



# Pooled Time Series Modeling Reveals Smoking Habit Memory Pattern

Jesús F. Rosel<sup>1</sup>, Marcel Elipe-Miravet<sup>2</sup>, Eduardo Elósegui<sup>3</sup>, Patricia Flor-Arasil<sup>4</sup>, Francisco H. Machancoses<sup>5</sup>, Jacinto Pallarés<sup>1</sup>, Sara Puchol<sup>6</sup> and Juan J. Canales<sup>7\*</sup>

<sup>1</sup> Department of Evolutionary Psychology, Educative, Social Studies and Methodology, Universitat Jaume I, Castellón, Spain, <sup>2</sup> Department of Basic and Clinical Psychology and Psychobiology, Universitat Jaume I, Castellón, Spain, <sup>3</sup> Department of Theory and History of Education and Methods of Research and Diagnosis in Education, Universidad de Málaga, Málaga, Spain, <sup>4</sup> Department of Health Science, Valencian International University, Castelló de la Plana, Spain, <sup>5</sup> Predepartmental Unit of Medicine, Universitat Jaume I, Castellón, Spain, <sup>6</sup> IATS, Consejo Superior de Investigaciones Científicas, Castellón, Spain, <sup>7</sup> School of Psychological Sciences, University of Tasmania, Hobart, TAS, Australia

## OPEN ACCESS

### Edited by:

Susan Schenk,  
Victoria University of Wellington,  
New Zealand

### Reviewed by:

Delin Sun,  
Duke University, United States  
Li-Zhuang Yang,  
Hefei Institutes of Physical Science  
(CAS), China

### \*Correspondence:

Juan J. Canales  
juan.canales@utas.edu.au

### Specialty section:

This article was submitted to  
Addictive Disorders,  
a section of the journal  
Frontiers in Psychiatry

**Received:** 05 June 2019

**Accepted:** 21 January 2020

**Published:** 19 February 2020

### Citation:

Rosel JF, Elipe-Miravet M, Elósegui E,  
Flor-Arasil P, Machancoses FH,  
Pallarés J, Puchol S and Canales JJ  
(2020) Pooled Time Series Modeling  
Reveals Smoking Habit  
Memory Pattern.  
Front. Psychiatry 11:49.  
doi: 10.3389/fpsy.2020.00049

Smoking is a habit that is hard to break because nicotine is highly addictive and smoking behavior is strongly linked to multiple daily activities and routines. Here, we explored the effect of gender, age, day of the week, and previous smoking on the number of cigarettes smoked on any given day. Data consisted of daily records of the number of cigarettes participants smoked over an average period of 84 days. The sample included smokers (36 men and 26 women), aged between 18 and 26 years, who smoked at least five cigarettes a day and had smoked for at least 2 years. A panel data analysis was performed by way of multilevel pooled time series modeling. Smoking on any given day was a function of the number of cigarettes smoked on the previous day, and 2, 7, 14, 21, 28, 35, 42, 49, and 56 days previously, and the day of the week. Neither gender nor age influenced this pattern, with no multilevel effects being detected, thus the behavior of all participants fitted the same smoking model. These novel findings show empirically that smoking behavior is governed by firmly established temporal dependence patterns and inform temporal parameters for the rational design of smoking cessation programs.

**Keywords:** tobacco, pooled time series, panel time series, intensive data analysis, memory, multilevel regression

## INTRODUCTION

Patterned behavior can be predicted on the basis of the behavior displayed in the immediate, or even remote, past. However, very few longitudinal studies in psychology and health sciences have used this approach to characterize specific behaviors, problematic or otherwise, although several authors have highlighted the usefulness of this kind of methodology (1–4).

We propose here that habitual smoking builds a temporally distinct smoking “memory”, defined, from a point of view of statistical time series, as the maximal distance (lags) from which the values of the series can be confidently predicted. For example, if a temporary process was measured in days, and the maximum delay to accurately forecast the series was five lags, it would follow that the memory of

the series, this is, the autoregressive process (AR), is 5 days (AR (5)). This statistical concept is increasingly used in time series as applied in physical, social, engineering, and statistical sciences (5–8). Indeed, the designers of time series analysis have used this notion to define short and long memory processes (9). Our aim here is to temporally define the smoking habit memory pattern, understood as a statistical modeling of habitual smoking, and determine the maximum delay (behavioral memory) of the model.

Although the number of longitudinal studies has risen in the last 15 years (2, 4), very little research involving pooled time series (i.e. panel analysis, or different subjects measured several times over time) provides evidence that nicotine dependence is a function of smokers' AR memory. Three models have been proposed to explain the possible relationships between nicotine and the smoking habit (10). Velicer et al. (11) later reviewed the three models and assigned a time series model to each of them (11). The first model was the "nicotine fixed-effect" model, which has positive AR coefficients. A second model was the "nicotine regulation model," which assumes a "zero" autocorrelation. The third model was the "multiple regulation model," where the levels of smoking would fluctuate so that a reduction in smoking would give rise to a greater desire to smoke, which has negative AR coefficients.

The study by Velicer et al. (11) is one of only two studies in the literature that are similar to the present research. Velicer et al. (11) worked with a sample of ten smokers who recorded data twice a day for 62 days. In the other study (12), a sample of 29 participants was used, who recorded the number of cigarettes smoked over a period of 84 days. These previous studies showed that the memory for smoking spans 1 or 2 days, in one case, and 7 and 14 days, in the other. It is worth noting, however, that Velicer et al. (11) did not analyze the data in search for weekly seasonality, unlike Rosel and Elósigui (12), which can impact on the number of significant lags that are found.

