JOINT EUROPEAN COMMISSION – OECD WORKSHOP ON INTERNATIONAL DEVELOPMENT OF BUSINESS AND CONSUMER TENDENCY SURVEYS

BRUSSELS

14 – 15 NOVEMBER 2005

Performance of Seasonal Adjustment Procedures: Simulation and empirical results*

Philip Hans Franses, Richard Paap, and Dennis Fok[1]

Economic Institute

Erasmus University Rotterdam

- May 27 2005 -

[1] This is the final report of the study on "Seasonal adjustment of business and consumer survey data" (Contract number ECFIN-195-2004/SI2.385615)

Contents

1 Introduction

This report contains the results of a study concerning the seasonal adjustment of business and consumer survey [BCS] data. We aim to evaluate three methods of seasonal adjustment, that is, Census X12-ARIMA, TRAMO/SEATS and Dainties, for these series. We first compare these methods for simulated data. It is important to note that we do not compare the seasonally adjusted data with the raw data, as is typically done in various studies, see for example Ghysels *et al.* (1996), but we choose to compare each method relative to the other methods.

Business and consumer survey data are qualitative data. The consumer surveys consist of answers to various questions, where the answers concern five possible categories. These surveys are held amongst thousands of individuals in a specific country, and over time other groups of individuals are surveyed. The business surveys mostly contain three answer categories and the sample sizes are often much smaller. The final reported figures are balances of percentages, these are bounded from above and below by 100 and -100, respectively. Minimum and maximum values can only be attained when all respondents give the same extremely negative or extremely positive answer. Hence, it is unlikely that these values will ever be reached.

Due to their qualitative nature, the BCS data have different properties than quantitative data like industrial production and employment. First, we discuss trends. The fact that the data are bounded from below and above renders a deterministic trend model implausible. Along similar arguments, one can state that a trend model of the unit root type is equally implausible, however a feature of this model is that it allows that sequences of shocks may cause the data never to come close to the boundary values. It is likely though that the data experience short periods where the levels of the BCS data are different than before or after those periods. For example, such level changes can be caused by sentiments that may have nothing to do with the questions in the survey but which can cause people to feel overly pessimistic or optimistic for a while. One way to capture this is to include temporary level shifts in the data, but this inclusion assumes that the researcher knows when these level shifts start and end. A better way to capture such time varying patterns is to impose a unit root in the AR part of the model. In other words, the data do not need to truly have a stochastic trend, but an imposition of a unit root allows one to handle various short-term level shifts at unknown occasions.

Second, and this is mainly the topic of the present report, there is the issue of seasonality. The surveys involve questions where respondents should compare the past months to future months. For the consumer surveys the comparison period is 12 months, whereas for the business surveys it is mostly 3 months. In all cases respondents are explicitly asked to disregard seasonal influences. Hence, strictly speaking, there should be no seasonality caused by the phenomenon featuring in the question (financial situation, employment, purchase of new durable goods). In other words, if there is seasonality, it is either due to the respondent's inability to disregard seasonality or to features from the outside, like the weather or festivals. Various psychological studies have shown that moods change with the season, and that people tend to be more optimistic in the beginning of a new calendar year and less optimistic by the end of summer. Therefore, seasonal variation in BCS data is probably not strong, and also, it is unlikely to change much over time. This notion is important to take along in the study below.

Third, BCS data may have outliers. Usually, we discern two types of outliers. The first and most common is the additive outlier. This is often caused by an outside phenomenon and it is often termed as "a typing error". For example, due to an event as the Exxon Valdez disaster, the outbreak of wars in the Middle East and in Africa, but also due to the winning of a soccer tournament, people may be in worse or better moods, and in that case the survey data could get unexpected higher of lower values. When such outliers occur in the first difference of a series, they mark the beginning or end of periods with level shifts.

The second type of outlier is the innovation outlier. An example for BCS data would be that, due to unforeseen events, many industrialized countries suddenly enter recessions, which may affect the level of BCS data for a long while, as people stay pessimistic for a long time. One may say that the cause of an innovation outlier should be found inside the economic system. If the data are of the unit root type, such an innovation outlier can lead to a level shift that lasts for a long time, sometimes even forever.

There are many methods to detect outliers. These are often used to downplay the relevance of the corresponding observations in estimation routines. Another angle to this issue is that one may want to design methods that are resistant (robust) to a certain amount of outliers, without having to bother about their location and origin. For BCS data it is likely that temporary level shifts occur as well as additive outliers. When a unit root is imposed, all outlier observations are then of the additive type. In this paper we seek to compare methods for seasonal adjustment, and we tend to favor those methods that are robust to an unknown but not too large amount of additive outliers at unknown locations.

The main idea behind our approach is that we examine the link between models for seasonal data and the assumptions underlying three seasonal adjustment methods. Even though the methods do not specifically assume a certain model, it can be envisaged that some methods would work best for data which could be described by a certain model. For example, if the seasonal adjustment method would simply be that one subtracts seasonal means from the data, then one can quickly understand that data according to a model with constant deterministic seasonality would be best adjusted using that particular method. In sum, we aim to see which type of data would be "best adjusted" by which method.

The first and largest part of this report contains the simulation study. This simulation study is divided in three parts. We first consider the situation where there are *no outliers* in the simulated data. We compare the relative performance of the three adjustment methods. To analyze the performance of the three adjustment methods in case there are *outliers*, we introduce innovation and additive outliers to our simulated data. We compare the results with the case of no outliers. In the third part of the simulation study we analyze the effects of *aggregation* on seasonal adjustment. We compare the situation where we first aggregate time series and then apply the seasonal adjustment methods with the situation where we apply the seasonal adjustment methods first to the individual series and then aggregate the series.

In the second part of this report we consider a large amount (that is, 300 series) of business and consumer survey data. We analyze the properties of the series in order to be able to recommend the "best adjustment method" for these data.

The outline of our report is as follows. In Section 2 we describe various general aspects of the three seasonal adjustment methods under scrutiny. These methods are Census X12-ARIMA, TRAMO/SEATS, and Dainties. For each method we describe the main idea of the method. As all methods require some human input in terms of parameter configurations, we next describe the settings we have chosen.

In Section 3, we outline the diagnostics we use to evaluate the seasonal adjustment methods. These diagnostics are statistical tests, which are based on the statistical relevance of parameters in certain regression models. These diagnostics either focus on (i) does the seasonal adjustment method effectively remove seasonality?, or on (ii) does the seasonal adjustment method change features of the time series other than seasonality? For the sake of convenience, we will use the standard or tabulated critical values of these tests, where we thus pretend we do not know the features of the data generating process

[DGP]. Indeed, in practice one is also usually uncertain of the best model for the data, so our simulation experiments mimic every day practice.

In Section 4, we describe the six data generating processes we consider in our simulations, where we briefly discuss the degree of nonstationarity of the data implied by these models. Note that we focus on quarterly data to save space. There are many indications that findings for quarterly data carry over to monthly data or data at other frequencies. We also hypothesize which adjustment method would be most associated with each of these DGPs. The first five DGPs represent the most commonly found models for seasonal data, as can be noticed from the relevant literature, see the surveys in Franses (1996b) and Ghysels and Osborn (2001). The sixth DGP is simply white noise. We choose parameters that are also commonly found in practice. In Appendix A, we give graphical impressions of the typical data that correspond to these DGPs.

In Sections 5 to 7, we present the results of our simulations. In Section 5, we directly simulate data from the DGPs. In the other two sections we consider the influence of outliers and aggregation, respectively. One of our main findings is that the Census X12-ARIMA and TRAMO/SEATS methods are most robust to variations in the data generating process. This implies that in case one would not have any strong indications as to which model could best describe the raw (unadjusted) data, then these two methods are to be preferred. On the other hand, the Dainties method performs relatively well when the data show patterns that are close to deterministic seasonality. Hence, if one suspects that the raw data are best described as such, then the Dainties method suffices.

The simulation results in Section 6 show that the effect of additive outliers on the performance of the adjustment methods is relatively small. Especially, the Census X-12 method seems to handle these outliers quite well. The adjustment methods are however very sensitive to innovation outliers. It seems that none of the methods is capable of removing these types of outliers adequately from the series before seasonal adjustment. This is especially true if the series contains seasonal unit roots.

The main conclusion of Section 7 is that the order of seasonal adjustment and aggregation hardly has an influence of the performance of the seasonal adjustment methods.

In Section 8, we compare the performance of the three seasonal adjustment methods on 300 series from the business and consumer surveys. A main conclusion from this section is that there are no marked differences in the performance of the three seasonal adjustment methods. Another conclusion is that the data are close to the "deterministic seasonality case", which implies that the Dainties methods is very useful in this context. In Section 9 we conclude with a discussion of the main results.

2 Seasonal adjustment procedures

In this section we describe the three seasonal adjustment methods under scrutiny. In the last subsection we discuss their merits relative to each other.

2.1 Census X12-ARIMA

The X12-ARIMA method is one of the most popular seasonal adjustment procedures around. It is a major improvement over the X11 method. These improvements were made by Statistics Canada and the US Census bureau, see Findley *et al.* (1998) for a detailed discussion of its capabilities. Here we give a basic summary of the methodology of the X12-ARIMA method based on Ghysels and Osborn (2001).

The core of the procedure is the X11 program, which consists of a set of moving average filters that are applied to the data. Before the data are put through these filters they are adjusted by the routine known as regARIMA. In this routine an ARIMA model is fitted to the data. This routine also allows for additional regressors to capture, for example, calendar effects.

We use the X12-ARIMA procedure with all the default settings. In accordance with the data generating processes we consider, we impose an additive seasonal pattern (so, no natural logs are taken). As regressor variables we only use a constant. The ARIMA model selection is done using the automatic procedure. We let X12-ARIMA select the best specification out of a (default) set of options.

2.2 Dainties

The Dainties procedure is the method currently used by ECFIN. In the Dainties method a seasonally adjusted series is obtained through a moving window regression. For quarterly data, the regressors are a constant, time trend, quadratic time trend, a cubic time trend and 3 seasonal dummies. Parameter estimates corresponding to the seasonal dummies are used to remove the seasonal patterns in the data.

The above regressions are performed for three different window lengths. For quarterly data the window lengths are 13, 17 or 21 quarters. Furthermore, in case the original series only contains positive values, the Dainties procedure also considers these regressions for the log-transformed series (multiplicative seasonality). For the original and the log-transformed series one obtains a possible seasonally adjusted series for each window length.

In the Dainties procedure these six (or three) series are combined by taking a quality weighted average. The weight of an adjusted series based on multiplicative seasonality is set to zero if the weight does not exceed a threshold value, where the threshold is based on the performance of the three adjusted series based on additive seasonality.

