

Санкт-Петербургский государственный университет

ПРОХОРЕНКО Филипп Александрович

Выпускная квалификационная работа

Моделирование многомерного динамического процесса с вероятностной классификацией текущего состояния

Уровень образования:

Направление *01.04.02 «Прикладная математика и информатика»*

Основная образовательная программа *ВМ.5504 «Исследование операций и системный анализ»*

Научный руководитель:
кафедра МЭС
д.ф.-м.н., профессор
Прасолов Александр
Витальевич

Рецензент:
Руководитель направления
отделения Сбербанка,
к.ф.-м.н.,
Замураев Константин
Александрович

Санкт-Петербург
2020

Оглавление

Введение.....	3
Постановка задачи	4
Обзор литературы	5
Глава 1. Предварительная обработка данных.....	8
1.1. Отбор показателей и их визуальный анализ	8
1.2. Визуальный анализ и учет инфляции.....	13
1.3 Нормализация отклонений	27
Глава 2. Робастность показателей	30
2.1 Тест Грэнджера	30
2.2 Применение теста Грэнджера.....	32
2.3 Робастность влияющих показателей	36
Глава 3. Прогнозирование притока/оттока инвестиций.....	39
3.1 Подбор моделей	39
3.2 Интервалы инвестирования	47
3.3 Итоговый алгоритм.....	52
Выводы.....	56
Заключение.....	58
Список литературы	59

Введение

Благосостояние любой отрасли экономики региона напрямую зависит от объема инвестиций, который был вложен в эту отрасль. Чем больше денежных средств пошло на развитие хозяйства, чем больший «положительный рост» данная отрасль покажет с течением времени.

При управлении регионом (областью, городом, округом, страной) органам исполнительной власти, как лицам принимающим решение, необходимо выбрать стратегию о том, как распределять инвестиции между отраслями экономики. Ведь недостаток вложенных средств в определенное хозяйство может повлечь за собой катастрофические последствия в виде упадка отрасли, потери рабочих мест и ухудшению качества жизни людей. Конечно, в подавляющем большинстве случаев, до таких кардинальных последствий не доходит, но это не умаляет значимость правильного планирования потока денег из бюджета.

Если рассмотреть разные виды экономической деятельности то можно заметить, что в некоторые из них вкладываются большие инвестиции, и они гораздо активнее развиваются (например: ядерная отрасль, нефтегазовая промышленность), а в некоторые меньше, и они либо стагнируют, либо их динамика развития отрицательная. Проблема заключается в том, что экономика состоит из огромного числа всевозможных отраслей, которые друг с другом очень тяжело сопоставить, то есть сравнить, как изменение числа инвестиций в одной сфере повлияет на другую отрасль. Данная работа предлагает алгоритм для лица, принимающего решение, который показывает, как распределить инвестиций между отраслями таким образом, чтобы улучшить ситуацию в одних отраслях, но при этом не допустить ухудшения в других

Постановка задачи

Задача диссертации состоит в том, чтобы разработать математический алгоритм, который поможет органам исполнительной власти принять решение о распределении бюджетных средств между отраслями экономики города Санкт-Петербурга. В текущих реалиях существует различное множество методик о том, как лица, ответственные за перенаправление потока инвестиций, распределяют его по отраслям. Для такого типа задача важно правильно задать метрику, которая оценивает качество построенного алгоритма[1]. Методика, предложенная в данной работе, использует в качестве метрики, показывающей, как изменяются инвестиции между отраслями, наклон линейного тренда для показателя, который характеризует данную отрасль. Каждая отрасль в работе описывается показателем, который относится к ней. Для каждого показателя находится группа самых влияющих на него.

Построенный алгоритм позволит количественно оценить, как изменится значение показателя, характеризующего отрасль, при оптимальном перераспределении инвестиций между отраслями, которые больше всего влияют на него. Эта величина и покажет какой эффект даст для отрасли такое перераспределение.

Причинные связи между показателями построены с помощью статистического теста Грэнджера. Количественная оценка эффективности перераспределения рассчитана с помощью моделей множественной линейной регрессии.

Обзор литературы

Для получения достоверных статистических данных по всем видам экономической деятельности была использована Единая межведомственная информационно-статистическая система (ЕМИСС)[2]. Платформа позволяет получить быстрый доступ к любой статистической информации, касающейся социально-экономического развития, как страны, так и любого региона. Сервис позволяет выгрузить данные для любых показателей за разные промежутки времени (год, квартал, месяц). С помощью системы я мог в кратчайшие сроки найти всю необходимую статистику по нужным мне показателям и загрузить в удобном для меня формате для дальнейшего анализа данных.

Федеральная служба государственной статистики – это российский федеральный орган исполнительной власти, осуществляющий функции по формированию официальной статистической информации о социальных, экономических, демографических, экологических и других общественных процессах в Российской Федерации [3]. Официальный сайт службы содержит определение каждого социально-экономического показателя, его подробное описание, а также подробную информацию о сферах его применения и важности его для отрасли экономики. Используя этот источник, я сформировал четкую картину того, как устроена каждая отрасль, какие показатели являются для нее определяющими, что позволило выбрать показатели, участвующие в данной работе.

В книге Joel Grus «Data science from Scratch» описаны основные методы анализа и визуализации данных с использованием языка программирования Python. Книга разделена на 3 основные главы, которые посвящены математическому анализу и статистике; сбору, обработке и хранению данных и машинному обучению соответственно. В книге описано большое количество статистических методов, один из которых (тест Грэнджера) был

использован в работе. Книга раскрывает базовые понятия моделей машинного обучения, их преимущества и недостатки и на примерах показывает, в каких задачах какой метод будет более приемлем, и покажет более точный результат.

Книга Wes McKinney «Python for Data Analysis» является современным практическим руководством по разработке научных приложений на языке программирования Python. Данная книга повествует о том, как разрабатывать оптимизированные программы на Python для решения задач анализа данных. В книге разобраны основные библиотеки, помогающие производить всевозможные математические вычисления между переменными, показывает, как максимально эффективно подготовить данные для последующего их применения в моделях машинного обучения. Также, в этой книге представлены техники визуализации данных и построения графиков, ускоряющие анализ данных.

Книга Прасолова А.В. «Математические методы экономической динамики» описывает ту часть экономической теории, которая устанавливает причины изменений в экономике, основываясь на количественных оценках. В книге подробно изложены математические методы, такие как методы исследования линейных и нелинейных уравнений, методы приближения, анализ влияния временных лагов, задачи идентификации и прогнозирования, которые позволяют решать всевозможные экономические задачи. Книга содержит практические примеры, в которых применяются изложенные методы с использованием реальных данных. На мой взгляд, данная книга показывает, как с помощью математических методов принимать управленческие решения в экономике, что свидетельствует о наличии тесной связи между двумя сферами.

Книга Кистанова В.В. и Копылова Н.В. «Региональная экономика России» описывает особенности развития хозяйства экономических районов и зон. По каждому макрорегиону раскрываются основные черты хозяйства,

рыночная обстановка, а также перспективы развития. Данная книга дает представление о том, как функционируют регионы страны, как развивается каждая из отраслей, какие есть в них особенности.

