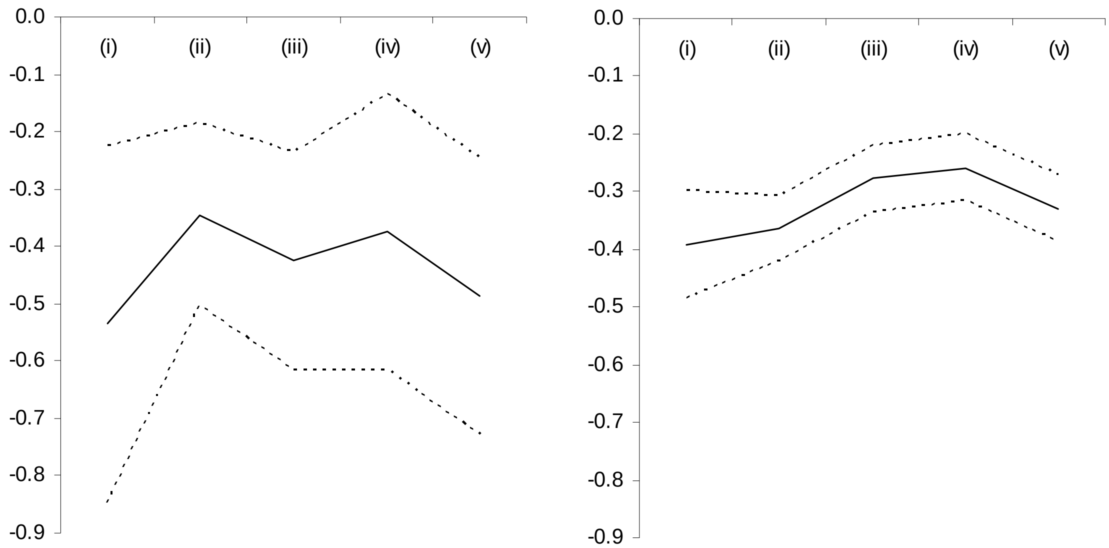
Table 5: Correlation across the estimated θ_4 parameters

Vintage	(*i*)	(*ii*)	(*iii*)	(*iv*)	(*v*)
(*i*)	1.000	0.582	0.382	0.139	0.114
(*ii*)		1.000	0.565	0.509	0.471
(*iii*)			1.000	0.894	0.777
(*iv*)				1.000	0.842
(*v*)					1.000

Table 6: Regression results for the θ_4 parameters (estimated standard errors are in parentheses)

Explanatory Variables	Dependent variable: θ_4 (*ii*)	θ_4 (*iii*)	θ_4 (*iv*)	θ_4 (*v*)
Intercept	-0.185 (0.059)	-0.188 (0.127)	0.078 (0.078)	-0.137 (0.120)
θ_4 (*i*)	0.298 (0.080)	0.051 (0.164)	-0.243 (0.095)	-0.087 (0.170)
θ_4 (*ii*)		0.634 (0.320)	0.263 (0.207)	0.157 (0.325)
θ_4 (*iii*)			1.137 (0.149)	0.261 (0.503)
θ_4 (*iv*)				0.612 (0.397)
R^2	0.339	0.324	0.864	0.718

Figure 1: Estimated average "deterministic seasonality" parameters for each of the five vintages of data. The dashed lines represent the 95%'s confidence bounds.



(a) Univariate approach

(b) Panel data model

References

Bell, W. (1987), A Note on Overdifferencing and the Equivalence of Seasonal Time Series Models with Monthly Means and Models with $(0,1,1)_{12}$ Seasonal Parts When θ = 1, *Journal of Business & Economic Statistics* **5**, 383-387

Corradi, V., A. Fernandez and N.R. Swanson (2008), Information in the Revision Process of Real Time Datasets, *Journal of Business & Economic Statistics,* in print

Croushore, D. (2006), Forecasting with Real-Time Macroeconomic data, in C.W.J. Granger, G. Elliott and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Amsterdam: Elsevier

Franses, P.H. (1996), Recent Advances in Modelling Seasonality, *Journal of Economic Surveys* **10**, 299-345

Franses, P.H. and R. Paap (2004), *Periodic Time Series Models*, Oxford: Oxford University Press

Ghysels, E. and D.R. Osborn (2001), *The Econometric Analysis of Seasonal Time Series*, Cambridge: Cambridge University Press

Harvey, A.C. (1989), *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press

Hylleberg, S, R.F. Engle, C.W.J. Granger, and B.S. Yoo (1990), Seasonal Integration and Cointegration, *Journal of Econometrics* **44**, 215-238

Kavajecz, K. and S. Collins (1995), Rationality of Preliminary Money Stock Estimates, *Review of Economics and Statistics* **77**, 32-41

Osborn, D.R. (1990) A Survey of Seasonality in UK Macroeconomic Variables, *International Journal of Forecasting* **6**, 327-336

Swanson, N.R. and D. van Dijk (2006), Are Statistical Reporting Agencies Getting it Right? Data Rationality and Business Cycle Asymmetry, *Journal of Business & Economic Statistics* **24,** 24-42