Journal of Official Statistics, Vol. 18, No. 4, 2002, pp. 511–530

Measures to Evaluate the Discrepancy Between Direct and Indirect Model-Based Seasonal Adjustment

Edoardo Otranto[1] *and Umberto Triacca*[2]

In this article we deal with the problem of the evaluation of the discrepancy between direct and indirect seasonal adjustment. In a model-based framework, the direct seasonally adjusted series seems to be preferable, but a large discrepancy over the indirect seasonally adjusted series can cause confusion among the users. This is a crucial problem in respect of dissemination policy for the National Statistical Institutes. We propose a new approach to evaluate the size of the discrepancy, based on the idea that the two data generating processes of the alternative series (the direct and the indirect seasonally adjusted series) can be compared in terms of dissimilarity measures between RegARIMA models. A small dissimilarity implies that the difference between direct and indirect series is negligible. The procedure is performed in terms of classical hypothesis tests.

Key words: Time series; ARMA; distance; forecastability.

1. Introduction

A seasonal time series can be the result of adding up two or more sub-series (possibly weighted). From the point of view of the seasonal adjustment policy, two natural alternatives arise to seasonally adjust the aggregated series:

1) the aggregated series is seasonally adjusted on its own (direct method);
2) the aggregated seasonally adjusted series is obtained as a sum of seasonally adjusted sub-series (indirect method).

It is clear that the two approaches may give different results, and a problem of choice of method arises.

Some criteria are proposed in the literature with regard to choosing the direct or the indirect method. In the seminal paper of Geweke (1978), the direct and indirect methods are compared, using the mean squared error (MSE) criterion, whilst Dagum (1979) uses two measures of lack of smoothness of the seasonally adjusted series. Lothian and Morry (1977) indicate small revision errors as an important factor in choosing the method for seasonal adjustment, and Ghysels (1997) suggests the final estimation error. A criterion based on the stability of the seasonally adjusted series is used in the X-12-RegARIMA

[1] Istituto Nazionale di Statistica, Via Tuscolana, 1776, I-00173 Roma, Italy. Email: otranto@istat.it
[2] Università degli Studi di L'Aquila, Piazzale del Santuario, 19, I-67040, Roio Poggio (AQ), Italy. Email: triacca@ec.univaq.it
Acknowledgments: We thank the participants in the conference ''Seasonality in Economic and Financial Variables'' (Faro, October 5–6, 2000) for their useful suggestions and comments. The article benefits from the useful comments of the Associate Editor E. B. Dagum and two anonymous referees. We acknowledge the use of computer routines reported in Corduas (2000).

program, based on sliding spans and month-to-month changes (Findley et al. 1998). Another approach was proposed by den Butter and Fase (1991), allocating the discrepancy between direct and indirect methods among the seasonally adjusted sub-series, proportionally to the variance of the sub-series. In practice, they create a new seasonally adjusted series, different from those obtained from direct or indirect methods.

The most recent developments refer directly to the model-based approach. Planas and Campolongo (2000) base their analysis both on final estimation errors and on total revisions in concurrent estimates, applying this procedure to the industrial production series of European Monetary Union countries. Gómez (2000) has proposed a criterion based on empirical revisions, measured with three alternative statistics.

The limitation of these approaches is that the choice is based on a single aspect of the seasonal adjustment (smoothness, errors, revisions, stability, etc.), which can vary with the kind of series or subjectively, according to the point of view of the researcher.

The direct method can be preferable since the aggregated adjusted series is clearly of a higher quality. Furthermore, following a model-based approach, the direct method is the natural choice because the seasonally adjusted series is derived from the ARIMA model of the original series. On the other hand the indirect method allows consistency in aggregation.

In general, the National Statistical Institutes consider quality and consistency equally important.

A typical common criterion (recommended by Eurostat) is the use of direct seasonal adjustment if the discrepancy between methods is acceptable, and the use of indirect seasonal adjustment if this discrepancy is relevant. But when is the discrepancy relevant?

The simple evaluation of the size of the discrepancies is not sufficient; it depends on the series analyzed and its magnitude. In addition, two series can be similar for the absolute discrepancies but they can have different behaviours in terms of period-to-period variations (the period-to-period variations are relevant in dealing with seasonally adjusted data).

If the seasonal adjustment method is a model-based one, we believe that a correct approach should consider the stochastic properties of the series. In other terms, we regard a discrepancy as relevant when the data generating process (DGP) of the indirect seasonally adjusted series is different from the DGP of the direct seasonally adjusted one. This approach is different from the others because we do not consider a particular aspect of the seasonal adjustment, but its stochastic properties, concerning the DGP. We propose the use of Piccolo's (1990) distance as a measure of difference between the two DGP's.[3] Another approach to evaluating the difference between DGP's is based on the comparison of the forecasts produced from the two models; for this purpose we use a Diebold and Mariano (1995) style test.

We want to stress that we do not propose a solution to the problem of the choice between direct or indirect seasonal adjustment, but only some instruments to evaluate the discrepancy between them in a model-based framework; these instruments can support the choice of diffusion policy in respect of seasonally adjusted data.

In the next section we formalize the consistency problem. In Section 3 we derive

[3] Of course other measures of dissimilarity could be used; for example, an interesting alternative is the distance between filters used in Depoutot and Planas (1998).

the DGP for direct and indirect seasonally adjusted series, and then in Section 4 the two procedures are described. In Section 5 an application of the approaches is provided. Concluding remarks follow.

2. The Consistency Problem

The purpose of this section is to provide a formal framework to analyze the so-called consistency problem. For the sake of simplicity, let us consider an observed seasonal time series Y_t composed of two sub-series X_t and Z_t, via the relationship

$$Y_t = \lambda_x X_t + \lambda_z Z_t, \quad t = 1, \ldots, T$$

where λ_x and λ_z are known constants. Furthermore, we assume that the observable time series Y_t, X_t, Z_t can be decomposed as

$$Y_t = Y_t^{ns} + Y_t^s$$
$$X_t = X_t^{ns} + X_t^s$$
$$Z_t = Z_t^{ns} + Z_t^s$$

where Y_t^{ns}, X_t^{ns}, Z_t^{ns} are the nonseasonal components containing the trend, the cycles and the irregular components, and Y_t^s, X_t^s, Z_t^s are the seasonal components.

The results of this article will be valid even if the aggregation formula is multiplicative:

$$Y_t = X_t^{\lambda_x} Z_t^{\lambda_z} \quad t = 1, \ldots, T$$

and the decomposition follows a multiplicative model too:

$$Y_t = Y_t^{ns} Y_t^s$$
$$X_t = X_t^{ns} X_t^s$$
$$Z_t = Z_t^{ns} Z_t^s$$

In fact using the log-transformation we are in the presence of additive models and additive aggregation.

A desired property of seasonal adjustment procedures is that

$$Y_t^{ns} = \lambda_x X_t^{ns} + \lambda_z Z_t^{ns} \tag{1}$$

In general the consistency requirement (1) is not satisfied: the seasonally adjusted series obtained directly from the composed series Y_t is not equal to the sum of the seasonally adjusted components. In this case, a natural question arises: is it ‘‘better’’ to seasonally adjust the aggregated series (*direct* method) or to aggregate the seasonally adjusted sub-series (*indirect* method)?

If the discrepancy between direct and indirect seasonal adjustment

$$D_t^{ns} = Y_t^{ns} - (\lambda_x X_t^{ns} + \lambda_z Z_t^{ns}), \quad t = 1, \ldots, T$$

is ‘‘negligible,’’ the direct seasonal adjustment is preferable, because the seasonally adjusted composite series is clearly of a higher quality, especially when a model-based approach is used. Furthermore, the correlation structure between X_t and Z_t cannot be captured with the indirect method, whereas this problem does not exist if there is direct modelling of the aggregate series. Finally the seasonally adjusted series obtained by using the indirect method may still present spurious seasonality (for this reason, a common procedure is to test for residual seasonality). On the other hand, a strong difference between the direct seasonally adjusted series and the sum of the seasonally adjusted sub-series (a large D_t^{ns}) can cause confusion among the users of data.

