

СТАТЬИ

УДК 519.6:004:338.2
DOI 10.17513/snt.40164

**ПРИМЕНЕНИЕ ЦИФРОВЫХ ТЕХНОЛОГИЙ ДЛЯ ПОСТРОЕНИЯ
МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ ИССЛЕДОВАНИЯ
СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ
В ПРОЦЕССЕ ОБУЧЕНИЯ СТУДЕНТОВ
ЭКОНОМИЧЕСКИХ СПЕЦИАЛЬНОСТЕЙ**

Березняк И.С.

*ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации»,
Смоленский филиал, Смоленск, e-mail: bis1605@mail.ru*

Научная публикация посвящена разработке методики, с помощью которой студенты, обучающиеся на экономических специальностях и не обладающие достаточными навыками программирования, могут проводить исследования макроэкономических показателей с использованием языка программирования Python. Цель исследования – создание поэтапного плана анализа макроэкономических показателей, который включает в себя необходимые коды языка Python для построения математических моделей различных экономических процессов и их дальнейшей интерпретации. В качестве материалов данного исследования использованы экономические показатели, представленные Росстатом и характеризующие развитие экономики России. Методами исследования выступают экономические и статистические методы, а также программные коды на языке программирования Python. В работе приведены программные коды и обращения к специализированным библиотекам, которые позволяют провести начальный этап исследования, включающий в себя установку всех необходимых для дальнейшей работы библиотек, преобразование исходного файла с данными в датафрейм, вывода на экран загруженных данных, расчет основных статистических характеристик, способы изменения типа полей. Рассмотрены коды для проведения однофакторного анализа, в ходе которого оценивается сила связи между признаками, строится тепловая карта, временные ряды проверяются на стационарность, строятся модели временных рядов, уравнения парной регрессии. Заключительный этап представляет собой использование статистических библиотек для построения модели множественной регрессии. Представлены также коды, с помощью которых можно визуализировать полученные результаты для их большей наглядности. По предложенному алгоритму с помощью представленных кодов проведено исследование ряда макроэкономических показателей, характеризующих развитие экономики России в период с 2005 по 2023 г. Выявлены наиболее сильно влияющие на валовой внутренний продукт факторы, проведена проверка временных рядов на стационарность, построены уравнения парной и множественной регрессий.

Ключевые слова: обработка больших данных, среда Python в экономических исследованиях, тепловая карта, многофакторный корреляционный анализ, расчет основных характеристик временных рядов, регрессионные модели макроэкономических показателей

**APPLICATION OF DIGITAL TECHNOLOGIES FOR CONSTRUCTING
MATHEMATICAL MODELS FOR RESEARCHING SOCIAL
AND ECONOMIC INDICATORS IN THE PROCESS
OF TEACHING STUDENTS OF ECONOMIC SPECIALTIES**

Bereznyak I.S.

*Financial University under the Government of the Russian Federation, Smolensk branch,
Smolensk, e-mail: bis1605@mail.ru*

The scientific publication is devoted to the development of a methodology by which students studying economic specialties and who do not have sufficient programming skills had the opportunity to conduct research on macroeconomic indicators using the Python programming language. The purpose of the study is to create a step-by-step plan for analyzing macroeconomic indicators, which includes the necessary Python codes for constructing mathematical models of various economic processes and their further interpretation. The materials of this study are economic indicators provided by Rosstat and characterizing the development of the Russian economy. The research methods are economic and statistical methods, as well as program codes in the Python programming language. The work provides program codes and calls to specialized libraries that allow you to carry out the initial stage of the study, including the installation of all the libraries necessary for further work, converting the original data file into a data frame, displaying the downloaded data on the screen, calculating the main statistical characteristics, and ways to change the field type. The article considers codes for conducting a single-factor analysis, during which the strength of the relationship between features is estimated, a heat map is built, time series are checked for stationarity, time series models and pair regression equations are built. The final stage is the use of statistical libraries to build a multiple regression model. The article also presents codes that can be used to visualize the results for greater clarity. According to the proposed algorithm, using the presented codes, a study was conducted of a few macroeconomic indicators characterizing the development of the Russian economy in the period from 2005 to 2023. The factors that most strongly affect the gross domestic product were identified, time series were checked for stationarity, and pair and multiple regression equations were built.

