Seasonal Adjustment Procedures Using a Related Series: An Application on the Industrial Production Index

Filippo Moauro
Marco Marini
SEASONAL ADJUSTMENT PROCEDURES USING A RELATED SERIES: AN APPLICATION ON THE INDUSTRIAL PRODUCTION INDEX

BY MARCO MARINI AND FILIPPO MOAURO

Istat, Direzione Centrale della Contabilità Nazionale

(Preliminary and incomplete version)

The paper discusses the results of an empirical application of seasonal adjustment when the information on a related series is taken into account. We consider the ARIMA model-based decomposition in comparison with the structural time series approach to seasonal adjustment. In the latter approach the related series is modelled together with the target series in a bivariate seemingly unrelated set-up.

The illustration is performed on the Italian industrial production index released by Istat. The related series is a composite index extracted from the most relevant business survey indicators released by Isae. We compare the accuracy of seasonal adjusted figures from the univariate ARIMA model-based and structural time series approaches against that resulting from the use of the related series in terms of both forecasting and revision errors.

KEYWORDS: ARIMA Models; Seemingly unrelated time series equation models; Seasonal adjustment; Calendar effects; Kalman filter and smoother.

JEL classification: C5, C22, E32.

1 Introduction

Seasonal adjustment of macroeconomic variables is carried out by National Statistical Institutes in order to satisfy the needs coming from a number of directions: from one side there is the need of seasonal adjusted figures for domestic short term analysis; from another, being in the European Union, a large set of indicators must be produced from all the Member States for compilation of macroarea indicators.

The official practice of seasonal adjustment consists in applying a signal extraction procedure to a univariate time series in order to get an estimate of the seasonal component and to remove it from the actual observations. The most applied approaches are implemented into the program X12 developed by the Bureau of the Census (see Findley at al., 1998) and the programs TRAMO-SEATS by Gomez and Maravall (1997). In summary, X12 consists in the application of a sequence of two-sided filters to the original observations of the series, where the final and initial values of the sample are supplemented respectively from forecasting future and initial observations by fitting an appropriate seasonal ARIMA (SARIMA) model. On the other hand through TRAMO-SEATS a SARIMA

1Marco Marini, Via Agostino Depretis 74/B, 00184 Roma, Italy, tel. (+39) 0646733209, e-mail marini@istat.it. Filippo Moauro, Via Agostino Depretis 74/B, 00184 Roma, Italy, tel. (+39) 0646733206, e-mail moauro@istat.it.
model is fitted, performing at a second stage a decomposition of the series into trend, seasonal and irregular components following the approach adopted by Burman (1980) and Hillmer and Tiao (1982).

An alternative approach is given by the structural time series models proposed by Harvey and Todd (1983). In this case the components are explicitly defined, and seasonal adjustment simply consists into removing from the original observations the estimates of the seasonal component. Another advantage in the use of this approach is the possibility to move from the univariate to the multivariate representation quite straightforwardly, following Harvey (1989) and Harvey and Koopman (1997).

Seasonal adjusted releases are for their nature subject to revisions. Two main reasons are mentioned from the literature: the first source of revision arises since last observations of a time series are often preliminary, becoming ‘final’ or ‘fully revised’ only after a number of periods; the latter simply depends on the full use of the information coming from the new observations at the end of the sample from a release date to another. Data revision and the rationality of late predictions and early releases have been object of discussion in many works: very shortly, from the first works by Morgenstern (1963), to Howrey (1983), till the recent paper by Swanson and van Dijk (2006).

The use of a related series to perform seasonal adjustment is another important issue to take into account. An explanatory variable might enter into the seasonal adjustment strategy basically into two alternative ways: 1) using it as a regressor in a univariate model, then assuming weakly exogeneity with respect to the target series. In this case, the related series is seasonal adjusted at a first stage and then the desired univariate procedure is applied on the residuals of the regression; 2) estimating a multivariate model, thus removing the assumption of exogeneity of the explanatory variable. The latter strategy is for sure less complicated and more transparent of the first, with the advantage of simultaneity in model estimation.

