МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий
(наименование института полностью)

Кафедра «Прикладная математика и информатика»
(наименование)
01.04.02 Прикладная математика и информатика
(код и наименование направления подготовки)

Математическое моделирование
(направленность (профиль))

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА (МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ)

на тему Построение математических моделей валового внутреннего продукта на основе регрессионного анализа

Студент А. Б. Анорова
(И.О. Фамилия) (личная подпись)

Научный руководитель к. ф-м. н., доцент, Г. А. Тырыгина
(ученое звание, степень, И.О. Фамилия)

Тольятти 2020

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Актуальность работы состоит в том, что в 2014 году российская экономика испытала ряд негативных моментов, связанных с падением курса национальной валюты и введением внешних экономических санкций против России. В связи с этим определенный интерес представляет анализ динамики валового внутреннего продукта (ВВП) российской экономики и выявление ряда факторов, определяющих тенденции его развития.

Объектом исследования является математическая модель валового внутреннего продукта.

Предметом исследования является математическая модель валового внутреннего продукта на основе регрессионного анализа.

Целью данной работы является построение математических моделей валового внутреннего продукта на основе регрессионного анализа.

Гипотеза исследования: построить математические модели валового внутреннего продукта, учитывающие факторы, определяющие его тенденцию.

Для достижения цели были сформулированы следующие **задачи:**

1. Исследование математической модели множественной регрессии в условиях нарушения предпосылок классической линейной модели множественной регрессии (КЛММР).
2. Анализ факторов, влияющих на валовой внутренний продукт.
3. Разработка математических моделей валового внутреннего продукта.

Методы исследования: компьютерное моделирование.

Публикации по теме исследования. Основные результаты теоретической части исследования изложены в статьях:

1. Анорова А. Б., Тырыгина Г. А. Анализ мультиколлинеарности и способы её преодоления: сб. науч. тр. / Прикладная математика и информатика: современные исследования в области естественных и

технических наук: материалы V Международной научно-практической конференции (школы-семинара) молодых ученых: 22-24 апреля 2019 г. – Тольятти: Издатель Качалин Александр Васильевич, 2019. 660 с.

2. Анорова А. Б. Анализ гетероскедастичности и способы её преодоления : сб. науч. тр. / Прикладная математика и информатика: современные исследования в области естественных и технических наук: материалы VI Международной научно-практической конференции (школы-семинара) молодых ученых: 23-25 апреля 2020 г. – Тольятти: Издатель Качалин Александр Васильевич, 2020.

На защиту выносятся:

Математические модели ВВП построенные на основе регрессионного анализа.

Структура. Работа состоит из введения, четырех глав, заключения.

Первая глава является теоретической и описывает процесс построения регрессионных моделей. Вторая глава является теоретической и описывает анализ регрессионных моделей с нарушением классических предпосылок. Описаны явления мультиколлинеарности, гетероскедастичности и автокорреляции. Описаны различные виды тестирования данных явлений, последствия и способы их преодоления. В третьей главе представлен обзор существующих решений данной проблемы. Четвертая глава является практической и в ней представлено построение регрессионных моделей ВВП. Описано решение проблемы мультиколлинеарности. Проведена проверка качества итоговой модели и интерпретация полученных результатов.

В заключении представлены выводы и результаты проделанной работы.

Работа изложена на 107 страницах и включает 14 рисунков, 36 таблиц, 33 источника и 3 приложения.

1 ПОСТРОЕНИЕ И АНАЛИЗ РЕГРЕССИОННЫХ МОДЕЛЕЙ

1.1 Проблемы и ошибки спецификации

Процесс построения эконометрической модели подразделяется на шесть основных этапов:

«I этап (постановочный) – постановить цели моделирования, отобрать переменные, и определить их влияние;

II этап (априорный) – проанализировать исследуемый объект, структурировать априорную информацию;

III этап (параметризация) – моделирование (выбор вида модели, вида связи, вида функции);

IV этап (информационный) – сбор статистических данных;

V этап (идентификация модели) – статистический анализ модели;

VI этап (верификация модели) – проверка адекватности данных» [1].

Рассматриваемая нами проблема решается на первых трех этапах моделирования. Именно спецификация играет решающую роль в успехе всего эконометрического исследования.

Процесс эконометрического моделирования является многоступенчатым, что позволяет постепенно улучшать и расширять модель, основываясь на статистический анализ и экспериментальную верификации результатов.

После проведения теоретического эконометрического анализа, выбора переменных, обработки статистических рядов и характеристики взаимосвязи приступают к численной оценке модели.

Экспериментальная проверка всех вариантов уравнений и переменных может повлечь за собой рассмотрение новых переменных, уравнений и верификаций. Далее проводится экспериментальный анализ модели, который может внести коррективы и вызвать необходимость в дополнительных проверках (рисунок 1).

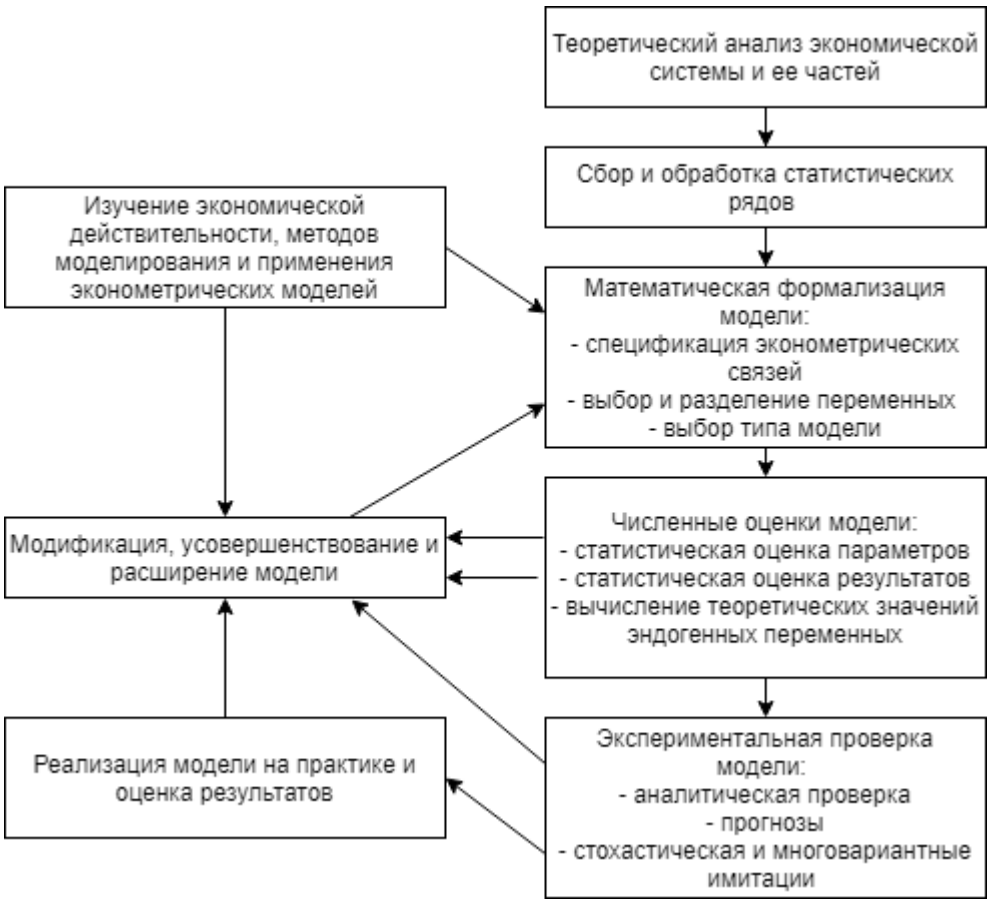

Рисунок 1.1 – Процесс построение и применения эконометрической модели.

Для понимания эконометрических процессов и связей моделирование необходимо сначала проанализировать простые – малоразмерные модели [2].

Спецификация модели – это математически описанная форма зависимости зависимой переменной от одного или нескольких объясняющих факторов. Таким образом, она подразумевает отбор факторов, включаемых в модель, и выбор формы уравнения регрессии [3].

Главная задача эконометрики – количественно описать взаимосвязи между экономическими переменными. Понятно, что она тесно связана с методами регрессии и корреляции.

Парная регрессия – регрессия между двумя переменными – x, y:

$$y = f(x), \qquad 1)$$

где y – «зависимая (эндогенная) переменная, результативный признак»;

x – «независимая (экзогенная) переменная или объясняющая переменная, факторный признак»

«Множественная регрессия – регрессия зависимой переменной с двумя и более числом факторов»:

$$y = f(x_1, x_2, x_3, \dots, x_n).$$

2)

Из всех возможных факторов, которые имеют влияние на результативный признак необходимо выбрать наиболее существенные. Парная регрессия уместна, в том случае, когда выявлен доминантный признак, который используют в качестве независимой переменной.

Уравнение парной регрессии описывает некую усредненную связь двух переменных по множеству наблюдений.

В каждом случае зависимая переменная (y_i) складывается из теоретического значения (y_{x_i}) и возмущения, которое показывает отклонение фактического значения от вычисленного (Δ_i):

$$y_i = y_{x_i} + \Delta_i.$$

Присутствие возмущения вызвано такими факторами, как: спецификация, выборочные характеристики данных, особенности измерения переменных. Оно включает влияние неучтенных в модели факторов. Иногда объясняющие переменные оказывают столь малое влияние на объясняющую переменную, что ими можно пренебречь. Но порой происходит обратная ситуация – находятся переменные, между которыми существует тесная связь. При наличии данного явления невозможно выяснить какое воздействие оказывает конкретная переменная на объясняющую переменную. Такое явление называется мультиколлинеарностью и одной из переменных необходимо пренебречь [4]. Из этого можно сделать вывод, что возмущение зависит от правильного проведения этапа спецификации.

Проблемами спецификации являются три вида задач:

1. определение набора объясняющих переменных;

2. выбор формы уравнения модели;
3. выбор модели случайного члена.

Когда с уверенностью можно сказать, что некая переменная должна входить в уравнение, то остается только определить коэффициенты и найти доверительные интервалы, а также провести проверку различных гипотез. В реальности такие ситуации очень редки, и невозможно гарантировать правильность спецификации. Однако именно от нее зависят свойства оценок коэффициентов.

К ошибкам спецификации относят:

1. пропущенные переменные;
2. включение в уравнение лишних переменных;
3. неправильный выбор вида зависимости между переменными.

При спецификации уравнения применяются разные стратегии: путем добавления переменных или путем их исключения.

Как известно, одной из причин мультиколлинеарности являются неоправданно большие значения коэффициентов с точки зрения экономической теории. Это происходит по причине суммирования влияния зависимости между переменными. Выходом из данной ситуации может стать подбор других независимых факторов или преобразование ранее выбранных переменных. Если же их присутствие необходимо, то следует применить условный метод наименьших квадратов.

Правильность спецификации модели поможет проверить коэффициент детерминации, корреляции и стандартного отклонения. Наилучшим считается уравнение, которое имеет наименьшее значение стандартного отклонения и наибольшее значение коэффициента детерминации. Остальные уравнения в неполной мере описывают существующие зависимости, следовательно, необходимо внести изменения в набор факторов.

Также верным признаком неправильной спецификации является автокорреляция возмущения, так как это указывает на пропущенные

переменные. Если определение данных переменных не представляется возможным необходимо воспользоваться квазиразностными преобразованиями или ввести трендовые переменные.

Напоследок рассмотрим еще два вида ошибок, которые не связаны с уравнением:

- Ошибки выборки – если выборка неоднородна, то уравнение регрессии не будет иметь смысла. Для решения этой проблемы необходимо увеличить объем выборки.
- Ошибки измерения – данный вид ошибок зачастую сводит на нет все усилия, по количественной оценке, связи между признаками.

1.2 Методы отбора факторов при построении регрессионных моделей

Выбор «оптимальных» факторов осуществляется на основе содержательного и количественного анализа тенденций социально-экономических процессов.

Происходит по разным причинам, например, нет необходимых данных или не получается подобрать хорошую proxy-переменную. Так возникают пропущенные переменные (omitted variables). Данная ошибка приводит к смещению оценок коэффициентов регрессии. Коэффициент при регрессоре показывает изменение зависимой переменной, вызванное единичным изменением регрессора при условии, что значения всех остальных независимых переменных, включенных в регрессию фиксированы. Смещение вызвано тем, что если переменная пропущена, то изменяется интерпретация коэффициентов при оставшихся регрессорах. Смещения не произойдет, если пропущенная переменная не входит в истинную модель или, если она не коррелирует с остальными переменными.

Помочь найти пропущенные переменные может теоретический фундамент – включать в уравнение необходимо те переменные, которые подсказывает эконометрическая теория.

На этапе верификации возникает проблема отбора группы «наилучших» независимых переменных из всего набора.

Для реализации отбора используют следующие подходы:

1. «Априорный подход» – перед началом построения модели исследуют характер и силу взаимосвязей между переменными. Переменную включают в модель в случае, когда присутствует непосредственное влияние на зависимую переменную, в противном случае – исключают. Изначальное сильное влияние фактора на зависимую переменную должно подтверждаться определенными количественными характеристиками, например, парный линейный коэффициент корреляции, позволяющий говорить о наличии связи меду переменными *y, x. «*Если два и более факторов выражают одно и то же явление, то между ними также должна существовать сильная взаимосвязь. В таких ситуациях один из факторов целесообразно исключить из модели, чтобы одна и та же причина не указывалась дважды».

2. «Статистический анализ факторов» – на основе качественных характеристик.

Совмещая различные методы и подходы задача отбора переменных решается наиболее эффективно.

При таком наборе опираются на опыт и ориентируются на содержательную сторону проблемы. Содержательный анализ позволяет решить вопрос о целесообразности включения в модель факторов, основываясь на допущениях экономической теории. Содержательный анализ помогает выявить наличие связей между переменными.

При отборе переменных нередко встречается явление ложной корреляции. На это явление обычно указывают значения коэффициентов парной корреляции переменных, которые на первый взгляд не связаны между собой, но совпали случайно.

Избежать такого рода ошибки позволяет проведение качественного анализа проблемы, который направлен на обоснование правильности содержания и формы модели.

Чем меньше факторов включаем в модель, тем более адекватной она получается. Это связано с тем, что в более сложных моделях зачастую происходит дублирование связей между переменными.

При подходе, основанном на статистическом анализе построенного варианта эконометрической модели (апостериорный подход), группу количественных характеристик образуют значения *t*-критерия Стьюдента, рассчитываемые для параметров уравнения регрессии. С помощью *t*-критерия проверяется гипотеза о значимости (существенности) влияния фактора на зависимую переменную, тем самым выявляются факторы, удаление которых целесообразно.

Используя апостериорный подход модель строится следующим образом:

2 Сначала выбираем все переменные, которые были отобраны на этапе содержательного анализа;

3 Начинают постепенное удаление тех переменных, у которых значение *t*-критерия Стьюдента наименьшее. Удаление происходит именно постепенно, т.к. на каждом последующем шаге необходимо снова производить проверку значимости переменных. Может оказаться, что ранее незначимый фактор станет значимый по причине удаления наихудшего признака ранее.

4 Данный процесс останавливается лишь тогда, когда в модели останутся лишь значимые переменные и она будет удовлетворять критериям качества. Иначе необходимо искать другой набор переменных.

Видно, что этап выбора переменных является наиболее важным при эконометрическом моделировании. Существует несколько методов выбора переменных:

- метод исключения – отсев факторов из полного его набора;
- метод включения – дополнительное введение фактора;

- пошаговый отбор переменных [5].

Пропущенные переменные (omitted variables) - происходит по разным причинам, например, нет необходимых данных или не получается подобрать хорошую proxy-переменную. Данная ошибка приводит к смещению оценок коэффициентов регрессии. Коэффициент при регрессоре показывает изменение зависимой переменной, вызванное единичным изменением регрессора при условии, что значения всех остальных независимых переменных, включенных в регрессию фиксированы. Смещение вызвано тем, что если переменная пропущена, то изменяется интерпретация коэффициентов при оставшихся регрессорах. Смещения не произойдет, если пропущенная переменная не входит в истинную модель или, если она не коррелирует с остальными переменными.

Помочь найти пропущенные переменные может теоретический фундамент – включать в уравнение необходимо те переменные, которые подсказывает эконометрическая теория [6].

Метод включения.

Например, отобраны *n* переменных.

1. Проведем *n* парных регрессий *Y* на $X_1, \dots, X_n$ и выберем переменную с наибольшим коэффициентом детерминации - R_1^2. На этом шаге необходимо выбрать лишь одну переменную.

2. Проведем *n*(n-1)* регрессий, каждый раз включая две из *n* переменных и выберем переменную с наибольшим значением R_2^2 – пара $(X^{(1)}, X^{(2)})$ – наиболее информативная пара переменных: эта пара будет иметь тесную статистическую связь с зависимой переменной *Y*. Переменная, отобранная на первом шаге не обязательно будет входить в эту пару.

3. Далее найдем тройку наилучших переменных, проведя *n*(n-1)*(n-2)* регрессий и выбирая переменные, у которых значение R_3^2 наибольшее – пара $(X^{(1)}, X^{(2)}, X^{(3)})$ – наиболее информативная пара переменных.

Строгих правил остановки нет, но есть некоторые критерии. Необходимо построить график зависимости скорректированного коэффициента детерминации выбранной совокупности переменных от числа этих переменных. Одновременно будет откладывать следующую величину:

$$R^2_{min} = R^2_{adj}(k) - 2\sqrt{\frac{2k(N-k-1)}{(N-1)(N^2-1)}}\left(1 - R^2(k)\right).$$

Оптимальное число объясняющих переменных равно числу, при котором R^2_{min} достигает своего максимума.

Понятно, что для данного метода необходимы значительные мощности, так как количество регрессий для оценки равно, например, для $p = 20$: $2^p - 2 = 1048576$.

Пошаговый отбор переменных.

В данном методе результаты каждого шага учитываются на последующих шагах.

1. Выберем переменную, имеющую наибольший коэффициент корреляции.

2. Затем, необходимо перебрать все пары, в которых будет участвовать переменная, полученная на первом шаге. Пара, которая имеет наибольший коэффициент частой корреляции, очищенный от влияния переменной, полученной на первом шаге и будет той самой информативной парой.

Когда коэффициент корреляции будет уже очень близок к нулю и, когда величина R^2_{min} достигнет своего максимума процесс следует остановить.

Описанные процедуры не гарантируют абсолютно оптимальный набор переменных, но конечный набор будет очень близок нему.

1.3 Выбор формы уравнения множественной регрессии

Для того, чтобы выбрать вид аппроксимирующей функции можно воспользоваться методами:

- графическим;
- аналитическим;
- экспериментальным;

Провести оценку тесноты связи эндогенной переменной с каждой из независимых переменных помогают диаграммы рассеивания, которые выявляют зависимость, её вид и тесноту в исследуемом соотношении.

Поле корреляции является важным и довольно простым инструментом при исследовании эконометрических взаимосвязей, но для окончательной спецификации необходимы более точные критерии.

Существует «аналитический метод выбора вида уравнения, который основан на изучении материальной природы связи исследуемых признаков».

При автоматизированной обработке информации «выбор вида уравнения регрессии обычно осуществляется экспериментальным методом, путем сравнения величины остаточной дисперсии $S^2_{\text{ост}}$, рассчитанной при разных моделях»:

$$S^2_{\text{ост}} = \frac{1}{n}\sum(y - y_x)^2.$$

Известно, что «наилучшее уравнение будет иметь наименьшую величину остаточной регрессии, и соответственно меньшее влияние неучтенных переменных».

При количественной оценке связи между двумя переменными используются следующие классы математических функций:

$$y_x = a_0 + a_1 x,$$

$$y_x = a_0 + a_1 x + a_2 x^2,$$

$$y_x = a_0 + \frac{a_1}{x},$$

$$y_x = a_0 + a_1 x + a_2 x^2 + a_3 x^3,$$

$$y_x = a_0 + x^{a_1},$$

$$y_x = a_0 + a_1^x.$$

Линейная регрессия является наиболее распространенным видом. Зная значения переменных *x* по уравнению вида $y_x = a_0 + a_1 x$ можно определить значение зависимой переменной *y*. Построение линейной регрессии сводится к оценке ее параметров - a_0, a_1. Оценить параметры такого уравнения позволит графический способ или метод наименьших квадратов.

Как и в парной регрессии, возможны «разные классы аппроксимирующих функций для множественной регрессии: как линейные, так и нелинейные».

На практике зачастую используют линейную и степенную функции. В линейной множественной регрессии

$$y_x = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$$

параметры при *x* называются коэффициентами «чистой» регрессии. Коэффициенты при независимых переменных показывают изменение соответствующей переменной на единицу, при условии, что все остальные факторы остались без изменений. Коэффициент a_0 показывает усредненное влияние факторов, не включенных в модель.

В степенной функции

$$y_x = a_0 + x_1^{a_1} \cdot x_2^{a_2} \cdot x_3^{a_3} \cdot \ldots \cdot x_n^{a_n}$$

коэффициенты $a_1, a_2, \ldots, a_n$ являются «коэффициентами эластичности, показывающими, на сколько процентов изменится в среднем результативный признак с изменением соответствующего фактора на 1% при неизменности действия других факторов». Этот вид уравнения регрессии получил наибольшее распространение в производственных функциях, исследованиях спроса и потребления.

В производственных функциях имеют смысл не только коэффициенты эластичности каждого фактора, но и их сумма – сумма эластичностей: $A = a_1, a_2, \ldots, a_n$. Величина *A* фиксирует обобщенную характеристику эластичности производства, т.е. с ростом каждого фактора на 1% выпуск продукции увеличивается на величину, равную *A*.

В эконометрическом моделировании применяются и другие линеаризуемые функции для построения уравнения множественной регрессии:

- экспоненциальная:

$$y_x = e^{a_0+a_1x_1+a_2x_2+\cdots+a_nx_n};$$

- гиперболическая:

$$y_x = \frac{1}{a_0+a_1x_1+a_2x_2+\cdots+a_nx_n},$$

которая используется при синтезировании моделей с обратными связями.