2.3 TRAMO/SEATS

The TRAMO/SEATS (Time Series Regression with ARIMA Noise, Missing Observations, and Outliers/Signal Extraction in ARIMA Time Series) (Gómez and Maravall, 1997) consists of two parts. The first part is much like the regARIMA procedure in X-12-ARIMA. It deals with preadjustment of the data. The second part (SEATS) deals with the extraction of the seasonal pattern from the selected ARIMA model.

We apply the two procedures with the default settings. To fit the data, we do not use the standard data transformation to logs. With the default settings the TRAMO/SEATS procedure uses an Airline model (see Section 4.4 below) to estimate the seasonal component of a series.

2.4 A brief comparison of the methods

An advantage of the Census X-12 method is that it goes back a long while and finds it links with smoothing and exponentially weighting methods. In principle, the method can be done by hand (at least, when there are no outliers, level shifts, etcetera), and first-order approximations to the full Census X-12 filter are easy to compute, see Franses (1996a, Table 4.1). A drawback of the method is that it needs past and future observations of a series, in order to seasonally adjust this month's observation. As these future data are not available, one usually relies on forecasts, that replace the true observations. These forecasts are often obtained from an Airline model, see below. As the forecasts improve when new observations become available, the Census X-12 program implies that one will have to revise the seasonally adjusted data in the future. Note that in a comparison study, the "final values" are obtained for almost all periods. In a real-time context the provisional concurrent estimates will have to be used, but we do not consider this situation. A second drawback of this method is that it does not deliver standard errors. Indeed, seasonally adjusted data are estimates, and it can be useful to have their associated standard errors, see Koopman and Franses (2002).

The Dainties method is based on a local regression. An advantage is that it can be useful for data where seasonality changes only very slowly or not at all. Another advantage of Dainties is that it does not lead to revisions. Indeed, data revisions are difficult to justify in the case of opinion data as opinions expressed in the past will not change due to incoming new data. A drawback of (the current version) of Dainties is that it does not include a outlier-correction method. This would also not be straightforward, as it concerns local regressions. Hence, what could be an outlier in a certain window of data, might not be so when the next few data points are included and the first few data points are discarded. One can now rely on full sample outlier detection methods, but this discards the notion behind the local regression, because these methods assume a certain model for the data. A possible solution could be to rely on winsorizing techniques, that is, to set certain observations at the value of the mean plus or minus k times the standard error. Although, again, this potentially is in conflict with the notion that outliers should be defined for the entire series and not just for a part of the series.

Finally, TRAMO/SEATS also implies data revisions, so this is a drawback, but on the other hand, it has good procedures for handling outliers and level shifts.

3 Diagnostic tests

In this section we discuss a number of diagnostic and specification tests that can be applied to judge the quality of a seasonal adjustment procedure. Each test focusses on a property that should (or should not) be present in seasonally adjusted data. We consider tests for the presence of seasonal unit roots, the presence of changing seasonal means, the presence of deterministic seasonality, the presence of correlation in the seasonal lag, the presence of periodicity in the autoregressive parameters and the presence of seasonality in the variance of the series. In this section we focus on tests for quarterly data, most tests can easily be extended to monthly data.

3.1 HEGY test

The unit roots in seasonal data, which can be associated with changing seasonality, are the so-called seasonal unit roots, see Hylleberg *et al.* (1990). For quarterly data, these roots are -1, i, and $-$i. For example, data generated from the model $y_t = -y_{t-1} + \varepsilon_t$ would display seasonality. Similar observations hold for the model $y_t = -y_{t-2} + \varepsilon_t$, which can be written as $(1 + L^2)y_t = \varepsilon_t$, where the autoregressive polynomial $1 + L^2$ corresponds to the seasonal unit roots i and $-$i, as these two values solve the equation $1 + z^2 = 0$. Hence, when a model for y_t contains an autoregressive polynomial with roots -1 and/or i, $-$i, the data are said to have seasonal unit roots.

To test for the presence of seasonal unit roots, we consider in this section the approach of Hylleberg *et al.* (1990), henceforth abbreviated by HEGY. The HEGY method amounts to a regression of $\Delta_4 y_t$ on deterministic terms like seasonal dummies and a trend and on $x_{1t} = (1 + L + L^2 + L^3)y_{t-1}$, $x_{2t} = (-1 + L - L^2 + L^3)y_{t-1}$, $x_{3t} = -(1 + L^2)y_{t-1}$, $x_{4t} = -(1 + L^2)y_{t-2}$, and on lags of $\Delta_4 y_t$, where $\Delta_i y_t = y_t - y_{t-i}$. The exact test regression reads

$$\Delta_4 y_t = \sum_{s=1}^{4} \beta_s D_{s,t} + \gamma t + \pi_1 x_{1t} + \pi_2 x_{2t} + \pi_3 x_{3t} + \pi_4 x_{4t} + \sum_{i=1}^{p} \phi_i \Delta_4 y_{t-i} + \varepsilon_t, \qquad (1)$$

where $D_{s,t} = 1$ if t corresponds to season s and 0 otherwise. The t-test for the significance of the parameter for x_{1t} (π_1) is denoted by t_1, the t-test for π_2 by t_2, and the joint significance test for π_3 and π_4 is denoted by F_{34}. If the test statistics are not significant, this corresponds to the presence of the associated root(s), which are 1, -1, and the pair i, $-$i, respectively. In Table B.1 we provide the asymptotical critical values of these test statistics.

In a properly seasonally adjusted times series seasonal unit roots -1, i and $-$i should not be present. Ideally, the unit root 1 should not be altered if it is present, as this root is associated with the stochastic trend in the series.

For monthly data a similar approach can be developed, see Franses (1991) and Beaulieu and Miron (1993). Only in this case more seasonal frequencies can be defined and the test regression (1) therefore contains more terms.

The presence of seasonal unit roots at some frequency and not at other frequencies can lead to problems of interpretation. One way to look at this is given by the following, see also Franses and Kunst (1999). The presence of a seasonal unit root at a certain frequency implies that there is no deterministic cycle at that frequency but a stochastic cycle. The usual seasonal dummies can be written as sine and cosine functions, see also Ghysels and Osborn (2001, Chapter 2). If one applies a seasonal difference filter to a time series, these deterministic functions are canceled out by corresponding seasonal unit roots in the filter. For example, the filter $(1+L)$, for a unit root -1, cancels the constant cycle implied by $\cos(\pi t)$. In other words, if $(1+L)$ is needed, there is no constant cycle of length 2 quarters within 4 quarters. And, hence, it may be that there is a non-constant cycle length, which may even increase over time.

The literature on the HEGY test is huge. Many modifications have been proposed, but for the present purpose and the type of data under scrutiny, the above version suffices. There are two issues that need to be mentioned. First, and as with all unit root tests, the critical values of the test statistics are obtained through simulation, and are often interpolated from different sample sizes. This means that in simulation studies, when one holds a 5% significance level, this level is not exactly obtained in size experiments. Also, the power of unit root tests is low, that is, it is not easy to distinguish between genuine unit roots and near-unit roots. The literature suggests that this might not be too large a problem, as erroneously imposing a unit root seems better than not imposing it when one should. Second, the HEGY test, as with many tests, assumes the adequacy of a certain model, in this case an AR model, preferably with lags beyond the number of seasons. For example, an AR(5) model for quarterly data would do well. When this assumption is violated, all kinds of things can happen. Size may go up, and power too, or the other way around. Usually, overlooking level shifts gives more evidence of unit roots (Perron and Vogelsang, 1992) while overlooking additive outliers gives the opposite outcome (Franses and Haldrup, 1994). A devastating effect on unit root tests is generated by overlooking moving average (MA) components, see also Schwert (1989). Hence, one may expect size distortions for the HEGY test to be largest for the airline model.

3.2 Canova-Hansen test

The test developed by Canova and Hansen (1995) takes as the null hypothesis that the seasonal pattern is deterministic. To explain the test, consider the process

$$y_t = \sum_{s=1}^{4} \delta_{st} D_{s,t} + \varepsilon_t \tag{2}$$

with

$$\begin{aligned} \delta_{1t} &= \mu_t + \alpha_{1t} - \alpha_{3t} \qquad & \delta_{2t} &= \mu_t - \alpha_{2t} + \alpha_{3t} \\ \delta_{3t} &= \mu_t - \alpha_{1t} - \alpha_{3t} \qquad & \delta_{4t} &= \mu_t + \alpha_{2t} + \alpha_{3t}, \end{aligned} \tag{3}$$

where the stochastic trend is defined as

$$\mu_t = \mu + \mu_{t-1} + \xi_t \tag{4}$$

with $\xi_t \sim N(0, \sigma_\xi^2)$ and the stochastic seasonal terms are given by

$$\alpha_{jt} = \beta_j + \alpha_{j,t-1} + \eta_{jt} \tag{5}$$

with $\eta_{jt} \sim N(0, \sigma_j^2)$, for $j = 1, \ldots, 3$. The process has a stochastic seasonal pattern if some $\sigma_j^2 > 0$. If $\sigma_j^2 = 0$ for all j, we have deterministic seasonality. The Canova-Hansen test corresponds to jointly testing for $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0$. The asymptotical critical values are given in Canova and Hansen (1995). For the quarterly case and a significance level of 5%, the critical value is 1.010.

The Canova-Hansen test also allows for testing for stationarity of the process itself, that is, testing for $\sigma_\xi^2 = 0$. However, this is not considered here as we only focus on the seasonal properties of the data.

The null hypothesis in the Canova-Hansen test is rejected in case seasonality of a series is not constant. After seasonal adjustment the Canova-Hansen test therefore should not reject. Note that having no seasonal pattern at all also implies constant seasonality.

3.3 Test for equal seasonal dummies

A basic test for the presence of seasonality in a time series is to regress the time series on four seasonal dummies. If there is no seasonality in the series, the four coefficients for these dummies should be equal. This property can easily be tested with an F-test.

This test can be performed for the level of the series as well as for the first difference, depending on the behavior of the series. The test regression equals

$$\Delta_1 y_t = \sum_{s=1}^{4} \beta_s D_{s,t} + \varepsilon_t, \tag{6}$$

where $D_{s,t} = 1$ if t corresponds to season s and 0 otherwise. If there is no seasonality, the F-test for $\beta_1 = \beta_2 = \beta_3 = \beta_4$ should not be rejected.