The analyses performed to date on smoking habits have been mostly of a univariate nature, this is, they include only one person per analysis and, as established by Makridakis (13) and Box and Jenkins (14), have little statistical (9). The statistical power of a test is the probability that the test rejects the null hypothesis when the specific alternative hypothesis is true (15), so if a statistical test has low power (as in univariate times series estimation of parameters) it is more difficult to obtain significant results. The statistical power increases, however, with the use of pooled time series analysis (16–24).

From a cognitive approach to the study of addiction habits, there are two models of behavior subject to regularities. The first model includes components that allow it to be considered an incremental, persistent habit, or a "model-free system," that is largely insensitive to outcome devaluation and, in this case, the AR coefficients will all be positive. However, if the behavior also involves evaluation and decision making directed toward achieving certain objectives, such as trying to reduce the number of cigarettes smoked per day, or quitting the habit, then the behavior would be an example of a "model-based system" (25–29). In time series this implies that they will present one or several negative AR coefficients.

## Hypotheses

Several hypotheses were postulated for this study. The main hypothesis was that the smoking habit is expected to have an AR component with respect to previous days (9, 30). In other words, the number of cigarettes smoked on any given day will be a function of the number of cigarettes smoked on previous days, but could also vary according to weekly memory patterns<sup>1</sup>. Please note we have removed the equations from the main text and included them as footnotes to facilitate readability for readership not specialized in time series.

Similarly, it is assumed that the values of the coefficients will be positive in accordance with the nicotine fixed-effect model (11) or with the model-free system (25, 28, 29).

The data collection model had two levels. The first refers to the number of cigarettes smoked daily, while the second level is the person, and thus each daily item of data belongs to a particular smoker (31–38). Hence, the hypothesis here is that the coefficients  $b_0$ ,  $b_1$  of the lag  $Y_{t-1,j}$  and  $b_7$  of the lag  $Y_{t-7,j}$  differ from one participant to another<sup>2</sup>.

It is also hypothesized that the number of cigarettes smoked is a function of the day of the week, for which six dummy variables were created, taking Sunday as the reference<sup>3</sup>.

We tested whether the variables gender and age influenced the number of cigarettes smoked each. Both are level 2 variables because they have a fixed value for each person, this is, these values do not change for each of the participants ( $\gamma_0$ ), but there is a random component for each individual ( $u_{Gj}$ )<sup>4</sup>.

<sup>1</sup>This could be expressed with the following equation:

$$Y_{t,j} = f(Y_{t-1,j}, Y_{t-2,j}, \dots, Y_{t-7,j}, Y_{t-14,j}, \dots, Y_{t-7,k,j}), \quad (1)$$

<sup>2</sup>This could be expressed with the following set of equations:

$$\begin{aligned} b_{0j} &= \gamma_{00} + u_{0j} \\ b_{1j} &= \gamma_{10} + u_{1j} \\ b_{7j} &= \gamma_{70} + u_{7j} \end{aligned} \quad (2)$$

where  $\gamma_{00}$  would now be the fixed part of the intercept and is equivalent to the mean of the level 2 intercepts. The value  $u_{0j}$  would be the random part of the second-level intercept of the participants in the sample, and there would be as many intercepts as there are participants. The same interpretation is assumed for  $b_{1j}$  and  $b_{7j}$ .

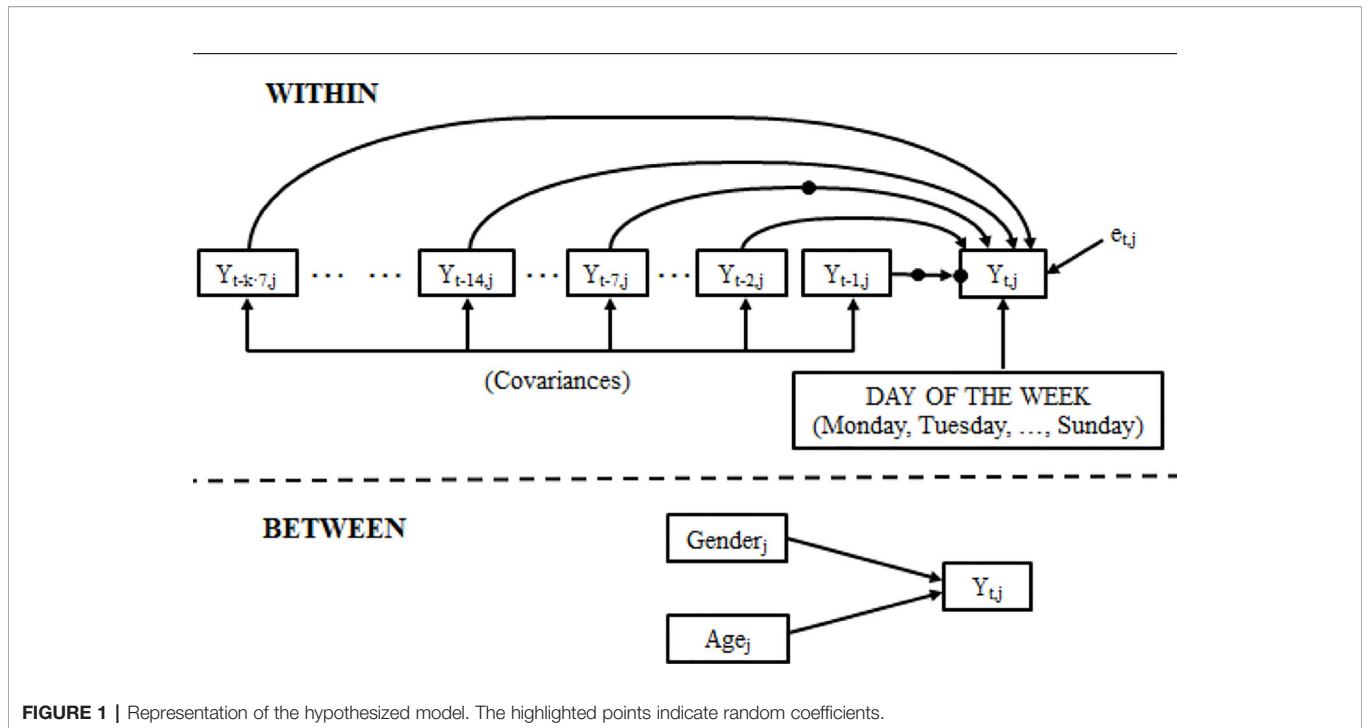
<sup>3</sup>Being the codes:

