Journal of Official Statistics, Vol. 26, No. 2, 2010, pp. 361–369

Seasonality in Revisions of Macroeconomic Data

Philip Hans Franses[1] *and Rene Segers*[1]

We analyze the revision history of quarterly and monthly (seasonally unadjusted) macroeconomic variables for the Netherlands, Ireland, Luxemburg and the United States, where we focus on the degree of deterministic seasonality in these revisions. We document that the data show most deterministic seasonality in the final revision. The first-release data and the in-between revisions show a variety of seasonal patterns. The consequences of these findings for the interpretation and modeling of macroeconomic data are discussed.

Key words: Seasonality; real-time data; JEL Classification Codes: C32, C52, C82, E20.

1. Introduction

With the advent of real-time databases, carefully compiled by academics and statistical institutes alike, one can observe a growing interest in analyzing the properties of various revisions of data. It is of course tremendously relevant to understand what first-release data actually tell us about the economy, and what later revisions 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. (2009). So far, the literature has not addressed the issue of seasonality in revisions of data, including whether such seasonality is constant across subsequent revisions or not. It is this issue that we address in the current article.

One reason why seasonality is rarely considered is that quite often only seasonally adjusted data are available. Indeed for the U.S., data on various revisions are available but only after seasonal adjustment. We could find seasonally unadjusted real-time data only for Industrial Production (see below). In contrast, Statistics Netherlands, the National Statistical Institute of Luxemburg and the Central Statistics Office Ireland have compiled real-time databases which give only (or partly) seasonally unadjusted data. Even though these data are usually reported in terms of annual growth rates (in time series jargon: after applying the fourth-differencing filter to log-transformed quarterly data), we can still analyze seasonal patterns when we employ 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.

[1] Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands. Emails: franses@few.eur.nl and rsegers@few.eur.nl
Acknowledgments: We thank Dick van Dijk for his helpful comments. We also thank the Assocoate Editor and an anonymous referee for helpful comments.

In Section 3 we apply our methodology to 18 (Netherlands) plus 5 (Luxemburg) plus 17 (Ireland) plus 1 (US) series with four or five revisions of important macroeconomic quarterly (and monthly) observed variables such as Gross Domestic Product, Consumption by Households and Exports. We find that there is much variation in seasonality across the revisions for virtually all variables. We also document that there is strong variation across the different seasonality parameters for each of the revisions. Generally, the finding is that the final releases show most deterministic seasonality. 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 it is this parameter that 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.[2] The way we can find the degree of deterministic seasonality in y_t follows from the results in Bell (1987). In that study, the model

$$\Delta_4 y_t = \mu + (1 + \theta_4 L^4)\varepsilon_t \tag{1}$$

is analyzed, where L denotes the familiar lag operator and where ε_t is a standard white noise variable. Bell (1987) shows that when $\theta_4 = -1$ in this moving average model, the model

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

appears. Note that when $\theta_4 = 0$ in (1), y_t is a seasonal random walk describing data that has seasonal fluctuations that can vary widely over the sample period. And, given (1), at the other end, when $\theta_4 = -1$ the seasonal fluctuations 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 our interest is θ_4 in a model as in (1).

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

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

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 nonnormality. Hence, Model (3) seems to fit most of the data fairly well.

Real-time data sets are usually presented in a two-dimensional grid as shown in Figure 1. In this figure, each observation is represented by a square. The vertical axis shows the dates to which the observations pertain and the horizontal axis shows their publication dates. A column of observations represent the vintage of data that is released in one particular month. Thus, a vintage consists of a first estimate for the most recent quarter at

[2] For the Industrial Production series of the U.S. and most of the series for Luxemburg we have monthly data available and hence we consider $\Delta_{12} y_t$.

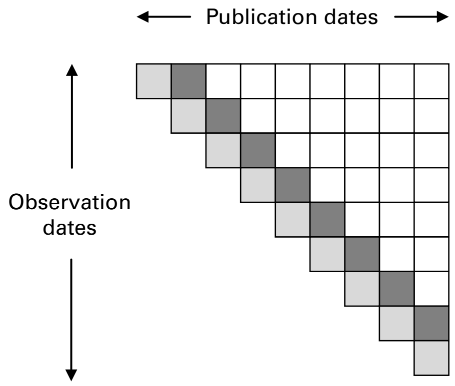

Fig. 1. The structure of real-time data sets

the date of publication, a first revision for the second most recent quarter, a second revision for the third most recent quarter, and so on. Looking at a particular row one observes the revision history of the estimate for the corresponding calendar date. Finally, starting from the bottom right corner, the diagonals group all first release data, marked by light grey squares in Figure 1, all first revised data, marked by dark grey squares in Figure 1, all second revised data, and so on. Models are typically estimated on the latest-available vintage, that is, on the data in the rightmost column in the figure. One can easily appreciate that if the underlying data generation process is different for first releases and for subsequent revisions, inference and especially forecasting on the basis of vintage data is complicated. In this article, we investigate whether indeed these data generation processes are different. In particular, we look at the degree of deterministic seasonality across the first release data and different revisions by estimating the parameter θ_4 in (3) separately for each diagonal in the real-time data set.