In this article we utilize formal statistical tests to evaluate if the discrepancy is negligible. In particular, we consider the model-based approach, so that every series follows a RegARIMA model; it implies that nonseasonal and seasonal components follow RegARIMA models too. If the RegARIMA model for the observed series is known, the models for the components can be derived through the canonical decomposition (Hillmer and Tiao 1982). This approach is followed in the routine TRAMO-SEATS, developed by Gómez and Maravall (1997).

3. DGP for Direct and Indirect Seasonally Adjusted Series

From what was said in the previous section it should be clear that a problem in choosing between direct and indirect methods is the evaluation of the size of the discrepancy. A natural solution can be obtained by use of some instruments to measure the differences between the DGP of Y_t^{ns} and the DGP of $(\lambda_x X_t^{ns} + \lambda_z Z_t^{ns})$. In a model-based framework the DGP's can be derived from the identification and estimation of an ARIMA model for each series considered.

The ARIMA model for the direct seasonally adjusted series, Y_t^{ns}, is immediately obtained. Details can be found in Box et al. (1978), Burman (1980), and Hillmer and Tiao (1982). In this framework it is supposed that the aggregated series Y_t is decomposable in a trend-cycle, a seasonal component and an irregular component; all these components are unobserved. If the following assumptions hold:

1. the unobserved components are uncorrelated;
2. the unobserved components follow ARIMA models;
3. the AR polynomials of the components do not have common roots;
4. the model for the observed series is known;

then it is possible to consistently identify the models for each component with the overall model for the observed series. There are infinite admissible decompositions, differing from each other by the amounts of noise in each component (a constant in the spectra of the components); the so-called canonical decomposition, in which the irregular part contains the maximum amount of noise, is chosen. In this way, the canonical decomposition produces an invertible ARIMA model for the seasonally adjusted series. For example, a typical seasonal model, such as the ARIMA$(0, 1, 1)(0, 1, 1)$, produces a seasonally adjusted series following an IMA$(2, 2)$ model.

The DGP of the indirect seasonally adjusted series can be obtained by summing up the ARIMA models relative to the seasonally adjusted sub-series (obtained from the canonical decomposition too). The state-space representation and the Kalman filters can be utilized to obtain the implicit model relative to the aggregated series. The techniques to derive it are summarized in the Appendix.

Note that this solution is valid both in the case of additive aggregation and additive decomposition and in the case of multiplicative aggregation and multiplicative decomposition. For the cases of additive aggregation-multiplicative decomposition and multiplicative aggregation-additive decomposition, we cannot derive the ARIMA model for the indirect seasonally adjusted series. In this case we suggest an approximation of the DGP, explained in the next section.

4. Comparing Direct and Indirect Seasonal Adjustment

Dealing with model-based methods for seasonal adjustment, we have shown that the DGP's can be represented by ARIMA models. More in general, it is supposed that the series Y_t is composed of the sum of a deterministic part and a stochastic part, such as:

$$Y_t = \boldsymbol{D}_t \beta^{(y)} + \varepsilon_t^{(y)}, \quad t = 1, \ldots, T \tag{2}$$

where $\boldsymbol{D}_t$ is a known row vector containing h regressors (generally representing outliers and calendar effects), $\beta^{(y)} = (\beta_1^{(y)}, \ldots, \beta_h^{(y)})'$ is a vector containing unknown coefficients and $\varepsilon_t^{(y)}$ is a disturbance which follows an ARIMA model.

The sub-series X_t and Z_t will follow similar models:

$$\begin{aligned} X_t &= \boldsymbol{D}_t \boldsymbol{\beta}^{(x)} + \varepsilon_t^{(x)} \\ Z_t &= \boldsymbol{D}_t \boldsymbol{\beta}^{(z)} + \varepsilon_t^{(z)} \end{aligned} \tag{3}$$

Note that we are considering the same regressors for the three models; if they are not fully equal, $\boldsymbol{D}_t$ will contain the union of the regressors, constraining to zero some coefficients.

The aggregation constraint

$$Y_t = \lambda_x X_t + \lambda_z Z_t$$

implies

$$\boldsymbol{D}_t \boldsymbol{\beta}^{(y)} + \varepsilon_t^{(y)} = \boldsymbol{D}_t \left(\lambda_x \boldsymbol{\beta}^{(x)} + \lambda_z \boldsymbol{\beta}^{(z)}\right) + \lambda_x \varepsilon_t^{(x)} + \lambda_z \varepsilon_t^{(z)} \tag{4}$$

As a matter of fact, the comparison between direct and indirect methods can be performed in two phases; first, we *linearize* the original series, subtracting the deterministic parts[4] and verifying their equality; then we compare the stochastic parts.

4.1. *Comparing the deterministic parts*

The first step is to verify if the deterministic effects identified on the aggregated series are equivalent to the sum of the deterministic effects identified on the sub-series. This property can be verified by a classical statistical test. Let us recall that the direct and indirect series are obtained independently, so that we can represent the equations in (2)–(3) in the following compact form:

$$\begin{bmatrix} Y_t \\ X_t \\ Z_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}_t & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D}_t & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{D}_t \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \varepsilon_t^{(y)} \\ \varepsilon_t^{(x)} \\ \varepsilon_t^{(z)} \end{bmatrix}$$

[4] To construct the final seasonally adjusted series, some regression effects (in particular, the calendar effects) will be assigned to the seasonal component, whereas the others (for example a constant term or a deterministic slope) may stay with the nonseasonal component.

where $\beta = (\beta^{(y)'}, \beta^{(x)'}, \beta^{(z)'})'$. Given the constraint of aggregation (4), the null hypothesis of equal deterministic effects can be verified as:

$$H_0: \boldsymbol{\beta}^{(y)} = \lambda_x \boldsymbol{\beta}^{(x)} + \lambda_z \boldsymbol{\beta}^{(z)}$$

This is a particular case of the well-known econometric problem of testing a set of linear restrictions:

$$H_0: \boldsymbol{R}\beta = \mathbf{0} \tag{5}$$

where $\boldsymbol{R} = [\boldsymbol{I}_h, -\lambda' \otimes \boldsymbol{I}_h]$, with $\boldsymbol{I}_h$ representing the $h \times h$ identity matrix, $\lambda = (\lambda_x, \lambda_z)'$ and $\otimes$ the Kronecker product. Let $\tilde{\boldsymbol{V}}^{(i)}$ $(i = y, x, z)$ be the covariance matrix of the estimator of $\beta^{(i)}$ (say $\tilde{\beta}^{(i)}$); for the independence assumption the covariance matrix of the full set of parameters $\tilde{\beta}$ will be $\tilde{\boldsymbol{V}} = diag\left(\tilde{\boldsymbol{V}}^{(y)}, \tilde{\boldsymbol{V}}^{(x)}, \tilde{\boldsymbol{V}}^{(z)}\right)$. Using the maximum likelihood estimator, $\boldsymbol{R}\tilde{\beta}$ is asymptotically Normal with mean $\boldsymbol{R}\beta$ and covariance matrix $\boldsymbol{RVR}'$, $\boldsymbol{V}$ being the true covariance matrix of β. An appropriate statistic to test the null hypothesis (5) is:

$$F = \left(\mathbf{R}\tilde{\beta}\right)'\left(\mathbf{R}\tilde{\mathbf{V}}\mathbf{R}'\right)^{-1}\left(\mathbf{R}\tilde{\beta}\right) \tag{6}$$

which follows a Chi-squared distribution with degrees of freedom equal to the number of constraints in (5) (h in this case).