$$\text{Day of the week} = [D_{\text{Mon}}, D_{\text{Tue}}, D_{\text{Wed}}, \dots, D_{\text{Sat}}] \quad (3)$$

<sup>4</sup>Hence, if the coefficients for gender and age are  $b_{G,j}$  and  $b_{Ed,j}$ , their corresponding multilevel coefficients will be:

$$\begin{aligned} b_{G,j} &= \gamma_{0G} + u_{G,j} \\ b_{Ed,j} &= \gamma_{0Ed} + u_{Ed,j} \end{aligned} \quad (4)$$

the properties of the terms in Equation 4 being similar to those in Equation 2.



The hypotheses with the final model can be expressed with the following equation<sup>5</sup> and are shown in a simplified manner in **Figure 1** (39).

We have included hypothetical *constraints* for the data analysis, which were laid out in order to simplify the model in coherence with it being a time series. The constraints were:

- a. the greater the elapsed time, the smaller the non-seasonal *time effects* are for the days immediately prior to the measurement of  $Y_t$  ( $Y_{t-1}$ ,  $Y_{t-2}$ , ...); this is, the greater the time interval between the moment of the lagged measurement ( $Y_{t-j}$ ) and the moment of the forecast ( $Y_t$ ), the smaller the temporal memory of the system will be. Likewise, the greater the amount of time elapsed, the less effect of the weekly seasonality ( $Y_{t-7}$ ,  $Y_{t-14}$ ,  $Y_{t-21}$ , ...,  $Y_{t-7k}$ ) will have on  $Y_t$  (9, 40)<sup>6</sup>.
- b. equality of the *variances* of the temporal independent variables, since the same lagged variable is used<sup>7</sup>.
- c. the *covariances* between the pairwise temporal independent variables are assumed to be equal, based on the distance (in days) between the lags<sup>8</sup>.

<sup>5</sup>General equation with the hypotheses:

$$Y_{t,j} = (\gamma_{00} + u_{0j}) + (\gamma_{10} + u_{1j}) \cdot Y_{t-1,j} + b_2 \cdot Y_{t-2,j} + b_3 \cdot Y_{t-3,j} + \dots + (\gamma_{70} + u_{7j}) \cdot Y_{t-7,j} + b_{14} \cdot Y_{t-14,j} + \dots + b_{k \cdot 7} \cdot Y_{t-7k,j} + [b_{Mon} \cdot D_{Mon} + b_{Tue} \cdot D_{Tue} + \dots + b_{Sat} \cdot D_{Sat}] + (\gamma_{0G} + u_{Gj}) \cdot Gender + (\gamma_{0Ed} + u_{Edj}) \cdot Age + e_{t,j} \quad (5)$$

<sup>6</sup>Constraints on the coefficients

$$b_1 > b_2 > b_3 > \dots \text{ (for the simple AR components)}$$

$$b_7 > b_{14} > \dots > b_{7 \cdot k} \text{ (for the 7 - day seasonal AR components)} \quad (6)$$

<sup>7</sup>Equations:

$$\text{Var}(Y_{t-1}) = \text{Var}(Y_{t-2}) = \dots = \text{Var}(Y_{t-7}) =$$

$$\text{Var}(Y_{t-14}) = \text{Var}(Y_{t-21}) = \dots = \text{Var}(Y_{t-7k}) \quad (7)$$

<sup>8</sup>Hence, for the covariances between successive days:

$$\text{Cov}(Y_{t-1}, Y_{t-2}) = \text{Cov}(Y_{t-2}, Y_{t-3}) = \dots = \text{Cov}(Y_{t-h}, Y_{t-h-1}).$$

For the covariances between each set of 7 days, 7-k being the last lag in the series:

$$\text{Cov}(Y_{t-7}, Y_{t-14}) = \text{Cov}(Y_{t-14}, Y_{t-21}) = \text{Cov}(Y_{t-21}, Y_{t-28}) = \dots$$

$$= \text{Cov}(Y_{t-7k+7}, Y_{t-7k}).$$

For the covariances between each set of 14 days:

$$\text{Cov}(Y_{t-7}, Y_{t-21}) = \text{Cov}(Y_{t-14}, Y_{t-28}) = \dots = \text{Cov}(Y_{t-7k+14}, Y_{t-7k}).$$

In the same way, the covariances with a distance of 21 days between each other have been equalised:

$$\text{Cov}(Y_{t-7}, Y_{t-28}) = \dots = \text{Cov}(Y_{t-7k+21}, Y_{t-7k});$$

$$\text{between 28 days: } \text{Cov}(Y_{t-7}, Y_{t-35}) = \dots = \text{Cov}(Y_{t-7k+28}, Y_{t-7k});$$

$$\text{and between 35 days: } \text{Cov}(Y_{t-7}, Y_{t-42}) = \dots = \text{Cov}(Y_{t-7k+35}, Y_{t-7k}). \quad (8)$$

## METHOD

### Participants

Smokers were first- or second-year students of the Psychology and Child Education degrees of the University of Málaga (Spain). Participation in the study was voluntary. The research was carried out with the approval of the University of Málaga Human Ethics Committee, and each participant signed an informed consent document prior to taking part in the study. In order to be included in the sample participants had to meet three requirements: (i) they must have been smoking for more than two consecutive years prior to data collection, (ii) they had to smoke more than five cigarettes a day, and (iii) they had to fill in a smoking record sheet for at least 63 consecutive days. Participants were not screened for health conditions, including psychological or neurological disorders. Of the 310 students that expressed an interest in the study, 99 (31.9%) stated that they were habitual smokers.

