

9. ESTIMATION OF SEASONAL FACTORS

9.1. Introduction

Economic time series are generally found to exhibit regular, intra-year seasonal movements around its annual trend path. Such repetitive seasonal variations can result from climatic conditions, production cycle characteristics, seasonal nature of economic activity, festivals, vacation practices, etc. While the seasonal variations occur regularly, yet they may vary in magnitude from year to year. From the policy perspective, information on seasonal factors of an economic variable is useful as it enables the policy maker to differentiate between the seasonal changes and long-run changes in a variable and thereby design appropriate policy responses. An article regarding monthly seasonal factors for selected economic and financial time series of the Indian economy is being regularly published in the RBI Bulletin from 1980 onwards. This article presents the monthly seasonal factors of selected economic/financial time series classified into major five groups, namely,

- A. Monetary and Banking Indicators;
- B. Wholesale Price Index (WPI);
- C. Consumer Price Index for Industrial workers (CPI-IW);
- D. Index of Industrial Production (IIP);
- E. External Trade.

The seasonal factors are being estimated using the X-12 auto-regressive integrated moving average (ARIMA) methodology. X-12 ARIMA is the software for seasonal adjustment developed by the US Census Bureau. It is an enhanced version

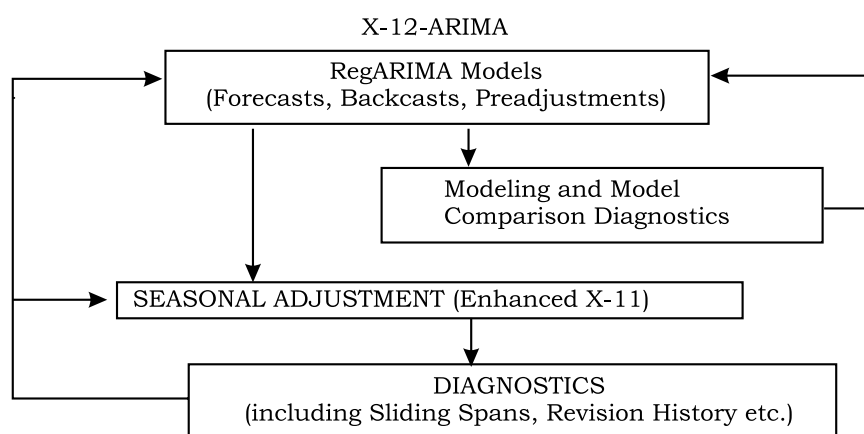
of the X-11 Variant of the Census Method II seasonal adjustment programme. The procedure makes multiplicative/additive adjustments of a time series and creates an output data set containing the adjusted time series and intermediate calculations.

The main source of these new tools is the extensive set of time series model building facilities built into the programme for fitting the regARIMA models. These are regression models with ARIMA (Autoregressive Integrated Moving Average) errors. More precisely, they are models in which the mean function of the time series (or its lags) is described by a linear combination of regressors, and the covariance structure of the series is that of an ARIMA process. If no regressors are used, indicating that the mean is assumed to be zero, the regARIMA model reduces to an ARIMA model. There are regressors for modeling certain kinds of disruption in the series, or sudden changes in level, whose effects need to be removed from the data before the methodology can adequately estimate seasonal adjustments. The regARIMA-modeling module of X-12-ARIMA was adopted from the regARIMA programme developed by the Time Series Staff of Census Bureau's Statistical Research Division.

9.2. Methodology

The X-12-ARIMA is developed following the operation-flow diagram of Figure 9.1. This posits a regARIMA (linear regression model with ARIMA time series errors) modeling subprogram that can

