# Моделирования социально-экономических процессов

* 1. Прогнозирование значения индекса потребительских цен на период до конца 2021 года.

В качестве исходных файлов использовался файл «Индекс потребительских цен». Данные визуализированы в виде графика (рис. 1). За 2015-2106 год наблюдается сильное отклонение (~12%) от среднего значения по всему временному диапазону. Данное отклонение обусловлено финансовым кризисом на фоне снижения цен на нефть и введённых санкций против России.



Рис. . Индекс потребительских цен в зависимости от месяца (временной период 2010 г. – 2021 г.)

На рис.1 видны редкие выбросы (выделены красным цветом). В связи с этим, для построения лучшей модели, необходимо сначала обработать исходные данные. Данные выбросы можно заменить на среднее арифметическое на всём временном промежутке или удалить из наших данных, т.к. временной промежуток имеет достаточное количество данных. В ходе проделанной работы, рассматривались оба способа обработки данных, ниже будет представлен сравнительный анализ полученных результатов. Изначально выбросы заменены на среднее арифметическое исходных данных.

Для прогнозирования использовались следующие модели:

* Авторегрессионная модель;
* Модель Хольта-Винтерса;
* ARIMA.

Для прогнозирования с использованием модели ARIMA строилась коррелограмма (рис. 2), на которой наблюдается большое количество значимых лагов, а также не стационарность ряда. Данный ряд имеет единичные корни, критерий Дики-Фуллера не выполняется *(p = 0,06).* В исходном ряду присутствует тренд, т.к. график АКФ сходится к нулю достаточно медленно. Дополнительно, на графике отсутствуют периодические всплески, значит, в ряду отсутствует сезонность.

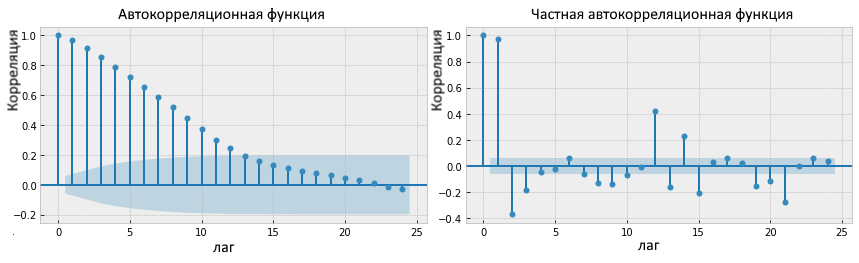
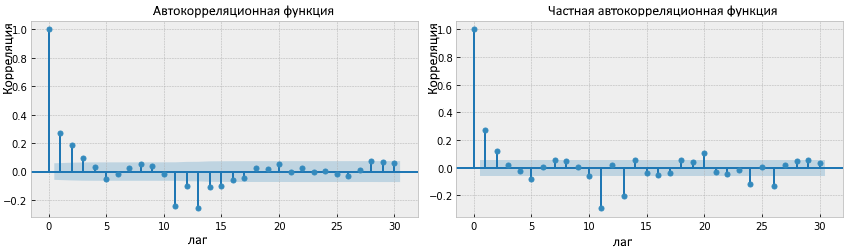


Рис. 2. Графики автокорреляционной функции и частной автокорреляционной функции

Получить стационарный ряд возможно путем взятия первых разностей (рис. 3), таким образом, исходный ряд является интегрированным рядом первого порядка. Критерий Дики-Фуллера для данного ряда выполняется, единичных корней нет. По автокорреляционной и частично автокорреляционной функцией определяем параметры для модели ARIMA *(p = 1, d = 1, q = 1)*.



В качестве тренировочной выборки, использовался временной промежуток от января 2010 г. до июня 2020 г. Оставшиеся значения использовались для тестовых данных. В результате прогнозирования, получены следующие данные (рис.3 - 5).