Finally, 80 students took part in the study. However, 13 of them did not submit fully completed records, and 5 participants were excluded because they smoked an average of fewer than five cigarettes per day during the study. The final study sample therefore consisted of 62 participants, of whom 36 were male and 26 were female (57.1% and 42.9% of the final sample, respectively). The participants were between 18 and 26 years of age, the mean being 20.44 years for the male, 20.37 years for the female and 20.41 years for the sample as a whole. A sample of 62 participants is considerable, generating a total sample with a chain of 5159 longitudinal data.

### Procedure

Each participant recorded the number of cigarettes smoked by means of a daily self-report chart that was filled in each night. Recordings took place over a period that ranged between 69 to 91 days, the mode being 84 days, as 46 participants submitted a record of 84 days (74.2%).

A total of 5159 daily smoking data items were submitted but some items, 1.08% of the total, were missing. We did not estimate missing values because statistical systems for estimating missing values are designed for cross-sectional data, not longitudinal data.

### Data Analysis

The data for the analysis in this research were transcribed and entered in a statistics program (41). Descriptive analyses were also performed in the same program. Once the data had been introduced with a “person-level data set” system, they were transferred to a “longitudinal time record” system (42), each record corresponding to a day of smoking, and were organized

for analysis with a pooled time series system (17, 43, 44). After organizing the data, they were analyzed using the statistics software package Mplus (39). The value  $\alpha = .05$  was taken as the level of significance. Data, input, and output files in Mplus can be obtained from this website: <http://repositori.uji.es/xmlui/handle/10234/180682>.

## RESULTS

**Table 1** shows the mean and the SD of the number of cigarettes smoked by the participants every day of the week. The day on which the participants in the sample smoked the least was Monday, the figure rising slowly day by day until Saturday, which was the day they smoked the most. The mean of the number of cigarettes smoked per day by the participants was 15.664 ( $SD = 8.257$ ). Levene's test was performed ( $F(1, 46) = 1.187, p = .310$ ) on the number of cigarettes smoked daily as a function of the day of the week. The assumption of homogeneity of the variances was met.

The intraclass correlation (ICC) was .802. The consistency between the measures was above 75% and therefore it represented excellent intraclass reliability (45).

**Table 1** also shows the means and the SD of the four participants with the lowest mean and the four with the highest mean number of cigarettes smoked per day. Due to the large difference in the SD of the participants, we tested the homoscedasticity of the dependent variable. The result showed that the data were heterogeneous ( $F(61, 5097) = 291.04, p < .001$ ), and therefore the variance of the series as a function of the subjects was not constant.

The systems for estimating parameters based on the normality of the residuals are not very efficient when there is

**TABLE 1B** | Participants ordered according to their direct and homoscedastic mean and SD, from lowest to highest.

Participant	Raw data	Homoscedastic data
	$Y_{t,j}$	$Y_{t,j}^H$
	M (SD)	M (SD)
18	5.56 (2.99)	1.86 (1.00)
35	5.79 (2.29)	2.53 (1.00)
19	6.21 (1.73)	3.59 (1.00)
1	6.27 (1.88)	3.34 (1.00)
–	–	–
45	26.11 (5.45)	4.72 (1.00)
51	30.57 (5.77)	5.29 (1.00)
71	35.13 (5.28)	6.656 (1.00)
24	42.11 (7.56)	5.567 (1.00)

**TABLE 1A** | Means and SD of the number of cigarettes smoked by the sample.

Measure	Day of the week						
	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
Mean	14.38	15.03	15.11	15.03	16.99	17.62	15.46
SD	8.35	8.06	7.88	8.07	8.29	8.42	8.25

no homogeneity of variances, as they can lead to type II errors, or failure to reject a false null hypothesis (46, 47). Additionally, because the subjects have different variances, the precision of the forecasts will be different depending on the individuals, thereby giving rise to variances of different magnitudes in the residuals. The decision was therefore made to transform each value of  $Y_{t,j}$  by dividing it by the corresponding standard deviation of each participant ( $s_j$ ).

$$\frac{Y_{t,j}}{s_j} = Y_{t,j}^H \tag{9}$$

By so doing, each raw value is converted into a new one, but now the variances of  $Y_{t,j}^H$  are homogeneous and equal to unity:  $\text{Var}(Y_{t,j}^H) = \text{Constant} = 1$ . We will call the new transformed variable “homoscedastic number of cigarettes per day” ( $Y_{t,j}^H$ ).

Note in the column  $Y_{t,j}^H$  in **Table 1** that each participant now has a variance equal to one, the new homoscedastic mean is 4.546, the minimum homoscedastic mean is 1.860 and the maximum is 6.657, while the total variance of  $Y_{t,j}^H$  is 2.253, which is not equal to one because not all the subjects have the same mean.

In order to determine the maximum number of lags to be introduced into the model, the simple autocorrelation function and the partial autocorrelation function of the series of  $Y_{t,j}^H$  were calculated. It was found that the lags are significant up to 56 days, with a seasonal pattern being repeated every 7 days.