Как отмечалось ранее, «компьютерные программы обработки регрессионного анализа позволяют перебирать различные функции и выбрать наиболее подходящую – экспериментальный метод выбора аппроксимирующей функции».

Выводы по главе.

В результате работы были рассмотрены этапы построения эконометрических моделей, проблемы и ошибки спецификации:

- определение набора объясняющих переменных;
- выбор формы уравнения и модели случайного члена.

Также рассмотрены методы и подходы решения данных проблем:

- априорный и апостериорный подходы;
- метод включения/исключения и пошагового отбора переменных;
- графический, аналитический и экспериментальный методы выбора формы уравнения.

2 АНАЛИЗ РЕГРЕССИОННЫХ МОДЕЛЕЙ С НАРУШЕНИЕМ КЛАССИЧЕСКИХ ПРЕДПОСЫЛОК

В математической модели множественной линейной регрессии заложены некоторые предпосылки. Одной из таких предпосылок является предположение о том, что объясняющие переменные линейно независимы, т.е. независимы столбцы матрицы регрессоров X'. При её нарушении, т.е. когда существует тесная (сильная) корреляционная связь между экзогенными (независимыми) переменными возникает явление, которое называется мультиколлинеарностью [7]. Впервые на данную проблему внимание обратил Р. Фриш. При наличии данного явления невозможно выяснить какое воздействие оказывает конкретная переменная на объясняющую переменную.

При наличии строгой функциональной зависимости между регрессорами модели $\Upsilon = X\beta + \varepsilon$, т.е. при нарушении одного из требований Гаусса-Маркова, говорят о полной (совершенной) мультиколлинеарности. Она возникает, если хотя бы одна из независимых переменных коррелирует с другими. Такого вида мультиколлинеарность не позволяет дать однозначную оценку параметрам исходной модели.

Особым случаем полной мультиколлинеарности является наличие доминирующей переменной. Доминирующая переменная полностью определяет зависимую переменную. Зная значение доминирующей переменной, можно вычислить значение зависимой переменной без какой-либо дополнительной информации. Например, если в уравнение производственной функции обувной фабрики включить регрессор, изменяющий количество сырья, то его коэффициент будет иметь очень большое наблюдаемое значение $t-\text{статистики}^2$ на фоне традиционных значений t-статистик оценок остальных коэффициентов. Такая переменная должна быть исключена из уравнения, поскольку она идентична зависимой переменной. Зная количество поступившего на фабрику сырья, можно практически безошибочно вычислить

выпуск продукции. Включение доминирующей переменной в уравнение – это тавтология.

На практике чаще возникает другой вид мультиколлинеарности – частичная. Он имеет место, когда между объясняющими переменными точной линейной зависимости не существует, но между ними существует более или менее сильная корреляционная связь. Матрица $\mathrm{A^T A}$, где $A = (X^T X)^{-1}(X^T Y)$, будет иметь полный ранг, но ее определитель будет близок к нулю.

Природу мультиколлинеарности легко понять, если вспомнить смысл коэффициентов регрессии. Коэффициент при регрессоре показывает влияние на зависимую переменную единичного изменения регрессора при сохранении значений всех остальных переменных уравнения неизменными. При наличии явной корреляционной связи между регрессорами невозможно независимо изменять значение лишь одной переменной. Другими словами, невозможно правильно оценить коэффициент регрессии и сохранить его эконометрический смысл. Правильную оценку выполнить тем сложнее, чем выше степень мультиколлинеарности.

Причиной появления мультиколлинеарности в эконометрических моделях является наличие большого количества взаимосвязей между факторами. Данное явление встречается как в линейной регрессии, так и в регрессии временных рядов. Когда имеют дело с трендом, не требующим независимости наблюдений зачастую автоматически возникает явление мультиколлинеарности.

Необходимость исследования мультиколлинеарности возникает только при наличии ее серьезного влияния на результаты оценки регрессии. Наибольшую роль играет не столько вид мультиколлинеарности, сколько ее степень выраженности. Оценка регрессии не пострадает лишь в том случае, если все независимые переменные окажутся абсолютно некоррелированными. Итак, явление мультиколлинеарности становится проблемой, когда тесная

корреляционная зависимость между регрессорами ведет к получению ненадежных оценок регрессии.

Перечислим последствия, к которым приводит мультиколлинеарность регрессоров:

- оценки коэффициентов остаются несмещенными;
- дисперсии оценок коэффициентов возрастают – вследствие этого возрастает и вероятность получения неправильного знака оценки коэффициента;
- абсолютные значения t-статистики могут уменьшится – это происходит не только по причине роста стандартных ошибок оценок коэффициентов, но и из-за увеличения вероятности попадания оценок в интервал вблизи нулевого значения;
- оценки коэффициентов становятся очень чувствительными к изменениям спецификации – в результате иногда исключение незначимой переменной приводит к драматическим изменениям оценок коэффициентов;
- оценки коэффициентов становятся очень чувствительными к изменениям выборки – изменение исходных данных может привести к существенному изменению оценок коэффициентов модели;
- общая значимость уравнения значительно ухудшается – t-тест показывает хорошее качество подгонки, несмотря на низкие наблюдаемые значения t-статистик оценок коэффициентов. Мультиколлинеарность не снижает качество прогноза или предсказания до тех пор, пока выбранное для прогноза значение независимой переменной не усилит степень мультиколлинеарности, наблюдавшуюся при оценке коэффициентов регрессии;
- оценки коэффициентов немультиколлинеарных переменных не ухудшаются [8].

2.1 Тестирование мультиколлинеарности

Строго говоря, мультиколлинеарность присутствует в любом наборе данных, собранных для статистического обоснования корректно составленной теоретической модели. Невозможно в реальном мире найти набор независимых переменных, полностью некоррелируемых между собой. Поэтому перечисленные ниже рекомендации необходимо использовать для оценки степени, а не обнаружения мультиколлинеарности.

Степень мультиколлинеарности зависит не только от спецификации уравнения, но и от выборки наблюдений. Сильная мультиколлинеарность может пропасть при смене набора данных. Искусством эконометрики является нахождение теоретически существенной переменной, статистически немультиколлинеарной с другими переменными.

Укажем некоторые способы определения наличия мультиколлинеарности. Какой именно способ будем использоваться на практике решают, исходя из размерности задачи и имеющихся программ. Следует помнить, что различные методы могут давать разные результаты.

2. Тест на мультиколлинеарность (Фаррара-Глоубера) – близость к нулю определителя матрицы X'X .

Тест содержит три вида статистических критериев проверки наличия мультиколлинеарности.

I. Проверка наличия мультиколлинеарности всей матрицы переменных (критерий хи-квадрат).

Построить корреляционную матрицу регрессоров R и найти ее определитель *det[R]*. Вычислить наблюдаемое значение статистики Фаррара-Глоуберга по формуле:

$$FG_{\text{набл}} = -\left[n - 1 - \frac{1}{6}(2k + 5)\right]\ln(\det[R]),$$

где n – количество наблюдений; k – количество факторов. Доказано, что эта статистика имеет распределение χ^2 (хи-квадрат). Фактическое значение этого

критерия сравнивается с табличным значением χ^2 с *0,5k (k-1)* степени свободы и уровнем значимости *α.* Если $FG_{\text{набл}}$ больше табличного, то считаем, что в матрице объясняющих переменных существует мультиколлинеарность.

II. Проверка наличия мультиколлинеарности каждого фактора с другими (F-критерий).

Вычислить обратную матрицу $C = R^{-1}$. Вычислить F-критерии для каждого фактора $F_j = (c_{jj} - 1)\frac{n-k-1}{k}$, где c_{jj}- диагональные элементы матрицы *C.* Фактические значения F-критериев сравниваются с табличным значением $v_1 = k$ и $v_2 = (n - k - 1)$ степенями свободы и с уровнем значимости *α,* где *k* – количество факторов. Если $F_i > F_{\text{табл}}$, то считаем, что X_j имеет тесную корреляционную связь с другими регрессорами.

III. Проверка мультиколлинеарности каждой пары переменных (t-тест).

Найти частные коэффициенты корреляции: $r_{ij()} = \frac{-c_{ij}}{\sqrt{c_{ii} \cdot c_{jj}}}$, c_{ij} – элементы *i*-й строки и *j*-го столбца матрицы *C*. Вычислить *t*-критерии:

$$t_{ij} = \frac{r_{ij()}\sqrt{n-k-1}}{\sqrt{1-r_{ij()}^2}}$$

Фактические значения критериев t_{ij} сравниваются с табличным $t_{\text{табл}}$ с $(n - k - 1)$ степенями свободы и с уравнением значимости *α.* Если $|t_{ij}| > t_{\text{табл}}$, то считаем, что между независимыми переменными i и j существует мультиколлинеарность [9].

3. Обнаружение мультиколлинеарности на основе анализа матрицы парных корреляций между факторами:

$$R = \begin{bmatrix} r_{x_1x_1} & r_{x_1x_2} & \cdots & r_{x_1x_p} \\ r_{x_2x_1} & r_{x_2x_2} & \cdots & r_{x_2x_p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{x_px_1} & r_{x_px_2} & \cdots & r_{x_px_p} \end{bmatrix} = \begin{bmatrix} 1 & r_{x_1x_2} & \cdots & r_{x_1x_p} \\ r_{x_2x_1} & 1 & \cdots & r_{x_2x_p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{x_px_1} & r_{x_px_2} & \cdots & 1 \end{bmatrix} \qquad 1)$$

При помощи коэффициентов парной корреляции $r_{x_i x_j}$ между независимыми переменными можно выявить дублирующие факторы. При значении коэффициента $r_{x_i x_j} \geq 0.8$, можно с уверенностью сказать, что между независимыми переменными x_i и x_j существует линейная зависимость, а факторы будут называться явно коллинеарными. Данное правило является эмпирическим.

4. Высокий фактор роста дисперсии (VIF). Метод оценки фактор роста дисперсии основан на вычислении степени объяснения рассматриваемой независимой переменной другими независимыми переменными.

Пусть уравнение содержит *k-1* независимых переменных:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_{k-1} X_{k-1t} + \varepsilon_t.$$

Вычислить OLS-оценку регрессии, выбрав в качестве зависимой переменной X_t, исключив ее из первой части каждой регрессии:

$$X_{it} = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \cdots + \alpha_{i-1} X_{i-1t} + \alpha_{i+1} X_{i+1t} + \cdots + \alpha_k X_{kt} + v_t.$$

где *v* – случайный член, *i – 1,2,…,k-1*.

Вычислить фактор роста дисперсии (VIF) для каждой $\widehat{\beta_l}$ по формуле:

$$VIF\left(\widehat{\beta_l}\right) = (1 - R_i^2)^{-1}.$$

Сравним VIF всех $\widehat{\beta_l}$. Чем выше VIF рассматриваемой оценки коэффициента, тем выше влияние мультиколлинеарности на оценку *i-го* коэффициента регрессии $\left(\widehat{\beta_l}\right)^1$.

К сожалению, нет однозначного критерия принятия решения на основе VIF. Критическое значение может колебаться от 5 до 10. В то же время возможно наблюдение VIF=4,4 в случае, когда коэффициент парной корреляции независимых переменных равен 0,88.VIF не превышает критическое значение, но коэффициент корреляции говорит о высокой вероятности коллинеарности.

5. Мультиколлинеарность будет считаться доказанной, если гипотеза H_0 о независимости переменных, т.е. $Det|R| = 1$ будет отклонена. Учитывая,

что величина $\left[n-1-\frac{1}{6}(2n+5)lgDetR\right]$ имеет приближенное распределение χ^2 с $df=\frac{1}{2}p(p-1)$ степенями свободы. Если фактическое значение χ^2 превосходит табличное (критическое) $\chi^2_{\text{факт}} > \chi^2_{\text{табл}(df,a)}$, то гипотеза H_0 отклоняется.

6. Выявить мультиколлинеарность можно по коэффициенту множественной детерминации $R^2_{x_1 \vee x_2 x_3 \ldots x_p}; R^2_{x_2 \vee x_1 x_3 \ldots x_p}$ .... Его получают по уравнениям регрессии, в которых в качестве зависимой переменной рассматривают один из факторов. Чем ближе окажется значение коэффициента детерминации к единице, тем сильнее проявляется мультиколлинеарность факторов. Например, эмпирическое правило гласит что, при значении коэффициента множественной детерминации $R^2_{x_1 \vee x_2 x_3 \ldots x_p} > 0.6$ мультиколлинеарность факторов считается установленной [10].

7. Неправильный знак или значение (с точки зрения здравого смысла или экономической теории) оценок коэффициентов регрессии.

Перечисленные признаки не являются необходимыми условиями сильной мультиколлинеарности. Иногда сильная мультиколлинеарность может существовать и при выполнении перечисленных выше условий.

2.2 Методы устранения мультиколлинеарности

Для преодоления явления мультиколлинеарности между факторами используются следующие способы**:**

1 Отбор наиболее информативных объясняющих переменных в модель регрессии. Для начала из модели исключаются независимые переменные, имеющие сильную корреляционную связь, и модель оценивается заново. Переменные исключаются на основании коэффициента корреляции, а именно - по значению оценки значимости коэффициентов корреляции. Согласно эмпирическому правилу при значении коэффициента парной корреляции больше 0,8 одну из переменных можно исключить. Но какую

именно переменную удалить из анализа, решают исходя из экономических соображений. При исключении переменных из модели возможны ошибки спецификации. Поэтому в прикладных эконометрических модели начинают исключать только когда коллинеарность становится серьезной проблемой.

2 Переход с помощью линейного преобразования к новым некоррелирующим независимым переменным. Способ заключается в переходе к регрессии приведенной формы путем замены коллинеарных переменных на их линейную комбинацию.

3 Исключение тренда. При построении регрессии по данным, полученным из временных рядов, рекомендуется исключить тренд или компенсировать изменение последовательных значений переменных. С помощью этого предпосылки регрессионного анализа будут выполняться.

4 Переход к смещенным оценкам, имеющим меньшую дисперсию.

5 «Использование предварительной информации. Обычно на основе ранее проведенного регрессионного анализа или в результате эконометрических исследований уже имеется более или менее точное представление о величине или соотношении двух, или нескольких коэффициентов регрессии. Эта предварительная информация может быть использована исследователем при построении регрессии. В связи с тем, что часть оценок, полученных на основе вне выборочных данных уже имеет достаточно четкую интерпретацию, это облегчает путь обнаружения взаимных влияний изменений различных переменных» [11].

6 Метод дополнительных регрессий. Строятся уравнения регрессии, которые связывают каждую независимую переменную со всеми остальными. Затем необходимо вычислить коэффициенты детерминации R^2 для каждого уравнения регрессии и проверить статистическую гипотезу $H_0: R^2 = 0$с помощью F-теста. Вывод: если гипотеза H_0 не отвергается, то данный регрессор не приводит к мультиколлинеарности.

7 Метод последовательного присоединения. «Регрессионная модель строится, учитывая все предполагаемые регрессоры. По признакам делается вывод о возможном присутствии мультиколлинеарности. Затем вычисляется матрица корреляций и выбирается регрессор, который имеет наибольшую корреляцию с выходной переменной. Добавляя последовательно оставшиеся регрессоры к выбранному регрессору для каждой из моделей вычисляются скорректированные коэффициенты детерминации. К модели необходимо присоединить регрессор, обеспечивающий наибольшее значение скорректированного R^2. Процесс присоединения регрессоров прекращается, когда значение скорректированного R^2 становится меньше достигнутого на предыдущем шаге» [12].

8 Метод предварительного центрирования. Суть метода сводится к тому, что перед нахождением параметров математической модели проводится центрирование исходных данных: из каждого значения в ряде данных вычитается среднее по ряду: $Y_t' = Y_t - \acute{Y}$. В результате этого оценки модели становятся устойчивыми. Каким бы образом не осуществлялся отбор факторов, уменьшение их числа приводит к улучшению обусловленности матрицы, а, следовательно, и к повышению качества оценок параметров модели.

Следует также учитывать ограничение, накладываемое на количество факторов, имеющимся числом наблюдений. Количество наблюдений должно превышать количество факторов более чем в 6-7 раз.

Итак, проблема мультиколлинеарности на сегодня еще окончательно не решена. Однако, используя различные подходы, мы пытаемся определить наличие мультиколлинеарности, чтобы затем по возможности с помощью того или иного метода ее уменьшить. Если же это не удается, то к оценкам коэффициентов регрессии и значениям регрессии надо относиться с большой осторожностью.

Итак, проблема мультиколлинеарности на сегодня еще окончательно не решена. Однако, используя различные подходы, мы пытаемся определить

наличие мультиколлинеарности, чтобы затем по возможности с помощью того или иного метода ее уменьшить. Если же это не удается, то к оценкам коэффициентов регрессии и значениям регрессии надо относиться с большой осторожностью.

2.3 Гетероскедастичность.

Рассмотрим один из случаев обобщенной регрессионной модели – модель с гетероскедастичностью. Гетероскедастичность является нарушением 5-го условия классической модели регрессии. При анализе неоднородных данных данная проблема возникает довольно часто. Различают чистую и нечистую гетероскедастичность.

Чистая гетероскедастичность наблюдается в правильно специфицированных уравнениях.

Нечистая гетероскедастичность вызывается ошибками спецификации. Самой распространенной причиной является пропуск существенной переменной. Неправильная функциональная форма вызывает гетероскедастичность гораздо реже.

Пропущенная существенная переменная может вызывать нечистую гетероскедастичность по причине, что ее влияние на зависимую переменную частично поглощается случайным членом.

Символьная запись нарушения гомоскедастичности имеет вид:

$$V(\varepsilon_t) = \sigma_t^2 \qquad t = 1,2,\dots,n. \tag{1}$$

Эта запись означает, что дисперсия случайного члена может зависеть от номера наблюдения.

Гетероскедастичность (чистая) имеет огромное число форм. Рассмотрим лишь основные принципы.

Пусть дана простейшая модель:

$$V(\varepsilon_t) = \sigma^2 Z_t^2 \qquad t = 1,2,\dots,n, \tag{2}$$

где Z – фактор пропорциональности – экзогенная переменная, не обязательно являющаяся регрессором уравнения [13].

2.4 Последствия гетероскедастичности

Рассмотрим на примере данной модели возможные последствия гетероскедастичности:

Чистая гетероскедастичность не вызывает смещений оценок коэффициентов. В оценках уравнений с нечистой гетероскедастичностью эти смещения возможны.

Гетероскедастичность увеличивает дисперсии распределений оценок коэффициентов.

Гетероскедастичность вызывает занижение стандартных ошибок OLS-оценок коэффициентов уравнения. Это смещение существует до тех пор, пока дисперсия случайного члена положительно коррелирована с абсолютными значениями наблюдений, что характерно для экономических примеров. В результате исследователь не должен полагаться на значения t- и F-статистик, поскольку это может привести к ошибочным выводам [14].

2.5 Диагностика гетероскедастичности

Разнообразию форм гетероскедастичности соответствует и разнообразие тестов, определяющих ее. Однако выбор необходимо теста осложняется тем, что природа гетероскедастичности в каждом конкретном случае практически не известна. Модель (2) является одним из примеров ее спецификации.

2.5.1Тест Парка

Данный тест используется для выявления правдоподобия гетероскедастичности вида (2) и основан на анализе остатков.

Перед проведением теста необходимо выполнить следующие действия:

Выбрать из всех моделей наилучшую с точки зрения устранения ошибок спецификации.

Оценить методом OLS коэффициенты модели.

Выявить фактор пропорциональности, построив график остатков:

$$e_t = Y_t - \widehat{\beta_0} - \widehat{\beta_1}X_{1t} - \widehat{\beta_2}X_{2t} - \cdots - \widehat{\beta_{k-1}}X_{k-1t} \quad (3)$$

Последовательность выполнения теста следующая:

Оценивается парная регрессия логарифма квадратов остатков (3) на логарифм фактора пропорциональности Z:

$$\ln(e_t^2) = \alpha_0 + \alpha_1 lnZ_t + u_t, \quad (4)$$

где u_t – классический (гомоскедастичный) случайный член.

Тестируется значимость коэффициента α_1 уравнения (4) с помощью t-теста. Обычно используют двусторонний тест. Это объясняется тем, что на практике может наблюдаться зависимость дисперсии остатков от фактора пропорциональности, отличная от квадратичной. Значимое отличие коэффициента от нуля свидетельствует о наличие гетероскедастичности с фактором пропорциональности Z.

Основная проблема при проведении теста Парка заключается в необходимости выявления фактора пропорциональности. Довольно часто он является одной из объясняющих переменных, но это не обязательно.

Порядок проведения теста показывает, что доказать отсутствие гетероскедастичности в наблюдениях невозможно [15].

2.5.2Тест Голдфелда-Квандта

Наиболее широко распространенный тест.

Выполняется в следующей последовательности:

1. Наблюдения сортируются по возрастанию значения предполагаемого фактора пропорциональности Z.

2. Отдельно выполняются OLS-оценки первой и последней трети отсортированных наблюдений. При этом используется спецификация оригинального уровня.

3. Вычисляют $GQ = RSS_3/RSS_1$, где RSS_1 и RSS_3 – суммы квадратов остатков оценок регрессий по первой и последней трети отсортированных наблюдений соответственно.

4. Используется F-тест для проверки нулевой гипотезы о гомоскедастичности. Сравнивается наблюдаемое значение GQ с критическим значением F-статистики для $n_3 - k$ и $n_1 - k$ степеней свободы $F(n_3 - k), F(n_1 - k), k$ – число оцениваемых коэффициентов регрессии (число коэффициентов наклона, плюс постоянный член), n_1 и n_3 – число наблюдений в каждой из оцененных регрессий.

Правило принятия решений. Если GQ больше критического значения F-статистик, делают вывод о том, что сумма квадратов остатков регрессии, построенной по последней трети наблюдений, значимо превышает сумму, соответствующую первой трети наблюдений, гипотезу о гомоскедастичности отвергают в пользу альтернативной (гетероскедастичности). Если GQ меньше критического значения F-статистики, это говорит, что проведенные наблюдения не дают оснований отклонить нулевую гипотезу. Но это не означает, что наблюдения гомоскедастичны. Возможно, фактор пропорциональности выбран неправильно.

