



Seasonality in COVID-19 times

Juan Bógalo^a, Martín Llada^{b,*}, Pilar Poncela^a, Eva Senra^c

^a Universidad Autónoma de Madrid, Spain

^b IIEP-BAIRES (UBA-CONICET), Argentina

^c Universidad de Alcalá, Spain



ARTICLE INFO

Article history:

Received 1 October 2021

Received in revised form 22 November 2021

Accepted 27 November 2021

Available online 15 December 2021

JEL classification:

C22

E17

Keywords:

CiSSA

COVID-19

Deseasonalizing

Outlier

X-13ARIMA-SEATS

ABSTRACT

COVID-19 hit the economy in an unprecedented way, changing the data generating process of many series. We compare different seasonal adjustment methods through simulations, introducing outliers in the trend and seasonality to reproduce the heterogeneity in the series during COVID-19.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

COVID-19 has destroyed the dynamics of many economic time series. In particular, trend and seasonality have been greatly hit by the effects of the lockdown. In this context, seasonal adjustment, a frequently used tool to monitor the state of the economy in real time, has become a difficult task generating great distress about the reliability of the alternative estimations. As a consequence, the different statistical offices reacted quickly and provided some guidelines to adapt seasonal adjustment methodologies to cope with this new and unexpected situation; see, for instance, the [Australian Bureau of Statistics \(2020\)](#), the European Statistical Office ([European Commission, 2020](#)), or the updated information provided by the United States Census Bureau accompanying the release of some indicators ([United States Census Bureau, 2021](#)). Two main options stand out within the recommendations: the use of concurrent seasonal adjustment with additive outlier interventions or the projection of the previous year's estimated seasonal factors. All in all, there is great uncertainty about how to proceed to deseasonalize economic time series nowadays.

This paper sheds some light onto this process by means of simulations where we know the true seasonal and deseasonalized time series. On the basis of a basic structural time series model, we simulate a distortion in the trend and seasonal component to

replicate the effect of COVID-19. In the simulations we know both, the raw data as well as the data clean of seasonality. Therefore, the application of alternative seasonal adjustment methods on the simulated time series enables us to approximate the effect that COVID-19 has had on the deseasonalized data as well as the relative reliability between different procedures.

The benchmark seasonal adjustment strategy considered is the well-known X-13ARIMA-SEATS of the Census Bureau (<https://www.census.gov/data/software/x13as.html>). We have also included as an alternative the newly introduced non-parametric Circulant Singular Spectrum Analysis (CiSSA), see [Bógalo et al. \(2021\)](#). The comparison exercise also considers the projection of last year's seasonal factors.

The remainder of the paper consists of four sections. In Section 2, we revise the different methodologies for seasonal adjustment. In Section 3, we introduce our simulation strategy. In Section 4, we present the results. Finally, in Section 5, we draw our conclusions.

2. Deseasonalizing strategies

The most common methods for seasonal adjustment used nowadays are included in X-13ARIMA-SEATS, developed by the US Bureau of Census. This comprises enhanced versions of both, the non-parametric X-11 ([Shiskin et al., 1967](#)) and the ARIMA model based TRAMO-SEATS ([Maravall, 1993](#); [Gómez and Maravall, 1996](#)). Given, for simplicity, a zero mean time series x_t , $t = 1, \dots, T$, X-11, first, estimates an initial trend using a moving

* Corresponding author.

E-mail address: lladamartin@gmail.com (M. Llada).

average. After removing the trend from the time series, it estimates the seasonal component by also moving average filters. Re-estimation of the previous components is done in an iterative way. For more details, see: <https://www.census.gov/topics/research/seasonal-adjustment.html>. As an alternative, TRAMO-SEATS (TS) is based on the ARIMA model decomposition. TS derives ARIMA models for each unobserved component and identifies and estimates the different signals according to the identified specifications.