3.4 Test for correlation at the seasonal lag

Seasonal time series typically display autocorrelation at the seasonal lags. To test for significant autocorrelation at the seasonal lag we consider the following regression model

$$\Delta_1 y_t = \mu + \phi_1 \Delta_1 y_{t-1} + \phi_2 \Delta_1 y_{t-2} + \phi_3 \Delta_1 y_{t-3} + \phi_4 \Delta_1 y_{t-4} + \varepsilon_t \tag{7}$$

and we test for $\phi_4 = 0$ using a t-test. Insignificant values of the t-test denotes absence of correlation at the seasonal lag. One has to be a little cautious with this approach. Autocorrelation at the seasonal lag does not have to imply seasonality as the true lag order of the series may be 4 or higher.

3.5 Test for periodicity in AR parameters

Another property which may indicate the presence of seasonality in time series is different autoregressive parameters across the seasons, see Franses and Paap (2004). To investigate this periodicity we consider the PAR(p) model

$$\Delta_1 y_t = \mu + \sum_{i=1}^{p} \psi_{is} D_{s,t} \Delta_1 y_{t-i} + \varepsilon_t. \tag{8}$$

Absence of periodicity corresponds to the restriction $\psi_{i1} = \psi_{i2} = \psi_{i3} = \psi_{i4}$ for $i = 1, \ldots, p$. This can be tested with a standard F-test. If the F statistic is not significant, there is no statistical evidence for periodicity in the autoregressive parameters. The value of p can be determined using an information criterion such as the Bayesian Information Criterion [BIC].

3.6 Test for seasonality in the variance

The previous tests mainly consider the presence of seasonality in the mean of the series. To test for the presence of seasonality in the variance of the series we consider the residuals

ε_t of an AR(p) model for $\Delta_1 y_t$

$$\Delta_1 y_t = \mu + \sum_{i=1}^{p} \phi_i \Delta_1 y_{t-i} + \varepsilon_t. \tag{9}$$

The LM-test for seasonality in the residuals amount to testing for $\beta_1 = \beta_2 = \beta_3 = \beta_4$ in the auxiliary regression

$$\hat{\varepsilon}_t^2 = \sum_{s=1}^{4} \beta_s D_{s,t} + \sum_{i=1}^{p} \rho_i \Delta_1 y_{t-i} + \eta_t \tag{10}$$

using a standard F-test, where $\hat{\varepsilon}_t$ denotes the estimated residuals of (9), see Franses and Paap (2004). Insignificant values of the F-test imply the absence of seasonality in the variance. The value of p can be determined using BIC.

4 Data generating processes

For our first experiment we generate 5000 quarterly time series with 55 years of data for six different DGPs. All series are generated with standard normal innovations, that is, $\varepsilon_t \sim N(0,1)$. The first five years are discarded to reduce the dependence on the initial observations (which are set to zero). The analysis below is based on the remaining 50 years. The series are generated without outliers. The DGPs with outliers are discussed in Section 6.

Below we present the exact specification of the six DGPs we consider. The first five DGPs are chosen such that they mimic series which are frequently encountered in reality. The sixth DGP is a white noise series, this process is included to see what the adjustment procedures do to a nonseasonal series. For each DGP we briefly discuss the expected performance of the different seasonal adjustment procedures. We expect a particular procedure to perform well if the actual DGP is close to the DGP that underlies the seasonal adjustment method. In Appendix A we graphically present an example time series for each DGP.

4.1 DGP1: Constant expected yearly growth

The first data generating process assumes a fixed (expected) yearly growth for each quarter, that is,

$$\Delta_4 y_t = 0.25 + \varepsilon_t. \tag{11}$$

This model assumes three seasonal unit roots, that is, -1 and $\pm$i. We expect all methods to perform reasonably well, although the Dainties method may perform slightly less than the others. Its underlying trend plus seasonal dummies specification does not fit with this DGP, where the seasonal unit roots imply changing seasonality. On the other hand, as Dainties uses local regressions, such changes could be captures sufficiently.

4.2 DGP2: Deterministic seasonality

The second process we consider specifies different growth rates for each quarter, that is,

$$\Delta_1 y_t = 10D_{1,t} - 4D_{2,t} + 4D_{3,t} - 9.8D_{4,t} + \varepsilon_t. \tag{12}$$

This particular specification implies a small but positive expected growth over an entire year. On average the first and third quarter will show positive growth, whereas the other quarters will generally show a decline.

Also for this DGP, we expect all seasonal adjustment procedures to perform well, although now Dainties may be better as its local regression looks very much like (12). The other two seasonal adjustment methods will impose seasonal unit roots and these must then also appear as MA seasonal unit roots.

4.3 DGP3: Stochastic seasonality

For some economic series the seasonal pattern changes over time. The third DGP in our simulation experiment mimics this feature through stochastic seasonality, that is,

$$\begin{aligned} y_t &= \sum_{s=1}^{4} \alpha_{st} D_{s,t} + 0.1\varepsilon_t \\ \alpha_{st} &= \alpha_{s,t-1} + \eta_{st} \qquad \eta_{st} \sim N(0, \frac{1}{4}) \\ \alpha_{11} &= 10, \alpha_{21} = -4, \alpha_{31} = 4, \alpha_{41} = -9.8. \end{aligned} \tag{13}$$

This DGP comes close to the model in DGP1, although now seasonality does not change as quickly.

DGP3 does not assume seasonal unit roots in the series, but it does assume random walk like patterns in the parameters. When the variances of the error terms are large, it is quite likely that the data from this model can be approximated by a model with seasonal unit roots. When the variances are zero, this model collapses to DGP2. When the variances are very small, the data from this model can display slowly changing seasonal patterns, and this may well be captured by Dainties. Of course, when the variances are zero, that is DGP2, Dainties should work well, as the model matches with the local regression model.

4.4 DGP4: Airline model

The fourth data generating process in our simulation experiment is exactly the model underlying the TRAMO/SEATS method, that is, the Airline model. This model is specified as

$$\Delta_1 \Delta_4 y_t = (1 - 0.5L)(1 - 0.9L^4)0.1\varepsilon_t. \tag{14}$$

It is to be expected that TRAMO/SEATS will yield the best seasonally adjusted series for this DGP.

DGP4 assumes 3 three seasonal unit roots. Bell (1987) shows that when the MA(4) parameter gets closer to -1, the model generates data that are close to those of DGP2.

In principle, the airline model can describe data that show varying patterns of changing seasonality over time.

4.5 DGP5: Generalized deterministic seasonality

The fifth DGP generalizes DGP2, that is, the growth rate is related to seasonal dummies as well as lagged growth rates. The exact specification reads

$$\Delta_1 y_t = 10D_{1,t} - 4D_{2,t} + 4D_{3,t} - 9.8D_{4,t} + 0.6\Delta_1 y_{t-1} + 0.3\Delta_1 y_{t-4} + \varepsilon_t. \quad (15)$$

So, again, as for the second DGP we conjecture that Dainties would perform well for this DGP.

4.6 DGP6: White noise

The final DGP we use does not contain a seasonal pattern, that is, this series is white noise. For completeness we give the specification

$$y_t = \varepsilon_t. \quad (16)$$

We should hope that all methods work equally well. If not this would mean that the method *introduces* seasonality in an otherwise nonseasonal series, which does not seem useful.

5 Simulation results

In this section we discuss the results of our first simulation experiment in words. In this case we do not yet consider aggregation of series nor the cases with outliers. The results of this baseline simulation are presented in the first panel of the tables in Appendix C[1]. To assist in the interpretation of the results, we give in Table B.2 an overview of how to interpret the numbers in the tables.

5.1 DGP1 – Constant expected yearly growth

First of all, we consider the first DGP of Section 4. The first panel of Table C.1 presents the percentage of rejections of the null of the various diagnostic tests of Section 3 for 5000 generated series according to DGP1.

The HEGY test correctly identifies the nonseasonal unit root in the data. For this DGP, the three seasonal adjustment methods do not lead to large differences in detecting the nonseasonal root in the series. However, in general some evidence of this root is removed by all procedures. The Dainties method alters the unit root the least as compared to X12-ARIMA and TRAMO/SEATS.

The third and the fourth column of Table C.1 show the results of the test for a unit root at one of the seasonal frequencies. The test results for the unadjusted data shows that this DGP has a stochastic seasonal component, as we know. All methods effectively remove this stochastic seasonality. Compared to the other methods, the Dainties procedure performs worst on this test as some stochastic seasonality remains after correction.

The next column shows the results of the Canova-Hansen test. For all adjustment procedures we should ideally find very few rejections of H_0 after seasonal adjustment. This is the case for the X12 and TRAMO/SEATS methods. The Dainties method however gives adjusted series in which there still is stochastic seasonality.

For the test for equal seasonal dummies we should also find as few rejections as possible after adjustment. Again, this is the case for X12-ARIMA and TRAMO/SEATS while for Dainties we find a substantive number of rejections. This indicates that for this DGP Dainties does not succeed in removing the seasonal pattern in the first differences.

[1] All simulations were done in Ox 3.4 (Doornik, 1999). The actual seasonal adjustment was done through calls to the original procedures of X12-ARIMA, TRAMO/SEATS and Dainties, where the first two are shipped with EViews and the latter was programmed by Hendyplan SA for the European Commission.

The remaining three properties, seasonal correlation, periodicity in the AR component, and seasonal variance, should be absent after correction. X12-ARIMA and TRAMO/SEATS turn out to produce series in which there is seasonal correlation, even when this was not present in the original data. Furthermore, X12-ARIMA and to a lesser extent TRAMO/SEATS give series with in which absence of periodicity in the AR component is rejected in about 15% and 10% of the cases, respectively.

Overall, none of the three methods shows a perfect performance on all diagnostic tests. The series obtained with X12-ARIMA and TRAMO/SEATS show seasonal correlation and periodicity in the AR component, while Dainties fails to remove all evidence for seasonal dummies in the first difference.

5.2 DGP2 – Deterministic seasonality

For this DGP and the ones presented below, we will not present the results in as much detail as for DGP1. Instead we highlight the most important results. The reader is again referred to Table B.2 for an overview of the favorable results of the tests.

The results of the HEGY tests do not show problems with any of the adjustment methods. X12-ARIMA and Dainties remove some of the non-seasonal unit root in the data. The Canova-Hansen test and the test for seasonal dummies show a perfect performance for all methods. The only test that gives different results is the test for seasonal correlation. TRAMO/SEATS performs best on this test, followed by Dainties. X12-ARIMA produces adjusted series in which in 75% of the cases our test indicates the presence of seasonal correlation.

For this DGP all methods perform relatively well. TRAMO/SEATS and Dainties perform equally well, perhaps with a slight advantage for TRAMO/SEATS. X12-ARIMA only fails to remove seasonal correlation.

