# Flexibilisation of X-11 for higher-frequency data

**Keywords:** *X-11, daily time series, seasonal adjustment, moving averages, JDemetra+*

## 1 Introduction

The X-11 algorithm has been used for seasonal adjustment of official statistics for decades [1]. It is the seasonal adjustment core in X-12-ARIMA, X-13ARIMA-SEATS and x13 in JDemetra [2],[3],[4]. Despite its popularity, its use is restricted to half-yearly, quarterly and monthly data. Yet, new data structures play an increasingly important role. The number of weekly and daily time series has surged in the last years and more frequently they are used for economic analysis [5],[6]. Here, we detail the X-11 algorithm and show how it can be generalised to an arbitrary integer-valued frequency, such as those relevant for daily data. For this study we use an extension of JDemetra+ version 3.0.

## 2 Methods

### 2.1 X-11

X-11 applies a series of moving averages of order $(3 \times \kappa)$, $\mathrm{MA}_{3\times\kappa}(\cdot)$, and Henderson-type moving averages of order $\gamma$, $\mathrm{HMA}_{\gamma}(\cdot)$, to the linearised time series $(Y_t)$ - either directly or period-wise - to separate out its unobservable components given in the basic time series model

$$Y_t = T_t + S_t + I_t. \tag{1}$$

The components comprise the trend-cycle $(T_t)$, seasonal $(S_t)$, and irregular $(I_t)$ component. The seasonal component captures periodically recurring influences with a cycle length $\tau$. Linearisation refers to any pre-processing, including the removal of outlier and calendar effects and potential transformations of the series such as applying the logarithm, so that $Y_t = f(Y_t^\star)$, where $Y_t^\star$ is the unadjusted time series.[1]

The X-11 algorithm consists of three iterations of sequentially applying weighted moving averages to the time series and its preliminary and final estimated components. These iterations are named 'B: Preliminary weights of extreme values', 'C: Final weights of extreme values', and 'D: Final estimates of trend-cyclical, seasonal and irregular component'. For our purposes, the steps in the B and C iteration are similar to those computed in the D iteration. But as in B and C solely the estimated extreme values are used in the next steps, we will only present the D iteration in detail to identify which parts need to be adapted to fully flexibilise X-11. An extensive introduction to X-11 can be found in [7] and [8].

Most steps in X-11 depicted in algorithm 1 do not require any adaptation, though the exceptions are particularly important. The first exception is the $\gamma$ parameter in $D_7$ and $D_{12}$ that governs the estimation of the trend-cycle. For monthly and quarterly data, $\gamma$ is set by the user or determined based on heuristics relying on the estimated variance of a preliminary irregular and trend component (see below). Here, we will show how these heuristics might be simplified without a loss of accuracy.

[1] For a thorough discussion of additional transformations and further modelling strategies, see [3].

**Algorithm 1** X-11: steps in D iteration

**procedure** (Using an extreme values correction ($C_{20,t}$) calculated in previous steps, the linearised time series $Y_t$ is decomposed.)
  $D_{1,t} \leftarrow Y_t - C_{20,t}$
  $D_{2,t} \leftarrow \text{MA}_{\tau\times 2}(D_{1,t})$
  $D_{4,t} \leftarrow D_{1,t} - D_{2,t}$
  **for** period $\in \{1,\dots,\tau\}$ **do**
    $D^{\star}_{5,t\in\text{period}} \leftarrow \text{MA}_{3\times\kappa}(D_{4,t\in\text{period}})$
  $D_{5,t} \leftarrow \text{MA}_{\tau\times 2}(D^{\star}_{5,t})$
  $D_{6,t} \leftarrow D_{1,t} - D_{5,t}$
  $D_{7,t} \leftarrow \text{HMA}_{\gamma}(D_{6,t})$
  $D_{8,t} \leftarrow Y_t - D_{7,t}$
  Estimate extreme value correction weights $w_t$
  **for** period $\in \{1,\dots,\tau\}$ **do**
    $D^{\star}_{10,t\in\text{period}} \leftarrow \text{MA}_{3\times\kappa}(D_{8,t\in\text{period}} \times w_t)$
  $S_t = D_{10,t} \leftarrow \text{MA}_{\tau\times 2}(D^{\star}_{10,t})$
  $D_{11,t} \leftarrow Y_t - D_{10,t}$
  $T_t = D_{12,t} \leftarrow \text{HMA}_{\gamma}(D_{11,t})$
  $I_t = D_{13,t} \leftarrow D_{11,t} - D_{12,t}$

The second relevant parameter is $\kappa$ which impacts the volatility of the estimated seasonal component. Until now, $\kappa$ has been restricted so that $\kappa \in \{1, 3, 5, 9, 15\}$. For higher frequency time series a wider range of values may be necessary. Deriving a flexible variant of the seasonal filters is not straight forward. To see this, note that for any odd $m$, an MA($\gamma$) filter for the observation $X_t$ is computed based on the observations $t^{\star} \in \{t - (\gamma/2 - .5), \dots, t + (\gamma/2 - .5)\}$. A filter is said to be symmetric, if all of $t^{\star}$ can be observed. Otherwise, an asymmetric filter needs to be used so that the available observations become a higher weight. While the derivation of the symmetric weights is trivial, as we will see, the derivation of the asymmetric weights relies on parameters, for which there is no default.