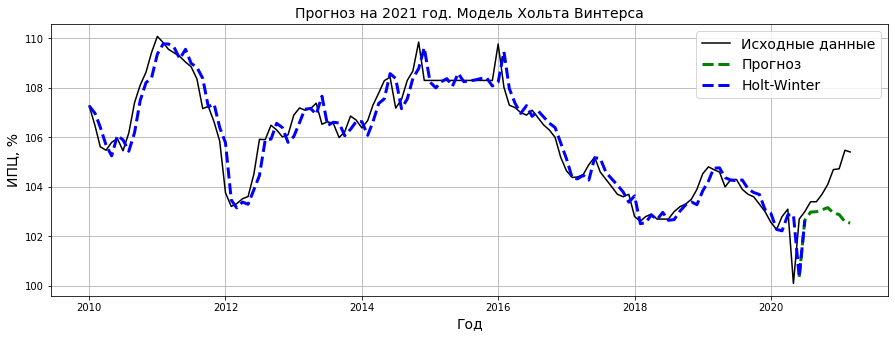


Рис. 3. Модель Хольта Винтерса. Выбросы заменены на среднее

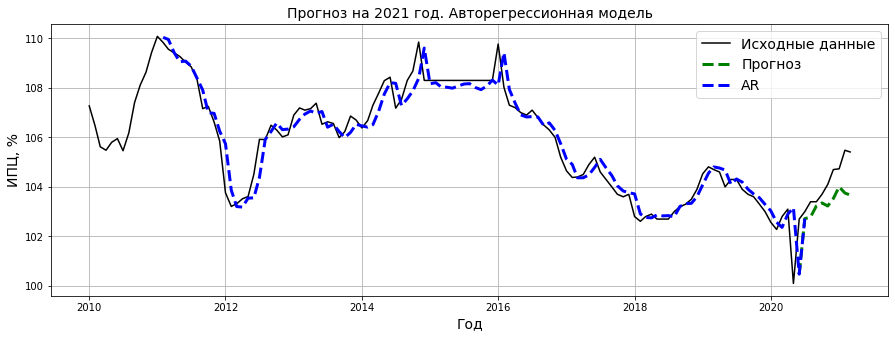


Рис. 4. Авторегрессионная модель. Выбросы заменены на среднее

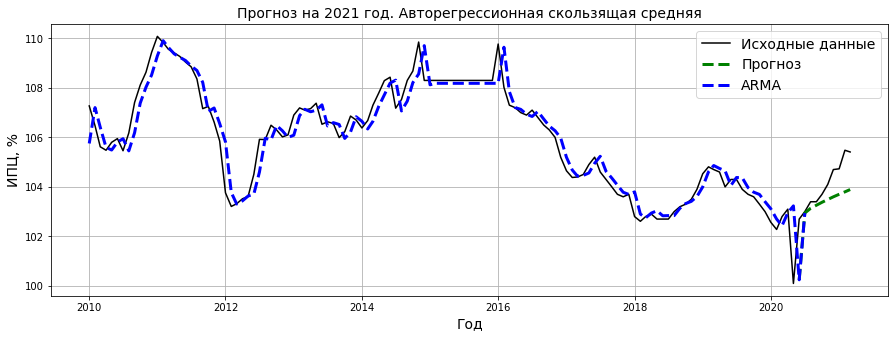


Рис. 5. Авторегрессионная скользящая средняя. Выбросы заменены на среднее

Таблица . Оценка точности моделей

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | MAE | RMSE |
| Выбросы заменены на среднее значение | | | |
| **Модель Хольта Винтерса** | -2,49 | 1,44 | 3,02 |
| **Авторегрессионная модель** | -0,64 | 0,99 | 1,42 |
| **Авторегрессионная скользящая средняя** | -0,61 | 0,92 | 1,40 |
| Выбросы удалены из данных | | | |
| **Модель Хольта Винтерса** | -4.27 | 1.71 | 3.87 |
| **Авторегрессионная модель** | **0.05** | **0.68** | **0.69** |
| **Авторегрессионная скользящая средняя** | -0.45 | 0.79 | 0.90 |

Лучше всего показала себя авторегрессионная модель с данными, в которых удалены значения выбросов (таблица 1). График тестирования такой модели представлен на рис. 6. По данной модели построено предсказание на конец 2021 года (см. рис. 7). Результаты прогнозирования представлены в таблице 2.