If the null hypothesis is accepted, it is possible to consider only the stochastic parts to compare direct and indirect seasonal adjustments.

4.2. *Comparing the stochastic parts: a distance-based approach*

The comparison between direct and indirect linearized series can be achieved using a dissimilarity measure between ARIMA models. In order to obtain this measure, a useful tool is the AR metrics introduced by Piccolo (1989, 1990).

Let V_t be a zero-mean stochastic process and F the class of ARIMA invertible processes. It is well-known that if $V_t \in F$ then there exists a sequence of constants $\{\pi_i\}$ such that

$$\sum_{i=1}^{\infty} |\pi_i| < \infty$$

and

$$V_t = \sum_{i=1}^{\infty} \pi_i V_{t-1} + \varepsilon_t \tag{7}$$

where $\varepsilon_t \sim WN(0, \sigma^2)$.

Following Piccolo (1989, 1990) we define the *distance* between two processes V_{1t}, $V_{2t} \in F$ as

$$d(V_{1t}, V_{2t}) = \left[\sum_{i=1}^{\infty} (\pi_{1i} - \pi_{2i})^2\right]^{1/2} \tag{8}$$

To economize on notation we write d to indicate (8). In real applications, considering a convenient finite approximation of the $AR(\infty)$ representation and using suitable estimates of π_{ji} $(j = 1, 2)$, say $\tilde{\pi}_{ji}$, we obtain the distance estimator:

$$\tilde{d}_k = \left[\sum_{i=1}^{k} (\tilde{\pi}_{1i} - \tilde{\pi}_{2i})^2\right]^{1/2} \tag{9}$$

In practice a suitably large k is used (in our applications we will use $k = 200$).

Piccolo (1989) shows that if the two processes are independent, the asymptotic distribution of $\tilde{d}_k^2$ is a linear combination of independent Chi-squared variables. To test the null hypothesis:

$$H_0\colon d = 0 \tag{10}$$

we use the procedure explained by Corduas (1996), who approximates the distribution of $\tilde{d}_k^2$ with a single Chi-squared distribution.

Let us note that (10) is equivalent to testing:

$$\pi_1 = \pi_2 \tag{11}$$

where π_1 and π_2 are vectors containing the first k autoregressive coefficients of the representation (7) of the processes V_{1t} and V_{2t}, respectively.

Let us denote by $\tilde{\boldsymbol{\pi}}_1$ and $\tilde{\boldsymbol{\pi}}_2$ the coefficients obtained as functions of the maximum likelihood estimators of the parameters of the ARIMA models (say $\tilde{\theta}_1$ and $\tilde{\theta}_2$), for V_{1t} and V_{2t} respectively:

$$\tilde{\pi}_J = \pi\left(\tilde{\theta}_J\right), \quad J = 1, 2 \tag{12}$$

For the property of invariance they are the maximum likelihood estimates of π_1 and π_2. Thus $\tilde{\boldsymbol{\pi}}_1$ and $\tilde{\boldsymbol{\pi}}_2$ are asymptotically multivariate normally distributed with mean π_1 and π_2 and variance matrices $\tilde{\boldsymbol{\Sigma}}_1$ and $\tilde{\boldsymbol{\Sigma}}_2$, that can be obtained analytically. In particular:

$$\tilde{\boldsymbol{\Sigma}}_J = B_J \tilde{\boldsymbol{V}}_J B_J'$$

where $\tilde{\boldsymbol{V}}_1$ and $\tilde{\boldsymbol{V}}_2$ are the covariance matrices of the estimated ARIMA coefficients and $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$ are matrices containing the derivatives of the functions (12), for V_{1t} and V_{2t} respectively. Setting $\tilde{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Sigma}}_1 + \tilde{\boldsymbol{\Sigma}}_2$ and $\varsigma = \left(\tilde{\boldsymbol{\Sigma}}\right)^{-1/2}\left(\tilde{\pi}_1 - \tilde{\pi}_2\right)$, the statistic

$$\tilde{d}_k^2 = \varsigma' \tilde{\boldsymbol{\Sigma}} \varsigma \tag{13}$$

is a linear combination of r Chi-squared variables with 1 degree of freedom. r is the rank of $\tilde{\boldsymbol{\Sigma}}$ and the weights of the linear combination are equal to the positive eigenvalues of $\tilde{\boldsymbol{\Sigma}}$. Under the null hypothesis (11), this distribution can be approximated with $a\chi_c^2 + b$, where χ_c^2 is a Chi-squared random variable with c degrees of freedom (see also Mathai and Provost 1992) and, setting $t_i = trace\left(\tilde{\boldsymbol{\Sigma}}_i\right)$:

$$a = t_3/t_2, \quad b = t_1 - t_2^2/t_3, \quad c = t_2^3/t_3^2 \tag{14}$$

This approximation has a good performance, as shown in Corduas (1996).

We will now provide a result that will be utilized in the next section. Denoting by d_1 and d_2 the orders of integration of the processes V_{1t} and V_{2t}, respectively, we can show that if $d = 0$, then $d_1 = d_2$. In order to do this, we remember that the processes V_{1t} and V_{2t} can be represented in the following form:

$$\phi_1(B)(1 - B)^{d_1} V_{1t} = \vartheta_1(B)\varepsilon_{1t}, \quad \varepsilon_{1t} \sim WN\left(0, \sigma_1^2\right)$$

$$\phi_2(B)(1 - B)^{d_2} V_{2t} = \vartheta_2(B)\varepsilon_{2t}, \quad \varepsilon_{2t} \sim WN\left(0, \sigma_2^2\right)$$

where d_1 and d_2 are two integers, $\phi_1(z)$, $\vartheta_1(z)$, $\phi_2(z)$, *and* $\vartheta_2(z)$ are finite polynomials in z (z is a complex variable) of degrees p_1, q_1, p_2, and q_2 respectively, with $\phi_1(z) \neq 0$, $\vartheta_1(z) \neq 0$,

$\phi_2(z) \neq 0$ and $\vartheta_2(z) \neq 0$ for $|z| \leq 1$. Now, if $d = 0$, then $\pi_{1i}^{(v)} = \pi_{2i}^{(v)}$, $i = 1, 2, \ldots$. Since

$$1 - \pi_{11}^{(v)} z - \pi_{12}^{(v)} z^2 - \ldots = \frac{\phi_1(z)(1-z)^{d_1}}{\vartheta_1(z)}$$

and

$$1 - \pi_{21}^{(v)} z - \pi_{22}^{(v)} z^2 - \ldots = \frac{\phi_2(z)(1-z)^{d_2}}{\vartheta_2(z)}$$

we have that

$$\frac{\phi_1(z)(1-z)^{d_1}}{\vartheta_1(z)} = \frac{\phi_2(z)(1-z)^{d_2}}{\vartheta_2(z)}$$

that is

$$\phi_1(z) = \frac{\vartheta_1(z)\phi_2(z)}{\vartheta_2(z)} \frac{(1-z)^{d_2}}{(1-z)^{d_1}}$$

Let us assume that $d_1 \neq d_2$. In particular, without loss of generality, we can set $d_2 = d_1 + k$ (k integer). We have that

$$\phi_1(z) = \frac{\vartheta_1(z)\phi_2(z)}{\vartheta_2(z)} (1-z)^k$$

Thus $\phi_1(z)$ has k unit roots. This result is absurd since, by assumption, $\phi_1(z) \neq 0$, for $|z| \leq 1$; it follows that $d_1 = d_2$.