Keywords: big data processing, Python environment, heat map, correlation analysis, time series characteristics calculation, regression models

Введение

На современном этапе реализации основных положений национальной цели «Цифровая трансформация», которая утверждена Указом Президента РФ от 21 июля 2020 г. № 474 «О национальных целях развития Российской Федерации на период до 2030 г.» [1], в процессе полномасштабного перехода к цифровой экономике компетенции, связанные с грамотным использованием информационных технологий, становятся все более востребованными. В то же время студенты, обучающиеся на экономических специальностях, достаточно часто сталкиваются со сложностями, связанными с необходимостью всестороннего анализа различных статистических показателей с применением современных информационных технологий [2]. Сам процесс автоматизации построения различных моделей, расчет большого количества разнообразных показателей связан с необходимостью эффективной автоматизации рутинных расчетов с применением современных информационных технологий [3]. В связи с этим применение различных языков программирования, в частности Python, поможет избежать вычислительных ошибок и существенно сократит время построения, а также повысит качество анализа различных математических моделей, описывающих те или иные социально-экономические процессы и явления. Выбор языка программирования Python обусловлен, с одной стороны, его относительной простотой, с другой – наличием специализированных библиотек и фреймворков [4]. Стоит учесть, что для студентов непрофильных специальностей программирование зачастую является достаточно сложной задачей, поэтому наличие примерного поэтапного плана исследования экономических показателей с приведением основных кодов и объяснением полученных результатов, может оказаться весьма полезным навыком в свете необходимости приобретения актуальных цифровых компетенций [5, 6].

Целью исследования является разработка методики, которая позволила бы студентам, не владеющим навыками программирования, с одной стороны, приобрести актуальные цифровые компетенции, с другой – используя необходимые функции для построения временных рядов, выявления наличия различных форм связи между признаками, выбора наиболее адекватных моделей, описывающих парную и множественную связи, а также построение различных типов графиков и диаграмм, которые позволяют наглядно увидеть распределения значений признаков, их взаимосвязи,

тенденции развития, наличие аномальных значений с использованием языка программирования Python, проводить исследования макроэкономических показателей, что необходимо в процессе обучения студентов экономических специальностей.

Материалы и методы исследования

Многомерные временные ряды, составленные на основе данных Росстата, представленных на официальном сайте и представляющих собой ряд макроэкономических показателей, характеризующих некоторые аспекты развития экономики России в течение 19 последних лет (с 2005 по 2023 г.). Методы исследования включают в себя классические экономические и системные методы, которые дополнены применяемыми в среде программирования Python специальными методами и алгоритмами многомерного анализа и статистического моделирования.

Результаты исследования и их обсуждение

Изучение трансформации во времени ряда экономических показателей ставит перед собой следующие основные задачи; изучение динамики анализируемых показателей и построение линий тренда, выявление взаимосвязей между показателями, оценка силы и направления этой связи, построение уравнений парной и множественной регрессии.

В эпоху цифровизации экономики одним из важнейших навыков становится не только знание и грамотное использование соответствующего математического аппарата, но и умение продуктивно использовать самые современные инструменты, позволяющие автоматизировать рутинные расчеты и существенно повышающие эффективность работы за счет снижения временных затрат и существенного уменьшения вычислительных ошибок. Языки программирования, в частности Python и различные специализированные библиотеки, позволяют всесторонне анализировать самые разнообразные экономические показатели на микро- и макроуровнях.