Our new proposal Circulant Singular Spectrum Analysis (CISSA), is a novel variant of SSA. CISSA is a signal extraction algorithm that decomposes the original time series into the sum of oscillatory components at known frequencies. In what follows, we briefly describe the CISSA algorithm. In the first step, we define a window length L , and transform the original vector of data $x_t, t = 1, \dots, T$ into a related trajectory matrix \mathbf{X} of size $L \times N$ where $N = T - L + 1$, given by:

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & x_3 & \dots & x_N \\ x_2 & x_3 & x_4 & \dots & x_{N+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_T \end{pmatrix}.$$

In the second step, we find the eigenstructure of a matrix of second moments related to \mathbf{X} to obtain the so called elementary matrices of rank 1. The trajectory matrix \mathbf{X} can be recovered as the sum of the elementary matrices which are associated to different frequencies. In particular, CISSA builds a circulant matrix related to the second moments of the time series, \mathbf{S}_C given by:

$$\mathbf{S}_C = \begin{pmatrix} \hat{c}_0 & \hat{c}_1 & \hat{c}_2 & \dots & \hat{c}_{L-1} \\ \hat{c}_{L-1} & \hat{c}_0 & \hat{c}_1 & \dots & \hat{c}_{L-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{c}_1 & \hat{c}_2 & \hat{c}_3 & \dots & \hat{c}_0 \end{pmatrix}$$

where the elements \hat{c}_j are $\hat{c}_m = \frac{L-m}{L} \hat{\gamma}_m + \frac{m}{L} \hat{\gamma}_{L-m}, m = 0, 1, \dots, L-1$, being $\hat{\gamma}_m = \frac{1}{T-m} \sum_{t=1}^{T-m} x_t x_{t+m}, m = 0, \dots, L-1$ the estimated autocovariances

The eigenvalues and eigenvectors of \mathbf{S}_C are given by (see Lancaster, 1969):

$$\hat{\lambda}_{L,k} = \sum_{m=0}^{L-1} \hat{c}_m \exp\left(i2\pi m \frac{k-1}{L}\right) \tag{1}$$

and $\mathbf{u}_k = L^{-1/2}(u_{k,1}, \dots, u_{k,L})^H$ respectively, for $k = 1, \dots, L$, where H indicates the conjugate transpose and $u_{k,j} = \exp(-i2\pi(j-1)\frac{k-1}{L})$.

The diagonalization of \mathbf{S}_C allows us to write \mathbf{X} as sum of elementary matrices \mathbf{X}_k of rank 1 as:

$$\mathbf{X} = \sum_{k=1}^L \mathbf{X}_k = \sum_{k=1}^L \mathbf{u}_k \mathbf{w}_k^T,$$

where $\mathbf{w}_k = \mathbf{X}^T \mathbf{u}_k$.

In a third step, we go back from the matrices to the vectors of the time series and transform the elementary matrices \mathbf{X}_k into elementary signals of the same length as the original series for each frequency $w_k, k = 1, \dots, L$ by

$$x_t^{(k)} = \begin{cases} \frac{1}{t} \sum_{i=1}^t \tilde{x}_{i,t-i+1}, & 1 \leq t < L \\ \frac{1}{L} \sum_{i=1}^L \tilde{x}_{i,t-i+1}, & L \leq t \leq N \\ \frac{1}{T-t+1} \sum_{i=L-N+1}^{T-N+1} \tilde{x}_{i,t-i+1}, & N < t \leq T \end{cases},$$

where $\tilde{x}_{i,j}$ are the elements of the elementary matrix \mathbf{X}_k .

In the final fourth step, we group the extracted elementary signals according to the frequency they represent. Notice that the k th eigenvalue in (1) is an estimate of the spectral density at $w_k = \frac{k-1}{L}, k = 1, \dots, L$ and, therefore, the k th eigenvalue and corresponding eigenvector are associated with this frequency.

