

*Итаева Надежда Михайловна
Бакалавр 3 курса института экономики и управления Волгоградского
государственного университета, г. Волгоград*

*Черкунова Мария Евгеньевна,
Магистр 1 курса Высшей школы менеджмента Санкт-Петербургского
государственного университета, г. Санкт-Петербург*

*Черкунова Наталья Анатольевна,
старший преподаватель кафедры прикладной информатики и
математических методов в экономике Волгоградского государственного
университета, г. Волгоград*

*Itaeva Nadezhda Mikhailovna
3rd year Bachelor's degree, Institute of Economics and Management, Volgograd
State University, Volgograd*

*Cherkunova Mariia Evgeniena,
1st year Master's degree, Graduate School of Management, St. Petersburg
University, St. Petersburg*

*Cherkunova Natalia Anatolyevna,
Senior Lecturer, Department of Applied Informatics and Mathematical Methods in
Economics, Volgograd State University, Volgograd*

**Анализ и прогнозирование динамики поступления налога на доходы
физических лиц по Волгоградской области**
**Analysis and forecasting of the dynamics of personal income tax receipts in the
Volgograd region**

Аннотация: Статья посвящена исследованию динамики поступлений налога на доходы физических лиц (НДФЛ) в Волгоградской области за период 2018-2023 гг. Актуальность исследования обусловлена необходимостью повышения эффективности налоговой политики и обеспечения стабильного финансового развития региона.

В работе использовались ежемесячные данные, скорректированные на инфляцию, что позволило получить более точную картину реальной динамики поступлений НДФЛ. Для прогнозирования будущих значений НДФЛ была выбрана модель SARIMA, которая учитывает сезонные и автокорреляционные свойства временного ряда.

Статья описывает анализ стационарности временного ряда и определение порядка интегрируемости. Результаты исследования позволяют оценить влияние сезонных факторов на динамику поступлений НДФЛ и разработать более точные прогнозы для планирования бюджета и оптимизации налоговой политики региона.

Annotation: The article is devoted to the study of the dynamics of personal income tax (PIT) revenues in the Volgograd region for the period 2018-2023. The relevance of the study is due to the need to improve the efficiency of tax policy and ensure stable financial development of the region.

The work used monthly data adjusted for inflation, which made it possible to obtain a more accurate picture of the real dynamics of PIT revenues. To forecast future PIT values, the SARIMA model was chosen, which takes into account the seasonal and autocorrelation properties of the time series.

The article describes the analysis of the stationarity of the time series and the determination of the order of integrability. The results of the study make it possible to assess the influence of seasonal factors on the dynamics of PIT revenues and develop more accurate forecasts for budget planning and optimization of the tax policy of the region.

Ключевые слова: Налог на доходы физических лиц (НДФЛ), Волгоградская область, динамика поступлений, прогнозирование, модель SARIMA, сезонность, временной ряд, стационарность, дефляция

Key words: Personal income tax (PIT), Volgograd region, revenue dynamics, forecasting, SARIMA model, seasonality, time series, stationarity, deflation

Формирование доходной части бюджета является ключевой задачей для любого региона, обеспечивая финансирование социальных программ, инфраструктуры и других государственных функций. Налог на доходы физических лиц (НДФЛ) является одним из главных источников доходов бюджета России. Изучение динамики НДФЛ позволяет оценить эффективность налоговой политики, выявить тенденции в уровне доходов населения.

Цель исследования – проанализировать динамику НДФЛ в Волгоградской области и спрогнозировать будущие поступления. Результаты исследования могут быть использованы для оптимизации налоговой политики региона и обеспечения его устойчивого финансового развития.

В данной статье исследуется динамика НДФЛ в Волгоградской области за период 2018-2023гг. [1]. Для анализа были использованы ежемесячные данные, скорректированные на инфляцию. Данная корректировка необходима, поскольку инфляция искажает реальную картину динамики. Номинальный рост поступлений НДФЛ может быть обусловлен не только увеличением количества налогоплательщиков или ростом их доходов, но и повышением цен. Была проведена дефляция с использованием индекса потребительских цен (ИПЦ) [2]. Дефляция - это процесс пересчета номинальных значений в реальные, т.е. с учетом изменения цен. Формула дефляции: Реальный показатель = Номинальный показатель / ИПЦ *100.