With regards to the estimation of the hypothesized model, Equation 5, which we will call Model 1 (M1), was tested (**Table 2**). Data showed that this model did not converge. Due to this, each of the level 2 independent variables were removed one by one: gender and age (since the estimation did not converge), and each of the random coefficients were also removed separately from the intercept, from lag 1 and from lag 7. Even so, the estimation model did not converge or was not significant. The fixed part of the model was left, and each of the level 2 variables (gender and age) and each of the random coefficients from Equations 2 and 4 were introduced separately, although it still failed to converge. The overall fit indicators of each estimated model are shown in **Table 2**.

Finally, only the level 1 variables (intercept, days of the week dummy and lags) were left, and this model was called model 2 (M2). **Table 2** shows the results of M2, where it can be observed that all the variables are significant at  $\alpha = .05$ .

Additionally, in order to determine the exact probability of the overall variable day of the week, model 3 (M3) was estimated, which was made up of only the intercept and the AR values.

Thus, as M3 is nested in M2 (48, 49), the difference in  $\chi^2$  is 511,394 ( $\Delta\chi^2$  (M2,M3)), with a difference in the number of degrees of freedom ( $\Delta df$ ) of 60, which yields a  $p < .001$ . Thus, the variable day of the week is significant ( $p < .001$ ). Likewise, in order to test the exact significance of the overall fit of M2, the null model (M0) of M2 was used as the reference model, this is, with all the M2 variables, but indicating that the covariances of all the variables are equal to zero. It was found that  $\Delta\chi^2$  (M0,M2) = 50849.003, and that  $\Delta df$  (M0,M2) = 47, which means that the probability of these differences is  $p < .001$ , indicating that the overall fit of M2 is significant.

Residuals ( $e_{t,j}$  values) must be normal, and they should not be autocorrelated. The normality of the residuals was tested by applying the Kolmogorov–Smirnov test ( $p = .064$ ), and it can therefore be concluded that the residuals of M2 follow a normal distribution. To check that the residuals are “white noise” we applied the Ljung–Box test (50), the result of the  $\chi^2$  for three seasons, or 21 lags, being 27.055 ( $df = 21, p = .169$ ), which indicates that the residuals are indeed “white noise.”

Of all the models analyzed, we accept M2 as the model that best fits the data and our hypotheses because: (i) the overall indicators are good, except CFI, which is still acceptable, and, moreover, the test of the exact probability of the overall fit and the  $R^2$  of the dependent variable are both significant ( $R^2 = .837, p < .001$ ); (ii) the independent variables are significant; (iii) the AIC of M2 is lower than that of M3 (**Table 2**); and (iv) the residuals are normal and “white noise.” **Figure 2** show the results of M2.

## DISCUSSION

The results obtained in this research confirm the main hypotheses of the proposed model, this is, that the smoking habit has an AR component, specifically of 56 days, and that the day of the week leads to a change in the number of cigarettes smoked. As stated in the introduction, previous studies found that participants followed AR models with a maximum of 14 days. In this study, however, we have found that participants follow a simple AR model of 2 days, to which we must add the seasonal pattern of 7 days with eight seasons, this is, a total of 56 days. This finding suggests that the number of cigarettes a person smokes on a given day is influenced by the day of the week and by the number of cigarettes smoked 1, 2, 7, 14, 21, and so forth up to 56, days before. This implies that the implicit memory for smoking spans 56 days. Furthermore, all the AR coefficients were

**TABLE 2A** | Summary of the models.

Models	df	$\chi^2$	p	RMSEA (90% CI <sup>a</sup> )	CFI	TLI	SRMR	AIC
M1	Does not converge							
M2	89	554.797	<.001	.056 (.051–.060)	.848	.973	.036	56002.439
M3	29	43.403	<.042	.017 (.003–.027)	.995	.998	.014	56016.744
M0 <sup>b</sup>	136	51403.800	<.001	.355 (.353–.358)	.000	.000	.471	160384.664

<sup>a</sup>CI, Confidence Interval. <sup>b</sup>M0, Null model with respect to M2.



**TABLE 2B** | Coefficients of the regression model, homoscedastic scores.

Variable	Estimate	SE	t	$\beta$ (standardized)	p-value
Intercept	.2618	.078	3.348	.1164	.001 <sup>a</sup>
$Y_{t-1}^H$	.2468	.023	10.624	.2485	<.001 <sup>b</sup>
$Y_{t-2}^H$	.0983	.022	4.501	.0990	<.001 <sup>b</sup>
$Y_{t-7}^H$	.1934	.023	8.494	.1947	<.001 <sup>b</sup>
$Y_{t-14}^H$	.0867	.024	3.662	.0873	<.001 <sup>b</sup>
$Y_{t-21}^H$	.0716	.023	3.065	.0721	.001 <sup>b</sup>
$Y_{t-28}^H$	.0653	.016	4.179	.0657	<.001 <sup>b</sup>
$Y_{t-35}^H$	.0652	.016	4.173	.0656	<.001 <sup>b</sup>
$Y_{t-42}^H$	.0523	.011	4.792	.0526	<.001 <sup>b</sup>
$Y_{t-49}^H$	.0522	.011	4.783	.0525	<.001 <sup>b</sup>
$Y_{t-56}^H$	.0521	.011	4.778	.0524	<.001 <sup>b</sup>
Day of the week ( $\Delta(\chi^2) = 511.394, \Delta(df) = 60$ )					<.001 <sup>c</sup>
D <sub>Mon</sub>	-.3845	.084	-4.563	-.0590	<.001 <sup>a</sup>
D <sub>Tue</sub>	-.2052	.085	-2.403	-.0316	.016 <sup>a</sup>
D <sub>Wed</sub>	-.1910	.084	-2.264	-.0299	.024 <sup>a</sup>
D <sub>Thu</sub>	-.1479	.084	-1.765	-.0231	.078 <sup>a</sup>
D <sub>Fri</sub>	-.0449	.087	-.519	-.0070	.604 <sup>a</sup>
D <sub>Sat</sub>	-.0732	.086	-.847	-.0114	.397 <sup>a</sup>

<sup>a</sup>Two-tailed probability.