3. Simulation strategies

We simulate time series, modify their trend and seasonal dynamics to approximate the impact of COVID-19 and state the basis for understanding and comparing the effect that the pandemic might have had on real economic time series data. We take as the series free from COVID-19 effects, the addition of the simulated trend, cycle, seasonal, and irregular components. This can be considered as the data that we should have obtained, had the pandemic not taken place, i.e. a counterfactual time series. Notice that the sum of all the components except the seasonal one constitutes the “true” time series, free of seasonality, denoted as $x_{SA,t}^0$.

The knowledge of the original components enables us to assess the reliability of the different methods (denoted by the subindex m) by comparing precisely the “true” seasonally adjusted time series ($x_{SA,t}^0$) with the deseasonalized series resulting from each estimation method ($x_{m,t}^0$).

In the free-COVID-19 BSM model the data are generated as the sum of trend, cycle, seasonality and irregular components as follows¹ (we denote this counterfactual time series by x_t^0):

$$x_t^0 = \mu_t + c_t + s_t + e_t \tag{2}$$

where μ_t is the changing level or trend component, c_t is the cycle, s_t is the seasonal component and e_t is the irregular component. The “true” deseasonalized time series will be given by $x_{SA,t}^0 = x_t^0 - s_t$.

Regarding the data generating process for x_t^0 given by (2), we assume an integrated random walk for the trend, see, for instance, Young (1984), given by

$$T_t = T_{t-1} + \beta_{t-1} \tag{3}$$

$$\beta_t = \beta_{t-1} + \eta_t \tag{4}$$

with $\eta_t \sim N(0, \sigma_\eta^2)$. The cyclical and seasonal components are specified according to Durbin and Koopman (2012). The cycle is given by the first component of the bivariate VAR(1):

$$\begin{pmatrix} c_t \\ \tilde{c}_t \end{pmatrix} = \rho_c \begin{pmatrix} \cos(2\pi w_c) & \sin(2\pi w_c) \\ -\sin(2\pi w_c) & \cos(2\pi w_c) \end{pmatrix} \begin{pmatrix} c_{t-1} \\ \tilde{c}_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \tilde{\varepsilon}_t \end{pmatrix} \tag{5}$$

with $\begin{pmatrix} \varepsilon_t \\ \tilde{\varepsilon}_t \end{pmatrix} \sim N(\mathbf{0}, \sigma_\varepsilon^2 I)$ and $\frac{1}{w_c}$ is the cyclical (business cycle) period, $w_c \in [0, 1]$. The seasonal component is given by

$$s_t = \sum_{j=1}^{[s/2]} a_{j,t} \cos(2\pi w_j t) + b_{j,t} \cos(2\pi w_j t) \tag{6}$$

with $w_j = \frac{j}{s}, j = 1, \dots, [s/2]$ and s the seasonal period, where $[\cdot]$ denotes the integer part, and $a_{j,t}$ and $b_{j,t}$ are two independent random walks with noise variances equal to σ_j^2 . Finally, the irregular component is white noise with the variance σ_e^2 . All the components are independent of each other. We set $\rho_c = 1$, so the trend, cycle and seasonal components all have unit roots.

To approximate the COVID-19 effects on the simulated time series we address both, the impact on the trend and the seasonal components. Regarding the trend, we considered a shock at time $T_{COVID-19}$ of magnitude δ_{Trend} drawn from a uniform distribution in $[a, b], 0 < a < b < 1$ where b and a would represent the minimum and maximum impact on the level of the series, respectively. This heterogeneity in the value of δ_{Trend} will capture the alternative effects of COVID-19 in different economic time series.

¹ There is no clear consensus about how the modelling strategy should be approached and some authors have advocated the use of multiplicative rather than additive time series models given that some quantities are zero or close to zero. We do not take a strong stand on this because were the model is multiplicative, just by taking logs we will have produced an additive model.

To represent the dynamics of the shock we include the possibility of transitory effects by means of including a Transitory Change intervention.

