

# Оценка на ефекта от пандемията COVID-19 върху броя на туристическите посещения на чужденци в България през 2020 г.

*гл. ас д-р Виктор Аврамов*

*Нов български университет*

**Резюме.** Целта на този анализ е оценката на ефекта върху броя на посещения на чужденци в България. Използвани са данни от НСИ, които са месечни в интервала от януари 2008 г. до декември 2020 г. В анализа се демонстрира, че данните имат силно изразена годишна сезонност. Това допуска те да бъдат прогнозираны за период от година напред сравнително лесно. Оценява се точността на прогнозирането и направената прогноза се използва за оценка на ефекта на пандемията за намаляването на туристическите пътувания.

**Ключови думи:** туристически посещения, пандемия, анализ на времеви редове.

## Assessment of the Effect of the COVID-19 Pandemic on the Number of Tourist Visits of Foreigners in Bulgaria in 2020

*Chief Assistant PhD Victor Avramov*

*New Bulgarian University*

**Summary.** The goal of this analysis is to quantify the effect on the number of visits of foreign tourists to Bulgaria. Time series data with monthly frequency in the interval from January 2008 to December 2020 is used. The analysis demonstrates that the data have a strong annual seasonality. This allows to predict tourist visits for a year ahead relatively easy. The accuracy of the forecast is assessed, and the forecast is used to quantify the

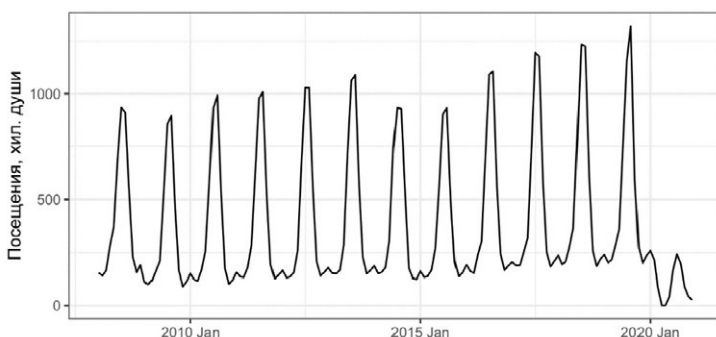
effect of the pandemic on the reduction of tourist travel.

**Keywords:** tourist visits, pandemic, time series analysis.

## 1. ДИНАМИКА НА ТУРИСТИЧЕСКИТЕ ПОСЕЩЕНИЯ НА ЧУЖДЕНЦИ В БЪЛГАРИЯ

Точната дефиниция на данните за туристически пътувания от категорията „Туризм“ в НСИ е „посещения на чужденци в България с цел почивка и екскурзия“ и „пътувания на български граждани в чужбина с цел почивка и екскурзия“<sup>1</sup>. Те са с честота от един месец. Съществено ограничение пред анализа на вторите е липсата на пълни данни за пътуванията на българските туристи извън граница след март 2020 г. Това поставя акцента тук върху чуждестранните туристи, но е естествено да се очаква, че именно те са от критично значение за българския туристически сектор.

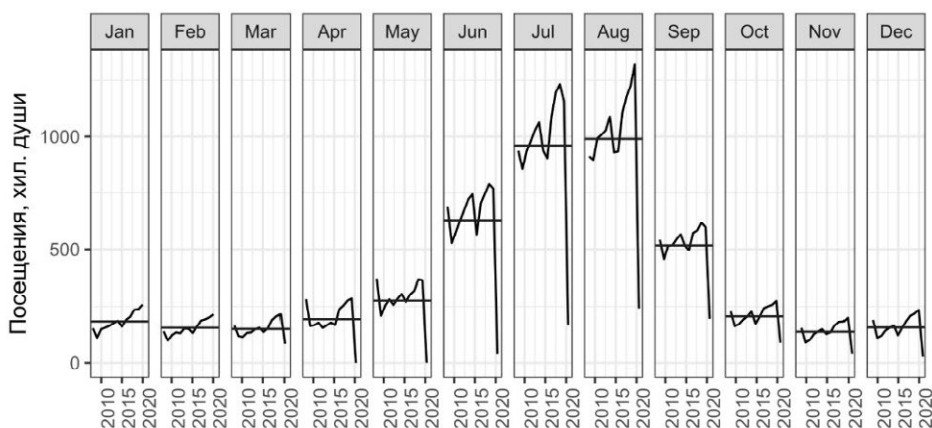
На Фигура 1 са визуализирани данните за посещения на чужденци в България за целия период, в който са налични. От графиката на фигурата веднага се вижда силно изразената повторямост на тези посещения. Съществуват силно изразени пикове, както и по-малки нараствания, които изглеждат регулярни и повтарящи се през фиксирани периоди. Вижда се, че има и по-дългосрочна динамика. На годишна база посещенията на чужденците нарастват до 2013 г., когато са общо 4 913 773, намаляват до 4 532 307 в следващата година и 4 269 102 в 2015 г. Оттогава растат до 2019 г., когато е отбелязан максимумът в наличните данни – 5 860 447. За 2020 г. те са 1 364 688, което е намаление с близо 77%.