2.1. Hypotheses

The applied econometrics literature contains many studies that examine the type of seasonality in macroeconomic data. Tests for seasonal unit roots -1 and $\pm$ 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, for example Harvey 1989). 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 nonperiodic 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

Table 1a. Effective samples, data for the Netherlands

Variable	Estimation sample First release	Final release
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

H_1: The final revision of quarterly (or monthly) 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 revisions and the first release? A substantial part of the first-release data necessarily concerns forecasted values. This means that the nature of the seasonality in the first-release data 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 release of quarterly (or monthly) seasonally unadjusted macroeconomic data shows close to deterministic seasonality as the forecasting models used for the components are models with substantial deterministic seasonality, that is, as in (2).

Table 1b. Effective samples, data for Luxemburg

Variable	Estimation sample First release	Final release
Current Account Balance	1998Q4 -2007Q3	1995Q1-2006Q4
Total Industrial Production	1999M4 -2007M10	1995M1-2007M7
Industrial Producer Prices	1999M10-2007M11	1990M1-2008M8
National Index of Consumer Prices	2000M1 -2007M12	1996M1-2007M9
Unemployment Rate	1998M4 -2007M8	1996M3-2007M5

Table 1c. Effective samples, data for Ireland

Variable	Estimation sample	
	First release	Final release
Current Account Balance	1999Q4-2007Q2	1997Q1-2006Q3
Nominal Changes in Inventories	1999Q4-2007Q2	1997Q1-2006Q3
Real Changes in Inventories	1999Q1-2007Q2	1997Q1-2006Q3
Government Consumption Deflator	1999Q1-2007Q2	1997Q1-2006Q3
Nominal Government Consumption	1999Q1-2007Q2	1997Q1-2006Q3
Real Government Consumption	1999Q1-2007Q2	1997Q1-2006Q3
Nominal Investments	1999Q1-2007Q2	1997Q1-2006Q3
Real Investments	1999Q1-2007Q2	1997Q1-2006Q3
Nominal Imports	1999Q1-2007Q2	1997Q1-2006Q3
Real Imports	1999Q1-2007Q2	1997Q1-2006Q3
Personal Consumption Deflator	1999Q1-2007Q2	1997Q1-2006Q3
Nominal Consumption	1999Q1-2007Q2	1997Q1-2006Q3
Real Consumption	1999Q1-2007Q2	1997Q1-2006Q3
Nominal Exports	1999Q1-2007Q2	1997Q1-2006Q3
Real Exports	1999Q1-2007Q2	1997Q1-2006Q3
Nominal GDP	1999Q1-2007Q2	1997Q1-2006Q3
Real GDP	1999Q1-2007Q2	1997Q1-2006Q3

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

> H_{2b}: The first release of quarterly (or monthly) 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 release and the final revision, national statistical agencies typically compute the quarterly (or monthly) data from updated forecasts for annual time series. Hence, in between the first release and the final revision, the data are reallocated across quarters (months). Therefore one may expect more randomness in these releases.

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

Table 1d. Effective samples, data for the United States

Variable	Estimation sample	
	First release	Final release
Industrial Production	1951M4-2008M7	1950M4-2006M3

Table 2a. "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 for the Netherlands

