DEUTSCHE BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 16/2026

Selecting seasonal filters in X–13–ARIMA via cross-validation

Daniel Ollech

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank, Press and Public Relations Division, at the above address or via email: www.bundesbank.de/contact

Internet http://www.bundesbank.de

DOI https://doi.org/10.71734/DP-2026-16
ISBN 978-3-98848-073-6
ISSN 2941-7503

Bundesbank Discussion Paper No 16/2026

Selecting seasonal filters in X–13–ARIMA via cross-validation*

Daniel Ollech
Deutsche Bundesbank

Abstract

Official statistics routinely employs the X–13–ARIMA method to seasonally adjust economic time series. A key step is choosing the length of the seasonal moving average. Traditionally, this choice relies on ad hoc criteria and expert judgement. We propose a cross-validation-based filter selection scheme that offers greater flexibility, including the possibility of incorporating novel filters. This approach is particularly promising for the seasonal adjustment of weekly, daily, and high-frequency time series. We demonstrate how to integrate cross-validation into the X–13–ARIMA method and discuss the advantages of various implementation options. Evaluation on monthly and quarterly time series demonstrates that this selection method performs at least as well as, and often better than, conventional selection criteria.

Keywords: Seasonal adjustment; time series characteristics; non-parametric methods

JEL classification: C13, C14, C22, C53

*Contact address: Daniel Ollech, Deutsche Bundesbank, Central Office, Directorate General Statistics, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main, Germany. Tel.: +49 69 9566 36250. Email: daniel.ollech@bundesbank.de. The author thanks Christiane Hofer for implementing the code into Java and JDemetra+, Luca Brilhaus and Rouven Stäcker for help with R computations, Julian LeCrone for verifying the visual inspections, and conference participants of the 4th Workshop on Time Series Analysis for Official Statistics at the OECD for their helpful comments.

1 Introduction

Although cross-validation has been widely used for model selection in forecasting and machine learning for decades, its application to filter selection in seasonal adjustment remains uncommon. The calendar and seasonal adjustment procedure X–13–ARIMA combines a RegARIMA-based pre-processing – which is employed to estimate and adjust calendar as well as certain outlier effects and produces forecasts of the time series – with the X–11 seasonal adjustment method. This procedure is implemented in the software packages JDemetra+ (Grudkowska, 2017) and Win X–13 (U.S. Census Bureau, 2016).

In X–11, automatic filter selection continues to rely on heuristic rules that restrict the set of admissible seasonal filters. For instance, the moving seasonality ratio (MSR) – the default rule in the automatic mode – chooses among only three of the six seasonal filters available in the aforementioned software packages. This limitation becomes more consequential as seasonal adjustment is extended to daily and other high-frequency series, where additional seasonal filters must be created and considered (see Ollech, 2023).

The contributions of this paper are threefold. Firstly, we show how cross-validation can be introduced into the X–11 method in order to select the appropriate seasonal filter. Secondly, we examine at which stage of the algorithm – across alternative X–11 tables – cross-validation should be applied to achieve a favourable balance of accuracy and stability. Thirdly, we compare cross-validation-based selections to established rules (MSR; airline model parameter estimates; and expert visual assessment of seasonal adjustment results), documenting relative performance in terms of accuracy, deviance, bias, relative error, and stability.

The remainder of the paper is organised as follows. In Section 2 we present the X–11 method used to decompose linearised time series into seasonal, trend and irregular component in X–13–ARIMA. We also discuss the hitherto used selection rules for the seasonal filters. Section 3 introduces the cross-validation framework, discusses robustness considerations, and explains how to integrate cross-validation into X–11 for seasonal filter selection. Section 4 details our evaluation strategy: we describe the simulation design that generates series with known seasonality and define accuracy, deviance, bias, relative error, and stability metrics; we then compare cross-validation applied at different X–11 tables with traditional rules and assess stability on a large set of real-world series. We also discuss which X–11 table we recommend as the basis for cross-validation in practice. A set of real-world examples further illustrates differences across methods in representative applications. Section 5 concludes with practical recommendations for implementing cross-validation in X–11 and outlines directions for future work.

2 The seasonal adjustment method X–11

Let the linearised time series – i.e. a time series adjusted for calendar and outlier effects – be represented by

$$Y_t = T_t + S_t + I_t \quad \forall \quad t = 1, \ldots, T \tag{1}$$

where T_t denotes the trend-cycle component, capturing the medium- and long-term progression of the series. S_t represents the seasonal component with seasonal length τ, encompassing all periodically repeating patterns with a cycle length of less than or equal to one year. I_t is the irregular component that captures all remaining short-term dependent and random fluctuations.

The X–11 method utilises weighted moving averages to iteratively estimate and adjust the trend-cycle and seasonal components as illustrated in Figure 1. Each of the intermediate results is stored in so-called tables. For example, table B2 contains the first estimation of the trend-cyclical component.

To estimate the trend component in the different stages of X–11, the trend filters employed include a $2 \times \tau$ moving average, where τ is the number of observations per year, and Henderson moving averages[1]. To automatically select the lengths of the Henderson filters, a statistic, namely the I/C ratio, based on the relationship between the variability of the estimated preliminary trend-cycle and the irregular components is used (Ladiray and Quenneville, 2001). However, the selection of the trend filter exerts only a minor influence on the estimated seasonally adjusted time series. Consequently, we focus this study on identifying the optimal configuration for cross-validation, specifically for selecting the seasonal filter.[2] This is also in line with the default algorithms implemented in X–11 that do not select trend and seasonal filters jointly.

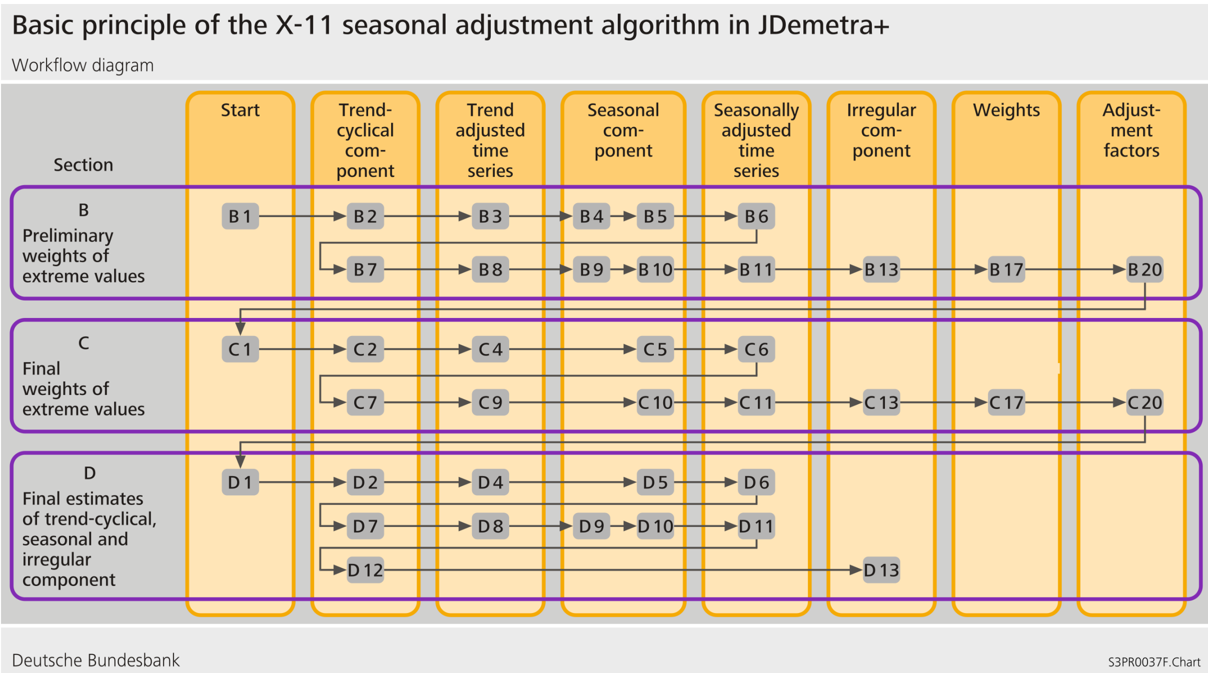


Figure 1: The X–11 method

The seasonal filters available in X–11 are symmetric moving averages of length $3 \times K$. This means that an MA(3) followed by a MA(K) with $K \in \{1, 3, 5, 9, 15\}$ is applied on

[1]Henderson moving averages are filters used to extract the trend-cycle component. Their weights are constructed so that the filter reproduces cubic polynomials exactly within the span of the filter (Henderson, 1916)

[2]Haller, Daniel, and Bellone (2025) show how to use cross-validation in the STAHL-algorithm, a seasonal and holiday estimation procedure inspired by the seasonal adjustment method STL, to simultaneously select trend and seasonal filters.

a period-by-period basis to the trend- and potentially extreme-value-adjusted series.[3] At the ends of the series, where we do not have enough past or future observations to apply symmetric filters, asymmetric filters are used (cf. Musgrave, 1964; Findley, Monsell, Bell, Otto, and Chen, 1998).