Глава 1. Предварительная обработка данных

1.1. Отбор показателей и их визуальный анализ

Экономика каждого субъекта Российской Федерации состоит из нескольких сотен показателей, каждый из которых характеризует ту или иную сторону жизни региона. Некоторые показатели, такие как, например, «Валовый региональный продукт», «Оборот инвестиций в основной капитал», «Численность постоянного населения на начало года» играют важнейшую роль в благополучии региона. Их «положительное» изменение с течением времени, будь то рост, будь то падение, позволяет идентифицировать динамику развития социально-экономического объекта [4]. Рассмотрим на примере, почему для некоторых показателей хорошо, когда их значение уменьшается, что на первый взгляд кажется странным. Возьмем показатель – «Уровень преступности на 100 тысяч человек», который рассчитывается по формуле:

$$level = \frac{c * 10000}{N},$$

где *level* – уровень преступности на 100 тысяч человек населения

c – количество преступлений

N – численность населения

Наблюдается прямо пропорциональная зависимость между показателем «Уровень преступности на 100 тысяч человек» и количеством совершенных преступлений. Соответственно, при увеличении числа преступлений будет наблюдаться увеличение значения рассматриваемого показателя при условии, что численность населения не изменилась. Чем меньше число преступлений в регионе, тем более процветающим он выглядит, что и является первоочередной целью при управлении субъектом. Таким образом,

позитивная динамика изменения показателя заключается в уменьшении его значения с течением времени.

Помимо показателей, играющих важную роль и определяющих экономику региона, существуют факторы, которые не имеют сильного влияния на субъект в целом, но существенны для какой-то определенной отрасли. Например, такие показатели, как «Незавершенное строительство/Количество приостановленных объектов» и «Индекс цен строительной продукции» играют колоссальную роль в принятии решений касательно отрасли «Строительство», так как позволяют понять, например, как ведет себя стоимость строительной продукции и в зависимости от ее цены выбрать оптимальное количество товара, но для экономики в целом эти знания дадут крайне малый эффект [5].

В связи с огромным количеством факторов в данной работе я рассмотрю 10 показателей, которые, на мой взгляд, в достаточной мере описывают экономику региона. Выбранные показатели обобщают основные сферы жизнедеятельности региона: социально-экономическую, социально-политическую, социально-демографическую, производственно-экономическую, финансово-кредитную сферы. Все последующие суждения несложно обобщить для любого числа показателей, что в будущем я и планирую сделать. Я рассматриваю в работе экономику Санкт-Петербурга, поэтому мною были взяты данные по Санкт-Петербургу для следующих 10 показателей:

1. Оборот розничной торговли / всего (млн. руб.)
2. Численность рабочей силы (экономически активного населения) / всего (тыс. чел.)
3. Денежные доходы населения / всего (млн. руб.)
4. Общая площадь жилищного фонда (тыс. м²)
5. Валовой региональный продукт в основных ценах / всего (млрд. руб.)

6. Ввод жилых домов нарастающим итогом с начала года / тыс. кв. м (тыс. м²)
7. Инвестиции в основной капитал на душу населения нарастающим итогом с начала года (млн. руб.)
8. Общая заболеваемость населения на 1000 чел. (чел.)
9. Численность постоянного населения, на начало периода / всего (тыс. чел.)
10. Индекс потребительских цен годовой / сводный (%)

Экономическая ситуация в любом регионе очень не стабильна. Политическая обстановка может кардинально поменяться буквально за 10 лет, поэтому статистические данные, взятые с большим временным промежутком между ними, не могут быть одинаково интерпретируемы. Например, денежные доходы населения в 1990 году и в 2018 году имеют серьезные различия, обусловленные тяжелой ситуацией в стране в 90-ые годы. Поэтому всевозможные зависимости, полученные с помощью математических методов, не могут быть трактованы в верном ключе. В связи с этим, мною были взяты статистические данные за период 2000-2018 г. По некоторым показателям сбор статистики задерживается, поэтому, чтобы использовать реальные данные, я рассматриваю данные до 2018 года включительно. Все данные были взяты единой межведомственной информационно-статистической системы (ЕМИСС) [2].

В Таблице 1 представлены значения выбранных показателей за указанный временной промежуток.

Таблица 1.

Год	Оборот розничной торговли / всего (млн. руб)	Численность рабочей силы (экономически активного населения) / всего (тыс. чел.)	Денежные доходы населения / всего (млн. руб)	Общая площадь жилищного фонда (тыс. м ²)	Валовой региональный продукт в основных ценах / всего (млрд руб)	Ввод жилых домов нарастающим итогом с начала года / тыс. кв. м (тыс. м ²) (хб)	Инвестиции в основной капитал на душу населения нарастающим итогом с начала года (млн руб)	Общая заболеваемость населения на 1000 чел. (чел.)	Численность постоянного населения, на начало периода / всего (тыс. чел.)	Индекс потребительских цен годовой / сводный (%)
2000	86493	2451,8	145000	93471	188,2	1080,9	29029,5	1331,8	4741,9	123,5
2001	112300	2465,9	193300	94000	251,7	1118,2	47145	1324,1	4714,8	118,1
2002	139300	2543,3	253100	96000	336,7	1213,7	76045,6	1326,7	4688,4	114,7
2003	161878	2516,3	381500	97000	409,6	1757,9	111677,9	1316,1	4656,5	112,2
2004	215893	2552,8	507800	98650	542,4	2032	112944	1426,0	4662,0	112,7
2005	292504	2619,5	675700	100326	666,4	2273,4	156854	1700,5	4686,0	112
2006	357373	2725,1	776900	102547	825,1	2376	177994	1846,4	4713,0	110
2007	447928	2812,1	926200	106000	1109,3	2637	303448	1900,0	4748,0	113,2
2008	582235	2808,8	972400	108000	1431,8	3212	372637	1910,0	4765,0	114,4
2009	605691	2847,3	1219300	109937	1475,8	2603,2	324710	2155,5	4799,0	108,5
2010	685051	2807,3	1397500	112875	1699,5	2656,5	401537	2182,3	4833,0	109,4
2011	742104	2857,9	1536400	115417	2091,9	2705,7	360368	2118,4	4899,3	105,9

2012	84475 9	2895,9	16669 00	11967 5	2280,4	2577	352116	2151, 8	4953,2	106,1
2013	92072 1	2848,9	19146 00	11967 5,2	2491,4	2584	475149	2178, 9	5028,0	106,7
2014	10176 23	2885,1	21530 00	11100 0	2661,2	3262	523331	2206, 0	5132,0	113,3
2015	11446 07	2967,2	24905 00	12310 0	3387,4	3030 ,7	483423	2276, 5	5191,7	113,2
2016	12343 24	3020,3	25952 00	12830 0	3666,0	3116	678642	2388, 0	5225,7	105,2
2017	13263 16	3048,6	26881 00	13310 0	3866,4	3536	672365	2369, 0	5281,6	103,7
2018	14124 00	3061,1	28986 00	13676 2	4193,5	3950 ,3	747407	2430, 0	5351,9	103,9

1.2. Визуальный анализ и учет инфляции

Инфляция – это общий прирост цен в стране с течением времени. Вместе с инфляцией говорят об обесценивании денег, то есть за одинаковую сумму со временем можно купить меньшее количество товара. Причины инфляции в большинстве своем связаны с количеством денег в экономике и их доступностью, с так называемыми монетарными факторами. Также, есть другая группа факторов, называемая немонетарные, к которой можно отнести: погоду, влияющую на урожай, конкуренцию на рынке, курсы валют и другое.

Для приведения данных к единому формату необходимо убрать инфляцию из всех показателей, содержащих денежную составляющую. Таких показателей в работе четыре: «Оборот розничной торговли / всего», «Денежные доходы населения / всего», «Валовой региональный продукт в основных ценах / всего», «Инвестиции в основной капитал на душу населения нарастающим итогом с начала года». За базовый год был взят первый отчетный год - 2000 год.

Дадим краткую интерпретацию каждого из выбранных показателей и проведем визуальный анализ за рассматриваемый промежуток времени. Также определим для каждого показателя его положительную динамику изменения, и установим, как ведет себя линейный тренд для рассматриваемых временных рядов.