Для исключения выбросов и повышения точности данных было применено усреднение.

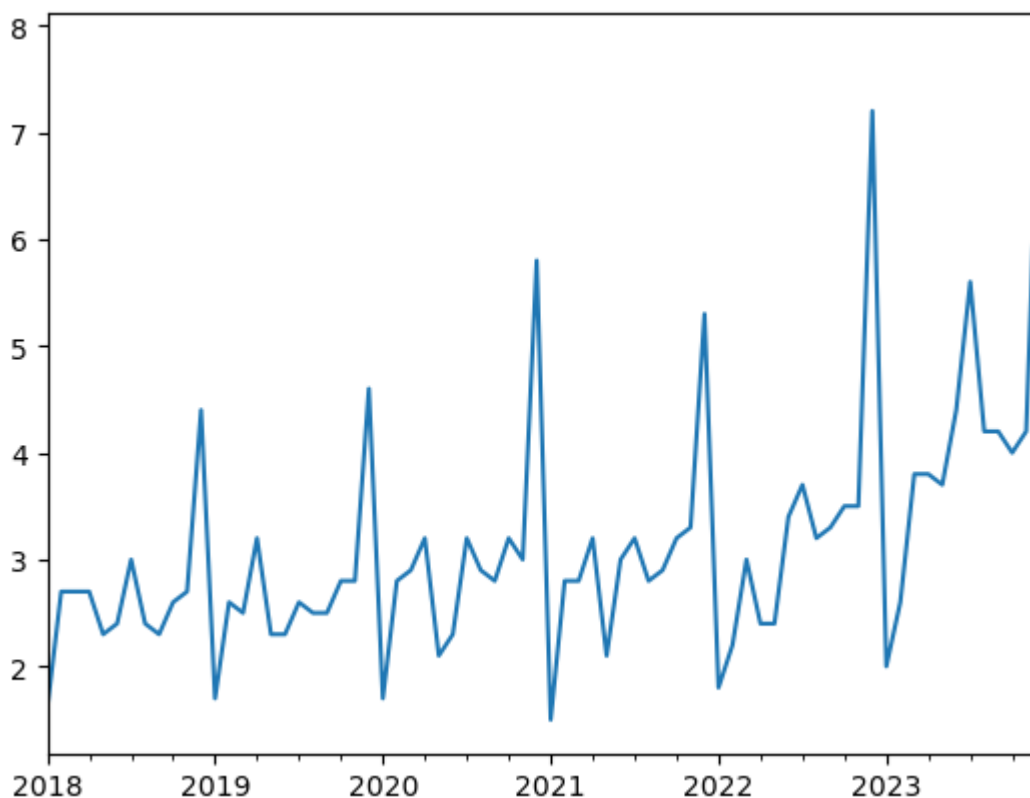


Рис.1. НДФЛ скорректированный на инфляцию (млрд. руб).

В исследовании решено применить модель SARIMA, позволяющую прогнозировать будущие значения НДФЛ на основе анализа прошлых данных и выявления сезонных и трендовых составляющих. Модели SARIMA (Seasonal Autoregressive Integrated Moving Average) - это мощный инструмент для прогнозирования временных рядов, имеющих сезонные и автокорреляционные свойства [3]. SARIMA модели включают в себя следующие компоненты:

- AR (Autoregressive): Модель использует предыдущие значения временного ряда для прогнозирования будущих.
- MA (Moving Average): Модель использует прошлые ошибки прогнозирования для прогнозирования будущих значений.
- I (Integrated): Модель применяется к дифференцированным данным, что позволяет устранить тренд и сделать данные стационарными.
- Сезонные компоненты (SAR, SMA, SI): Модель учитывает сезонные изменения в данных, используя аналогичные AR, MA и I компоненты, но на сезонном уровне.

Перед оценкой модели и поиском параметров был проведен тест Дики-Фуллера на стационарность и определение порядка интегрируемости. Выявлено, что данный ряд имеет второй порядок интегрируемости. Для определения периода колебаний было рассмотрено сезонное разложение.