An assessment on how the add of the information of related series might improve seasonal adjustment of a target variable is treated in this paper from an empirical standpoint; the illustration is on the Italian industrial production index. The multivariate approach has been preferred to the two-step procedure, presenting a bivariate BSM set-up where the external information is summarized in a unique composite indicator. The performances of this model in terms of forecasts and revision errors of seasonal adjusted releases have been compared to two univariate specifications: the first referred to the univariate BSM without explanatory variables; the latter to the ARIMA model-based seasonal adjustment procedure.

The plan of the paper is the following: the next section introduces the statistical models for seasonal adjustment focusing on the BSM and the ARIMA model-based decomposition. Section 3 presents the dataset, the model estimates and discusses the results of the rolling forecasting experiment. Section 4 shortly concludes.
2 Statistical models for seasonal adjustment

This section reviews two popular formulations of models to handle seasonal adjustment. In particular we consider the basic structural model (BSM henceforth) discussed in Harvey (1989), and the SARIMA models proposed by Box and Jenkins (1970). Despite the fact that the models present similar reduced forms, the way in which seasonal adjustment is performed is quite different: from one side structural models have the advantage to model unobserved components which have a direct interpretation, so that it only requires to remove the estimated seasonality from the original observations to obtain a seasonal adjusted series; from another the ARIMA form is more parsimonious in terms of parameters to be estimated, even if it requires a signal extraction procedure like the canonical decomposition to obtain a series adjusted for seasonality; further, the structural approach presents a multivariate extension which allows a contemporaneous treatment of seasonality for a cross section of time series.

2.1 The basic structural model

The BSM considers an additive decomposition of the series into a trend, a seasonal and an irregular component. It might be formulated for both univariate or multivariate time series. Given a set of time series $y_t = (y_{1t}, ..., y_{Nt})'$ it is such that:

$$y_t = \mu_t + \gamma_t + \Lambda x_t + \epsilon_{it}, \quad t = 1, \ldots, n, \quad \epsilon_{it} \sim \text{NID}(0, \Sigma_{\epsilon})$$

where the trend, $\mu_t$, is a local linear component:

$$\begin{align*}
\mu_{t+1} &= \mu_t + \beta_t + \eta_t, \quad \eta_t \sim \text{N}(0, \Sigma_\eta) \\
\beta_{t+1} &= \beta_t + \zeta_t, \quad \zeta_t \sim \text{N}(0, \Sigma_\zeta)
\end{align*}$$

(1)

This representation is such that each element $\eta_{it}, \zeta_{it}$ and $\epsilon_{it}$ for $i \neq j$, respectively.

The seasonal component is represented by $\gamma_t = (\gamma_{1t}, ..., \gamma_{Nt})'$. For monthly time series the $i$-th element $\gamma_{it}$ arises from the combination of six stochastic cycles defined at the seasonal frequencies $\lambda_j = 2\pi j/12$, $j = 1, \ldots, 6$, $\lambda_1$ representing the fundamental frequency (corresponding to a period of 12 monthly observations) and the remaining being the five harmonics (corresponding to periods of 6 months, i.e. two cycles in a year, 4 months, i.e. three cycles in a year, 3 months, i.e. four cycles in a year, 2.4, i.e. five cycles in a year, and 2 months):

$$\gamma_{it} = \sum_{j=1}^{6} \gamma_{ij,t} = T^{\gamma_j} \begin{bmatrix}
\gamma_{ij,t+1} \\
\gamma_{ij,t+1} \\
\end{bmatrix} = T^{\gamma_j} \begin{bmatrix}
\gamma_{ij,t} \\
\omega_{ij,t} \\
\end{bmatrix}, j = 1, \ldots, 5,$$

(2)

$$T^{\gamma_j} = \begin{bmatrix}
\cos \lambda_j & \sin \lambda_j \\
-\sin \lambda_j & \cos \lambda_j
\end{bmatrix}, \quad j = 1, \ldots, 5.$$
and $\gamma_{i6,t+1} = T^{76}\gamma_{i6t} + \omega_{i6t}$, with $T^{76} = -1$. The disturbances $\omega_{jt}$ and $\omega^*_jt$ are normally and independently distributed with common covariance matrix $\Sigma_w$ for $j = 1, \ldots, 6$. The symbol $\epsilon_t$ denotes the irregular component.