5.3 DGP3 – Stochastic seasonality

For the DGP with stochastic seasonality we again see that X12-ARIMA and TRAMO/SEATS remove evidence for a non-seasonal unit root. At the seasonal frequencies the HEGY test does not show large differences. Dainties performs slightly worse than its two competitors.

The results of the Canova-Hansen test and the test on seasonal dummies also show the poor performance of Dainties for this DGP. Surprisingly, X12-ARIMA and TRAMO/

SEATS perform very bad on the test for seasonal correlation. For the series produced by TRAMO/SEATS we even find seasonal correlation in 97% of the cases. The test on periodicity in the AR parameters also shows some problems for these two methods.

5.4 DGP4 – Airline model

For the Airline model the size of the HEGY test is not correct. As this DGP clearly contains a unit root, we should reject the null of this test in 5% of the cases. The fact that we find a much larger rejection frequency is not surprising due to the $\Delta_1\Delta_4$ filter in the DGP.

Seasonal adjustment should ideally not change the rejection frequency for this test. This is not the case for any of the methods, for Dainties the difference is the smallest while for TRAMO/SEATS we find a very large change. However, we see the reverse for the seasonal HEGY test. Here, Dainties performs worst while X12-ARIMA and TRAMO/SEATS show a good performance. For this test we also see a size distortion. The test for seasonal dummies shows the same pattern.

Although this DGP does have stochastic seasonality, we reject the null of the Canova-Hansen test for the unadjusted data in only 37% of the cases. This can be explained by the fact that the Moving Average component in the Airline process is difficult to capture adequately in the test procedure. All three methods perform well on this diagnostic.

Finally, all methods perform poorly on the test for seasonal correlation and periodicity in the AR parameters. TRAMO/SEATS also fails on these tests, but it shows the best performance among the three methods.

Overall, all methods have difficulties with this process. For TRAMO/SEATS this is especially surprising as the Airline model forms the basis for this adjustment procedure. Dainties performs worst for this DGP.

5.5 DGP5 – Generalized deterministic seasonality

For the DGP with generalized deterministic seasonality, the methods perform quite well. The only problem is in the seasonal correlation. All methods yield series that have some seasonal correlation. This is especially a problem for Dainties. This method also fails to some extent to remove all evidence for periodicity in the AR parameters.

5.6 DGP6 – White noise

The final DGP we consider does not contain seasonality. Ideally, seasonal adjustment should not alter the series in this case. The rejection frequencies before and after adjustment should therefore be the same. Note however that this series does not contain a unit root, while some of the tests are based on the first difference of the series. For a white noise series, the tests are therefore based on an overdifferenced series. The size of the tests will not be correct. This is best seen in the rather high rejection frequency for the test of seasonal correlation. Before adjustment we reject the null in 83% of the cases. These test results should therefore be taken with care.

Overall we see that all methods change at least some of the properties of the series. Large differences are seen for the test of seasonal dummies and the test for seasonal correlation. It is however difficult to give a ranking of the methods.

6 Outliers

In this section we consider the effect of outliers on the performance of the seasonal adjustment routines. As an outlier is by definition not part of the seasonal component of a time series, outliers in the unadjusted data should be there, even after adjustment. The X12-ARIMA and the TRAMO/SEATS procedures are both designed to deal with outliers. The Dainties method however does not have this feature.

Our simulation study now proceeds as follows (i) generate data without outliers, (ii) generate data with outliers (using exactly the same draws for the disturbance terms), (iii) seasonally adjust the series with outliers (with outlier detection enabled), (iv) remove the effect of the outlier from the unadjusted data and the adjusted data, (v) perform all the diagnostic tests on the latter series. To remove the effect of the outlier we simply subtract the difference between the generated series with and without the outlier from the adjusted series. With simulated data one knows the position and magnitude of the outliers exactly. Note that this final step of removing the outlier from the series is necessary to obtain valid results from the diagnostic tests.

We consider two types of outliers, that is, additive outliers and innovation outliers. For each series we add three outliers, at $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$ of the length of the time series. As the length of the time series equals 50 years, not all outliers occur in the same quarter. The size of the outlier is four times the standard deviation of the innovations in all cases. The simulation of series with an additive outlier is straightforward. To simulate data with innovation outliers, one first generates the complete series of disturbance terms. The outlier is then added to this disturbance series before one generates the actual data from the disturbances.

6.1 Results for additive outliers

The second panel of the tables in Appendix C gives the rejection frequencies for all diagnostic tests based on the DGPs with three innovation outliers. Below we will only discuss those cases in which we see a marked difference with respect to the case without outliers (the top panel of each table).

For the additive outlier case, the results for the HEGY test are much like those for the baseline case. The only difference is in the results for Dainties in case of the Airline model (DGP4). For this combination we see that Dainties increases the evidence for a nonseasonal unit root. More importantly, quite some evidence remains for seasonal unit roots. We reject the null of the presence of the seasonal root -1 in only 25% of the cases.

For the Canova-Hansen test we do not find any differences. The poor performance of Dainties for DGP4 based on the test for seasonal dummies in the baseline case is enforced in case of additive outliers, as now we reject the absence of seasonal dummies for more than 90% of the generated series.

If we look at the correlation at the seasonal lag, we see that the results for X12-ARIMA hardly change. Surprisingly, the results for TRAMO/SEATS and Dainties become slightly better. We now find less evidence of correlation at the seasonal lag.

For the series without outliers we found that not all methods correctly remove periodicity in autoregressive parameters. We find the same for the case with additive outliers. For Dainties and TRAMO/SEATS the evidence is even stronger. Dainties performs poorly on DGP4 and 5, while TRAMO/SEATS does not perform well on DGP1 and 3.

Finally, the test on seasonality in the variance does not show large effects of outliers. The only difference we find is for TRAMO/SEATS with DGP1 and Dainties with DGP4. For these two cases we now hardly find any evidence for seasonality.

6.2 Results for innovation outliers

The same comparison as above can be made for the case of innovation outliers. These results are presented in the bottom panel of the tables in Appendix C. Clearly, it turns out that for some DGPs all methods are rather sensitive to this type of outlier. This is especially the case for the DGPs in which annual differences are taken, that is, DGP1 and 4. For these two processes all methods change some of the (nonseasonal) unit root properties. Except for DGP4 with TRAMO/SEATS we find more evidence for a unit root after adjustment.

The results for the seasonal unit roots are even more worrisome. For all generated series, all three methods leave seasonal unit roots in the series. Note that this could have two causes. First of all, the methods could truly leave seasonality in the series. However, a second, more plausible, reason is that the methods do not adequately detect the innovation outliers. If we subtract the effect of the innovation outlier, which contains a seasonal pattern, from the adjusted times series we may introduce seasonality again. This reasoning also explains why we only find this for the DGPs that contain $\Delta_4 y_t$. The same arguments hold for the results of the Canova-Hansen test and the test for seasonal dummies. A more positive conclusion from these results is that for the other DGPs we do no find any effect of the innovation outliers.

The results for the test on correlation at the seasonal lag are more mixed. In this case we do not find extreme differences with the baseline case. As for the case with additive outliers, the results are in fact better than those for the baseline.

The results shows that all methods still do not fully remove periodicity in AR parameters. Overall, the problems worsen under innovation outliers compared to no outliers. Especially the TRAMO/SEATS methods generates series with periodic AR components.

Finally, the results on the test on seasonality in the variance is not affected by the presence of innovation outliers.

7 Aggregated series

In practice, one often considers aggregated series, where aggregation means cross-sectional aggregation here. In case the disaggregates are also available there are two options to obtain a seasonally adjusted aggregate series. One could (i) first adjust the individual series and aggregate afterwards, or one could do the reverse, that is, (ii) first aggregate and then adjust the resulting series. Although both methods will generate different series, there are no theoretical grounds to prefer one method over the other. In this section we will run another simulation experiment to see whether the seasonal properties of the resulting series differ. Again we will generate 5000 series using one of our DGPs. From these series we construct 1000 aggregates by summing five series each time. To make our simulation more realistic we allow for correlations between each set of series that constitutes an aggregate.

The tables in Appendix D show the test results. Overall it turns out that there are very few differences between the two aggregation procedures. Moreover, the test results also hardly differ from the baseline case. Below, we will go into some of the noticeable differences.

For TRAMO/SEATS, the HEGY test on the nonseasonal root shows that first adjusting the components of the aggregate removes more evidence of a unit root than the reverse procedure. The results of the Canova-Hansen test for the two procedures do not differ. However, there is one difference with the results for the baseline case. For the baseline case, Dainties did not perform very well for DGP1. When we consider aggregates, this problem disappears.

For the test on seasonal dummies we only find interesting results for the Dainties procedure. For DGP1, first aggregating leads to marginally better adjusted series. The test on correlation at the seasonal lag, also does not show major differences.

For the results of the test on periodicity in the AR parameters we find the most differences. The most striking difference is in the performance of Dainties on DGP5. For this combination it is best to aggregate before seasonal adjustment. However, for some of the other combinations it is best to first do the seasonal adjustment. Based on the test on seasonality in the variance, for the combination Dainties with DGP2, it is best to first do the adjustment. The other rejection frequencies are very much alike.

Overall, the order of adjustment and aggregation hardly has an influence on the performance of the adjustment procedures. It seems difficult to give an advice on the appropriate order.

8 Business and consumer survey data

Finally, we consider series from the "Joint Harmonized EU Programme of Business and Consumer Surveys" [BCS]. In this section we focus on the series that are available on a monthly basis. In principle the series in this data set are available from January 1980 to September 2004. However, for many series fewer observations are available. A few series have missing values. For these we perform a simple imputation to complete the series, that is, we replace a missing value at time t by the average of the values at time $t-1$ and $t+1$. Another difficulty is that, especially at the start of the time series, observations are only available at a quarterly basis. For some series this leads to problems with our diagnostic tests, and these series are therefore excluded from further analysis. After removing these series, 300 series remain. These 300 series have been used before in previous exercises and here they form the basis on which we will compare the seasonal adjustment methods. In Section A.7 we show time series plots for a typical example. Clearly, seasonality is not strong, and also imposing a unit root does not seem harmful.

Table E.1 shows the rejection frequencies of our diagnostic tests for the 300 BCS series. We first focus on the first line, which shows the results of the tests on the unadjusted data. The test results indicate that almost all series contain a unit root. The series usually do not contain seasonal unit roots or a nonstationary seasonal process of the Canova-Hansen type. About two-thirds of the series contain seasonally varying growth rates. For about one third of the series we find seasonality in variance, significant AR parameters at the seasonal lag and significant periodicity in AR parameters. This final finding can partly be explained by the fact that for quite a large number of series a large part of the time series actually only contain quarterly observations. For example, Dezhbakhsh and Levy (1994) show that interpolation of quarterly data to monthly data may induce periodicity in the autocorrelations. In sum, based on these results in the first line, we may conclude that DGP2 (and to some extent DGP5) comes close the series in the BCS, where perhaps there is additional seasonality in variance[2].