К сожалению, использование тестов Парка и Голфелда-Квандта требует знания факторов пропорциональности. Часто их не удается найти до проведения теста. Если несколько величин являются возможными кандидатами на роль факторов пропорциональности, нет смысла проводить тесты Парка или

Голфелда-Квандта отдельно для каждой из этих величин. Лучше воспользоваться тестом Бреуша-Пагана или тестом Уайта [16].

2.5.3Тест Бреуша-Пагана

Позволяет проверить ответственность за гетероскедастичность нескольких факторов пропорциональности одновременно, одним тестом.

Проводится следующим образом:

Вычисляются остатки оцененного уравнения, подозреваемого на гетероскедастичность.

Оценивается регрессия квадратов полученных остатков на все переменные, от которых может зависеть дисперсия случайного члена тестируемого уравнения:

$$e_t^2 = \alpha_0 + \alpha_1 Z_{1t} + \alpha_2 Z_{2t} + \cdots + \alpha_m Z_{mt} + u_t. \quad 5)$$

В качестве факторов пропорциональности Z могут быть выбраны любые переменные, не обязательно содержащиеся в исходном уравнении. Их функциональная форма в (5) может быть произвольна.

Тестируется общая значимость оценки уравнения (5):

$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_m = 0,$

$H_1: H_0$ не верна.

Наблюдаемое значение статистики вычисляется по формуле:

$$BP = ESS / \left[2\left(\sum \frac{e_t^2}{n}\right)^2\right], \quad 6)$$

где ESS – объясненная сумма квадратов уравнения (5); e – остатки (3); n – число наблюдений.

Распределение BP асимптотически стремится к распределению χ^2 с числом степеней свободы m.

Правило принятия решения. Если BP превышает критическое значение, делают вывод о том, что оценка уравнения (5) в целом значима, нулевая гипотеза отвергается в пользу альтернативной. Этот вывод означает, что

дисперсия остатков тестируемого уравнения непостоянна, возможна гетероскедастичность. Если BP меньше критического значения, наблюдения не дают оснований отвергнуть нулевую гипотезу о гомоскедастичности.

Недостатки данного метода:

- Необходимость задания формы тестируемой гетероскедастичности.
- Большое число наблюдений.

2.5.4Тест Уайта

Преимуществом теста является абсолютное отсутствие требований о природе гетероскедастичности. Ввиду универсальности данный метод реализован во многих пакетах.

Последовательность выполнения:

- Вычисляются остатки оцененного исходного уравнения, подозреваемого на гетероскедастичность.
- Оценивается регрессия квадратов полученных остатков на все независимые переменные исходного уравнения, их квадраты и перекрёстные произведения:

$$e_t^2 = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \cdots + \alpha_{k-1} X_{k-1t} + \alpha_k X_{1t}^2 + \\ + \alpha_{k+1} X_{1t} X_{2t} + \cdots + u_1. \quad (7)$$

- Тестируется нулевая гипотеза об отсутствии гетероскедастичности в форме Уайта против альтернативной, предполагающей гетероскедастичность. С этой целью вычисляется наблюдаемое значение $Obs * R^2$ – число наблюдений умноженное на R^2. Эта статистика асимптотически распределена по закону χ^2 с числом степеней свободы, равным числу коэффициентов наклона уравнения (7).

Правило принятия решения. Если $Obs * R^2$ превышает критическое значение, нулевая гипотеза отвергается в пользу альтернативной. Делается вывод о наличии гетероскедастичности. Если $Obs * R^2$ меньше критического значения, наблюдения не дают оснований отвергнуть нулевую гипотезу о гомоскедастичности в форме Уайта.

Иногда (когда наблюдений слишком мало) уравнение (7) не может быть оценено из-за отрицательного числа степеней свободы. Число степеней свободы можно увеличить, исключив из уравнения (7) все перекрёстные произведения (сохранив регрессоры исходного уравнения и их квадраты).

2.6 Преодоление гетероскедастичности

Поскольку OLS-оценка гетероскедастичной модели не является BLUE-оценкой, но всё-таки остается несмещенной, то возможны случаи, когда не следует бороться с гетероскедастичностью. Искусство эконометрики заключается в обнаружении таких случаев.

Первый шаг в борьбе с гетероскедастичностью – определение ее типа. Если гетероскедастичность нечистая, для ее устранения необходимо включить в уравнение пропущенные существенные переменные и подобрать правильную функциональную форму.

При подозрении на чистую гетероскедастичность можно предпринять одно из следующих действий:

1. Воспользоваться взвешенным МНК.
2. Переопределить переменные.
3. Пересчитать оценки стандартных ошибок оценок коэффициентов регрессии с поправками в форме Уайта на гетероскедастичность.

2.6.1 Взвешенный метод наименьших квадратов

Фактически данный метод является одной из версий GLS.

В случае, если исходное уравнение

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_{k-1} X_{k-1t} + \varepsilon_t \qquad 8)$$

имеет гетероскедастичный случайный член (2) с известным фактором пропорциональности Z, то оценка взвешенного МНК реализуется как OLS-оценка уравнения (8) после умножения обеих частей на 1/Z:

$$Y_t/Z_t = \beta_0/Z_t + \beta_1 X_{1t} Z_t + \beta_2 X_{2t} Z_t + \cdots + \beta_{k-1} X_{k-1t} Z_t + \varepsilon_t Z_t$$

9)

Если фактор пропорциональности не выявлен, рекомендуется следующий двухшаговый метод борьбы с гетероскедастичностью.

1. Выполняется OLS-оценка уравнения (8), подозреваемого на гетероскедастичность.

2. Вычисляются остатки (3).

3. Методом OLS оценивается регрессия квадратов этих остатков на те регрессоры, от которых может зависеть дисперсия случайного члена. В случае, если нет гипотезы о спецификации гетероскедастичности, в качестве объясняющих переменных этой регрессии используются регрессоры уравнения (8), их квадраты и перекрестные произведения.

$$e_t^2 = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \cdots + \alpha_{k-1} X_{k-1t} + \alpha_k X_{1t}^2 + \ldots + u_1.$$

10)

4. Вычисляются прогнозные значения квадратов остатков (зависимой переменной в (10)) на основе оценок коэффициентов, полученных в предыдущем шаге:

$$\widehat{e_t^2} = \widehat{\alpha_0} + \widehat{\alpha_1} X_{1t} + \widehat{\alpha_2} X_{2t} + \cdots + \widehat{\alpha_{k-1}} X_{k-1t} + \widehat{\alpha_k} X_{1t}^2 + \cdots$$

11)

5. Выполняется OLS-оценка уравнения (9), где в качестве Z_t используется корень квадратный из соответствующего прогнозного значения $\widehat{e_t^2}$ (11).

Оценка взвешенного МНК регрессии с нечистой гетероскедастичностью уменьшит смещение, вызванное пропущенной переменной, но обычно оказывается хуже оценки уравнения с правильной спецификацией.

Переопределение данных подразумевает выбор переменных, снижающих вероятность возникновения гетероскедастичности. Это выполняется на основе теоретических моделей [17].

2.7 Автокорреляция

Автокорреляция представляет собой нарушение классического предположения о том, что значения случайного члена в различных наблюдениях некоррелированы.

Автокорреляция более вероятна в той выборке, для которой имеет значение последовательность полученных данных. Таким образом, она чаще встречается во временных рядах.

Чистая автокорреляция – нарушение требования условия $E(\varepsilon_j \varepsilon_t) = 0$, при $j \neq t$, при отсутствии ошибок спецификации.

В модели с чистой автокорреляцией первого порядка ошибки (ε) подчинены следующему рекуррентному соотношению:

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t \qquad (12)$$

где ρ – параметр, отражающий функциональную связь между последовательными наблюдениями случайного члена $(-1 < \rho < 1)$; u – классический случайный член.

Функциональную форму (12) часто называют марковской схемой первого порядка. Параметр ρ – коэффициент автокорреляции первого порядка.

Существуют и другие формы чистой автокорреляции. Например, квартальная или сезонная:

$$\varepsilon_t = \rho \varepsilon_{t-4} + u_t, \qquad (13)$$

автокорреляция второго порядка:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + u_t. \qquad (14)$$

Аналогично могут быть записаны выражения для автокорреляционных процессов более высоких порядков.

Нечистая автокорреляция является следствием ошибок спецификации, таких, как пропуск существенной переменной или неправильная функциональная форма. Поскольку этот тип корреляции вызван

неправильными действиями исследователя, то его всегда можно устранить, по крайней мере, теоретически.

Каким образом ошибки спецификации могут вызвать автокорреляцию? Вспомним, что случайный член содержит влияние пропущенных переменных, нелинейности, ошибок измерения и стохастических возмущений зависимой переменной. Если мы пропускаем существенную переменную или используем неправильную функциональную форму, то часть упущенных эффектов, необъяснимая оставшимися регрессорами, поглощается случайным членом.

2.7.1Последствия

Каждая из описанных выше проблем имеют собственные внешние проявления. Они обычно дают достаточную информацию для диагностики и устранения возможных проблем. Автокорреляция имеет ряд внутренних проявлений, которые не столь легко наблюдаемы. Существуют следующие характерные последствия автокорреляции.

Чистая автокорреляция не вызывает смещений оценок коэффициентов регрессии.

Автокорреляция увеличивает дисперсии распределений оценок коэффициентов.

Автокорреляция является причиной того, что OLS-оценки стандартных ошибок оценок коэффициентов регрессии являются заниженными.

Последнее последствие иногда маскирует второе. По наблюдаемым завышенным значениям t-статистик оценок коэффициентов регрессии можно сделать ошибочные выводы о значимости этих коэффициентов.

К сказанному следует добавить, что в условиях автокорреляции OLS-оценка не является BLUE.

2.7.2 Диагностика автокорреляции

Тесты Дарбина-Уотсона.

Они основаны на анализе остатков оценки уравнения регрессии.

Следующие предположения должны быть выполнены для того, чтобы можно было воспользоваться d-статистикой Дарбина-Уотсона.

1. Уравнение регрессии должно включать постоянный член.
2. Наблюдения имеют автокорреляцию первого порядка.
3. Независимые переменные уравнения регрессии не содержат лаговую зависимую переменную.

При больших выборках наблюдаемое значение d-статистик Дарбина-Уотсона может быть рассчитано по формуле:

$$d \approx 2(1 - \hat{\rho}), \quad (15)$$

где ρ – коэффициент регрессии остатков на их значения с лагом в один временной период.

Из выражения (15) видно, что значение d-статистики лежит в интервале (0;4), приближаясь к нулю при увеличении положительной автокорреляции, к четырем – отрицательной и равно двум в случае отсутствия автокорреляции первого порядка.

Особенность традиционного d-теста Дарбина-Уотсона – наличие зоны неопределенности. При попадании наблюдаемого значения d-статистики в эту зону не удается прийти к выводу о наличии или отсутствии оснований для отклонения нулевой гипотезы.

Порядок проведения d-теста Дарбина-Уотсона совпадает с порядком проведения t- и F-тестов.

1. Формулируются нулевая и альтернативная гипотезы в виде:

$H_0: \rho \leq 0$ (нет положительной автокорреляции);

$H_1: \rho > 0$ (нет положительной автокорреляции); (16)

или

$H_0: \rho = 0$ (нет автокорреляции);

$H_1: \rho \neq 0$ (автокорреляция); (17)

2. Выбирается уровень значимости, определяются размер выборки, число регрессоров и по соответствующим таблицам вычисляются верхнее d_U и нижнее d_L – критические значения d-статистики Дарбина-Уотсона.

3. Оценивается регрессия и вычисляется наблюдаемое значение d-статистики.

4. Принимается решение о возможности отклонения нулевой гипотезы.

Правило принятия решения для гипотез (16) (односторонний тест, рис. 1):

- если $d < d_L$, то нулевая гипотеза отвергается;
- если $d_L \leq d \leq d_U$ – тест не дает ответа;
- если $d > d_U$, то нулевая гипотеза не отвергается.

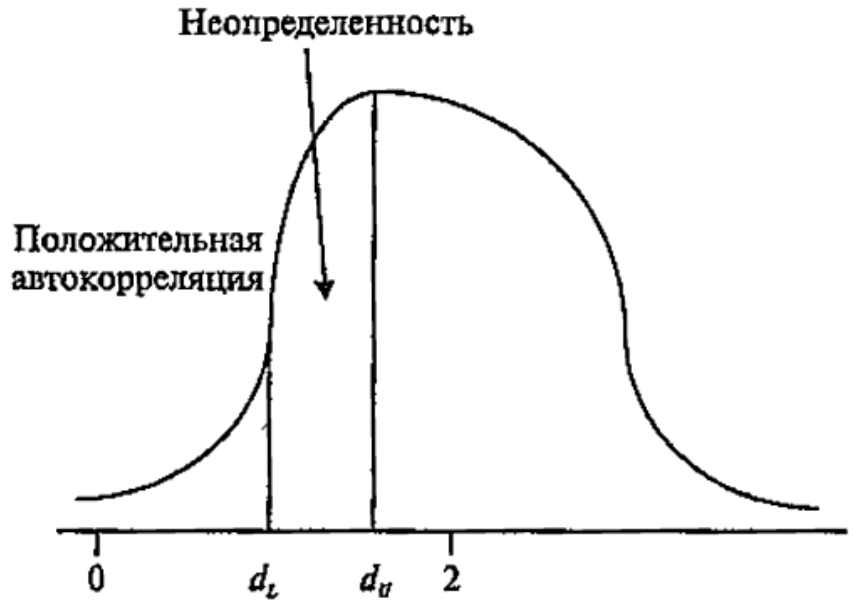

Рисунок 2.1 – Вид плоскости вероятности d-статистики и критические значения для одностороннего теста

Правило принятия решения для гипотез (17) (двусторонний тест, рис. 2):

- если $d < d_L$, то нулевая гипотеза отвергается;
- если $d_U < d < 4 - d_U$, то нулевая гипотеза не отвергается;
- если $d > 4 - d_U$, то нулевая гипотеза отвергается;
- в противном случае тест не дает ответа.

Рисунок 2.2 – Вид плоскости вероятности d-статистики и критические значения для двустороннего теста

Тест Дарбина-Уотсона не делает различия между чистой и нечистой автокорреляцией. Однако, если наблюдения позволяют, есть смысл упорядочить их по возрастанию значений одной из переменных. В этом случае значимая статистика Дарбина-Уотсона, указывающая на отрицательную автокорреляцию, является достаточно сильным признаком нечистой автокорреляции.

2.7.3 Обобщенный метод наименьших квадратов

Действия по устранению автокорреляции необходимо начать с проверки спецификации модели, поскольку всегда существует вероятность того, что обнаруженная автокорреляция является нечистой. Нечистая автокорреляция устраняется включением в уравнение пропущенной переменной или использованием правильной функциональной формы.

Эффективная оценка коэффициентов уравнения с чистой автокорреляцией ошибок может быть получена с помощью обобщенного метода наименьших квадратов. Иногда этот метод называют оценкой Айткена.

Если в уравнении

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_{k-1} X_{k-1t} + \varepsilon_t$$

случайный член ε подчиняется авторегрессионной схеме первого порядка, реализация GLS сводится к OLS-оценке коэффициентов квазиразностного уравнения:

$$Y_t^* = Y_t - \rho Y_{t-1}, X_{1t}^* = X_{1t} - \rho X_{1t-1}, \dots, X_{k-1t}^* = X_{k-1t} - \rho X_{k-1t-t},$$

$$\beta_0^* = \beta_0 - \rho\beta_0,$$

u_t – классический случайный член. Уравнение (18) иногда называют GLS-уравнением.

Недостатком этого метода является потеря в системе одного уравнения: в системе (18) на одно уравнение меньше, чем в (17). В результате при анализе выборок с небольшим числом наблюдений повышение эффективности при устранении автокорреляции может быть потеряно из-за снижения числа степеней свободы. Число степеней свободы можно сохранить, если в систему (18) добавить первое уравнение системы (17), умножив его левую и правую части на $(1-\rho)^{1/2}$ (поправка Прайса-Уинстена).

Выполнение GLS-оценки требует знания значения коэффициента авторегрессии ρ. Если этот коэффициент неизвестен, его заменяют оценкой $\hat{\rho}$. Простой способ ее получения заключается в использовании наблюдаемого значения d-статистики Дарбина-Уотсона. Из (15) следует

$$\hat{\rho} \approx 1 - \left(\frac{d}{2}\right).$$

Более точную оценку ρ и, следовательно, коэффициентов уравнения можно получить, воспользовавшись итерационным методом Кохрейна-Оркатта. Его выполнение можно разбить на следующие этапы.

Методом OLS оценивается модель (18) и находятся остатки.

Полученные остатки используются для нахождения OLS-оценки ρ как коэффициента наклона уравнения (12).

Полученные оценки коэффициентов используются для нахождения остатков (19).

Возврат на п.2.

Процедура повторяется до тех пор, пока оценка ρ не перестанет изменяться.

Не всегда высокая вероятность автокорреляции, полученная в тесте Дарбина-Уотсона, является поводом для проведения GLS-оценок.

Значимая d-статистика Дарбина-Уотсона может быть результатом нечистой автокорреляции. GLS уменьшит смещения оценок коэффициентов (по сравнению со смещениями OLS-оценок) только в случае, если пропущенная переменная значимо коррелирована хотя бы с одним из оставленных регрессоров. В случае неправильной спецификации и некоррелированности пропущенной переменной с оставленными OLS-оценки предпочтительнее.

Автокорреляция набора данных может быть уменьшена в результате агрегирования данных во времени. Обычно вероятность встретить автокорреляцию в среднегодовых наблюдениях и временных срезах ниже, чем в наблюдениях, усредненных за неделю, месяц, квартал. Иногда снижение числа степеней свободы в результате усреднения не приводит к потере информации, поскольку удается избавиться от «краткосрочного шума» (шума с малым временем корреляции). Однако следует помнить, что усреднение наблюдений приводит к потере эффективности: возрастанию дисперсий оценок коэффициентов.

Последствия автокорреляции могут быть незначительны. GLS-оценка эффективна только при условии, если оценка ρ совпадает с истинным значением. При оценивании ρ по малым выборкам значение практически всегда оказывается смещенным. Поскольку чистая автокорреляции не вызывает смещений оценок коэффициентов, оценка OLS может оказаться лучше, чем GLS-оценка, использующая смещенную $\hat{\rho}$. Примером может являться случай, когда OLS-оценки коэффициентов хорошо согласуются с теорией (являются

теоретически значимыми), в результате нет необходимости полагаться на t-статистику для сохранения тех или иных независимых переменных в уравнении.

3 ПОСТРОЕНИЕ РЕГРЕССИОННОЙ МОДЕЛИ ВАЛОВОГО ВНУТРЕННЕГО ПРОДУКТА

3.1 Понятие валового внутреннего продукта

Описать экономическое положение страны позволяет такой фактор как ВВП. При формировании бюджета основываются на уровне этого показателя, также, анализ его динамики показывает качество принятых управленческих решений, нацеленных на экономический рост [18].

Финансовый словарь дает следующее определение понятию ВВП: «Внутренний валовой продукт – это уровень совокупного производства в экономике в целом».

Существует два вида: валовой национальный и валовой внутренний продукт. Валовой национальный продукт — это рыночная стоимость всего объёма товаров и услуг, созданных производителями данной страны за один год как внутри своей страны, так и на территории других стран.

Валовой внутренний продукт — это стоимость всех созданных за год в стране товаров и услуг конечного потребления, оценённая в рыночных ценах.

В состав ВВП включаются/ не включаются (таблица 1).

Таблица 3.1 Состав ВВП.

Включаются	Не включаются
• конечные товары и услуги; • стоимость только тех товаров и услуг, которые были произведены для продажи, т. е. рыночная стоимость; • всё, что произведено внутри страны независимо от того, какой стране принадлежали факторы производства; • стоимость только тех товаров и услуг, которые были	• финансовые операции (покупка акций, облигаций, проценты по государственным облигациям); • трансферты (от лат. «переводить», «переносить») (пенсии, стипендии); • стоимость товаров и услуг, произведённых за пределами страны.

произведены в данном году.	

При производстве товаров и услуг в конечную стоимость продукта закладывают стоимость промежуточных товаров, которые необходимы для производства данного товара. Таким образом, в ВВП будет включатся только стоимость конечных товаров, которые готовы непосредственно к употреблению. Иначе, стоимость промежуточных товаров будет учитываться дважды.

Если показатель реального валового внутреннего продукта не увеличивается каждый год, то это еще не говорит об экономическом спаде. В данном вопросе допустимы циклические падения. Важнее оценивать направление движения показателя в целом, которое, для экономического роста, должно идти вверх.

При возникновении такого явления как инфляция цены могут значительно возрасти, а производство, наоборот, уменьшится. Чтобы избежать ошибочных суждений принято выполнять подсчет в так называемых постоянных ценах. Только таким образом получается реальный результат роста ВВП.

Экономический рост ведет к экономическому и социальному прогрессу. Он означает рост прибавочного продукта в стране, следовательно, рост прибыли — источника дальнейшего расширения и обновления производства и увеличения благосостояния населения.

ВВП может быть номинальным и реальным (рисунок 1).

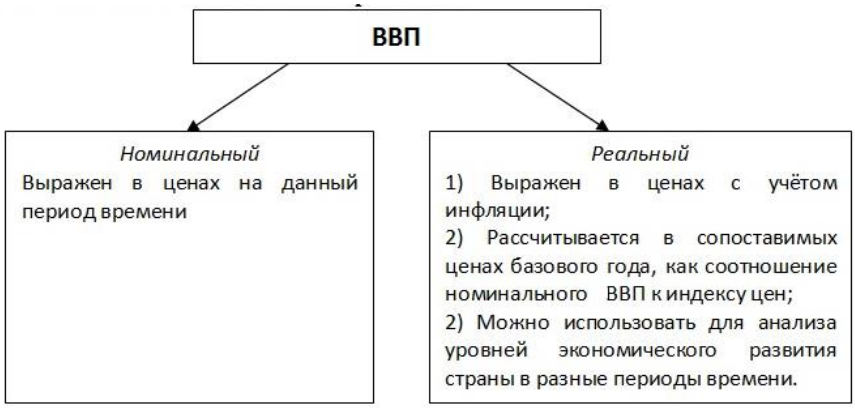

Рисунок 3.1 – Номинальный и реальный ВВП

В наших расчетах мы будем использовать реальное значение ВВП. При нем учитывается рост производства без учета финансовой составляющей. Он выражается в ценах любого года, взятого за основание. При его расчетах не учитывается инфляция, зато появляется возможность отследить, насколько серьезные изменения произошли в экономической ситуации за год [19].