**Фигура 1.**  
**Посещения на чужденци в България**

<sup>1</sup> Тези данни са достъпни на сайта на НСИ [прегледан на 13.04.2021 г.]: <https://www.nsi.bg/bg/content/1847/туризъм>

Използвайки Фигура 1 е трудно да се види спецификата на сезонността на посещенията на чужденци. Преди формално да се оцени сезонността в данните, на Фигура 2 се виждат средните нива на посещенията на чужденци, организирани по месеци за всички години. Този тип визуализация е полезна за открояване на повтарящата се месечна динамика. От графиките за всеки месец се вижда, че пиковете на посещенията на чужденци са основно в летните месеци юли и август, с по-значителни нива и в юни и септември. За останалите месеци от годината е трудно да се открие видимо нарастване, което да се асоциира с друг годишен сезон. Фигурата е създадена с използването на пакета *feasts* в R [1].



**Фигура 2. Месечна динамика на посещенията на чужденци в България, хил.**

## 2. СЕЗОННОСТ И СЕЗОННО ДЕКОМПОЗИРАНЕ

Преди да се пристъпи към използването на статистически критерии е добре да се дефинира сезонността: това е систематична, но не винаги постоянна динамика, причинена директно от метеорологичното време, календара или вземането на решения на икономическите субекти [2]. Данни, в които има изразена сезонност са например тримесечните данни за брутният вътрешен продукт, месечните данни за потреблението на електрическа енергия и много други. Важно е да се добави, че сезонността винаги настъпва на известни времеви интервали. За разлика от нея, цикличността

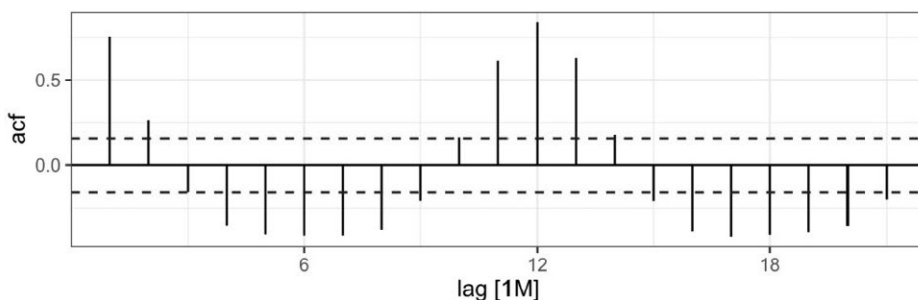
може да настъпва на нерегулярни такива [3]. Затова е добре „сезонност“ и „цикличност“ да не се употребяват като синоними.

В контекста на сезонността, един времеви ред обичайно се представя като група компоненти: сезонен, тренд и остатък:

$$y_t = S_t + T_t + R_t \quad (1)$$

където  $y_t$  е елемент от времевия ред във време  $t$ ,  $S_t$  е сезонния компонент,  $T_t$  е компонента на тренда и  $R_t$  е остатъка. Трендът е просто дългосрочната тенденция в данните, а остатъкът е този компонент от тях, които не се обяснява от предишните два.