2.1 Selection of seasonal filters in X–11

Several traditions coexist in practice for selecting the X–11 seasonal filter: (a) empirical rules based on the MSR statistic; (b) mappings from estimated airline model parameters to X–11 filters; and (c) expert visual inspection of seasonal adjustment results. The following sections evaluate the comparative advantages and limitations of these approaches and introduces cross-validation as an alternative.

Moving seasonality ratio (MSR) Mirroring the logic of the $\bar{I}/\bar{C}$ ratio for trend filters, the MSR compares the variability of preliminary irregular and seasonal components.

Marris (1960) posits that the importance of the irregular component compared to the seasonal component can guide the choice of an appropriate seasonal filter. The intuition is that the more the irregular component dominates and thereby obscures the seasonal component, the more observations are needed to estimate the latter. He derives a statistical criterion based on the ratio of these components and the filter weights. Lothian (1984) uses a variant that omits the filter weights. For each period $i \in (1, \cdots, \tau)$ with T_i observations, the moving seasonality ratio (MSR) is calculated as the ratio between period specific measures of the variability of the irregular component

$$v(I_i) = (T_i - 1)^{-1} \sum_{t=2}^{T_i} |\hat{I}_{i,t} - \hat{I}_{i,t-1}|$$

and the seasonal component

$$v(S_i) = (T_i - 1)^{-1} \sum_{t=2}^{T_i} |\hat{S}_{i,t} - \hat{S}_{i,t-1}|.$$

Then, the MSR for the whole series is given by

$$\mathrm{MSR} = \frac{\sum_i^\tau T_i v(I_i)}{\sum_i^\tau T_i v(S_i)} \tag{2}$$

Thresholds translate MSR values into seasonal-filter choices. The cutoffs implemented in X–11, however, differ from those originally proposed by Lothian (1984); see Dagum, Chhab, and Solomon (1991) and Dagum (1999) for a discussion of revision concerns near boundaries. In the X–11 decision rule, if the MSR falls between cutoffs, up to five years are omitted and the MSR is recomputed; if the result remains borderline, the S3×5 filter is chosen, otherwise the indicated filter is selected. The original and modified cutoffs are

[3] There is also a stable seasonal filter, which sets the seasonal filter to the period-specific mean over the available span. To the best of our knowledge, the stable filter is rarely used in official statistics, except in cases where the chosen $3{\times}K$ is too long for the series and X–11 falls back to the stable filter.

shown in Table 1. This omission and recomputation decreases the propensity to change the seasonal filter when new observations are added.

SMA	MSR ranges acc. to Lothian (1984) 0–15 years	15+ years	MSR ranges in X–11
S3×1	0.0 – 2.3	0.0 – 2.1	
S3×3	2.3 – 4.1	2.1 – 3.8	0.0 – 2.5
S3×5	4.1 – 5.2	3.8 – 5.0	3.5 – 5.5
S3×9	5.2 – 6.5	5.0 – 6.9	6.5+
Stable	6.5 – 7.1	6.9 – 7.1	

In the X–11 decision rule, if the MSR falls between the cutoffs, up to 5 years of the series are omitted and the MSR is recalculated. If it still falls between the cutoffs, the S3×5 filter is chosen.

Table 1: Seasonal filter selection based on moving seasonality ratio (MSR)

The MSR criterion is applied to table D8b, which constitutes the seasonal-irregular component of table D8 after incorporating extreme value replacements from table D9 in X–11. This implementation, however, presents a key limitation: the user-selected filter is only utilised in the final estimation of the seasonal component. In all preceding estimations, the default seasonal filters remain in effect. Consequently, if the recommended filter is selected, it supersedes the default filters in the steps prior to D8, thereby producing corresponding variations in the results.

Airline model parameter estimate (AMPE) based selection A time series Y_t following an ARIMA$(p, d, q)(P, D, Q)\tau$ model satisfies

$$\phi_p(B)\Phi_P(B^\tau)\nabla^d\nabla_\tau^D Y_t = \theta_q(B)\Theta_Q(B^\tau)\varepsilon_t, \tag{3}$$

where B is the backshift operator, such that $B^k Y_t = Y_{t-k}$, and $\nabla_k = 1 - B^k$ represents the differencing operator. Here, d and D denote the non-seasonal and seasonal orders of differencing, respectively. The polynomials ϕ_p and Φ_P are the non-seasonal and seasonal autoregressive (AR) components of orders p and P, while θ_q and Θ_Q are the non-seasonal and seasonal moving average (MA) components of orders q and Q. The term ε_t represents white noise with zero mean and finite variance. The airline model is the parsimonious (0,1,1)(0,1,1) specification, so that only a non-seasonal MA and a seasonal MA parameter need to be estimated. Depoutot and Planas (1998) compare X–11 filters with Wiener–Kolmogorov filters implied by the airline model, using the integrated squared difference between frequency responses as a distance metric, $d(c_1, c_2) = \int_0^\pi |c_1(e^{i\omega}) - c_2(e^{i\omega})|^2 d\omega$, to identify, for given parameter estimates, the X–11 filter closest to the model-based ideal; see Table 2 for the resulting mapping for the seasonal filter (associated Henderson trend filters are omitted here). In operational terms, this provides a simple rule: estimate the airline model and select the seasonal filter associated with the estimated (θ, Θ) pair.

Selection based on visual inspection of seasonal adjustment results A third approach is expert-driven: the seasonal filter is chosen by visually inspecting SI ratios,

θ	**0.95**	**0.8**	**0.7**	**0.6**	**0.5**	**0.4**	**0.2**	**0.0**
	Θ							
0.95	S3×15	S3×15	S3×9	S3×5	S3×3	S3×3	S3×3	S3×3
0.8	S3×15	S3×15	S3×9	S3×5	S3×3	S3×3	S3×3	S3×3
0.7	S3×15	S3×15	S3×9	S3×5	S3×3	S3×3	S3×3	S3×3
0.6	S3×15	S3×15	S3×9	S3×5	S3×5	S3×3	S3×3	S3×3
0.5	S3×15	S3×15	S3×9	S3×5	S3×5	S3×3	S3×3	S3×3
0.4	S3×15	S3×15	S3×9	S3×5	S3×5	S3×3	S3×3	S3×3
0.2	S3×15	S3×15	S3×9	S3×5	S3×3	S3×3	S3×3	S3×3
0.0	S3×15	S3×15	S3×9	S3×5	S3×3	S3×3	S3×3	S3×3

Table 2: Relationship between airline model parameter estimates (AMPE) and optimal seasonal filters for a monthly time series as derived by Depoutot and Planas (1998). The shading is used solely to highlight the pattern in the optimal filters.

which juxtapose the trend-adjusted series (raw seasonal component) with the estimated seasonal component on a period-by-period basis; see, e.g., Kirchner (1999) and Kirchner, Ladiray, and Mazzi (2018). The guiding principle is to select a filter that tracks genuine changes in the seasonal pattern without overreacting to irregular movements. Shorter filters can be too responsive, absorbing irregular movements into the seasonal estimate, whereas longer filters may be too sluggish and fail to capture evolving seasonality.

As an illustration, consider Figure 2 that shows different filter outcomes for the third month of the year. Generally, the March values of the raw seasonal component of industrial production show high variability and a steady upward tendency over the years. Shorter seasonal filters result in a seasonal component that reacts too strongly to single seasonal-irregular observation, thereby capturing too much irregularity into the seasonal component. Economically speaking, there is no reason to believe that the seasonal effect should fluctuate almost periodically around a longer-term trend, as seen with the S3×1 and S3×3 seasonal moving averages.

Conversely, longer seasonal filters may respond too slowly to upward trends over time. In this example, the stable seasonal filter does not track the long-term increase in the raw seasonal component. Even the S3×15 filter remains below all but one of the seasonal-irregular values during the last seven years and may thus may be regarded as being to inflexible.

The example presented here also illustrates the difficulty of this approach: multiple filters may produce similar results and serve as acceptable candidates. In such cases, the selection may depend on subjective preferences and idiosyncrasies. In this instance, we would favour the S3×9 filter, although the S3×5 filter could also be considered.

3 Cross-validation

Cross-validation is widely used for model selection, particularly in forecasting and predictive modelling (Hastie, Tibshirani, and Friedman, 2009). For this purpose, the available dataset is repeatedly split into training and test sets. For each training set, the candidate

Figure 2: SI ratios for industrial production with different seasonal filters for March

models are estimated and used to forecast the target in the corresponding test set. The discrepancy between forecasts and realised values is computed for every partition and aggregated to yield an empirical criterion for choosing the configuration that generalises best to unseen data.