The vector $x_t$ is a $K \times 1$ vector of regressors accounting for calendar effects and $\Lambda$ is a matrix of unknown regression coefficients, each row representing the effect of a single series. The regressors considered in this paper for the calendar effects are 3: the first concerns the working/trading days in the single regressor representation; the second the Easter effect; the third the length of the month or leap-year effect.

According to this model specification, the variables $y_{1t}, \ldots, y_{Nt}$ form a *Seemingly Unrelated Time Series Equations* system (Harvey, 1989). There is no cause and effect relationship among them, but they are subject to the same underlying economic environment.

The univariate formulation results quite straightforward from the representation above substituting the vectors with scalars for the structural components and the covariance matrices with scalar values of disturbance variances.

### 2.2 The ARIMA model-based decomposition

A representation for non-stationary time series with seasonal component can be derived within the class of ARIMA models. Consider the simple ARIMA model for the process $y_t$

$$\phi(B)(1 - B)^dy_t = \theta(B)\epsilon_t \quad (3)$$

where $\phi(B) = 1 - \phi_1B - \ldots - \phi_pB^p$ and $\theta(B) = 1 - \theta_1B - \ldots - \theta_qB^q$ are polynomials in the lag operator $B$ and $\epsilon_t$ is a stochastic process containing a seasonal component of period $s$. Then, it is convenient to write

$$\Phi(B^s)(1 - B^s)^Da_t = \Theta(B^s)\epsilon_t \quad (4)$$

where $\Phi(B) = 1 - \Phi_1B - \ldots - \Phi_PB^P$, $\Theta(B) = 1 - \Theta_1B - \ldots - \Theta_QB^Q$ and $\epsilon_t$ is a white-noise process. Combining (3) and (4), the overall model for $y_t$ becomes

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^Dy_t = \theta(B)\Theta(B^s)\epsilon_t \quad (5)$$

which is called the *seasonal ARIMA* model for $y_t$ (Box, Jenkins, and Bacon, 1967). The representation (5) is generally denoted with ARIMA$(p, d, q) \times (P, D, Q)_s$ to indicate the orders of seasonal and non-seasonal filters.

As shown by the Wold theorem, the process $y_t$ can also be determined by deterministic components. Typical examples in economic time series are calendar components (working days, Easter, leap-year effects) and outliers. These effects are generally searched by the researcher in a preliminary analysis of the series, in order to detach them from the original series and obtain an estimate of its stochastic part. Denoting by $x_t$ the $k \times 1$ vector containing such deterministic effects, model (5) assumes the more appropriate form

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D(y_t - \lambda'x_t) = \theta(B)\Theta(B^s)\epsilon_t \quad (6)$$
where $\lambda$ is the $k \times 1$ vector with the coefficients associated to $x_t$.

Suppose now that the observable series $y_t$ is composed by unobservable components as

$$y_t = T_t + S_t + \varepsilon_t$$

where $T_t$ is the trend-cycle, $S_t$ is the seasonal and $\varepsilon_t$ is the irregular component. Then, it is assumed that each of the components follows an ARIMA model

$$\phi_T(B)T_t = \theta_T(B)a_{1t}$$
$$\phi_S(B)S_t = \theta_S(B)a_{2t}$$
$$\phi_\varepsilon(B)\varepsilon_t = \theta_\varepsilon(B)a_{3t}$$

where each polynomial has its roots on or outside the unit circle. Obviously, equations in (7) cannot be specified since the components are unknown. They can be derived from the observable model (6) solving a system of covariance equations: since the number of unknowns is smaller, the system is underidentified. This implies that there exists an infinite number of ways to decompose the series $y_t$ (Planas, 1998). Assuming the canonical decomposition, Box, Hillmer and Tiao (1978) found a single admissible solution with the property of maximize the variance of the irregular component. Given the component models, their estimation can be obtained through the application of appropriate (infinite) linear filters. The Wiener-Kolomogorov filter is a natural choice in signal extraction, providing minimum MSE estimates of the components. The seasonal adjusted series is finally derived removing $S_t$ from the original series.

