

Seasonal Adjustment Algorithm

Kirill Spiridonov

Saratov State University, Saratov, Russian Federation,
kir.spiridonov@gmail.com

Abstract—The paper presents an algorithm for smoothing time series with seasonality. The proposed method is based on median smoothing and the Hodrick–Prescott decomposition. Using a software implementation in the R language, the correctness of the developed algorithm is checked; it is also compared with other seasonal smoothing methods.

Index Terms—Seasonal adjustment, Median adjustment, Hodrick–Prescott Filter

I. INTRODUCTION

The dynamics of some economic processes in the same time periods in different years may be similar, that is, have seasonality. The presence of seasonality in the time series can complicate the analysis and even lead to incorrect conclusions. And when studying the dependencies between economic processes, the presence of a similar seasonal component in several completely unrelated series may lead to the erroneous conclusion that there is a stable, statistically significant relationship between these processes. Thus, the task of eliminating seasonality is very important in the analysis of time series.

In the general case, the time series y_t consists of the trend-cycle – tc_t , seasonal – s_t and irregular – i_t components. Some authors (for example, [1] and [2]) separate the trend-cycle component that corresponds to the potential level of a certain economic indicator and the cyclical component associated with fluctuations in the economic situation. However, in the present work, the difference between the trend and the trend-cyclic component is bypassed. In addition, most modern seasonal smoothing algorithms also do not allocate the cycle into a separate component.

By the method of component aggregation, time series are additive

$$y_t = TR_t + S_t + I_t$$

and multiplicative

$$y_t = TR_t \cdot S_t \cdot I_t.$$

Most economic processes have a multiplicative structure. The use of the additive model is advisable in the case when as the level of the series increases, the amplitude of seasonal fluctuations is constant.

The exponential smoothing, proposed by Brown [3], has been one of the first methods for smoothing time series. This method is applicable for series fluctuating around a certain constant level, without trends and seasonal components. However, the presence of a structure in time series in some way changes the state of affairs. For example, for series with

a linear trend, Brown showed [4] that exponential smoothing does give not satisfactory results.

To smooth time series with a linear trend, the Brown method [5] and the Holt method [6] have proven themselves well. The described methods aren't suitable for smoothing time series with seasonality. Winters proposed a modification of the Holt method [7], taking into account the presence of a multiplicative seasonal component in the series. For the additive factor, there is another modification of the Holt method, called the Theil–Wedge method [8].

The most modern time series smoothing algorithms also take into account the structure. These algorithms are based on the decomposition of the initial series into a trend, seasonal and irregular components. Thus, the smoothed series is a combination of a trend and a random component. The most popular seasonal adjustment algorithms are the non-parametric X-series algorithms ([9]–[11]) and the parametric TRAMO/SEATS method [12].

II. SEASONAL ADJUSTMENT ALGORITHM

Before the description of the seasonal adjustment procedure, we note that the algorithm involves working only with monthly and quarterly data.

Suppose that we have a time series – x_1, \dots, x_n , then the number of full years for which there are observations can be defined as

$$T = \left[\frac{n}{p} \right] + \gamma, \quad p = 4, 12,$$

where $[x]$ – the integer part of a number x , and

$$\gamma = \begin{cases} 0, & n \bmod p = 0; \\ 1, & n \bmod p = 1. \end{cases}$$

A. Data Transformation

Depending on the aggregation of the main components of the time series, the series can have an additive or multiplicative structure. Working with the additive model seems more preferable because the seasonal wave becomes more uniform and does not evolve over time.

Let x_1, \dots, x_n be the time series with a multiplicative structure, then the transition to the additive model y_1, \dots, y_n is carried out

$$y_t = \ln(x_t + |\alpha|), \text{ where}$$

$$\alpha = \begin{cases} \min_t(x_t) - 1, & \text{if } \exists t : x_t \leq 0; \\ 0. & \end{cases}$$

Thus, at the first step, data linearization is carried out in order to move to an additive model

$$y_t = C_t + S_t + I_t.$$

It is also worth noting that if the series has an additive structure, then data conversion is not required.

B. Hodrick–Prescott Filter

The Hodrick–Prescott filter is used to remove the cyclical component of a time series. This method was first used by Hodrick and Prescott to analyze post-war US business cycles [13].