Variants of cross-validation differ primarily in the number and structure of splits and in whether an observation can appear in multiple test sets. Common choices include leave-one-out cross-validation (LOOCV), k-fold, repeated or Monte Carlo splits, and time-series-specific schemes such as rolling-origin and h- or hv-block cross-validation designed for dependent data (Arlot and Celisse, 2010; Racine, 2000).

Cross-validation can be sensitive to the presence of outliers (Leung, 2005), but robustness can be improved by using robust estimators or robust loss functions. In the X–11 framework, cross-validation benefits from the use of linearised series: RegARIMA pre-treatment typically identifies and adjusts for calendar effects and outliers (Findley et al., 1998), and, as described below, some candidate target series are corrected for extreme values, which further reduces sensitivity.

Generalized cross-validation (GCV) is a variant of cross-validation applicable to bandwidth and filter selection (Craven and Wahba, 1979; Golub, Heath, and Wahba, 1979). It provides a computationally efficient approximation to leave-one-out cross-validation (LOOCV), which is described below.

In the usual GCV setting, the estimator can be written in the linear smoother form

$$\hat{X} = S(\lambda)\, X,$$

where $S(\lambda)$ is a *fixed* (for given λ) hat matrix that maps the observed data X to the fitted values $\hat{X}$. This matrix encapsulates the influence of each observation on each fitted value and allows the prediction errors to be expressed in closed form via its diagonal elements, leading to the familiar GCV criterion

$$\mathrm{GCV}(\lambda) = \frac{\|X - \hat{X}\|^2}{\{T - \mathrm{tr}(S(\lambda))\}^2}.$$

In the case of smoothing splines, GCV-type estimators of the smoothing parameter are well established (e.g. Bottegal and Pillonetto, 2018; Lukas, de Hoog, and Anderssen, 2016). Paige and Trindade (2010) develop a GCV-type parameter estimator for the well-known Hodrick–Prescott trend filter by reformulating the filter as a linear penalised spline model. For X–11, however, a direct GCV formulation is problematic. The procedure relies on sequences of time-varying symmetric and asymmetric seasonal moving averages with endpoint adjustments, meaning that the fixed hat matrix $S(\lambda)$ cannot be easily derived in closed form. X–11's adaptive, time-dependent filters introduce structural variability, making it difficult to express the smoothing operation as a single fixed linear transformation. Any reformulation would, at best, capture only the symmetric interior filters, omitting a substantial portion of the series at the beginning and at the ends where asymmetric filters are applied. Moreover, X–11 is computationally fast, so the main advantage of GCV over exact LOOCV, namely, reduced computational cost, is limited in this case.

In general, the cross-validation statistic, based on some error metric F that quantifies the distance between an estimate $\hat{X}$ and its target X over P partitions, can be written as

$$CV = \sum_{p=1}^{P} \frac{T_p}{T} \mathrm{F}\left(X_p, \hat{X}_p\right)$$

where T_p is the number of observations in the p-th partition, T is the total number of observations, X_p denotes the realised target values in the test set of partition p, and $\hat{X}_p$ the corresponding predictions.

A special case is leave-one-out cross-validation (LOOCV), where each test set contains a single observation ($P = T$), yielding:

$$CV = \sum_{p=1}^{T} \frac{1}{T} \mathrm{F}\left(X_p, \hat{X}_p\right)$$

Friedman (1984) introduces a scatterplot smoother based on locally weighted linear fits with a variable span, which can also be used to estimate a time-series trend-cycle. A key feature is pointwise LOOCV to choose the local span: at each x-location it evaluates the absolute LOOCV residual for several candidate spans and selects the one with the smallest locally averaged value. Because the underlying fit is a linear smoother, the LOOCV residual at index j can be computed analytically, so the smoother need not be refitted for each left-out point. Note, this is exact LOOCV for linear smoothers, not the global GCV approximation.

Replicating this strategy within the X–11 framework is not straightforward either. X–11 estimates of the seasonal component are produced by sequences of time-varying linear filters with endpoint adjustments, so there is no simple, fixed hat matrix from which per-point estimates can be obtained.

Advantages of cross-validation for filter selection Each of the established filter selection strategies has clear merits, but also limitations. Both the MSR-based rule and the airline model-based selection are operationally simple yet restricted by design to a subset of filters. The latter presumes the airline model is an adequate approximation, which often but not always will be the case. The visual inspection, while often insightful, is subjective and will thus vary across analysts. Cross-validation offers a complementary, unified alternative: it evaluates candidate filters by their pseudo-out-of-sample performance in recovering the raw seasonal component itself. In this paper we tailor cross-validation to X–11, assess where in the algorithm it should be applied, and benchmark its empirical properties against established rules.

Beyond methodological coherence, our results preview two practical messages. First, cross-validation is competitive with MSR and airline model rules in identifying filters that minimise deviations from the true seasonal signal in simulations and from user judgements in real-world applications. Second, among the candidate X–11 tables, applying cross-validation to the B3 table provides a particularly favourable balance of accuracy and stability while insulating the choice from user-defined settings elsewhere in the procedure.

3.1 Incorporating cross-validation into X–11

To account for all observations and for both symmetric and asymmetric seasonal filters, we employ standard (non-generalised), model-independent cross-validation. Because seasonal moving averages are local, pointwise linear filters, leave-one-out cross-validation (LOOCV) is the most natural choice: each fold omits a single observation and evaluates the filter's estimate at that index.

We use an imputation-based LOOCV approximation because X?11 does not accept missing values. Assuming the trend as given, the target series is the trend-adjusted series, i.e., the raw seasonal component, denoted S^{raw}. It is sometimes also referred to as the seasonal-irregular component. The estimated seasonal component, $\hat{S}^{(-p),K}$, is obtained by applying seasonal moving averages of order $\text{S3}\times K$ to the raw seasonal component, $S^{\text{raw},(-p)}$, from which observation p has been omitted. The cross-validation statistic based on an error metric F is then given by

$$\text{CV}_{(K)} = \sum_{p=1}^{T} \text{F}\left(S_p^{\text{raw}},\, \hat{S}_p^{(-p),K}\right),$$

where $\hat{S}_p^{(-p),K}$ is the p-th observation, which is calculated based on $S^{\text{raw},(-p)}$. By default, X–11 cannot handle missing values. To compute the seasonal component when an observation is omitted, we replace the missing value using the built-in extreme-value correction routine in X–11. Specifically, the imputed value is the simple average of the four nearest observations in the same period (for details, see U.S. Census Bureau, 2016; Ladiray and Quenneville, 2001).

Our suggested implementation of cross-validation into X–11 is summarised in Algorithm 1.

Our simulations suggest that the preferred error metric is the mean squared error (MSE)[4], i.e.

$$\mathrm{CV}_{(K)} = \sum_{p=1}^{T} \frac{1}{T} \big(S_p^{\mathrm{raw}} - \hat{S}_p^{(-p),K} \big)^2 \tag{4}$$

Given its iterative nature, X–11 produces several estimates of S^{raw} (see Figure 1), notably in tables B3, B8, C4, C9, D4, and D8. It also computes additional extreme-value corrections, for example, those reported in table D9.

It is not obvious a priori which table is preferable. Consider the first and the last candidates. The B3 raw seasonal is obtained by subtracting a simple trend estimate from the linearised series (B1), where B1 is the RegARIMA pre-treated series (adjusted for calendar effects and outliers). The trend is estimated by applying a simple $2{\times}\tau$ moving average to B1 (with τ the seasonal period) and is therefore independent of the user's choice of Henderson filter, which is only applied at a later stage. By contrast, D8 is obtained by subtracting a Henderson-filter-based trend estimate from table D1, the linearised series corrected for extreme values. Extreme-value detection uses limits based on a moving standard deviation, and these limits are user-configurable. The D8b includes an additional extreme-value correction. While D8b incorporates final trend and extreme value estimates – potentially offering greater accuracy – it is particularly dependent on user-specified parameters such as the aforementioned sigma limits. In contrast, the B3 table offers a significant advantage: it remains invariant to user-defined X–11 specifications, as it relies solely on a $2{\times}\tau$-moving average type trend-adjustment. In short, B3 has the advantage of being largely unaffected by user-specified X–11 options, whereas D8b uses the final estimates of trend and extreme values and may therefore be more accurate.

[4]Using MAE yields qualitatively similar conclusions; however, MSE identifies the correct seasonal filter slightly more often and with greater reliability. Here, we will report the root of the MSE, i.e. the RMSE, for ease of interpretation. Detailed results are available upon request.