This theoretical result is confirmed in Table 1, in which the statistic (13), with $k = 200$, and the corresponding 95% critical value are reported for various lengths (T) of the series and various couples of $ARIMA(1, \delta, 1)$ processes,[5] different only on the order of the differences $\delta = 0$, 1. Note that the null H_0: $d = 0$ is always largely rejected.

The distance measure d can be used to compare the DGP of the direct seasonally adjusted series and the DGP of the indirect seasonally adjusted series.[6]

If this distance does not differ significantly from zero, the discrepancy between direct and indirect methods can be considered negligible.

The procedure that we propose follows these steps:

1. seasonally adjust the aggregated series and each sub-series with the model-based procedure; the former is the direct seasonally adjusted series;
2. sum the ARIMA models relative to the seasonally adjusted sub-series, obtaining the coefficients via the procedure described in the Appendix; they are the coefficients of the indirect seasonally adjusted model;
3. express the direct and indirect seasonally adjusted series in the AR form (7);
4. calculate the distance d_k between the two AR models and check if d_k^2 differs from zero with the test procedure explained in this section; if the null hypothesis is accepted the discrepancy between direct and indirect method can be considered negligible.

[5] Knowing the length of the series T, the covariance matrix can be calculated as a function of T and the ARMA coefficients.
[6] We note that the direct and indirect series are independent regarding construction, in the sense that they are obtained separately, not using any information about the other series.

Table 1. Squared distance and relative 95% critical value (cv) for various ARIMA models

ARIMA process	95% cv		$\tilde{d}^2_{200}$
$\phi_1 = \phi_2 = 0.3,\ \vartheta_1 = \vartheta_2 = 0$	0.192	$(T = 50)$	1.090
$d_1 = 0,\ d_2 = 1$	0.960	$(T = 100)$	
	0.064	$(T = 150)$	
$\phi_1 = \phi_2 = 0.6,\ \vartheta_1 = \vartheta_2 = 0$	0.135	$(T = 50)$	1.360
$d_1 = 0,\ d_2 = 1$	0.067	$(T = 100)$	
	0.045	$(T = 150)$	
$\phi_1 = \phi_2 = 0.9,\ \vartheta_1 = \vartheta_2 = 0$	0.040	$(T = 50)$	1.810
$d_1 = 0,\ d_2 = 1$	0.020	$(T = 100)$	
	0.013	$(T = 150)$	
$\phi_1 = \phi_2 = 0,\ \vartheta_1 = \vartheta_2 = 0.3$	0.161	$(T = 50)$	1.099
$d_1 = 0,\ d_2 = 1$	0.081	$(T = 100)$	
	0.054	$(T = 150)$	
$\phi_1 = \phi_2 = 0,\ \vartheta_1 = \vartheta_2 = 0.6$	0.279	$(T = 50)$	1.563
$d_1 = 0,\ d_2 = 1$	0.139	$(T = 100)$	
	0.093	$(T = 150)$	
$\phi_1 = \phi_2 = 0,\ \vartheta_1 = \vartheta_2 = 0.9$	3.863	$(T = 50)$	5.263
$d_1 = 0,\ d_2 = 1$	1.932	$(T = 100)$	
	1.288	$(T = 150)$	
$\phi_1 = \phi_2 = 0.3,\ \vartheta_1 = \vartheta_2 = 0.6$	0.287	$(T = 50)$	1.141
$d_1 = 0,\ d_2 = 1$	0.143	$(T = 100)$	
	0.096	$(T = 150)$	
$\phi_1 = \phi_2 = 0.6,\ \vartheta_1 = \vartheta_2 = 0.3$	0.285	$(T = 50)$	1.099
$d_1 = 0,\ d_2 = 1$	0.143	$(T = 100)$	
	0.095	$(T = 150)$	
$\phi_1 = \phi_2 = 0.9,\ \vartheta_1 = \vartheta_2 = 0.3$	0.282	$(T = 50)$	1.396
$d_1 = 0,\ d_2 = 1$	0.141	$(T = 100)$	
	0.094	$(T = 150)$	
$\phi_1 = \phi_2 = 0.3,\ \vartheta_1 = \vartheta_2 = 0.9$	2.157	$(T = 50)$	2.895
$d_1 = 0,\ d_2 = 1$	1.079	$(T = 100)$	
	0.719	$(T = 150)$	

As previously noted, this procedure is valid in the case of additive models. In the case of multiplicative models we cannot make explicit the DGP of the indirect seasonally adjusted series. In fact, we maintain the relationship:

$$Y_t = \lambda_x X_t + \lambda_z Z_t$$

but the multiplicative model implies:

$$Y_t = Y_t^{ns} Y_t^s$$

$$X_t = X_t^{ns} X_t^s$$

$$Z_t = Z_t^{ns} Z_t^s$$

The ARIMA models identified are referred to the logs of the series and components. Thus we cannot obtain the ARIMA model for the indirect seasonally adjusted series as the sum of two ARIMA models. In this case, we suggest directly estimating an ARIMA model on the indirect seasonally adjusted series and comparing this model with that of the direct seasonally adjusted series. In the application of Section 6 we will note that the introduction of this approximation does not imply differences in the final results. This suggests that we can approximate the DGP of the indirect seasonally adjusted series, estimating an ARIMA model in the case of an additive relationship too, bypassing the second step of the procedure. The computational advantage of this approximation is particularly significant when the number of sub-series is large.

4.3. Comparing the stochastic parts: a forecasting accuracy approach

Another criterion when it comes to evaluating the discrepancy between the DGP of the indirect seasonally adjusted series and the DGP of the direct seasonally adjusted one is based on the forecastability of the stationary part of the two series.[7]

Let us consider two processes V_{1_t} and V_{2_t} in F; they can be represented in the following form:

$$\phi_1(B)(1-B)^{d_1}V_{1_t} = \vartheta_1(B)\varepsilon_{1_t}, \quad \varepsilon_{1_t} \sim WN(0, \sigma_1^2)$$

$$\phi_2(B)(1-B)^{d_2}V_{2_t} = \vartheta_2(B)\varepsilon_{2_t}, \quad \varepsilon_{2_t} \sim WN(0, \sigma_2^2)$$

where d_1 and d_2 are two integers, $\phi_1(z)$, $\vartheta_1(z)$, $\phi_2(z)$, and $\vartheta_2(z)$ are finite polynomials in z of degrees p_1, q_1, p_2, and q_2 respectively, with $\phi_1(z) \neq 0$, $\vartheta_1(z) \neq 0$, $\phi_2(z) \neq 0$ and $\vartheta_2(z) \neq 0$ for $|z| \leq 1$. The processes $W_{1_t} = (1-B)^{d_1}V_{1_t}$ and $W_{2_t} = (1-B)^{d_2}V_{2_t}$ can be represented in the following form:

$$W_{1t} = \sum_{i=1}^{\infty} \pi_{1i}^{(w)} W_{1t-i} + \varepsilon_{1t}$$

$$W_{2t} = \sum_{i=1}^{\infty} \pi_{2i}^{(w)} W_{2t-i} + \varepsilon_{2t}$$

We note that

$$E(W_{1t} \mid W_{1t-1}, \ldots) = \sum_{i=1}^{\infty} \pi_{1i}^{(w)} W_{1t-i}$$

$$E(W_{2t} \mid W_{2t-1}, \ldots) = \sum_{i=1}^{\infty} \pi_{2i}^{(w)} W_{2t-i}$$

Thus

$$\hat{W}_{1t} = \sum_{i=1}^{\infty} \pi_{1i}^{(w)} W_{1t-i}$$

and

$$\hat{W}_{2t} = \sum_{i=1}^{\infty} \pi_{2i}^{(w)} W_{2t-i}$$

[7] We are clearly supposing again working with linearized series, given that the equality of the deterministic parts has been verified previously.

are the best mean squared predictors of W_{1t} and W_{2t}, respectively, and the innovations ε_{1t} and ε_{2t} are the correspondent forecast errors.