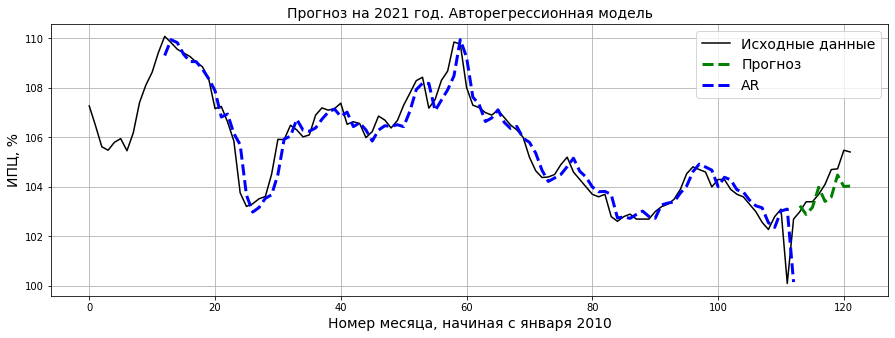


Рис. 6. Оптимальная модель для прогнозирования - авторегрессионная модель. Выбросы удалены

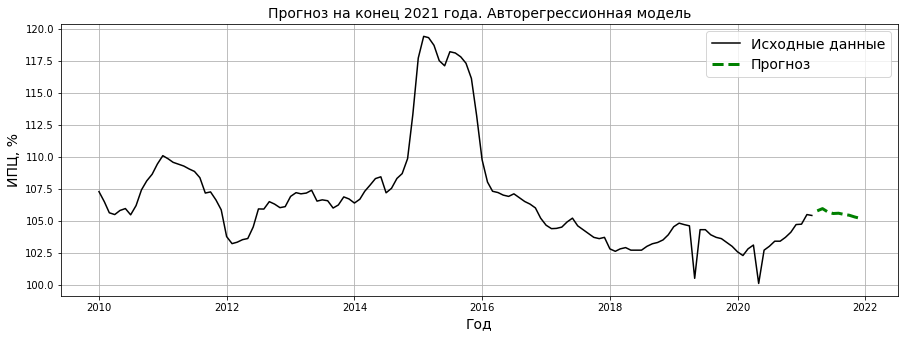


Рис. 7. Предсказание на конец 2021 года

|  |  |
| --- | --- |
| Апрель 2021 | 105,77 |
| Май 2021 | 105,94 |
| Июнь 2021 | 105,68 |
| Июль 2021 | 105,56 |
| Август 2021 | 105,58 |
| Сентябрь 2021 | 105,49 |
| Октябрь 2021 | 105,43 |
| Ноябрь 2021 | 105,29 |
| Декабрь 2021 | 105,19 |

Таблица 2. Прогноз ИПЦ на 2021 год

* 1. Постройте лучшую многофакторную регрессионную модель

На основе исходных данных, проведена корреляционная зависимость. В файле *“показатели СЭР.xlsx”* представлена полученная корреляционная таблица.

У многих признаков наблюдается, по матричным диаграммам рассеивания, линейная зависимость. Все пропущенные значения, имеющие такую зависимость с годом, найдены с помощью линейной регрессии.

На основе корреляционной зависимости и линейной регрессии производилось прогнозирование некоторых признаков до 2023 года. Заполненные столбцы представлены в таблице 3.

Таблица . Формирование признаков на 2021-2023 года

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ГОД | ЧИСЛЕННОСТЬ ПОСТОЯННОГО НАСЕЛЕНИЯ | ОБОРОТ РОЗНИЧНОЙ ТОРГОВЛИ | ДОХОДЫ НА ДУШУ НАСЕЛЕНИЯ В МЕСЯЦ | СРЕДНЕМЕСЯЧНАЯ ЗАРАБОТНАЯ ПЛАТА | ОЦЕНКА ПРОЖИТОЧНОГО МИНИМУМА В РАСЧЕТЕ НА ДУШУ НАСЕЛЕНИЯ | ВРП В ОСНОВНЫХ ЦЕНАХ | ВРП НА ДУШУ НАСЕЛЕНИЯ |
| 2021 | **5386,00** | **1636698,71** | **52197,98** | **70316,66** | **12941,72** | **4864,47** | **914,22** |
| 2022 | **5507,57** | **1721694,39** | **54766,22** | **74217,37** | **13585,92** | **5137,39** | **963,51** |
| 2023 | **5558,90** | **1806690,07** | **57334,47** | **78118,08** | **14230,12** | **5410,31** | **1012,80** |

Для прогнозирования остальных значений использовалась многофакторная регрессия. Изначально в качестве прогнозируемого столбца *Y* был выбран столбец “Объём платных услуг населению”. С помощью автоматического перебора признаков, были выделены 3 признака, имеющие наибольшую значимость для предсказываемого столбца: “ Оборот розничной торговли”, “Среднемесячная заработная плата” и “Стоимостная оценка прожиточного минимума”. С данными признаками среднеквадратическая ошибка прогноза была минимальной (), а коэффициент детерминации был максимальным ().