Algorithm 1 Cross-validation for X–11 Filter Selection

Require: Raw seasonal component $S^{\text{raw}} = (S_1^{\text{raw}}, \dots, S_T^{\text{raw}})$ with τ observations per year, error metric F (e.g., squared error)

for each K in $\{1, 3, 5, 9, 15\}$ **do**
  Initialize $\text{CV}_{(K)} \leftarrow 0$
  for $p = 1$ to T **do**
    Create $S^{\text{raw},(-p)}$ by omitting S_p^{raw} from S^{raw}
    Impute $S_p^{\text{raw},(-p)} \leftarrow \frac{1}{4}(S_{p-2\tau}^{\text{raw},(-p)} + S_{p-\tau}^{\text{raw},(-p)} + S_{p+\tau}^{\text{raw},(-p)} + S_{p+2\tau}^{\text{raw},(-p)})$
    Apply the S3 $\times$ K moving average to $S^{\text{raw},(-p)}$ to obtain $\hat{S}_p^{(-p),K}$
    Compute error $e_p \leftarrow \text{F}(S_p^{\text{raw}}, \hat{S}_p^{(-p),K})$
    Update $\text{CV}_{(K)} \leftarrow \text{CV}_{(K)} + \frac{1}{T} e_p$
  end for
  Store $\text{CV}_{(K)}$
end for
Select $K^* = \arg\min_K \text{CV}_{(K)}$
Output: Use S3 $\times$ K^* as the seasonal filter

Note: At the beginning and end of the series, where not all two past or two future period-specific observations are available, the four closest period-specific observations are used for imputation (U.S. Census Bureau, 2016).

4 Evaluating the seasonal filter selection algorithms

4.1 Design of simulation study

To evaluate the performance of cross-validation across different X–11 tables, we simulate monthly time series for which the true seasonal component is known. Knowing the ground truth seasonality allows us to quantify the estimation error and, for each series, identify the optimal S3$\times K$ seasonal filter. Optimal means that estimated seasonal component, $\hat{S}_t$, for which the mean absolute error between the simulated S_t and $\hat{S}_t$ is minimal.

We follow the simulation framework of Ollech (2021); data generation is implemented in the R package `tssim` (version 0.2.7).[5]

Let $Y_{t=1}^{T}$ be a monthly series of length $T = 12N$, so N is the number of years. We decompose

$$Y_t = Y_t^{SA} + S_t, \tag{5}$$

where Y_t^{SA} is the non-seasonal component and S_t is the seasonal component. The com-

[5]As a robustness check, we replicated the simulation study using the time series simulation algorithms of Bandara, Hyndman, and Bergmeir (2025) and Cuevas and Quilis (2023). The results are qualitatively similar to those presented here and are available upon request.

Algorithm 2 Simulating time series for which the optimal seasonal filter in X–11 is known

for $N \in \{5, 10, \ldots, 30\}$ years **do**
  Set $T \leftarrow 12N$.
  for $K \in 1, 3, 5, 9, 15$ **do**
    Initialize counts $m_{N,K} \leftarrow 0$
    while $m_{N,K} < 500$ **do**
      Simulate Y_t and S_t according to Equation 5.
      for $L \in \{1, 3, 5, 9, 15\}$ **do**
        Estimate $\hat{S}_{t,L}$ using X–11 with an S3×L seasonal filter.
        Compute $\text{MAE}_L \leftarrow T^{-1} \sum_{t=1}^{T} |\hat{S}_{t,L} - S_t|$.
      end for
      $L_{\text{optimal}} \leftarrow \arg\min_{L \in 1,3,5,9,15} \text{MAE}_L$.
      if $L_{\text{optimal}} = K$ **then**
        $m_{N,K} \leftarrow m_{N,K} + 1$.
      end if
    end while
  end for
end for

ponents are generated independently as follows:

$$Y_t^{SA} \sim \text{ARIMA}(p, d, q), \tag{6}$$

$$\tilde{S}_t = \tilde{S}_{t-12} + \epsilon_t, \tag{7}$$

$$\epsilon_t = \beta \epsilon_{t-1} + u_t, \quad u_t \sim \mathcal{N}(0, \sigma_u^2), \quad |\beta| < 1, \tag{8}$$

$$S_t = \tilde{S}_t \cdot \frac{\sigma_S}{\sigma_{\tilde{S}}}. \tag{9}$$

The recursion $\tilde{S}_t = \tilde{S}_{t-12} + \epsilon_t$ defines a seasonal random walk with AR(1) innovations. The scaling in (9) enforces a user-specified standard deviation σ_S for the seasonal component (with $\sigma_{\tilde{S}}$ denoting the sample standard deviation of $\tilde{S}_t$). The ARIMA model for the non-seasonal part is the {tssim} packages default ARIMA(3,1,1) model.

Based on Equation 5, we conduct a Monte Carlo experiment as follows. For each sample length $N \in \{5, 10, \ldots, 30\}$ years we repeatedly simulate monthly series and evaluate the X–11 seasonal component obtained with each candidate S3×K filter, $K \in 1, 3, 5, 9, 15$. For a given N, the filter yielding the smallest MAE is deemed optimal for that replication. We continue sampling until every filter has been selected as optimal in 500 replications. This results in 2,500 series per N (i.e., R=$15,000$ series overall). The detailed procedure is provided in 2.

We employ a balanced simulation design, generating an equal number of time series for each optimal seasonal filter (S3×1 through S3×15). Although the distribution of optimal seasonal filters in real-world applications is unknown and therefore not considered here, practical experience suggests that the S3×5 and S3×9 filters are likely to be selected by many statisticians for a substantial proportion of time series. We provide detailed results to ensure that the conclusions drawn from the simulation study are not unduly influenced by the number of specific series-filter combinations simulated.

Table 3 summarises the evaluation metrics used in this study. The accuracy metric indicates the percentage of times the optimal filter was correctly identified across all time series. Since the MSR- and AMPE-based filter selectors are limited to a subset of available filters, we also report accuracy calculated only for those series where these specific filters are optimal.

When a suboptimal filter is selected, it is important to assess whether the chosen filter is substantially different from the optimal one or relatively similar. The deviance metric quantifies, on average, how close the selected filter is to the optimal filter in terms of absolute rank difference. The bias metric, on the other hand, reveals whether there is a systematic tendency to select filters that are either too short or too long compared to the optimal choice.

The consequences of not selecting the optimal filter can vary. For some series, several filters may yield similar mean absolute errors (MAE) between the estimated and simulated seasonal components, while for others, the choice of filter can have a more pronounced impact. The relative MAE measures the MAE of the chosen filter as a percentage of the MAE of the optimal filter. A value of 100 indicates that the chosen filter performs as well as the optimal filter, while values above 100 indicate worse performance.

Finally, we assess the stability of the filter selection process. In official statistics, it is generally desirable for the filter selection algorithm to produce consistent results as new data become available, since changes in the selected filter can lead to additional technical revisions. The stability metric reports the percentage of cases in which the chosen filter remains unchanged when an additional year of observations is added.

Metric	Explanation
Accuracy	How often the optimal filter was found, in percent
Accuracy w/o S3$\times K$	Accuracy with series excluded for which specified filters is optimal
Relative MAE	$\frac{\text{MAE}_{K_{\text{chosen}}}}{\text{MAE}_{K_{\text{optimal}}}} \times 100$
Deviance	$\frac{1}{R}\lvert\text{Rk}(K_{\text{chosen}}) - \text{Rk}(K_{\text{optimal}})\rvert$
Bias	$\frac{1}{R}[\text{Rk}(K_{\text{chosen}}) - \text{Rk}(K_{\text{optimal}})]$
Stability	How often K_{chosen} remains the same when one year of observations is added

$Rk(K) \in (1, 2, 3, 4, 5)$ is the rank of K, K_{chosen} is the filter chosen, $K_{optimal}$ is the filter that would have yielded the seasonal component closest to the simulated seasonal component, MAE_K is mean absolute difference between true seasonal component and seasonal component if filter K is used, in percent. R is number of simulated series.

Table 3: Evaluation metrics and their explanations

4.2 Results of simulation study

Comparing cross-validation with other metrics From Table 4, we observe that cross-validation is generally a suitable alternative to the established selection rules. The accuracy of identifying the optimal model is qualitatively similar for cross-validation – regardless of the tables used – to the AMPE and MSR rules. Even when restricting the analysis to those time series for which a seasonal filter is optimal that AMPE and/or MSR is able to detect, cross-validation remains competitive.[6]

Focusing on the traditional X–11 filters S3×3, S3×5, and S3×9, that is, excluding time series for which S3×1 or S3×15 is optimal – cross-validation based on the B3 table achieves an accuracy of 47.7%, compared to 38.5% for AMPE and 44.6% for MSR. Conversely, when including series for which S3×15 is optimal, AMPE is correct 46.8% of the time, exceeding the 35.8% accuracy of cross-validation (B3 table) and the 33.5% achieved by MSR.

A noteworthy feature of cross-validation is its tendency to select filters more evenly across the available options. As shown in Table 9, the AMPE rule tends to favour more extreme filters, selecting the S3×3 filter in 60.0% of cases and the S3×15 filter in 21.7% of cases.