Формалната проверка за наличие на сезонност може да се направи с помощта на статистически критерии. Когато данните имат изразена сезонност, трябва да се очаква, че автокорелацията между наблюденията  $y_t$  и  $y_{t-s}$  трябва да бъде по-висока от тази между наблюденията  $y_t$  и  $y_{t-k, k \neq s}$ . Използвайки това, графиката на функцията на автокорелацията на месечните посещения на чужденци е визуализирана на Фигура 3. Силно изразеното нарастване на автокорелациите на 12 лаг показва, че времевия ред има месечна сезонност.



**Фигура 3. Графика на функцията на автокорелацията на посещенията на чуждестранни туристи в България**

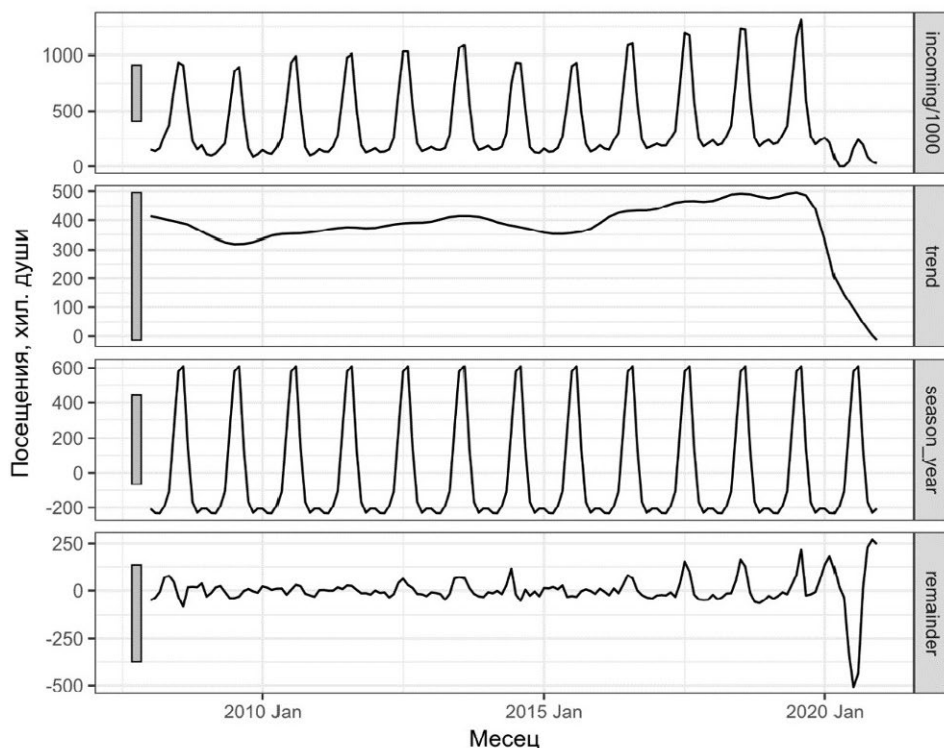
За проверка за наличие на автокорелация тук се използва критерият на Юнг–Бокс [4], в който нулевата хипотеза е за липса на автокорелация в серията. Броят на използваните лагове за проверката е равен на  $T/5$  и  $2s$ , където  $T$  е дължината на серията (общо 156 наблюдения) и  $s$  е дължината на сезонния цикъл (12 наблюдения). Данните в Таблица 1 показват, че и в двата случая нулевата хипотеза е отхвърлена, т.е. данните са автокорелирани, както се вижда и на Фигура 3.

**Таблица 1. Резултати от проверка на автокорелацията на серията с критерия на Юнг-Бокс**

lag	lb_stat	lb_pvalue
31	998	0
24	820	0

Сезонността може да се демонстрира и чрез сезонно декомпозиране. Тук се приема за достатъчно да се приложи сезонно и тренд декомпозиране, базирано на локална регресия“ („STL-декомпозиране“ оттук нататък) [5]. Отличаваща характеристика на STL-декомпозирането пред други методи е изглаждането на функцията на тренда, което е описано в [5]. За декомпозирането се използва пакетът feasts [1], който позволява процедурата да се автоматизира.