In this sense, the simulated series is of the form:

$$X_t^{Trend-Seas*} = \begin{cases} X_t^0 & \text{if } t < T_{COVID-19} \\ \mu_t^* + c_t + s_t^* + e_t & \text{otherwise} \end{cases} \quad (7)$$

with

$$\mu_t^* = \begin{cases} \mu_t & \text{if } t < T_{COVID-19} \\ \mu_t \times \delta_{Trend} \times D_t & \text{otherwise} \end{cases}$$

where D_t is of the form $D_t = \frac{I_{T_{COVID-19}}}{1-\omega B}$, with B representing the backshift operator such that $BX_t = X_{t-1}$, $I_{T_{COVID-19}}$ an impulse variable that takes the value 1 at time $T_{COVID-19}$ and 0 otherwise and $\omega \in [0, 1)$. For seasonality, we consider

$$s_t^* = s_t \times u_t, t = T_{COVID-19}+1, \dots, T$$

and u_t a random shock from a uniform [0, 1] distribution with a possible effect from total disruption to a non-significant dampening of the seasonality. This represents a complete disruption of the component during the time the intervention lasts and we call this effect total (seasonality) disruption.

4. Simulation results

To understand the possible effect of the COVID-19 pandemic on seasonal adjustment, we performed a set of simulations under different conditions. In the first set of simulations, the different procedures for seasonal adjustment were put to work for the non-contaminated BSM with the idea of understanding their performance and relative accuracy gains, if any. In a second set of simulations, we contaminate the 12 last observations to try to replicate the effects of the COVID-19 in an economic time series and analyse the results with respect to the previous stable period and the relative accuracy amongst the different deseasonalizing procedures.

The Basic Structural Model is generated by considering Eqs. (2) to (6), $s = 12$ (monthly time series) and the cyclical period equal to $\frac{1}{w_c} = 48$ months. The noise variances of the different components are given by $\sigma_n^2 = 0.0006^2$, $\sigma_j^2 = 0.004^2$, $\sigma_\varepsilon^2 = 0.008^2$ and $\sigma_e^2 = 0.06^2$. We considered three different sample sizes, $T = 97, 193$ and 243 .

We generated $R = 1000$ replications from the base model and on each replication we also generated the different contamination schemes. Then, for each replication, we assessed the accuracy of the alternative procedures by means of the root mean square error (RMSE) between the generated time series free of seasonality and the estimated seasonally adjusted time series:

$$RMSE_m = \sqrt{\frac{1}{T} \sum_{t=1}^T (X_{SA,t}^0 - X_{m,t})^2} \quad (8)$$

Afterwards, we computed the average of the RMSEs across replications.

4.1. Seasonal adjustment in normal times

To assess the relative performance of the different parametric and non-parametric seasonal adjustment strategies, the first set of simulations were run under the non-contaminated original Basic Structural Model.

All the models considered were applied in an automated way with no outlier treatment. Regarding CiSSA, we needed to choose the window length L , such that $L < T/2$, because the trajectory matrices with window length L and $N = T - L + 1$ are transposed.

Table 1

Non-COVID-19 basic structural model. Average RMSE between seasonally adjusted simulated and estimated time series, 1000 replications for different strategies and values of L in CiSSA estimation.

T	CiSSA				SEATS	X-11
	L = 24	L = 48	L = 96	L = 144		
97	0.056	0.049			0.053	0.058
193	0.056	0.048	0.054		0.050	0.055
289	0.056	0.047	0.053	0.062	0.049	0.054

Table 2

Basic structural model with trend and seasonality COVID-19 proxy shocks. Average RMSE between seasonally adjusted simulated and estimated time series, 1000 replications.

