Seasonal Adjustment of the Spanish Daily Sales Data

Ángel Cuevas *Ramiro Ledo* *Enrique M. Quilis*

Research and Statistics Department
Tax Agency

Eurostat, 9-11 March 2021 NTTS

Acknowledgment and Disclaimer

- Work in progress. Comments and suggestions are welcomed!!

- We thank A. Abad, J. Abelaira, R. Frutos, D. García, F. Melis, G. Pérez-Quirós, F. Pizarro, J. Polo and S. Relloso for their input at different stages of this project.

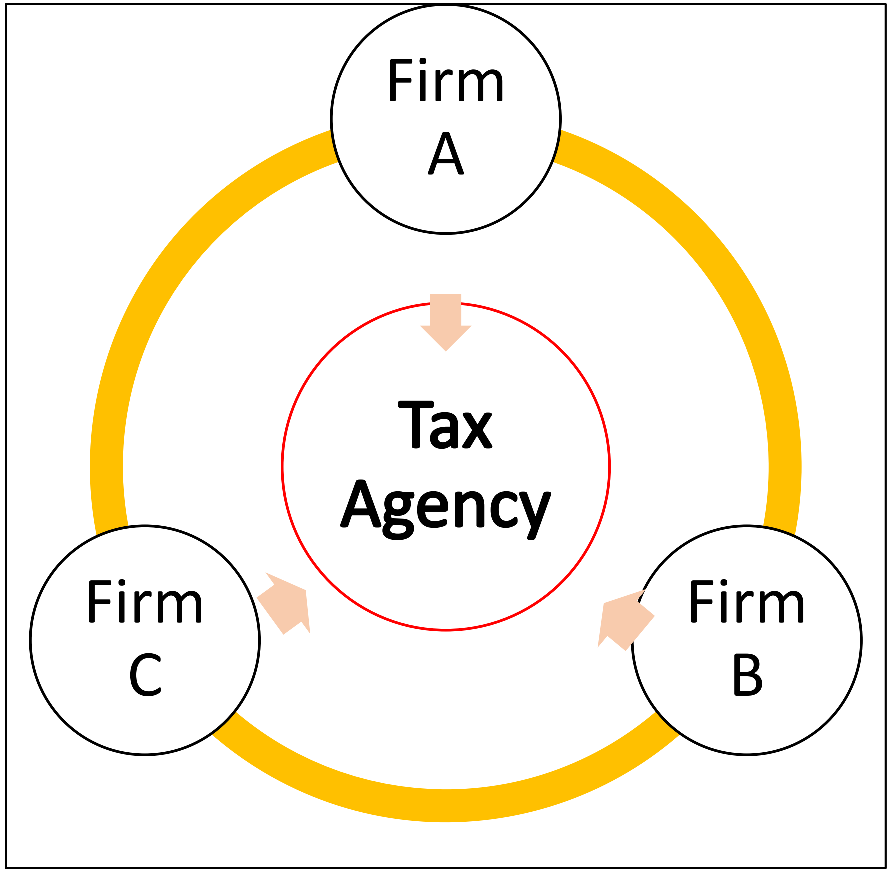
- The opinions are those of the authors and do not necessarily reflect the views of the Tax Agency.

Outline

- Introduction
- Data
- Seasonal adjustment: econometric methodology
- Seasonal adjustment: results
- Conclusions and future work

Data

- Sales data are derived from the Value Added Tax (VAT) system.
- VAT Instantaneous Information System (SII, *Sistema Inmediato de Información*):
  - Fully digital, ingrained in the firms' invoicing systems.
  - Operates on a continuous time basis.
  - Each firm has a tax account with the Tax Agency ➔ periodic settlement.
- The system provides daily data for sales, AOT.
- Operates as a reservoir ➔ (usually) upward revisions.
- Revisions stabilize ~10d after reference day.
- Coverage ~70% of the Spanish total firms turnover.
- Short homogeneous time series (since 2017:07). Longer but inhomogeneous time series is also availabe (since 2010).



Seasonal adjustment: econometric methodology

- The complete methodology has two steps (Hillmer et al., 1983; De Livera, 2011):

- **Step 1: Pre-preprocessing (linearization)**
  - Intervention analysis by means of exogenous deterministic variables designed to control for the presence of outliers and specific calendar effects that, due to their moving nature, do not fit into the structural representation considered by TBATS. This analysis is a preprocessing step of the observed series that renders it suitable for TBATS: y = z - xb.