As expected, given the design of the cutoff values (see Section 2.1), the MSR criterion has a strong propensity to select the S3×5 filter, doing so in 62.1% of cases. This explains its high accuracy for series where S3×5 is the optimal filter (74.7% accuracy).

By contrast, across all X–11 tables, cross-validation almost never selects any single filter more than 30% of the time, which is remarkable even when accounting for the larger number of filters available for selection.

A useful metric for assessing whether the selection propensity for certain filters is skewed is the *bias*, as defined in Table 3. Overall, the bias values are generally small; however, the infrequent selection of the S3×9 filter by the MSR rule results in the largest absolute bias, at -0.3. With respect to deviance, all selection rules exhibit comparable performance, with AMPE and most cross-validation based selections yielding the lowest deviance values.

Finally, to evaluate the extent to which each method approaches the best possible seasonal adjustment, we consider the relative MAE. According to this metric, AMPE achieves the lowest errors, closely followed by cross-validation – except for a few cases, such as selection based on table D8b – while the MSR rule exhibits the highest errors.

Reliability of the Identified Filters The accuracy of the selected filters ought to be the primary criterion for evaluating a selection algorithm. In practice, however, another important consideration is the reliability of the selection. If the chosen filter changes frequently with the addition of new data points, this instability can be problematic and is generally regarded as an undesirable property of the algorithm. Therefore, we assess the stability of different selection methods by measuring how often the selected filter remains unchanged when one year of data is added to the time series.

[6] The simulation design implicitly assumes that each filter is used with equal frequency. However, this is not the case for the monthly data adjusted by the Deutsche Bundesbank, where the distribution of filters is as follows: S3×1 $<$ 0.0%, S3×3 = 1.7%, S3×5 = 26.2%, S3×9 = 70.9%, S3×15 = 1.1%. Using this empirical distribution, the accuracy of cross-validation is 36.3% based on the B3 table and 40.0% based on the D8 table, while AMPE achieves 17.1% accuracy and MSR 15.0%.

	Cross-validation based on X–11 table								
	B3	B8	C4	C9	D4	D8	D8b	AMPE	MSR
Overall	42.4	43.1	41.5	39.1	40.2	44.0	33.2	37.4	26.8
Optimal filter									
S3×1	64.2	60.0	57.3	59.6	51.3	67.8	44.5	0.0	0.0
S3×3	26.2	29.1	27.7	31.4	25.8	29.5	32.9	85.1	56.5
S3×5	19.3	23.4	22.3	31.4	22.0	24.3	40.8	16.2	74.7
S3×9	42.4	45.4	50.7	44.1	53.5	43.1	35.0	14.1	2.6
S3×15	60.0	57.7	49.3	29.1	48.5	55.2	12.7	71.6	0.0
all w/o									
S3×1	35.8	36.0	35.6	36.4	34.9	36.7	33.6	46.8	33.5
S3×1, S3×15	47.7	48.0	47.4	48.5	46.5	48.9	44.7	38.5	44.6
Filter selected									
S3×1	54.9	51.7	53.2	46.1	53.0	52.0	47.1	—	—
S3×3	34.9	34.7	33.9	32.7	32.1	36.5	29.3	28.4	32.4
S3×5	29.9	31.6	29.0	32.6	26.9	33.1	28.0	27.6	24.0
S3×9	32.2	36.6	33.6	34.9	33.5	37.0	29.8	42.9	17.9
S3×15	53.8	56.5	59.5	55.6	59.8	56.0	42.4	66.1	—

Remark: Optimal filter refers to the filter that minimises the absolute difference between the estimated and simulated seasonal component. All w/o S3×K restricts both the simulated time series considered as well as the candidate filters in cross-validation, thus excluding S3×K.

Table 4: Accuracy (%) of identifying the optimal S3×K seasonal moving average

For this analysis, we use real-world data comprising a large set of time series that are regularly seasonally adjusted by the Deutsche Bundesbank. The dataset includes aggregates and component time series from industrial production, industry turnover, industry orders, monetary indicators, and trade, encompassing 596 series with observation lengths ranging from 8 to 56 years (23 years on average). These are not real-time data, i.e. we do not consider revisions due to changes to the unadjusted data here.

Table 6 demonstrates that filter selection based on cross-validation is generally less stable than selection using traditional rules. For example, the AMPE rule achieves a stability rate of 80.9% based on a restricted set of candidate filters, whereas cross-validation based on the B3 table – the most stable among the cross-validation approaches – recommends the same filter after adding one year of data in 75.3% of cases. The difference is

	Cross-validation based on X–11 table								
	B3	B8	C4	C9	D4	D8	D8b	AMPE	MSR
Relative MAE	105.8	105.7	105.8	106.8	105.8	105.8	107.5	105.4	108.6
Deviance	0.8	0.8	0.8	0.9	0.8	0.8	1.0	0.8	1.0
Bias	0.1	0.0	0.0	−0.2	0.1	−0.1	−0.2	−0.1	−0.3

Table 5: Additional evaluation metrics (see details in Table 3)

even more pronounced when comparing cross-validation to the MSR rule: in the scenario where only the S3×3, S3×5 and S3×9 filter can be selected, MSR maintains its choice in 94.3% of cases, while the best-performing cross-validation approaches yield stabilities of approximately 82%.

The high stability observed in the MSR rule is intrinsic to its algorithmic design, specifically the use of wide indecisive zones between cutoff values and the iterative re-estimation protocol. As discussed in Section 2.1, when the MSR falls between cutoffs (i.e., in the gaps), X–11 omits up to five years of data and recalculates the statistic on the shortened series; if it remains between cutoffs, S3×5 is selected. Consequently, when a new year is appended, the truncation step effectively down-weights the influence of the added observations, and the selected filter typically remains unchanged unless the MSR moves decisively across a cutoff into another filter's range. This design feature contributes considerably to the high stability reported for MSR.

Among the cross-validation results, B3 stands out as the best-performing table in terms of stability, when all filters or all except the S3×1 are available for selection. When the selection is restricted to all filters except S3×1 and S3×15, the D4 table performs slightly better. Overall, the differences in stability between the various tables are mostly small.

	Cross-validation based on X–11 table								
	B3	B8	C4	C9	D4	D8	D8b	AMPE	MSR
all	74.0	72.0	72.5	70.1	72.6	71.0	64.1	–	–
all w/o S3×1	75.3	73.8	74.4	72.5	74.5	72.8	68.0	80.9	–
all w/o S3×1, S3×15	81.7	81.6	81.8	79.3	82.2	80.6	73.0	–	94.3

Table 6: Stability analysis based on real-world data.

Selecting the ideal X–11 table as the basis for cross-validation Cross-validation generally performs well across all X–11 tables, allowing users flexibility in their choice of table. However, we recommend the use of the B3 table for several reasons. First, cross-validation based on the B3 table demonstrates strong performance across all evaluation metrics, including accuracy, relative MAE, and, in particular, stability. Additionally, the selection is not particularly biased towards any specific filter. Second, the estimate of the seasonal-irregular component in the B3 table is independent of all user-defined X-11 settings – most notably, the choice of Henderson filter, which affects the trend estimate, and the sigma limits, which influence the detection of extreme values. Third, when the selection is based on the B3 table, the chosen seasonal filter can be consistently applied to all subsequent tables, ensuring that the final seasonally adjusted series remains unchanged, even if the user explicitly specifies that filter.

In contrast, the MSR rule selects the seasonal filter based on the D8b table. If the user subsequently fixes this filter, it will be applied to all relevant tables, resulting in a seasonally adjusted series that may differ slightly from the automatic mode. In automatic mode, all estimates of the seasonal component use the default filters, with only the final estimation employing the filter selected by the MSR rule. The same applies if cross-validation is based on any table later than B3.

		Industrial Production	Retail sales	HICP: clothes and shoes	Deflator of orders received (NACE 28.15)	Registrations of new cars
Cross-validation (B3 table)	S3×1	12.75	11.00	5.95	2.55	51.14
	S3×3	12.65	10.09	5.87	2.55	49.85
	S3×5	12.24	10.02	5.97	2.52	49.02
	S3×9	11.62	10.04	6.24	2.47	48.20
	S3×15	11.54	10.76	6.68	2.37	48.72
	Choice	S3×15	S3×5	S3×3	S3×15	S3×9
AMPE	θ	0.36	0.71	0.40	0.00	0.67
	Θ	0.71	0.56	0.34	0.87	0.65
	Choice	S3×9	S3×5	S3×3	S3×15	S3×5
MSR	Statistic	5.23	6.03	3.27	6.22	4.96
	Choice	S3×5	S3×5	S3×3	S3×9	S3×5
Visual inspection		S3×9	S3×5	S3×3	S3×15	S3×9

Note: Cross-validation statistics (RMSE) have been multiplied by 1000 for readability. Visual inspection is based on the intersubjective assessment of SI ratios (see, e.g., Figure 3). For HICP: clothes and shoes and the deflator, the model was simplified to an ARIMA (0,1,1)(0,1,1) model.