1. «Оборот розничной торговли/ всего» - выручка от продажи товаров населению для личного потребления или использования в домашнем хозяйстве за наличный расчет или оплаченных по кредитным карточкам, расчетным чекам банков, по перечислениям со счетов вкладчиков, по поручению физического лица без открытия счета, посредством платежных карт. Стоимость товаров, проданных (отпущенных) отдельным категориям населения со скидкой, или полностью оплаченных органами социальной защиты, включается в оборот розничной торговли в полном объеме. В оборот

розничной торговли не включается стоимость товаров, отпущенных из розничной торговой сети юридическим лицам (в том числе организациям социальной сферы, спецпотребителям и т.п.) и индивидуальным предпринимателям, и оборот общественного питания [3]. Соответственно, данный фактор экономики показывает уровень платежеспособности населения страны. Поэтому положительной динамикой его изменения является увеличение значения от года к году. В работе показатель измеряется в миллионах рублей.

График показателя «Оборот розничной торговли» представлен на рисунке 1.



Рисунок 1.

При анализе графика временного ряда можно заметить положительную тенденцию, заключающуюся в постепенном увеличении оборота розничной торговли, начиная с 2000 (первого отчетного) года. Построим линейный тренд. Он также возрастает с течением времени, наклон линейного тренда (коэффициент при x) положительный.

В данной работе изменение наклона тренда обозначает изменение количества инвестиций, вкладываемых в отрасль, поэтому оборот розничной

торговли может быть использован, как фактор, представляющий отрасль из которой можно взять инвестиции для поддержания других отраслей.

2. «Численность рабочей силы (экономически активного населения)».

Экономически активное население - часть населения, обеспечивающая предложение рабочей силы для производства товаров и услуг. Численность экономически активного населения включает занятых и безработных. Занятые лица в экономике – лица мужского и женского пола в возрасте 16 лет и старше. Лица младше 16 лет относятся к данной категории, только если они:

а) Выполняли работу по найму за вознаграждение на условиях полного или неполного рабочего времени, а также за любую иную работу приносящую доход этим лицам

б) временно отсутствовали на работе по причине болезни, травмы, отпуска, отгулов, возмещения сверхурочных работ и так далее

в) выполняли работу без оплаты на семейном предприятии.

Безработные лица в экономике – лица мужского и женского пола в возрасте 16 лет, которые в рассматриваемый период: не имели работу, либо занимались поиском работы, либо готовы были приступить к работе. К этой группе лиц также относятся студенты, которые ведут поиск работы [3].

Увеличение количества работающего населения стимулирует экономику расти. В связи с этим, для данного показателя положительной динамикой его изменения будет увеличение значения от года к году. В работе единица измерения показателя – тысячи человек.

График показателя «Численность рабочей силы (экономически активного населения)» представлен на рисунке 2.



Рисунок 2.

Визуальный анализ графика позволяет заметить, что на протяжении всего рассматриваемого периода (2000-2018) наблюдается небольшой волнообразный рост показателя, к 2018 году значение показателя составило 3061,1 тысяч человек, что на 609,3 тысячи человек (24,8%) больше, чем в базовом 2000 году. Линейный тренд для временного ряда имеет положительный наклон.

3. «Денежные доходы населения» - это сумма денежных средств, получаемые домохозяйствами за определенный промежуток времени и предназначенных для приобретения благ и услуг в целях личного потребления. В зависимости от циклов жизнедеятельности человека доходы подразделяются на получаемые:

- a. до участия в трудовой деятельности (до достижения трудоспособного возраста)
- b. от участия в трудовой, предпринимательской, общественной деятельности
- c. временно неработающими (безработными, беженцами и т.д.)

d. после завершения трудовой деятельности (пенсионерами)

Все доходы населения можно подразделить на две большие группы: заработная плата и не денежные доходы.

Увеличение денежных доходов населения увеличит покупательную способность граждан региона, что стимулирует предприятия производить больше продукции. А это в свою очередь приводит к росту экономики. В связи с этим, для данного показателя положительной динамикой его изменения будет увеличение значения от года к году. В работе единица измерения показателя – миллион рублей.

График показателя «Денежные доходы населения» представлен на рисунке 3.



Рисунок 3.

Визуальный анализ графика позволяет заметить, что на протяжении всего рассматриваемого периода (2000-2018) наблюдается сильный рост показателя даже в перерасчете на деньги 2000 года.

4. «Общая площадь жилищного фонда». Для начала дадим пояснение, что подразумевается под жилищным фондом. *Жилищный фонд* – это совокупность всех жилых помещений независимо от форм собственности, включая жилые дома, специализированные дома

(общежития, приюты, дома маневренного фонда, специальные дома для одиноких престарелых, детские дома, дома-интернаты для инвалидов, ветеранов, интернаты при школах и школы-интернаты), квартиры, служебные жилые помещения, иные жилые помещения в других строениях, пригодные для проживания. По формам собственности жилищный фонд подразделяется на частный, государственный, муниципальный, общественный, а также жилищный фонд в коллективной собственности [3]. Тогда по *общей площади жилищного фонда* понимается полная сумма площадей всей совокупности жилых помещений, описанных в определении *жилищного фонда*.

Увеличение общей площади жилищного фонда ведет к повышению зданий, которые могут быть использованы гражданами для проживания, что приводит к улучшению качества жизни. Также, новые сооружения могут быть использованы, например, предприятиями для повышения их производительности или развитие новых технологий. Таким образом, увеличение значений данного показателя с течением времени стимулирует экономику и является положительной динамикой его изменения. В работе, единица измерения показателя – тыс. м².

График показателя «Общая площадь жилищного фонда» представлен на рисунке 4.

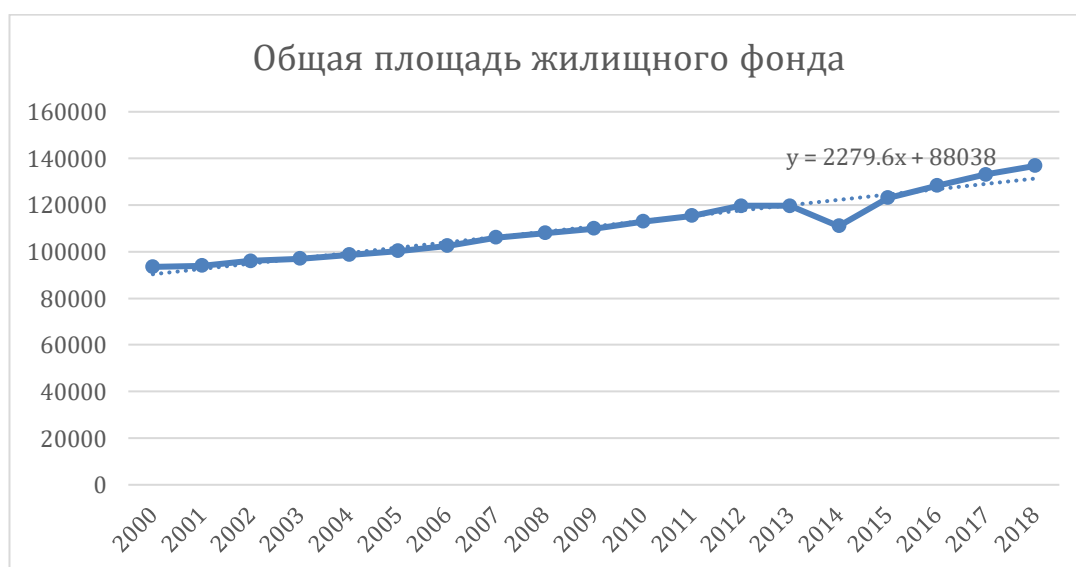


Рисунок 4.