Figure 9.1: Flow Diagram for Seasonal Adjustment with X-12-ARIMA

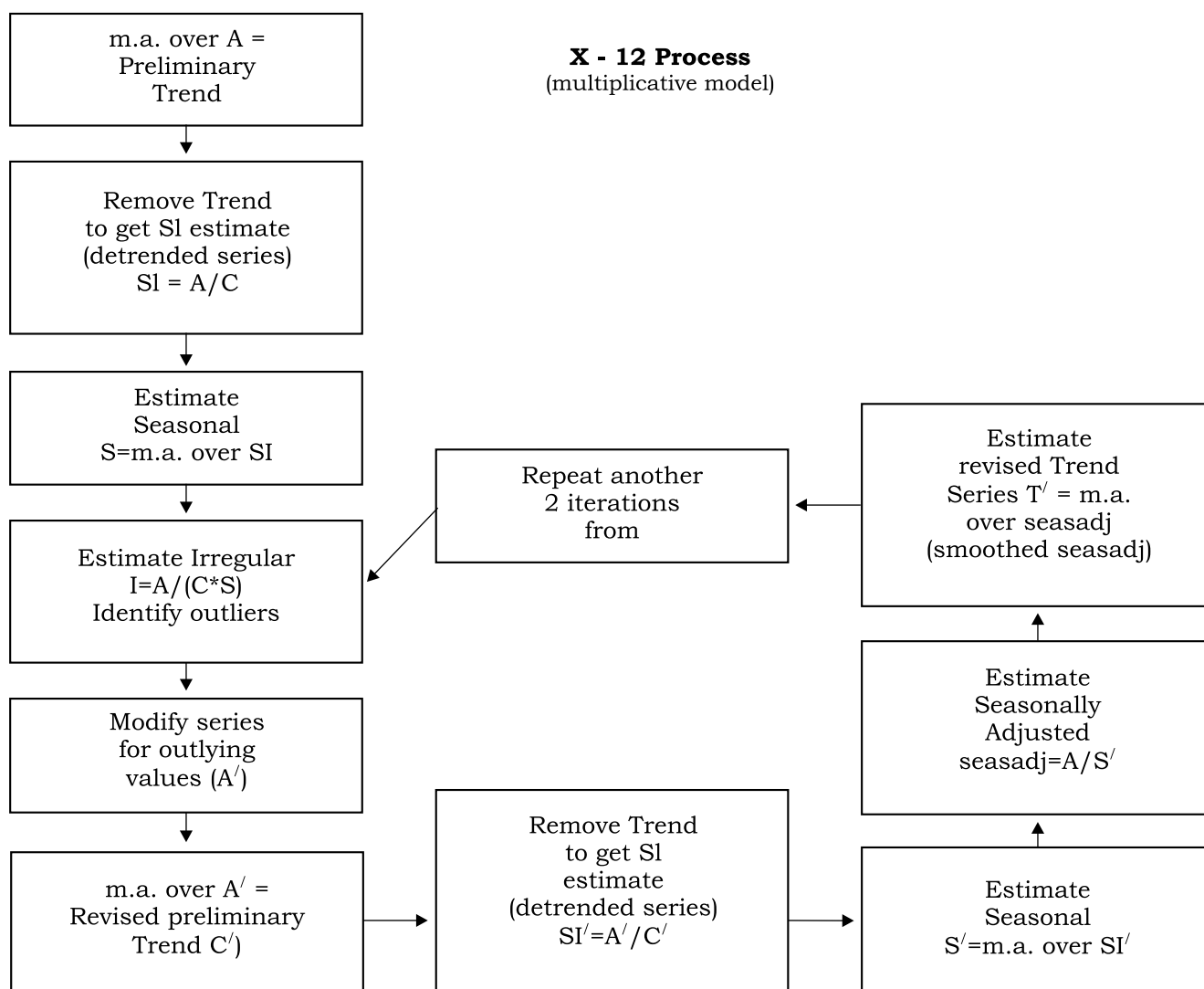


provide forecasts, backcasts, and prior adjustments for various effects before the seasonal adjustment subprogram in the central box is invoked. The final box in Figure 9.1 represents a set of post-adjustment diagnostic routines that can be used to obtain indicators of the effectiveness of both the modeling and the seasonal adjustment options chosen. The modeling module of X-12-ARIMA is designed for regARIMA model building with seasonal economic time series. To this end, several categories of predefined regression variables are available in X-12-ARIMA (for further details user may refer to the manual available at <http://www.census.gov/srd/www/x12a/>). User-defined regression variables can also

be easily read in and included in models. For a multiplicative model, Figure 9.2 describes X-12 process for computing seasonal factors.