Расчет ВВП производится несколькими способами:

1) по добавленной стоимости

2) по доходам = [национальный доход =общ прибыль компании + процентные платежи + арендная плата + з/п) + амортизация] – косвенные налоги – субсидии государства – чистый факторный доход.

3) по расходам = конечное потребление + инвестиционные капиталы + чистый экспорт + государственные расходы.

На ВВП влияют: инвестиционные средства, физический и человеческий капитал, вложение денег в инфраструктуру и основной капитал, расходы.

Чаще всего используется следующая формула: потребительские расходы + валовые инвестиции + государственные расходы + (экспорт – импорт).

3.2. Обзор существующих моделей валового внутреннего продукта

Для анализа взаимосвязи показателей, которые характеризуют социально-экономическое развитие России может быть использован такой математический инструментарий как корреляционно-регрессионный анализ. При использовании данного метода на первом этапе необходимо правильно специфицировать предполагаемую математическую модель, которая будет отражать взаимосвязь ряда показателей развития российской экономики.

В модели результативным признаком является показатель, анализ динамики которого интересен исследователю. В рамках данной задачи анализа развития экономики России целесообразно выбрать в качестве такого показателя объем ВВП в рыночных ценах, в миллиардах рублей

Факторы, оказывающие влияние на ВВП – это различные показатели социально-экономического развития России. В рамках исследования, описанного в статье «Эконометрический анализ статистической взаимосвязи показателей социально-экономического развития России» Гусаровой О.М., в качестве факторов были выбраны: численность экономически активного населения России (фактор X1, тысяч человек), численность занятых в экономике (фактор X2, тысяч человек), прибыль организаций (фактор X3, млрд руб.), численность научных организаций (фактор X4), среднемесячная зарплата (фактор X5, тысяч рублей). Результатом исследования является определение статистически значимого фактора – прибыль организаций [20].

В дополнение к существующим решениям авторами статьи «Эконометрический анализ валового внутреннего продукта России» была построена множественная регрессионная модель с тремя переменными: безработица, инвестиции, цены на нефть. Высокий уровень безработицы — это последствие экономического спада и признак социального неблагополучия: бедности, социальной напряженности, роста криминальной активности. Более того, в силу специфики структуры экономики и экспортного потенциала важнейшим экзогенным фактором экономической

динамики для России являются мировые цены на нефть: повышение мировых цен на нефть положительно влияет на динамику российской экономики как за счет роста спроса на результаты ее текущего функционирования, так и за счет повышения инвестиционной активности, напротив, снижение мировых цен на нефть практически неизбежно влечет за собой падение реального ВВП и объема инвестиций. Еще одну из важнейших наиболее изменчивых экономических категорий, определяющих развитие экономики, представляют собой инвестиции. Благодаря им осуществляется накопление общественного капитала, внедрение достижений науки и техники, вследствие чего создаётся база для расширения производственных возможностей страны и экономического роста. В результате исследования выявлена слабая взаимосвязь безработицы и ВВП. Можно предположить, что такая ситуация сложилась из-за значительных масштабов скрытой безработицы: испытывающие финансовые затруднения предприятия предпочитают сдерживать рост заработной платы и сокращать вакансии, но не работников, переводить работников на неполный рабочий день и отправлять в вынужденные отпуска. В исследовании было выяснено, что ВВП является неэластичным (коэффициент эластичности меньше единицы) показателем по отношению к ценам на нефть. Высокие цены на нефть непосредственно расширяют инвестиционные возможности бюджета и системообразующих российских компаний, работающих в нефтегазовом секторе. Если ожидания цены на нефть в будущем во многом формируются на основе текущего уровня цен, то при росте цен на нефть корректируются в сторону повышения оценки перспектив прибыли от проектов, не только реализуемых в энергосырьевом секторе и ориентированных на внешний спрос, но и ориентированных на удовлетворение внутреннего спроса в связи с ожидаемым ростом покупательной способности населения и смягчением финансовых ограничений для бизнеса и государственного сектора. Существование тесной причинно-следственной связи между инвестициями

и экономическим ростом общепризнанно. С одной стороны, инвестиции — главный «мотор» экономического роста: чем больше страна накопляет, тем выше темпы роста её экономики, поэтому государству следует поощрять сбережения и ограничивать потребление. С другой стороны — высокий спрос ведёт к росту производства, что заставляет инвестировать товаропроизводителей, а, значит, государству следует способствовать увеличению спроса, в том числе потребительского [21].

Автор статьи «Эконометрический анализ динамики макроиндикаторов экономики России в период с 2003 по 2016 г.» Балашова С. А. проводит анализ динамики основных макроэкономических показателей на основе выявления и идентификации сформировавшихся трендов. В статье рассмотрена динамика ВВП, инвестиций в основной капитал, объемов промышленного производства, объема экспорта и импорта, показателей потребительского рынка, таких как реальные доходы населения и заработанная плата [22].

Цвиль М. М. в своей статье «Эконометрический анализ валового внутреннего продукта на душу населения в Российской Федерации» в качестве результирующего фактора выбирает показатель ВВП, рассчитанный на душу населения. Он рассчитывается как результат деления ВВП на численность населения страны и показывает, какой объем ВВП в стоимостном выражении произведен за год на одного жителя данной страны. Экономические явления, как правило, определяются большим числом одновременно и совокупно действующих факторов. Задача исследования зависимости одной эндогенной переменной Y от нескольких объясняющих (экзогенных) переменных решается с помощью множественного регрессионного анализа. Для эконометрического анализа ВВП на душу населения авторы использовали данные Росстата за период 2004-2016 гг. В качестве эндогенной переменной выступает Y – ВВП на душу населения в РФ, а в качестве экзогенных: динамика инвестиций в основной капитал,

число занятых в экономике, число прибывших в РФ, число выбывших из РФ, номинально начисленная средняя заработанная плата, число безработных, экспорт РФ, реальная зарплата. В результате исследования авторы приходят к следующему выводу: «Стране необходимо встать на путь инновационного развития, чтобы быть конкурентоспособной на международном рынке и тем самым увеличить экспорт. Кроме того, необходимо обеспечить достойную заработанную плату, что приведет к росту производительности труда, будет способствовать сохранению интеллектуального потенциала как одного из элементов экономической безопасности государства» [23].

Министерство Финансов РФ прогнозирует рост ВВП в 2019 году на 1,3%, об этом в интервью федеральному телеканалу «Россия 24» заявил Глава Минэкономразвития Максим Орешкин. Также, он заметил, что факторами, оказывающими негативное влияние на рост ВВП в начале года, будут являться повышение ставки НДС, волатильность на сырьевых рынках и снижение спроса на экспортную российскую продукции.

Подведем итог анализа факторов на основе проведенных ранее исследований (таблица 3.2).

Таблица 3.2 – Факторы

Статья/ Факторы	Гусарова О.М.	Храмов А.В.	Цвиль М.М.	Озмитель К.В.
Число экономически активного населения (X_1)	+			
Число занятых в экономике (X_2)	+		+	
Прибыль организаций (X_3)	+ значим			
Число научных	+			

организаций (X_4)				
Среднемесячная з/п (X_5)	- сильная корреляция с x3, значим, но исключен		+ значим	
Уровень безработицы, в % от экономически активного населения (X_6)		- незначим		- слабое влияние
Прямые иностранные инвестиции млн. дол. (X_7)		+		
Цены на нефть марки Brent, долл/бар. (X_8)		+		
Динамика инвестиций в основной капитал, в%, (X_9)			+	
Число прибывших в РФ, (X_{10})			+	
Число выбывших, (X_{11})			+ значим	
Число			+	

безработных, (X_{12})				
Экспорт в млн дол, (X_{13})			+ значим	
Реальная з/п, % (темп роста к предыд.год), (X_{14})			+	
Общий объем инвестиций, млрд дол, (X_{15})				- исключен ввиду мульт-ти
Валовые нац. Сбережения млрд дол, (X_{16})				- исключен ввиду мульт-ти
Общий гос. доход млрд дол, (X_{17})				- исключен ввиду мульт-ти
Общий гос. расход млрд дол, (X_{18})				+ значим
Совокупный гос. долг, млрд дол, (X_{19})				- исключен ввиду мульт-ти
Сальдо платежного баланса, млрд				- слабое влияние

дол., (X_{20})				

В различные периоды экономического развития государства являются значимыми различные факторы. Попробуем построить модель актуальную на сегодняшний день. Для этого необходимо отобрать переменные и определить их влияние.

Выводы по главе.

Задача исследования зависимости одной эндогенной переменной Y – показателя ВВП от нескольких объясняющих (экзогенных) переменных решается с помощью множественного регрессионного анализа. Для эконометрического анализа ВВП авторы использовали различные факторы. Для построения современной актуальной модели необходимо произвести отбор наиболее значимых переменных и построить собственною модель.

4 ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Имеются статистические данные федеральной службы государственной статистики макроэкономического показателя, отражающего рыночную стоимость всех конечных товаров и услуг, произведенных за год во всех отраслях экономики на территории страны – ВВП. Данные за период с 2005 по 2017 г. приведены в приложении А [24-26].

Определим переменные:

- N– год;
- X_1 – число экономически активного населения, тыс. чел.;
- X_2 – число занятых в экономике, тыс. чел.;
- X_3 – прибыль организаций, млрд. руб. [28];
- X_4 – численность научных организаций;
- X_5 – номинальная среднемесячная заработанная плата, руб.;
- X_6 – уровень безработицы, в процентах [29];
- X_7 –инвестиции в основной капитал, млрд руб. [30];
- X_8 – потребительские расходы, млн. руб. [31];
- X_9 – общегосударственные доходы, млрд. руб. [32];
- X_{10} – реальная заработанная плата, темп роста к предыдущему году, в %;
- X_{11} – число выбывших из РФ, чел.;
- X_{12} – экспорт по всем странам мира, млн. дол. [33];
- X_{13} – общегосударственные расходы, млрд. руб. [33];
- Y – ВВП в рыночных ценах, млрд. руб.

Есть несколько подходов к понятию валовых инвестиций: первый подход выражается на макроэкономическом уровне. Валовые инвестиции – это общий объем инвестиций в экономике страны в совокупности. Второй подход рассматривает валовые инвестиции как вложения, которые направлены на поддержку и увеличение объемов основного капитала

предприятия и запасов. Третий подход подразумевает под валовыми инвестициями все суммарные вложения инвестора, которые осуществлялись в инвестиционный проект [33]. Выбираем второй подход.

Числовые характеристики (описательные статистики) рассматриваемых наблюдений приведены на рисунке 1.

Средний объем ВВП в выборке (Mean) составляет 57713,78 млрд. руб. Среднеквадратическое отклонение (Std. Dev.) - 24250,5. Минимальное значение показателя ВВП (Minimum) составляет 21609,8 млрд. руб. (2005 год). Максимальное значение показателя ВВП (Maximum) составляет 92089,3 (2017 год). Значение 60282,5 является медианной вариантой (Median).

4.1 Регрессионный анлиз показателя валового внутреннего продукта

Матрица парных коэффициентов корреляции исходных переменных приведена на рисунке 2.

Анализируя значения коэффициентов корреляции между зависимой и факторными переменными, которые расположены в первом столбце или первой строке корреляционной матрицы, можем проранжировать факторы по их линейной зависимости с регрессором. Таким образом, на ВВП в наибольшей степени влияют следующие факторы:

- X_5 – номинальная среднемесячная з\п – $r(Y, X_5) = 0.99$;
- X_9 – общегосударственные доходы – $r(Y, X_9) = 0.98$;
- X_{13} – общегосударственные расходы – $r(Y, X_{13}) = 0.98$;
- X_3 – прибыль организаций – $r(Y, X_3) = 0.94$;
- X_2 – число занятых в экономике – $r(Y, X_2) = 0.905$;
- X_{11} – число выбывших из РФ – $r(Y, X_{11}) = 0.87$;
- X_1 – число экономически активного населения – $r(Y, X_1) = 0.79$;
- X_6 – уровень безработицы – $r(Y, X_6) = -0.77$;

- X_{10} – реальная заработанная плата – $r(Y, X_{10}) = -0.54$.

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
Mean	57713.78	75561.91	70871.75	8867.385	3726.692	23988.73	6.230769	20.34377	373149.5	10619.57	105.5308	152032.1	373149.5	10893.19
Median	60282.50	75676.10	71003.10	8794.000	3622.000	23369.20	6.000000	18.66600	357817.0	11367.70	105.2000	69798.00	357817.0	10925.60
Maximum	92089.30	76858.00	72755.00	15823.00	4175.000	39167.00	8.300000	32.53900	527266.0	15088.90	117.2000	377155.0	527266.0	16420.30
Minimum	21609.80	73581.00	68339.00	3674.000	3492.000	8554.900	5.200000	6.853000	29190.00	5127.200	91.00000	32458.00	29190.00	3539.450
Std. Dev.	24250.50	850.6809	1322.252	3622.221	220.6144	10057.14	0.965561	7.642079	141978.1	3398.921	7.345677	137220.6	141978.1	4508.359
Skewness	-0.054487	-0.794200	-0.466391	0.522039	0.870536	-0.022751	0.750670	-0.045986	-0.943650	-0.213177	-0.296851	0.629547	-0.943650	-0.299699
Kurtosis	1.544749	3.675030	2.219082	2.283605	2.301809	1.755934	2.499114	1.932135	3.603752	1.563756	2.427420	1.658239	3.603752	1.790058
Jarque-Bera	1.153550	1.613451	0.801622	0.868465	1.906017	0.839460	1.356826	0.622264	2.126809	1.215811	0.368513	1.833889	2.126809	0.987586
Probability	0.561707	0.446317	0.669777	0.647762	0.385579	0.657224	0.507422	0.732617	0.345278	0.544490	0.831723	0.399739	0.345278	0.610307
Sum	750279.1	982304.8	921332.7	115276.0	48447.00	311853.5	81.00000	264.4690	4850944.	138054.4	1371.900	1976417.	4850944.	141611.4
Sum Sq. Dev.	7.06E+09	8683895.	20980193	1.57E+08	584048.8	1.21E+09	11.18769	700.8165	2.42E+11	1.39E+08	647.5077	2.26E+11	2.42E+11	2.44E+08
Observations	13	13	13	13	13	13	13	13	13	13	13	13	13	13

Рисунок 4.1 – Описательные статистики

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
Y	1.000000	0.790487	0.905131	0.935092	0.497093	0.992573	-0.768320	0.323753	0.016560	0.979807	-0.539247	0.871059	0.016560	0.980322
X1	0.790487	1.000000	0.867360	0.807000	0.606768	0.817300	-0.474415	0.355764	-0.109061	0.750663	-0.409538	0.578042	-0.109061	0.837420
X2	0.905131	0.867360	1.000000	0.878004	0.652307	0.889027	-0.846391	0.496691	0.030896	0.918227	-0.315057	0.757567	0.030896	0.872086
X3	0.935092	0.807000	0.878004	1.000000	0.660072	0.940246	-0.699430	0.362414	-0.162974	0.871294	-0.393998	0.875631	-0.162974	0.912856
X4	0.497093	0.606768	0.652307	0.660072	1.000000	0.494303	-0.522877	0.220605	-0.607259	0.441680	0.096038	0.636653	-0.607259	0.441082
X5	0.992573	0.817300	0.889027	0.940246	0.494303	1.000000	-0.713285	0.337794	-0.018571	0.957275	-0.564226	0.872649	-0.018571	0.989835
X6	-0.768320	-0.474415	-0.846391	-0.699430	-0.522877	-0.713285	1.000000	-0.500636	-0.164050	-0.835374	0.095611	-0.740811	-0.164050	-0.656654
X7	0.323753	0.355764	0.496691	0.362414	0.220605	0.337794	-0.500636	1.000000	0.441681	0.365641	0.108318	0.227137	0.441681	0.279704
X8	0.016560	-0.109061	0.030896	-0.162974	-0.607259	-0.018571	-0.164050	0.441681	1.000000	0.149427	-0.061278	-0.309892	1.000000	-0.006267
X9	0.979807	0.750663	0.918227	0.871294	0.441680	0.957275	-0.835374	0.365641	0.149427	1.000000	-0.514795	0.824299	0.149427	0.944720
X10	-0.539247	-0.409538	-0.315057	-0.393998	0.096038	-0.564226	0.095611	0.108318	-0.061278	-0.514795	1.000000	-0.417394	-0.061278	-0.633328
X11	0.871059	0.578042	0.757567	0.875631	0.636653	0.872649	-0.740811	0.227137	-0.309892	0.824299	-0.417394	1.000000	-0.309892	0.823326
X12	0.016560	-0.109061	0.030896	-0.162974	-0.607259	-0.018571	-0.164050	0.441681	1.000000	0.149427	-0.061278	-0.309892	1.000000	-0.006267
X13	0.980322	0.837420	0.872086	0.912856	0.441082	0.989835	-0.656654	0.279704	-0.006267	0.944720	-0.633328	0.823326	-0.006267	1.000000

Рисунок 4.2 – Матрица парных коэффициентов

Перечисленные переменные имеют прямое влияние на ВВП. Две последние пары – (Y, X_6) и (Y, X_{10}) имеют обратную зависимость: чем выше уровень безработицы, тем меньше объем ВВП, и чем выше реальная заработанная плата, тем меньше ВВП.

Анализуруя элементы матрицы межфакторной корреляции, которая получается вычеркиванием первой строки и первого столбца, можем заметить наличие существенной корреляции $(r \geq 0.8)$ между следующими переменными:

- $r(X_1, X_2) = 0.86$;
- $r(X_1, X_3) = 0.807$;
- $r(X_1, X_5) = 0.817$;
- $r(X_1, X_{13}) = 0.84$;
- $r(X_2, X_3) = 0.88$;
- $r(X_2, X_5) = 0.89$;
- $r(X_2, X_9) = 0.92$;
- $r(X_2, X_{13}) = 0.87$;
- $r(X_3, X_5) = 0.94$;
- $r(X_3, X_9) = 0.87$;
- $r(X_3, X_{10}) = 0.88$;
- $r(X_3, X_{13}) = 0.91$;
- $r(X_5, X_9) = 0.95$;
- $r(X_5, X_{11}) = 0.87$;
- $r(X_5, X_{13}) = 0.99$;
- $r(X_9, X_{11}) = 0.82$;
- $r(X_9, X_{13}) = 0.94$;
- $r(X_{11}, X_{13}) = 0.82$.

Таким образом, имеются пары коррелированных переменных, что позволяет сделать вывод о наличии мультиколлинеарности факторов.

Проверим также наличие мультиколлинеарности с помощью критерия χ^2.

Для преодоления явления мультиколлинеарности используется метод пошагового отбора переменных.

В данном методе результаты каждого шага учитываются на последующих шагах.

1. Выберем переменную, имеющую наибольший коэффициент корреляции.

2. Затем, необходимо перебрать все пары, в которых будет участвовать переменная, полученная на первом шаге. Пара, которая имеет наибольший коэффициент частой корреляции, очищенный от влияния переменной, полученной на первом шаге и будет той самой информативной парой.

Когда коэффициент корреляции будет уже очень близок к нулю и, когда величина R^2_{min} достигнет своего максимума процесс следует остановить.

4.2 Построение модели 1

Для практического применения данного метода используется пакет прикладных программ Matlab. Реализуем данный метод с возможностью выбора удаляемых переменных на основе экономических соображений (Приложение В).

Результат представлен на рисунке 4.3.

```
MODEL: 1 ------------------------------
Критерий Хи^2
Выбранная гипотеза: H1 - Есть мультиколлинеарность
x(1)x(2) = 0.86736
x(1)x(3) = 0.807
x(1)x(5) = 0.8173
x(1)x(13) = 0.83742
x(2)x(3) = 0.878
x(2)x(5) = 0.88903
x(2)x(6) = -0.84639
x(2)x(9) = 0.91823
x(2)x(13) = 0.87209
x(3)x(5) = 0.94025
x(3)x(9) = 0.87129
x(3)x(11) = 0.87563
x(3)x(13) = 0.91286
x(5)x(9) = 0.95728
x(5)x(11) = 0.87265
x(5)x(13) = 0.98983
x(6)x(9) = -0.83537
x(9)x(11) = 0.8243
x(9)x(13) = 0.94472
x(11)x(13) = 0.82333
Переменные с мультиколлинеарностью: x(1)x(2)x(3)x(5)x(6)x(9)x(11)x(13)
```

Рисунок 4.3 – Результат работы программы.

Удалим из модели переменные: $X_5, X_6, X_8, X_9, X_{11}, X_{13}$ (рисунок 4.4).

```
Переменные с мультиколлинеарностью: x(1)x(2)x(3)x(5)x(6)x(9)x(11)x(13)
Введите индекс удаляемой переменной x (0 - выход): 5
Введите индекс удаляемой переменной x (0 - выход): 6
Введите индекс удаляемой переменной x (0 - выход): 8
Введите индекс удаляемой переменной x (0 - выход): 9
Введите индекс удаляемой переменной x (0 - выход): 11
Введите индекс удаляемой переменной x (0 - выход): 13
Введите индекс удаляемой переменной x (0 - выход): 0
Выбранные переменные: 1   2   3   4   7   10   12
Количество наблюдений (n): 13 Количество переменных (p): 7
```

Рисунок 4.4 – Удаление переменных

Для оставшихся 7 переменных необходимо вычислить матрицу парных коэффициентов корреляции R_{xx} и R_{xy} (Таблица 4.1, Таблица 4.2).