На протяжении всего ретроспективного периода наблюдается постепенный рост показателя за исключением небольшого падения значения фактора в 2014 году. Линейный тренд у временного ряда возрастающий.

5. «Валовой региональный продукт в основных ценах» представляет собой стоимость товаров и услуг, произведенных для конечного использования. Валовой региональный продукт рассчитывается производственным методом как разница между выпуском и промежуточным потреблением [3].

Выпуск товаров и услуг представляет собой суммарную стоимость товаров и услуг, являющихся результатом производственной деятельности единиц-резидентов региональной экономики в отчетном периоде. Реализованные товары и услуги включаются в выпуск по фактической рыночной стоимости, нереализованная готовая продукция – по средним рыночным ценам. Изменение незавершенного производства включается в валовой выпуск по себестоимости.

Промежуточное потребление состоит из стоимости товаров и услуг, которые трансформируются или полностью потребляются в процессе производства в отчетном периоде [3].

Валовой региональный продукт (ВРП) является одним из основных показателей, оценивающих эффективность экономики региона. Прежде всего, ВРП – это показатель произведенного продукта, который представляет собой стоимость произведенных конечных товаров и услуг.

Увеличение конечного производства товаров и услуг является одной из главных задач экономики, поэтому рост показателя служит критерием эффективного развития экономики. В связи с этим, положительной динамикой изменения значений показателя является повышение значений фактора с течением времени.

График показателя «Валовой региональный продукт в основных ценах» представлен на рисунке 5.

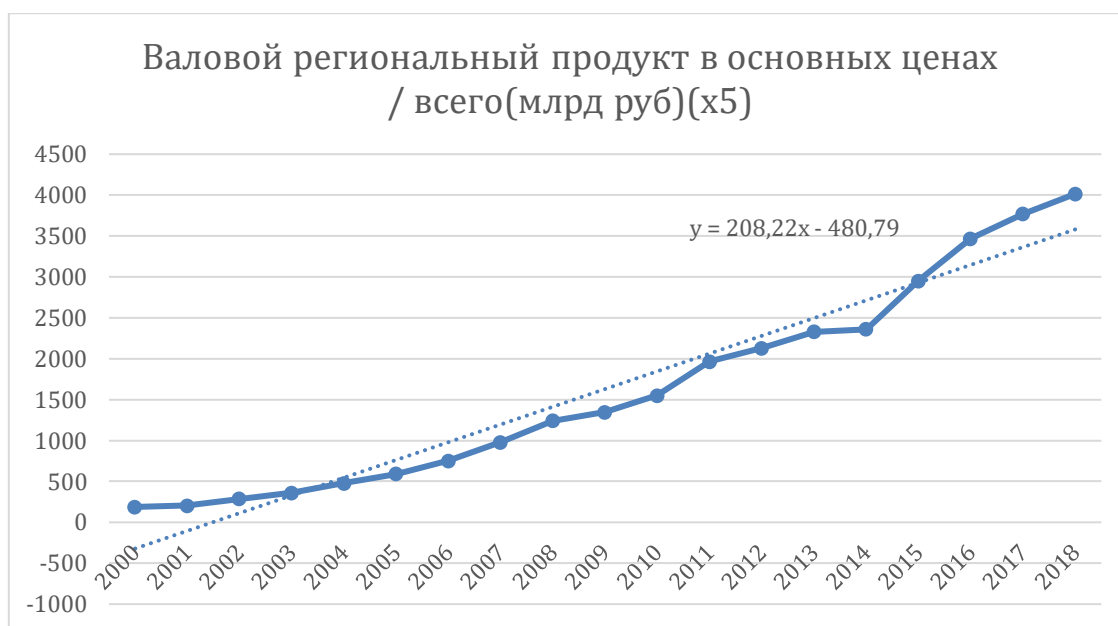


Рисунок 5.

На протяжении всего ретроспективного периода (2000-2018 года) наблюдается постепенный рост показателя. Линейный тренд для временного ряда положительный. Это означает, что отрасль развивается в правильном направлении.

6. «Ввод жилых домов нарастающим итогом с начала года» - это общая площадь, введенных в действие жилых домов на территории региона. Термин «ввести в действие» означает начать хозяйственное использование, применение вновь построенных, созданных объектов строительства, основных средств [3].

Увеличение введенных в действие жилых домов приведет к повышению количества помещений, которые могут быть использованы, например, для более комфортного проживания граждан региона, что в свою очередь стимулирует многие отрасли экономики к росту. В работе, положительной динамикой изменения показателя с течением времени является его рост. Единица измерения показателя – тыс. м².

График показателя «Ввод жилых домов нарастающим итогом с начала года» представлен на рисунке 6.



Рисунок 6.

На протяжении всего отчетного периода (2000-2018 год) наблюдается скачкообразный рост показателя и к 2018 году его значение составит 3950,3 тыс. кв.м. Линейный тренд для временного ряда имеет положительный наклон.

7. «Инвестиции в основной капитал на душу населения нарастающим итогом с начала года». Дадим определение инвестиций. *Инвестиции* - денежные средства, ценные бумаги, иное имущество, в том числе имущественные права, иные права, имеющие денежную оценку, вкладываемые в объекты предпринимательской и (или) иной деятельности в целях получения прибыли и (или) достижения иного полезного эффекта. Исходя из этого определения, инвестиции в основной капитал представляют собой затраты на строительство, реконструкцию (включая расширение и модернизацию) объектов, которые приводят к увеличению их первоначальной стоимости, приобретение машин, оборудования, транспортных средств, производственного и хозяйственного инвентаря, бухгалтерский учет которых осуществляется в порядке, установленном для учета вложений во внеоборотные активы, инвестиции в объекты

интеллектуальной собственности; культивируемые биологические ресурсы. К инвестициям в основной капитал относятся затраты на создание и приобретение новых основных средств, а также поступивших по импорту, осуществляемые за счет всех источников финансирования, включая средства бюджетов на возвратной и безвозвратной основе, кредиты, техническую и гуманитарную помощь [3].

Увеличение инвестиций в основной капитал приводит к повышению количества свободных денежных средств для перенаправления их в отрасли экономики. Дополнительные деньги позволяют экономике развиваться в более быстрой темпе, что и является целью для людей, управляющих экономической деятельностью. Единица измерения показателя – миллионы рублей.

График показателя «Инвестиции в основной капитал на душу населения нарастающим итогом с начала года» представлен на рисунке 7.



Рисунок 7.

К 2018 году наблюдается постепенное увеличение значений показателя до 715492,72 млн. руб. Линейный тренд для данного временного ряда возрастающий. Наклон линейного тренда положительный.

8. «Общая заболеваемость населения на 1000 чел.». Данный показатель рассчитывается как отношение числа вновь возникших заболеваний (заболеваний впервые установленным диагнозом) к средней численности населения, умноженное на 1000. Единицей наблюдения при изучении общей заболеваемости в текущем календарном году по поводу данного заболевания являются: первичное обращение больного в лечебное учреждение и первое обращение по поводу хронического заболевания выявленного в предыдущие годы. Хронические и длительно протекающие заболевания учитываются только раз в году.[3]

Позитивной динамикой изменения данного показателя является снижение общей заболеваемости населения, что приведет к повышению здоровья граждан и улучшению их качества жизни. Фактор измеряется в «чел.».

График показателя «Общая заболеваемость населения на 1000 чел.» представлен на рисунке 8.