Table 7: Cross-validation applied to a set of economic time series

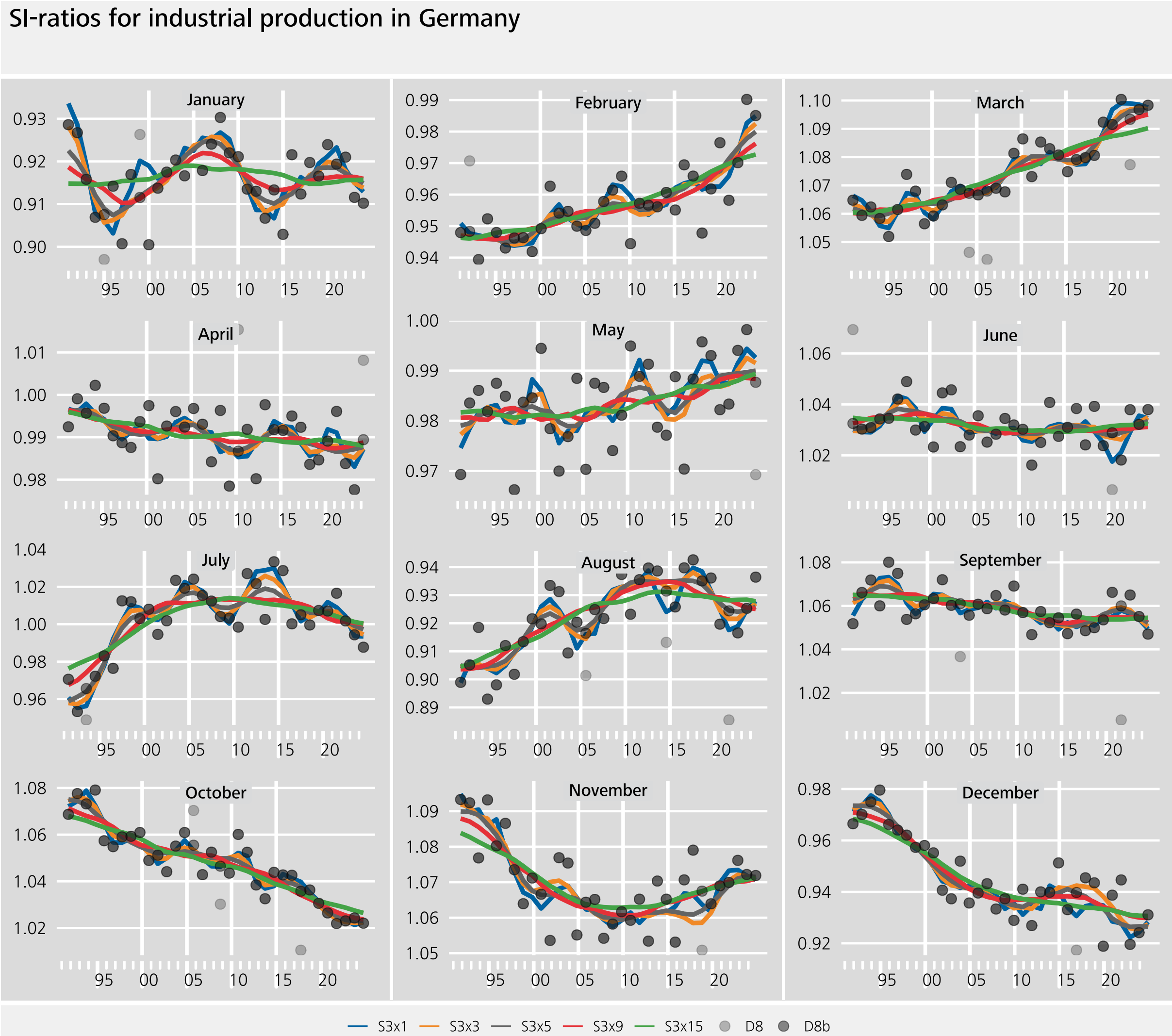
4.3 Real-world examples

To illustrate the use of cross-validation in a real-world application, we selected five monthly German time series representing different sectors of the economy. These series include both aggregate and component series, as well as varying time spans. Specifically, we analyse: industrial production (Jan 1991 – Dec 2024), turnover in retail trade (Jan 2017 – Aug 2025), the HICP for clothes and shoes (Jan 2008 – Feb 2025), the deflator of domestic orders received in manufacture of bearings and gears (NACE 28.15, Jan 1991 – Aug 2025) and the registration of new cars (Jan 1991 – Oct 2025).

To obtain the estimates, we model these series using X–13–ARIMA in JDemetra+ (version 3.4.1). Where necessary, we include calendar variables for working days, trading days and/or moving holidays, as well as outlier regressors – particularly for the financial crisis in 2008-09 and the Covid-19 pandemic. Cross-validation is applied to the B3 table, that is, the linearised and preliminarily trend-adjusted series. All five seasonal moving averages are considered. For the shortest series, retail sales, the S3×15 filter becomes a stable filter.

To obtain airline model parameter estimates (AMPE), we use the estimates from the pre-processing stage. For the deflator of orders received and the HICP series, we needed to simplify the ARIMA model used in production to the SARIMA(0,1,1)(0,1,1) model. For all other series, the airline model was already employed.

JDemetra+ reports the MSR for each individual month; the MSR statistic reported in the table is the mean across all months. The selected filter is the one chosen in JDemetra+



Note: The raw seasonal components w/o and with additional extreme value replacement (D8 and D8b, resp.) are based on the estimation with a S3x9 filter.
Deutsche Bundesbank

Figure 3: SI-ratios for selected months for industrial production in Germany

when the automatic choice is based on MSR.

Visual inspection is based on SI plots. A version with all estimated seasonal components derived from different seasonal filters for industrial production is included in Figure 3; excerpts for all other series are included in the appendix. The assessments are intersubjective, conducted independently by multiple practitioners.

The collected results are shown in Table 7.

For some series the smallest and second (and sometimes third) smallest cross-validation error statistics are very close, as for industrial production and retail sales. Consequently, adding a few observations can change the selected filter. The same holds for the AMPE criterion. In particular, for $\Theta \approx 0.6$, where the S3×5 filter is always selected, the margin to parameter values that would recommend a different filter is narrow. This is evident for retail sales and registrations of new cars, both of which lie close to a decision boundary.

As discussed in Section 2.1, the MSR criterion mitigates such instability by using non-adjacent cut-offs, re-estimating on a shortened span when the statistic falls between cut-offs, and, if no clear decision emerges, defaulting to S3×5. As discussed above, MSR strongly favours S3×5 filters by design (see also Table 9). A similar, but more balanced, stabilisation could be incorporated into the cross-validation procedure to improve robustness.

Relative to visual inspection cross-validation generally selects sensible filters. The main exception is industrial production, where S3×15 is selected and appears overly smooth for several months (e.g. January and March). A pragmatic remedy is to constrain the admissible set (e.g. cap K at 9). While month-specific cross-validation metrics can be computed, this reduces temporal coherence and stability.

5 Summary

Cross-validation is an effective method for selecting seasonal moving averages in X–11. For daily and high-frequency time series, the fixed set of standard X–13–ARIMA filters is often insufficient (cf. Ollech, 2021, 2023). Cross-validation's flexibility has a practical advantage for these kinds of time series, because the framework readily accommodates additional candidate filters.

Future research should investigate how to stabilise filter selection when new observations are added – at least in situations where the filter does change appropriately. Such stabilisation mechanisms may include: (a) the requirement of a minimum improvement threshold before switching filters; (b) breaking near-ties in favour of the shorter filter; (c) re-running the selection on a shortened span when the margin is below the threshold, or (d) when ties persist, default to the most commonly used filter, such as the S3×9. These steps would increase stability without affecting cases where one filter clearly outperforms others.

This kind of stabilisation mechanism may also be adopted to allow cross-validation to be used to select period-specific filters. Although this is computationally feasible, the limited number of observations within individual periods typically prevents reliable period-specific choices at present.

References

Arlot, S. and A. Celisse (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys 4*, 40–79.

Bandara, K., R. J. Hyndman, and C. Bergmeir (2025). MSTL: A seasonal-trend decomposition algorithm for time series with multiple seasonal patterns. *International Journal of Operational Research 51*(1), 79–98.

Bottegal, G. and G. Pillonetto (2018). The generalized cross validation filter. *Automatica 90*, 130–137.

Craven, P. and G. Wahba (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische mathematik 31*(4), 377–403.

Cuevas, Á. and E. M. Quilis (2023). Seasonal adjustment methods for daily time series. a comparison by a monte carlo experiment.