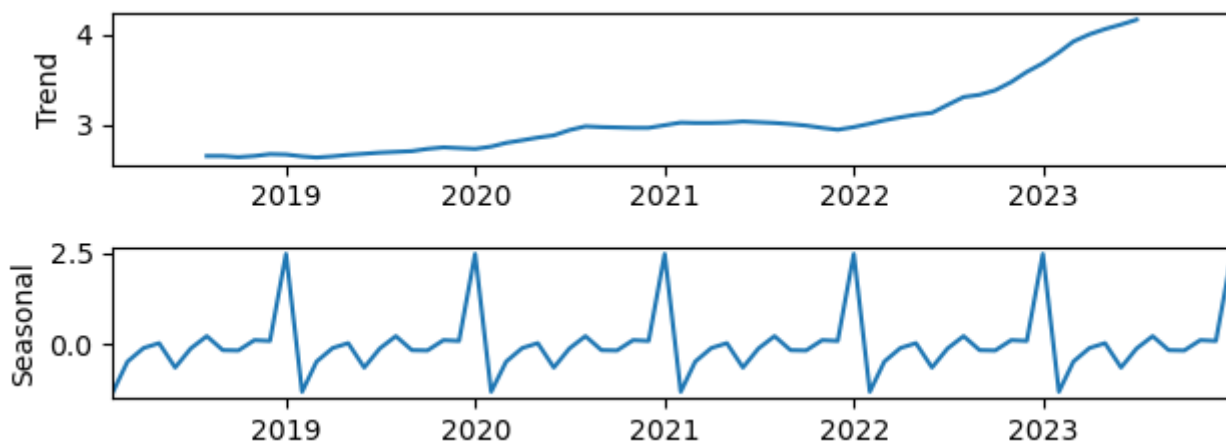


Рис. 2. Сезонное разложение.

Визуально можно сделать вывод, что ряд имеет линейный тренд, ежегодную сезонность, порядок сезонного интегрирования равен единице.

С помощью библиотеки PMDARIMA (Auto-ARIMA) [4] (библиотека на Python, которая автоматизирует процесс выбора оптимальных параметров для модели ARIMA, широко используемого метода прогнозирования временных рядов) были найдены параметры модели для прогнозирования. В качестве критерия использовался информационный критерий AIC. Наилучшей моделью оказалась SARIMA (4,2,0)(0,1,0,12). Это означает, что модель, которая лучше всего описывает динамику данных, использует 4 предыдущих значения временного ряда, дифференцирует данные два раза, не использует прошлые ошибки прогнозирования, дифференцирует данные на сезонном уровне один раз и учитывает сезонность с периодом 12. Далее были оценены коэффициенты данной модели. Можно отметить, что все они статистически значимы.

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	-1.4169	0.110	-12.859	0.000	-1.633	-1.201
ar.L2	-1.2570	0.214	-5.869	0.000	-1.677	-0.837
ar.L3	-0.9822	0.204	-4.815	0.000	-1.382	-0.582
ar.L4	-0.4244	0.114	-3.725	0.000	-0.648	-0.201
sigma2	0.2509	0.040	6.267	0.000	0.172	0.329

Рис. 3. Коэффициенты оцененной модели.

С использованием оцененной модели были получены прогнозные значения и доверительный интервал при 5% уровне значимости. При этом, если прогноз или доверительный интервал принимали отрицательные значения, то они заменялись нулевыми.