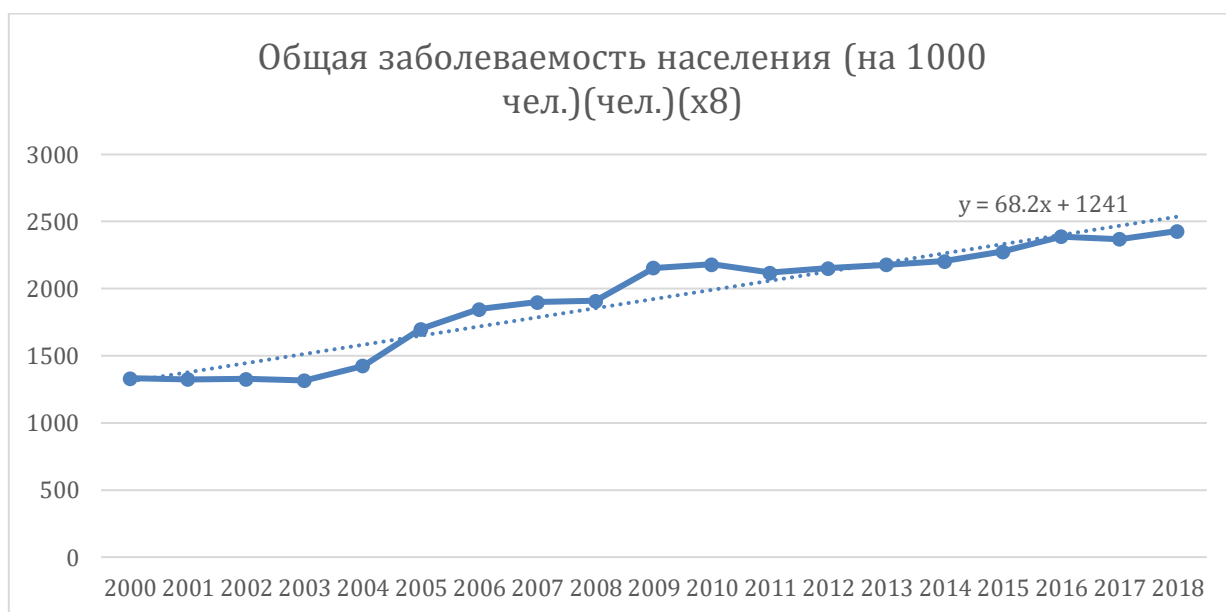


Рисунок 8.

На протяжении всего отчетного периода наблюдается постепенное увеличение значений показателя, что является негативной тенденцией его изменения. Линейный тренд возрастающий, наклон линейного тренда

положительный. Необходимо, чтобы значение показателя снижалось с течением времени.

9. «Численность постоянного населения на начало года» - это количество людей, проживающих в данном регионе на начало года.

Рассчитывается значение данного показателя по следующей формуле:

$$S_{t+1} = S_t + N_t - M_t + П_t - B_t, \text{ где}$$

S_t и S_{t+1} – численность населения на начало года t и $t + 1$

N_t – число родившихся в году t

M_t – число умерших в году t

$П_t$ – число прибывших на данную территорию в году t

B_t – число выбывших с данной территории в году t

Задача региона состоит в увеличении численности его населения, так как это, например, приводит к увеличению числа рабочей силы, а также это показывает престиж региона (если количество мигрантов увеличивается, то регион привлекателен для проживания) [3]. Положительной динамикой изменения показателя с течением времени является его рост. Единица измерения показателя – «чел.».

График показателя «Численность постоянного населения на начало года» представлен на рисунке 9.



Рисунок 9.

Начиная с 2003 года, наблюдается постепенное увеличение численности населения на начало года в городе Санкт-Петербурге. Линейный тренд возрастающий, наклон линейного тренда положительный.

10. «Индекс потребительских цен/годовой» измеряет среднее изменение цен, уплачиваемых потребителями за корзину товаров и услуг. Показатель количественно оценивает совокупный уровень цен в экономике, и таким образом измерить покупательную способность денежной единицы страны. Величина >100 означает, что цены за расчетный период выросли. Величина <100 означает, что цены за расчетный период снизились [3].

Чем меньше значения индекса потребительских цен, тем медленнее растут цены на товары и услуги. Таким образом, положительной динамикой изменения показателя с течением времени является его снижение. Единица измерения показателя - %.

График показателя «Индекс потребительских цен/годовой» представлен на рисунке 10.



Рисунок 10.

На протяжении всего отчетного периода наблюдается постепенное снижение значений показателя. В 2018 году значение индекса

потребительских цен составило 103,9%. Линейный тренд убывает, наклон линейного тренда отрицательный.

1.3 Нормализация отклонений

Для последующего проведения теста Грэнджера[3] на причинность для рассматриваемых временных рядов необходимо привести данные к одному порядку. В предыдущем параграфе денежные единицы были трансформированы в соответствии с инфляцией. Рассматривать будем совокупность преобразованных временных рядов.

Для каждого временного ряда посчитаем линейный тренд.

Формула линейного тренда имеет следующий вид:

$$\hat{y}_i = a_i + b_i * \hat{x}_i, \text{ где}$$

\hat{y}_i – значение тренда в i – ый момент времени

a_i – свободный член

b_i – угловой коэффициент

\hat{x}_i – значение временного ряда в i – ый момент времени

Значение тренда для каждого из временных рядов:

1. Оборот розничной торговли / всего (млн. руб.)

$$\hat{y}_i = -131551 + 72898 * \hat{x}_i$$

2. Численность рабочей силы (экономически активного населения) / всего (тыс. чел.)

$$\hat{y}_i = 2435,7 + 34 * \hat{x}_i$$

3. Денежные доходы населения / всего (млн. руб.)

$$\hat{y}_i = -284582,8 + 151488,6 * \hat{x}_i$$

4. Общая площадь жилищного фонда (тыс. м2)

$$\hat{y}_i = 88038 + 2279,6 * \hat{x}_i$$

5. Валовой региональный продукт в основных ценах / всего (млрд. руб.)

$$\hat{y}_i = -480,8 + 208,2 * \hat{x}_i$$

6. Ввод жилых домов нарастающим итогом с начала года / тыс. кв. м (тыс. м²)

$$\hat{y}_i = 1218,8 + 129,3 * \hat{x}_i$$

7. Инвестиции в основной капитал на душу населения нарастающим итогом с начала года (млн. руб.)

$$\hat{y}_i = -55020 + 36526 * \hat{x}_i$$

8. Общая заболеваемость населения на 1000 чел. (чел.)

$$\hat{y}_i = 1241 + 68,2 * \hat{x}_i$$

9. Численность постоянного населения, на начало периода / всего (тыс. чел.)

$$\hat{y}_i = 4515,9 + 38,2 * \hat{x}_i$$

10. Индекс потребительских цен годовой / сводный (%)

$$\hat{y}_i = 117,9 - 0,7 * \hat{x}_i$$

Для нормализации отклонений от линейного тренда выполним следующие шаги:

- Вычтем линейный тренд из каждого временного ряда
- Посчитаем для полученного ряда выборочное среднее и среднеквадратичное отклонение
- Вычтем для каждого значения ряда его выборочное среднее
- Поделим полученную разность на среднеквадратичное отклонение

После выполнения выше описанной процедуры для всех 10 показателей, получились временные ряды (с округлением до второго знака после запятой), представленные в Таблице 2.

Таблица 2.