- **Step 2: Unobserved components model (TBATS)**
  - The pre-processed time series is decomposed into: trend-cycle (cp), seasonality (s) and a stationary innovation (u): y = *p + s + u.*

Seasonal adjustment: econometric methodology
Step 1: Pre-processing

$$z_t = \beta x_t + \alpha_0 + \alpha_1 t + \sum_{i=1}^{3} \sum_{j=1}^{k_i} \left[\gamma_j \sin\left(\frac{2j\pi t}{m_i}\right) + \varphi_j \cos\left(\frac{2j\pi t}{m_i}\right)\right] + e_t \qquad [1]$$

- z_t: (log-transformed) observed variable.
- x_t: m deterministic (dummy) variables linked to the bank holidays and to the inexact periodicity of the monthly seasonality.
- m_i: periodicity of the i-th seasonal component: weekly, monthly and yearly with periodicities: 7d, 30.4375d and 360.25d.
- k_i: number of harmonics for each seasonal component.
- e_t: Gaussian error term.

- Model [1] is a regression-based approximation to TBATS

Seasonal adjustment: econometric methodology

- We follow the **TBATS** structural approach of De Livera et al. (2011).
- TBATS stands for:
  - **T**rigonometric representation of seasonality.
  - **B**ox-Cox transformation.
  - **A**RIMA model for the irregular (remainder) component.
  - **T**rend: very general structural representation that ranges from I(1) to I(2).
  - **S**easonality: multiple seasonal patterns with possibly fractional periodicity.
- The model is cast in state space form and estimated via Kalman filtering and numerical optimization methods (Nelder-Mead).

Seasonal adjustment: econometric methodology
Step 2: Stochastic decomposition (TBATS)

- The TBATS model assumes that the linearized time series (y_t) results from the aggregation of three unobserved components: trend (p_t), seasonality (s_t) and a stationary innovation (u_t).

$$y_t = p_t + s_t + u_t$$

$$\left(1 - \phi_1 B - \cdots - \phi_p B^p\right) u_t = \left(1 - \theta_1 B - \cdots - \theta_q B^q\right) e_t$$

$$e_t \sim iid\ N(0, v)$$

- The remainder follows a stationary and invertible ARMA model.
- It is much more than a simple added noise, since it is the stochastic input for all the other components.

Seasonal adjustment: econometric methodology
Step 2: Stochastic decomposition (TBATS)