X-12-ARIMA uses the standard $(p d q) (P D Q)_s$ notation for seasonal ARIMA models. The $(p d q)$ refers to the orders of the nonseasonal autoregressive (AR), differencing, and moving average (MA) operators, respectively. The $(P D Q)_s$ refers to the seasonal autoregressive, differencing, and moving average orders. The s subscript denotes the seasonal period, e.g., $s = 12$ for monthly data, or 4 for quarterly data. Great flexibility is allowed in the specification of the ARIMA structures: any number of AR,

Figure 9.2: Flow Diagram for Seasonal Adjustment X-12 Process in Multiplicative models



MA, and differencing operators may be used; missing lags are allowed in AR and MA operators; and AR and MA parameters can be fixed at user-specified values. The specification of a regARIMA model requires specification both of regression variables to be included in the model and also the type of ARIMA model for regression errors (i.e., the orders $(p\ d\ q)$ $(P\ D\ Q)_s$). Specification of the regression variables depends on user knowledge about the series being modelled.

For the user who wishes to fit customised time series models, X-12-ARIMA provides capabilities for the three modeling stages of *identification*, *estimation*, and *diagnostic checking*. *Identification* of the ARIMA model for the regression errors follows well-established procedures based on examination of various sample autocorrelation and partial autocorrelation functions produced by X-12-ARIMA. Once a regARIMA model has been specified, X-12-ARIMA will *estimate* its parameters by maximum likelihood using an iterative generalised least squares (IGLS) algorithm. *Diagnostic checking* involves estimation of residuals from the fitted model for signs of model inadequacy.

The major improvements in X-12 ARIMA are new modeling capabilities using regARIMA models, sliding spans diagnostic procedure, ability to produce revisions history of a given seasonal adjustment, several new outlier detection options for the irregular component of the seasonal adjustment, etc. RegARIMA is a linear regression model with ARIMA that can provide forecasts and prior adjustments for various effects. X-12 ARIMA uses regARIMA models to preadjust a series by removing effects such as outliers, etc., before the seasonal adjustment program is invoked.

9.3. RegARIMA

Let B denote the backshift operator, $By_t = y_{t-1}$. X-12-ARIMA can estimate regARIMA models of order $(p\ d\ q)(P\ D\ Q)_s$ for y_t . These are models of the form

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D(y_t - \sum_{i=1}^r \beta_i x_{it}) = \theta_q(B)\Theta_Q(B^s)a_t \quad (1)$$

where 's' is the length of the seasonal period, $s = 4$ or 12 . The polynomials $\phi_p(z)$, $\Phi_p(z)$, $\theta_q(z)$, $\Theta_Q(z)$ with degrees p ; P ; q ; and Q respectively have constant terms equal to one. These polynomials are constrained so that the zeroes of $\theta_q(z)$ and

$\Theta_Q(z)$ have magnitudes greater than or equal to one, so that the zeroes of $\phi_p(z)$ and $\Phi_p(z)$ have magnitudes greater than one. Because 'a_t' is assumed to be a sequence of independent variables with mean '0' and constant variance σ_a^2 , it follows from these constraints that

$$w_t = (1 - B)^d (1 - B^s)^D (y_t - \sum_{i=1}^r \beta_i x_{it}) \quad (2)$$

is a covariance stationary time series that satisfies the difference equation

$$\phi_p(B)\Phi_p(B^s)w_t = \theta_q(B)\Theta_Q(B^s)a_t. \quad (3)$$

Consequently, we can re-express the model (3) for y_t as,

$$(1 - B)^d (1 - B^s)^D y_t = \sum_{i=1}^r \beta_i \{(1 - B)^d (1 - B^s)^D x_{it}\} + w_t \quad (4)$$

This is a regression model with stationary ARMA errors w_t for suitably differenced y_t . Its regressors result from applying the same differencing operations to the x_{it} . The model (6), together with an assumption that the innovations at in the model for w_t are i.i.d $N(0, \sigma^2)$, determine the likelihood function that is maximized to estimate the regression coefficients β_i , σ^2 , and the coefficients of $\phi_p(B)$, $\Phi_p(B^s)$, $\theta_q(B)$, and $\Theta_Q(B^s)$. The default likelihood in X-12-ARIMA is the fully exact Gaussian likelihood.