T	CiSSA		TRAMO-SEATS	X-11	CiSSA*	TRAMO-SEATS*	X11*
	L = 48	L_{max}					
COVID-19 period (Last 12 observations)							
97	0.073	0.073	0.083	0.095	0.249	0.255	0.251
193	0.109	0.092	0.099	0.117	0.302	0.319	0.317
289	0.146	0.109	0.115	0.136	0.333	0.369	0.367
Whole sample							
97	0.059	0.059	0.062	0.068	0.106	0.107	0.107
193	0.059	0.064	0.057	0.062	0.091	0.096	0.098
289	0.062	0.074	0.056	0.062	0.088	0.092	0.094

We chose several values of L , multiples of $s = 12$, according to the different sample sizes T analysed. Table 1 shows the average RMSE for the 1000 replications for the different seasonal adjustment strategies and sample sizes.

Table 1 suggests that in normal times, there are no noteworthy differences among the alternative procedures. Table 1 also allows us to assess the sensibility to the different values of the window length L . The best results were obtained with $L = 48$, that is a multiple of the seasonality and matches the simulated cycle.

4.2. Seasonal adjustment in COVID-19 times

In this section we show the impact of the COVID-19 pandemic whose effect was simulated with changes in trend and seasonality according to (7) for the last 12 observations. The new parameters were δ_{Trend} and w in (7). δ_{Trend} is related to the COVID-19 effect of the trend which is drawn from a uniform [0.2, 0.8]. These limits on the magnitude of the intervention attempt to accommodate the heterogeneity of the impact in different time series according to the relationship with the mobility of their corresponding sector. The other parameter takes the value $w = 0.8$ for a Temporary Change (TC), to account for the long-term effects of the pandemic. Seasonality is totally disrupted with each alternative seasonal factor multiplied by a uniform [0,1] random value. Regarding the different alternatives m , we considered the automated versions of X-13ARIMA-SEATS (both, SEATS and X-11) with an outlier intervention. In the case of CiSSA, we included the results for $L = 48$ and $L_{max} = T/2 + 1$ as an alternative choice when no information about the cycle periodicity was known. No outlier correction was made for CiSSA due to its non-parametric nature and the good performance showed by Bógalo et al. (2021) in nonlinear and nonstationary signals. Finally, we also considered the results of projecting the seasonality of the previous year and have denoted this option by CiSSA*, TRAMO-SEATS* and X11*.

Table 2 shows the results of the average RMSE for the 1000 replications for each method. The results are presented not only for the last 12 observations but also for the whole sample to account for the distortion they may have caused on the previous estimations. The first column relates to the sample size, while the remaining columns include the average RMSEs across the replications of the different seasonal adjustment methodologies.