$$p_t = p_{t-1} + \phi\, g_{t-1} + \alpha\, u_t$$

$$g_t = (1-\phi)b + \phi\, g_{t-1} + \beta\, u_t$$

$$\phi \in [0,1]$$

$$\alpha \geq 0$$

$$b \geq 0$$

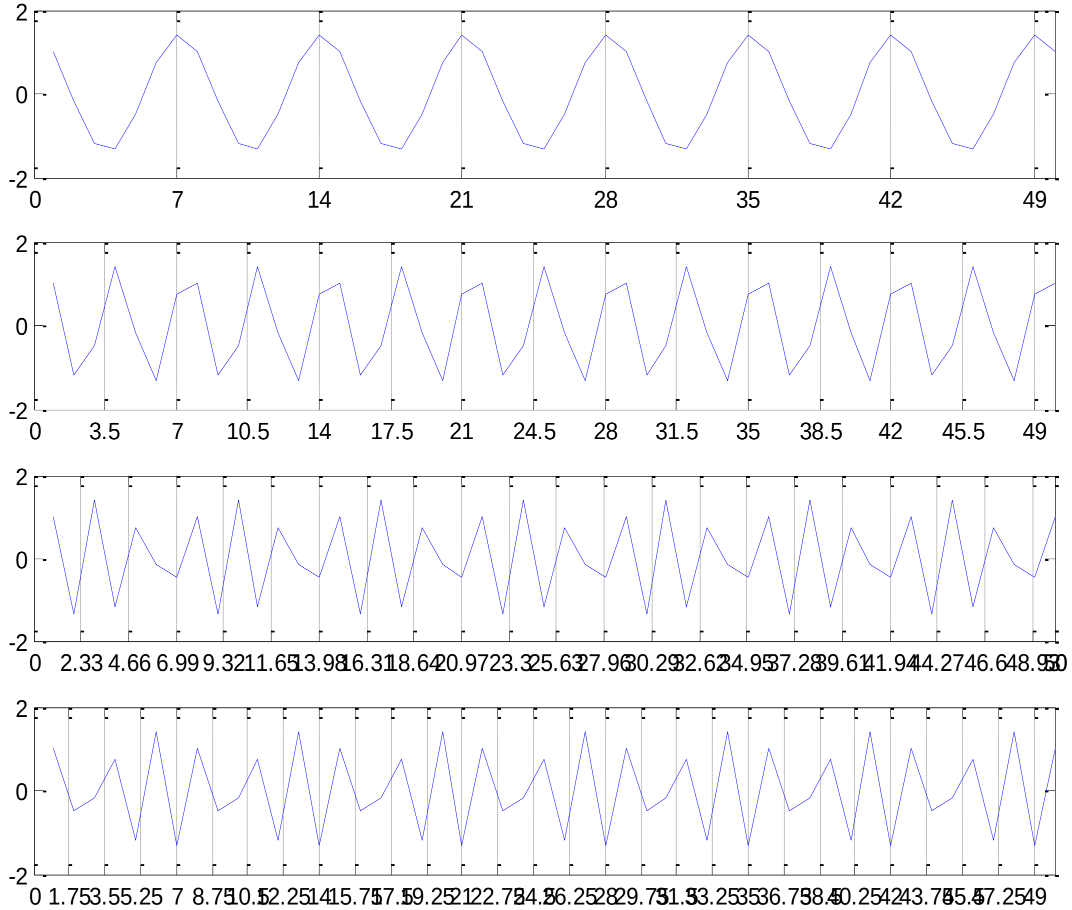
$$\beta \geq 0$$

- The trend consists of two elements:
  - Level ~ I(1) + drift
  - Drift ~ I(0) or I(1)
- The basic parameters are:
  - *Damping*: Φ
  - *Scale*: α, β
  - *Long run drift average*: b
- Depending on the combination of the parameters, the trend behavior ranges from a linear trend to an I(2), including all the intermediate cases (e.g. I(1)+drift).
- The level and the drift share the same innovation, appropriately scaled by α and β, respectively.
- The component that adds up to seasonality and irregularity is:

$$cp_t = p_{t-1} + \phi\, g_{t-1} = p_t + \alpha\, u_t$$

Seasonal adjustment: econometric methodology
Step 2: Stochastic decomposition (TBATS)

$$S_t = \sum_{i=1}^{I} S_t^{(i)} \qquad w_j^i = \frac{2\pi j}{m_i}$$



i: Seasonal sub-component (e.g. weekly).

m: Periodicity in time units (e.g., days). It can be fractional (e.g, 365.25 days for the annual seasonality)

j: Basic frequency (j=1) and its harmonics (j>1). For each sub-componente TBATS considers up to k harmonics.

Seasonal adjustment: econometric methodology
Step 2: Stochastic decomposition (TBATS)

For each i (its index is dropped):

$$\begin{bmatrix} S_{j,t} \\ s_{j,t} \end{bmatrix} = \begin{bmatrix} \cos(w_j) & \sin(w_j) \\ -\sin(w_j) & \cos(w_j) \end{bmatrix} \begin{bmatrix} S_{j,t-1} \\ s_{j,t-1} \end{bmatrix} + \begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix} u_t$$

- For each seasonal sub-component (e.g. weekly) we set k VAR(1) models.
- The VAR(1) transition matrix is known and depends only on the frequency (basic or harmonic, in rads).
- Given the initial conditions, the equation generates a deterministic seasonal path.
- These paths are buffeted by some shocks, linked to the common innovation of the model. The size of the shocks is determined by the scale parameters. If both are zero, the seasonality becomes deterministic.

Parameters (for each i):