Now, in order to compare the forecastability of W_{1t} and W_{2t}, we can consider the following functions of the forecast errors

$$g(\varepsilon_{1t}) = \frac{\varepsilon_{1t}^2}{var(W_{1t})}$$

and

$$g(\varepsilon_{2t}) = \frac{\varepsilon_{2t}^2}{var(W_{2t})}$$

and define

$$f_t = g(\varepsilon_{2t}) - g(\varepsilon_{1t})$$

We note that

$$E(f_t) = \frac{var(\varepsilon_{2t})}{var(W_{2t})} - \frac{var(\varepsilon_{1t})}{var(W_{1t})} \tag{15}$$

The quantity $R_1^2 = 1 - var(\varepsilon_{1t})/var(W_{1t})$ is a reasonable measure of the forecastability of W_{1t} relative to the information set $\{W_{1t-1}, \ldots\}$. In particular, R_1^2 is 1 when the variance of the forecast error is 0; R_1^2 is 0 when W_{1t} is a white noise (not forecastable); the same considerations hold for $R_2^2 = 1 - var(\varepsilon_{2t})/var(W_{2t})$. Thus the null hypothesis of equal forecastability for the two processes can be formulated as follows:

$$H_0\colon R_1^2 - R_2^2 = 0$$

Hence, the equal forecastability null hypothesis is equivalent to the null hypothesis that the population mean of f_t is 0. Since f_t is a stationary and short memory process, we can verify H_0: $E(f_t) = 0$ using the Diebold-Mariano (1995) test. The statistic is:

$$S_1 = \sqrt{T}\frac{\bar{f}}{\sqrt{\gamma_0}} \tag{16}$$

where $\bar{f} = T^{-1}\sum_{t=1}^{T} f_t$ and $\gamma_0 = T^{-1}\sum_{t=1}^{T}(f_t - \bar{f})^2$. Under the null hypothesis, S_1 is asymptotically normally distributed with unit variance (see Harvey et al. 1997).

When the seasonally adjusted series are stationary or integrated of the same order, this test can be used to compare the direct and indirect seasonally adjusted series following the criterion of the forecast performance. Instead, it is likely that in cases where the stochastic nonstationary trends are different, the proposed measure may accept the null hypothesis H_0: $E(f_t) = 0$, even if the direct and indirect adjustments will give very different results. However, in these cases we can continue to use this procedure to make inferences regarding the distance between direct and indirect seasonally adjusted series. In fact, we can show that the equal forecastability condition for the two stationary processes, W_{1t} and W_{2t}, is necessary for a null distance between V_{1t} and V_{2t}.

We have proved in the previous section that if $d(V_{1t}, V_{2t}) = 0$, the following holds:

$$\frac{\phi_1(z)(1-z)^{d_1}}{\vartheta_1(z)} = \frac{\phi_2(z)(1-z)^{d_2}}{\vartheta_2(z)}$$

with $d_1 = d_2$; thus:

$$\frac{\vartheta_1(z)}{\phi_1(z)} = \frac{\vartheta_2(z)}{\phi_2(z)}$$

On the other hand, we note that by developing the variances in the denominators of (15) we obtain

$$E(f_t) = (1 + \psi_{21}^2 + \psi_{22}^2 + \ldots)^{-1} - (1 + \psi_{11}^2 + \psi_{12}^2 + \ldots)^{-1}$$

where the coefficients $\{\psi_{1i}\}$ and $\{\psi_{2i}\}$ are determined by the relationships

$$\sum_{i=1}^{\infty} \psi_{1i} z^i = \frac{\vartheta_1(z)}{\phi_1(z)}$$

$$\sum_{i=1}^{\infty} \psi_{2i} z^i = \frac{\vartheta_2(z)}{\phi_2(z)}$$

respectively. Hence we have

$$1 + \psi_{11} z + \psi_{12} z^2 + \ldots = 1 + \psi_{21} z + \psi_{22} z^2 + \ldots$$

Thus we can conclude that $R_1^2 - R_2^2 = 0$. The condition of equal predictability for the stationary processes W_{1t} and W_{2t}, $R_1^2 - R_2^2 = 0$, is necessary for $d(V_{1t}, V_{2t}) = 0$. The importance of this result consists in the fact that it can be utilized to make inferences regarding $d(V_{1t}, V_{2t})$. In particular, if the null hypothesis H_0: $E(f_t) = 0$ is rejected we can also reject the hypothesis H_0': $d(V_{1t}, V_{2t}) = 0$. However, it is important to remember that if we accept H_0: $E(f_t) = 0$, we cannot accept H_0': $d(V_{1t}, V_{2t}) = 0$. The condition $E(f_t) = 0$ is only necessary for $d(V_{1t}, V_{2t}) = 0$.

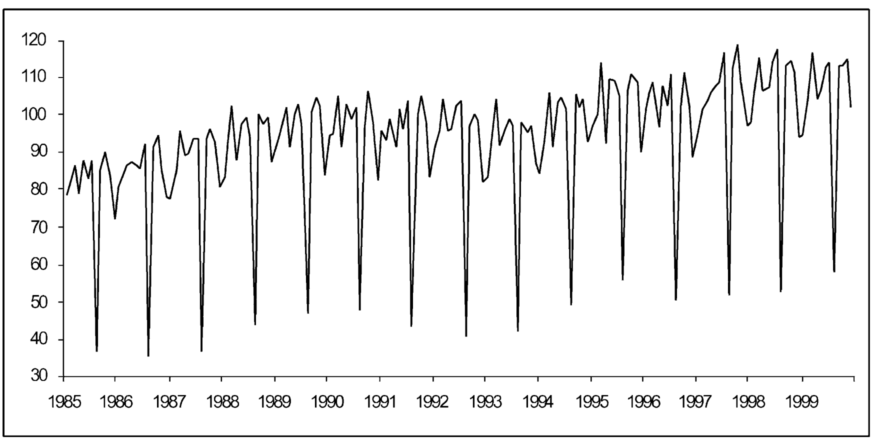

Fig. 1. General Industrial Production Index (Istat source)

Table 2. Weights of the sub-series to obtain $IPI^{(0)}$

Years	$IPI^{(1)}$	$IPI^{(2)}$	$IPI^{(3)}$
1985–1989	0.267	0.177	0.556
1990–1994	0.257	0.158	0.585
1995–1999	0.232	0.165	0.603

5. An Example: The Italian Industrial Production Index

Let us consider the monthly series of the general Italian Industrial Production Index ($IPI^{(0)}$) from January 1985 to December 1999, plotted in Figure 1. This series, which presents a clear seasonal behavior, is obtained by weighted aggregation of the indices relative to consumer ($IPI^{(1)}$), investment ($IPI^{(2)}$) and intermediate goods ($IPI^{(3)}$). The weights change with the different bases used to construct the indices and are reported in Table 2.