Variable	Estimated θ_4 parameters for release				
	(*i*)	(*ii*)	(*iii*)	(*iv*)	(*v*)
Gross Domestic Product	−0.78 (0.11)	−0.26 (0.16)	−0.68 (0.12)	−0.62 (0.13)	−0.93 (0.03)
Gross Investments Industry	−0.90 (0.06)	−0.56 (0.12)	−0.91 (0.05)	−0.59 (0.12)	−0.52 (0.14)
Gross Investments Government	−0.91 (0.06)	−0.33 (0.14)	−0.40 (0.14)	−0.42 (0.14)	−0.44 (0.14)
Consumption Households	−0.36 (0.15)	−0.19 (0.16)	−0.13 (0.16)	−0.32 (0.15)	−0.26 (0.16)
Consumption Government	0.37 (0.30)	0.03 (0.16)	0.06 (0.16)	−0.05 (0.18)	−0.41 (0.20)
Imports	−0.88 (0.05)	−0.97 (0.04)	−0.85 (0.07)	−0.84 (0.07)	−0.89 (0.05)
Value Added, Construction	−0.95 (0.07)	−0.79 (0.09)	−0.34 (0.15)	−0.37 (0.14)	−0.41 (0.14)
Value Added, Energy	−0.93 (0.05)	−0.86 (0.06)	−0.35 (0.19)	−0.38 (0.16)	−0.50 (0.13)
Value Added, Commerce	−0.93 (0.05)	−0.23 (0.15)	−0.97 (0.03)	−0.92 (0.04)	−0.93 (0.04)
Value Added, Industry	0.89 (0.08)	−0.08 (0.17)	−0.70 (0.11)	−0.92 (0.03)	−0.92 (0.02)
Value Added, Agriculture	−0.93 (0.05)	−0.22 (0.15)	0.14 (0.14)	0.33 (0.13)	−0.08 (0.14)
Value Added, Government	0.37 (0.34)	0.17 (0.16)	0.11 (0.17)	0.16 (0.17)	0.56 (0.12)
Value Added, Transport	−0.89 (0.11)	−0.30 (0.14)	−0.37 (0.15)	0.02 (0.16)	−0.87 (0.08)
Value Added, Health Care	−0.91 (0.05)	0.05 (0.17)	0.18 (0.16)	0.92 (0.02)	0.91 (0.03)
Exports	−0.87 (0.05)	−0.45 (0.13)	−0.64 (0.11)	−0.91 (0.04)	−0.74 (0.10)
Value Added, Financial Sector	0.91 (0.04)	−0.05 (0.16)	0.00 (0.17)	0.00 (0.17)	−0.53 (0.15)
Value Added, Mining	−0.99 (0.09)	−0.91 (0.02)	−0.91 (0.02)	−0.90 (0.03)	−0.90 (0.03)
Total Value Added	−0.97 (0.05)	−0.28 (0.15)	−0.91 (0.04)	−0.93 (0.03)	−0.91 (0.03)
Mean	−0.54	−0.34	−0.43	−0.37	−0.49

Table 2b. "Deterministic seasonality" parameters for each of the four releases of data (estimated standard errors are in parentheses) estimated separately for each of the 5 series for Luxemburg

	Estimated θ_4/θ_{12} parameters for release			
Variable	(*i*)	(*ii*)	(*iii*)	(*iv*)
Current Account Balance	−0.45 (0.20)	−0.82 (0.11)	−0.72 (0.15)	−0.93 (0.00)
Total Industrial Production[a]	−0.87 (0.06)	−0.97 (0.05)	−0.84 (0.07)	−1.02 (0.18)
Industrial Producer Prices[a]	−0.91 (0.03)	−0.85 (0.06)	−0.86 (0.07)	−0.93 (0.04)
National Index of Consumer Prices[a]	−0.97 (0.03)	−0.94 (0.03)	−0.95 (0.03)	−1.02 (0.05)
Unemployment Rate[a]	−0.92 (0.04)	−0.82 (0.04)	−0.82 (0.04)	−0.82 (0.03)
Mean	−0.83	−0.88	−0.83	−0.94

[a] Monthly series.

3. Results for Many 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 1a. Similarly, Table 1b summarizes the data we have available for Luxemburg. These data are compiled by the National Statistical Institute of Luxemburg. Note that all series are measured monthly, except for the series Current Account Balance. For Ireland we have 17 series available, measured quarterly. These are compiled by the Central Statistics Office Ireland, and are summarized in Table 1c. Finally, we will consider monthly seasonally unadjusted U.S. Industrial Production (see Table 1d). These data are compiled from the Federal Reserve Bulletin and the Survey of Current Business (see Corradi et al. 1999 for details).

Our Dutch data consist of first-release data and four subsequent revisions. The first-release data (*i*) are so-called Flash estimates, which are released 45 days after the end of each quarter. The first revision (*ii*) is the regular quarterly estimate, released 90 days after the end of each quarter. The second revision (*iii*) concerns the preliminary annual estimates for each quarter, released six months after the end of the fourth quarter, from which new quarterly data are constructed. The third revision (*iv*) concerns the second preliminary annual estimates for each quarter, released 18 months after the end of the fourth quarter. Finally, the fourth revision (*v*) involves the final annual estimates for each quarter, released 30 months after the end of the fourth quarter. Similar agendas of data releases occur for the other countries.

As said, for each of the variables and for each of the releases we estimate the parameters of Model (3). We report the estimates of θ_4 (or θ_{12} if we consider monthly data) in Tables 2a–d. Looking at the minimum and maximum values of these estimates in the component tables in Table 2 we see a strong variation.