<sup>b</sup>One-tailed probability, positive value.

<sup>c</sup>Calculated by means of the difference between M2 and M3.

positive, confirming the “nicotine fixed-effect model” (11), also referred to as the model-free smoking habit (27).

The fact that people smoke with a regular weekly pattern indicates the extent to which their habits are conditioned and subject to unwritten behavioral norms and other norms that are driven by social demands with respect to their routines, e.g. meeting up with friends to go out at the weekend and smoke. Likewise, when a person smokes, he or she incorporates different aspects of his or her behavior into daily life: physical behavior, behavior aimed at apparently reducing stress through smoking tobacco, and habits associated with tobacco (hours of smoking;

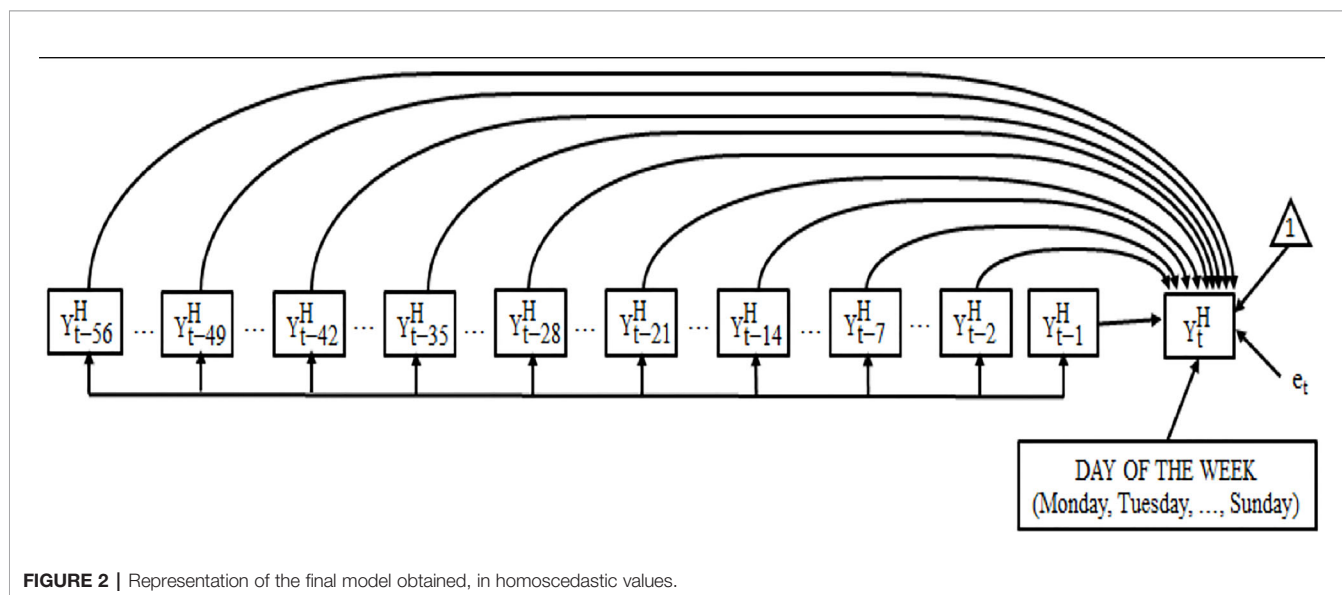
other simultaneous, previous or subsequent consumptions, such as alcohol or coffee; smoking after certain meals; etc.).

In the present study all subjects presented the same statistical model of smoking, with daily and weekly components, suggesting that participants responded similarly to a constellation of personal (Pavlovian habits, abstinence syndrome, particular response to stress, etc.), contextual (break-time, drinking coffee or alcohol, after meals, etc.) and social (co-workers, friends who smoke, parties, holidays, etc.) stimuli (28, 29).

One of the possible reasons that may explain why the multilevel hypotheses of the model were not confirmed could be the actual memory of the data. Lagging the dependent variable significantly up to 56 days means that the data have a very high memory load, thereby reproducing the level of the y-intercept through the level of lags, and this would therefore mean that the model does not require a multilevel model for the y-intercept or for the coefficients of lags 1 and 7.

The findings of this research must be evaluated within the context of several methodological limitations. One of them is the generalizability or external validity, since the sample used consisted exclusively of university students, this is, mostly young people. Furthermore, the model obtained is the same for all the participants in our sample, regardless of the smoker's gender or age. Given that the sample age was between 18 and 26 years, it is possible that if we had taken a sample with more variability in age, this variable could have been significant. More broadly, the fact that the sample was composed of young adults (i.e. university students) limits the generalizability of the findings. As an additional drawback, it is difficult to carry out a direct comparison between our results and those of other studies because, first, few studies have been conducted on this topic and, second, different methods of data analysis were used.