The remaining lines of Table E.1 give the test results for the seasonally adjusted

[2]This result can be formalized by comparing the test results numerically. In case one calculates the sum of the squared difference between the test results for the BCS and the results for each of the DGPs, one finds that the BCS data indeed come closest to DGP2. The second best fit is given by DGP5 (generalized deterministic seasonality). The test results for the other four DGPs are very different from those for the BCS data. This can also be seen graphically by comparing Figure A.7 to the other figures in Appendix A.

series. For X12-ARIMA and TRAMO/SEATS we enable all outlier correction features. Overall there are no large differences between the adjustment methods. Although the test results on the unadjusted data show that the Airline model does not fit the BCS data, we surprisingly find that incorrectly imposing this model (as is done in TRAMO/SEATS) does not lead to large differences in performance.

According to the first four tests, all methods perform very well. The last three tests on seasonal correlation, periodic autocorrelation and seasonal variance show that all methods fail to remove all seasonal properties of the series. As discussed before, these findings may be due to the fact that the first observations of a large number of series are actually interpolated quarterly observations. For example, consider the seasonal variance for interpolated monthly data. The monthly growth rates will exactly equal to zero for all months within a quarter. The variance of these monthly growth rates will therefore also be zero. Naturally, one would find seasonal variances in growth rates in such a data set. It is not surprising that none of the adjustment methods succeeds in removing these seasonal features.

9 Conclusion

In this report we have compared three seasonal adjustment methods using simulated data and 300 Business and Consumer Survey [BCS] data. We first outlined the specifics of these methods, and we discussed the link between these methods and potentially useful models to describe the data. Next, we carried out a simulation experiment to highlight the differing performances of the three methods. We found that the Census X12-ARIMA and TRAMO/SEATS methods were most robust to variations in the data generating process. This implies that in case one would not have any strong indications as to which model could best describe the raw (unadjusted) data, then these two methods are to be preferred. On the other hand, the Dainties method performed relatively well for two types of DGPs, that is, series with deterministic seasonality. Hence, if one suspects that the data are best described by these models, then the Dainties method is preferable.

It turns out that additive outliers do not have a serious effect on the performance of the seasonal adjustment procedures. If a procedures generates proper corrected series in the case without outliers, it also performs well when additive outliers are present. For Dainties this is somewhat surprising as the method does not explicitly deal with outliers. However, the adjustment methods are very sensitive to innovation outliers. None of the seasonal adjustment methods is capable of remove this type of outlier adequately before seasonal correction. This is especially true if the series contains seasonal unit roots. For cross-sectional aggregated series we have found that the order of seasonal adjustment and aggregation does not influence the seasonal properties of the resulting series.

Summarizing, for the series in the BCS the choice of the seasonal adjustment method is not very important. We find that the BCS data closely resemble those generated by DGP2, that is, deterministic seasonality. Given the way the data are collected this is not surprising. A priori we expected seasonal variation in BCS data not to be strong, and unlikely to change much over time. Our simulation results showed that for this DGP, Dainties and TRAMO/SEATS all perform well. We find the same for the actual data. The Dainties method does a good job. A disadvantage of this method is however that it does not take outliers into account. Our final recommendation is therefore to extend the Dainties method with a method for outlier correction. The details of such a procedure are left for further research.

A Data generating processes

In this appendix we present one example time series for each DGP. The graphs give the level (y_t), the first difference ($\Delta_1 y_t$), and the seasonal difference ($\Delta_4 y_t$) of the series. Furthermore, we give so-called Vector-of-Quarters [VQ] plots (Franses, 1994), also known as seasonal splits (as in the EViews package), of the level (y_{st}) and the first difference ($\Delta_1 y_{st}$). By splitting the time series into separate series for each quarter, VQ-plots give an insightful overview of the development of the seasonal patterns in a time series.

A.1 Constant expected yearly growth

DGP: $\Delta_4 y_t = 0.25 + \varepsilon_t$

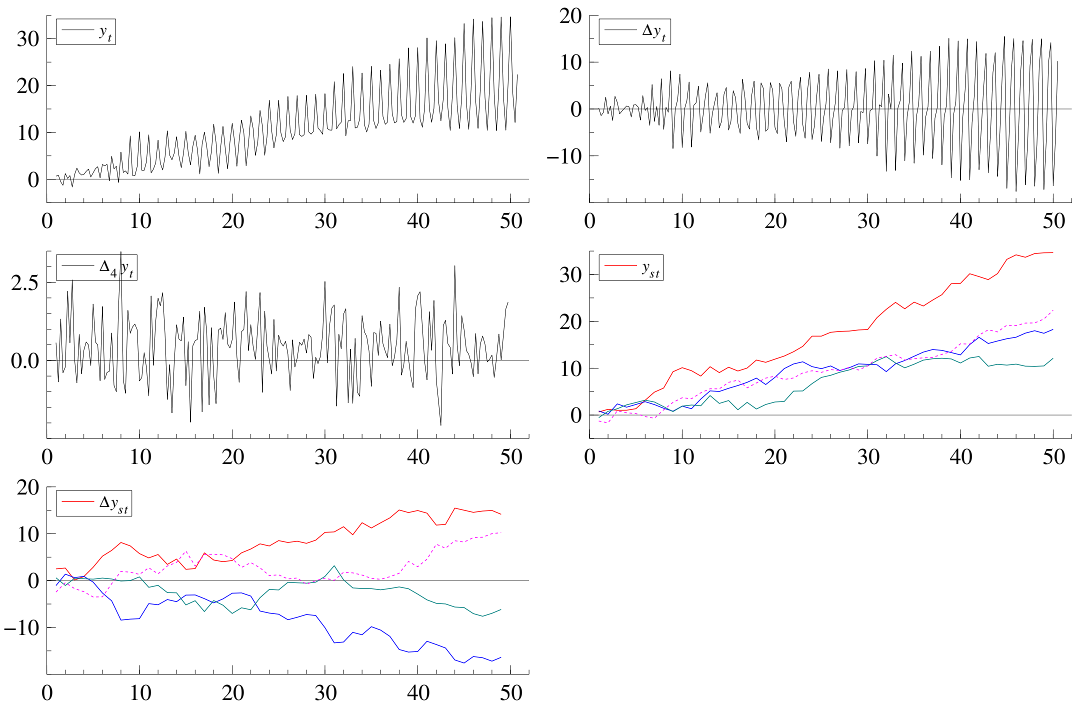

Figure A.1: Example of time series with DGP1: Constant expected yearly growth

This DGP displays an upward trend and changing seasonality, as can be seen from the graphs for $\Delta y_{s,t}$.

A.2 Deterministic seasonality

$\Delta_1 y_t = 10D_{1,t} - 4D_{2,t} + 4D_{3,t} - 9.8D_{4,t} + \varepsilon_t$

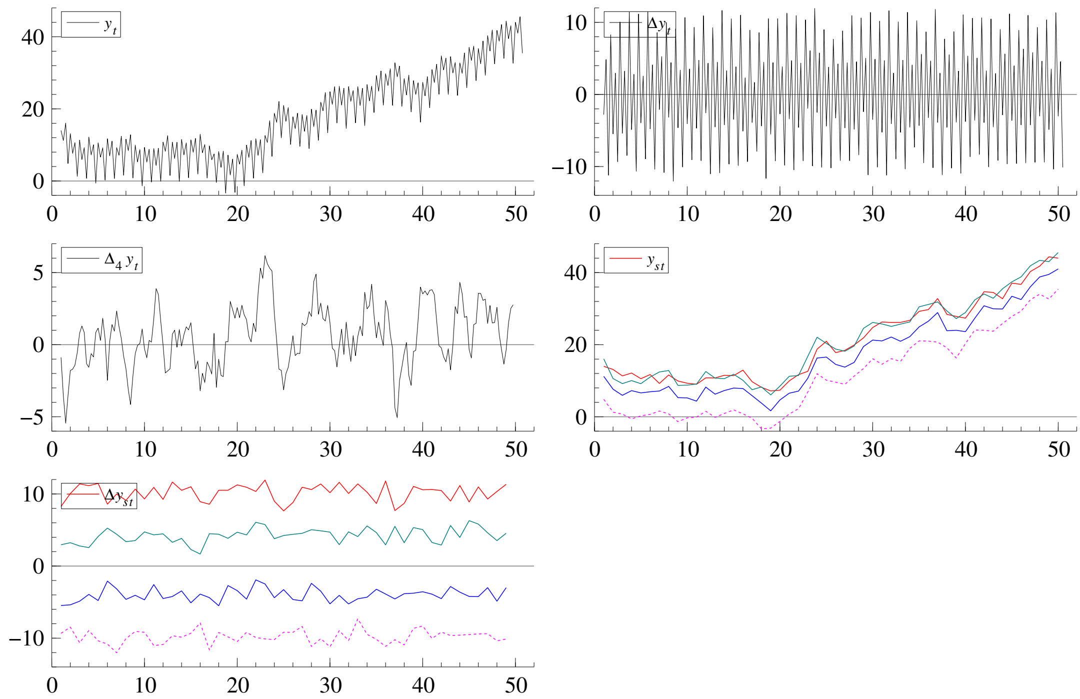

Figure A.2: Example of time series with DGP2: Deterministic seasonality

This DGP assumes that seasonal variation in the growth rate ($\Delta_1 y_t$) is constant over time, as is particularly evident from the four graphs for Δy_{st}.