Таблица 4.1 – Матрица парных коэффициентов корреляции R_{xx}

	X_1	X_2	X_3	X_4	X_7	X_{10}	X_{12}
X_1	1.0000	0.8674	0.8070	0.6068	0.3558	-0.4095	-0.1091
X_2	0.8674	1.0000	0.8780	0.6523	0.4967	-0.3151	0.0309
X_3	0.8070	0.8780	1.0000	0.6601	0.3624	-0.3940	-0.1630
X_4	0.6068	0.6523	0.6601	1.0000	0.2206	0.0960	-0.6073

X_7	0.3558	0.4967	0.3624	0.2206	1.0000	0.1083	0.4417
X_{10}	-0.4095	-0.3151	-0.3940	0.0960	0.1083	1.0000	-0.0613
X_{12}	-0.1091	0.0309	-0.1630	-0.6073	0.4417	-0.0613	1.0000

Таблица 4.2 – Матрица парных коэфициентов корреляции R_{yx}

	Y	X_1	X_2	X_3	X_4	X_7	X_{10}	X_{12}
Y	1.0000	0.7905	0.9051	0.9351	0.4971	0.3238	-0.5392	0.0166
X_1	0.7905	1.0000	0.8674	0.8070	0.6068	0.3558	-0.4095	-0.1091
X_2	0.9051	0.8674	1.0000	0.8780	0.6523	0.4967	-0.3151	0.0309
X_3	0.9351	0.8070	0.8780	1.0000	0.6601	0.3624	-0.3940	-0.1630
X_4	0.4971	0.6068	0.6523	0.6601	1.0000	0.2206	0.0960	-0.6073
X_7	0.3238	0.3558	0.4967	0.3624	0.2206	1.0000	0.1083	0.4417
X_{10}	-0.5392	-0.4095	-0.3151	-0.3940	0.0960	0.1083	1.0000	-0.0613
X_{12}	0.0166	-0.1091	0.0309 -	0.1630	-0.6073	0.4417	-0.0613	1.0000

Для построения модели воспользуемся обобщенным методом наименьших квадратов. Необходиые расчеты представлены в таблице 4.3.

$$\hat{Y} = -3.6754 - 0.0000X_1 + 0.0001X_2 + 0.0000X_3 - 0.0004X_4 - 0.0014X_7 - 0.0041X_{10} - 0.0000X_{12} + \varepsilon(1.0e{+}05\ *).$$

Проверим значимость полученного уравнения по F-критерию на 5% уровне значимости. Вычисление суммы квадратов дало следующие результаты: общая (Q_{sum}) – 7057041183.0631, обусловленная регрессией (Q_r) – 43301440607.4469, остаточная (Q_e) – 50358481790.51. Таким образом

$$F = \frac{Qr*(n-p-1)}{Qe*p} = 0.61419.$$

$F_{0.05:8:5}$ = 4.8183, $F<F_{0.05:8:5}$, следовательно уравнение незначимо.

Общие характеристики модели:

Коэфициент детерминации (R^2_{yx}): 0.96968, коэфициент детерминации скорректированный ($\widehat{R^{*2}_{yx}}$): 0.92724, наименьшее значение коэфициента детерминации (R^2_{min}): 0.91594.

Шаг 1. Выбор первого предиката (переменной).

В классе моделей регрессии *y* по единственной объясняющей переменной выбиратеся наиболее информативный (с максимальным коэффициентом детерминации) предиктор. Поскольку на первом шаге величина $R^2_{y,X}$ совпадает с квадратом обычного (парного) коэффициента корреляции $r(y,x)$, а $\max_{1\le j\le 12} r^2(y,x^{(j)}) = r^2\left(y,x^{(3)}\right) = 0.93509^2 = 0.8744$, то наиболее информативным предиктором в классе однофакторных (парных) регрессионных моделей оказывается переменная $x^{(3)}$ – прибыль организаций. Подсчет скорректированного (на несмещенность) значения $r^{*^2}\left(y,x^{(3)}\right) = \widehat{R^{*^2}}(1)$ и его нижней доверительной границы $R^{(1)}_{min}$дает следующие значения:

$$\widehat{R^{*^2}}(1) = 0{,}86298 \text{ и } R^2_{min} = 0{,}83674$$

Таблица 4.3 – Расчеты для модели 1 (1.0e+09 *).

Y_i	X_{i1}	X_{i2}	X_{i3}	X_{i4}	X_{i7}	X_{i10}	X_{i12}	$\widehat{Y}_i$	$Y_i - \widehat{Y}_i$	$(Y_i - \widehat{Y}_i)^2$	$(Y_i - Y_{\text{ср}})^2$	$(\widehat{Y}_i - Y_{\text{ср}})^2$
-0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0002	0	0.0000	0.4670	1.3035	3.3309
0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0004	0	0.0000	0.7245	0.9484	3.3309
0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0004	0	0.0000	1.1054	0.5986	3.3309
0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0005	0	0.0000	1.7038	0.2702	3.3309
0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0003	0	0.0000	1.5060	0.3575	3.3309
0.0000	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0004	0	0.0000	2.1445	0.1301	3.3309
0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0005	0	0.0001	3.6340	0.0066	3.3309
0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0005	0	0.0001	4.6463	0.1092	3.3309
0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0005	0	0.0001	5.3486	0.2378	3.3309
0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0005	0	0.0001	6.2726	0.4616	3.3309
0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0000	0	0.0001	6.9277	0.6512	3.3309
0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0003	0	0.0001	7.3978	0.8007	3.3309
0.0001	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000	0.0004	0	0.0001	8.4804	1.1817	3.3309

Шаг 2.

Среди всех возможных пар объясняющих переменных $(x^{(3)}, x^{(j)}), j = 1,2,4,7,10,12$, выбирается наиболее информативная пара предикторов. Для этого необходимо провести соответствующие расчеты:

Проверяемые переменные: $x^{(1)}$ и $x^{(3)}$. Парные коэффициенты корреляции представлены в таблице 4.4.

Таблица 4.4 – Парные коэффициенты кореляции

	y	$x^{(1)}$	$x^{(3)}$
y	1.0000	0.7905	0.9351
$x^{(1)}$	0.7905	1.0000	0.8070
$x^{(3)}$	0.9351	0.8070	1.0000

Коэффициент детерминации $\widehat{R^2_{y,(x^{(1)},x^{(3)})}} = 0.87809.$

Проверяемые переменные: $x^{(2)}$ и $x^{(3)}$. Парные коэффициенты корреляции представлены в таблице 4.5.

Таблица 4.5 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(3)}$
y	1.0000	0.9051	0.9351
$x^{(2)}$	0.9051	1.0000	0.8780
$x^{(3)}$	0.9351	0.8780	1.0000

Коэффициент детерминации $\widehat{R^2_{y,(x^{(2)},x^{(3)})}} = 0{,}90528.$

Проверяемые переменные: $x^{(3)}$ и $x^{(4)}$. Парные коэффициенты корреляции представлены в таблице 4.6.

Таблица 4.6 – Парные коэффициенты кореляции

	y	$x^{(3)}$	$x^{(4)}$
y	1.0000	0.9351	0.4971
$x^{(3)}$	0.9351	1.0000	0.6601
$x^{(4)}$	0.4971	0.6601	1.0000

Коэффициент детерминации $\widehat{R^2_{y,(x^{(3)},x^{(4)})}} = 0{,}89997.$

Проверяемые переменные: $x^{(3)}$ и $x^{(7)}$. Парные коэффициенты корреляции представлены в таблице 4.7.

Таблица 4.7 – Парные коэффициенты кореляции

	y	$x^{(3)}$	$x^{(7)}$
y	1.0000	0.9351	0.3238
$x^{(3)}$	0.9351	1.0000	0.3624
$x^{(7)}$	0.3238	0.3624	1.0000

Коэффициент детерминации $\widehat{R^2_{y,(x^{(3)},x^{(7)})}} = 0{,}87466.$

Проверяемые переменные: $x^{(3)}$ и $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.8.

Таблица 4.8 – Парные коэффициенты кореляции

	y	$x^{(3)}$	$x^{(10)}$
y	1.0000	0.9351	-0.5392
$x^{(3)}$	0.9351	1.0000	-0.3940
$x^{(10)}$	-0.5392	-0.3940	1.0000

Коэффициент детерминации $\widehat{R^2_{y,(x^{(3)},x^{(10)})}} = 0{,}90894.$

Проверяемые переменные: $x^{(3)}$ и $x^{(12)}$. Парные коэффициенты корреляции представлены в таблице 4.9.

Таблица 4.9 – Парные коэффициенты кореляции

	y	$x^{(3)}$	$x^{(12)}$
y	1.0000	0.9351	0.0166
$x^{(3)}$	0.9351	1.0000	-0.1630
$x^{(12)}$	0.0166	-0.1630	1.0000

Коэффициент детерминации $\widehat{R^2_{y,(x^{(3)},x^{(12)})}} = 0{,}90372.$

Поскольку

$$\max_{\substack{1\le j\le 12\\(j\neq 3)}} \widehat{R^2_{y,(x^{(3)},x^{(J)})}} = \widehat{R^2_{y,(x^{(3)},x^{(10)})}} = 0.90894,$$

то наиболее информативной парой предикторов оказываются объясняющие переменные: $x^{(3)}$ – прибыль организаций и $x^{(10)}$ – реальная заработанная плата. Подсчет нижней доверительной границы $R^{(1)}_{min}$дает следующие значения:

$$R^2_{min} = 0{,}86507.$$

Целесообразность включения в модель в качестве второй объясняющей переменной предиктора $x^{(10)}$ подтверждается сравнением нижних доверительных границ $R^2_{min}(1)$ и $R^2_{min}(2)$, т. к. $R^2_{min}(1) < R^2_{min}(2)$.

Шаг 3.

Проверяемые переменные: $x^{(1)}$, $x^{(3)}$ и $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.10.

Таблица 4.10 – Парные коэффициенты кореляции

	y	$x^{(1)}$	$x^{(3)}$	$x^{(10)}$
y	1.0000	0.7905	0.9351	-0.5392
$x^{(1)}$	0.7905	1.0000	0.8070	-0.4095
$x^{(3)}$	0.9351	0.8070	1.0000	-0.3940
$x^{(10)}$	-0.5392	-0.4095	-0.3940	1.0000

Коэффициент детерминации $\widehat{R^2_{y,(x^{(1)},x^{(3)},x^{(10)})}} = 0{,}90983.$

Проверяемые переменные: $x^{(2)}$, $x^{(3)}$и $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.11.

Таблица 4.11 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(3)}$	$x^{(10)}$
y	1.0000	0.9051	0.9351	-0.5392
$x^{(2)}$	0.9051	1.0000	0.8780	-0.3151
$x^{(3)}$	0.9351	0.8780	1.0000	-0.3940
$x^{(10)}$	-0.5392	-0.3151	-0.3940	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(2)},x^{(3)},x^{(10)})}} = 0{,}94475.$

Проверяемые переменные: $x^{(3)}$, $x^{(4)}$ и $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.12.

Таблица 4.12 – Парные коэффициенты кореляции

	y	$x^{(3)}$	$x^{(4)}$	$x^{(10)}$
y	1.0000	0.9351	0.4971	-0.5392
$x^{(3)}$	0.9351	1.0000	0.6601	-0.3940
$x^{(4)}$	0.4971	0.6601	1.0000	0.0960
x^{10}	-0.5392	-0.3940	0.0960	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(3)},x^{(4)},x^{(10)})}} = 0{,}91453.$

Проверяемые переменные: $x^{(3)}$, $x^{(7)}$ и $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.13.

Таблица 4.13 – Парные коэффициенты кореляции

	y	$x^{(3)}$	$x^{(7)}$	$x^{(10)}$
y	1.0000	0.9351	0.3238	-0.5392
$x^{(3)}$	0.9351	1.0000	0.3624	-0.3940
$x^{(7)}$	0.3238	0.3624	1.0000	0.1083
$x^{(10)}$	-0.5392	-0.3940	0.1083	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(3)},x^{(7)},x^{(10)})}} = 0{,}91054.$

Проверяемые переменные: $x^{(3)}$ и $x^{(10)}$, $x^{(12)}$. Парные коэффициенты корреляции представлены в таблице 4.14.

Таблица 4.14 – Парные коэффициенты кореляции

	y	$x^{(3)}$	$x^{(10)}$	$x^{(12)}$
y	1.0000	0.9351	-0.5392	0.0166
$x^{(3)}$	0.9351	1.0000	-0.3940	-0.1630
$x^{(10)}$	-0.5392	-0.3940	1.0000	-0.0613
$x^{(12)}$	0.0166	-0.1630	-0.0613	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(3)},x^{(10)},x^{(12)})}} = 0{,}93053.$

Среди всевозможных троек объясняющих переменных $(x^{(3)}, x^{(10)}, x^{(j)}), j = 1{,}2{,}4{,}7{,}12,$ наиболее информативной оказалась тройка $(x^{(3)}, x^{(10)}, x^{(2)})$, поскольку

$$\max_{\substack{1\le j\le 12 \\ (j\neq 3{,}10)}} \widehat{R^2_{y,(x^{(3)},x^{(10)},x^{(j)})}} = \widehat{R^2_{y,(x^{(3)},x^{(10)},x^{(2)})}} = 0{,}94475,$$

то наиболее информативной парой предикторов оказываются объясняющие переменные: $x^{(3)}$ – прибыль организаций и $x^{(10)}$ – реальная заработанная плата, $x^{(2)}$ – число занятых в экономике. Подсчет нижней доверительной границы $R^{(1)}_{min}$дает следующие значения:

$$R^2_{min} = 0.90825$$

Целесообразность включения в модель в качестве третьей объясняющей переменной предиктора $x^{(2)}$ подтверждается сравнением нижних доверительных границ $R^2_{min}(2)$ и $R^2_{min}(3)$, т. к. $R^2_{min}(2) < R^2_{min}(3)$.

Шаг 4.

Проверяемые переменные: $x^{(1)}$, $x^{(2)}$, $x^{(3)}$, $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.15

Таблица 4.15 – Парные коэффициенты кореляции

	y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(10)}$
y	1.0000	0.7905	0.9051	0.9351	-0.5392
$x^{(1)}$	0.7905	1.0000	0.8674	0.8070	-0.4095
$x^{(2)}$	0.9051	0.8674	1.0000	0.8780	-0.3151
$x^{(3)}$	0.9351	0.8070	0.8780	1.0000	-0.3940
$x^{(10)}$	-0.5392	-0.4095	-0.3151	-0.3940	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(1)},x^{(2)},x^{(3)},x^{(10)})}} = 0{,}9546.$

Проверяемые переменные: $x^{(2)}$, $x^{(3)}$, $x^{(4)}$ и $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.16.

Таблица 4.16 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(10)}$
y	1.0000	0.9051	0.9351	0.4971	-0.5392
$x^{(2)}$	0.9051	1.0000	0.8780	0.6523	-0.3151
$x^{(3)}$	0.9351	0.8780	1.0000	0.6601	-0.3940
$x^{(4)}$	0.4971	0.6523	0.6601	1.0000	0.0960
$x^{(10)}$	-0.5392	-0.3151	-0.3940	0,0960	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(2)},\, x^{(3)},\, x^{(4)},\, x^{(10)})}} = 0{,}95769.$

Проверяемые переменные: $x^{(2)}$, $x^{(3)}$, $x^{(7)}$ и $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.17.

Таблица 4.17 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(3)}$	$x^{(7)}$	$x^{(10)}$
y	1.0000	0.9051	0.9351	0.3238	-0.5392
$x^{(2)}$	0.9051	1.0000	0.8780	0.4967	-0.3151
$x^{(3)}$	0.9351	0.8780	1.0000	0.3624	-0.3940
$x^{(7)}$	0.3238	0.4967	0.3624	1.0000	0.1083
$x^{(10)}$	-0.5392	-0.3151	-0.3940	0.1083	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(2)},\,x^{(3)},\,x^{(7)},\,x^{(10)})}} = 0{,}94623.$

Проверяемые переменные: $x^{(2)}$, $x^{(3)}$, $x^{(10)}$ и $x^{(12)}$. Парные коэффициенты корреляции представлены в таблице 4.18.

Таблица 4.18– Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(3)}$	$x^{(10)}$	$x^{(12)}$
y	1.0000	0.9051	0.9351	-0.5392	0.0166
$x^{(2)}$	0.9051	1.0000	0.8780	-0.3151	0.0309
$x^{(3)}$	0.9351	0.8780	1.0000	-0.3940	-0.1630
$x^{(10)}$	-0.5392	-0.3151	- 0.3940	1.0000	-0.0613
$x^{(12)}$	0.0166	0.0309	-0.1630	-0.0613	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(2)},\,x^{(3)},\,x^{(10)},\,x^{(12)})}} = 0{,}95126.$

Среди всевозможных четверок объясняющих переменных $(x^{(2)}, x^{(3)}, x^{(10)}, x^{(j)}), j = 1{,}4{,}7{,}12,$ наиболее информативной оказалась четверка $(x^{(2)}, x^{(3)}, x^{(4)}, x^{(10)})$, поскольку

$$\max_{\substack{1\le j\le 12\\ (j\neq 2,3,10)}} \widehat{R^2_{y,(x^{(3)},x^{(4)},x^{(10)},x^{(j)})}} = \widehat{R^2_{y,(,x^{(2)},x^{(3)},x^{(4)},x^{(10)})}} = 0{,}95769,$$

то наиболее информативной парой предикторов оказываются объясняющие переменные: $x^{(3)}$ – прибыль организаций и $x^{(10)}$ – реальная заработанная плата, $x^{(2)}$ – число занятых в экономике, $x^{(4)}$ – численность научных организаций. Подсчет нижней доверительной границы $R^{(1)}_{min}$дает следующие значения:

$$R^2_{min} = 0.92146$$

Целесообразность включения в модель в качестве четвертой объясняющей переменной предиктора $x^{(4)}$ подтверждается сравнением нижних доверительных границ $R^2_{min}(3)$ и $R^2_{min}(4)$, т. к. $R^2_{min}(3) < R^2_{min}(4)$.

Шаг 5 - Выбор наилучшей переменной

Проверяемые переменные: $x^{(1)}$, $x^{(2)}$, $x^{(3)}$, $x^{(4)}$, $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.19.

Таблица 4.19 – Парные коэффициенты кореляции

	y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(10)}$
y	1.0000	0.7905	0.9051	0.9351	0.4971	-0.5392
$x^{(1)}$	0.7905	1.0000	0.8674	0.8070	0.6068	-0.4095
$x^{(2)}$	0.9051	0.8674	1.0000	0.8780	0.6523	-0.3151
$x^{(3)}$	0.9351	0.8070	0.8780	1.0000	0.6601	-0.3940
$x^{(4)}$	0.4971	0.6068	0.6523	0.6601	1.0000	0.0960
$x^{(10)}$	-0.5392	-0.4095	-0.3151	-0.3940	0.0960	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(1)},x^{(2)},x^{(3)},\ x^{(4)},x^{(10)})}} = 0{,}96325.$

Проверяемые переменные: $x^{(2)}$, $x^{(3)}$, $x^{(4)}$, $x^{(7)}$, $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.20.

Таблица 4.20 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(7)}$	$x^{(10)}$
y	1.0000	0.9051	0.9351	0.4971	0.3238	-0.5392
$x^{(2)}$	0.9051	1.0000	0.8780	0.6523	0.4967	-0.3151

$x^{(3)}$	0.9351	0.8780	1.0000	0.6601	0.3624	-0.3940
$x^{(4)}$	0.4971	0.6523	0.6601	1.0000	0.2206	0.0960
$x^{(7)}$	0.3238	0.4967	0.3624	0.2206	1.0000	0.1083
$x^{(10)}$	-0.5392	-0.3151	-0.3940	0.0960	0.1083	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(\,x^{(2)},x^{(3)},\,x^{(4)},\,\,x^{(7)},\,x^{(10)}\,)}} = 0{,}96411.$

Проверяемые переменные: $x^{(2)}$, $x^{(3)}$, $x^{(4)}$, $x^{(10)}$, $x^{(12)}$. Парные коэффициенты корреляции представлены в таблице 4.21.

Таблица 4.21 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(10)}$	$x^{(12)}$
y	1.0000	0.9051	0.9351	0.4971	-0.5392	0.0166
$x^{(2)}$	0.9051	1.0000	0.8780	0.6523	-0.3151	0.0309
$x^{(3)}$	0.9351	0.8780	1.0000	0.6601	-0.3940	-0.1630
$x^{(4)}$	0.4971	0.6523	0.6601	1.0000	0.0960	-0.6073
$x^{(10)}$	-0.5392	-0.3151	-0.3940	0.0960	1.0000	-0.0613
$x^{(12)}$	0.0166	0.0309	-0.1630	-0.6073	-0.0613	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(\,x^{(2)},x^{(3)},\,x^{(4)},\,\,x^{(10)},\,x^{(12)}\,)}} = 0{,}95897.$

Среди всевозможных пятерок объясняющих переменных $(x^{(2)}, x^{(3)}, x^{(4)}, x^{(10)}, x^{(j)}), j = 1{,}7{,}12,$ наиболее информативной оказалась пятерка $(\,x^{(2)}, x^{(3)}, x^{(4)}, x^{(7)}, x^{(10)})$, поскольку

$$\max_{\substack{1\le j\le 12\\ (j\neq 2,3,4,10)}} \widehat{R^2_{y,(x^{(2)},x^{(3)},x^{(4)},x^{(10)},x^{(j)})}} = \widehat{R^2_{y,(,x^{(2)},x^{(3)},x^{(4)},x^{(7)},x^{(10)})}} = 0{,}96411,$$

то наиболее информативной парой предикторов оказываются объясняющие переменные: $x^{(3)}$ – прибыль организаций и $x^{(10)}$ – реальная заработанная плата, $x^{(2)}$ – число занятых в экономике, $x^{(4)}$ – численность научных организаций, $x^{(7)}$ – инвестиции в основной капитал. Подсчет нижней доверительной границы $R^{(1)}_{min}$дает следующие значения:

$$R^2_{min} = 0.9251$$

Целесообразность включения в модель в качестве пятой объясняющей переменной предиктора $x^{(7)}$ подтверждается сравнением нижних доверительных границ $R^2_{min}(4)$ и $R^2_{min}(5)$, т. к. $R^2_{min}(4) < R^2_{min}(5)$.

Шаг 6 - Выбор наилучшей переменной.

Проверяемые переменные: $x^{(1)}$, $x^{(2)}$, $x^{(3)}$, $x^{(4)}$, $x^{(7)}$, $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице 4.22.

