

УДК 519.254

ПРОГНОЗИРОВАНИЕ РЕЙТИНГА ТЕЛЕВИЗИОННЫХ КАНАЛОВ НА ОСНОВЕ ДАННЫХ МЕДИАИССЛЕДОВАНИЙ

¹Полежаев В.Д., ²Юсупова К.О.¹ФГБОУ ВО «Российский экономический университет им. Г.В. Плеханова»,
Москва, e-mail: vpolej@gmail.com;²ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)», Москва, e-mail: uko2304@mail.ru

Задача прогнозирования временных рядов имеет высокую актуальность для многих предметных областей и является неотъемлемой частью повседневной работы многих компаний. С использованием данных исследовательской компании АО «Медиаскоп» проведен анализ поведения телеаудитории на основе одного из важнейших показателей в медиапланировании – средней доли зрительской аудитории телеканалов. В ходе исследований было выявлено, что для прогнозирования доли зрителей телеканалов наиболее подходящими оказались методы, которые позволяют объединить модели авторегрессии и скользящего среднего. Применение таких методов является особенно актуальным для описания и прогнозирования процессов, в которых проявляются однородные колебания вокруг среднего значения. Показано, что построение прогнозов с учетом максимального количества факторов и особенностей, таких как конкуренция и сезонность, может быть реализовано с использованием адаптивных методов прогнозирования. При таком подходе особый интерес представляет использование методов идентификации параметров модели для составления максимально точного прогноза. Представлена прогнозная модель, построенная на основе реальных данных. Адекватность модели была проверена путем построения автокорреляционной функции для остатков. Автокорреляции не выходят за пределы допустимых интервалов, что говорит о независимости остатков. Также было проверено, что остатки распределены по нормальному закону. Выполнение этих двух критериев подтверждает адекватность полученной модели.

Ключевые слова: медиаконтент, медиапланирование, зрительская аудитория, прогнозирование, среднесуточная доля телеканала, рейтинг, телепрограмма

PREDICTION OF THE RATING OF TELEVISION CHANNELS BASED ON DATA FROM MEDIA RESEARCH

¹Polezhaev V.D., ²Yusupova K.O.¹Plekhanov Russian University of Economics, Moscow, e-mail: vpolej@gmail.com;²Moscow State Technical University, Moscow, e-mail: uko2304@mail.ru

Time series forecasting is of high relevance for many subject areas and is an integral part of the daily work routine for many companies. Using data from the research company Mediaskop, the behavior of the television audience was analyzed on the basis of one of the most important indicators in media planning – the average audience share of the TV channels. In the course of the research, it was found that for predicting the share of viewers of TV channels, methods that allow combining autoregressive models and a moving average were the most suitable. The use of such methods is particularly relevant for the description and prediction of processes in which homogeneous fluctuations around the mean value occur. It is shown that building forecasts based on the maximum number of factors and features, such as competition and seasonality, can be implemented using adaptive forecasting methods. With this approach, the use of methods for identifying model parameters to compile the most accurate forecast is of particular interest. The forecast model based on real data is presented. The adequacy of the model was tested by constructing an autocorrelation function for residuals. Autocorrelations do not go beyond the permissible intervals, which indicate the independence of the residuals. It was also verified that the residues are distributed according to the normal law. The fulfillment of these two criteria confirms the adequacy of the model obtained.

Keywords: media content, media planning, audience, forecasting, average daily share of the TV channel, rating, TV program

Мир вступил в новый, информационный век. Сегодня появляется все больше новых производителей медиаконтента и все больше новых, в том числе технологически инновационных, каналов распространения этого контента. Однако с появлением новых СМИ сам объем медиапотребления не меняется. На практике в основном происходит лишь перераспределение времени и расходов на медиапотребление между растущим числом производителей контента и телезрителями.

В результате происходит дробление аудитории СМИ. Ее становится все сложнее охватить. Именно для решения данной проблемы и возникло медиапланирование как отрасль, призванная находить более эффективные пути донесения информации до аудитории различных медиа [1–3]. Для этого надо использовать различные исследовательские подходы и, в частности, прогнозирование [4].

