

Черкунова М. Е.,
Санкт-Петербургский государственный университет
st076309@student.spbu.ru

Cherkunova M. E.,
St Petersburg University
st076309@student.spbu.ru

ПРИМЕНЕНИЕ МЕТОДОВ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ ПРИ ПРОГНОЗИРОВАНИИ СПРОСА НА СЕЗОННЫЕ ТОВАРЫ

Аннотация. Статья посвящена исследованию возможностей применения методов анализа временных рядов для прогнозирования спроса на сезонные товары. Для анализа были использованы данные о продажах средства от насекомых за 10 лет. Проведено исследование с использованием модели SARIMA с учётом экзогенной переменной. Исследование показало, что модель способна эффективно учитывать сезонные колебания и давать адекватные прогнозы. Определены условия масштабируемости модели для прогнозирования спроса на другие сезонные товары. Результаты исследования имеют практическую ценность, позволяя оптимизировать управление запасами и повысить эффективность розничной сети.

Ключевые слова: анализ временных рядов, прогнозирование спроса, сезонность, модель SARIMA, экзогенная переменная, управление запасами.

APPLICATION OF TIME SERIES ANALYSIS METHODS IN FORECASTING DEMAND FOR SEASONAL PRODUCTS

Abstract. The article is devoted to the study of the possibilities of using time series analysis methods to forecast demand for seasonal goods. The analysis used data on insect repellents sales for 10 years. The study was conducted using the SARIMA model taking into account the exogenous variable. The study showed that the model is able to effectively take into account seasonal fluctuations and provide adequate forecasts. The scalability conditions of the model for forecasting demand for other seasonal goods are determined. The results of the study have practical value, allowing to optimize inventory management and improve the efficiency of the retail network.

Keywords: time series analysis, demand forecasting, seasonality, SARIMA model, exogenous variable, inventory management.

Прогнозирование спроса на сезонные продукты – ключевой фактор успешного розничного бизнеса. Ошибки в прогнозах могут привести к нехватке товара на складе и упущенной выгоде или к избыточным запасам, что влечёт дополнительные затраты на хранение и утилизацию.

В данной статье исследуются возможности применения методов анализа временных рядов для прогнозирования спроса на сезонный товар. В качестве объекта исследования выбран товар из группы средств от насекомых, продаваемых в розничной сети.

Для проведения анализа использованы два временных ряда:

- Ряд 1: количество проданных единиц товара за период с 31.01.2014 по 30.06.2024 (ежемесячные данные) (рис. 1). Этот ряд является целевым, для него необходимо построить прогноз.
- Ряд 2: количество магазинов в городе за тот же период. Этот ряд рассматривается в качестве экзогенной переменной в модели, так как количество магазинов может влиять на продажи товара.