A.3 Stochastic seasonality

$y_t = \sum_{s=1}^{4} \alpha_{st} D_{st} + 0.1\varepsilon_t$

$\alpha_{st} = \alpha_{s,t-1} + \eta_{st}$

$\alpha_{11} = 10, \alpha_{21} = -4, \alpha_{31} = 4, \alpha_{41} = -9.8$

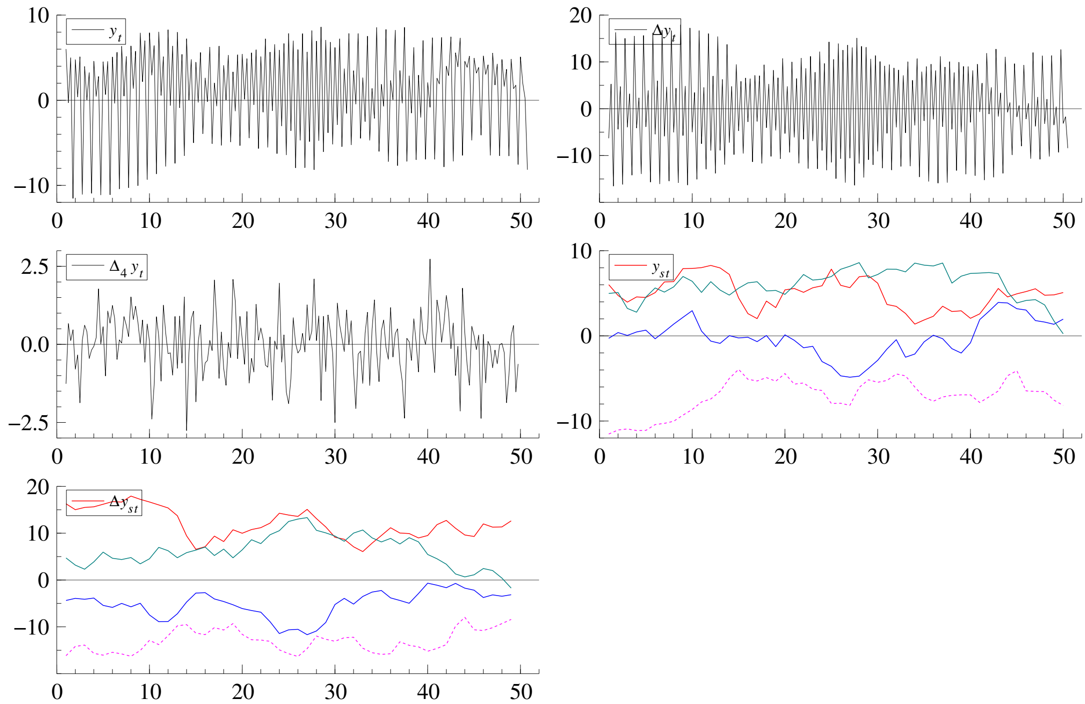

Figure A.3: Example of time series with DGP3: Stochastic seasonality

This DGP allows for changing seasonality, like DGP1, but here the changes are not quick. A model with seasonal unit roots, as in DGP1, could fit these data to some extent, but such a model would require a MA component with (near) seasonal unit roots. A model as in DGP4 comes closer to DGP3.

A.4 Airline model

$\Delta_1 \Delta_4 y_t = (1 - 0.5L)(1 - 0.9L^4)0.1\varepsilon_t$

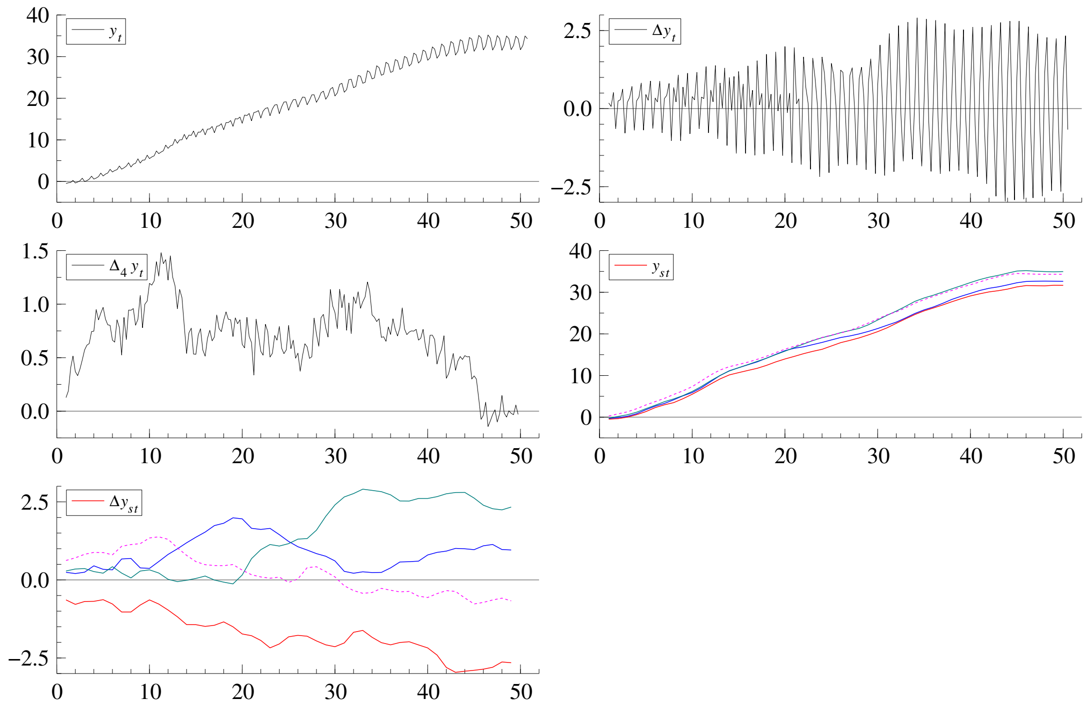

Figure A.4: Example of time series with DGP4: Airline model

This DGP allows for rapid changes in seasonality, as well as for increasing seasonal variation. The Airline model is very flexible. As the MA parameters come closer to -1, models like DGP1 and DGP2 appear.

A.5 Generalized deterministic seasonality

$$\Delta_1 y_t = 10D_{1,t} - 4D_{2,t} + 4D_{3,t} - 9.8D_{4,t} + 0.6\Delta_1 y_{t-1} + 0.3\Delta_1 y_{t-4} + \varepsilon_t$$

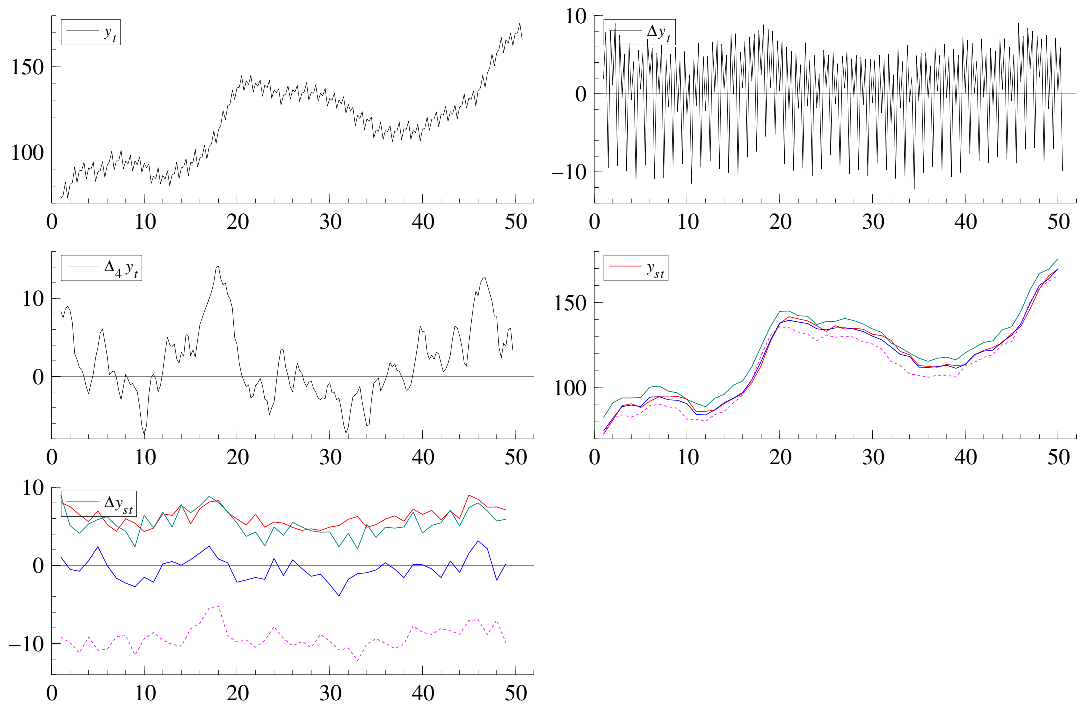

Figure A.5: Example of time series with DGP5: Generalized deterministic seasonality

As DGP2 is quite likely to be a reasonable model for the BCS data, we also include an extended version of it, which is this DGP. Seasonality is constant over time, but due to the inclusion of $\Delta_1 y_{t-4}$ small changes can occur.

A.6 White noise

$y_t = \varepsilon_t$

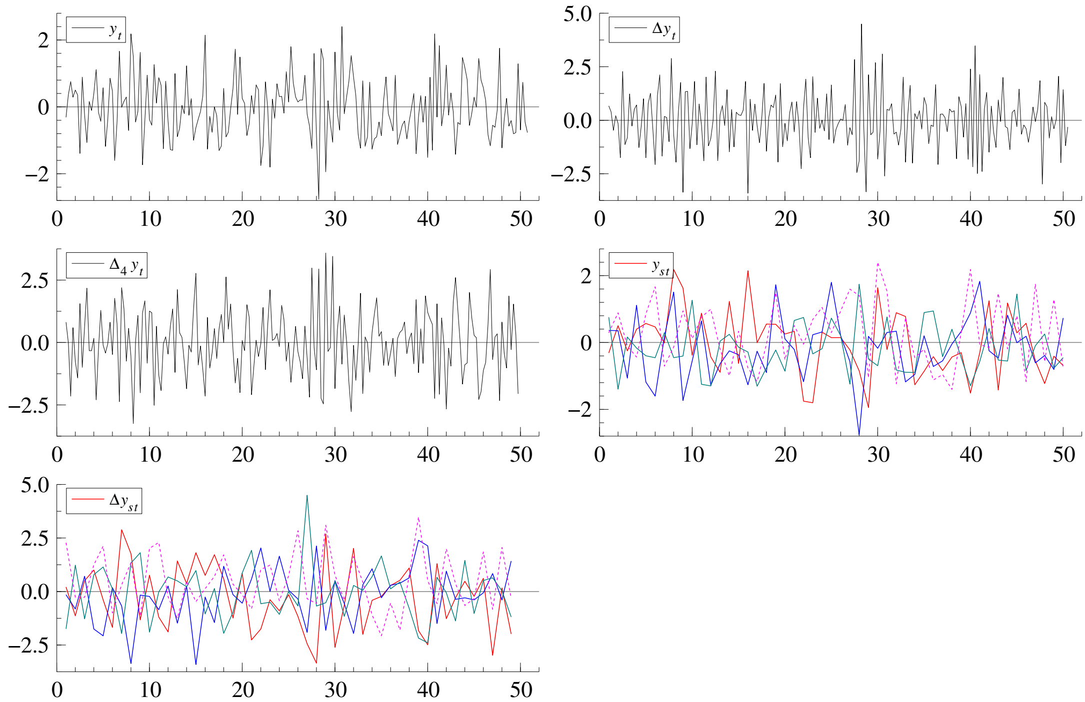

Figure A.6: Example of time series with DGP6: White noise

This DGP does not include any form of seasonality, and it is hoped that seasonal adjustment methods do not introduce any seasonality.