The programs TRAMO and SEATS (Gomez and Maravall, 1997) implements the seasonal adjustment procedure described in this section. This approach is known in the literature as the **ARIMA model-based** decomposition of time series. The use of TRAMO-SEATS in official statistics is widespread, mainly in European statistical agencies: its flexibility and robustness have made it very appealing for NSIs as an automatic and easy-to-run tool in SA production process.

### 3 The application

The following sections give a detailed description of the experiment. First we present the time series involved in the empirical analysis; then, the results of the factor analysis for computation of the composite indicator used in the experiment; third, the three estimated models for the industrial production index using both the SARIMA and the BSM approaches; finally, the rolling forecasting exercise to make an assessment on the utility of a related series for seasonal adjustment and to compare the forecasting accuracy and revision process of the resulting seasonal adjusted series.
The target series is given by the Italian industrial production index (IPI henceforth) released monthly by Istat. It is referred to the total of the industrial economic activities with base year $2000 = 1000$, is seasonal unadjusted with sample period from January 1981 to December 2005. Its plot is given in figure 1.

The related indicator has been computed as a synthetic composite index from a set of series of the business survey on manufacturing and extractive firms in Italy released monthly from Isae. Five series have been taken into account: the current trend of production and inventories, the domestic and foreign order books together with their total. All the series refer to the net value between positive and negative answers provided monthly from the survey in the period from January 1981 to December 2005. The preference to these indicators is due to the circumstance that these series are an alternative measure of the trend of production, resulting strongly related to it, with the appealing property to be available before the end of the relevant month and then in advance of around 30 days with respect to IPI.

The composite index has been built extracting the first principal component from the factorial analysis of the five series, which explained the 87.6% of total variance. The weights of the five series (the eigenvector associated to the first eigenvalue) resulted all with the expected sign, i.e. all positive except the inventories, estimated equal to $-0.40$; the trend of production was equal to 0.46, whereas the domestic and foreign order books
resulted 0.46 and 0.45 respectively, with their total equal to 0.47. The original series together with the first factor are plotted in figure 2.

Figure 2: The factor associated to the first eigenvector (upper-left panel) from selected business survey variables (January 1981-December 2005)

3.2 The estimated ARIMA model

The procedure proposed by Box and Jenkins (1970) was followed to identify and estimate an appropriate ARIMA model for IPI. The programs TRAMO-SEATS are used for estimation and decomposition into unobserved components. The sample period is the same used for the structural model (January 1981-December 2005).

The variability of the series appears relatively constant over this period, so we decided to model the levels of the series. Firstly, we obtain a linearization of the series \( y_t \) by adjusting for three calendar effects (working days \( wd_t \), Easter \( e_t \)) and leap year \( ly_t \)) and additive outliers in August 1984 \( (o1) \) and August 1987 \( (o2) \), automatically detected by the procedure of Tsay (1986) available in TRAMO. The next equation shows the estimated coefficients (by exact ML estimation) with their relative standard errors in parenthesis

\[
y_t = 7.152 \, wd_t + 22.432 \, ly_t - 11.254 \, e_t + 38.409 \, o_1 - 21.005 \, o_2 + g_t^f \tag{8}
\]
An ARIMA model was then searched for the linearized series $y_t$. From its global and partial autocorrelation function we noted that both difference and seasonal difference operators are needed to make the series stationary. The differenced series is adequately represented by an ARMA $(2, 0) \times (0, 1)$: the final estimated equation is

$$
(1 + 0.574B + 0.319B^2) (1 - B)(1 - B^{12}) y_t = (1 - 0.651B^{12}) \varepsilon_t. \tag{9}
$$