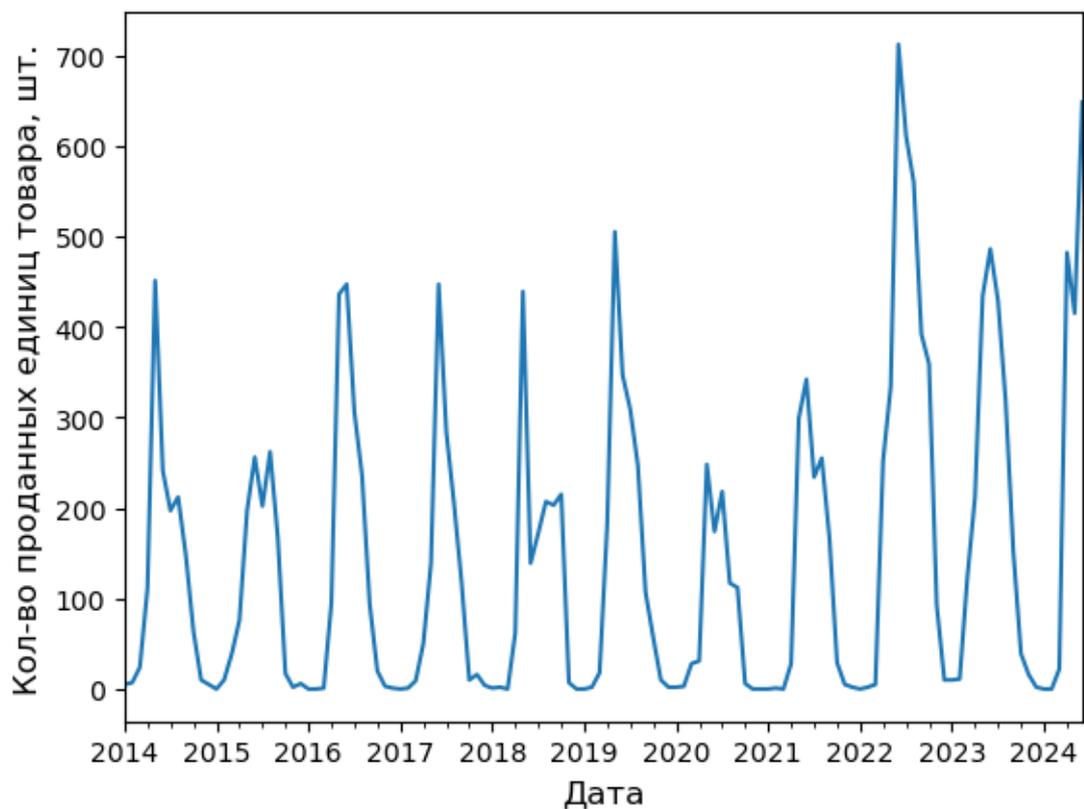


Рис. 1. График динамики спроса на товар

Исходные данные были разделены на обучающую и тестовую выборки. В качестве тестовой выборки взяты данные за последний год (с 31.07.2023 по 30.06.2024), что позволило оценить точность полученных прогнозов.

Цель работы – исследовать применимость моделей анализа временных рядов для прогнозирования спроса на товары с сезонностью, построить прогноз на основе выбранной модели и проанализировать возможность масштабирования данного процесса для других сезонных товаров. Результаты исследования позволят оптимизировать процессы управления запасами и повысить эффективность розничной сети.

Предварительный анализ исходных данных показал ярко выраженную сезонность. Была проведена проверка на выбросы и пропуски, данные были приведены к необходимому формату. Далее решено было применить модель SARIMAX (Seasonal Autoregressive Integrated Moving Average with eXogenous Regressors) для прогнозирования. Эта модель объединяет в себе методы авторегрессии, интегрированного скользящего среднего, сезонности и возможности добавления экзогенных переменных, что позволяет прогнозировать будущие значения ряда на основе его предыдущих значений с учетом сезонных паттернов [1]. Таким образом, модель SARIMAX состоит из следующих частей:

- AR (Autoregressive): используются прошлые значения самого временного ряда для прогнозирования.
- I (Integrated): применяется дифференцирование для устранения нестационарности в данных.
- MA (Moving Average): используются прошлые ошибки прогнозирования для улучшения точности прогнозов.
- Seasonal components: добавляются сезонные компоненты.

- **exogenous Regressors**: позволяет включать внешние факторы, которые могут влиять на временной ряд.

Перед непосредственно оценкой модели был проведен тест Дики-Фуллера для определения порядка интегрируемости ряда продаж. Этот метод представляет количественную оценку стационарности временного ряда. Основная гипотеза теста предполагает, что ряд не стационарен [2]. Для исходного ряда данный тест не отверг нулевую гипотезу, а вот первая разность оказалась стационарной, следовательно, ряд первого порядка интегрируемости ($I=1$).

На следующем этапе анализа были построены графики ACF (Autocorrelation Function) (рис. 2) и PACF (Partial Autocorrelation Function) (рис. 3). Это инструменты анализа временных рядов, которые помогают нам понять структуру зависимости данных во времени [3]. Данный вид графиков подтверждает наличие сезонности в данных.