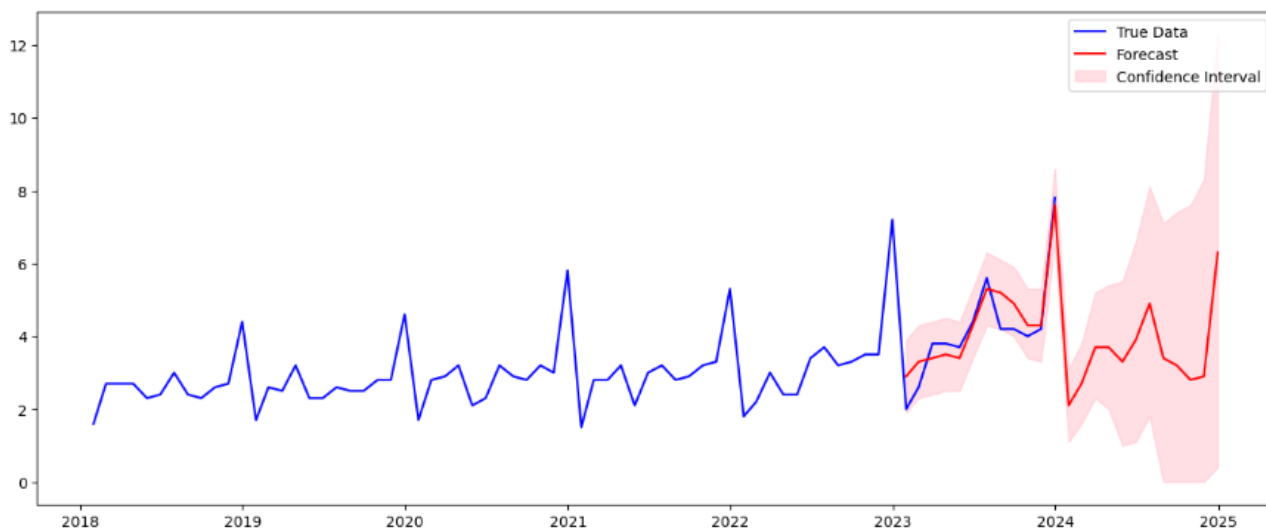


Рис. 3. Фактические и прогнозные значения динамики поступления НДФЛ, доверительный интервал (млдр. руб).

На момент окончания исследования стали доступны данные по поступлениям НДФЛ за первые 5 месяцев 2024 года, и можно заметить, что почти все фактические значения лежат в доверительном интервале, что говорит о высокой прогнозной способности модели.

Таблица 1

Сравнение фактических значений с прогнозными, полученными по модели

	Нижняя граница доверительного интервала	Верхняя граница доверительного интервала	Прогноз, полученный по модели	Фактические значения (с учетом корректировки на инфляцию)
Январь 2024	1,1	3,1	2,1	2,0
Февраль 2024	1,6	3,8	2,7	4,7
Март 2024	2,3	5,2	3,7	3,3
Апрель 2024	2,0	5,4	3,7	3,7
Май 2024	1,0	5,5	3,3	5,4
Июнь 2024	1,1	6,6	3,9	
Июль 2024	1,8	8,1	4,9	
Август 2024	0,0	7,1	3,4	
Сентябрь 2024	0,0	7,4	3,2	
Октябрь 2024	0,0	7,6	2,8	
Ноябрь 2024	0,0	8,3	2,9	
Декабрь 2024	0,4	12,3	6,3	

В заключение данной работы можно сделать выводы, что анализ динамики поступлений НДФЛ в Волгоградской области за период 2018-2023 гг. выявил сезонность временного ряда. Это обусловлено колебаниями доходов населения, которые связаны с особенностями производственных циклов, сезонными расходами и выплатами бонусов и премий.

Для прогнозирования будущих поступлений НДСЛ была применена модель SARIMA(4,2,0)(0,1,0,12), которая учитывает сезонные и автокорреляционные свойства временного ряда.

Результаты исследования подтверждают важность учета сезонных факторов при прогнозировании поступлений НДСЛ. Это позволяет более точно планировать бюджетные доходы и эффективно управлять налоговой политикой региона.

Список использованной литературы

1. Аналитический портал ФНС России. – URL: <https://analytic.nalog.gov.ru/> (дата обращения: 16.02.2024). – Текст: электронный.

2. Федеральная служба государственной статистики: официальный сайт. – URL: <https://rosstat.gov.ru/statistics/price#> (дата обращения: 02.06.2024). – Текст: электронный.

3. Михайлов А. Анализ временных рядов / А. Михайлов. – Текст электронный // Habr. – URL: <https://habr.com/ru/companies/otus/articles/732080/> (дата обращения: 15.06.2024).

4. Himanshu Sharma Unveiling the Future: Mastering Stock Market Prediction with PMDARIMA / Himanshu Sharma // Towards AI Inc. – URL: <https://towardsai.net/p/data-analysis/unveiling-the-future-mastering-stock-market-prediction-with-pmdarima> (дата обращения: 17.06.2024).