In model estimation, any of the ARMA coefficients can be held at fixed values, such as zero. The program produces asymptotic standards errors, correlations, and t-statistics for the estimated coefficients as well as confidence intervals for forecasts.

Following are the different types of regressors used in RegARIMA:

- 1) *Constant term* (for stationary variable i.e., if the ARIMA model does not use differencing) represents the mean of the (stationary) series provided there are no other regression variables in the model.
- 2) *Trend term* captures the deterministic trend present in the data. The polynomial trend term reduces to a constant term after differencing.
- 3) *Fixed seasonal regressors* can be in one of the following category:
 - a) 11 contrasts in the 12 indicator variables.

- b) 11 variables taken from the Fourier series representation of a fixed monthly pattern.
- 4) X-12 ARIMA provides four other types of regression variables to deal with abrupt changes in the level of a series of a temporary or permanent nature: *additive outliers (AOs)*, *level shifts (LSs)*, *temporary changes (TCs)* and *ramps*. AOs affect only one observation in the whole series and hence this effect is removed by a dummy variable, which takes '0' at break and '1' for other period. LSs increases or decreases all observations from a certain time point onward by some constant amount, this LS effect is removed by introducing a dummy variable which takes value '-1' for all the time point up-to the break point and '0' for all the time points afterwards. TCs allow for an abrupt increase or decrease in the level of the series that returns to its previous level exponentially, this effect is captured by a variable which takes value '0' for all observation before the change point and α_i ($0 < \alpha < 1$) thereafter. Ramps allow for a linear increase or decrease in the level of the series over a specified time interval (say $t_0 - t_1$). Ramps are smoothed out by introducing a variable which take, three values '-1' for time $t < t_0$, $(t-t_0)/((t_1-t_0)-1)$ for $t_0 < t < t_1$, and '0' after the time point $t > t_1$. In the default procedure, appropriate AO and LS regressors are fit at (almost) all time points of the series (or of a chosen subspan), and their corresponding *t-statistics* are compared against specified critical values. Such large critical values are appropriate because of the large number of regressors to which individual significance tests are applied.

9.3. Diagnostic Checking

Diagnostic checking of a reg- ARIMA model is performed through various analyses of the residuals from model estimation, the objective being to check if the true residuals appear to be white noise – i.i.d. $N(0, \sigma^2)$. (Note: Normality of the a_t 's is not needed for large sample estimation and inference results; it is most important for validity of prediction intervals produced in forecasting). The check spec is used to produce various diagnostics statistics using the residuals from the fitted model. To check for autocorrelation, X-12-ARIMA can produce ACFs and PACFs of the

residuals (with standard errors), along with Ljung and Box (1978) summary Q-statistics. X-12-ARIMA can also produce basic descriptive statistics of the residuals and a histogram of the standardised residuals.

An important aspect of diagnostic checking of time series models is outlier detection. The outlier spec of X-12-ARIMA provides the automatic detection of additive outliers (AOs), temporary change outliers (TCs) and level shifts (LSs). X-12-ARIMA's approach of outliers detection involves computing *t-statistics* for significance of each outlier type at each time point, searching through these *t-statistics* for significant outlier(s), and adding the corresponding AO, LS, or TC regression variable(s) to the model.

A seasonal adjustment that leaves detectable residual seasonal and calendar effects in the adjusted series is usually regarded as unsatisfactory. Even if no residual effects are detected, the adjustment will be unsatisfactory if the adjusted values (or important derivative statistics, such as the percent changes from one month to the next) undergo large revisions when they are recalculated as future time series values become available. Frequent, substantial revisions cause data-users to lose confidence in the usefulness of adjusted data. Indeed, such instabilities in the adjustments should cause the producers of adjustments to question their meaning. Unstable adjustments can be the unavoidable result of the presence of highly variable seasonal or trend movements in the series being adjusted. X-12-ARIMA includes two types of stability diagnostics, sliding spans and revision histories.