If the memory for smoking is 56 days, we might wonder whether, from a biological point of view, plasticity changes occur in brain areas associated with habit formation, including the



**FIGURE 2** | Representation of the final model obtained, in homoscedastic values.

basal ganglia and the prefrontal cortex (27, 51) that sustain such mnemonic and behavioral processes. This warrants further investigation. The data presented here with regards to the timeframe that allows smoking behavior to be confidently predicted (i.e., 8 weeks), which underlies the presence of a temporally resilient memory of the smoking habit, combined with the elucidation of the underlying neurobiological processes involved, could potentially lead to the design of more effective smoking cessation interventions. Such interventions should take into consideration these temporal parameters. Indeed, previous studies have shown that for cessation of smoking to be more effective, patients must engage for at least eight weeks in some kind of smoking cessation program because otherwise they are more likely have a relapse in the 12 months following the start of the program (52, 53). The present findings would therefore be in agreement with the conclusions of these studies. Taken together, these findings suggest that healthcare professionals should plan therapies or treatments, with their associated follow-ups, to last no less than 8 weeks in order to achieve higher percentages of success in giving up the smoking habit and preventing relapses both during that time and in the following 12 months.

In summary, the present study shows a range of novel findings. First, we found significant results until lag 56, this is, more lags than in previous studies, indicating that a person's memory for smoking is longer-lasting than previously thought. Second, we observed that the day of the week also influences the smoking habit. Third, the multilevel hypotheses on the coefficients were not confirmed, and no differences in smoking

habits were found according to gender or age, indicating that the behavior is common to all the individuals in the sample. Further research is needed to be able to generalize the results presented here and to apply them in future studies on prevention as well as treatment for smokers.

## DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the Repositori Universitat Jaume I at <http://repositori.uji.es/xmlui/handle/10234/180682>.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Málaga Human Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

EE collected the data. JR, ME-M, PF-A, FM, JP and SP analysed the data. JR, ME-M and JC wrote the manuscript.

## REFERENCES

- Weiner I, Weiner I, Schmitt N, Highhouse S, Schinka J, Velicer W. *Handbook of psychology*. Hoboken, NJ: Wiley (2013).
- Liu Y, West S. Weekly cycles in daily report data: an overlooked issue. *J Pers* (2015) 84(5):560–79. doi: 10.1111/jopy.12182
- Schuurman N, Ferrer E, de Boer-Sonnenschein M, Hamaker E. How to compare cross-lagged associations in a multilevel autoregressive model. *psychol Methods* (2016) 21(2):206–21. doi: 10.1037/met0000062
- Hamaker E, Wichers M. No time like the present. *Curr Dir In psychol Sci* (2017) 26(1):10–5. doi: 10.1177/0963721416666518
- Gottman JM. *Time-series analysis: A comprehensive introduction for social scientists*. Cambridge, UK: Cambridge University Press (1981).
- Mudelsee M. *Climate time series analysis: Classical statistical and bootstrap methods*. Heidelberg, Germany: Springer (2014).
- Hassler U. *Time series analysis with long memory in view*. Hoboken, NJ: J. Wiley (2019). doi: 10.1002/9781119470380
- Sahoo BB, Jha R, Singh A, Kumar D. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. *Acta Geophys*. (2019) 67:1471–81. doi: 10.1007/s11600-019-00330-1
- Box GEP, Jenkins GM, Reinsel GC, Ljung GM. *Time series analysis: Forecasting and control*. 5th edition. Hoboken, NJ: Wiley (2016).
- Leventhal H, Cleary PD. The smoking problem: A review of the research and theory in behavioral risk modification. *psychol Bull* (1980) 88:370–405. doi: 10.1037/0033-2909.88.2.370
- Velicer W, Redding C, Richmond R, Greeley J, Swift W. A time series investigation of three nicotine regulation models. *Addictive Behav* (1992) 17(4):325–45. doi: 10.1016/0306-4603(92)90039-X
- Rosel J, Elósegui E. Daily and weekly smoking habits: a Box-Jenkins analysis. *psychol Rep* (1994) 75(3):1639–48. doi: 10.2466/pr0.1994.75.3f.1639
- Makridakis S. *Forecasting: Methods and applications*. New York: Wiley. (1998).
- Box GEP, Jenkins GM. *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day. (1976).
- Kraemer HC, Blasey C. *How many subjects?: Statistical power analysis in research*. Newbury Park, CA: Sage (2016). doi: 10.4135/9781483398761
- Dielman TE. *Pooled cross-sectional and time series data analysis*. New York: Marcel Dekker (1989).
- Says LS. *Pooled time series analysis*. Newbury Park, CA: Sage (1989). doi: 10.4135/9781412985420
- Moore K, Osgood D, Larzelere R, Chamberlain P. Use of pooled time series in the study of naturally occurring clinical events and problem behavior in a foster care setting. *J Consult. Clin Psychol* (1994) 62:718–28. doi: 10.1037/0022-006X.62.4.718
- Johnson DR. Alternative methods for the quantitative analysis of panel data in family research: pooled time-series models. *J Marriage Family* (1995) 57:1065–77. doi: 10.2307/353423
- Soliday E, Moore K, Lande M. Daily reports and pooled time series analysis: Pediatric psychology applications. *J Pediatr Psychol* (2002) 27:67–76. doi: 10.1093/jpepsy/27.1.67
- Sun Y, Li Y. Children's well-being during parents' marital disruption process: a pooled time-series analysis. *J Marriage Family* (2002) 64:472–88. doi: 10.1111/j.1741-3737.2002.00472.x
- Hoepfner B, Goodwin M, Velicer W, Heltshe J. An applied example of pooled time series analysis: cardiovascular reactivity to stressors in children with autism. *Multivariate Behav Res* (2007) 42:707–27. doi: 10.1080/00273170701755291
- Winquist A, Klein M, Tolbert P, Sarnat S. Power estimation using simulations for air pollution time-series studies. *Environ Health* (2012) 11(1):1–12. doi: 10.1186/1476-069X-11-68