Резултатите от декомпозирането са визуализирани на Фигура 4.



**Фигура 4. STL-декомпозиране на посещенията на чуждестранни туристи**

STL-декомпозирането тук позволява специфициране на сезонен прозорец и прозорец на тренда – заедно с други опции. Изборът на сезонния прозорец зависи от бързината, с която теоретично се допуска да се изменя сезонността. Тук се приема за уместно да се избере периодична сезонност (s.window = „periodic”) при което сезонността не се изменя във времето.

Относителната тежест на тренда и на сезонния компонент може да се дефинират като:

$$F_t = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(T_t + R_t)}\right)$$

$$F_s = \max\left(0, 1 - \frac{\text{Var}(R_t)}{\text{Var}(S_t + R_t)}\right),$$

където  $F_t$  е мярка за тежестта на тренда, а  $F_s$  мярка за тежестта на сезонния компонент. И двете са в интервала между 0 и 1. За използваните данни тези тежести са съответно  $F_t = 0,572$  и  $F_s = 0,939$ . Това показва, че ролята на сезонния компонент във времевата серия е по-значителна от тази на тренда.

### 3. МОДЕЛИРАНЕ И ПРОГНОЗИРАНЕ

Моделирането може да се направи с използването на сезонен ARIMA( $p, d, q$ )( $P, D, Q$ ) $m$  модел, където  $p$  е степента на авторегресивния компонент,  $d$  е степента на диференциране,  $q$  е степента на пълзящата средна, като и трите се отнасят за несезонната част на модела.  $P, D, Q$  са съответните степени на сезонния модел, и  $m$  е броят на годишните наблюдения.

Времевият ред се тества за стационарност. Използва се критерият на Квятковски, Филипс, Шмит и Шин [6]. При два варианта на брой на лаговете, използвани за проверка –  $\sqrt[4]{4\left(\frac{n}{100}\right)}$  и  $\sqrt[4]{12\left(\frac{n}{100}\right)}$ , нулевата хипотеза за стационарност не може да се отхвърли.

Данните се разделят на два периода – за „трениране“ и за „тестване“. На първи етап, периодът за трениране е от януари 2008 до декември 2018 г., а за тестване – от януари 2019 до декември 2019 г. Това позволява на базата на историческите данни да се оцени точността на използвания модел.

Моделът, с който се извършва прогнозирането е селектиран на база минимизиране на критерия на Акаике (AICCS). Избраният по този критерий модел е ARIMA(1,0,1)(0,1,2)<sub>12</sub>. Критерият на Юнг-Бокс се използва за тестване за автокорелираност на остатъците от регресията. Добре е това

да се направи за група от лагове, защото може да се очаква че нивото на значимост няма да бъде константно навсякъде. Резултатите са систематизирани в Таблица 2. Те показват, че остатъците не са автокорелирани, както трябва да бъде.

**Таблица 2. Резултати от проверка на автокорелацията на серията с критерия на Юнг–Бокс**

lag	lb_stat	lb_pvalue
24	24,8	0,209
36	38,7	0,194
48	43,5	0,494

Точността на модела обичайно се оценява с използването на корена на средната квадратична грешка RMSE, която се дефинира като  $RMSE = \sqrt{mean(e_t^2)}$ , където  $e_t$  е моделната грешка. Точността на избрания модел е систематизирана в Таблица 3.

**Таблица 3. Прогнозна точност на използваният ARIMA модел**

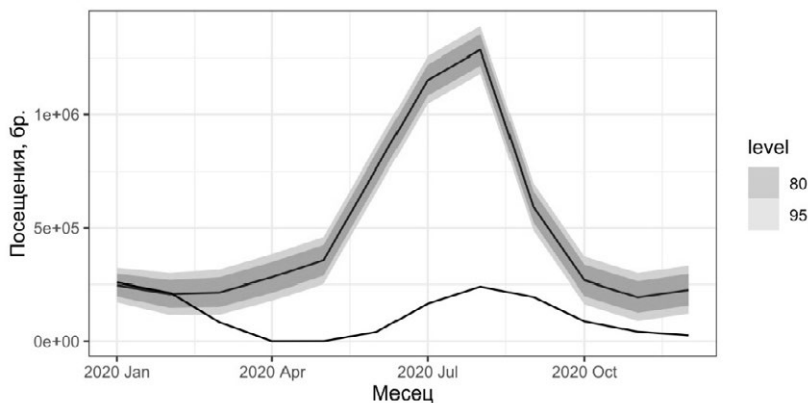
ME	RMSE	MAE	MPE	MAPE	MASE	RMSSE
10979	41557	19533	1.57	3.16	0.505	0.747