Для иллюстрации предлагаемого теоретического материала в качестве исследуемых экономических показателей рассмотрим ряд макроэкономических показателей [7, 8], один из которых (валовой внутренний продукт (ВВП), млрд руб.) будет выступать в качестве результативного признака (Y), а ряд других показателей, которые обозначим X_i и число которых может быть произвольным, в зависимости от поставленной задачи, выступают в качестве факторных признаков [9].

Таблица 1

Основные команды начального этапа исследования макроэкономических показателей

Код	Результат выполнения
<code>i!pip install pandas numpy matplotlib seaborn statsmodels</code>	Инсталляция всех необходимых для исследования библиотек, включая библиотеку построения статистических моделей
<code>df = pd.read_excel('/content/drive/MyDrive/Название файла.xlsx')</code>	Создание датафрейма (имя может быть задано произвольно) из исходного файла Excel, содержащего значения анализируемых показателей
<code>df.head()</code> <code>df.info()</code>	Вывод на экран нескольких (количество может быть указано в качестве аргумента) строк датафрейма и полной информации о датафрейме
<code>df.describe()</code>	Вывод на экран структурных средних и величины среднеквадратического отклонения
<code>df['Имя столбца с временной компонентой'] = pd.to_datetime(df['Имя столбца с временной компонентой'])</code>	Изменение типа столбца, содержащего даты (годы, месяцы и т.п.) на временной для возможности построения рядов динамики

Таблица 2

Основные команды однофакторного анализа макроэкономических показателей

Код	Результат выполнения
<code>plt.figure(figsize=(5, 10))</code> <code>sns.heatmap(df.corr()[['Имя результативного столбца']], vmin=-1, vmax=1, annot=True);</code>	Построение тепловой карты (матрицы корреляции)
<code>from statsmodels.tsa.stattools import adfuller</code> <code>result = adfuller(df['Имя столбца'])</code> <code>for key, value in result[4].items():</code> <code>print('\t%s: %.3f' % (key, value))</code>	Применение теста Дики-Фуллера для возможности отнесения анализируемого временного ряда к стационарным
<code>df['Имя столбца без тренда'] = df['Имя столбца'] - df['Имя столбца'].rolling(window=2).mean()</code> <code>df['Имя столбца стационарные'] = df['Имя столбца без тренда'].diff()</code>	Дополнение датафрейма столбцами, в которых представлены данные показателя без тренда и приведенные к стационарному виду
<code>from statsmodels.tsa.arima.model import ARIMA</code> <code>modelvr = ARIMA(df['Имя столбца результативного показателя'], order=(1, 1, 1))</code> <code>modelvr_fit = modelvr.fit()</code> <code>print(modelvr_fit.summary())</code>	Построение и обучение модели временных рядов ARIMA, вывод статистики построенной модели, включая уравнение тренда
<code>import statsmodels.api as sm</code> <code>y = df['Имя столбца результативного показателя']</code> <code>x = df[['Имя столбца выбранного факторного признака']]</code> <code>x = sm.add_constant(x)</code> <code>model = sm.OLS(y, x).fit()</code> <code>print(model.summary())</code>	Построение уравнений регрессии результативного показателя с выбранными факторными признаками (построение парных регрессий, модель OLS() из библиотеки statsmodels) с выводом полной статистики построенной модели

На первом этапе исследования необходимо преобразовать наш исходный файл с данными, представляющий собой файл Excel, в датафрейм, предварительно импортировав необходимые для работы с датафреймами библиотеки и указав путь к файлу. Это осуществляется при помощи

соответствующих кодов, выполнение которых приведет к созданию датафрейма, содержащего исходные статистические данные. Кроме этого, на начальном этапе целесообразно вывести загруженные данные на экран, чтобы убедиться в корректности их загрузки, рассчитать основные

статистические характеристики по каждому показателю и, при необходимости, визуализировать изменения показателей, построив гистограммы или линейные графики. Коды, которые могут быть применены для проведения начального этапа исследования, и результаты их выполнения приведены в табл. 1.