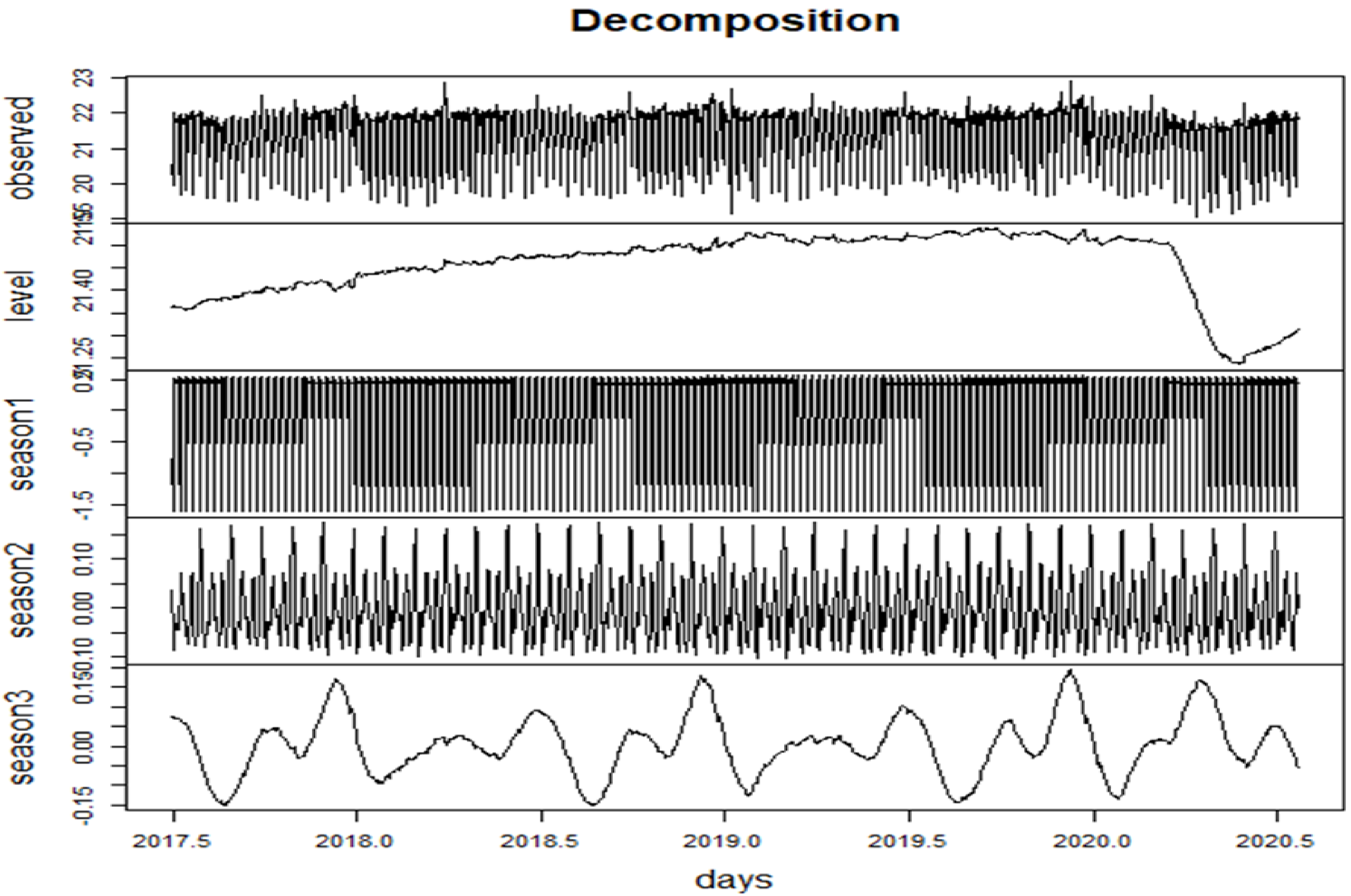
- *Initial conditions:*
  $$\begin{bmatrix} S_{j,0} \\ s_{j,0} \end{bmatrix}$$
- *Scale parameters:*
  $$\begin{bmatrix} \gamma_1 \\ \gamma_2 \end{bmatrix}$$
- *Number of harmonics:*
  j=1..k

Seasonal adjustment: econometric methodology
Results

$$z_t = \beta x_t + \alpha_0 + \alpha_1 t + \sum_{i=1}^{3} \sum_{j=1}^{k_i} \left[\gamma_j \sin\left(\frac{2j\pi t}{m_i}\right) + \varphi_j \cos\left(\frac{2j\pi t}{m_i}\right) \right] + e_t$$

		Monthly component					
		Basic effect			Interaction with the weekends		
	Holiday	End of month	Beginning of month	15th day	End of month	Beginning of month	15th day
β	-1.40	1.67	0.50	0.40	1.24	0.65	0.67
t(β)	-34.00	31.44	9.17	8.11	16.30	8.63	8.84

Note: The number of harmonics associated with each seasonal component (weekly, monthly and annual) is 3, 9 and 5, respectively. These values are derived from the preliminary estimate of a TBATS model applied to the original time series.