Цель исследования: анализ основных подходов и разработка методов совершенствования методического инструментария

прогнозирования в медиапланировании рейтинга телевизионного контента, выявление возможностей применения адаптивных методов прогнозирования с использованием методов идентификации параметров модели для составления максимально точного прогноза.

Материалы и методы исследования

Исследование просмотра телевизионного и видеоконтента – задача, важность которой интенсивно растет в последнее время. Исследованием объема и характеристик аудитории СМИ занимается Mediascore – ведущая исследовательская компания на российском рынке в сфере медиаисследований и мониторинга рекламы и СМИ. Mediascore исследует просмотр ТВ-контента на всех платформах – от классического телевизора до digital среды. Одним из преимуществ технологий, применяемых Mediascore в телевизионных исследованиях, является возможность получения социально-демографических характеристик зрителя, для чего каждый участник исследования должен регистрировать вход и выход в помещение с телевизором членов семьи или гостей. Это дает возможность с высокой точностью определять аудиторию конкретных телеканалов и эфирных событий. Полученные данные помогают вещателям показывать наиболее востребованный зрителями контент, а рекламному рынку – планировать, размещать и оценивать рекламные кампании.

На сайте компании [5] представлены основные показатели ведущих телеканалов за последние годы:

- рейтинг телепрограмм – среднее количество человек, смотревших телепрограмму, выраженное в процентах от общей численности исследуемой аудитории;
- доля телепрограммы – среднее количество человек, смотревших телепрограмму, выраженное в процентах от общего количества телезрителей (тех, кто смотрел любую программу, включая оцениваемую программу) в данный момент времени;
- среднесуточная доля телеканала – количество человек, смотревших телеканал в среднем в сутки, выраженное в % от всех телезрителей.

В компаниях-медиацеллерах и рекламных агентствах обычно выполняют прогноз аудитории телевидения по следующей схеме:

1. Считают среднюю долю зрительской аудитории для каждого месяца рассматриваемого периода (обычно год);

2. Далее определяют сезонный коэффициент K для каждого месяца:

$$K = \frac{Share_i}{Share_{баз.}}$$

где $Share_i$ – это доля телезрителей за месяц ($i = 1$: январь, $i = 2$: февраль и т.д.), а $Share_{баз.}$ – это доля зри-

телей за месяц, который аналитик принимает за базовый. То есть коэффициент сезонности для каждого месяца определяется как отношение размера аудитории в текущий месяц к ее размеру в каком-либо фиксированном месяце.

3. Чтобы построить прогноз, берут значение доли телезрителей за последний известный месяц и умножают на полученные в п. 2 коэффициенты сезонности.

Продемонстрируем это на примере. Пусть нам известны гипотетические помесечные значения доли зрителей канала X в течение 2018 г. и января 2019 г. (табл. 1). Чтобы спланировать бюджет, необходимо спрогнозировать аудиторию канала на февраль – апрель 2019 г.

Принимаем за базовый месяц январь 2018 г. Значит, коэффициенты сезонности для февраля, марта и апреля будут равны соответственно: $K_{02} = 13/15$, $K_{03} = 12/15$, $K_{04} = 14/15$.

Крайний известный месяц – январь 2019 г. Значит, прогноз на последующие месяцы строится относительно этого месяца.

$$Share_{02.19} = K_{02} \cdot Share_{01.19} = 13/15 \cdot 16 = 13,86,$$

$$Share_{03.19} = K_{03} \cdot Share_{01.19} = 12/15 \cdot 16 = 12,8,$$

$$Share_{04.19} = K_{04} \cdot Share_{01.19} = 14/15 \cdot 16 = 14,9.$$

Считается, что такой прогноз сохраняет наличие сезонных изменений, учитывает отношение аудитории каждого месяца к остальным, а на значение прогнозируемого месяца влияет последний фактический. Как правило, прогнозы строятся на базе данных за 1–2 года, предшествующих прогнозируемому. Считается, что более ранние периоды включают в себя некоторые другие факторы, которые влияли на аудиторию (был другой набор каналов в городе, немного другие предпочтения у людей, другое качество программ и фильмов и т.д.). Тем самым прогноз на базе небольшого набора данных дает более грубую оценку. В итоге это приводит к тому, что прогнозируемое значение, как правило, отличается от зафиксированного позже реального результата.