Dagum, E. B. (1999). *The X11ARIMA/2000 Seasonal Adjustment Method. Foundations and User´s Manual.*

Dagum, E. B., N. Chhab, and B. Solomon (1991). The autocorrelation of residuals from the X11ARIMA method. *Journal of Official Statistics 7*(2), 181.

Depoutot, R. and C. Planas (1998). Comparing seasonal adjustment and trend extraction filters with application to a model-based selection of X11 linear filters. *Tech. Paper TP361, Joint Research Centre, Ispra, Italy*.

Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto, and B.-C. Chen (1998). New capabilities and methods of the X-12-ARIMA seasonal-adjustment program. *Journal of Business & Economic Statistics 16*(2), 127–152.

Friedman, J. H. (1984). A variable span smoother. Technical Report 5, Laboratory for Computational Statistics, Stanford University.

Golub, G. H., M. Heath, and G. Wahba (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics 21*(2), 215–223.

Grudkowska, S. (2017). *JDemetra+ Reference Manual* (2.2 ed.). Department of Statistics, Narodowy Bank Polski.

Haller, V., S. Daniel, and B. Bellone (2025). STAHL: Seasonal, trend, and holiday decomposition with loess. *Journal of Official Statistics 41*(4), 1112–1153.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning – Data Mining, Inference, and Prediction* (Second ed.). Heidelberg: Springer.

Henderson, R. (1916). Note on graduation by adjusted average. *Transactions of the actuarial society of America 17*(55), 43–8.

Kirchner, R. (1999). Auswirkungen des neuen Saisonbereinigungsverfahrens Census X-12-ARIMA auf die aktuelle Wirtschaftsanalyse in Deutschland. (7), 1–107.

Kirchner, R., D. Ladiray, and G. L. Mazzi (2018). Quality measures and reporting for seasonal adjustment. In G. L. Mazzi, D. Ladiray, and D. A. Rieser (Eds.), *Handbook on Seasonal Adjustment*, Chapter 22, pp. 591–627. Publications Office of the European Union, Luxembourg.

Ladiray, D. and B. Quenneville (2001). *Seasonal Adjustment with the X-11 Method*, Volume 158 of *Lecture Notes in Statistics*. Springer.

Leung, D. H.-Y. (2005). Cross-validation in nonparametric regression with outliers. *Annals of Statistics 33*(5), 2291–2310.

Lothian, J. (1984). The identification and treatment of moving seasonality in the X-11 seasonal adjustment method. In *Proceedings of the Business and Economic Statistics Section*, pp. 166–171.

Lukas, M. A., F. R. de Hoog, and R. S. Anderssen (2016). Practical use of robust GCV and modified GCV for spline smoothing. *Computational Statistics 31*(1), 269–289.

Marris, S. N. (1960). The treatment of moving seasonality in census method II. In *Seasonal Adjustment on Electronic Computers*, pp. 257–309. Organisation for Economic Co-operation and Development, Paris.

Musgrave, J. C. (1964). A set of end weights to end all end weights. Technical report.

Ollech, D. (2021). Seasonal adjustment of daily time series. *Journal of Time Series Econometrics 13*(2).

Ollech, D. (2023). Economic analysis using higher-frequency time series: Challenges for seasonal adjustment. *Empirical Economics 3*, 1375–1398.

Paige, R. L. and A. A. Trindade (2010). The Hodrick-Prescott filter: A special case of penalized spline smoothing. *Electronic Journal of Statistics 4*, 856–874.

Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of econometrics 99*(1), 39–61.

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

A Appendix

A.1 Further tables and graphics

		Cross-validation based on X–11 table								
		B3	B8	C4	C9	D4	D8	D8b	AMPE	MSR
Overall		42.4	43.1	41.5	39.1	40.2	44.0	33.2	37.4	26.8
Series length	5y	26.0	27.8	26.6	27.6	26.4	28.6	27.8	25.1	20.1
	10y	39.4	40.3	38.6	36.0	38.9	40.0	30.6	33.6	24.3
	15y	48.2	48.8	46.5	42.6	45.9	48.8	37.2	41.0	27.0
	20y	44.2	45.4	43.1	42.1	42.4	46.5	35.4	41.0	29.0
	25y	48.8	47.3	47.4	42.4	44.8	50.0	33.9	42.2	30.1
	30y	48.0	49.0	46.4	44.1	42.9	50.2	34.2	41.6	30.1

Table 8: Accuracy (%) of identifying the optimal S3×K filter by time series length

	Cross-validation based on X–11 table								
	B3	B8	C4	C9	D4	D8	D8b	AMPE	MSR
S3×1	23.4	23.2	21.5	25.9	19.3	26.1	18.9	0.0	0.0
S3×3	15.0	16.7	16.4	19.2	16.1	16.2	22.5	60.0	34.9
S3×5	12.9	14.8	15.3	19.2	16.4	14.7	29.2	11.8	62.2
S3×9	26.3	24.8	30.2	25.2	32.0	23.3	23.4	6.6	2.9
S3×15	22.3	20.4	16.6	10.5	16.2	19.7	6.0	21.7	0.0

Table 9: Filter selection frequency

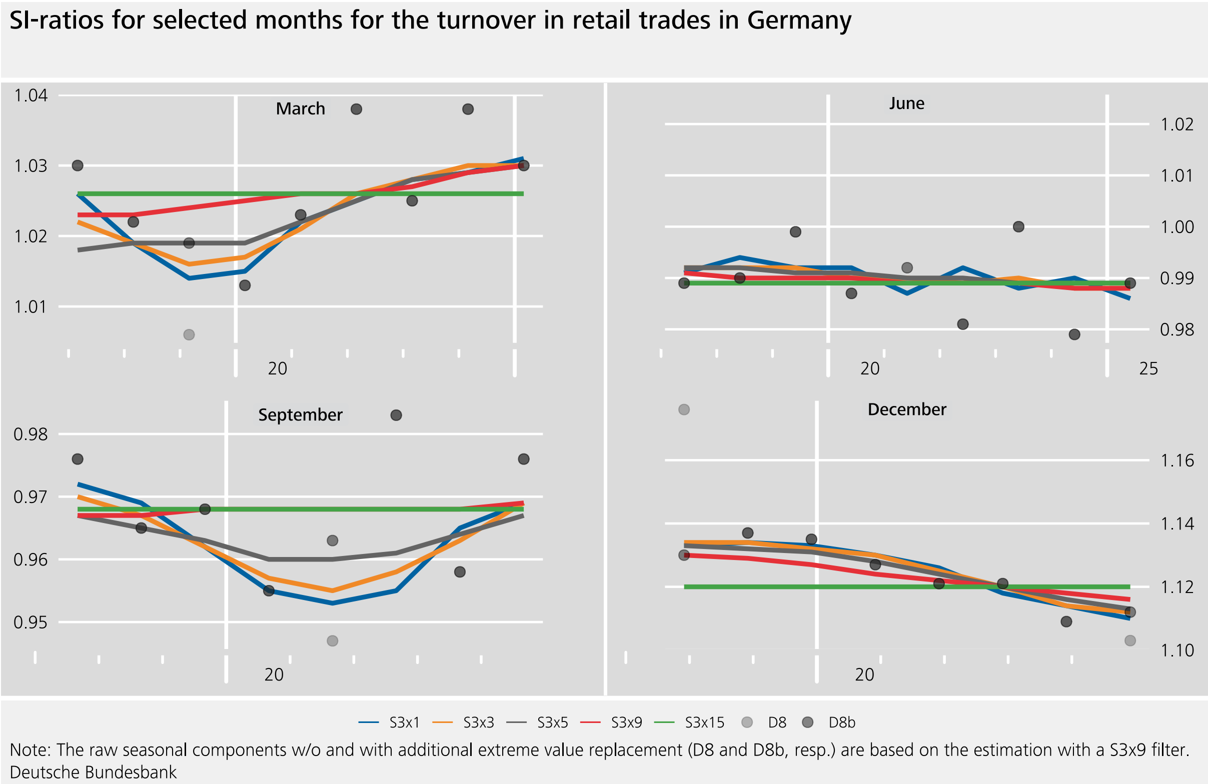


Figure 4: SI-ratios for selected months for the turnover in retail trades in Germany