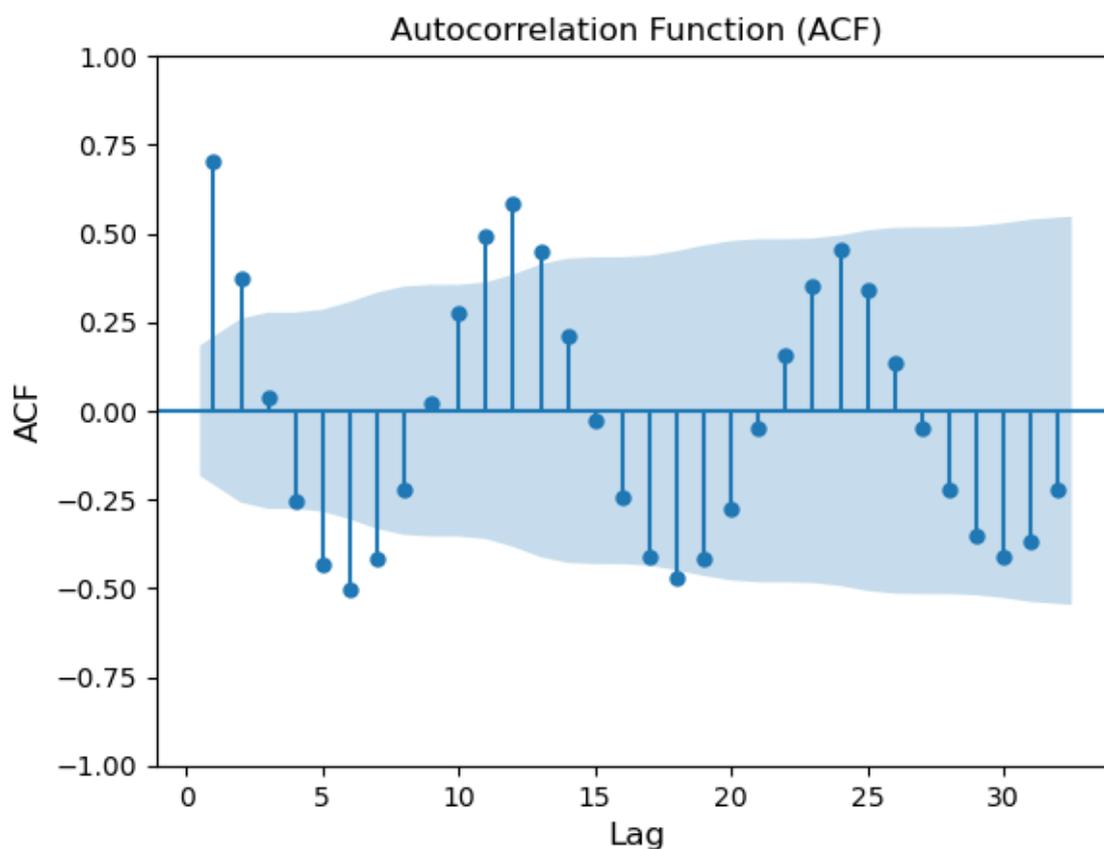


Рис. 2. График ACF ряда спроса на товар

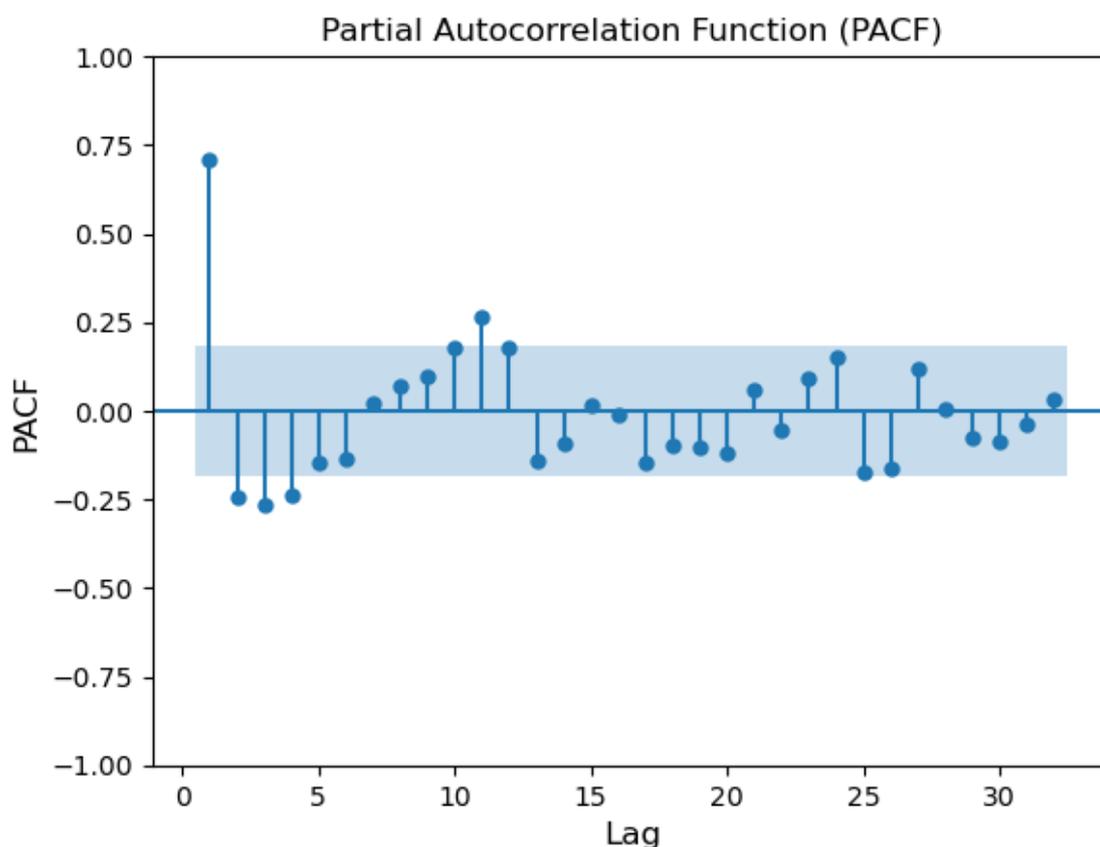


Рис. 3. График PACF ряда спроса на товар

С помощью библиотеки PMDARIMA (Auto-ARIMA) [4] (библиотека на Python, которая автоматизирует процесс выбора оптимальных параметров для модели ARIMA) были найдены параметры модели. Наилучшей моделью оказалась SARIMA (0,1,1)(4,1,0,12). Далее для оценки влияния константы и экзогенной переменной (количество магазинов) на продажи, были рассмотрены четыре варианта модели: модель с константой (в этой модели использовалась только константа, без учета экзогенной переменной); модель с экзогенной переменной (без константы); модель с константой и экзогенной переменной; модель без константы и без экзогенной переменной. Сравнительный анализ информационных критериев показал, что наилучшие результаты (минимальные значения) были получены для модели с экзогенной переменной, но без константы. Итоговая модель была оценена на обучающей выборке (рис. 4). Коэффициенты модели статистически значимы. Согласно тесту Льюнг-Бокса на автокорреляцию остатков, остатки модели случайны, что подтверждает адекватность модели. К сожалению, результат теста Харке-Бера говорит о том, что остатки не соответствуют нормальному распределению. Это не является критичным фактором, так как в реальных