Предложим другой, более точный метод предварительной оценки и последующего прогноза средней доли телезрителей на некотором канале Y на первые несколько месяцев 2017 г. Выполним анализ по данным Mediascore за 2015–2016 гг. Регион исследования – «Москва», исследуемая аудитория – «население в возрасте от 4 лет и старше». Воспользуемся методом, который объединяет модель авторегрессии и скользящего среднего, которые оказываются особенно эффективными для описания и прогнозирования процессов, обнаруживающих однородные колебания вокруг среднего значения.

Таблица 1

Доля зрителей ($Share_i$) канала X (помесечно январь 2018 г. – январь 2019 г.)

Месяц	01.18	02.18	03.18	04.18	05.18	06.18	07.18	08.18
$Share_i$	15	13	12	14	15	11	10	11
Месяц	09.18	10.18	11.18	12.18	01.19	02.19	03.19	04.19
$Share_i$	12	13	14	14	16	?	?	?

В авторегрессионной модели наиболее сильные связи наблюдаются у соседних состояний и быстро уменьшаются с увеличением расстояния между ними (влияние предыдущего состояния процесса на будущее). Математически это свойство можно выразить уравнением

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t,$$

где y_t – значение y в момент времени t ; ϕ_i – коэффициенты уравнения ($i = 1, 2, \dots, p$); p – порядок авторегрессии; ε_t – случайная величина.

В модели скользящего среднего в отличие от предыдущего способа предполагается, что каждый элемент ряда подвержен суммарному воздействию случайных предыдущих ошибок ε_i :

$$y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

где y_t – значение y в момент времени t ; θ_i – коэффициенты уравнения ($i = 1, 2, \dots, q$); q – порядок модели скользящего среднего; ε_t – случайная величина.

В программе анализа данных Statistica модель, являющаяся комбинацией двух вышеперечисленных, имеет название ARIMA. Данная модель подходит только для стационарных рядов, т.е. для тех, у которых среднее и дисперсия примерно постоянны во времени. Поэтому одним из этапов построения модели (кроме восстановления пропущенных данных, определения порядка модели p и q , оценки параметров модели, проверки адекватности и прогноза) является преобразование ряда к стационарному виду. Возможность реализации этих задач заложена в пакете прикладных программ Statistica [6].

Из рис. 1 видно, что в поведении аудитории наблюдается некоторая сезонная закономерность. В летний период наблюдается сильное уменьшение значения доли зрителей.

Будем строить прогноз исходя из сезонной последовательной зависимости, т.е. учитывать, что

на значение доли каждый месяц влияет ее значение в предыдущий месяц. На следующем рисунке представлена автокорреляционная функция, которая показывает, что автокорреляция достигает пиков при лаге, равном 12, что очевидно из соображений сезонности. Но из рис. 2 видно, что корректные результаты также достигаются при сезонном лаге, равном 1, 5, 6 и 7.

Проведем процесс оценивания параметров модели (рис. 3).

На рис. 3 красным цветом выделены наиболее значимые коэффициенты (как правило, это те параметры, которые более чем в два раза превосходят свои стандартные ошибки). Остаточная сумма квадратов и средний квадрат остатков малы, что является подтверждением того, что полученная оценочная модель достаточно близка к фактическим данным.

Прогноз, который выдает программа, строится на период, равный 12 месяцам (рис. 4). Приведем сводку прогнозируемых значений (рис. 5):

Сравнение полученных результатов с фактическими данными за 2017 г. (табл. 2) демонстрирует, что в первые 4 месяца прогноза совпадение происходит с минимальной погрешностью.

Таблица 2

Фактические данные и ошибки прогноза средней доли телезрителей на канале Y за 2017 г.