A.7 Example series from the BCS

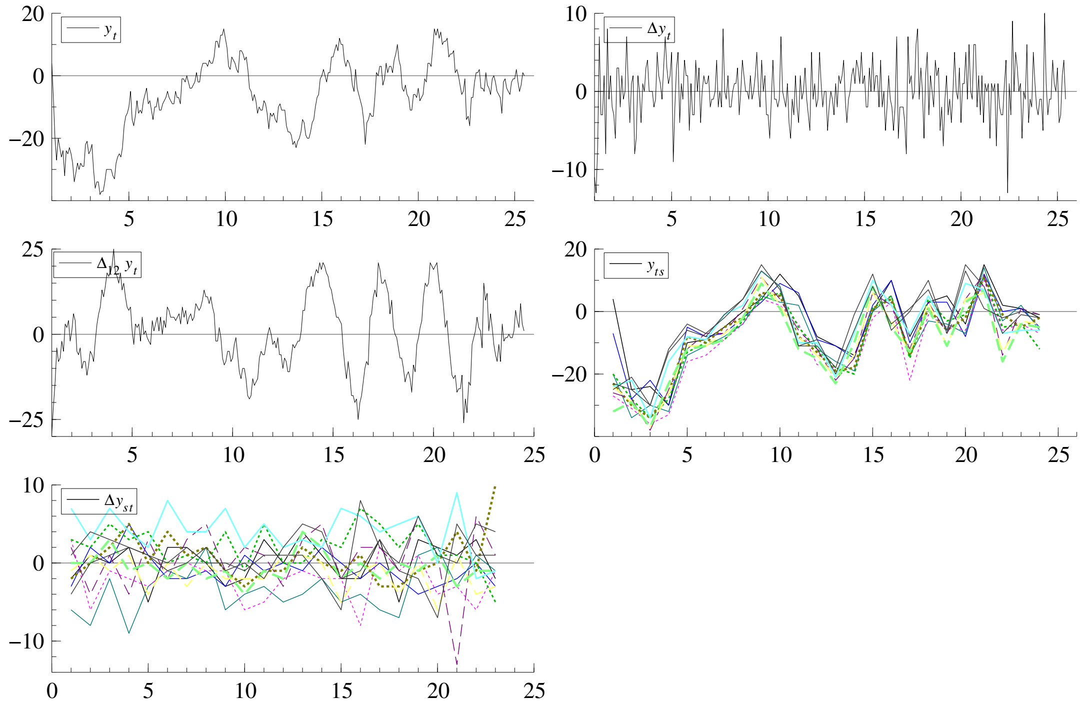

Figure A.7: Example of a time series from the Business and Consumer Surveys

B Critical values and favorable outcomes

Table B.1: 5 % critical values of the HEGY tests (1) for $T = 200$[1]

Auxiliary regressions	t_1	t_2	F_{34}
intercept	-2.87	-1.92	3.12
seasonal dummies	-2.91	-2.89	6.61
seasonal dummies + trend	-3.49	-2.91	6.57

[1] Source Hylleberg *et al.* (1990).

Table B.2: Null hypotheses of tests and favorable outcomes after seasonal adjustment

Test	H_0	Favorable rejection frequency
HEGY (zero frequency, 1)	presence of unit root	unchanged
HEGY (seasonal frequency, -1, $\pm$i)	presence of seasonal unit root	high
Canova-Hansen	stationary seasonal process	low
Seasonal dummies	equal seasonal dummies	low
Seasonal correlation	absence of correlation	low
Periodicity in AR parameters	absence of periodicity	low
Seasonality in variance	absence of seasonality	low

C Basic Simulations and Outliers

Table C.1: Rejection frequencies of diagnostic tests for DGP1 – Constant yearly growth

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				No outliers				
Unadjusted data	4.16	4.56	5.78	99.46	99.70	5.00	16.98	5.96
X12-ARIMA	9.76	100.00	100.00	0.00	0.24	53.34	14.90	4.76
Dainties	6.72	91.82	99.94	10.56	57.14	7.28	5.60	5.48
TRAMO/SEATS	13.94	100.00	100.00	0.00	0.00	96.94	9.60	6.90
				Additive outliers				
Unadjusted data	3.96	4.96	5.32	99.12	99.88	5.22	16.54	5.50
X12-ARIMA	9.20	100.00	100.00	0.00	0.98	50.86	13.34	4.52
Dainties	6.38	90.42	99.92	9.10	58.24	7.18	6.04	4.88
TRAMO/SEATS	13.30	100.00	100.00	0.00	0.96	73.80	46.54	2.12
				Innovative outliers				
Unadjusted data	4.64	4.54	5.52	99.40	99.74	5.36	16.90	5.60
X12-ARIMA	2.92	0.00	96.62	100.00	100.00	22.16	8.04	7.30
Dainties	5.38	0.00	71.80	100.00	100.00	8.56	9.72	6.24
TRAMO/SEATS	1.14	0.00	35.94	100.00	100.00	99.74	94.56	1.48

Table C.2: Rejection frequencies of diagnostic tests for DGP2 – Deterministic seasonality

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				No outliers				
Unadjusted data	4.66	100.00	100.00	0.12	100.00	100.00[1]	47.80[1]	9.04
X12-ARIMA	7.98	100.00	100.00	0.00	0.00	74.62	7.86	4.88
Dainties	7.02	100.00	100.00	0.00	0.00	17.54	6.62	5.48
TRAMO/SEATS	4.92	100.00	100.00	0.00	0.00	8.10	5.38	5.26
				Additive outliers				
Unadjusted data	4.98	100.00	100.00	0.12	100.00	100.00[1]	47.68[1]	9.50
X12-ARIMA	7.72	100.00	100.00	0.00	0.02	68.70	7.20	4.70
Dainties	5.60	100.00	100.00	0.00	4.88	4.02	6.02	4.08
TRAMO/SEATS	5.26	100.00	100.00	0.00	0.00	7.56	5.14	4.96
				Innovative outliers				
Unadjusted data	4.72	100.00	100.00	0.14	100.00	100.00[1]	48.14[1]	9.66
X12-ARIMA	7.04	100.00	100.00	0.00	0.00	67.08	7.38	3.96
Dainties	5.88	100.00	100.00	0.00	0.26	6.84	5.14	4.32
TRAMO/SEATS	5.02	100.00	100.00	0.00	0.00	7.70	4.90	4.86

[1] Rejection of H_0 is due to seasonal dummies in the DGP.

Table C.3: Rejection frequencies of diagnostic tests for DGP3 – Stochastic seasonality

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				No outliers				
Unadjusted data	4.98	4.98	6.10	100.00	100.00	5.38	13.78	5.86
X12-ARIMA	11.16	100.00	100.00	0.00	0.04	57.48	15.00	5.04
Dainties	7.40	92.24	99.92	35.54	55.06	8.22	5.06	5.64
TRAMO/SEATS	16.08	100.00	100.00	0.20	0.00	97.38	10.74	6.92
				Additive outliers				
Unadjusted data	4.94	4.96	5.42	99.98	100.00	4.82	14.42	5.58
X12-ARIMA	10.58	100.00	100.00	0.00	0.88	53.88	14.00	5.00
Dainties	7.46	90.84	99.92	34.58	58.78	7.74	5.50	5.00
TRAMO/SEATS	14.84	100.00	100.00	0.02	0.98	74.04	48.14	2.04
				Innovative outliers				
Unadjusted data	4.36	4.60	5.76	99.94	100.00	4.86	14.00	5.70
X12-ARIMA	10.48	100.00	100.00	0.02	1.00	52.10	14.30	4.52
Dainties	7.16	91.30	99.70	34.18	58.64	7.50	5.32	4.86
TRAMO/SEATS	14.26	100.00	100.00	0.06	1.04	73.60	48.46	2.02

Table C.4: Rejection frequencies of diagnostic tests for DGP4 – Airline model

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				No outliers				
Unadjusted data	16.90	11.84	12.00	37.10	99.84	100.00	22.36	10.60
X12-ARIMA	27.06	99.68	99.96	0.62	0.04	100.00	19.04	8.98
Dainties	21.98	82.44	95.40	1.10	48.32	100.00	15.70	9.86
TRAMO/SEATS	6.06	100.00	100.00	1.08	0.00	27.62	14.92	9.94
				Additive outliers				
Unadjusted data	18.24	11.36	9.90	36.06	99.92	100.00	22.50	10.50
X12-ARIMA	28.60	99.74	99.10	0.54	0.32	100.00	18.86	9.18
Dainties	4.54	24.64	65.68	0.28	91.32	99.98	86.60	0.12
TRAMO/SEATS	7.06	99.94	100.00	0.96	0.00	24.34	14.54	8.94
				Innovative outliers				
Unadjusted data	18.82	11.22	10.90	36.38	99.88	100.00	23.28	10.20
X12-ARIMA	12.80	0.00	35.36	46.64	100.00	100.00	22.16	12.02
Dainties	15.52	0.32	71.18	45.58	100.00	100.00	19.48	11.40
TRAMO/SEATS	20.38	0.00	0.00	46.84	100.00	100.00	85.38	12.74

Table C.5: Rejection frequencies of diagnostic tests for DGP5 – Generalized deterministic seasonality

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				No outliers				
Unadjusted data	2.42	100.00	100.00	3.62	100.00	100.00	62.90	11.32
X12-ARIMA	3.68	100.00	100.00	3.60	0.00	52.40	1.24	4.22
Dainties	2.48	100.00	100.00	3.50	0.00	91.58	10.70	6.04
TRAMO/SEATS	3.54	100.00	100.00	3.78	0.00	22.36	0.78	6.08
				Additive outliers				
Unadjusted data	2.06	100.00	100.00	3.82	100.00	100.00	61.94	10.76
X12-ARIMA	3.08	100.00	100.00	3.80	0.00	55.60	1.60	4.32
Dainties	1.92	100.00	100.00	3.64	0.02	99.66	28.18	5.64
TRAMO/SEATS	3.22	100.00	100.00	3.84	0.00	27.56	1.96	4.82
				Innovative outliers				
Unadjusted data	2.00	100.00	100.00	4.08	100.00	100.00	62.86	10.70
X12-ARIMA	3.20	100.00	100.00	4.36	0.00	58.86	1.88	4.12
Dainties	1.78	100.00	100.00	4.26	0.00	97.44	16.74	4.68
TRAMO/SEATS	2.96	100.00	100.00	4.46	0.00	32.48	2.14	5.32