Decomposition
observed
level
season1
season2
season3
days
2017.5
2018.0
2018.5
2019.0
2019.5
2020.0
2020.5

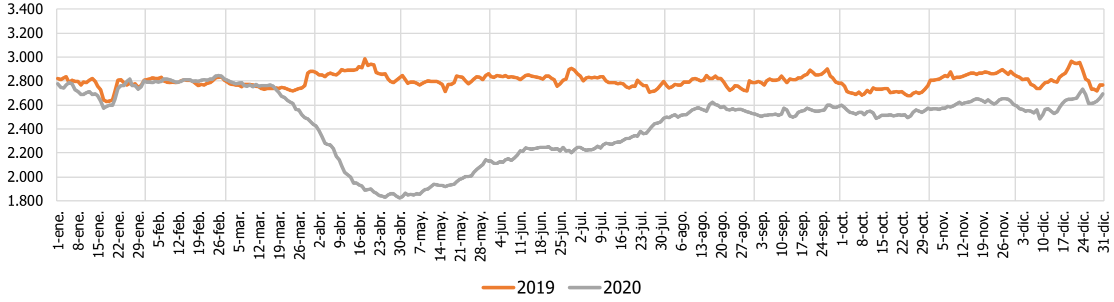
Results by sectorial breakdown

- The firms included in the Immediate Information Supply System (SII) are classified according to their declared activity. The activities are classified according to the epigraphs of the Economic Activities Tax (IAE).

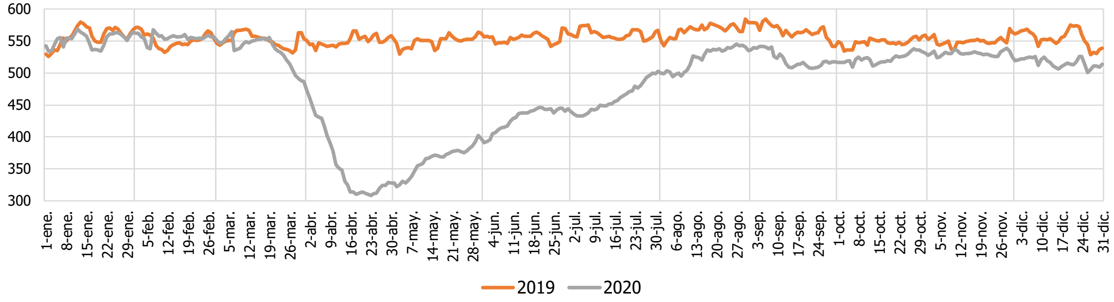
- For the purpose of data dissemination, firms are classified to four digits of the NACE-2009 and are grouped, based on this classification, at 34 branches level.

Results by sectorial breakdown: SAC series (MA-28)