данных редко встречаются идеально нормальные остатки. Важнее, чтобы модель давала приближенные к реальности прогнозы. Модели, обеспечивающие нормальное распределение остатков, могут быть сложнее в реализации и требовать большего объема данных, не гарантируя при этом существенное улучшение точности прогнозирования. Гетероскедастичность в оцененной модели отсутствует, что также хорошо.

	coef	std err	z	P> z 	[0.025	0.975]
am_shops	16.9178	8.396	2.015	0.044	0.461	33.374
ma.L1	-0.3292	0.082	-4.017	0.000	-0.490	-0.169
ar.S.L12	-0.7856	0.079	-9.923	0.000	-0.941	-0.630
ar.S.L24	-0.6345	0.139	-4.574	0.000	-0.906	-0.363
ar.S.L36	-0.6425	0.112	-5.750	0.000	-0.862	-0.424
ar.S.L48	-0.5681	0.077	-7.357	0.000	-0.719	-0.417
sigma2	5647.5695	775.607	7.281	0.000	4127.409	7167.730
Ljung-Box (L1) (Q):	0.01	Jarque-Bera (JB):	14.45			
Prob(Q):	0.91	Prob(JB):	0.00			
Heteroskedasticity (H):	1.57	Skew:	-0.11			
Prob(H) (two-sided):	0.19	Kurtosis:	4.84			

Рис. 4. Сводка оценённой модели для прогнозирования спроса на товар

С помощью оцененной модели были получены прогнозные значения для показателя продаж и доверительные интервалы при 5% уровне значимости для периода из тестовой выборки (рис. 5). При этом, в случаях, когда прогнозное значение или границы доверительного интервала оказывались отрицательными, они заменялись нулевыми значениями.