Table C.6: Rejection frequencies of diagnostic tests for DGP6 – White noise

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				No outliers				
Unadjusted data	100.00	100.00	100.00	4.00	12.60	83.00	7.24	5.10
X12-ARIMA	99.76	100.00	100.00	0.00	0.02	96.60	11.36	5.36
Dainties	99.98	100.00	100.00	0.00	0.00	91.88	9.18	5.84
TRAMO/SEATS	99.98	100.00	100.00	0.02	0.00	87.82	7.76	4.98
				Additive outliers				
Unadjusted data	100.00	100.00	100.00	3.66	12.38	82.48	6.94	5.18
X12-ARIMA	99.96	100.00	100.00	0.00	0.26	95.92	11.20	5.60
Dainties	100.00	100.00	100.00	0.00	0.02	87.94	9.18	5.88
TRAMO/SEATS	100.00	100.00	100.00	0.04	0.00	86.18	7.24	5.30
				Innovative outliers				
Unadjusted data	100.00	100.00	100.00	4.96	12.50	82.96	6.90	5.24
X12-ARIMA	99.94	100.00	100.00	0.00	0.30	96.08	11.10	5.08
Dainties	100.00	100.00	100.00	0.00	0.00	88.18	9.32	5.32
TRAMO/SEATS	100.00	100.00	100.00	0.02	0.00	86.80	7.44	5.22

D Aggregation and Seasonal Adjustment

Table D.1: Rejection frequencies for diagnostic tests DGP1 – Constant yearly growth

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				Aggregation followed by SA				
Unadjusted data	3.90	4.50	6.40	95.20	99.80	4.30	16.30	6.20
X12-ARIMA	9.20	100.00	100.00	0.00	0.10	56.60	13.00	4.40
Dainties	7.20	95.00	100.00	2.50	51.30	7.10	6.70	7.90
TRAMO/SEATS	12.00	100.00	100.00	0.00	0.00	96.80	4.90	6.00
				SA followed by Aggregation				
Unadjusted data	3.90	4.50	6.40	95.20	99.80	4.30	16.30	6.20
X12-ARIMA	9.00	100.00	100.00	0.00	0.80	53.20	9.00	4.80
Dainties	6.20	91.70	99.90	3.00	57.00	7.80	6.00	4.90
TRAMO/SEATS	14.30	100.00	100.00	0.00	0.00	99.90	4.10	5.60

Table D.2: Rejection frequencies for diagnostic tests DGP2 – Deterministic seasonality

	Diagnostic test							
	HEGY (1)	HEGY (-1)	HEGY ($\pm$i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				Aggregation followed by SA				
Unadjusted data	3.80	100.00	100.00	0.10	100.00	100.00	48.10	10.80
X12-ARIMA	6.40	100.00	100.00	0.00	0.00	71.80	7.60	4.40
Dainties	7.20	100.00	100.00	0.00	1.20	15.90	9.80	15.80
TRAMO/SEATS	3.90	100.00	100.00	0.00	0.00	8.30	5.50	4.60
				SA followed by Aggregation				
Unadjusted data	3.80	100.00	100.00	0.10	100.00	100.00	48.10	10.80
X12-ARIMA	6.60	100.00	100.00	0.00	0.20	69.20	6.60	5.10
Dainties	6.50	100.00	100.00	0.00	0.00	15.50	5.30	5.10
TRAMO/SEATS	4.00	100.00	100.00	0.00	0.00	7.60	5.50	4.60

Table D.3: Rejection frequencies for diagnostic tests DGP3 – Stochastic seasonality

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				Aggregation followed by SA				
Unadjusted data	5.30	5.10	5.40	100.00	100.00	5.20	14.60	6.30
X12-ARIMA	10.00	100.00	100.00	0.00	0.10	54.20	15.80	3.90
Dainties	6.70	91.60	99.90	34.00	59.00	6.80	5.10	4.70
TRAMO/SEATS	13.90	100.00	100.00	0.10	0.00	97.90	9.90	6.30
				SA followed by Aggregation				
Unadjusted data	5.30	5.10	5.40	100.00	100.00	5.20	14.60	6.30
X12-ARIMA	10.10	100.00	100.00	0.00	0.30	53.90	10.20	3.90
Dainties	7.00	91.40	100.00	34.40	59.50	7.20	5.00	5.30
TRAMO/SEATS	16.80	100.00	100.00	0.00	0.00	99.80	10.00	6.00

Table D.4: Rejection frequencies for diagnostic tests DGP4 – Airline model

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
	Aggregation followed by SA							
Unadjusted data	16.20	12.50	11.60	34.60	99.80	100.00	22.40	10.90
X12-ARIMA	23.40	99.80	100.00	0.60	0.20	100.00	18.50	8.30
Dainties	20.70	80.60	95.20	1.50	50.10	100.00	15.90	9.20
TRAMO/SEATS	5.10	100.00	100.00	1.20	0.00	28.00	12.40	10.50
	SA followed by Aggregation							
Unadjusted data	16.20	12.50	11.60	34.60	99.80	100.00	22.40	10.90
X12-ARIMA	22.80	99.50	100.00	0.60	0.20	100.00	15.90	7.80
Dainties	20.40	81.30	95.40	1.50	50.00	100.00	16.70	8.70
TRAMO/SEATS	5.50	100.00	100.00	1.20	0.00	29.00	13.40	10.30

Table D.5: Rejection frequencies for diagnostic tests DGP5 – Generalized deterministic seasonality

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				Aggregation followed by SA				
Unadjusted data	4.70	100.00	100.00	0.90	100.00	100.00	64.00	9.30
X12-ARIMA	5.30	100.00	100.00	0.90	0.00	56.70	0.50	4.50
Dainties	4.70	100.00	100.00	0.90	0.00	92.90	12.30	6.10
TRAMO/SEATS	4.70	100.00	100.00	0.90	0.00	20.60	0.40	6.10
				SA followed by Aggregation				
Unadjusted data	4.70	100.00	100.00	0.90	100.00	100.00	64.00	9.30
X12-ARIMA	5.10	100.00	100.00	0.90	0.00	57.20	0.70	6.10
Dainties	4.60	100.00	100.00	0.90	0.00	95.70	27.90	6.70
TRAMO/SEATS	4.50	100.00	100.00	0.90	0.00	25.00	0.40	6.30

Table D.6: Rejection frequencies for diagnostic tests DGP6 – White noise

	Diagnostic test							
	HEGY (1)	HEGY (−1)	HEGY (±i)	CH	Seas. dum.	Seas. corr.	Per. AR	Seas. var.
				Aggregation followed by SA				
Unadjusted data	100.00	100.00	100.00	4.20	11.10	81.50	6.60	6.40
X12-ARIMA	99.70	100.00	100.00	0.00	0.00	95.60	10.60	5.90
Dainties	100.00	100.00	100.00	0.00	0.00	91.10	8.70	6.20
TRAMO/SEATS	100.00	100.00	100.00	0.00	0.00	86.30	6.90	6.30
				SA followed by Aggregation				
Unadjusted data	100.00	100.00	100.00	4.20	11.10	81.50	6.60	6.40
X12-ARIMA	99.80	100.00	100.00	0.10	0.20	95.10	10.20	6.80
Dainties	100.00	100.00	100.00	0.00	0.00	91.30	9.00	6.40
TRAMO/SEATS	100.00	100.00	100.00	0.10	0.00	86.30	6.90	5.80

E Actual data

Table E.1: Rejection frequencies actual data

	HEGY nonseasonal	HEGY[1] seasonal	Canova Hansen	Seasonal dummies	AR seas. freq.	PAR	Seasonal variance
Unadjusted data	2.00	89.67	2.33	67.67	34.67	34.67	34.00
X12-ARIMA	2.67	92.33	0.00	0.33	51.67	38.00	24.00
Dainties	2.00	89.67	0.00	0.67	29.33	31.00	37.33
TRAMO/SEATS	2.33	91.00	0.00	0.33	25.00	34.00	26.00

[1] Value gives the minimum over the rejection frequencies of the HEGY test per seasonal frequency.

References

Beaulieu, J. J. and J. A. Miron (1993), Seasonal Unit Roots in Aggregate U.S. Data, *Journal of Econometrics*, **55**, 305–328.

Bell, W. R. (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.

Canova, F. and B. E. Hansen (1995), Are Seasonal Patterns Constant over Time? A Test for Seasonal Stability, *Journal of Business and Economic Statistics*, **13**, 237–252.

Dezhbakhsh, H. and D. Levy (1994), Periodic Properties of Interpolated Time Series, *Economics Letters*, **44**, 221–228.

Doornik, J. A. (1999), *Object-Oriented Matrix Programming Using Ox*, 3rd edn., London: Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik.

Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto, and B.-C. Chen (1998), New Capabilities and Methods of the X-12-ARIMA Seasonal-Adjustment Program, *Journal of Business and Economic Statistics*, **16**, 127–177.

Franses, P. H. (1991), Seasonality, Nonstationarity and the Forecasting of Monthly Time Series, *International Journal of Forecasting*, **7**, 199–208.

Franses, P. H. (1994), A Multivariate Approach to Modeling Univariate Seasonal Time Series, *Journal of Econometrics*, **63**, 133–151.

Franses, P. H. (1996a), *Periodicity and Stochastic Trends in Economic Time Series*, Oxford University Press, Oxford.

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

Franses, P. H. and N. Haldrup (1994), The Effects of Additive Outliers on Tests for Unit Roots and Cointegration, *Journal of Business and Economic Statistics*, **12**, 471–478.

Franses, P. H. and R. M. Kunst (1999), On the Role of Seasonal Intercepts in Seasonal Cointegration, *Oxford Bulletin of Economics and Statistics*, **61**, 409–433.

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

Ghysels, E., C. W. J. Granger, and P. L. Siklos (1996), Is Seasonal Adjustment a Linear or Nonlinear Data Filtering Process?, *Journal of Business and Economic Statistics*, **14**, 374–386.

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

Gómez, V. and A. Maravall (1997), Programs TRAMO and SEATS, Instructions for the User, Beta Version: November 1997.

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

Koopman, S. J. and P. H. Franses (2002), Constructing Seasonally Adjusted Data with Time-Varying Confidence Intervals, *Oxford Bulletin of Economics and Statistics*, **64**, 509–526.

Perron, P. and T. J. Vogelsang (1992), Nonstationarity and Level Shifts with an Application to Purchasing Power Parity, *Journal of Business and Economic Statistics*, **10**, 301–320.

Schwert, G. W. (1989), Tests for Unit Roots - A Monte-Carlo Investigation, *Journal of Business and Economic Statistics*, **7**, 147–159.