In order to apply our procedure, we have estimated the following model for each series:

$$IPI_t^{(i)} = \beta_{TD}^{(i)} TD_t + \beta_{LY}^{(i)} LY_t + \beta_{EE}^{(i)} EE_t + \beta_H^{(i)} H_t + \varepsilon_t^{(i)}, i = 0, 1, 2, 3 \tag{17}$$

where $\varepsilon_t^{(i)}$ is a disturbance which follows an ARIMA(0, 1, 1)(0, 1, 1) model,

$$\varepsilon_t^{(i)} = \frac{(1 + \vartheta_1^{(i)} B)(1 + \vartheta_{12}^{(i)} B^{12}) w_t^{(i)}}{(1 - B)(1 - B^{12})}, \quad w_t^{(i)} \sim IIN(0, \sigma_{(i)}^2)$$

TD_t is a regressor which represents the trading days effect, LY_t is the leap year effect at time t, EE_t is the Easter Effect and H_t is the Holidays effect. The regressors are obtained as:

$TD_t =$ number of (Mon, Tue, Wed, Thu, Fri)-#(Sat, Sun)5/2 in the month t;

$$LY_t = \begin{cases} 0.75 & \text{if } t \text{ is referred to a February in a leap year} \\ -0.25 & \text{if } t \text{ is referred to a February in a nonleap year} \\ 0 & \text{otherwise} \end{cases}$$

$$EE_t = \frac{j}{6} - \frac{1}{2}$$

where j is the number of days of the month t that are in the temporal interval:

[(*date of Easter*) – (6 *days*), *date of Easter*]

$H_t =$ number of national holidays, not coincident with Saturday and Sunday, that lies in the month t.

Using the routine TRAMO-SEATS, we obtain the canonical decomposition of the model in a trend-cycle, a seasonal component and an irregular part.

In this case, the deterministic effects identified with regard to the aggregated series are equivalent to the sum of the deterministic effects identified on the sub-series. In fact we have applied test (6) to our series (the statistic follows a Chi-squared distribution with four degrees of freedom), verifying separately the three constraints of Table 2; note that the three rows are very similar, so that we cannot verify simultaneously the twelve constraints, because a singular matrix problem in (6) would arise. Each p-value obtained is

Table 3. Estimates of the parameters of models (17) (*standard errors in parentheses*)

	ϑ_1	ϑ_{12}	β_{TD}	β_{LY}	β_{EE}	β_H	σ^2
$IPI^{(0)}$	−0.484	−0.660	0.805	3.036	−1.768	−2.118	3.664
	(0.070)	(0.072)	(0.036)	(0.895)	(0.525)	(0.281)	(0.023)
$IPI^{(1)}$	−0.598	−0.543	0.951	2.507	−1.937	−2.487	7.843
	(0.063)	(0.075)	(0.048)	(1.112)	(0.663)	(0.359)	(0.736)
$IPI^{(2)}$	−0.539	−0.679	1.072	3.907	−1.791	−2.662	12.562
	(0.069)	(0.068)	(0.069)	(1.674)	(0.994)	(0.527)	(1.366)
$IPI^{(3)}$	−0.387	−0.664	0.669	2.851	−1.750	−1.813	3.131
	(0.073)	(0.075)	(0.033)	(0.830)	(0.478)	(0.259)	(0.220)

above 0.999, largely accepting the null hypothesis of equality of the deterministic effects in the direct and indirect seasonally adjusted series. Thus the presence of deterministic effects does not affect our analysis.

The estimated ARIMA parameters for the three series and the other estimated coefficients are reported in Table 3. They produce the following models for the seasonally adjusted series (D and I indicate direct and indirect, respectively):

$$\begin{aligned}
(1-B)^2 SA_t^{(0)D} &= (1-1.4550B+0.4724B^2)v_t^{(0)D}, & \sigma^2_{(0)D} &= 2.482 \\
(1-B)^2 SA_t^{(1)} &= (1-1.5575B+0.5770B^2)v_t^{(1)}, & \sigma^2_{(1)} &= 3.439 \\
(1-B)^2 SA_t^{(2)} &= (1-1.5119B+0.5264B^2)v_t^{(2)}, & \sigma^2_{(2)} &= 8.462 \\
(1-B)^2 SA_t^{(3)} &= (1-1.3597B+0.3801B^2)v_t^{(3)}, & \sigma^2_{(3)} &= 2.278
\end{aligned} \tag{18}$$

where $SA^{(i)}$ ($i=0$, 1, 2, 3) is the seasonally adjusted series for $IPI^{(i)}$ and $v^{(i)}$ is a white noise.

Summing the stationary parts of the processes relative to $SA^{(1)}$, $SA^{(2)}$, $SA^{(3)}$, weighted following Table 2 (their innovation variances are 0.159, 0.187 and 0.608, respectively), we obtain the model for the indirect seasonally adjusted series:

$$(1-B)^2 SA_t^{(0)I} = (1-1.4185B+0.4377B^2)v_t^{(0)I} \tag{19}$$

The distance between the direct and indirect seasonally adjusted series, using 200 AR coefficients, is 0.603. To test the null hypothesis of distance zero we use the statistic (13), obtaining $\tilde{d}_k^2$ =0.363. The parameters (14) to calculate the approximate distribution are:

$$a = 18.029, \quad b = 0.721, \quad c = 1.006$$

and the 95% critical value is equal to 70.223; hence the null hypothesis is largely accepted. Estimating directly an IMA(2,2) model on the indirect seasonally adjusted series, we obtain the following model:

$$(1-B)^2 SA_t^{(0)I^*} = (1-1.4381B+0.4851B^2)v_t^{(0)I^*} \tag{20}$$

The distance between (19) and (20) is equal to 4.117, that is not significantly different from zero, because the 95% critical value of (13) is 42.396 ($\tilde{d}_k^2 = 16.952$, $a=10.492$, $b=1.137$, and $c=1.037$). In the same way, the distance between (18) and (20) is negligible, being $\tilde{d}_k^2$ =12.642 with a 95% critical value equal to 35.550 ($a=8.763$, $b=1.007$, and $c=1.041$).

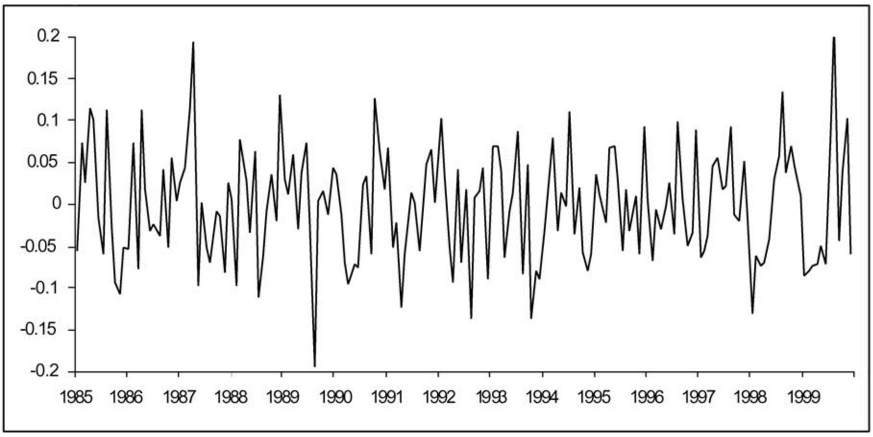

Fig. 2. Discrepancy between the direct and indirect seasonal adjusted series

Now, we expect a coherent result with the forecasting accuracy criterion. In this case, we compare the forecastability of the series $(1-B)^2 SA_t^{(0)D}$ and $(1-B)^2 SA_T^{(0)I}$. Using the forecast errors function described in Section 5, we obtain the S_1 statistic (16) equal to 0.1526, accepting the null hypothesis of equal forecasting accuracy of the direct and indirect seasonally adjusted models.

The inferential results are confirmed by the graphics comparison of Figure 2, in which the discrepancy between the direct and indirect seasonally adjusted series is plotted. We note that the maximum difference is less then 0.22. Note that, excluding the presence of different deterministic effects between the two series and of spurious seasonality in the indirect seasonally adjusted series, their dynamics is random. In Figure 3 the discrepancy of month-to-month percentage variations is plotted, confirming the similarity of the two series.