The sliding spans diagnostics display, and provide summary statistics for, the different outcomes obtained by running the program up to four overlapping subspans of the series. For each month that is common to at least two of the subspans, these diagnostics analyze the difference between the largest and smallest adjustments of the month's datum obtained from the different spans. They also analyze the largest and smallest estimates of month-to-month changes and of other statistics of interest. They improve upon, or complement in important ways, earlier diagnostics for (i) determining if a series is being adjusted adequately, (ii) for deciding between direct and indirect adjustments of an aggregate series, and (iii) for confirming option

choices such as the length chosen for the seasonal filter or showing that other option choices must be tried.

The second type of stability diagnostic in X-12-ARIMA considers the revisions associated with continuous seasonal adjustment over a period of years. The basic revision calculated by the program is the difference between the earliest adjustment of a month's datum obtained when that month is the final month in the series and a later adjustment based on all future data available at the time of the diagnostic analysis. Similar revisions are obtained for month-to-month changes, trend estimates, and trend changes. Sets of these revisions, calculated over a consecutive set of time points within the series, are called revisions histories.

9.4. Forecasting

For a given reg ARIMA model with parameters estimated by the X-12-ARIMA programme, the forecast spec will use the model to compute point forecasts, and associated forecast standard errors and prediction intervals. The point forecasts are Minimum Mean Squared Errors (MMSE) linear predictions of future y_t 's based on the present and past y_t 's assuming that,

- i) the regARIMA model form is correct,
- ii) the correct regression variables have been included,
- iii) no additive outlier or level shifts will occur in the forecast period,
- iv) the specified ARIMA orders are correct, and
- v) the parameter values used (typically estimated parameters) are equal to the true values.

These are standardised assumptions, though obviously unrealistic in practical applications. What is more realistically hoped is that the regARIMA model will be a close enough approximation to the true, unknown model for the results to be approximately valid. Two sets of forecasts errors are produced. One assumes that all parameters are known. The other allows for additional forecast error that comes for estimating the regression parameters, while still assuming that the AR and MA parameters are known.

If the series is transformed, the forecasting results are first obtained in the transformed scale, and

then transformed back to the original scale. For example, if one specifies a model of form (1) for $y_t = \log(Y_t)$, where Y_t is the original time series, then y_t is forecasted first, and the resulting point forecasts and prediction interval limits are exponentiated to produce point and interval forecasts in the original (Y_t) scale. The resulting point forecasts are MMSE for $y_t = \log(Y_t)$, but not for Y_t , under the standard assumptions mentioned above. If any prior adjustments are made, these will also be inverted in the process of transforming the point forecasts and prediction interval limits back to the original scale.

If there are any user-defined regression variables in the model, X-12-ARIMA requires that the user supply data for these variables for the forecast period. For the predefined regression variables in X-12-ARIMA, the programme will generate the future values required. If user-defined prior adjustment factors are specified, values for these should also be supplied for the forecast period.

9.5. Limitations

Some of the limitations of X-12-ARIMA procedure are listed below:

1. Observations (data) from a time series to be modeled and / or seasonally adjusted using X-12-ARIMA should be quantitative, as opposed to binary or categorical.
2. Observations must be equally spaced in time, and missing values are not allowed.
3. X-12-ARIMA handles only univariate time series models, i.e., it does not estimate relationships between different time series.
4. The set of automatically identified outliers can change if the regressor set or ARIMA model type is changed.
5. Regressors with *t*-statistic values just below the critical values can have their *t*-statistics increase above the critical values as new data are added to the series over time. Conversely, regressors can drop out of the set of identified outliers as new data are added. The printed output of X-12-ARIMA's automatic-outlier-identification option lists months whose AO or LS regressors are close to the critical values. This is done to enable the user to consider in advance whether to include such regressors in subsequent runs of the program.