A filter is a solution to an optimization problem. On the one hand, the smoothed series should not differ much from the original series, that is, it is necessary to minimize the sum of the squared deviations.

$$\sum_{t=1}^n (y_t - L_t)^2 \longrightarrow \min.$$

On the other hand, the series should be smooth enough and change not too sharply

$$\sum_{t=2}^{n-1} ((L_{t+1} - L_t) - (L_t - L_{t-1}))^2 \longrightarrow \min.$$

Thus, the filter is a two-sided linear filter that calculates the smoothed series L_t of the time series y_t by minimizing the difference of the series L_t and y_t , and minimizing the sum of the second differentiated series L_t

$$\sum_{t=1}^n (y_t - L_t)^2 + \lambda \sum_{t=2}^{n-1} ((L_{t+1} - L_t) - (L_t - L_{t-1}))^2 \longrightarrow \min.$$

The parameter $\lambda \geq 0$ is a measure of the smoothness of the resulting series. For $\lambda = 0$, the smoothed series exactly matches the original one. If $\lambda \rightarrow +\infty$, then the series approaches a linear trend. In the literature, there are different approaches to the definition of λ (for example, [14]), however, the following values are considered a “classic” variant [15]

- $\lambda = 100$, for annual data;
- $\lambda = 1600$, for quarterly data;
- $\lambda = 14400$, for monthly data.

As an alternative, the Baxter–King filter was considered. However, using this method, the observations at the ends of the sample fall out of consideration. Therefore, the results of this filter are not presented in the work. The asymmetric Christiano–Fitzgerald filter does not have this drawback, but an additional step is required before its application – checking the series for the presence of unit roots, since the filter works differently on stationary and integrated series.

Thus, the use of the Hodrick–Prescott filter is related to its simplicity (in fact, it is only necessary to set the parameter λ) and the lack of the need for preliminary work with the series.

C. Estimate of Seasonal Component

Let’s apply the Hodrick–Prescott filter to y_t . Now we get the first estimate of the trend L_t^1 and subtracting it from the original series, we get the sum of the seasonal and irregular components

$$S_t + I_t = y_t - L_t^1.$$

Now it is needed to separate the irregular component from the seasonal wave. For this, it is necessary to group seasonal components from $S_t + I_t$ by seasonality periods. We have p distributions of seasonal coefficients s_i^1, \dots, s_i^T , where $i \in [1, p]$. Meanwhile, in some sets the coefficient of the last period T may be absent.

Each of the series s_i^1, \dots, s_i^T is smoothed by a moving median, with a window width of three. In many algorithms, the first estimate of seasonality is obtained using a moving (weighted) average. Compared with the average, the median estimate of the distribution center is more robust, which allows us to more effectively exclude outliers from consideration. Thus, at this stage, we have the median smoothed coefficients S_t^1 .

A moving window with a width of three observations is used for several reasons. Firstly, such a window width filters out errors well, while the series does not become too smooth (the same window width is used, for example, in some X-series algorithms). Secondly, already for relatively short time series (from 4 years) you can get a quality estimate of seasonality.

Since this approximation of the seasonal wave is rather rough, it must be supplemented with errors limited to a certain value. In addition, in the absence of outbursts inside the window, the smoothed curve will have a “jagged” look, so it needs to be modified. To do this, we form a series of errors

$$e_t = S_t + I_t - S_t^1$$

and calculate the standard deviation σ for it. We compose a new series of errors from such members e_t that do not exceed 2σ , and instead of those that exceed, we use 2σ :

$$\hat{e}_t = \begin{cases} e_t, & e_t \leq 2\sigma; \\ 2\sigma, & e_t > 2\sigma. \end{cases}$$

Let us add the resulting series \hat{e}_t to the median-smoothed series S_t^1 and smooth it with a third-order moving average in three different ways:

- 1) The window consists of the current and 2 previous observations;
- 2) The window consists of the current and 2 subsequent observations;
- 3) The window consists of the current, previous and subsequent observations.

Thus, we get three smoothed series S_t^{21} , S_t^{22} and S_t^{23} . As the final estimate of the seasonal wave, we use the average value of the obtained series, i.e.

$$S_t^2 = \frac{S_t^{21} + S_t^{22} + S_t^{23}}{3}.$$

Such a construction of a series is dictated by two conditions: firstly, points at the ends of the original series do not drop out of consideration; secondly, obtained series has less volatility.

At this stage, the obtained estimate of the seasonal component is, as a rule, quite adequate and can be used for further processing of the series. However, in some cases, seasonality assessment may have high volatility. This situation is observed in disaggregated time series. In fact, at the next stage, it is necessary to reduce the volatility of the converted series, if it is excessive.