All the diagnostic checks in TRAMO do not reject the null hypothesis of randomness of the estimated residuals $\varepsilon_t$. The stationary AR polynomial has complex conjugates roots with argument 120.54, which corresponds to a seasonal harmonic of three-months period. It is therefore assigned to the seasonal component of the series. As we will see, this attribution turns out to be determinant for the smoothness of the seasonal adjusted series. The standard deviation of the residuals of model (9) is 17.59. The resulting models for the components are

$$
(1 - B)^2 T_t = (1 + 0.035B - 0.965B^2) a_{1t} \quad \sigma_{\alpha_1}^2 = 0.047
$$

$$
U(B) S_t = \theta_S(B) a_{2t} \quad \sigma_{\alpha_2}^2 = 0.175
$$

$$
\sigma_{\alpha_3}^2 = 0.171
$$

with

$$
U(B) = 1 + 1.574B + 1.893B^2 + 1.893B^3 + 1.893B^4 + \ldots + 1.893B^5 + 1.893B^6 + 1.893B^7 + 1.893B^8 + \ldots + 1.893B^9 + 1.893B^{10} + 1.893B^{11} + 1.893B^{12} + 0.319B^{13}
$$

$$
\theta_S(B) = 1 + 0.842B + 0.482B^2 + 0.323B^3 + 0.134B^4 + \ldots - 0.055B^5 - 0.228B^6 - 0.369B^7 - 0.454B^8 + \ldots - 0.475B^9 - 0.360B^{10} - 0.410B^{11} - 0.469B^{12} - 0.228B^{13}.
$$

Table 1 shows the percentage reduction in the standard error of the revision with respect to the concurrent estimator of $T_t$ and of the seasonal adjusted series ($SA_t$) after additional years. The process of convergence towards the final estimator is very fast for $T_t$, but the performance for $SA_t$ is also appreciable: more than 50% of revision is in fact eliminated after one additional year.
Table 1: Percentage reduction in the standard error of the revision in comparison with the concurrent estimator

<table>
<thead>
<tr>
<th>Additional obs.</th>
<th>$T_t$</th>
<th>$SA_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 months</td>
<td>72.63</td>
<td>52.19</td>
</tr>
<tr>
<td>24 months</td>
<td>82.18</td>
<td>68.88</td>
</tr>
<tr>
<td>36 months</td>
<td>88.40</td>
<td>79.74</td>
</tr>
<tr>
<td>48 months</td>
<td>92.45</td>
<td>86.81</td>
</tr>
<tr>
<td>60 months</td>
<td>95.09</td>
<td>91.42</td>
</tr>
</tbody>
</table>

3.3 The estimated bivariate BSM

Cast the bivariate BSM of Section 2.1 in state space form, maximum likelihood estimation of unknown parameters has been obtained using the Kalman filter and smoothing algorithms in Ox, version 3.0 (see Doornik 2001), and the SsfPack package (see Koopman, Shephard and Doornik 1999, 2002). Over the full sample period, January 1981-December 2005, the parameter estimates were as follows:

$$
\hat{\sigma}_{1\eta} = 6.636, \quad \hat{\sigma}_{2\eta} = 0.409, \quad \hat{\rho}_{\eta} = 0.654,
\hat{\sigma}_{1\omega} = 0.743, \quad \hat{\sigma}_{2\omega} = 0.003, \quad \hat{\rho}_{\omega} = 1,
$$

with $\hat{\sigma}_{1\epsilon} = 11.002$ and $\sigma_{1\zeta} = \sigma_{2\zeta} = 0$, where $\sigma_{1h}$ and $\sigma_{2h}$, $h = \eta, \zeta, \epsilon$, are the square root of the elements of $\Sigma_h$ on the principal diagonal, the suffix 1 denotes the industrial production index and 2 the composite index; finally, $\rho_h$ represents the correlation of the same component-disturbances among the two series; the maximised log-likelihood is equal to $-1402.906$.


These results show that for both the series the trend features a constant slope, since its disturbance variance is zero; as a result the trend is a bivariate random walk with a constant drift, with positively, but not perfectly, correlated disturbances ($\rho_\eta$ is estimated equal to 0.654). This suggests that the series are not cointegrated. The composite index has the variance of the irregular component equal to zero, hence resulting equal to the estimated trend.