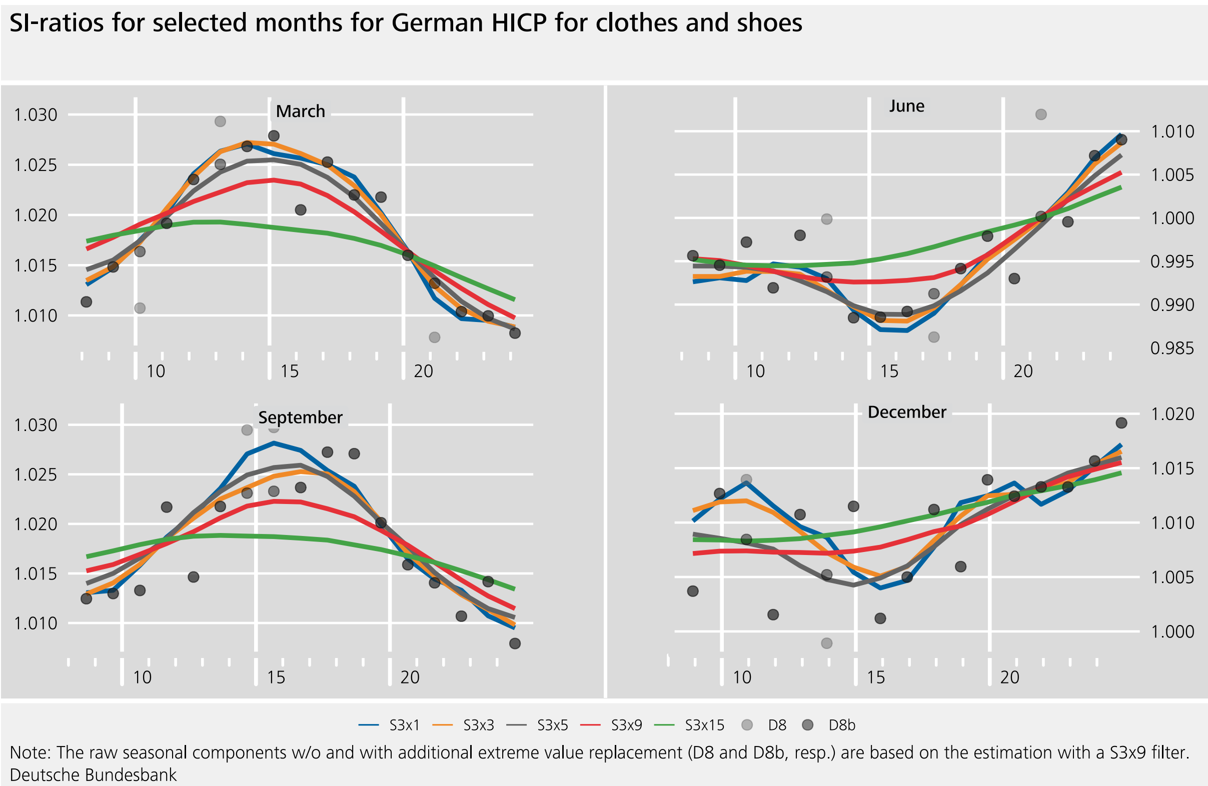


Figure 5: SI-ratios for selected months for German HICP for clothes and shoes

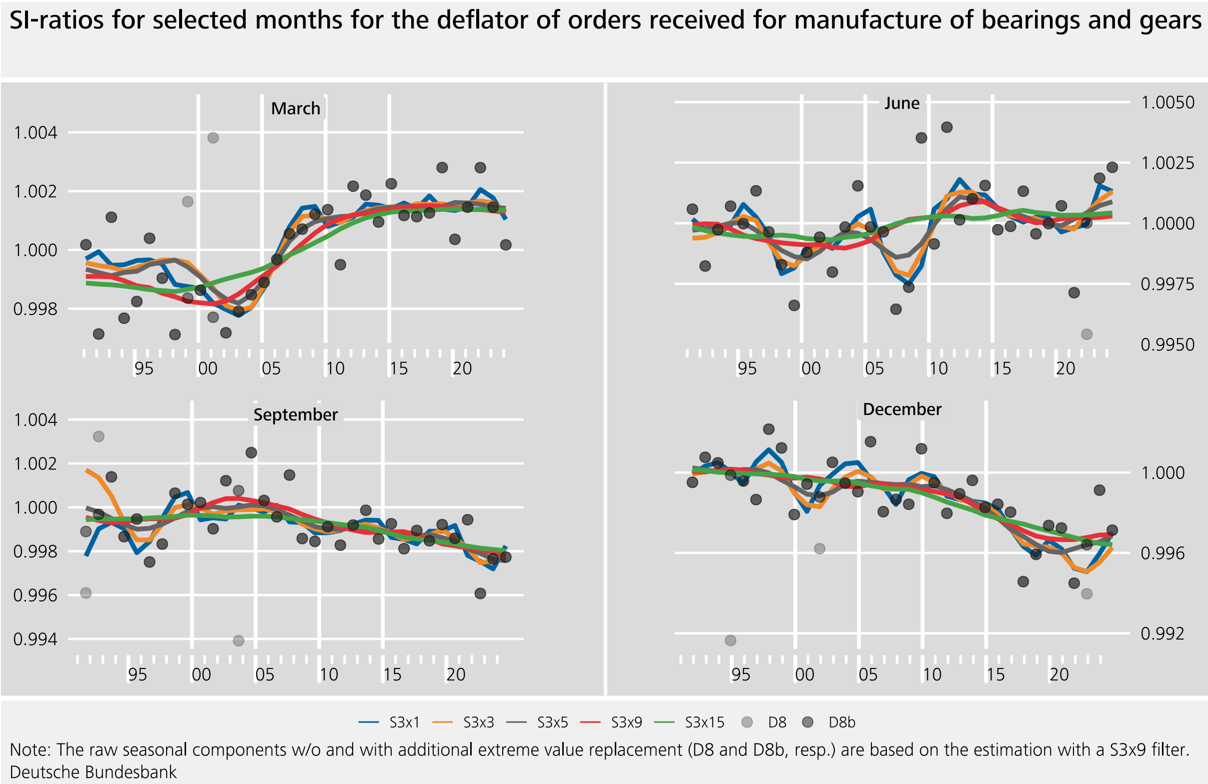


Figure 6: SI-ratios for selected months for the deflator of orders received for manufacture of bearings and gears

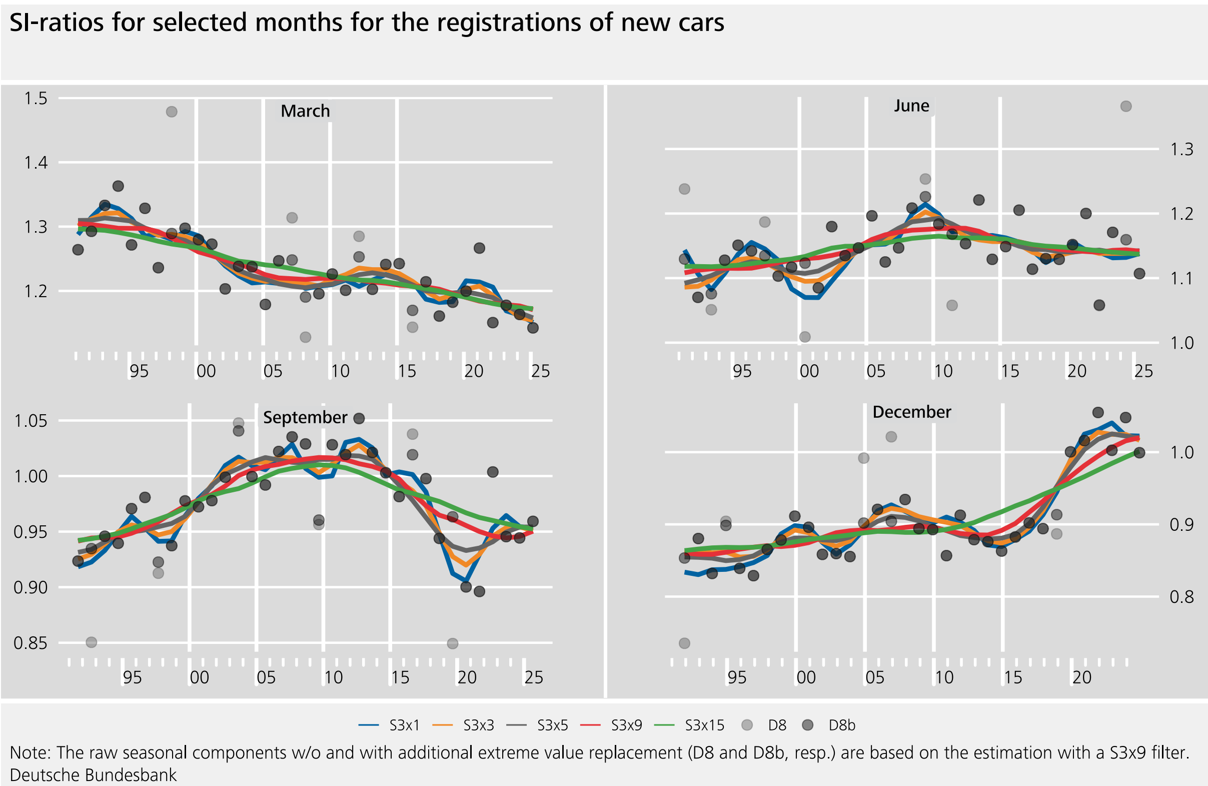


Figure 7: SI-ratios for selected months for registrations of new cars