For each of the p distributions of the transformed seasonal coefficients $\hat{s}_i^1, \dots, \hat{s}_i^T$, where $i \in [1, p]$, we calculate the variance $\hat{\sigma}_i^2$ (we will use the variance as a measure of the volatility of the series). To reduce the volatility of a set of seasonal coefficients, we can use the following approach: if some seasonal period has high variance, then it is smoothed out more. A measure of the smoothing force of a series can be the width of a sliding window or the distribution of weights inside it. If the series have high volatility, the more observations will be used for smoothing.

Now, for definiteness, we fix some $i \in [1, p]$ and $t \in [1, T]$ and consider only one seasonal coefficient \hat{s}_i^t . Each such coefficient will be replaced by a certain smoothed value, depending on the variance of the set of coefficients in which it is included, i.e.

$$\tilde{s}_i^t = \sum_{j=1}^T \hat{w}_{it}^j(\hat{\sigma}_i^2) \hat{s}_i^j, \quad (1)$$

where

$$\sum_{j=1}^T \hat{w}_{it}^j(\hat{\sigma}_i^2) = 1 \quad \forall i, t \quad \text{and} \quad \hat{w}_{it}^j(\hat{\sigma}_i^2) \geq 0 \quad \forall i, t, j, \quad (2)$$

$$\hat{w}_{it}^j(\hat{\sigma}_i^2) = \frac{w_{it}^j(\hat{\sigma}_i^2)}{\sum_{j=1}^T w_{it}^j(\hat{\sigma}_i^2)},$$

$$w_{it}^j(\hat{\sigma}_i^2) = \frac{1}{\sqrt{2\pi\hat{\sigma}_i^2}} \exp\left(-\frac{(j-t)^2}{2\hat{\sigma}_i^2}\right).$$

The dependence of the smoothing weights on the variance is dictated by the condition for reducing the series volatility: the higher variance, the slower weights decrease to zero, and therefore the series becomes smoother.

It is worth noting that when smoothing a series using these weights, a problem arises: observations that are closer to the ends of the set of seasonal coefficients, as a result of smoothing, strongly depend on their initial value. The solution to this problem is that in each of the p distributions of seasonal coefficients for the weights of each seasonal coefficient \hat{s}_i^t , the weight corresponding to the smoothed coefficient should be replaced by the minimum weight in this set, i.e.

$$\hat{w}_{it}^j(\hat{\sigma}_i^2) = \min_t(\hat{w}_{it}^j(\hat{\sigma}_i^2) : j = t) \quad \text{for} \quad j = t.$$

After that, the new weights are normalized again, as in the first part of (2). Finally, using the new weights, all seasonal coefficients are recalculated using (1).

Further, we will no longer work with each individual seasonal period, but we will work with a seasonal wave within each year under consideration. To filter out the remaining seasonal errors, we are smoothing out the seasonal components of each year with a third-order moving average in three different ways, as described above. As a result, we obtain three smoothed series (S_t^{31} , S_t^{32} and S_t^{33}), and use the average as the final estimate of the seasonal wave value from obtained series.

$$S_t^3 = \frac{S_t^{31} + S_t^{32} + S_t^{33}}{3}.$$

Note that this procedure is carried out only for series with a monthly frequency. For quarterly data, this step is omitted, because otherwise the final series would be too smooth, and the seasonal estimate would be distorted.

Now it remains to normalize the seasonal component. In the case of the additive model, the sum of seasonal coefficients during the year is 0, and in the case of the multiplicative model, their product is 1:

- for the additive model from each coefficient within one year, subtract their average value;
- for the multiplicative model, we divide each coefficient within one year by their geometric mean.

Thus, we have the normalized seasonal coefficients \tilde{S}_t^3 , and a smoothed series ($y_t - \tilde{S}_t^3$). The final trend estimate can be obtained by applying a Hodrick–Prescott filter to the smoothed series.

III. PRACTICAL IMPLEMENTATION

We use R language for the implementation of the seasonal adjustment algorithm. The choice of this software product is due to several factors. Firstly, at the moment, R is one of the most popular and most powerful languages for statistical calculations and modeling, while its syntax is quite simple. Secondly, due to its high popularity, there are a large number of user libraries that implement additional functionality and allow you to solve almost any task in a few simple steps.

To develop the seasonal smoothing script, we used both standard functions and procedures of the R language [16], as well as functions developed by third-party users. In particular, the functions included in the freeware packages `mFilter` [17] and `psych` [18] were used.