24. Schultzeberg M, Muthén B. Number of subjects and time points needed for multilevel time-series analysis: A simulation study of dynamic structural equation modeling. *Struct Equation Model.* (2017) 25(4):495–515. doi: 10.1080/10705511.2017.1392862
25. Neal DT, Wood W, Labrecque JS, Lally P. How do habits guide behavior? Perceived and actual triggers of habits in daily life. *J Exp Soc Psychol* (2012) 48:492–8. doi: 10.1016/j.jesp.2011.10.011
26. Dayan P, Berridge KC. Model-based and model-free pavlovian reward learning: reevaluation, revision and revelation. *Cognitive Affect Behav Neurosci* (2013) 14(2):473–92. doi: 10.3758/s13415-014-0277-8
27. Everitt BJ, Robbins TW. Drug addiction: updating actions to habits to compulsions ten years on. *Annu Rev Psychol* (2016) 67:23–50. doi: 10.1146/annurev-psych-122414-033457
28. Wood W, Rünger D. Psychology of habit. *Annu Rev Psychol* (2016) 67:289–314. doi: 10.1146/annurev-psych-122414-033417
29. O'Doherty JP, Cockburn J, Pauli WM. Learning, reward, and decision making. *Annu Rev Psychol* (2017) 68:73–100. doi: 10.1146/annurev-psych-010416-044216
30. McCleary R, McDowal D, Bartos BJ. *Design and analysis of time series experiments.* New York: Oxford University Press (2017). doi: 10.1093/oso/9780190661557.001.0001
31. Goldstein H, Healy MJR, Rasbash J. Multilevel time series models with applications to repeated measures data. *Stat In Med* (1994) 13:1643–55. doi: 10.1002/sim.4780131605
32. Leyland AH, Goldstein H. *Multilevel modelling of health statistics.* Chichester: John Wiley (2001).
33. Raudenbush SW, Bryk AS. *Hierarchical linear models: Applications and data analysis methods.* Newbury Park, CA: Sage (2002).
34. Goldstein H. *Multilevel statistical models.* Chichester, UK: Wiley (2011). doi: 10.1002/9780470973394
35. Snijders T, Bosker R. *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage (2011).
36. Hox JJ, Roberts J. *Handbook of advanced multilevel analysis.* Hoboken, NJ: Taylor and Francis (2015).
37. Hox JJ, Moerbeek M, Schoot RVD. *Multilevel analysis. Techniques and applications. 3rd ed.* Mahwah, NJ: Lawrence Erlbaum (2017). doi: 10.4324/9781315650982
38. Spieler I, Scheibe S, Stamov-Roßnagel C, Kappas A. Help or hindrance? Day-level relationships between flextime use, work–nonwork boundaries, and affective well-being. *J Appl Psychol* (2017) 102(1):67–87. doi: 10.1037/apl0000153
39. Muthén LK, Muthén BO. *Mplus user's guide. 8th ed.* Los Angeles, CA: Muthén & Muthén (2017).
40. Rosel J, Plewis I. Longitudinal data analysis with structural equations. *Methodol: Eur J Res Methods Behav Soc Sci* (2008) 4(1):37–50. doi: 10.1027/1614-2241.4.1.37
41. SPSS. *IBM SPSS statistics 25.* Chicago, IL: SPSS Inc. (2017).
42. Singer JD, Willett JB. *Applied longitudinal data analysis: Modeling change and event occurrence.* New York: Oxford University Press (2003).
43. Andreß HJ, Golsch JK, Schmidt AW. *Applied panel data analysis for economic and social surveys.* Berlin: Springer (2013). doi: 10.1007/978-3-642-32914-2
44. Rosel JF, Jara P, Machancoses FH, Pallares J, Torrente P, Puchol S, et al. Intensive longitudinal modelling predicts diurnal activity of salivary alpha-amylase. *PloS One* (2019) 14(1):e0209475. doi: 10.1371/journal.pone.0209475
45. Fleiss J, Cohen J. The equivalence of weighted Kappa and the intraclass correlation coefficient as measures of reliability. *Educ psychol Meas.* (1973) 33(3):613–9. doi: 10.1177/001316447303300309
46. Gujarati DN, Porter DC, Gunasekar S. *Basic econometrics. 5th ed.* New Delhi: Tata McGraw Hill (2013).
47. Draper NR, Smith H. *Applied regression analysis.* New York: Wiley (2014).
48. Little TD. *Longitudinal structural equation modeling.* New York: Guilford (2013).
49. Kline RB. *Principles and practice of structural equation modeling.* New York: Guilford (2016).
50. Ljung GM, Box GEP. On a measure of a lack of fit in time series models. *Biometrika* (1978) 65(2):297–303. doi: 10.1093/biomet/65.2.297
51. Canales JJ. Stimulant-induced adaptations in neostriatal matrix and striosome systems: transiting from instrumental responding to habitual behavior in drug addiction. *Neurobiol Learn Memory* (2005) 83(2):93–103. doi: 10.1016/j.nlm.2004.10.006
52. Barrueco M, Torrecilla M, Maderuelo JA, Jiménez C, Hernández MA, Plaza MD. Two-month results are predictors of outcome in a smoking cessation program. *Med. Clin.* (2001) 116(7):246–50. doi: 10.1016/S0025-7753(01)71786-1
53. McAfee T, Bush T, Deprey T, Mahoney L, Zbikowski S, Fellows J, et al. Nicotine patches and uninsured quitline callers. *Am J Prevent. Med* (2008) 35(2):103–10. doi: 10.1016/j.amepre.2008.04.017

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Rosel, Elipe-Miravet, Elósegui, Flor-Arasil, Machancoses, Pallarés, Puchol and Canales. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.