Оборот розничной торговли / всего (млн. руб)	Численность рабочей силы (экономически активной) / населения / всего (тыс. чел.)	Денежные доходы населения / всего (млн.руб)	Общая площадь жилищного фонда (тыс м2)	Валовой региональный продукт в основных ценах / всего (млрд руб)	Ввод жилых домов нарастающим итогом с начала года / тыс. кв. м (тыс м2)(х6)	Инвестиции в основной капитал на душу населения нарастающим итогом с начала года (млн руб)	Общая заболеваемость населения на 1000 чел. (чел.)	Численность постоянного населения, на начало периода / всего (тыс. чел.)	Индекс потребительских цен годовой / сводный (%)
2,21	-0,38	2,07	0,95	1,55	-0,82	0,93	0,20	2,16	1,96
1,18	-0,80	1,04	0,42	0,86	-1,10	0,40	-0,46	1,40	0,49
0,48	0,12	0,34	0,34	0,41	-1,20	0,20	-1,03	0,66	-0,35
-0,27	-1,17	0,11	-0,05	-0,07	0,07	0,14	-1,72	-0,14	-0,91
-0,65	-1,12	-0,18	-0,24	-0,39	0,51	-0,55	-1,36	-0,52	-0,53
-0,69	-0,43	-0,17	-0,42	-0,73	0,85	-0,48	0,44	-0,68	-0,53
-0,82	1,09	-0,51	-0,44	-0,91	0,77	-0,76	1,11	-0,81	-0,93
-0,87	2,22	-0,83	-0,08	-0,85	1,18	0,59	0,98	-0,85	0,28
-0,30	1,43	-1,76	-0,17	-0,64	2,54	0,97	0,48	-1,09	0,88
-0,69	1,52	-0,88	-0,27	-1,02	0,28	-0,28	2,02	-1,14	-0,74
-0,69	-0,05	-0,80	-0,07	-1,03	0,05	0,38	1,66	-1,19	-0,24
-0,71	0,31	-0,68	0,01	-0,29	-0,20	-0,88	0,51	-0,87	-1,11
-0,41	0,39	-0,95	0,60	-0,44	-0,99	-1,78	0,21	-0,69	-0,83
-0,42	-1,33	-0,34	-0,08	-0,47	-1,36	-0,23	-0,15	-0,27	-0,42
-0,91	-1,28	-0,59	-3,38	-1,12	0,32	-0,57	-0,50	0,49	1,86
-0,58	-0,26	0,22	-0,42	0,25	-0,79	-2,12	-0,48	0,73	2,04
0,92	0,15	1,23	0,45	1,37	-0,92	1,49	-0,11	0,68	-0,23
1,71	0,02	1,33	1,21	1,70	-0,03	1,04	-0,87	0,88	-0,48
1,50	-0,43	1,35	1,63	1,83	0,84	1,50	-0,93	1,25	-0,20

Такая процедура называется *нормализацией* отклонений. Она приводит дисперсии к одному порядку. Для получившихся новых показателей проведем тесты Грэнджера на причинность для выявления связей между ними. Речь об этом пойдет в следующей главе.

Глава 2. Робастность показателей

2.1 Тест Грэнджера

Для нахождения для каждого фактора группы факторов, который больше всего на него влияют, воспользуемся тестом Грэнджера на причинность [5].

Тест Грэнджера позволяет обнаружить причинные связи между временными рядами. Идея теста заключается в том, что значения одного временного ряда, являющегося причиной изменений другого временного ряда, должны предшествовать изменениям этого временного ряда и кроме того вносить весомый вклад в прогноз значений этого ряда. Это отражается в гипотезах данного теста. Тест Грэнджера проверяет нулевую гипотезу о том, что фактор не является причиной другого фактора и для каждой регрессии суть гипотезы в том, что коэффициенты при лагах второй переменной одновременно равны нулю.

Рассмотрим 2 временных ряда y_t и x_t .

$$y_t = a_0 + a_1 y_{t-1} + \dots + a_p y_{t-p} + b_1 x_{t-1} + \dots + b_p x_{t-p} + \varepsilon_t$$

$$x_t = c_0 + c_1 x_{t-1} + \dots + c_p x_{t-p} + d_1 y_{t-1} + \dots + d_p y_{t-p} + u_t$$

Для каждой авторегрессионной модели проверяются 2 нулевые гипотезы о равенстве нулю коэффициентов при значениях второго временного ряда:

$$H_0^1 : b_1 = \dots = b_p = 0$$

$$H_0^2 : d_1 = \dots = d_p = 0$$

Эквивалентная формулировка нулевых гипотез звучит так: x_t не является причиной y_t , y_t не является причиной x_t соответственно.

Альтернативная гипотеза гласит о том, что хотя бы один из коэффициентов не равен b_i и d_i равен нулю.

Эквивалентная формулировка альтернативной гипотезы звучит так:
фактор x_t является причиной y_t , фактор y_t является причиной x_t соответственно.

2.2 Применение теста Грэнджера

Проведем тест Грэнджера для нормализованных отклонений для нахождения зависимостей между показателями. Для каждого фактора найдем 3 других фактора, которые влияют на него больше всего. Была написана программа на языке программирования Python [7][8], реализующая данную процедуру. При выполнении каждого теста мною был найден оптимальный лаг для обоих рядов. Факторы для каждого показателя отбирались по следующему принципу (для остальных факторов – аналогично):

1. Проводилось 9 тестов Грэнджера с каждым из 9 оставшихся факторов
2. Были выбраны те 3, у которых значение p -value, полученное после проведения теста, было минимальным.

Результаты теста представлены в Таблице 3. В каждой строке таблицы выделены 3 фактора, которые больше всего влияют на рассматриваемый мною показатель. Цифра «1» в строке означает, что фактор, в столбце которого стоит «1» влияет на показатель больше всего. Цифра «2» в строке означает, что фактор, в столбце которого стоит «2» занимает «2 место» по влиянию на показатель. Цифра «3» в строке означает, что фактор, в столбце которого стоит «3» занимает «3 место» по влиянию на показатель.

Таблица 3.

	Обо рот роз нич ой торг овл и / всег о(м лн. руб)	Числ еннос ть рабо чей силы (экон омич ески акти вного насел ения) / всего (ттыс .чел.)	Дене жные доход ы насел ения / всего (млн. руб)	Об щая пло щадь жил ищн ого фон да(т ыс м2)	Вало вой реги ональ ный проду кт в основ ных ценах / всего(млрд руб)	Ввод жил ых домо в нарас тающ им итого м с нача ла года / тыс. кв. м(ты с м2)	Инве стиц ии в основ ной капи тал на душу насел ения нарас тающ им итого м с нача ла года(млн руб)	Обща я забол еваем ость насел ения (на 1000 чел.)(чел.)	Числе нност ь посто янног о насел ения, на начал о пери ода / всего(тыс.ч ел.)	Индек с потреб ительс ких цен годово й / сводн ый(%)
Оборот рознич ной торговл и / всего(м лн. руб)			2		1				3	
Числен ность рабочей силы (эконом ически активн ого населен ия) / всего(т тыс.чел .)	3			2		1				
Денежн ые доходы населен ия / всего(м лн.руб)					3				1	2
Общая	2		3		1					

площадь жилищного фонда(тыс м2)										
Валовой региональный продукт в основных ценах / всего(млрд руб)	3			1						2
Ввод жилых домов нарастающим итогом с начала года / тыс. кв. м(тыс м2)				3		2				1
Инвестиции в основной капитал на душу населения нарастающим итогом с начала года(млн руб)	3			2	1					
Общая заболеваемость населения (на 1000)			2			1			3	

чел.)(чел.)										
Численность постоянного населения, на начало периода / всего(тыс.чел.)		2			3	1				
Индекс потребительских цен годовой / сводный(%)	3			1	2					

2.3 Робастность влияющих показателей

Основная задача данной работы заключается в распределении инвестиций между отраслями путем изменения наклона линейного тренда. Но распределение инвестиций подразумевает, что из одной отрасли какое-то количество денежных средств будет изъято и пойдет на улучшение ситуации в остальных отраслях. Слишком резкий «забор» денег из одной отрасли может привести к непредвиденным изменениям в экономике и к ухудшению ситуации. Например, уменьшение численности активного населения в пользу численности постоянного населения может привести к тому, что население региона растет согласно разработанному плану, но численность рабочей силы сильно уменьшилась. Из-за нехватки работников многие предприятия могут прекратить свою работу, что отрицательным образом скажется на росте экономики [9].