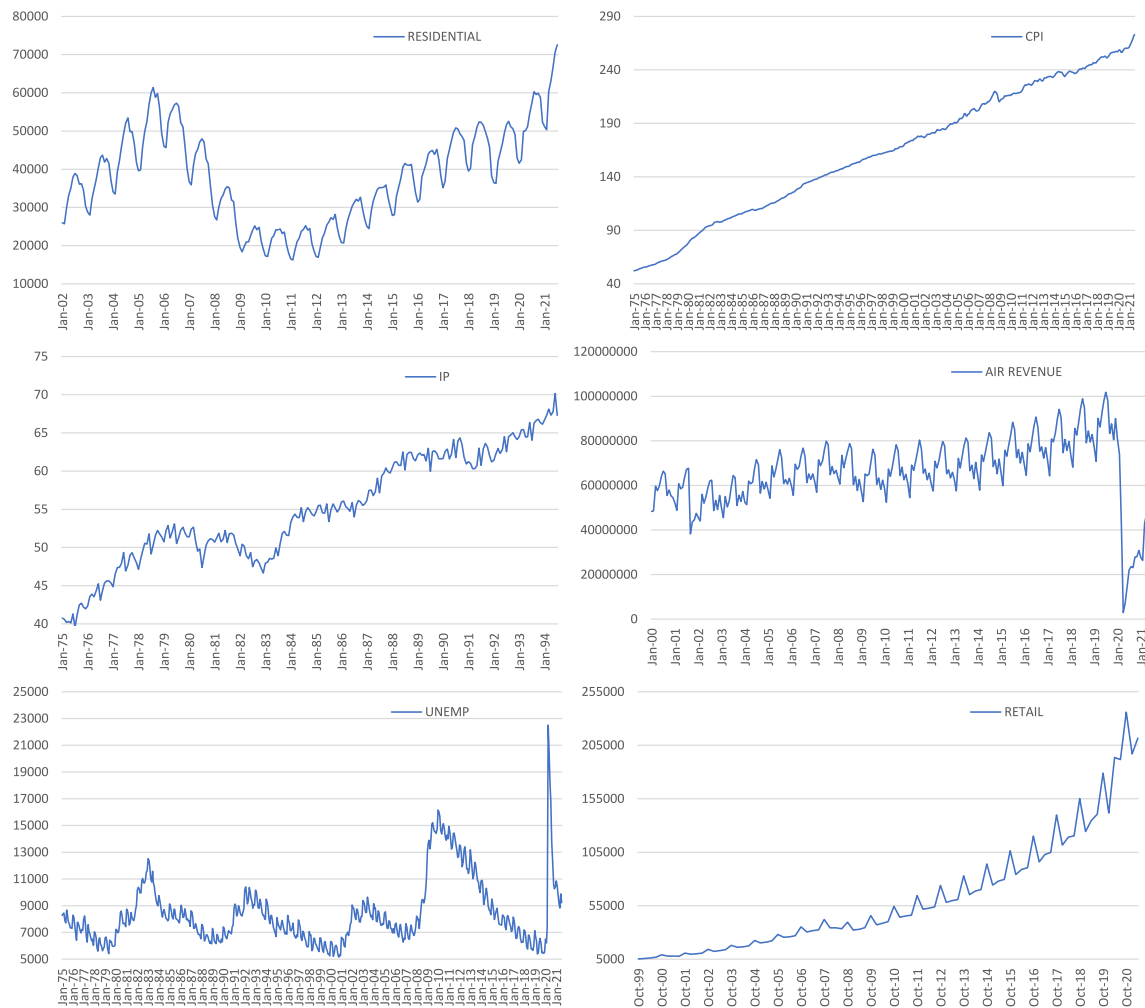


Fig. 1. Evolution of economic indicators of the U.S.

Columns two to five show the results of the estimations using the whole sample: CiSSA (columns 2 and 3) with $L = 48$ (the simulated cycle periodicity) and $L = L_{max} = T/2 + 1$ (the possible maximum value), TRAMO-SEATS (column 4) and X-11 (column 5). Finally, columns six to eight show the behaviour of projecting the previous year's factors during the simulated COVID-19 with each of the methodologies.

Comparing Table 2 with Table 1 we can see that the accuracy diminishes under the effects of the shocks in trend and seasonality with all the methods. The magnitude of the worsening ranges between 1.1 and 2.5 times the RMSEs presented in Table 1 for estimations made with the whole sample, but between 4.3 and 9.7 when projecting the previous year seasonal factors. The second conclusion is that projecting the previous year seasonal factors when abrupt changes in trend and seasonality take place is always the worst option. Looking at the results for the last 12 observations, we can see that the average RMSE increases with T . Also, we see that CiSSA shows the best results with $L = L_{max}$, diminishing the RMSE with respect to TRAMO-SEATS by 12% with $T = 97$, 7% for $T = 293$ and 4% for $T = 289$. It is worth highlighting that, contrary to Table 1, the best results are obtained with L_{max} and not with $L = 48$, which is in line with the disruption of both the seasonality and the simulated business cycle.

Looking at the RMSE results for the whole sample assesses the effects on the re-estimations of the previous values. The conclusion regarding the worst behaviour of projecting the seasonality

of the previous year remains. We can also see that CiSSA is only better than TRAMO-SEATS when the sample is short ($T = 97$). However, when the sample size increases, TRAMO-SEAT obtains better results than CiSSA. This might suggest that, although for the last 12 observations CiSSA yields the best results, estimations before COVID are also affected and some kind of interventions might be needed as for the other methods.