Таблица 4.22 – Парные коэффициенты кореляции

	y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(7)}$	$x^{(10)}$
y	1.0000	0.7905	0.9051	0.9351	0.4971	0.3238	-0.5392
$x^{(1)}$	0.7905	1.0000	0.8674	0.8070	0.6068	0.3558	-0.4095
$x^{(2)}$	0.9051	0.8674	1.0000	0.8780	0.6523	0.4967	-0.3151
$x^{(3)}$	0.9351	0.8070	0.8780	1.0000	0.6601	0.3624	-0.3940
$x^{(4)}$	0.4971	0.6068	0.6523	0.6601	1.0000	0.2206	0.0960
$x^{(7)}$	0.3238	0.3558	0.4967	0.3624	0.2206	1.0000	0.1083
$x^{(10)}$	-0.5392	-0.4095	-0.3151	-0.3940	0.0960	0.1083	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(1)},x^{(2)},x^{(3)},x^{(4)},x^{(7)},x^{(10)})}} = 0{,}96968$.

Проверяемые переменные: $x^{(2)}$, $x^{(3)}$, $x^{(4)}$, $x^{(7)}$, $x^{(10)}$, $x^{(12)}$. Парные коэффициенты корреляции представлены в таблице 4.23.

Таблица 4.23 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(7)}$	$x^{(10)}$	$x^{(12)}$
y	1.0000	0.9051	0.9351	0.4971	0.3238	-0.5392	0.0166
$x^{(2)}$	0.9051	1.0000	0.8780	0.6523	0.4967	-0.3151	0.0309
$x^{(3)}$	0.9351	0.8780	1.0000	0.6601	0.3624	-0.3940	-0.1630
$x^{(4)}$	0.4971	0.6523	0.6601	1.0000	0.2206	0.0960	-0.6073
$x^{(7)}$	0.3238	0.4967	0.3624	0.2206	1.0000	0.1083	0.4417

$x^{(10)}$	-0.5392	-0.3151	-0.3940	0.0960	0.1083	1.0000	-0.0613
$x^{(12)}$	0.0166	0.0309	-0.1630	-0.6073	0.4417	-0.0613	1.0000

Коэффициент детерминации: $R^2_{y,(\widehat{x^{(2)},x^{(3)},x^{(4)},\ x^{(7)},x^{(10)},x^{(12)}})} = 0{,}96424.$

Среди всевозможных наборов объясняющих переменных $(x^{(2)}, x^{(3)}, x^{(4)}, x^{(7)}\ x^{(10)}, x^{(j)}), j = 1{,}12,$ наиболее информативной оказался набор $(x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)}, x^{(7)}, x^{(10)})$, поскольку

$\max\limits_{\substack{1\le j\le 12 \\ (j\neq 2,3,4,7,10)}} R^2_{y,(\widehat{x^{(2)},x^{(3)},x^{(4)},x^{(7)},x^{(10)},x^{(j)}})} = R^2_{y,(\widehat{x^{(1)},x^{(2)},x^{(3)},x^{(4)},x^{(7)},x^{(10)}})} = 0{,}96968,$

то наиболее информативной парой предикторов оказываются объясняющие переменные: $x^{(3)}$ – прибыль организаций и $x^{(10)}$ – реальная заработанная плата, $x^{(2)}$ – число занятых в экономике, $x^{(4)}$ – численность научных организаций, $x^{(7)}$ – инвестиции в основной капитал, $x^{(1)}$ – число экономически активного населения. Подсчет нижней доверительной границы $R^{(1)}_{min}$дает следующие значения:

$$R^2_{min} = 0.9279$$

Целесообразность включения в модель в качестве шестой объясняющей переменной предиктора $x^{(1)}$ подтверждается сравнением нижних доверительных границ $R^2_{min}(5)$ и $R^2_{min}(6)$, т. к. $R^2_{min}(5) < R^2_{min}(6)$.

Шаг 7 - Выбор наилучшей переменной.

Проверяемые переменные: $x^{(1)}$, $x^{(2)}$, $x^{(3)}$, $x^{(4)}$, $x^{(7)}$, $x^{(10)}$, $x^{(12)}$. Парные коэффициенты корреляции представлены в таблице 4.24.

Таблица 4.24 – Парные коэффициенты кореляции

	y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(7)}$	$x^{(10)}$	$x^{(12)}$
y	1.0000	0.7905	0.9051	0.9351	0.4971	0.3238	-0.539	0.0166
$x^{(1)}$	0.7905	1.0000	0.8674	0.8070	0.6068	0.3558	-0.410	-0.1091
$x^{(2)}$	0.9051	0.8674	1.0000	0.8780	0.6523	0.4967	-0.315	0.0309

$x^{(3)}$	0.9351	0.8070	0.8780	1.0000	0.6601	0.3624	-0.394	-0.1630
$x^{(4)}$	0.4971	0.6068	0.6523	0.6601	1.0000	0.2206	0.0960	-0.6073
$x^{(7)}$	0.3238	0.3558	0.4967	0.3624	0.2206	1.0000	0.1083	0.4417
$x^{(10)}$	-0.5392	-0.4095	-0.3151	-0.3940	0.0960	0.1083	1.0000	-0.0613
$x^{(12)}$	0.0166	-0.1091	0.0309	-0.1630	-0.6073	0.4417	-0.061	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(1)},x^{(2)},x^{(3)},x^{(4)},x^{(7)},x^{(10)},x^{(12)})}} = 0{,}96968.$

Подсчет нижней доверительной границы $R^{(1)}_{min}$ дает следующие значения:

$$R^2_{min} = 0.91594$$

Нецелесообразность включения в модель в качестве седьмой объясняющей переменной предиктора $x^{(12)}$ подтверждается сравнением нижних доверительных границ $R^2_{min}(6)$ и $R^2_{min}(7)$, т. к. $R^2_{min}(6) > R^2_{min}(7)$.

Таким образом, наиболее информативной парой предикторов оказываются объясняющие переменные: $x^{(1)}$ – число экономически активного населения, $x^{(2)}$ – число занятых в экономике, $x^{(3)}$ – прибыль организаций, $x^{(4)}$ – численность научных организаций, $x^{(7)}$ – инвестиции в основной капитал и $x^{(10)}$ – реальная заработанная плата.

4.3 Построение модели 2

Проверим наличие мультиколлинеарности в новой модели с помощью критерия χ^2.

Мультиколлинеарность будет считаться доказанной, если гипотеза H_0 о независимости переменных, т.е. $Det|R| = 1$ будет отклонена. Учитывая, что величина $\left[n - 1 - \frac{1}{6}(2n + 5)lgDetR\right]$ имеет приближенное распределение χ^2 с $df = \frac{1}{2}p(p-1)$ степенями свободы. Если фактическое значение χ^2 превосходит табличное (критическое) $\chi^2_{\text{факт}} > \chi^2_{\text{табл}(df,a)}$, то гипотеза H_0 отклоняется.

$$\chi^2_{\text{факт}} = \left[n - 1 - \frac{1}{6}(2n + 5)lgDetR\right] = 45.5220.$$

$$\chi^2_{\text{табл}} = 24.9958.$$

$\chi^2_{\text{факт}} > \chi^2_{\text{табл}(df,a)}$, следовательно, гипотеза H_0 отклоняется.

Между оставшимися переменными по-прежнему присутствует явление мультиколлинеарности. Переменные с мультиколлинеарностью:

- $(X_1, X_2) = 0.86$;
- $(X_1, X_3) = 0.807$;
- $(X_2, X_3) = 0.878$.

Удалим переменную X_3.

Для оставшихся 5 $(X_1, X_2, X_4, X_7, X_{10})$ переменных необходимо вычислить матрицу парных коэффициентов корреляции R_{xx} и R_{yx} (Таблица 4.25, Таблица 4.26.).

Таблица 4.25 – Матрица парных коэффициентов корреляции R_{xx}

1.0000	0.8674	0.6068	0.3558	-0.4095
0.8674	1.0000	0.6523	0.4967	-0.3151
0.6068	0.6523	1.0000	0.2206	0.0960
0.3558	0.4967	0.2206	1.0000	0.1083
-0.4095	-0.3151	0.0960	0.1083	1.0000

Таблица 4.26 – Матрица парных коэфициентов корреляции R_{yx}

1.0000	0.7905	0.9051	0.4971	0.3238	-0.5392
0.7905	1.0000	0.8674	0.6068	0.3558	-0.4095
0.9051	0.8674	1.0000	0.6523	0.4967	-0.3151
0.4971	0.6068	0.6523	1.0000	0.2206	0.0960
0.3238	0.3558	0.4967	0.2206	1.0000	0.1083
-0.5392	-0.4095	-0.3151	0.0960	0.1083	1.0000

Для построения модели воспользуемся методом наименьших квадратов. Необходиые расчеты представлены в таблице .

$$\hat{Y} = -8.2683 - 0.0000X_1 + 0.0002X_2 - 0.0000X_4 - 0.0029X_7 - 0.0089X_{10} + \varepsilon(1.0e{+}05\ *).$$

Проверим значимость полученного уравнения по F-критерию на 5% уровне значимости. Вычисление суммы квадратов дало следующие результаты: общая (Q_{sum}) – 7057041183.0631, обусловленная регрессией (Q_r) – 43301440607.4469, остаточная (Q_e) – 50358481790.51. Таким образом

$$F = \frac{Qr*(n-p-1)}{Qe*p} = 1.2038.$$

$F_{0.05:6:7}$ = 3.866, F<$F_{0.05:8:5}$, следовательно уравнение незначимо.

Общие характеристики модели:

Коэффициент детерминации (R^2_{yx}): 0.90074, коэффициент детерминации скорректированный ($\widehat{R^2_{yx}}$): 0.82984, наименьшее значение коэффициента детерминации ($\widehat{R^2_{yx}}$): 0.79285.

Шаг 1 - Выбор первого предиката (переменной).

В классе моделей регрессии y по единственной объясняющей переменной выбиратеся наиболее информативный (с максимальным коэффициентом детерминации) предиктор. Поскольку на первом шаге величина $R^2_{y,X}$ совпадает с квадратом обычного (парного) коэффициента корреляции $r(y,x)$, а $\max_{1\le j\le 12} r^2(y,x^{(j)}) = r^2\left(y,x^{(2)}\right) = 0.90513^2 = 0.81926$, то наиболее информативным предиктором в классе однофакторных (парных) регрессионных моделей оказывается переменная $x^{(2)}$ – прибыль организаций. Подсчет скорректированного (на несмещенность) значения $r^{*^2}\left(y,x^{(2)}\right) = \widehat{R^{*^2}}(1)$ и его нижней доверительной границы $R^{(1)}_{min}$дает следующие значения:

$$\widehat{R^{*^2}}(1) = 0{,}80283 \text{ и } R^2_{min} = 0{,}76507$$

Шаг 2.

Среди всех возможных пар объясняющих переменных $(x^{(2)},x^{(j)}), j = 1,4,7,10$, выбирается наиболее информативная пара предикторов. Для этого необходимо провести соответствующие расчеты:

Проверяемые переменные: $x^{(1)}$ и $x^{(2)}$. Парные коэффициенты корреляции представлены в таблице:

Таблица 4.27 – Парные коэффициенты кореляции

	y	$x^{(1)}$	$x^{(2)}$
y	1.0000	0.7905	0.9051
$x^{(1)}$	0.7905	1.0000	0.8674
$x^{(2)}$	0.9051	0.8674	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(1)},x^{(2)})}} = 0{,}81938$.

Проверяемые переменные: $x^{(2)}$, $x^{(4)}$. Парные коэффициенты корреляции представлены в таблице:

Таблица 4.28 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(4)}$
y	1.0000	0.9051	0.4971
$x^{(2)}$	0.9051	1.0000	0.6523
$x^{(4)}$	0.4971	0.6523	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(2)},x^{(4)})}} = 0{,}83442.$

Проверяемые переменные: $x^{(2)}$, $x^{(7)}$. Парные коэффициенты корреляции представлены в таблице:

Таблица 4.29 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(7)}$
y	1.0000	0.9051	0.3238
$x^{(2)}$	0.9051	1.0000	0.4967
$x^{(7)}$	0.3238	0.4967	1.0000

Коэффициент детерминации: $\widehat{R^2_{y,(x^{(2)},x^{(7)})}} = 0{,}84028.$

Проверяемые переменные: $x^{(2)}$, $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице:

Таблица 4.30 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(10)}$
y	1.0000	0.9051	-0.5392
$x^{(2)}$	0.9051	1.0000	-0.3151
$x^{(10)}$	-0.5392	-0.3151	1.0000

Коэфициент детерминации: $\widehat{R^2_{y,(x^{(2)},x^{(10)})}} = 0{,}89093.$

Поскольку

$$\max_{\substack{1\le j\le 10\\(j\neq 2)}} \widehat{R^2_{y,(x^{(2)},x^{(J)})}} = \widehat{R^2_{y,(x^{(2)},x^{(10)})}} = 0.89093,$$

то наиболее информативной парой предикторов оказываются объясняющие переменные: $x^{(2)}$ –число занятых в экномике и $x^{(10)}$ – реальная заработанная плата. Подсчет нижней доверительной границы $R^{(1)}_{min}$дает следующие значения:

$$R^2_{min} = 0.83839.$$

Целесообразность включения в модель в качестве второй объясняющей переменной предиктора $x^{(10)}$ подтверждается сравнением нижних доверительных границ $R^2_{min}(1)$ и $R^2_{min}(2)$, т. к. $R^2_{min}(1) < R^2_{min}(2)$.

Шаг 3 - Выбор наилучшей переменной.

Проверяемые переменные: $x^{(1)}, x^{(2)}$, $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице:

Таблица 4.31 – Парные коэффициенты кореляции

	y	$x^{(1)}$	$x^{(2)}$	$x^{(10)}$
y	1.0000	0.7905	0.9051	-0.5392
$x^{(1)}$	0.7905	1.0000	0.8674	-0.4095
$x^{(2)}$	0.9051	0.8674	1.0000	-0.3151
$x^{(10)}$	-0.5392	-0.4095	-0.3151	1.0000

Коэффициент детерминации : $\widehat{R^2_{y,(x^{(1)},x^{(2)},x^{(10)})}} = 0{,}89574.$

Проверяемые переменные: $x^{(2)}, x^{(4)}$, $x^{(10)}$. Парные коэффициенты корреляции представлены в таблице:

Таблица 4.32 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(4)}$	$x^{(10)}$
y	1.0000	0.9051	0.4971	-0.5392
$x^{(2)}$	0.9051	1.0000	0.6523	-0.3151
$x^{(4)}$	0.4971	0.6523	1.0000	0.0960

$x^{(10)}$	-0.5392	-0.3151	0.0960	1.0000

Коэффициент детерминации : $\widehat{R^2_{y,(x^{(2)},x^{(4)},x^{(10)})}} = 0{,}89108.$

Проверяемые переменные: $x^{(2)}, x^{(7)}, x^{(10)}$. Парные коэффициенты корреляции представлены в таблице:

Таблица 4.33 – Парные коэффициенты кореляции

	y	$x^{(2)}$	$x^{(7)}$	$x^{(10)}$
y	1.0000	0.9051	0.3238	-0.5392
$x^{(2)}$	0.9051	1.0000	0.4967	-0.3151
$x^{(7)}$	0.3238	0.4967	1.0000	0.1083
$x^{(10)}$	-0.5392	-0.3151	0.1083	1.0000

Коэффициент детерминации : $\widehat{R^2_{y,(x^{(2)},x^{(7)},x^{(10)})}} = 0{,}8948.$

Поскольку

$$\max_{\substack{1\le j\le 10\\ (j\neq 2,10)}} \widehat{R^2_{y,(x^{(2)},x^{(j)})}} = \widehat{R^2_{y,(x^{(1)},x^{(2)},x^{(10)})}} = 0.89574,$$

то наиболее информативной тройкой предикторов оказываются объясняющие переменные: $x^{(1)}$ – число экономически активного населения, $x^{(2)}$ – число занятых в экномике и $x^{(10)}$ – реальная заработанная плата. Подсчет нижней доверительной границы $R^{(1)}_{min}$дает следующие значения:

$$R^2_{min} = 0.82685.$$

Нецелесообразность включения в модель в качестве третьей объясняющей переменной предиктора $x^{(1)}$ подтверждается сравнением нижних доверительных границ $R^2_{min}(2)$ и $R^2_{min}(3)$, т. к. $R^2_{min}(2) > R^2_{min}(3)$.

Таким образом, наиболее информативной парой предикторов оказываются объясняющие переменные: $x^{(2)}$ – число занятых в экономике, и $x^{(10)}$ – реальная заработанная плата.

4.4 Построение модели 3

Итак, в модели осталось две переменных, мультиколлинеарность отсутсвует, следовательно мы можем приступить к построению конечной модели методом наименьших квадратов. Необходимые расчеты представлены в таблице.

Проверим значимость полученного уравнения по F-критерию на 5% уровне значимости. Вычисление суммы квадратов дало следующие результаты: общая (Q_{sum}) – 7057041183.0631, обусловленная регрессией (Q_r) – 43301440607.4469, остаточная (Q_e) – 50358481790.51. Таким образом

$$F = \frac{Qr*(n-p-1)}{Qe*p} = 4.2993.$$

$F_{0.05:8:5}$ = 3.7083, F>$F_{0.05:8:5}$, следовательно уравнение значимо.

Общие характеристики модели:

Коэффициент детерминации: 0.89, коэффициент детерминации скорректированный: 0.87, наименьшее значение коэффициента детерминации: 0.84.

Итоговая модель регрессии

$$\hat{Y} = -904994{,}78 + 14{,}9705X_2 - 931{,}2343X_{10} + \varepsilon.$$

Таблица 4. 34 – Расчеты для метода наименьших квадратов

Y_i	X_{i2}	X_{i10}	$\widehat{Y}_i$	$Y_i - \widehat{Y}_i$	$(Y_i - \widehat{Y}_i)^2$	$(Y_i - Y_{\text{ср}})^2$	$(\widehat{Y}_i - Y_{\text{ср}})^2$
-0.0000	0.0000	0.0000	-0.0000	-0.0000	-0.0000	-0.0000	-0.0000
0.0000	0.0001	0.0000	0	0.0000	0.4670	1.3035	3.3309
0.0000	0.0001	0.0000	0	0.0000	0.7245	0.9484	3.3309
0.0000	0.0001	0.0000	0	0.0000	1.1054	0.5986	3.3309
0.0000	0.0001	0.0000	0	0.0000	1.7038	0.2702	3.3309
0.0000	0.0001	0.0000	0	0.0000	1.5060	0.3575	3.3309
0.0000	0.0001	0.0000	0	0.0000	2.1445	0.1301	3.3309
0.0001	0.0001	0.0000	0	0.0001	3.6340	0.0066	3.3309
0.0001	0.0001	0.0000	0	0.0001	4.6463	0.1092	3.3309
0.0001	0.0001	0.0000	0	0.0001	5.3486	0.2378	3.3309
0.0001	0.0001	0.0000	0	0.0001	6.2726	0.4616	3.3309
0.0001	0.0001	0.0000	0	0.0001	6.9277	0.6512	3.3309
0.0001	0.0001	0.0000	0	0.0001	7.3978	0.8007	3.33093
0.0001	0.0001	0.0000	0	0.0001	8.4804	1.1817	3.3309

```
Qsumm: 7057041183.0631
Qr: 43301440607.4469
Qe: 50358481790.51
Критерий Стьюдента
V (число степеней свободы): 10
t критический: 2.2281
|t наблюдаемый| < t критический: b(0) = 0.0019301
|t наблюдаемый| > t критический: b(2) = 152.3884 значимо
|t наблюдаемый| < t критический: b(10) = 0.84658 не значимо
Критерий Фишера
V1 (число степеней свободы): 3
V2 (число степеней свободы): 10
F наблюдаемый: 4.2993
F критический: 3.7083
F наблюдаемый > F критический: H0 отклоняется, принимается альтернативная – уравнение значимо
Коэфициент детерминации (R2): 6.1359
Коэфициент детерминации скорректированный (R2scorr): 7.1631
Общие характеристики модели
Коэфициент детерминации (Ryx2): 0.89093
Коэфициент детерминации скорректированный (Ryx2scorr): 0.86912
Наименьшее значение коэфициента детерминации (Ryx2min): 0.83839
Итоговая модель регрессии
-904994.7883+14.9705*x2-931.2343*x10
The end!
```

Рисунок 4.5 – Итоговая модель

4.5 Проверка качества итоговой модели

В модели остались две переменные: X_2, X_{10}. Динамика показателей отражена на рисунке 4.6.

Рисунок 4.6 – Данные.

Значения матрицы парных коэффициентов корреляции позволяют сделать вывод об отсутствии мультиколлинеарности (рисунок 4.7). Анализируя значения коэффициентов корреляции между зависимой и факторными переменными, которые расположены в первой строке (столбце)

корреляционной матрицы наблюдается прямое влияние фактора «число занятых в экономике» ($r(Y, X_2) = 0.905$) и обратное влияние фактора «реальная з/п» ($r(Y, X_{10}) = -0.54$).

Анализируя элементы матрицы межфакторной корреляции, которая получается вычеркиванием первого столбца и первой строки из приведенной матрицы, можем заметить, что $r(X_2, X_{10}) = -0.3151$. Коэффициент >0.5, что позволяет сделать вывод об отсутвиии мультиколлинеарности. Также данный вывод подтверждает критерий χ^2 (хи-квадрат).

```
MODEL: 3 ----------------------------
Критерий Хи^2
Выбранная гипотеза: H0 - Отсутствует мультиколлинеарность
Количество наблюдений (n): 13 Количество переменных (p): 2
Матрица парных коэфициентов корреляции Rxx
    1.0000   -0.3151
   -0.3151    1.0000

Матрица парных коэфициентов корреляции Ryx
    1.0000    0.9051   -0.5392
    0.9051    1.0000   -0.3151
   -0.5392   -0.3151    1.0000
```

Рисунок 4.7 – Матрица парных коэффициетов корреляции.