Задача состоит в распределении инвестиций таким образом, чтобы улучшить ситуацию в одних отраслях, но при этом не допустить ухудшения в других. В данном примере эта задача полностью провалена. Конечно, это всего лишь гипотетический пример, но он показывает, что может случиться, если перераспределение инвестиций будет необдуманным.

В связи с этим необходимо понять, как приток новых инвестиций в отрасль повлияет на группу самых влияющих на нее факторов. Если группа таких факторов претерпит изменения, в нашей работе это будет означать, что увеличение инвестиций в данную отрасль дестабилизирует экономическую ситуацию и приток такого количества инвестиций неэффективен. В работе за приток/отток инвестиций отвечает изменение наклона линейного тренда.

Поэтому необходимо изменять наклон тренда, пока группа самых влияющих факторов не изменится. Нужно найти верхнюю и нижнюю границу изменения наклона, при которых группа останется прежней, так называемую область робастности (устойчивости) для каждого показателя. Дадим определения этой области.

Определение: Область изменения количества вложений инвестиций в показатель называется областью устойчивости, если такое изменение не приведет к трансформированию группы из трех самых влияющих факторов для рассматриваемого показателя.

В предыдущем параграфе я сформировал такие группы для каждого показателя, содержащие по 3 фактора. Тест Грэнджера проводился для нормализованных отклонений. При формировании этих отклонений фигурировал линейный тренд, построенный для временных рядов. Именно наклон этого тренда я и буду менять.

Сильное приток/отток инвестиций из отрасли крайне редко встречается в реальной жизни, чтобы не допустить дестабилизации экономики, поэтому рассматриваю увеличение и уменьшение наклона линейного тренда не более чем на 10%. Большее изменение наклона я рассматривать не буду, даже если оно не приведет к изменению группы влияющих факторов.

Была написана программа на языке программирования Python, которая в зависимости от наклона тренда рассчитывает, изменилась ли группа самых влияющих факторов и выводит на экран новую группу факторов при ее изменении.

Области устойчивости для каждого показателя, показывающие на сколько %, мы можем увеличить приток или отток инвестиций из отрасли, не изменив взаимосвязи между отраслями, представлены в таблице 4.

Таблица 4.

Наименование показателя	Верхняя граница устойчивости(%)	Нижняя граница устойчивости(%)
Оборот розничной торговли / всего	10	5
Численность рабочей силы (экономически активного населения) / всего	3	10
Денежные доходы населения / всего	7	0,5
Общая площадь жилищного фонда	10	8
Валовой региональный продукт в основных ценах / всего	0,8	10
Ввод жилых домов нарастающим итогом с начала года / тыс. кв. м	10	10

Инвестиции в основной капитал на душу населения нарастающим итогом с начала года млн руб)	6	2
Общая заболеваемость населения на 1000 чел.(чел.)	10	10
Численность постоянного населения, на начало периода / всего (тыс. чел.)	7	0,1
Индекс потребительских цен годовой / сводный(%)	10	10

Дадим пояснения наименований столбцов, указанных в таблице:

Верхняя граница устойчивости - максимальное количество инвестиций, которое можно привлечь в отрасль, чтобы не изменилась группа из трех самых влияющих факторов.

Нижняя граница устойчивости - максимальное количество инвестиций, которое можно забрать из отрасли, чтобы не изменилась группа из трех самых влияющих факторов.

Для показателей «Ввод жилых домов нарастающим итогом с начала года», «Общая заболеваемость населения на 1000 человек» и «Индекс потребительских цен годовой» область устойчивости оказалась максимальной. Таким образом, для этих показателей любое изменение инвестиций в отрасль в пределах максимально выбранного диапазона не повлечет к изменению группы, состоящей из трех самых влияющих на показатель факторов. Для остальных показателей область существенно меньше, например, для показателя «Валовый региональный продукт в основных ценах» увеличение инвестиций даже на 0,9% приведет к изменению группы влияющих факторов, значит, данная отрасль имеет достаточно средств и не нуждается в дополнительном финансировании, нижняя граница устойчивости для данного показателя = 10%, это означает, что она готова «поделиться» своими инвестициями в пользу других отраслей в рамках установленного диапазона, и это не приведет к дестабилизации экономики.

Для 7 показателей, для которых область не оказалась максимальной, будем изменять наклон линейного тренда только в пределах значений, указанных в таблице.

Глава 3. Прогнозирование притока/оттока инвестиций

3.1 Подбор моделей

После проведения теста Грэнджера и нахождения для каждого показателя группы, состоящей из трех самых влияющих факторов, были найдены области устойчивости для каждого фактора, в которых группа самых влияющих на фактор показатель оставалась неизменной.

Задача этой главы состоит в количественной оценке прироста инвестиций в отрасль.

Утверждение: Математический алгоритм решения задачи состоит в следующем:

1. Рассмотреть показатель, характеризующий отрасль, для которого будет выполняться количественная оценка – y .
2. Взять группу, состоящую из трех самых влияющих на него показателей, которая была получена с помощью теста Грэнджера на причинность - x_1, x_2, x_3 .
3. Построить модель множественной линейной регрессии для показателя, где в качестве регрессоров будут факторы из группы, введенной в пункте 2.

$$y = a_0 + a_1 * x_1 + a_2 * x_2 + a_3 * x_3$$

4. Для каждого регрессора взять значение на прогнозном шаге, равное ретроспективному значению за последний отчетный год (2018 год) - $x_{1pred}, x_{2pred}, x_{3pred}$
5. Построить прогноз на 1 год вперед для зависимой переменной.

$$y_{pred} = a_0 + a_1 * x_{1pred} + a_2 * x_{2pred} + a_3 * x_{3pred}$$

6. Для каждого регрессора изменить прогнозное значение. Величина, на которую изменяем, значение берется из области устойчивости, которая была рассчитана для каждого показателя.

Например, для показателя «Оборот розничной торговли» верхняя граница устойчивости равна 10%, а нижняя граница устойчивости равна 5%. Получается, что если показатель «Оборот розничной торговли» является регрессором для какого-то другого показателя, то его прогнозное значение можно увеличить максимум на 10%, а уменьшить максимум на 5%. Обозначим эту величину за α_i .

7. Получим новые прогнозные значения регрессоров, которые будут использоваться при прогнозировании зависимой переменной.

$$\hat{x}_{1pred} = \alpha_1 * x_{1pred}$$

$$\hat{x}_{2pred} = \alpha_2 * x_{2pred}$$

$$\hat{x}_{3pred} = \alpha_3 * x_{3pred}$$

8. Построим прогноз на 1 год вперед для зависимой переменной y , взяв в качестве значений регрессоров на прогнозном периоде «новые» значения, рассчитанные на шаге 6.

$$\hat{y}_{pred} = a_0 + a_1 * \hat{x}_{1pred} + a_2 * \hat{x}_{2pred} + a_3 * \hat{x}_{3pred}$$

9. Рассчитаем, насколько изменится прогнозное значение переменной y по формуле. Обозначим это изменение, как ε . Тогда,

$$\varepsilon = 100 * \left(\frac{\hat{y}_{pred}}{y_{pred}} - 1 \right)$$

Используя этот алгоритм, я рассчитаю, как сильно изменится значение рассматриваемого показателя при увеличении/уменьшении инвестиций в факторы, от которых он зависит, при этом варьируя поток инвестиций, не дестабилизирую экономическую ситуацию в регионе.