Finally, we want to stress the fact that, following other criteria based only on an aspect of the characteristics of the seasonally adjusted series, we can achieve different results. For example, if the criterion chosen is the smoothness, we could calculate the

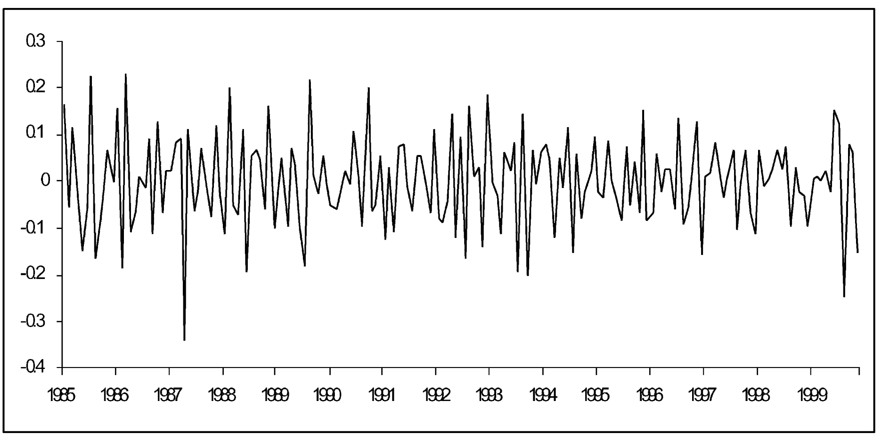

Fig. 3. Discrepancy between direct and indirect month-to-month variations

Table 4. Statistics on direct and indirect methods

Criteria	Direct	Indirect
Lack of smoothness	2.407	2.340
Final error variance	0.087	0.092
Revision error variance	0.087	0.092
Total error variance	0.174	0.184

roughness measure proposed by Dagum (1979) for both the direct and indirect seasonally adjusted series:

$$\sum_{t=2}^{T} (SA_t^{(0)i} - SA_{t-1}^{(0)i})^2, \quad i = D, I$$

This measure, expressed as an average, is 2.407 for the direct series and 2.340 for the indirect one, so that the direct shows a larger lack of smoothness (first row of Table 4).

On the other hand, if we are interested in the estimation errors and revisions, we can apply the criterion proposed by Campolongo and Planas (2000), obtaining the variances of the final, revision and total errors displayed in Table 4 (expressed in units of variance of the innovation of IPI^0). In this case the direct approach seems more precise.[8]

6. Final Remarks

Many of the economic series to be seasonally adjusted by statistical agencies are obtained as aggregations of a certain number of components. In this situation, the issue of adjusting the series directly or indirectly immediately arises. If the target of the National Statistical Institutes is to seasonally adjust the aggregated and the component series, a problem of consistency can arise too: the seasonally adjusted series obtained directly from the composed series is not equal to the sum of the seasonally adjusted components. A large difference between the direct seasonally adjusted series and the sum of the seasonally adjusted sub-series can cause confusion among the data users.

In this article, we have utilized two approaches to evaluate the relevance of this discrepancy. The first one is based on the Piccolo's distance between ARIMA models using the test procedure described in Corduas (1996), in which is also studied the capability of this approach to compare couples of series with Monte Carlo experiments.

The second approach proposed in this article concerns a comparison of the forecastability of the direct and indirect series, applying a Diebold and Mariano (1995) style test. Obviously, the functions of forecast errors utilized can be substituted by other functions; our choice is justified by the interpretation that is consistent with the seasonal adjustment operation and by the relationship with the Piccolo distance. The Diebold and Mariano test needs simple calculations and it does not imply the transformation of the seasonally

[8] An example where our procedure rejects the hypothesis of equality of direct and indirect seasonally adjusted series is represented by the quarterly (I 1993–IV 2001) Italian labor forces, obtained as the sum of male and female labor forces. The three series do not present deterministic effects; the series of males follows an additive ARIMA $(1,0,0)(0,1,0)$ model, whereas the female and total series follow an ARIMA$(0,1,1)(0,1,1)$ model. The statistic (16) is equal to -3.37, rejecting the null hypothesis; as a consequence, also the distance-based approach will reject the hypothesis of distance zero, as demonstrated in Section 4.3. Details of this application are available on request.

adjusted models in AR models, as the other approach does. On the other hand, it can only be applied to stationary series, whereas the Piccolo's distance is also used with integrated processes. In both cases, to compare only the stochastic parts, a test of equal deterministic effects has to be performed first.

The approaches are correctly applied only in the case of additive models with additive aggregations or multiplicative models with multiplicative aggregations. Their application to the cases of multiplicative models and additive aggregations or additive models and multiplicative aggregations requires the DGP of the indirect seasonally adjusted series to be approximated by estimating an ARIMA model directly rather than deriving it from the sum of the ARIMA models of the sub-series. Finally, we recall that the nature of our approach is inferential, being based on classical tests. As for the usefulness of these tools, we think that they can be applied to support the choice of diffusion policy in respect of seasonally adjusted data.

Appendix

For the sake of simplicity, we always consider the simple case (1), in which the aggregated series is the sum of 2 sub-series, but the results are easily extended by summing up more series. For example, let us consider two MA processes of order q_1 and q_2, respectively:

$$X_t = \zeta_t + \vartheta_1 \zeta_{t-1} + \ldots + \vartheta_{q_1} \zeta_{t-q_1}$$
$$Z_t = \eta_t + \theta_1 \eta_{t-1} + \ldots + \theta_{q_2} \eta_{t-q_2}$$

where ζ_t and η_t are white noises uncorrelated at all lags and leads. We recall that, in the indirect method, the sub-series are individually adjusted and then aggregated, without taking into account the correlation among them.

The process $Y_t = X_t + Z_t$ follows an MA($\max\{q_1, q_2\}$) model (for sake of simplicity we omit the weights λ_x and λ_z). In order to obtain the coefficients of this last model we express the process Y_t in the following state-space form:

$$\text{Observation equation:} \quad Y_t = \boldsymbol{a}\boldsymbol{\xi}_t$$
$$\text{State equation:} \quad \boldsymbol{\xi}_{t+1} = \boldsymbol{B}\boldsymbol{\xi}_t + \boldsymbol{e}_{t+1}$$

where:

$$\boldsymbol{a} = [1 \quad \vartheta_1 \quad \ldots \quad \vartheta_{q_1} \quad 1 \quad \theta_1 \quad \ldots \quad \theta_{q_2}]$$

$$\boldsymbol{B} = \begin{bmatrix} \underset{(1\times q_1)}{\mathbf{0}} & 0 & \underset{(1\times q_2)}{\mathbf{0}} & 0 \\ \mathbf{I}_{q_1} & \underset{(q_1\times 1)}{\mathbf{0}} & \underset{(q_1\times q_2)}{\mathbf{0}} & \underset{(q_1\times 1)}{\mathbf{0}} \\ \underset{(1\times q_1)}{\mathbf{0}} & 0 & \underset{(1\times q_2)}{\mathbf{0}} & 0 \\ \underset{(q_2\times q_1)}{\mathbf{0}} & \underset{(q_2\times 1)}{\mathbf{0}} & \mathbf{I}_{q_2} & \underset{(q_2\times 1)}{\mathbf{0}} \end{bmatrix}$$

$$\boldsymbol{\xi}_t' = [\zeta_t \quad \zeta_{t-1} \quad \ldots \quad \zeta_{t-q_1} \quad \eta_t \quad \eta_{t-1} \quad \ldots \quad \eta_{t-q_2}]$$
$$\boldsymbol{e}_t' = [\zeta_t \quad 0 \quad \ldots \quad 0 \quad \eta_t \quad 0 \quad \ldots \quad 0]$$

and $\underset{(h\times k)}{\mathbf{0}}$ is an $h \times k$ matrix with all the elements equal to zero, whereas I_k is the identity