Итак, оценим уравнение множественной линейной регрессии вида:

$$Y = \beta_0 + \beta_1 X_2 + \beta_3 X_{10} + \varepsilon.$$

Для построения уравнения регрессии воспользуемся статистическим пакетом Eviews. Эмпирическое уравнение регресси построенное методом наименьших квадратов имеет вид:

$$Y = -1130631 + 17{,}899 X_2 - 726{,}836 X_{10} + \varepsilon.$$

Естественно ожидать увеличение покзателя ВВП с увеличением количества числа занятых в экономике, поэтому при переменой X_2 ожидался положительный коэффициент. Также естественно ожидать, что при росте показателя реальной заработанной платы (X_{10}) произойдет снижение объема ВВП, так как з/п выплачивается из ВВП.

Таким образом, знаки полученных оценок коэффициентов соответствуют ожиданиям.

Проверка остатков на отсутствие автокорреляции проводится с помощью теста Дарбина-Уотсона.

Значение статистики Дарбина-Уотсона (Durbin-Watson stat) – $d = 0{,}66$ (рисунок 4.8). При $\alpha = 0.05$, $n = 15$, $m = 2$ по таблице пороговых точек статистики Дарбина-Уотсона определяем, что $d_l = 0{,}946$, $d_u = 1.543$. Статистика попадает в интервал $d < d_L$, что говорит о наличии автокорреляции.

Dependent Variable: Y
Method: Least Squares
Date: 05/26/20 Time: 09:40
Sample: 2005 2019
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X2	17.89896	2.953917	6.059399	0.0001
X10	-726.8360	562.5231	-1.292100	0.2207
C	-1130631.	234953.7	-4.812143	0.0004

R-squared	0.798865	Mean dependent var	64263.45
Adjusted R-squared	0.765342	S.D. dependent var	28360.22
S.E. of regression	13738.11	Akaike info criterion	22.07059
Sum squared resid	2.26E+09	Schwarz criterion	22.21220
Log likelihood	-162.5294	Hannan-Quinn criter.	22.06908
F-statistic	23.83069	Durbin-Watson stat	0.666688
Prob(F-statistic)	0.000066		

Рисунок 4.8 – Уравнение регрессии.

Расширим количество наблюдений: с 1998 по 2019 год для проведения более точного анализа (Приложение С). Для преодоления автокорреляции построим уравнение регрессии обобщенным методом наименьших квадратов.

Получаем следующее уравнение (рисунок 4.9):

$$Y = -409303{,}9 + 8{,}060718X_2 - 941{,}0225X_{10} + \varepsilon.$$

Dependent Variable: Y
Method: Least Squares
Date: 05/25/20 Time: 12:46
Sample: 1998 2019
Included observations: 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X2	8.060718	1.109965	7.262137	0.0000
X10	-941.0225	431.2683	-2.181988	0.0419
C	-409303.9	90211.51	-4.537158	0.0002

R-squared	0.753831	Mean dependent var	46754.17
Adjusted R-squared	0.727919	S.D. dependent var	35090.36
S.E. of regression	18303.63	Akaike info criterion	22.59371
Sum squared resid	6.37E+09	Schwarz criterion	22.74249
Log likelihood	-245.5308	Hannan-Quinn criter.	22.62876
F-statistic	29.09139	Durbin-Watson stat	0.735276
Prob(F-statistic)	0.000002		

Рисунок 4.9 – Уравнение регрессии.

Коэффициент демерминации данной модели $R^2 = 0{,}75$. Скорректированный коэффициент детерминации равен 0,72. Качество модели может быть охарактеризовано как высокое, построенное уравнение описывает не менее 72% вариации зависимой переменной ВВП. Мы можем говорить о хорошей подгонке регрессионной модели к наблюдаемым значениям ВВП.

Статистическая значимость полученного уравнения оценивается с помощью критерия Фишера:

F-statistic = 29.09

Prob(F-statistic) = 0.0000002.

F-критерий Фишера проверяет нулевую гипотезу, которая формулируется так: H_0: «уравнение в целом статистически не значимо». При её принятии все коэффициенты уравнения при независимых переменных можно считать равными нулю. В уравнении *Prob(F-statistic) = 0.0000002 < F-statistic = 29.09.* Значит H_0 отклоняется, уравнение в целом признается статистически значимым.

Помимо оценки уравнения необходимо оценить и его коэффициенты.

Значимость коэффициентов проверяется по критерию Стьюдента, если наблюдаемое значение *t-statistic* > *t-критич.*, то можно сделать вывод о значимости коэффициента. На рисунке 4.9 представлены оценки полученного уравнения, которые позволяют сделать вывод, что X_{10} статистически незначим.

Таким образом, независимую переменую X_2 следует признать значимо влияющей на ВВП. Значение коэффициента при этой переменной говорит о том, что при увеличении числа занятых в экономике на 1 тыс. человек показатель ВВП повысится на 8,06 млрд. руб.

Проверка остатков на гомоскедастичность осуществляется с помощью теста Бреуша-Пагана и теста Уайта (рисунки 4.10, 4.11). По тесту Бреуша-Пагана:

$$F\text{-}statistic = 0{,}54 < Prob.\ F(2.19) = 0{,}58,$$

следовательно, гетероскедастичность отсутствует.

По тесту Уайта

$$Prob.\ F(5.16) = 0{,}16,$$

что меньше $\alpha = 0.05$, следовательно, гетероскедастичность отсутствует.

Heteroskedasticity Test: Breusch-Pagan-Godfrey

F-statistic	0.543987	Prob. F(2,19)	0.5892
Obs*R-squared	1.191531	Prob. Chi-Square(2)	0.5511
Scaled explained SS	0.869107	Prob. Chi-Square(2)	0.6476

Рисунок 4.10 – Тест Бреуша-Пагана.

Heteroskedasticity Test: White

F-statistic	1.831277	Prob. F(5,16)	0.1635
Obs*R-squared	8.007528	Prob. Chi-Square(5)	0.1558
Scaled explained SS	5.840721	Prob. Chi-Square(5)	0.3220

Рисунок 4.11 – Тест Уайта.

При тестировании выборки на нормальное распределение основной гипотезой является:

H_0: «выборочная совокупность имеет нормальное распределение».

Для проверки нулевой гипотезы о нормальном распределении остатков может использоваться тест Жака-Бера, который основан на проверке статистической значимости расхождения фактических значений коэффициентов асимметрии и эксцесса с ожидаемыми для нормального распределения нулевыми значениями характеристик.

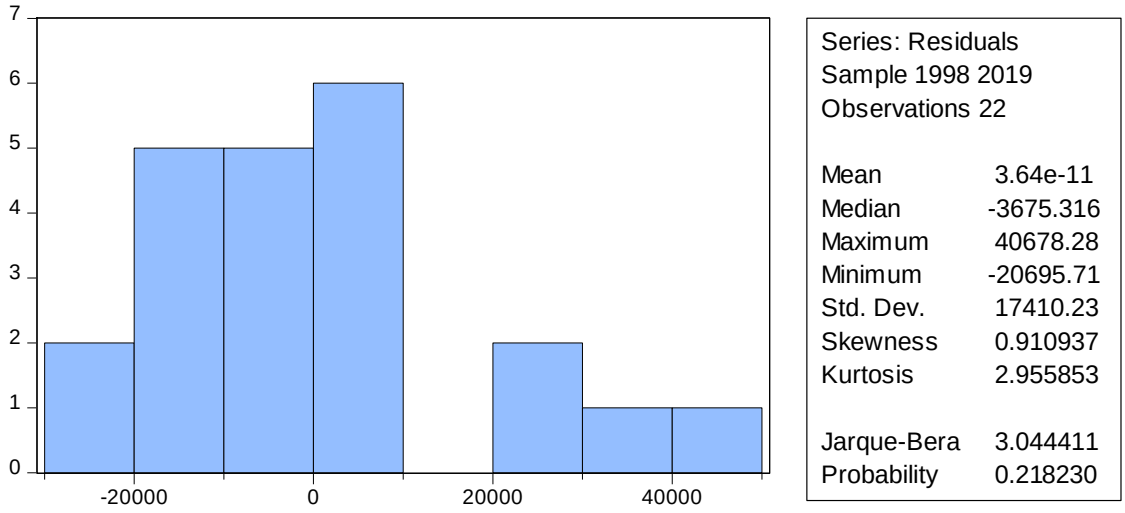

Рисунок 4.12 – Тест Жака-Бера

Так как *Probability = 0.218>0.05*, то гипотеза о нормальном распределении принимается.

ЗАКЛЮЧЕНИЕ

Задача исследования зависимости одной эндогенной переменной Y – показателя ВВП от нескольких объясняющих (экзогенных) переменных решается с помощью множественного регрессионного анализа. Для эконометрического анализа ВВП авторы использовали различные факторы. Для построения современной актуальной модели был произведен отбор наиболее значимых переменных и построена собственною модель.

В результате работы были рассмотрены этапы построения эконометрических моделей, проблемы и ошибки спецификации:

- определение набора объясняющих переменных;
- выбор формы уравнения и модели случайного члена.

Также рассмотрены методы и подходы решения данных проблем:

- априорный и апостериорный подходы;
- метод включения/исключения и пошагового отбора переменных;
- графический, аналитический и экспериментальный методы выбора формы уравнения.

Были рассмотрены проблемы гетероскедастичности, мультиколлинеарности и автокорреляции, а также способы их преодоления.

В ходе выполнения работы были исследованы статистические данные показателя ВВП РФ в период с 1998 по 2019 года.

Корреляционный анализ позволил выявить факторы, имеющие тесную линейную взаимосвязь с показателем ВВП. Дальнейшие исследования позволили установить оптимальный набор факторов для включения в модель регрессии. Проведена проверка предпосылок МНК. Установлена проблема автокорреляции. Проблема решена применением доступного обобщенного метода наименьших квадратов.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т. 2-е изд., испр. – Т. 2: Айвазян С.А. Основы эконометрики. – М.: ЮНИТИ-ДАНА, 2001. – 432с., ISBN 5-238-00305-6.
2. Гладилин А.В., Эконометрика : учебное пособие / А.В. Гладилин, А.Н. Герасимов, Е.И. Громов. – 3-е изд., стер. – М. : КНОРУС, 2014. – 228с., ISBN 978-5-406-03792-8.
3. Спецификация модели, [Электронный ресурс]. Режим доступа: https://studopedia.su/9_6398_spetsifikatsiya-modeli.html - Дата обновления:18 марта 2019 г.
4. Кисляк Н.В., Эконометрика, [Электронный ресурс]. Режим доступа: http://elar.urfu.ru/bitstream/10995/1479/6/1324633_lectures.pdf - Дата обновления:18 марта 2019 г.
5. Бабешко Л.О. Эконометрика и экономическое моделирование: учебник / Л.О. Бабешко, М.Г. Бич, И.В. Орлова. – М.: Вузовский учебник : ИНФРА-М. 2019. – 385 с. : ил. – (Высшее образование: Бакалавриат). ISBN 978-5-9558-0576-4.
6. Аистов А.В., Эконометрика шаг за шагом [Текст] : учебное пособие для студентов высших учебных заведений, обучающихся по направлению подготовки "Экономика" / А. В. Аистов, А. Г. Максимов. - Москва : Издательский дом ГУ ВШЭ, 2006. - 177, [1] с. : ил., табл. - Библиогр.: с. 171-178 (104 назв.). - ISBN 5-7598-0332-8.
7. Multicollinearity in Regression Analysis: Problems, Detection, and Solutions [Электронный ресурс]. Режим доступа: http://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/ - Дата обновления: 1 января 2019 г.
8. Multicollinearity [Электронный ресурс]. Режим доступа: https://onlinelibrary.wiley.com/doi/full/10.1002/wics.84 Дата обновления: 1 января 2019 г.

9. Определение мультиколлинеарности [Электронный ресурс]. Режим доступа: https://studme.org/198429/ekonomika/opredelenie_multikollinearnosti - Дата обновления: 1 января 2019 г.
10. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных [Электронный ресурс]. Режим доступа: http://www.machinelearning.ru/wiki/index.php?title=Мультиколлинеарность – Дата обновления: 1 января 2019 г.
11. Математика и трейдинг. Основы статистического анализа Корреляция, коллинеарность и мультиколлинеарность [Электронный ресурс]. Режим доступа: https://quantpro.ru/archives/5218 - Дата обновления: 1 января 2019 г.
12. Эконометрика и эконометрическое моделирование : учебник / Л.О. Бабешко, М.Г. Бич, И.В. Орлова. - М. : Вузовский учебник : ИНФРА-М, 2019. - 385 с. : ил. — (Высшее образование: Бакалавриат). - Режим доступа: http://znanium.com/catalog/product/1029152.
13. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. М.: ЮНИТИ, 1998.
14. Доугерти К. Введение в эконометрику: учебное пособие 2-е изд. М. ИНФРА-М, 2004.
15. Магнус Я.П., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: учеб. Пособие для вузов. М.: Дело, 2004.
16. Построение и анализ модели в среде EViews [Электронный ресурс]. Режим доступа: https://vuzlit.ru/733110/postroenie_analiz_modeli_srede_eviews - Дата обновления:18 октября 2019 г.
17. Понятие о ВВП и ВНП, [Электронный ресурс]. Режим доступа: HTTPS://FOXFORD.RU/WIKI/OBSCHESTVOZNANIE/PONYATIE-O-VVP-I-VNP - Дата обновления: 20 ноября 2019 г.
18. Эконометрический анализ статистической взаимосвязи показателей социально-экономического развития России, [Электронный ресурс]. Режим

доступа: https://www.fundamental-research.ru/ru/article/view?id=39937 - Дата обновления: 28 ноября 2019 г.

19. Статистический анализ факторов, формирующих ВВП, [Электронный ресурс]. Режим доступа: http://apej.ru/article/04-04-16 - Дата обновления:18 ноября 2019 г.
20. Храмов, А. В. Эконометрический анализ валового внутреннего продукта России / А. В. Храмов, А. А. Миннуллин, Н. Н. Нуруллин, Е. И. Кадочникова. — Текст : непосредственный // Молодой ученый. — 2014. — № 21 (80). — С. 452-454. — URL: https://moluch.ru/archive/80/14317/ (дата обращения: 22.05.2020).
21. Эконометрический анализ динамики макроиндикаторов экономики России в период с 2003 по 201 гг., [Электронный ресурс]. Режим доступа: https://cyberleninka.ru/article/n/ekonometricheskiy-analiz-dinamiki-makroindikatorov-ekonomiki-rossii-v-period-s-2003-po-2016-gg/viewer - Дата обновления: 2 декабря 2019 г.
22. Эконометрический анализ валового внутреннего продукта на душу населения в Российской Федерации, [Электронный ресурс]. Режим доступа: https://cyberleninka.ru/article/n/ekonometricheskiy-analiz-valovogo-vnutrennego-produkta-na-dushu-naseleniya-v-rossiyskoy-federatsii/viewer - Дата обновления: 12 декабря 2019 г.
23. Что такое ВВП простыми словами, [Электронный ресурс]. Режим доступа: https://mbfinance.ru/terminy/chto-takoe-vvp-prostymi-slovami/ - Дата обновления: 22 декабря 2019 г.
24. ВВП России по годам: 1991 – 2020 [Электронный ресурс]. Режим доступа: http://global-finances.ru/vvp-rossii-po-godam/ - Дата обновления:18 октября 2019 г.
25. ЕМИСС [Электронный ресурс]. Режим доступа: https://www.fedstat.ru - Дата обновления:28 января 2020 г.

26. Федеральная служба государственной статистики [Электронный ресурс]. Режим доступа: https://www.gks.ru/ - Дата обновления:02 февраля 2020 г.
27. Федеральная служба государственной статистики Финансы России [Электронный ресурс]. Режим доступа: https://www.gks.ru/folder/210/document/13237 - Дата обновления:02 февраля 2020 г.
28. Федеральная служба государственной статистики Уровень безработицы [Электронный ресурс]. Режим доступа: https://www.gks.ru/labour_force - Дата обновления:02 февраля 2020 г.
29. Статистический ежегодник 2019 [Электронный ресурс]. Режим доступа: https://gks.ru/bgd/regl/b19_13/Main.htm - Дата обновления:02 февраля 2020 г.
30. России в цифрах [Электронный ресурс]. Режим доступа: https://www.gks.ru/folder/210/document/12993 - Дата обновления:02 февраля 2020 г.
31. Краткая информация об исполнении федерального бюджета [Электронный ресурс]. Режим доступа: https://www.minfin.ru/ru/statistics/fedbud/execute/?id_65=80041-yezhegodnaya_informatsiya_ob_ispolnenii_federalnogo_byudzhetadannye_s_1_yanvarya_2006_g - Дата обновления: 02 февраля 2020 г.
32. Федеральная служба государственной статистики Внешняя торговля [Электронный ресурс]. Режим доступа: https://gks.ru/folder/11193?print=1- Дата обновления: 02 февраля 2020 г.
33. Валовые инвестиции [Электронный ресурс]. Режим доступа: https://spravochnick.ru/investicii/valovye_investicii/ - Дата обновления: 02 февраля 2020 г.

ПРИЛОЖЕНИЕ А

Статистические данные за период с 2005 по 2017 гг.

Период	ВВП в рыночных ценах млрд руб.	Число экономически активного населения.	Числ. занятных в экономике, тыс чел.	Прибыль организаций млрд руб.	Численность научных организаций	Номинальная среднемесячная зп руб.	Уровень безработицы, %	Прямые Инвест. в осн. капитал, млрд долл.	Потребит. расходы млн руб.	Общегос. доходы, млрд руб.	Реальная з/п темп роста к предыдущему году, в %.	Число выбывших.	Экспорт млн дол по всем странам мира.	Общегос расходы, млрд руб.
Период	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
2005	21609,8	73581	68339	3674	3566	8554,9	7,1	13,072	241 473	5127,2	112,6	69798	241 473	3539,45
2006	26917,2	74418,9	69168,7	6085	3622	10633,9	7,1	13,678	351 930	6278,9	113,3	54061	351 930	4284,8
2007	33247,5	75288,9	70770,3	6412	3957	13593,4	6	27,797	351 930	7781,1	117,2	47013	351 930	5986,6
2008	41276,8	75700,1	71003,1	5354	3666	17290,1	6,2	27,027	467 581	9275,9	111,5	39508	467 581	7570,9
2009	38807,2	75694,2	69410,5	5852	3536	18637,5	8,3	15,906	301 667	7337,8	96,5	32458	301 667	9660,1
2010	46308,5	75477,9	69933,7	7353	3492	20952,2	7,3	13,81	397 068	8305,4	105,2	33578	397 068	10117,5
2011	60282,5	75779	70856,6	8794	3682	23369,2	6,5	18,415	516 718	11367,7	102,8	36774	516 718	10925,6
2012	68163,9	75676,1	71545,4	9213	3566	26628,9	5,5	18,666	524 698	12855,5	108,4	122751	524 698	12895
2013	73133,9	75528,9	71545,4	9519	3605	29792	5,5	26,118	527 266	13019,9	101,2	186382	527 266	13342,9
2014	79199,7	75428,4	71539	10465	3604	32495,4	5,2	22,031	497 834	14496,9	91	310496	497 834	14831,6
2015	83232,6	76588	72324	12653	4175	34030	5,6	6,853	29 190	13659,2	100,8	353233	29 190	15620,3
2016	86010,2	76858	72755	15823	4032	36709	5,5	32,539	285 772	13460	102,9	313210	285 772	16416,4
2017	92089,3	76285,4	72142	14079	3944	39167	5,2	28,557	357817	15088,9	108,5	377155	357817	16420,3