За интуитивно оценяване на точността може да се използва и MAPE (средната абсолютна процентна грешка), която е 3,16%.

На втори етап, за същинското прогнозиране, избраният модел се използва за прогнозиране на 2020 г. За целта периодът за трениране е от януари 2008 до декември 2019 г., а за тестване – от януари 2020 до декември 2020 г. Прогнозата е визуализирана на Фигура 5. Тя показва драматичната разлика между прогнозните и реалните туристически посещения. В Таблица 4 са систематизирани същите данни по месеци. Това помага да се разликата по месеци.

Както се вижда и от последната фигура, до месец март няма съществена разлика в броя посещения. Най-значителни са в следващите два месеца, в резултат на ограниченията върху пътуванията. Общо, разликата между реалните посещения и оценените с направената прогноза за цялата 2020 г. са 4 428 376. За сравнение, посещенията в 2011 г. са 4 484 248, а в

2014 г, както бе споменато, са 4 532 307. Може да се твърди, че 2020 г. е пропуснатата година за туристическата индустрия.



**Фигура 5. Прогнозни и реални посещенията на чуждестранни туристи в България за 2020 г.**

**Таблица 4. Прогнозни и реални посещенията на чуждестранни туристи в България за 2020 г. по месеци**

	Посещения	Прогноза	Разлика
Януари 2020	259133	248007	-11126
Февруари 2020	215259	208854	-6405
Март 2020	84726	216537	131811
Април 2020	586	281667	281081
Май 2020	873	356575	355702
Юни 2020	41240	761464	720224
Юли 2020	166766	1151426	984660
Август 2020	242468	1284186	1041718
Септември 2020	196003	593075	397072
Октомври 2020	88571	268539	179968
Ноември 2020	42491	195215	152724
Декември 2020	26572	227519	200947
Общо	1364688	5793064	4428376



## ЗАКЛЮЧЕНИЕ

Направената прогноза показва, че разликата между реалните туристически пътувания и очакваните при нормални обстоятелства са близки до общия брой туристически пътувания, отчетени в някои от изминалите години. Това оправдава оценките, направени публично от представители на туристическия бизнес.

Този извод трябва да е правдоподобен, защото туристическите посещения, отчетени с месечни данни, се отличават със значителна сезонност и могат да се прогнозират лесно. Това се вижда от статистическото моделиране, направено тук. Приложеният модел не се отличава с особена сложност и при нужда от повишаване на прогнозната точност може да се замени с друг. Но за целта на аргументите тук, може би направеното е достатъчно.

## ЛИТЕРАТУРА

- [1] M. O'Hara-Wild, R. J. Hyndman and E. Wang, „feasts: Feature Extraction and Statistics for Time Series“, 2021.
- [2] S. Hylleberg, *Seasonality in Regression*, Orlando: Academic Press, 1986.
- [3] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 3 ed., OTexts: Melbourne, Australia, 2021.
- [4] G. M. Ljung and G. E. P. Box, „On a measure of lack of fit in time series models,“ *Biometrika*, vol. 65, 1978.
- [5] R. B. Cleveland, W. S. Cleveland, J. E. McRae and I. Terpenning, „STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion),“ *Journal of Official Statistics*, vol. 6, p. 3–73, 1990.
- [6] M. O'Hara-Wild, R. J. Hyndman and E. Wang, „fable: Forecasting Models for Tidy Time Series,“ 2021.
- [7] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt and Y. Shin, „Testing the null hypothesis of stationarity against the alternative of a unit root. How sure are we that economic time series have a unit root?,“ *Journal of Econometrics*, vol. 54, 1992.