## 2.2 Simulation design

To investigate how $\gamma$ could be determined automatically, we simulate a large set of daily time series based on the extended time series model

$$Y_t = T_t * S_t^{(7)} * S_t^{(365)} * I_t \tag{2}$$

with seasonal factors with frequency $\tau$ ($S_t^{(\tau)}$).
The non-seasonal (SA) part of the series will be simulated using a versatile ARIMA model with $d \in \{0, 1\}$

$$Y_t^{(SA)} = T_t * I_t \sim ARIMA(1, d, 1). \tag{3}$$

The seasonal factors are simulated based on a smoothed seasonal random walk. The initial seasonal factor is given by

$$\tilde{S}_t^{(\tau)} = S_{t-\tau}^{(\tau)} + \epsilon_t \sim ARIMA(1, 1, 1),$$
$$\epsilon_t = 0.9\epsilon_{t-1} + u_t, u_t \sim N(0, \sigma_u^2).$$

To ensure that the seasonal factor has the intended magnitude, the standard deviation is fixed to $\sigma_S$ as follows

$$S_t^{(\tau)} = \tilde{S}_t^{(\tau)} / \sigma_{\tilde{S}} * \sigma_S.$$

# 3 Results

## 3.1 A default for the Henderson filter

For quarterly and monthly data, the automatic choice of an appropriate Henderson filter length is determined based on $\bar{I}/\bar{C}$, with quotients given by

$$\bar{I} = \frac{1}{n-1}\sum_{t=2}^{n} \text{abs}\left(\frac{I_t - I_{t-1}}{I_{t-1}}\right) \tag{4}$$

and

$$\bar{C} = \frac{1}{n-1}\sum_{t=2}^{n} \text{abs}\left(\frac{C_t - C_{t-1}}{C_{t-1}}\right). \tag{5}$$

where C is the trend-cycle derived by applying a Henderson(13)-filter to the (preliminarily) seasonally adjusted series. For the final trend estimation, $\gamma$ is set as follows

$$\gamma = \begin{cases} 9, & \bar{I}/\bar{C} < 1 \\ 23, & \bar{I}/\bar{C} > 3.5 \\ 13, & else. \end{cases} \tag{6}$$

We investigate the optimal $\gamma$ for higher frequency time series, we start with simulating daily data. As we know the true seasonally adjusted series, our objective is to find that $\gamma$ for which the mean absolute difference between the estimated and optimal seasonally adjusted series is smallest. Defining $y = g(x)$ to find the smallest odd integer $y \geq x$, the range of potential values is $\gamma \in \{g(\tau * 3/4), g(\tau + 1), g(\tau * 1.5), g(\tau * 2 + 1)\} = \{275, 367, 549, 731\}$, loosely inspired by the rules used for monthly data.
As can be seen in figure 1, there is no apparent exploitable relationship between $\bar{I}/\bar{C}$ and the optimal filter length. We therefore suggest using $\gamma = g(\tau+1)$ as a default for an arbitrary frequency. In our sample of simulated time series, using this filter length leads to the lowest mean absolute difference, as it is long enough to ensure, that no seasonal behaviour is captured in $T_t$ while maintaining sufficient flexibility.

## 3.2 Flexible seasonal filters

The seasonal filters used for lower frequency time series may not be sufficient for daily data. [6] shows that the estimation of the day-of-the-week effect, often very stable filters are needed.
As discussed above, obtaining flexible symmetric seasonal MA filters is straight forward. The challenge lies in finding appropriate asymmetric filters.
The asymmetric filters use weights derived by [9]. These M weights $v_1, ...v_M$ can be calculated by

$$v_j = w_j + \frac{1}{M}\sum_{i=M+1}^{N} w_i + \frac{(j - \frac{M+1}{2})D}{1 + \frac{M(M-1)(M+1)}{12}D}\sum_{i=M+1}^{N}(i - \frac{M+1}{2})w_i. \tag{7}$$

[8] argue empirically that $D = 9.8$ should be used. We will investigate this further based on simulated time series.