Коэффициенты построенной многофакторной регрессионной модели представлены в таблице 4. В данной модели t-статистика и P-статистика соответствуют требованиям, t-статистика лежит вне диапазона , а значение P менее 0,05. В таблице 5 представлены оценки качества построенной модели. У данной модели значение является достаточно значимым, близко к 1, и значение F-статистики много больше 4. Таким образом, наша модель является оптимальной.

Таблица 4. Оценки качества коэффициентов

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Наименование признака | Значение коэффициента | Стандартное отклонение | t-статистика | P>|t| |
| Стоимостная оценка прожиточного минимума | -52,0387 | 12,765 | -4,077 | 0,002 |
| Оборот розничной торговли | 1,1004 | 0,181 | 6,082 | 0 |
| Среднемесячная заработная плата | -8,3839 | 2,599 | -3,226 | 0,007 |
| const | 41310 | 15500 | 2,673 | 0,020 |

Таблица 5. Оценки качества модели

|  |  |  |  |
| --- | --- | --- | --- |
|  | F-статистика | AIC | BIC |
| 0,987 | 301,5 | 361,4 | 364,2 |

В результате подстановки коэффициентов, был спрогнозирован столбец “Количество платных услуг населению” за 2021 – 2023 года.

Аналогичным образом спрогнозированы столбцы “Численность рабочей силы”, “Численность занятых в экономике”, “ Население, имеющее среднедушевые доходы ниже прожиточного минимума / доля”, “ Децильный коэффициент фондов” и “ Коэффициент Джини”. С помощью автоматического перебора признаков для каждого прогнозируемого столбца выделены соответствующие признаки.

Показатели, оценивающие качество и адекватность всех моделей представлены в таблице 6 – 7. Полученные построенные модели являются приемлемыми, значения показателей находятся в допустимых диапазонах.

Таблица 6. Оценки качества коэффициентов для всех моделей

|  |
| --- |
| Прогнозируемый столбец “Численность рабочей силы” |
|  |
| **Прогнозируемый столбец “ Численность занятых в экономике”** |
|  |
| **Прогнозируемый столбец “Население, имеющее среднедушевые доходы ниже прожиточного минимума / доля”** |
|  |
| **Прогнозируемый столбец “Децильный коэффициент фондов”** |
|  |
| **Прогнозируемый столбец “Коэффициент Джини”** |
|  |

Таблица 7. Оценки качества всех моделей

|  |  |
| --- | --- |
| **Численность рабочей силы** | **Численность занятых в экономике** |
|  |  |
| **Население, имеющее среднедушевые доходы ниже прожиточного минимума / доля** | **Децильный коэффициент фондов** |
|  |  |
| **Коэффициент Джини** | |
|  | |

Результат прогнозирования многокритериальной регрессией представлен в таблице 8. Остальные значение прогнозирования представлены в файле *“показатели СЭР.xlsx”*. Максимальная ошибка аппроксимации всех моделей составила 3,52% (“Объем платных услуг населению”), что подтверждает адекватность всех регрессионных моделей.

Таблица 8. Прогнозирование на 2021-2023 года

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Год | Объем платных услуг населению | Численность рабочей силы | Численность занятых в экономике | Население, имеющее среднедушевые доходы ниже прожиточного минимума | Децильный коэффициент фондов | Коэффициент Джини |
| **2021** | 579334,95 | 3129,19 | 3060,02 | 7,47 | 14,38 | 0,36 |
| **2022** | 606637,73 | 3156,44 | 3086,95 | 7,26 | 14,02 | 0,35 |
| **2023** | 633940,52 | 3183,69 | 3113,88 | 7,05 | 13,65 | 0,35 |