To illustrate the performance of the different seasonal adjustment methods, we apply them to several US series from the FRED Economic Data database: Residential Construction (RESIDENTIAL, millions of dollars), Consumer Price Index (CPI, index), Industrial Production (IP, index), Air passengers Revenues (AIR REVENUE, thousands of dollars), Unemployment (UNEMP, thousands of persons), and E-Commerce Retail Sales (RETAIL, millions of dollars). As it can be seen in Fig. 1 seasonality is a relevant feature in all of them, being CPI the only one where it is not clearly seen just with visual inspection. We can also see the different impact COVID-19 has had in the series.

Fig. 2 shows the monthly growth rates of the seasonally adjusted series since January 2020 (quarter over the previous quarter in the case of RETAIL) as well as the officially published seasonally adjusted series. Differences among the considered procedures appear in Residential, Unemployment, Air Revenues and Retail. Notice as well that CiSSA is the procedure that departs more from the official data. As our simulations show, this might be due the different treatment of COVID outliers.

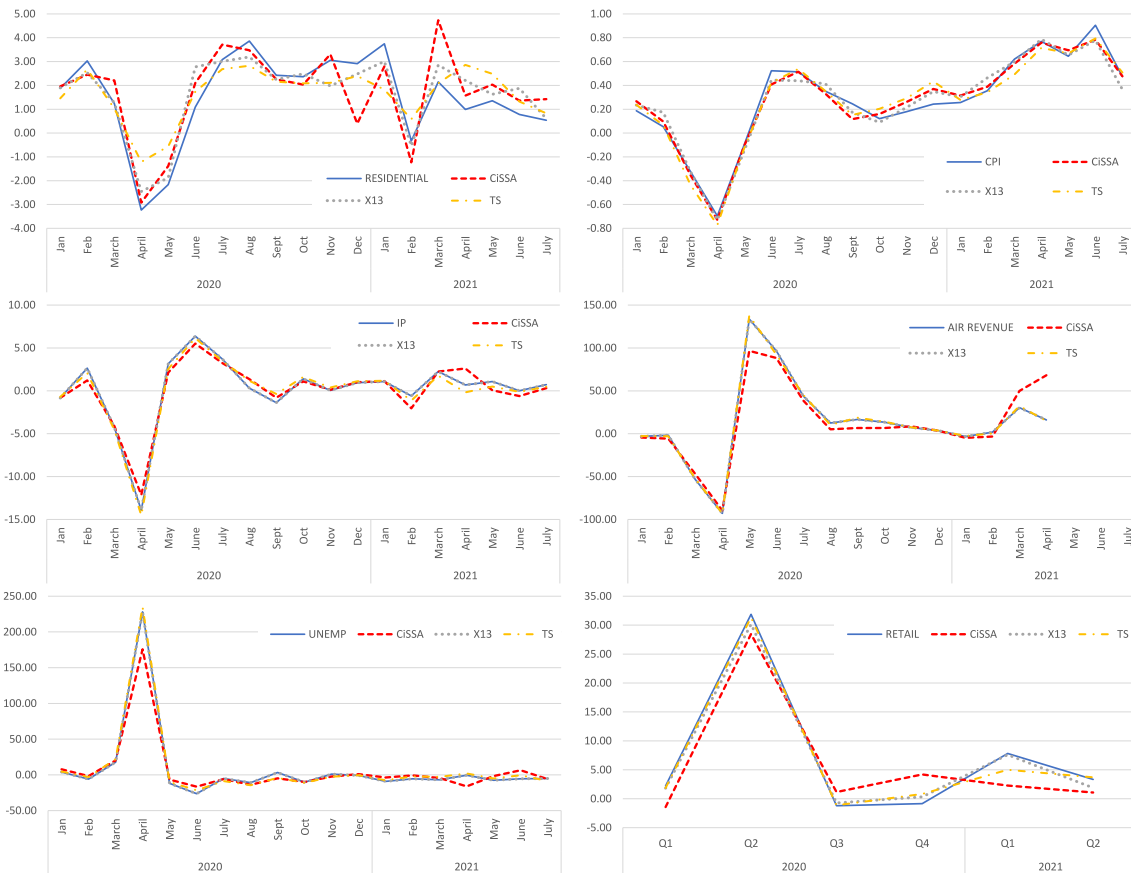


Fig. 2. Evolution of the growth rate of the seasonally adjusted economic indicators of the U.S.

5. Conclusion

Seasonal adjustment has become a very difficult task since the COVID-19 pandemic. Several proposals and recommendations have appeared to cope with this problem, among them, the use of outlier techniques and the projection of the estimated seasonal factors for 2019 in the subsequent months. We have run a set of simulations contaminating time series with shocks in the trend and seasonal components to emulate the type of shock that the COVID might have in economic time series and computed the seasonally adjusted time series with the widely used X-13ARIMA-SEATS and a new non-parametric technique based on subspace methods. From our set of simulations, projecting the estimated seasonality in 2019 in the following months gives the worst results for any of the procedures used for seasonal adjustment. On the contrary, the usual X-13ARIMA-SEATS with outlier detection seems a better option. Moreover, if the type of shock is a total disruption in seasonality combined with a shock in the trend, the