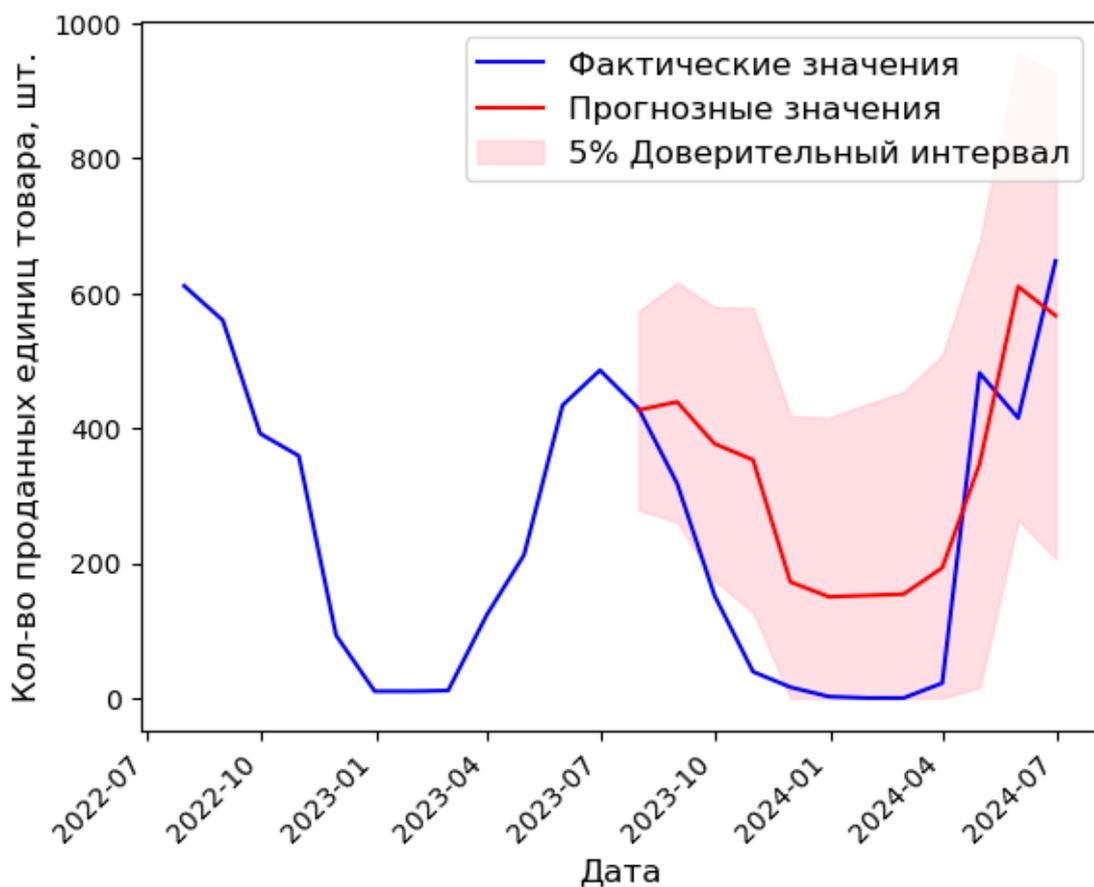


Рис. 5. График фактических и прогнозных значений продаж за последние 2 года, доверительный интервал

Визуально на представленном графике наблюдается, что фактические данные за последний год в большинстве случаев попадают в доверительный интервал за исключением некоторых периодов. Это не говорит о том, что модель плоха, так как интервал отражает вероятностный характер прогноза. Для практического использования модели рекомендуется регулярно обновлять ее с использованием новых фактических данных. Это позволит улучшить точность прогнозов, так как модель будет обучаться на более актуальных данных.

В заключение данной работы хотелось бы отметить, что проведенное исследование демонстрирует эффективность применения методов анализа временных рядов для прогнозирования спроса на сезонные товары. Использование модели SARIMA (0,1,1)(4,1,0,12) с экзогенной переменной (количество магазинов) позволило получить адекватные прогнозы продаж средства от насекомых с учетом сезонных колебаний. Модель продемонстрировала свою способность точно отражать исторические данные и давать прогнозы с приемлемым уровнем точности.

Результаты исследования указывают на возможность масштабирования модели для прогнозирования спроса на другие сезонные товары. Для этого необходимо адаптировать модель к конкретным особенностям товара, учитывать внешние факторы, влияющие на спрос. Для повышения точности модели могут быть рассмотрены влияния и других экзогенных переменных, например данных о погоде и экономических показателей.

Применение данной модели в практической деятельности может способствовать оптимизации управления запасами, снижению рисков дефицита или переизбытка товара, а

также увеличению прибыли розничной сети. Кроме того, модели SARIMAX имеют потенциал для применения и в других сферах бизнеса, например, оценке эффективности маркетинговых мероприятий с помощью анализа динамики продаж до и после их проведения. Это открывает возможности для исследования других применений моделей SARIMAX для ведения бизнеса.

Литература

1. Михайлов А. Анализ временных рядов [Эл. ресурс] / А. Михайлов. URL: <https://habr.com/ru/companies/otus/articles/732080/> (дата обращения: 20.09.2024).
2. Богомаз М. Прогнозирование временных рядов с помощью ARIMA в Python 3 [Эл. ресурс] / М. Богомаз. URL: <https://timeweb.cloud/tutorials/python/prognozirovanie-vremennyh-ryadov-python-3> (дата обращения: 23.09.2024).
3. Волков Н. Модели вида ARIMA [Эл. ресурс] / Н. Волков. URL: <https://education.yandex.ru/handbook/ml/article/modeli-vida-arima> (дата обращения: 23.09.2024).
4. Smith T. pmdarima.arima.auto_arima [Эл. ресурс]. / Т. Smith. URL: https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html (дата обращения: 23.09.2024).