	Фактические данные (значения Share)	Относительные ошибки прогноза
Март 2017	11,7	0,2%
Апрель 2017	11,4	1,7%
Май 2017	10,9	10%
Июнь 2017	10,3	2,9%
Июль 2017	10,1	6,9%

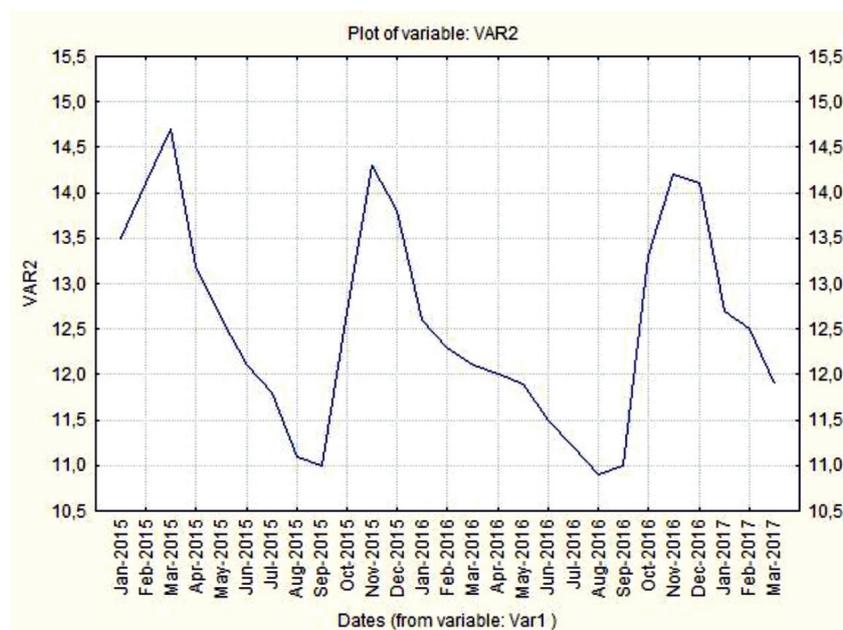


Рис. 1. Значения доли зрителей по месяцам исследуемого периода

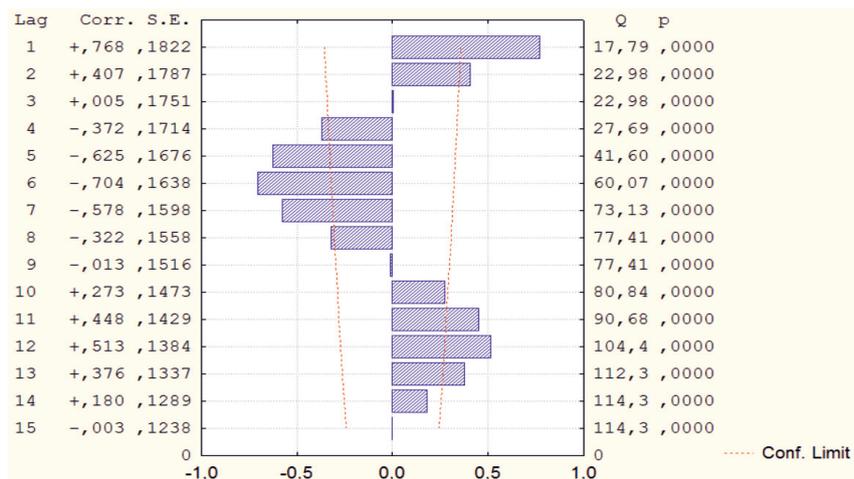


Рис. 2. График автокорреляционной функции

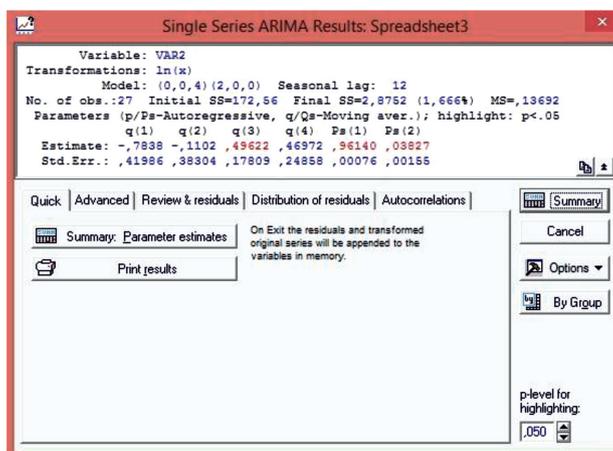


Рис. 3. Результаты расчета коэффициентов и параметров модели

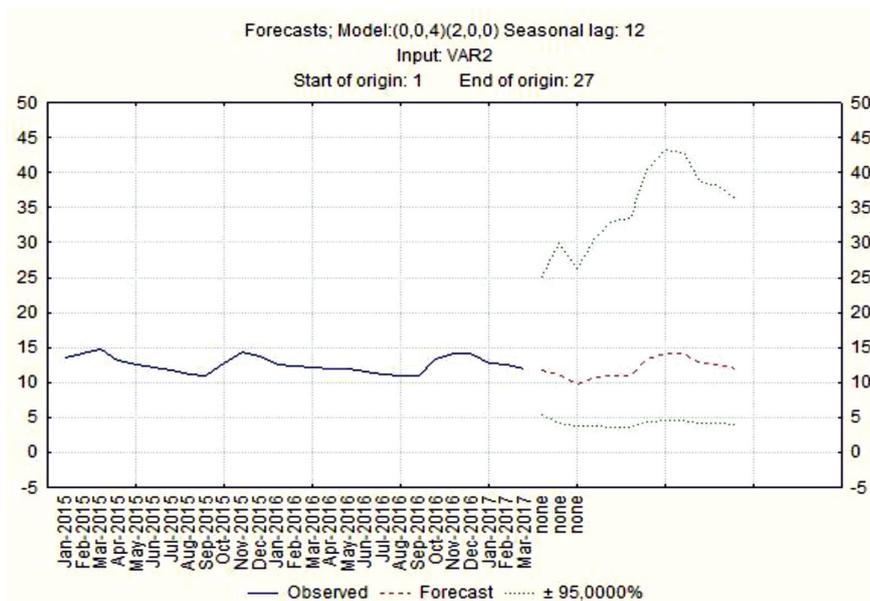


Рис. 4. График прогноза

Forecasts; Model:(0,0,4)(2,0,0) Seasonal lag: 12 (Spreadsheet3)			
Input: VAR2			
Start of origin: 1 End of origin: 27			
CaseNo.	Forecast	Lower 95,0000%	Upper 95,0000%
28	11,67231	5,407115	25,19695
29	11,21054	4,217145	29,80128
30	9,86116	3,695937	26,31067
31	10,69016	3,729534	30,64176
32	10,89909	3,579892	33,18259
33	10,99139	3,610211	33,46362
34	13,26531	4,357098	40,38664
35	14,19152	4,661318	43,20650
36	14,07624	4,623452	42,85552
37	12,68563	4,166696	38,62178
38	12,48199	4,099810	38,00179
39	11,89797	3,907984	36,22373

Рис. 5. Прогнозные значения доли зрителей

Результаты исследования и их обсуждение

Адекватность модели была проверена путем построения автокорреляционной функции для остатков. Автокорреляции не выходят за пределы допустимых интервалов, что говорит о независимости остатков. Также было проверено, что остатки распределены по нормальному закону. Выполнение этих двух критериев и подтверждает адекватность полученной модели.

Задача прогнозирования временных рядов имеет высокую актуальность для многих предметных областей и является неотъемлемой частью повседневной работы многих компаний. В настоящее время разработано множество моделей для решения задач прогнозирования временных рядов, среди которых наибольшую применимость имеют авторегрессионные и нейросетевые модели. При этом существенным недостатком авторегрессионных моделей является большое число свободных параметров, требующих идентификации; недостатками нейросетевых моделей является непрозрачность моделирования и сложность обучения сети.

Наиболее перспективным направлением развития моделей прогнозирования с целью повышения точности является создание комбинированных моделей, выполняющих на первом этапе кластеризацию, а затем прогнозирование временного ряда внутри установленного кластера. В ходе исследований было выявлено, что для прогнозирования доли зрителей телеканалов наиболее подходящими оказались методы, которые позволяют объединить модели авторегрессии и скользящего среднего [7, 8]. Применение

таких методов является особенно актуальным для описания и прогнозирования процессов, в которых проявляются однородные колебания вокруг среднего значения.