Следующим этапом исследования является однофакторный анализ, предполагающий оценку силы связи между факторами, проверку временных рядов по каждому показателю на стационарность, построение

временных рядов, устранение, при необходимости, сезонной компоненты и приведения их к стационарному виду. Коды, необходимые для осуществления этого этапа, приведены в табл. 2.

Заключительный этап исследования состоит в построении модели множественной регрессии, включающей в себя факторы, наиболее сильно связанные с результативным показателем, которые могут быть отобраны на основе анализа матрицы коэффициентов парных корреляций, полученной на предыдущем этапе.

```
y = df[ 'Имя столбца результативного показателя' ]
x = df[ [ 'Имя столбца факторного признака 1', 'Имя столбца факторного признака 2', 'Имя столбца факторного признака 3', 'Имя столбца факторного признака 4', 'Имя столбца факторного признака 5' ] ]
x = sm.add_constant(x)
model = sm.OLS(y, x).fit()
print(model.summary())
```

Сводка по построенной модели позволяет не только найти уравнение множественной регрессии, но и дает возможность оценить значимость как модели в целом, так и отдельных ее коэффициентов в нормальном и стандартизованном виде.

Следует помнить о том, что визуализация полученных результатов исследования может существенно улучшить его восприятие [10]. Представление, к примеру, матрицы корреляций в виде тепловой карты гораздо более наглядно в экономи-

ческих исследованиях, чем привычная таблица. Кроме того, полученные графики и диаграммы могут быть импортированы в интерактивные отчеты и использоваться для оперативного анализа и оценки эффективности компании. Библиотеки визуализации Python представляют широкие возможности построения разнообразных графиков, как простых, так и достаточно сложных. В табл. 3 представлены основные коды, позволяющие строить различные типы графиков.

Таблица 3

Основные команды для визуализации исследуемых показателей

Код	Результат выполнения
<code>df['Имя столбца'].hist(color='purple')</code>	Построение гистограммы распределения по выбранному столбцу
<code>plt.figure(figsize=(25,2)) sns.boxplot(df['Имя столбца'],fliersize=10, color='purple') sns.color_palette()</code>	Построение “ящика с усами” для выявления выбросов
<code>px.histogram(df, x='Имя столбца 1', color='Имя столбца 2')</code>	Возможность выбора цвета из палитры
<code>px.histogram(df, x='Имя столбца 1', color='Имя столбца 2')</code>	Построение совмещенной гистограммы по двум столбцам
<code>plt.plot(df['Временная переменная'], df['Имя столбца']) plt.title('Название диаграммы') plt.xlabel('Временная переменная') plt.ylabel('Имя столбца') plt.show()</code>	Построение графика по точкам с подписями осей и выводом названия графика
<code>import plotly.express as px px.scatter(df, x='Имя столбца 1', y='Имя столбца 2', size='Имя столбца 3', hover_name='Имя столбца 4', size_max=60, color='Имя столбца 5', facet_col='Имя столбца 6', log_x=True)</code>	Построение пузырьковой диаграммы для совместного анализа нескольких признаков

По рассмотренному алгоритму было проведено исследование зависимости одного из самых важных показателей – ВВП от ряда показателей, характеризующих занятость населения, уровень доходов и инвестиционных вложений, оборот малого и среднего бизнеса, число работников, занятых в этой сфере. Таким образом, было проведено исследование зависимости резуль­тативного показателя от пяти факторных признаков. Была построена тепловая карта, проведена оценка силы связи показателя ВВП с каждым из факторных признаков, показано, что практически все факторы, за исключением численности занятых в РФ и оборота малых и средних предприятий, существенно влияют на величину ВВП (значения коэффициентов корреляции остальных факторных признаков варьируются от 0,78 до 0,98). Анализ временных рядов на стационарность показал, что все ряды, кроме уровня доходов, являются стационарными. Нестационарный ряд был дополнен столбцами, в которых были удалены сезонные компоненты и непосредственно тренд. Для каждого факторного признака, имеющего сильную корреляционную связь с резуль­тативным показателем, были построены уравнения регрессии, а также получено уравнение множественной регрессии, включающее три из первоначально рассматриваемых пяти факторных признаков.