The non-zero values for the seasonal variance parameters $\sigma_{1\omega}$ and $\sigma_{2\omega}$ indicate that the seasonal pattern is time varying in the sample period for both the series; of course this source of volatility is very low for the composite index in comparison with IPI. Moreover, the relative disturbances resulted perfectly correlated indicating seasonal cointegration.

The model specification also included the regressors representing the calendar effects: the estimated coefficients for the industrial production index, denoted respectively by $\hat{\delta}_{i}$,
\[ \hat{\delta}_i^{\text{Easter}} \] and \[ \hat{\delta}_i^{\text{LOM}} \] have been

\[
\begin{align*}
\hat{\delta}_1^* & = 7.122 \
(0.289) \\
\hat{\delta}_1^{\text{Easter}} & = -22.463 \\
(4.233) \\
\hat{\delta}_1^{\text{LOM}} & = 28.674 \\
(6.208)
\end{align*}
\]

where in parenthesis are reported the standard errors. All the parameters of IPI are significant and have the expected sign, concluding that the calendar effect is significant for this series. For the composite index only the trading day regressor was included in model specification: its estimate resulted equal to 0.017 with a standard error of 0.005; all the other coefficients were removed from the model since not significant.

The diagnostics based on innovations for IPI are \( Q(9) = 9.347 \) and \( NORM = 0.090 \), while for the composite index are \( Q(9) = 26.663 \) and \( NORM = 0.258 \), where \( Q(P) \) is the Box-Ljung statistic based on the \( P \) autocorrelations and \( NORM \) is the Bowman-Shenton normality test statistic based on skewness and kurtosis of disturbances. If the model is correctly specified, the Box-Ljung statistics are asymptotically distributed with \( P-1 \) degrees of freedom, whereas \( NORM \) is distributed as a \( \chi^2 \) with 2 degrees of freedom. There is no evidence of non normality in both the series; concerning the serial correlation, the composite index disturbances show a high \( Q \)-statistic. However, we didn’t consider alternative specifications including cyclical or autoregressive components because the model shows a general good fit and we are generally interested in the performances in terms of forecasts and revision errors.

The monthly estimates adjusted for seasonality and calendar effects are presented and compared to those obtained using TRAMO-SEATS in figure ???. The plot highlights that seasonal adjustment performs quite different among the two procedures: in particular the SA estimates resulting from the application of the univariate procedure of TRAMO-SEATS are smoother than those obtained through the bivariate BSM.

A univariate BSM has been also fitted to IPI producing similar results in terms of parameter estimates, notably \( \hat{\sigma}_\eta = 7.176 \), \( \hat{\sigma}_\epsilon = 11.619 \), and \( \hat{\sigma}_\omega = 0.668 \); the parameters of calendar effects resulted \( \hat{\delta}_1^* = 7.014 \) \((0.297)\), \( \hat{\delta}_1^{\text{Easter}} = -24.784 \) \((4.549)\) and \( \hat{\delta}_1^{\text{LOM}} = 25.242 \) \((6.802)\) where in parenthesis are the standard errors.

### 3.4 The rolling forecasting exercise

The rolling forecasting exercise has been carried out computing 84 one-step-ahead seasonal adjusted forecasts of IPI: at the first step the sample is set to 216 observations, covering the period January 1981-December 1998; this sample has been used to provide an estimate of IPI adjusted for seasonality and calendar effects, together with the corresponding forecast over January 1999: in the bivariate BSM application the estimate includes the information of the composite index up until January 1999, simulating the occurrence that it leads the release of IPI for the same month.

From the second step on the same exercise is carried out, shifting of one month ahead the last observation; at the end, we have produced the history of 84 ‘vintages’ of seasonal adjusted series including 84 one-step-ahead forecasts for the months from January 1999.
to December 2005. In other terms, we have computed 84 one-step-ahead forecasts, 84 concurrent SA estimates for the months from December 1998 to November 2005, 83 first revisions for the months up until October 2005 and so on. The final estimate is provided using all the information available (that obtained from model estimation discussed in previous Sections), which in our exercise represents the ‘fully revised estimate’.