Комбинирование этих методов также эффективно в случае, когда рассматривается ситуация, в которой за промежуток времени, предшествующий прогнозируемому периоду, наблюдается аномальное отклонение данных от обычного поведения (среднего значения для соответствующего месяца). Например, в связи с каким-либо происшествием закрывается показ рекламы в эфире, что влечет резкий нехарактерный спад рейтинга в данном месяце. В таком случае в алгоритме построения прогноза добавляется пункт сглаживания отклонения, который должен производиться без нарушения общих тенденций поведения аудитории на всем периоде.

Заключение

В составлении прогнозов в сфере рекламы есть еще много моментов, требующих глубокого математического анализа [9]. Построение прогнозов с учетом максимального количества факторов и особенностей, таких как конкуренция и сезонность, может быть реализовано с использованием адаптивных методов прогнозирования. При таком подходе особый интерес представляет использование методов идентификации параметров модели для составления максимально точного прогноза [10]. Данная составляющая часть процесса моделирования нередко игнорируется, и производится оценка параметров стандартными приемами, что зачастую снижает точность прогноза. Применение комбинированных

моделей является направлением, которое при корректном подходе позволяет повысить точность прогнозирования. Главным недостатком комбинированных моделей является сложность и ресурсоемкость их разработки: нужно разработать модели таким образом, чтобы компенсировать недостатки каждой из них, не потеряв достоинств. Поэтому требуется продолжение исследований в указанном направлении.

Список литературы

1. Бузин В.Н., Бузина Т.С. Медиапланирование. Теория и практика: учеб. пособие для студентов вузов, обучающихся по специальностям «Реклама», «Маркетинг», «Психология», «Социология», «Журналистика». М.: ЮНИТИ-ДАНА, 2016. 495 с.
2. Назайкин А.Н. Современное медиапланирование: традиционные СМИ, а также реклама в интернете (медийная и контекстная): учеб. пособие. М.: СОЛОН-Пресс, 2016. 447 с.
3. Николаева М.А. Теория и практика медиапланирования: методический аспект // Педагогическое образование в России. 2015. № 10. С. 71–78.
4. Афанасьев В.Н., Юзбашев М.М. Анализ временных рядов и прогнозирование: учебник. Изд. 2-е, перераб. и доп. М.: Финансы и статистика, 2010. 320 с.
5. Данные по аудитории СМИ. TV Index [Электронный ресурс] // Медиаскоп. URL: <https://mediascope.net/data/> (дата обращения: 15.04.2019).
6. Куприенко Н.В., Пономарева О.А., Тихонов Д.В. Статистика. Временные ряды. Анализ тенденций и прогнозирование: учеб. пособие. СПб.: Изд-во Политехн. ун-та, 2015. 123 с.
7. Полежаев В.Д., Юсупова К.О. Моделирование прогнозных зависимостей на основе анализа данных медиаисследований // Математическое и компьютерное моделирование: сборник материалов VI Международной научной конференции, посвященной памяти Б.А. Рогозина (Омск, 23 ноября 2018 г.). Омск: Изд-во Ом. гос. ун-та, 2018. С. 127–129.
8. Юсупова К.О. Исследование наличия конкуренции между российскими телевизионными каналами и интернет-сайтами на основе количественного анализа их пользовательской аудитории // Международный студенческий вестник. 2018. № 3–1. С. 140–144.
9. Шматов Г.А. Математическая теория медиапланирования: монография / Российская акад. наук, Уральское отделение, Ин-т экономики. Екатеринбург: Ин-т экономики УрО РАН, 2009. 329 с.
10. Полежаев В.Д., Полежаева Л.Н. Обоснование выбора вида функций, применяемых для аппроксимации данных в экономико-математическом моделировании // Информационные технологии и математические методы в экономике и управлении (ИТиММ-2016) 26–27 мая 2016 г.: сборник научных статей. М.: ФГБОУ ВО «РЭУ им. Г.В. Плеханова», 2016. С. 18–25.