Several series of macroeconomic indicators were used to compare the smoothing algorithms. The information base of the study is the resources of the Federal State Statistics Service. Several series were smoothed out: retail turnover of non-food products, consumer price index for sugar, unemployment rate, volume of investments in fixed assets. The indicators were selected in such a way as to check all the options for the algorithm to work (on monthly and quarterly data, with and without logarithmic transformation).

The TRAMO/SEATS and X-12-ARIMA algorithms were chosen as alternative methods of seasonal smoothing to compare and verify the correctness of the developed algorithm. The choice of these methods is due to their popularity: both algorithms are integrated into most modern econometric

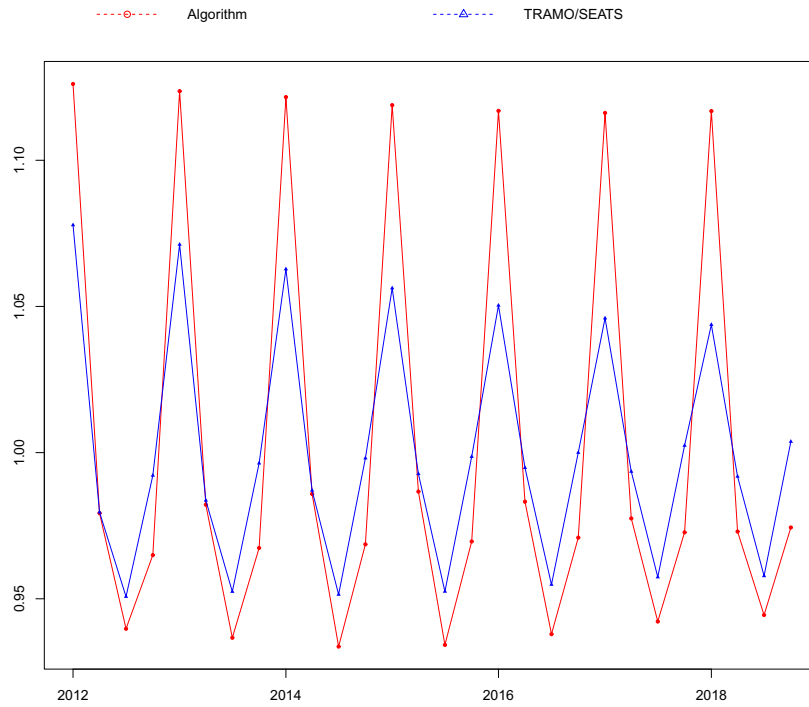


Fig. 1. Seasonal component: unemployment rate.

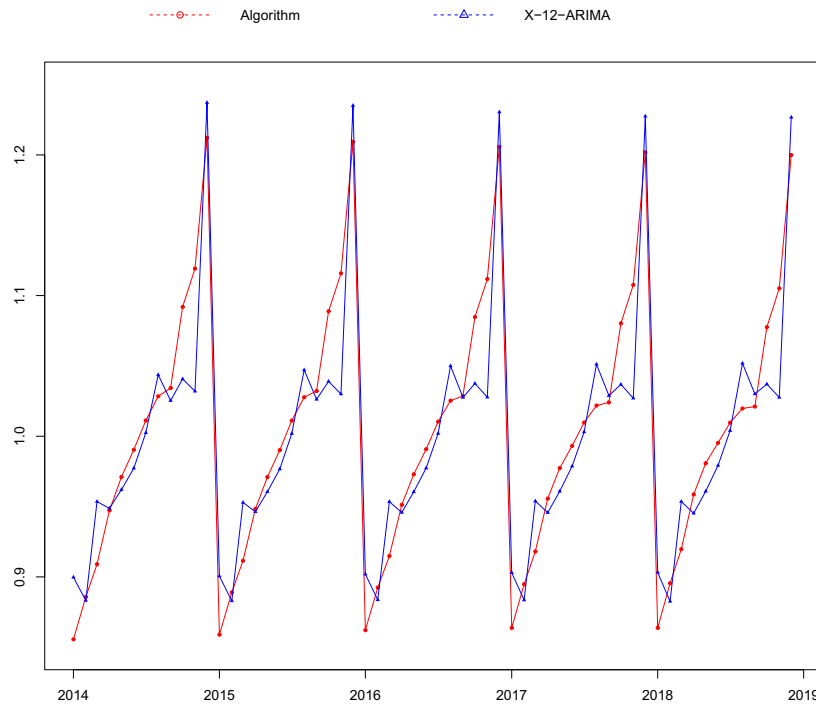


Fig. 2. Seasonal component: retail turnover.

packages and are used by many organizations for the seasonal smoothing procedure.