$k \times k$ matrix. Denoting by $\boldsymbol{K}$ the steady-state Kalman gain, defined as:

$$\boldsymbol{K} = \boldsymbol{BPa}'(\boldsymbol{aPa}')^{-1} \tag{21}$$

where $\boldsymbol{P}$ is the steady-state MSE matrix of the state vector ξ_t, it can be demonstrated that the coefficients of the MA process of Y_t are (see Hamilton, 1994, Chapter 13):

$$\delta_j = \boldsymbol{aB}^{j-1}\boldsymbol{K}, \quad j = 1, \ldots, \max(q_1, q_2) \tag{22}$$

The (21) does not imply burdensome calculations; in fact the steady-state MSE matrix is expressed as:

$$P = \lim_{t \to \infty} P_{t|t-1}$$

where $\{\boldsymbol{P}_{t|t-1}\}$ is the sequence of the variance matrices calculated in each step of the Kalman filter. In particular, this sequence can be calculated by:

$$\boldsymbol{P}_{t+1|t} = \boldsymbol{B}[\boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{a}'(\boldsymbol{aP}_{t|t-1}\boldsymbol{B})^{-1}\boldsymbol{aP}_{t|t-1}]\boldsymbol{B}' + \boldsymbol{Q} \tag{23}$$

where $\boldsymbol{Q}$ is the variance matrix of $\boldsymbol{e}_t$ (invariant with t). If $\boldsymbol{B}$ is a $k \times k$ matrix whose eigenvalues are all inside the unit circle and $\boldsymbol{P}_{1|0}$ is the initializing matrix of the sequence, satisfying:

$$vec(\boldsymbol{P}_{1|0}) = [\mathbf{I}_{k^2} - (\boldsymbol{B} \otimes \boldsymbol{B})]^{-1} vec(\boldsymbol{Q})$$

it can be demonstrated that $\{\boldsymbol{P}_{t|t-1}\}$ is a monotonically nonincreasing sequence that converges to:

$$\boldsymbol{P} = \boldsymbol{B}[\boldsymbol{P} - \boldsymbol{Pa}'(\boldsymbol{aPB})^{-1}\boldsymbol{aP}]\boldsymbol{B}' + \boldsymbol{Q}$$

In addition, if $\boldsymbol{Q}$ is strictly positive definite, the convergence is unique for any positive semidefinite symmetric matrix $\boldsymbol{P}_{1|0}$. In other terms, iterating (23) we can obtain $\boldsymbol{P}$, that provides the calculation of (21) and (22).

Hence, if the sub-series follow MA processes, we are able to make explicit the model representing the DGP of the indirect seasonally adjusted series.

Now let us suppose that X_t and Z_t are two AR processes of order p_1 and p_2 respectively:

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_{p_1} X_{t-p_1} + \zeta_t$$

$$Z_t = \phi_1 Z_{t-1} + \ldots + \phi_{p_2} Z_{t-p_2} + \eta_t$$

with ζ_t and η_t uncorrelated at each lead and lag.

In this case, the process $Y_t = X_t + Z_t$ follows an ARMA$(p_1 + p_2, \max\{p_1, p_2\})$ model:

$$\gamma(L)Y_t = \delta(L)\varepsilon_t$$

where $\gamma(L) = \varphi(L)\phi(L)$ and $\delta(L)\varepsilon_t = \phi(L)\zeta_t + \varphi(L)\eta_t$. The coefficients of the polynomial $\delta(L)$ are obtained by (22).

Finally, we remark that if X_t is an ARMA(p_1, q_1) process and Z_t an ARMA(p_2, q_2) process, then $Y_t = X_t + Z_t$ will follow an ARMA(p, q) model, with $p \leq p_1 + p_2$ and $q \leq \max\{p_1 + q_1, p_2 + q_2\}$, with coefficients deriving from the same rules previously described.

7. References

Box, G.E.P., Hillmer, S.C., and Tiao, G.C. (1978). Analysis and Modeling of Seasonal Time Series. In Seasonal Analysis of Time Series (A. Zellner ed.). Washington, D.C.: U.S. Department of Commerce, U.S. Census Bureau, 309–334.

Brockwell, P.J. and Davis, R.A. (1991). Time Series: Theory and Methods. New York: Springer-Verlag.

Burman, J.P. (1980). Seasonal Adjustment by Signal Extraction. Journal of the Royal Statistical Society, Ser. A, 143, 321–337.

Corduas, M. (1996). Uno Studio sulla Distribuzione Asintotica della Metrica Autoregressiva, Statistica, LVI, 3, 321–332.

Corduas, M. (2000). La Metrica Autoregressiva tra Modelli ARIMA: una procedura in linguaggio GAUSS, Quaderni di Statistica, 2, 1–37.

Dagum, E.B. (1979). On the Seasonal Adjustment of Economic Time Series Aggregates: A Case Study of the Unemployment Rate. In Counting the Labor Force, National Commission on Employment and Unemployment Statistics, Appendix, 2, 317–344.

Den Butter, F.A. and Fase, M.M. (1991). Seasonal Adjustment as a Practical Problem. North-Holland: Amsterdam.

Depoutot, R. and Planas, C. (1998). Comparing Seasonal Adjustment and Trend Extraction Filters with Application to a Model-based Selection of X11 Linear Filters. Eurostat Working Paper 9/1998/A/9, Luxembourg.

Diebold, F.X. and Mariano, R.S. (1995). Comparing Predictive Accuracy. Journal of Business and Economic Statistics, 13, 253–263.

Findley, D., Monsell, B., Bell, W., Otto, M., and Chen, B.C. (1998). New Capabilities and Methods of the X12-ARIMA Seasonal Adjustment Program. Journal of Business and Economic Statistics, 16, 127–152.

Geweke, J. (1978). The Temporal and Sectorial Aggregation of Seasonally Adjusted Time Series. In Seasonal Analysis of Economic Time Series (A. Zellner ed.). Washington, DC: U.S. Department of Commerce, U.S. Census Bureau, 411–427.

Ghysels, E. (1997). Seasonal Adjustment and Other Data Transformations. Journal of Business and Economic Statistics, 15, 410–418.

Gómez, V. (2000). Revision-based Test for Direct Versus Indirect Seasonal Adjustment of Aggregated Series. Doc. Eurostat/A4/SA/00/08.

Gómez, V. and Maravall, A. (1997). Programs TRAMO and SEATS: Instructions for the User (Beta Version: June 1997). Working Paper N. 97001, Dirección General de Análisis y P.P., Ministry of Economy and Finance.

Hamilton, J.D. (1994). Time Series Analysis. Princeton: University Press.

Harvey, D.I., Leybourne, J., and Newbold P. (1997). Testing the Equality of Prediction Mean Squared Errors. International Journal of Forecasting, 13, 281–291.

Hillmer, S.C. and Tiao, G.C. (1982). An ARIMA-model-based Approach to Seasonal Adjustment. Journal of the American Statistical Association, 77, 63–70.

Lothian, J. and Morry, M. (1977). The Problem of Aggregation: Direct or Indirect. Statistics Canada.

Mathai, A.M. and Provost, S.B. (1992). Quadratic Forms in Random Variables. New York: Marcell Dekker, Inc.

Piccolo, D. (1989). On the Measure of Dissimilarity Between ARIMA Models. Proceedings of the American Statistical Association, Section of Business and Economic Statistics, 231–236.

Piccolo, D. (1990). A Distance Measure for Classifying ARIMA Models. Journal of Time Series Analysis, 11, 153–164.

Planas, C. and Campolongo, F. (2000). The Seasonal Adjustment of Contemporaneously Aggregated Series. Doc. Eurostat/A4/SA/00/06.

Received June 2001
Revised March 2002