non-parametric CiSSA seems to render better results. It seems that if the type of outlier is not within those automatically contemplated by X-13ARIMA-SEATS, the non-parametric procedure developed in CiSSA might be preferred. Our proposal is just a starting point to check how the effect of the COVID-19 affects the different procedures for seasonal adjustment.

Acknowledgements

This paper was written while the second author was visiting the Department of Quantitative Economics of the Uni-

versidad Autónoma de Madrid (Spain). He thanks the financial support received from Fundación Carolina. The remaining authors also acknowledge valuable suggestions and comments from Arbués, I., Cuerpo, C., and Puch, L. and the financial support provided by the Spanish Ministry of Science and Innovation, contract grants: MINECO/FEDER PID2019-107161GB-C32 and PID2019-108079GB-C22 and the Spanish Ministry of Economic Affairs and Digital Transformation.

References

Australian Bureau of Statistics, 2020. Methods changes during the COVID-19 period. <https://www.abs.gov.au/articles/methods-changes-during-covid-19-period>.

Bógalo, J., Poncela, P., Senra, E., 2021. Circulant singular spectrum analysis: A new automated procedure for signal extraction. *Signal Process.* 179, 107824.

Durbin, J., Koopman, S.J., 2012. *Time Series Analysis by State Space Methods*, second ed. OUP Oxford.

European Commission, 2020. Guidance on treatment of COVID-19-CRISIS effects on data. https://ec.europa.eu/eurostat/cros/system/files/treatment_of_covid19_in_seasonal_adjustment_methodological_note.pdf.

Gómez, V., Maravall, A., 1996. Programs TRAMO and SEATS: Instructions for the User. Technical Report, Bank of Spain.

Lancaster, P., 1969. *Theory of Matrices*. Academic Press, New York.

Maravall, A., 1993. Stochastic linear trends: Models and estimators. *J. Econometrics* 56, 5–37.

Shiskin, J., Young, A.H., Musgrave, J.C., 1967. *The X-11 Variant of the Census Method II Seasonal Adjustment Program*. U.S. Department of Commerce, Bureau of the Census.

United States Census Bureau, 2021. COVID-19's effect on the advance monthly retail trade survey. <https://www.census.gov/retail/marts/www/covid19retailfaq.pdf>.

Young, C.P., 1984. *Recursive Estimation and Time-Series Analysis*. Springer-Verlag Berlin Heidelberg, Berlin.