Заключение

Описанная методика последовательного анализа макроэкономических показателей и оценка их взаимосвязи и взаимовлияния позволит студентам, обучающимся на экономических специальностях, освоить не только новый инструмент – язык программирования Python, что и позволит проводить исследования на более высоком профессиональном уровне, что особенно актуально в свете реализации национальной цели «Цифровая трансформация». С помощью предложенного алгоритма студенты смогут провести начальный этап исследования, установить все необходимые для работы библиотеки, загрузить данные, рассчитать статистические характеристики, при необходимости изменить тип загруженных данных, провести однофакторный анализ, исследовав как отдельные факторы, так и их взаимную связь, построить уравнения множественной регрессии. Полученные результаты исследования могут

быть дополнительно визуализированы, что существенно повысит качество анализа и сделает его более наглядным. Все предложенные этапы анализа были апробированы в ходе исследования показателей, характеризующих развитие экономики России в период с 2005 по 2023 г. Выявлены факторы, оказывающие наиболее сильное влияние на ВВП, проведена проверка временных рядов на стационарность, построены уравнения парной и множественной регрессий. Предложенный план анализа макроэкономических показателей позволит студентам перейти на более современный уровень исследования, способствует формированию новых цифровых компетенций и повышению уровня цифровой зрелости будущих специалистов.

Список литературы

1. О национальных целях развития Российской Федерации на период до 2030 года (Указ Президента Российской Федерации от 21.07.2020 г. № 474) [Электронный ресурс]. URL: <http://www.kremlin.ru/acts/bank/45726> (дата обращения: 20.09.2024).
2. Юдина О.В., Кириллова В.В. Выбор Python как инструмента решения задач цифровой экономики // Экономика и предпринимательство. 2023. № 2 (151). С. 1244–1246. DOI: 10.34925/EIP.2023.151.2.249.
3. Гусарова О.М., Березняк И.С., Денисов Д.Э. Математическое моделирование в среде Python активности предпринимательского сектора экономики // Современные наукоемкие технологии. 2024. № 1. С. 28–33. DOI: 10.17513/snt.39904.
4. Гусарова О.М., Березняк И.С. Теория и практика предпринимательства: экономический рост и тенденции развития малого и среднего бизнеса // Вестник Алтайской академии экономики и права. 2023. № 5–2. С. 219–224. DOI: 10.17513/vaael.2829.
5. Гусарова О.М., Березняк И.С., Попова В.В. Математическое моделирование с использованием цифровых технологий в решении прикладных задач анализа данных // Современные наукоемкие технологии. 2023. № 12–1. С. 10–15. DOI: 10.17513/snt.39853.
6. Половченко М.А., Клещева Ю.С. Значение бизнес-информатики в современной экономике // Управленческий учет. 2021. № 10–1. С. 215–220. DOI: 10.25806/uu10-12021215-220.
7. Росстат. Официальный сайт Федеральной службы государственной статистики. [Электронный ресурс]. URL: <https://rosstat.gov.ru/> (дата обращения: 15.09.2024).
8. Единый реестр субъектов малого и среднего предпринимательства. [Электронный ресурс]. URL: <https://tmssp.nalog.ru/> (дата обращения: 17.09.2024).
9. Зададаев С.А., Орлова И.В. Опыт применения эконометрического инструментария для прогнозирования показателей национальных целей развития РФ // Фундаментальные исследования. 2022. № 10–1. С. 54–59. DOI: 10.17513/fr.43343.
10. Dobryagina N.P. Behavioural economics contribution to the entrepreneurship theory and its application in entrepreneurship policies // Administrative consulting. 2021. № 1 (145). С. 50–60. DOI: 10.22394/1726-1139-2021-1-50-60.