The results of the comparison of the algorithms are shown in the Fig. 1 and 2. In general, it can be noted:

- the seasonal component obtained using the developed algorithm is similar to the seasonality identified by TRAMO/SEATS and X-12-ARIMA: the peaks of the seasonal wave fall on the same periods; trends observed

over one year are similar; the amplitude of the oscillations and some evolution of the wave in time are also consistent;

- the difference lies in the uniformity of the seasonal wave. In almost all cases, TRAMO/SEATS shows a more uniform result, which is due to the principle of its construction: when smoothing any series, the algorithm tries to minimize the evolution of seasonality over time. In cases where seasonality has evolved, the result of the developed algorithm is closer to X12-ARIMA.

On the one hand, the stability of the seasonal factor, like that of TRAMO/ SEATS, looks attractive, but on the other hand, the evolution of economic processes under the influence of many factors is ongoing (including the evolution of the seasonal component) and in this case the requirement for the stability of the seasonal component is quite strict. The result of the X-12-ARIMA in some cases, on the contrary, allows for too much seasonal variation. The developed algorithm in the sense of seasonal factor stability is located between these two smoothing methods and can be used for working with data.

Another advantage of the developed algorithm is that in the process of smoothing median smoothing is used, which allows to get rid of outliers and random errors that may be present in the data. This eliminates the need to separately model outliers (for example, using dummy variables) and generally simplifies series processing, since in some cases the quality of seasonal smoothing of the TRAMO/SEATS and X-12-ARIMA algorithms significantly depends on the selected outliers.

ACKNOWLEDGMENTS

This research was funded by Russian Science Foundation, grant number 19-18-00199.

REFERENCES

- [1] Lukashin, Y.P., *Adaptive methods for time series short term forecasting*. Finance and Statistics, 2003.
- [2] Nosko, V.P., *Econometrics. Book 1. Part 1 and 2*. Delo, 2011.
- [3] R. Brown, *Exponential Smoothing for Predicting Demand*. Little, 1956.
- [4] C. Lewis, *Industrial and Business Forecasting Methods: A Practical Guide to Exponential Smoothing and Curve Fitting*, ser. Butterworth scientific. Butterworths Scientific, 1982.
- [5] R. Brown, *Smoothing, forecasting and prediction of discrete time series*, ser. Prentice-Hall international series in management. Prentice-Hall, 1963.
- [6] C. Holt, "Forecasting trends and seasonal by exponentially weighted averages," *Office of Naval Research Memorandum*, vol. 20, 01 1957.
- [7] P. R. Winters, "Forecasting sales by exponentially weighted moving averages," *Management Science*, vol. 6, pp. 324–342, 04 1960.
- [8] H. Theil and S. Wage, "Some observations on adaptive forecasting," *Management Science*, vol. 10, pp. 198–206, 01 1964.
- [9] J. Shiskin, A. H Young, and J. C Musgrave, "The x-11 variant of the census method ii seasonal adjustment program," 01 1967.
- [10] *X-12-ARIMA Reference Manual*, U.S. Census Bureau, Washington, DC, 2011.
- [11] *X-13-ARIMA-SEATS Reference Manual*, U.S. Census Bureau, Washington, DC, 2015.
- [12] V. G'omez and A. Maravall, "Programs TRAMO and SEATS: instructions for the user," *Mimeo, Banco de España*, 01 1997.
- [13] R. Hodrick and E. Prescott, "Post-war US business cycles: An empirical investigation," vol. 8, pp. 231–247, 01 1981.
- [14] M. Ravn and H. Uhlig, "On adjusting the HP filter for the frequency of observations," *The Review of Economics and Statistics*, vol. 84, 02 1997.
- [15] C. Favero, *Applied Macroeconometrics*. Oxford University Press, 2001.
- [16] R. Kabacoff, *R in Action: Data Analysis and Graphics with R*. Greenwich, CT, USA: Manning Publications Co., 2015.
- [17] M. Balcilar, *mFilter: Miscellaneous time series filters*, 2007, r package version 0.1-3. [Online]. Available: <https://CRAN.R-project.org/package=mFilter>
- [18] W. Revelle, *psych: Procedures for Psychological, Psychometric, and Personality Research*, Northwestern University, Evanston, Illinois, 2018, r package version 1.8.12. [Online]. Available: <https://CRAN.R-project.org/package=psych>