ПРИЛОЖЕНИЕ В

Листинг программного кода

```
Regression_analysis.m
clear;
clc;
multiply = 1;
multicollinearAddition = 0.8;
alphaHi2 = 0.05; %Критерий Хи^2
alphaF = 0.05; %Критерий Фишера
alphaT = 0.025; %(Совпадает с 0,05) %Критерий Стьюдента
y = xlsread('3.xlsx', 'B3:B15', '', 'basic');
sourceX = xlsread('3.xlsx','C3:O15', '', 'basic');
%Уменьшим в н раз, чтобы не перегружать компьютер
y = y.*multiply;
sourceX = sourceX.*multiply;
%Добавляем индексы в первую строку
size_x = size(sourceX);
n = size_x(1);
p = size_x(2);
indexies = ones(1, p);
for i=1:p
    indexies(i) = i;
end
sourceX = [indexies; sourceX];
%Записываем индексы всех переменных x
size_x = size(sourceX);
p = size_x(2);
xIndexedForRyx = zeros(1, p);
for i=1:p
    xIndexedForRyx(i) = i+1;
end
indexesForRxx = xIndexedForRyx-1;
for o=1:p
    disp(['MODEL: ', num2str(o), ' ------------------------------']);
    x = removeXFromMAtrix(sourceX, indexesForRxx);
    [Rxx, Ryx] = calculate_Regression(x, y);
    %Проверяем на мультиколлинеарность, удаляем переменные
    [multicollinear, indexesForRxx] = test_Multicollinear(x, Rxx, indexesForRxx,
multicollinearAddition, alphaHi2);
    x = removeXFromMAtrix(sourceX, indexesForRxx);
    size_x = size(x);
    p = size_x(2);
    disp(['Количество наблюдений (n): ', num2str(n), ' Количество переменных
(p): ', num2str(p)]);
    if abs(n-p)<2
        disp(['Число наблюдений отличается от числа переменных всего на: ',
num2str(abs(n-p))]);
        stopCalculation = true;
    else
        stopCalculation = false;
    end
    if ~stopCalculation
        if multicollinear
            [Rxx, Ryx] = calculate_Regression(x, y);
        end
        disp('Матрица парных коэффициентов корреляции Rxx')
        disp(Rxx);
        disp('Матрица парных коэффициентов корреляции Ryx')
        disp(Ryx);
```

```
        X = calculate_X(x);
        disp('Матрица X');
        disp(X);

        [b, Am1, isFgls] = calculate_mnk(y, x, X);
        vs = 'Вычисление производится: ';
        if isFgls
            disp ([vs, 'Обобщенным методом наименьших квадратов']);
        else
            disp ([vs, 'Методом наименьших квадратов']);
        end
        disp('Коэффициенты регрессии (b)');
        disp (b);
        [ySumm, ySredn] = calculate_SummAndAvg(y);
        [yKr, sredn2, krysh, srednMkrysh2, krysh2, Qsumm, Qe , Qr] =
calculate_table(y, ySredn);
        %Таблица
        disp('Таблица расчётов');
        disp('1 - y, 2 - yKr, 3 - (y(i) - yKr(i)), 4 - (y(i) - yKr(i))^2, 5 -
(y(i) - ySredn)^2, 6 - (yKr(i) - ySredn)^2');
        allInOne = [[-1;y], x, [-2;yKr], [-3;krysh], [-4;krysh2], [-5;sredn2],
[-6;srednMkrysh2]];
        disp(allInOne);
        disp(['Qsumm: ', num2str(Qsumm)]);
        disp(['Qr: ', num2str(Qr)]);
        disp(['Qe: ', num2str(Qe)]);
        if ~isFgls
            disp('Критерий Стьюдента');
            [sigmaKr, sigmaKr2] = calculate_DispAndOtkl(x, Qsumm);
            sigmakrb2 = sigmaKr2 * Am1;
            sigmakrb = sigmakrb2.^(1/2);
            [tnabl, tkrit, V, znachimoS] = calculate_Student(b, x, sigmakrb,
alphaT);
            disp(['V (число степеней свободы): ', num2str(V)]);
            disp(['t критический: ', num2str(tkrit)]);
            for i = 1:length(tnabl)
                tn = abs(tnabl(i));
                if (tn>tkrit)
                    if (i==1)
                        disp(['|t наблюдаемый| > t критический: ', 'b(',
num2str(0), ') = ', num2str(tn)]);
                    else
                        disp(['|t наблюдаемый| > t критический: ', 'b(',
num2str(indexesForRxx(i-1)), ') = ', num2str(tn),' значимо']);
                    end
                else
                    if (i==1)
                        disp(['|t наблюдаемый| < t критический: ', 'b(',
num2str(0), ') = ', num2str(tn)]);
                    else
                        disp(['|t наблюдаемый| < t критический: ', 'b(',
num2str(indexesForRxx(i-1)), ') = ', num2str(tn),' не значимо']);
                    end
                end
            end
        else
            znachimoS = false;
        end
        disp('Критерий Фишера');
        [fnabl, fkrit, V1, V2, znachimoF] = calculate_Fisher(x, Qr, Qe, alphaF);
        disp(['V1 (число степеней свободы): ', num2str(V1)]);
        disp(['V2 (число степеней свободы): ', num2str(V2)]);
```

```
        disp(['F наблюдаемый: ', num2str(fnabl)]);
        disp(['F критический: ', num2str(fkrit)]);
        if (fnabl>fkrit)
            disp('F наблюдаемый > F критический: H0 отклоняется, принимается
альтернативная - уравнение значимо');
        else
            disp('F наблюдаемый < F критический: уравнение не значимо');
        end
        [R2, R2scorr] = calculate_Determinate(Qr, Qsumm, x);
        disp(['Коэфициент детерминации (R2): ', num2str(R2)]);
        disp(['Коэфициент детерминации скорректированный (R2scorr): ',
num2str(R2scorr)]);
        disp('Общие характеристики модели');
        l = p; % l = кол-ву переменных
        [Ryx2, Ryx2scorr, Ryx2min] = calculate_Determinate_R(Ryx, n, p, 1);
        disp(['Коэфициент детерминации (Ryx2): ', num2str(Ryx2)]);
        disp(['Коэфициент детерминации скорректированный (Ryx2scorr): ',
num2str(Ryx2scorr)]);
        disp(['Наименьшее значение коэфициента детерминации (Ryx2min): ',
num2str(Ryx2min)]);
        if (multicollinear || ((znachimoS ~= true) && (znachimoF ~= true)))
            %Шаг 1
            disp('Шаг 1 - Выбор первого предиката (переменной)');
            max = Ryx(1, 2);
            indexMax = 2;
            for i = 3:(p+1)
                if (Ryx(1, i)>max)
                    max = Ryx(1, i);
                    indexMax = i;
                end
            end
            disp(['Индекс x в матрицу Ryx: ', num2str(indexMax),' = ',
num2str(max)]);
            disp(['Лучший: x(', num2str(x(1, indexMax-1)),') = ',
num2str(max)]);
            [R2, R2scorr, R2min] = calculate_Determinate_r_mal(max, n, 1);
            disp(['Коэфициент детерминации (R2): ', num2str(R2)]);
            disp(['Коэфициент детерминации скорректированный (R2scorr): ',
num2str(R2scorr)]);
            disp(['Наименьшее значение коэфициента детерминации (R2min): ',
num2str(R2min)]);
            oldR2min = R2min;
            %Шаг 2
            %Начало цикла
            xIndexedForRyx = [indexMax];
            newP = length(xIndexedForRyx);
            %Избавляемся от мультиколлинеарности
            for step=2:p
                disp(['Шаг ',num2str(step),' - Выбор наилучшей переменной'])
                [bestRyx, newBestXindex] = calculate_PairCorrelation(Ryx,
xIndexedForRyx, x);
                newP = length(newBestXindex);
                xIndexedForRyx = newBestXindex;
                lastBestX = newBestXindex(length(newBestXindex));
                disp(['Индекс переменной с наибольшим значением парной
корреляции x: ', num2str(lastBestX)]);
                disp(['Переменная с наибольшим значением парной корреляции: x(',
num2str(x(1,lastBestX-1)),')']);
                disp('Описательные статистики (Ryx)')
                disp(bestRyx);
                [Ryx2, Ryx2scorr, Ryx2min] = calculate_Determinate_R(bestRyx, n,
newP, 1);% 1 - второй порядок
```

```
                newR2min = Ryx2min;
                disp(['Наименьшее значение коэфициента детерминации (Ryx2min):
', num2str(newR2min)]);

                if (newR2min<oldR2min)
                    disp('newR2min<oldR2min');
                    disp(['Переменная x(', num2str(x(1,lastBestX-1)),') не
подходит']);
                    disp('Останавливаем поиск');
                    disp('Объясняющие переменные:');
                    xIndexedForRyx(length(xIndexedForRyx)) = [];
                    xIndexedForRyx = sort(xIndexedForRyx);
                    disp(['Индексы (с учётом y): ', num2str(xIndexedForRyx(1,
:))]);
                    indexesForRxx = xIndexedForRyx;
                    for o=1:length(xIndexedForRyx)
                        disp(['x(', num2str(x(1,xIndexedForRyx(o)-1)), ')']);
                        indexesForRxx(o) = x(1, xIndexedForRyx(o)-1);
                    end
                    if (o==p)
                        stopCalculation = true;
                    end
                    break;
                else
                    disp('newR2min>oldR2min');
                    xIndexedForRyx = sort(xIndexedForRyx);
                    disp(['Индексы (с учётом y): ', num2str(xIndexedForRyx(1,
:))]);
                    indexesForRxx = xIndexedForRyx;
                    for o=1:length(xIndexedForRyx)
                        disp(['x(', num2str(x(1,xIndexedForRyx(o)-1)), ')']);
                        indexesForRxx(o) = x(1, xIndexedForRyx(o)-1);
                    end
                    if (o==p)
                        stopCalculation = true;
                    end
                    if (step==p)
                        break;
                    end
                end
                oldR2min = newR2min;
            end
        else
            stopCalculation = true;
        end
        if p==1
            stopCalculation = true;
        end
        if (stopCalculation)
            disp('Итоговая модель регрессии');
            if (length(b)>1 && b(2)>0)
                buffStr = strcat(num2str(b(1)), '+');
            else
                buffStr = num2str(b(1));
            end
            for k = 2:length(b)
                currentX = indexesForRxx(k-1);
                if k~=length(b)
                    if (b(k+1)>0)
                        buffStr = strcat(buffStr,
num2str(b(k)),'*x',num2str(currentX), '+');
                    else
```

```
                        buffStr = strcat(buffStr,
num2str(b(k)),'*x',num2str(currentX));
                    end
                else
                    buffStr = strcat(buffStr,
num2str(b(k)),'*x',num2str(currentX));
                end
            end
            disp(buffStr);
            if false
                disp('График - Выбор лучшего предиката (переменной)');
                max = Ryx(1, 2);
                indexMax = 2;
                for i = 3:(p+1)
                    if (ismember(i,xIndexedForRyx) && Ryx(1, i)>max)
                        max = Ryx(1, i);
                        indexMax = i;
                    end
                end
                disp(['Индекс x в матрице Ryx: ', num2str(indexMax),' = ',
num2str(max)]);
                disp(['Лучший: x(', num2str(x(1, indexMax-1)),') = ',
num2str(max)]);
                xclean = x;
                xclean(1,:) = [];
                xc = xclean(:, sourceX(1, indexMax));
                p=polyfit(xc,y,1);
                f=polyval(p,xc);
                plot(xc,y,'o',xc,f)
            end
            disp('The end!');
            break;
        end
    else
        break;
    end
end
```

calculate_AlgebracDop.m

```
function [nx] = calculate_AlgebraicDop(x, i, j)
nx = x;
nx(i, :) = [];
nx(:, j) = [];
```

calculate_Coef_r.m

```
function [r] = calculate_Coef_r(x1, x2, n)
x1x2Summ = 0;
for i=1:length(x1)
    x1x2Summ = x1x2Summ + x1(i)*x2(i);
end
x1Summ = 0;
for i=1:length(x1)
    x1Summ = x1Summ + x1(i);
end
x2Summ = 0;
for i=1:length(x1)
    x2Summ = x2Summ + x2(i);
end
x1Summ2 = 0;
for i=1:length(x1)
    x1Summ2 = x1Summ2 + x1(i)^2;
end
x2Summ2 = 0;
for i=1:length(x1)
    x2Summ2 = x2Summ2 + x2(i)^2;
end
r = (n*x1x2Summ-x1Summ*x2Summ)/(sqrt(n*x1Summ2-x1Summ^2) * sqrt(n*x2Summ2-
x2Summ^2));
```

calculate_Determinate.m

```
function [R2, R2scorr] = calculate_Determinate(Qr, Qsumm, x)
xclean = x;
xclean(1,:) = [];
size_x = size(xclean);
n = size_x(1);
p = size_x(2);
R2 = Qr/Qsumm;
R2scorr = 1-((1-R2)*((n-1)/(n-p-1)));
```

calculate_Determinate_R.m

```
function [Ryx2, Ryx2scorr, Ryx2min] = calculate_Determinate_R(Rxy, n, p, delta)
RxyPor = calculate_AlgebraicDop(Rxy, delta, delta);
Ryx2 = 1 - (det(Rxy)/det(RxyPor));
Ryx2scorr = 1 - (1-Ryx2) * ((n-1)/(n-p-1));
Ryx2min = Ryx2scorr - 2 * sqrt((2*p*(n-p-1))/((n-1)*(n^2-1))) * (1-Ryx2);
```

calculate_Determinate_r_mal.m

```
function [R2, R2scorr, R2min] = calculate_Determinate_r_mal(r, n, p)
R2 = r^2;
R2scorr = 1 - (1-R2) * ((n-1)/(n-p-1));
R2min = R2scorr - 2 * sqrt((2*p*(n-p-1))/((n-1)*(n^2-1))) * (1-R2);
```

calculate_DispAndOtkl.m

```
function [sigmaKr, sigmaKr2] = calculate_DispAndOtkl(x, Qsumm)
xclean = x;
xclean(1,:) = [];
size_x = size(xclean);
n = size_x(1);
```

```
p = size_x(2);
%Несмещенная оценка остаточной дисперсии
sigmaKr2 = (1/(n-p-1))*Qsumm;
%Оценка среднеквадратического отклонения
sigmaKr = sqrt(sigmaKr2);
```

calculate_fgls.m

```
function [coeff,se,EstCoeffCov] = calculate_fgls(y, x)
xclean = x;
xclean(1,:) = [];
[coeff,se,EstCoeffCov] = fgls(xclean,y);
```

calculate_Fisher.m

```
function [fnabl, fkrit, V1, V2, znachimo] = calculate_Fisher(x, Qr, Qe, alphaF)
xclean = x;
xclean(1,:) = [];
size_x = size(xclean);
n = size_x(1);
p = size_x(2);
%Проверяем значимость уравнения регрессии
fnabl = (Qr*(n-p-1))/(Qe*p);
V1 = p + 1;
V2 = n - p - 1;
fkrit=finv(1-alphaF,V1,V2);
znachimo = false;
if (fnabl>fkrit)
    znachimo = true;
end
```

calculate_Hi2.m

```
function [hi2, hi2krit] = calculate_Hi2(x, R, alpha)
xclean = x;
xclean(1,:) = [];
%H0: отсутствует мультиколлинеарность
%H1: есть мультиколлинеарность
size_x = size(xclean);
n = size_x(1);
p = size_x(2);
hi2 = -(n-1-(1/6)*(2*p+5))*log(det(R));
V = (1/2)*p*(p-1);
hi2krit = chi2inv(1-alpha, V);
```

calculate_mnk.m

```
function [b, Am1, isFgls] = calculate_mnk(y, x, X)
myWarn = warning('error','MATLAB:nearlySingularMatrix');
Xt = transpose(X);
XtX = Xt * X;
Xty = Xt * y;
b = zeros(0);
Am1 = zeros(length(XtX));
try
    Am1 = inv(XtX);
    b = XtX\Xty;
    isFgls = false;
catch
    try
        [b,se,EstCoeffCov] = calculate_fgls(y, x);
    catch
    end
    isFgls = true;
  end
```

```
warning(myWarn);
```

calculate_PairCorrelation.m

```
function [bestMatrix, bestXindex] = calculate_PairCorrelation(Ryx, indexes, x)
itR = 1;
for k = 2:length(Ryx)
    isMakeXTest = ~ismember(k,indexes);
    if isMakeXTest
        nx = Ryx;
        ind = [indexes, k];
        disp(['Индексы проверяемых переменных x (с учётом y): ',
num2str(ind(1,:))]);
        disp('Проверяемые переменные');
        for m=1:length(ind)
            disp(['x(', num2str(x(1, ind(m)-1)), ')']);
        end
        copensator = 0;
        for i = 2:length(Ryx)
            delete = true;
            for j = 1:length(ind)
                if (i == ind(j))
                    delete = false;
                end
            end
            if (delete)
                nx(i-copensator, :) = [];
                nx(:, i-copensator) = [];
                copensator = copensator + 1;
            end
            if (length(nx)==length(indexes)+2)
                size_x = size(nx);
                n = size_x(1);
                p = size_x(2);
                disp(nx);
                [Ryx2, Ryx2scorr, Ryx2min] = calculate_Determinate_R(nx, n, p,
1);
                disp(['Коэфициент детерминации (Ryx2): ', num2str(Ryx2)]);
                if itR == 1
                    bestXindex = ind;
                    bestRxx = Ryx2;
                    bestMatrix = nx;
                elseif Ryx2>bestRxx
                    bestXindex = ind;
                    bestRxx = Ryx2;
                    bestMatrix = nx;
                end
                itR = itR + 1;
                break;
            end
        end
    end
end
```

calculate_Regression.m

```
function [Rxx, Rxy] = calculate_Regression(x, y)
xclean = x;
xclean(1,:) = [];
size_x = size(xclean);
n = size_x(1);
p = size_x(2);
```

```
Rxx = ones(p);
for s1 = 1:p
    for s2 = 1:p
        [r] = calculate_Coef_r(xclean(:,s1), xclean(:,s2), n);
        Rxx(s1,s2) = r;
    end
end
yWithx = [y, xclean];
Rxy = ones(p);
for s1 = 1:p+1
    for s2 = 1:p+1
        [r] = calculate_Coef_r(yWithx(:,s1), yWithx(:,s2), n);
        Rxy(s1,s2) = r;
    end
end
```

calculate_Student.m

```
function [tnabl, tkrit, V, znachimo] = calculate_Student(b, x, sigmakrb, alphaT)
xclean = x;
xclean(1,:) = [];
size_x = size(xclean);
n = size_x(1);
p = size_x(2);
%t-критерий
tnabl = zeros(length(sigmakrb), 1);
for i = 1:length(tnabl)
    tnabl(i) = b(length(b))/(sigmakrb(i,i));
end
V = n-p-1;
tkrit=tinv(1-alphaT, V);
znachimo = false;
c = 0;
for i = 1:length(tnabl)
    tn = abs(tnabl(i));
    if (tn>tkrit)
        c = c+1;
    end
end
if c == length(tnabl)
    znachimo = true;
end
```

calculate_SummAndAvg.m

```
function [ySumm, ySredn] = calculate_SummAndAvg(y)

ySumm = 0;
for i=1:length(y)
    ySumm = ySumm + y(i);
end
ySredn = ySumm/length(y);
```

calculate_table.m

```
function [yKr, sredn2, krysh, srednMkrysh2, krysh2, Qsumm, Qe, Qr] =
calculate_table(y, ySredn)
n = length(y);
yKr = zeros(1, n);
krysh = zeros(1, n);
krysh2 = zeros(1, n);
sredn2 = zeros(1, n);
srednMkrysh2 = zeros(1, n);
```

```
%(У - среднее) в квадрате
for i=1:n
    sredn2(i) = (y(i) - ySredn)^2;
end
sredn2 = transpose(sredn2);
%(У - с крышечкой среднее) в квадрате
for i=1:n
    srednMkrysh2(i) = (yKr(i) - ySredn)^2;
end
srednMkrysh2 = transpose(srednMkrysh2);
for i=1:n
    krysh(i) = y(i) - yKr(i);
end
krysh = transpose(krysh);
for i=1:n
    krysh2(i) = krysh(i)^2;
end
krysh2 = transpose(krysh2);
%Qsum
Qsumm = 0;
for i=1:n
    Qsumm = Qsumm + sredn2(i);
end
%Остаточная сумма квадратов
Qe = 0;
for i=1:n
    Qe = Qe + krysh2(i);
end
%Qr сумма квадратов регрессии
Qr = 0;
for i=1:n
    Qr = Qr + srednMkrysh2(i);
end
```

calculate_X.m

```
function [X] = calculate_X(x)
xclean = x;
xclean(1,:) = [];
sizeX = size(xclean);
X = ones(sizeX(1), sizeX(2)+1);
for i=1:sizeX(1)
    for j = 1:sizeX(2)
        X(i,j+1) = xclean(i, j);
    end
end
```

test_ Multicollinear.m

```
function [multicollinear, indexesForRxxOut] = test_Multicollinear(x, Rxx,
indexesForRxxIn, multicollinearAddition, alphaHi2)
size_x = size(Rxx);
n = size_x(1);
p = size_x(2);
mIndexesCount = 1;
mIndexes = zeros(1, 0);
multicollinear = false;
[hi2, hi2krit] = calculate_Hi2(x, Rxx, alphaHi2);
disp('Критерий Хи^2');
vs = 'Выбранная гипотеза: ';
if (hi2>hi2krit)
    disp([vs, 'H1 - Есть мультиколлинеарность']);
    multicollinear = true;
```

```
else
    disp([vs, 'H0 - Отсутствует мультиколлинеарность']);
end
if multicollinear
    for i=1:n
        for j=i+1:p
            rxxAbs = abs(Rxx(i, j));
            if ((rxxAbs<1)&&(rxxAbs>multicollinearAddition))
                disp(['x(',num2str(indexesForRxxIn(i)), ')', 'x(',
num2str(indexesForRxxIn(j)), ') = ', num2str(Rxx(i, j))]);
                write = false;
                for k=1:length(mIndexes)
                    if
((mIndexes(k)~=indexesForRxxIn(i))||mIndexes(k)~=indexesForRxxIn(j))
                        write = true;
                    end
                end
                if write || isempty(mIndexes)
                    mIndexes(mIndexesCount) = indexesForRxxIn(i);
                    mIndexesCount = mIndexesCount + 1;
                    mIndexes(mIndexesCount) = indexesForRxxIn(j);
                    mIndexesCount = mIndexesCount + 1;
                end
            end
        end
    end
    mIndexes = sort(mIndexes);
    mIndexes = unique(mIndexes);
    xString = '';
    for i=1:length(mIndexes)
        xS = strcat('x(', num2str(mIndexes(i)), ') ');
        xString = strcat(xString, xS);
    end
    if (~isempty(mIndexes))
        disp(['Переменные с мультиколлинеарностью: ', num2str(xString(1,:))]);
        input_x = zeros(1,100);
        cnt = 0;
        while 1
            in = input('Введите индекс удаляемой переменной x (0 - выход): ');
            if in == 0
                break;
            else
                cnt = cnt+1;
                input_x(cnt) = in;
            end
        end
        input_x = input_x(1:cnt);
        indexesForRxxOut = setdiff(indexesForRxxIn, input_x);
        disp(['Выбранные переменные: ', num2str(indexesForRxxOut(1, :))]);
    else
        disp('Переменные не найдены');
        indexesForRxxOut = indexesForRxxIn;
    end
else
    indexesForRxxOut = indexesForRxxIn;
end
```

removeXFromMAtrix.m

```
function [x] = removeXFromMAtrix(x, indexesForRxx)
%disp('Удаление переменных их матрицы x');
size_x = size(x);
p = size_x(2);
```

```
comp = 0;
for i=1:p
    delete = ~ismember(i, indexesForRxx);
    if (delete)
        x(:,i-comp) = [];
        comp = comp + 1;
    end
end
```

ПРИЛОЖЕНИЕ С

Статистические данные за период с 1998 по 2019 год

Период	Y	X2	X10
1998	2629,6	58464	99
1999	4823,2	63633	83
2000	7305,6	65070,4	120,9
2001	8943,6	65122,9	120
2002	10830,5	66658,9	116,2
2003	13080,2	66339,4	110,9
2004	17027,2	67318,6	110,6
2005	21609,8	68339	112,6
2006	26917,2	69168,7	113,3
2007	33247,5	70770,3	117,2
2008	41276,8	71003,1	111,5
2009	38807,2	69410,5	96,5
2010	46308,5	69933,7	105,2
2011	60282,5	70856,6	102,8
2012	68163,9	71545,4	108,4
2013	73133,9	71545,4	101,2
2014	79199,7	71539	91
2015	83232,6	72324	100,8
2016	86010,2	72755	102,9
2017	92089,3	72142	108,5
2018	103626,6	72354	102,9
2019	110046,1	71933	107,5