Table 2c. "Deterministic seasonality" parameters for each of the four releases of data (estimated standard errors are in parentheses) estimated separately for each of the 17 series for Ireland

	Estimated θ_4 parameters for release			
Variable	(*i*)	(*ii*)	(*iii*)	(*iv*)
Current Account Balance	−0.91 (0.04)	−0.43 (0.18)	−0.91 (0.04)	−0.94 (0.03)
Nominal Changes in Inventories	−0.95 (0.05)	−0.95 (0.06)	−0.94 (0.04)	−0.91 (0.04)
Real Changes in Inventories	−0.96 (0.05)	−0.93 (0.05)	−0.94 (0.04)	−0.91 (0.04)
Government Consumption Deflator	−0.93 (0.07)	−0.99 (0.00)	−0.94 (0.05)	−0.97 (0.06)
Nominal Government Consumption	−0.99 (0.07)	−0.95 (0.03)	−0.92 (0.09)	−0.89 (0.08)
Real Government Consumption	−0.92 (0.06)	−0.99 (0.00)	−0.94 (0.05)	−0.98 (0.05)
Nominal Investments	0.95 (0.06)	0.03 (0.22)	0.35 (0.18)	−0.99 (0.06)
Real Investments	−0.94 (0.04)	−0.89 (0.07)	−0.94 (0.04)	−0.99 (0.08)
Nominal Imports	−0.89 (0.05)	−0.90 (0.07)	−0.97 (0.06)	−0.99 (0.08)
Real Imports	−0.91 (0.06)	−0.95 (0.06)	−0.96 (0.06)	−0.98 (0.07)
Personal Consumption Deflator	−0.95 (0.07)	−0.97 (0.06)	−0.95 (0.06)	−0.99 (0.06)
Nominal Consumption	−0.87 (0.06)	−0.87 (0.07)	−0.88 (0.06)	−0.89 (0.05)
Real Consumption	−0.94 (0.06)	−0.93 (0.05)	−0.97 (0.06)	−0.98 (0.06)
Nominal Exports	−0.89 (0.06)	−0.90 (0.09)	−0.88 (0.07)	−0.93 (0.06)
Real Exports	−0.89 (0.06)	−0.92 (0.06)	−0.93 (0.06)	−0.92 (0.07)
Nominal GDP	−0.96 (0.04)	−0.87 (0.09)	−0.21 (0.23)	−0.89 (0.05)
Real GDP	−0.95 (0.06)	−0.91 (0.07)	−0.99 (0.07)	−0.92 (0.06)
Mean	−0.82	−0.84	−0.82	−0.94

Table 2a shows that for the Netherlands we obtain an average pattern that looks like an inverted U shape. Tables 2b to 2d indicate that for the other series we observe an average pattern with a downward-sloping trend. Across all series, the common pattern is that the final release data show most deterministic seasonality. Also, across all series, we see that intermediate releases show lesser deterministic seasonality.

Table 2d. "Deterministic seasonality" parameters for each of the five releases of data (estimated standard errors are in parentheses) estimated for U.S. Industrial Production

	Estimated θ_{12} parameters for release				
Variable	(*i*)	(*ii*)	(*iii*)	(*iv*)	(*v*)
Industrial Production[a]	−0.83 (0.01)	−0.87 (0.01)	−0.91 (0.01)	−0.97 (0.01)	−0.97 (0.02)

[a] Monthly series.

4. Discussion and Conclusion

In this article we have analyzed various revisions of many quarterly macroeconomic variables (and one monthly), focusing on the degree of deterministic seasonality in these series. The data show most such deterministic seasonality for their first releases and final revisions.

What are the potential consequences of the 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 modeling. Third, potentially we need to rethink the seasonal adjustment methods of sequential vintages of data as these methods cannot be the same across revisions. 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).

5. 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 $\theta = 1$. Journal of Business and Economic Statistics, 5, 383–387.

Corradi, V., Fernandez, A., and Swanson, N.R. (2009). 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 Handbook of Economic Forecasting, C.W.J. Granger, G. Elliott, and A. Timmermann (eds). Amsterdam: Elsevier.

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

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

Ghysels, E. and Osborn, D.R. (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., Engle, R.F., Granger, C.W.J., and Yoo, B.S. (1990). Seasonal Integration and Cointegration. Journal of Econometrics, 44, 215–238.

Kavajecz, K. and Collins, S. (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 van Dijk, D. (2006). Are Statistical Reporting Agencies Getting it Right? Data Rationality and Business Cycle Asymmetry. Journal of Business and Economic Statistics, 24, 24–42.

Received April 2008
Revised July 2009