# 4 Conclusions

Based on the results presented here, a flexible X-11 is constructed, that can be used to seasonally adjust any integer-valued frequency. Next, we want to further analyse the defaults for $D$ and $\gamma$ based on a large scale simulation study. Additionally, we want to find a sensible default for $\kappa$, i.e. the order of the seasonal moving averages.

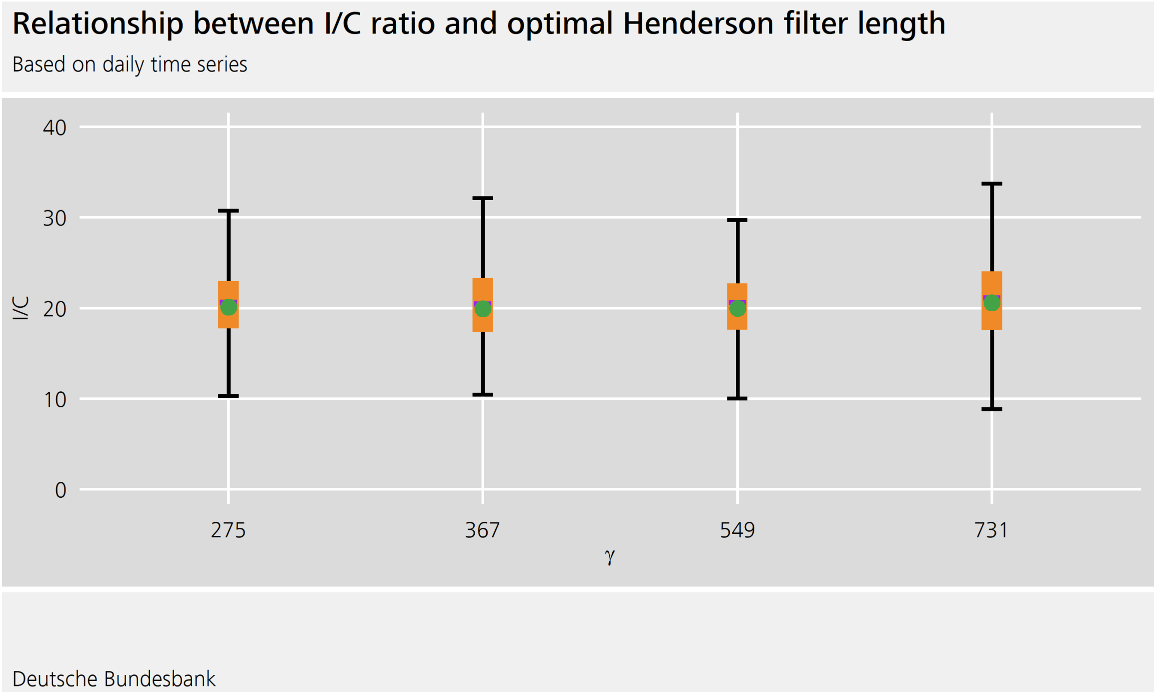


Figure 1: Relationship between I/C ratio and optimal Henderson filter length for daily data. Only includes the observations for that $\gamma$ for which the mean absolute difference between simulated and estimated seasonally adjusted series is minimal.

## References


[1] Allan H.; Musgrace John C. Shiskin, Julius; Young. The X-11 variant of the census method II seasonal adjustment program. Technical Report Technical Paper No. 15, 1967.

[2] Time Series Research Staff Census. *X-12-ARIMA Reference Manual*. Statistical Research Division, U.S. Census Bureau, 2011.

[3] U.S. Census Bureau. *X-13ARIMA-SEATS Reference Manual Version 1.1*. Time Series Research Staff, Statistical Research Division, U.S. Census Bureau, 2016.

[4] Silwia Grudkowska. *JDemetra+ Reference Manual*. NBP, 2015.

[5] S. Eraslan and T. Goetz. Weekly activity index for the german economy, 2020. URL https://www.bundesbank.de/wai.

[6] Daniel Ollech. Seasonal adjustment of daily data. *Deutsche Bundesbank, Discussion Paper Series 41/2018*, 2018. URL https://www.bundesbank.de/resource/blob/763892/0d1c33f19a204e2233a6fccc6e802487/mL/2018-10-17-dkp-41-data.pdf.

[7] Deutsche Bundesbank. The changeover from the seasonal adjustment method Census X-11 to Census X-12-ARIMA. *Monthly Report*, 51(9):39–50, September 1999.

[8] Dominique Ladiray and Benoît Quenneville. *Seasonal Adjustment with the X-11 Method*, volume 158 of *Lecture Notes in Statistics*. Springer, 2001.

[9] John C. Musgrave. A set of end weights to end all end weights. Technical report, 1964.