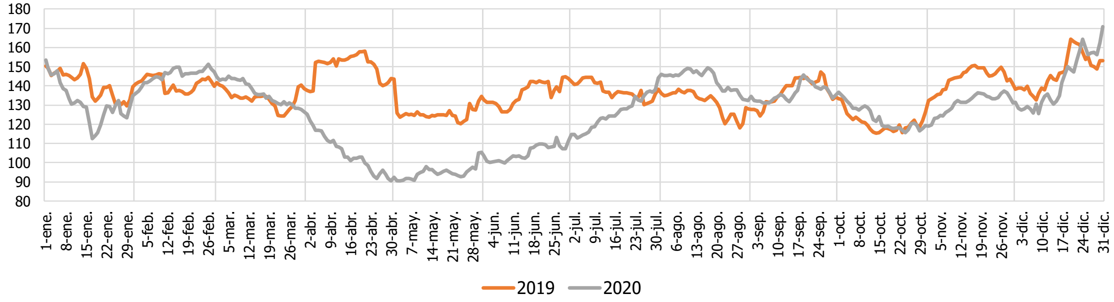
TOTAL



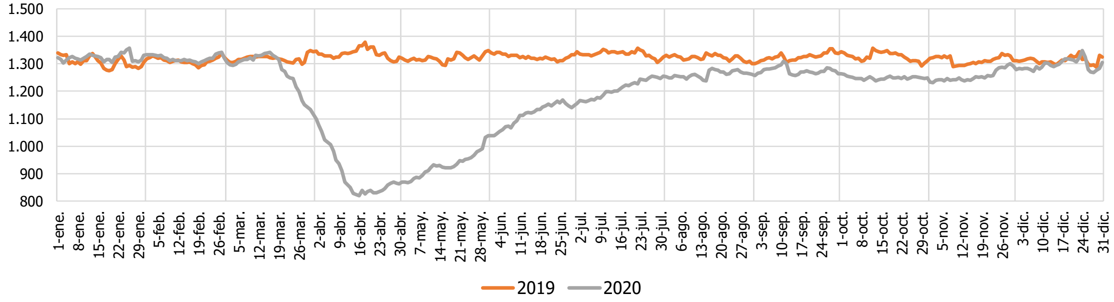
Manufacturing industry



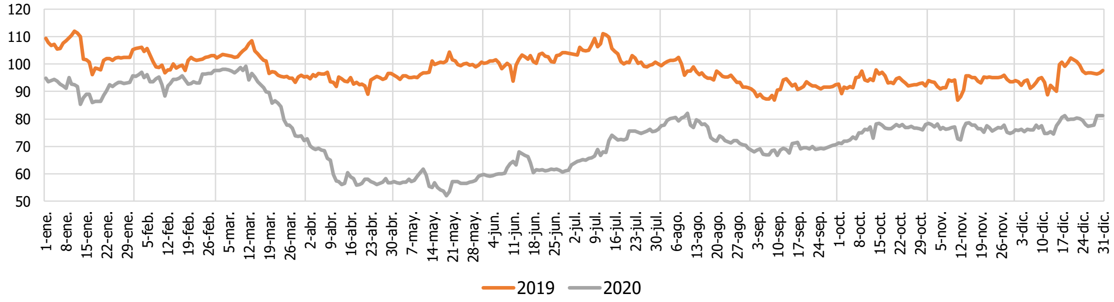
Construction



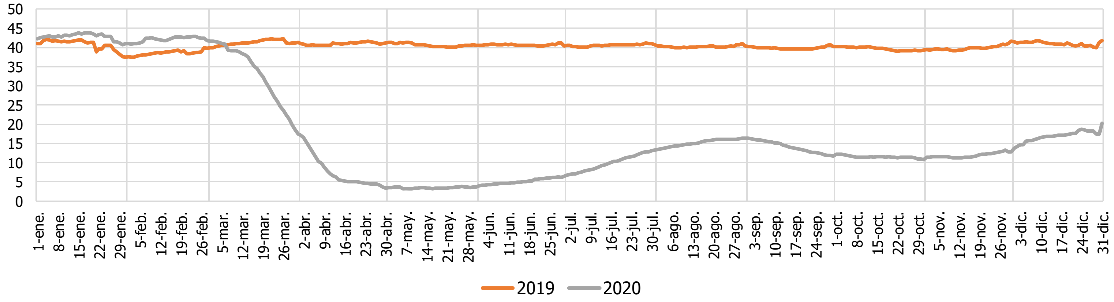
Commerce



Transport



Hostelry



Conclusions and future work

- The relevance and use of daily data has increased due to due to adverse and sudden economic developments, like the one due to COVID-19 (European Central Bank, 2020).
- Seasonal adjustment (SA) of daily data is necessary. This task is notably more difficult than in the case of monthly or quarterly data (Ladiray et al., 2018).
- The application of the two-step of approach of De Livera et al. (2011) provides sensible results, allowing a detailed sectorial analysis. In addition, it can be used for forecasting.
- The Spanish daily sales data are timely, reliable and have a detailed breakdown ➔ ideal for short-term monitoring.
- **Future work**: Comparison with alternative SA approaches (Ollech, 2018).

Thanks for your attention!

Ángel Cuevas: angel.cuevas@correo.aeat.es

Ramiro Ledo: ramiro.ledo@correo.aeat.es

Enrique M. Quilis: enrique.martinq@correo.aeat.es

Research and Statistics Department
Tax Agency

References

- A. De Livera, R.J. Hyndman and R.D. Snyder, "Forecasting time series with complex seasonal patterns using exponential smoothing", *Journal of the American Statistical Association*, 106(496), (2011), 1513-1527.
- European Central Bank (2020) "Tracking the economy with high-frequency data", Seminar Proceedings.
- S.C. Hillmer, W.R. Bell and G.C. Tiao, "Modeling considerations in the seasonal adjustment of economic time series". In: Zellner, A. (Ed.) Applied Time Series Analysis of Economic Data, p. 74-100. US Bureau of the Census, (1983).
- D. Ladiray, J. Palate, G.L. Mazzi and T. Proietti, "Seasonal adjustment of daily and weekly data". In: D. Ladiray and G.L. Mazzi (Eds.) Handbook on Seasonal Adjustment, Eurostat, (2018).
- D. Ollech, "Seasonal adjustment of daily time series", Deutsche Bundesbank, Discussion Paper n. 41/2018, (2018).