From the set of vintages we exclude the ones corresponding to the last 24 months, assuming that after 2 years the SA estimate of a given month gains a reasonably steady value. From the remaining 60 months, from January 1999 to December 2003, it has been computed a set of root mean squared error (RMSE) statistics based on the one-step-ahead forecast errors together with the RMSEs of the first up until the 13-th preliminary estimate. The errors are computed in terms of both the level and the monthly growth rates of the series.

The exercise can be formalised in terms of estimation errors of the preliminary seasonally adjusted series. Let \( y_{sa|T}^t \) denote the fully revised seasonally adjusted estimate of quarter \( t \) conditional to the full sample of length \( T \). For \( t << T \), this can be considered as a good approximation of the final estimation. In our exercise \( T = 300 \) and \( t = 217, ..., 276 \) (January 1998-December 2003), which represent the 60 vintages to compare with the final estimate. The preliminary estimation error is defined as

\[
\epsilon_{t|t+h}^{sa} = y_{t|t+h}^{sa} - y_{t|T}^{sa} \tag{10}
\]
where $y_{t|t+h}^{sa}$ is the preliminary estimate of the seasonally adjusted series at quarter $t$ conditional to the information up to quarter $t+h$. When $h = -1$, we obtain a forecasting error $e_{t|t-1}^{sa}$; for $h = 0$ the formula (10) defines the revision error in the so-called concurrent estimator. As $h$ increases, the errors $e_{t|t+h}^{sa}$ are expected to converge to the final estimator $y_{t|T}^{sa}$.

The RMSE statistics computed on the errors $e_{t|t+h}^{sa}$ are reported in Table 2: the values in the first two columns refer to the RMSEs computed on the history of SA estimates obtained from the fit of the bivariate and univariate BSM, whereas the third concerns the ARIMA specification; all these statistics are computed in terms of levels, whereas the values of the second term of columns reports the corresponding statistics based on monthly growth rates.

The assessment on the utility of the related series to seasonal adjustment is derived from comparing the performances among the two BSMs; in fact, the comparison with the third approach is affected from the smoother SA estimates performed from the canonical decomposition (see Figure 3). Then, it is clear that the bivariate BSM outperforms the univariate alternative both in terms of forecast and revision errors, either in the levels and monthly growth rates. The gain of the first on the latter approach in terms of RMSE of their forecasts is of 19.7% and 27.3% respectively for the levels and the monthly growth rates. We might then conclude that the related series is useful to reduce the forecasting and the revision errors of SA estimates.

At first sight what is surprising in Table 2 is the high value of RMSE of the SA forecasts referred to the TS exercise: notably, 59.3% and 98.3% higher of the univariate and bivariate specifications, respectively. On the other hand, moving to the preliminary releases the advantage of TS on the BSM goes from 40.2% for the second up until to 36.0% for the 13-th release.

The low performance of forecast errors for the ARIMA approach is clearly explained following the discussion in Harvey (1989, ch. 6.1 and 6.2): the canonical decomposition yields a forecast function of the trend which is not anchored to its estimate at the end of the sample, whereas it is for the BSM. On the opposite, inside the sample the trend of the first approach is smoother in comparison with the latter, then reducing its volatility and as a consequence the revision errors. The estimate of a smooth trend influences so much the results of the experiment at point that the advantage in the use of a related series completely disappears looking at the statistics in terms of growth rates.

### 4 Conclusions

This article has proposed an experiment on the Italian industrial production index pertaining the utility of using a related series in seasonal adjustment. Three alternative models have been fitted to monthly data: 1) a seasonal ARIMA model; 2) a univariate basic structural model; 3) a bivariate basic structural model of the industrial production together with a composite index of the business surveys. All the models include the regressors to enable the adjustment for calendar effects. For each model a rolling forecasting experiment has been carried out, producing a set of RMSE statistics to assess the
Table 2: RMSE of $e_{t|t+h}^{sa}$ at different steps (levels and monthly growth rates)

<table>
<thead>
<tr>
<th>step h</th>
<th>levels biv BSM</th>
<th>levels univ BSM</th>
<th>TS</th>
<th>monthly growth rates biv BSM</th>
<th>monthly growth rates univ BSM</th>
<th>TS</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>5.4653</td>
<td>6.8027</td>
<td>10.8378</td>
<td>0.7861</td>
<td>1.0815</td>
<td>0.7332</td>
</tr>
<tr>
<td>0</td>
<td>4.7143</td>
<td>5.4999</td>
<td>4.7798</td>
<td>0.7758</td>
<td>0.8309</td>
<td>0.4201</td>
</tr>
<tr>
<td>1</td>
<td>4.8010</td>
<td>4.9414</td>
<td>2.9533</td>
<td>0.7906</td>
<td>0.8299</td>
<td>0.3773</td>
</tr>
<tr>
<td>2</td>
<td>4.9102</td>
<td>4.9450</td>
<td>2.2646</td>
<td>0.7968</td>
<td>0.8238</td>
<td>0.3113</td>
</tr>
<tr>
<td>3</td>
<td>4.8645</td>
<td>5.0552</td>
<td>2.4000</td>
<td>0.7995</td>
<td>0.8317</td>
<td>0.3039</td>
</tr>
<tr>
<td>4</td>
<td>4.9181</td>
<td>4.9825</td>
<td>2.3164</td>
<td>0.8075</td>
<td>0.8348</td>
<td>0.2995</td>
</tr>
<tr>
<td>5</td>
<td>4.9485</td>
<td>5.0535</td>
<td>2.3741</td>
<td>0.8116</td>
<td>0.8387</td>
<td>0.3030</td>
</tr>
<tr>
<td>6</td>
<td>4.9239</td>
<td>5.0292</td>
<td>2.3497</td>
<td>0.8064</td>
<td>0.8384</td>
<td>0.3001</td>
</tr>
<tr>
<td>7</td>
<td>4.9205</td>
<td>5.0196</td>
<td>2.3277</td>
<td>0.8064</td>
<td>0.8338</td>
<td>0.2988</td>
</tr>
<tr>
<td>8</td>
<td>4.9371</td>
<td>5.0174</td>
<td>2.3367</td>
<td>0.8087</td>
<td>0.8344</td>
<td>0.3005</td>
</tr>
<tr>
<td>9</td>
<td>4.9414</td>
<td>5.0280</td>
<td>2.3217</td>
<td>0.8105</td>
<td>0.8379</td>
<td>0.2996</td>
</tr>
<tr>
<td>10</td>
<td>4.9407</td>
<td>5.0563</td>
<td>2.2654</td>
<td>0.7086</td>
<td>0.8388</td>
<td>0.3002</td>
</tr>
<tr>
<td>11</td>
<td>3.4740</td>
<td>5.0428</td>
<td>2.3543</td>
<td>0.4659</td>
<td>0.7553</td>
<td>0.2708</td>
</tr>
<tr>
<td>12</td>
<td>2.9741</td>
<td>3.7594</td>
<td>1.9041</td>
<td>0.4538</td>
<td>0.5303</td>
<td>0.1833</td>
</tr>
</tbody>
</table>

accuracy of seasonal adjustment in terms of forecasting and revision errors.

The main result of the experiment is that the seasonal adjusted estimates of the bivariate structural time series model produce a lower RMSE than the corresponding univariate form, both in terms of forecasting and revision errors. Then it might be concluded that the use of the related series might improve the seasonal adjustment process.

However the appealing of the decomposition based on ARIMA models remains high for Statistical Agencies since it produces very low revision errors of the seasonal adjusted series. This is due to the fact that the estimates of the SA series are particularly smooth, giving an idea of stability from a release to another.

5 Acknowledgments

We thank Giovanni Savio for a useful discussion on the issues covered in the paper and Marco Malgarini to have provided us the set of series on the business survey. The usual disclaimers apply.

References