Построим для каждого показателя модель множественной регрессии с 3-мя регрессорами

1. Модель множественной регрессии, построенная для показателя «Оборот розничной торговли»

$$y = 1054497,8 + 197,96 * x_1 + 0,25 * x_2 - 222,06 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Денежные доходы населения»

x_3 - «Численность постоянного населения, на начало периода».

Для показателя «Оборот розничной торговли» положительная динамика изменения это рост.

Коэффициенты при x_1 и $x_2 > 0$, поэтому наблюдается прямо пропорциональная связь между этими факторами и y . Если значения одного из регрессоров будут увеличиваться, то и значения зависимой переменной тоже будут увеличиваться, что и является целью (при условии, что остальные регрессоры не изменяются)

Обратная ситуация с регрессором x_3 . Коэффициент $a_3 < 0$, поэтому уменьшение значений x_3 приведет к увеличению значений y при условии, что остальные регрессоры не изменяются.

2. Модель множественной регрессии, построенная для показателя «Численность рабочей силы (экономически активного населения)»

$$y = 1866,8 + 0,0001 * x_1 + 0,05 * x_2 + 0,1 * x_3$$

где x_1 – «Оборот розничной торговли»

x_2 – «Общая площадь жилищного фонда»

x_3 - «Ввод жилых домов нарастающим итогом с начала года».

Для показателя «Численность рабочей силы (экономически активного населения)» положительная динамика изменения это рост.

Коэффициенты при x_1, x_2 и $x_3 > 0$, поэтому, несмотря на малые значения коэффициентов при регрессорах, наблюдается прямо пропорциональная связь между этими факторами и y . Если значения одного из регрессоров будут увеличиваться, то и значения зависимой переменной

тоже будут увеличиваться, что и является целью (при условии, что остальные регрессоры не изменяются)

3. Модель множественной регрессии, построенная для показателя «Денежные доходы населения»

$$y = 1518202 + 727,9 * x_1 - 259,53 * x_2 - 1774,58 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Численность постоянного населения на начало периода»

x_3 - «Индекс потребительских цен/годовой ».

Для показателя «Денежные доходы населения» положительная динамика изменения это рост.

Коэффициент при $x_1 > 0$, поэтому наблюдается прямо пропорциональная связь между этим фактором и y . Если значения регрессора будут увеличиваться, то и значения зависимой переменной тоже будут увеличиваться, что и является целью (при условии, что остальные регрессоры не изменяются)

Обратная ситуация с регрессорами x_2 и x_3 . Коэффициенты a_2 и $a_3 < 0$, поэтому уменьшение значений x_2 или x_3 приведет к увеличению значений y при условии, что остальные регрессоры не изменяются.

4. Модель множественной регрессии, построенная для показателя «Общая площадь жилищного фонда»

$$y = 93443 + 10,93 * x_1 - 0,01 * x_2 + 0,03 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Денежные доходы населения»

x_3 - «Оборот розничной торговли».

Для показателя «Общая площадь жилищного фонда» положительная динамика изменения это рост.

Коэффициенты при x_1 и $x_3 > 0$, поэтому наблюдается прямо пропорциональная связь между этими факторами и y . Если значения одного из регрессоров будут увеличиваться, то и значения зависимой переменной тоже будут увеличиваться, что и является целью (при условии, что остальные регрессоры не изменяются)

Обратная ситуация с регрессором x_2 . Коэффициент $a_2 < 0$, поэтому уменьшение значений x_2 приведет к увеличению значений y при условии, что остальные регрессоры не изменяются.

5. Модель множественной регрессии, построенная для показателя «Валовой региональный продукт в основных ценах»

$$y = -4616,1 + 0,03 * x_1 + 0,02 * x_2 + 17,3 * x_3$$

где x_1 – «Общая площадь жилищного фонда»

x_2 – «Оборот розничной торговли»

x_3 - «Индекс потребительских цен/годовой».

Для показателя «Валовой региональный продукт в основных ценах» положительная динамика изменения это рост.

Коэффициенты при $x_1, x_2, x_3 > 0$, поэтому наблюдается прямо пропорциональная связь между этими факторами и y . Если значения одного из регрессоров будут увеличиваться, то и значения зависимой переменной тоже будут увеличиваться, что и является целью (при условии, что остальные регрессоры не изменяются)

6. Модель множественной регрессии, построенная для показателя «Ввод жилых домов нарастающим итогом с начала года»

$$y = 6861,48 + 0,04 * x_1 - 27,1 * x_2 - 0,02 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Численность постоянного населения на начало периода»

x_3 - «Общая площадь жилищного фонда».

Для показателя «Ввод жилых домов нарастающим итогом с начала года» положительная динамика изменения это рост.

Коэффициент при $x_1 > 0$, поэтому наблюдается прямо пропорциональная связь между этим фактором и y . Если значения регрессора будут увеличиваться, то и значения зависимой переменной тоже будут увеличиваться, что и является целью (при условии, что остальные регрессоры не изменяются)

Обратная ситуация с регрессорами x_2 и x_3 . Коэффициенты a_2 и $a_3 < 0$, поэтому уменьшение значений x_2 или x_3 приведет к увеличению значений y при условии, что остальные регрессоры не изменяются.

7. Модель множественной регрессии, построенная для показателя «Инвестиции в основной капитал на душу населения нарастающим итогом с начала года»

$$y = -227867,3 - 116,8 * x_1 + 0,77 * x_2 + 2,41 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Оборот розничной торговли»

x_3 - «Общая площадь жилищного фонда ».

Для показателя «Инвестиции в основной капитал на душу населения нарастающим итогом с начала года» положительная динамика изменения это рост.

Коэффициенты при x_2 и $x_3 > 0$, поэтому наблюдается прямо пропорциональная связь между этим фактором и y . Если значения регрессора будут увеличиваться, то и значения зависимой переменной тоже будут увеличиваться, что и является целью (при условии, что остальные регрессоры не изменяются)

Обратная ситуация с регрессором x_1 . Коэффициент $a_1 < 0$, поэтому уменьшение значений x_1 приведет к увеличению значений y при условии, что остальные регрессоры не изменяются.

8. Модель множественной регрессии, построенная для показателя «Общая заболеваемость населения»

$$y = 8893,13 + 0,03 * x_1 + 0,0007 * x_2 - 1,65 * x_3$$

где x_1 – «Ввод жилых домов нарастающим итогом с начала года»

x_2 – «Денежные доходы населения»

x_3 - «Численность постоянного населения, на начало периода».

Для показателя «Общая заболеваемость населения» положительная динамика изменения его падение.

Коэффициенты при x_1 и $x_2 > 0$, поэтому наблюдается прямо пропорциональная связь между этими факторами и y . Если значения одного из регрессоров будет уменьшаться, то и значения зависимой переменной тоже будут уменьшаться, что и является целью (при условии, что остальные регрессоры не изменяются)

Обратная ситуация с регрессором x_3 . Коэффициент $a_3 < 0$, поэтому увеличение значений x_3 приведет к уменьшению значений y при условии, что остальные регрессоры не изменяются.

9. Модель множественной регрессии, построенная для показателя «Численность постоянного населения, на начало периода»

$$y = 5790,4 + 0,25 * x_1 - 0,03 * x_2 - 0,45 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Денежные доходы населения»

x_3 - «Численность постоянного населения, на начало периода».

Для показателя «Численность постоянного населения» положительная динамика изменения это рост.

Коэффициент при $x_1 > 0$, поэтому наблюдается прямо пропорциональная связь между этим фактором и y . Если значения регрессора будут увеличиваться, то и значения зависимой переменной тоже будут увеличиваться, что и является целью (при условии, что остальные регрессоры не изменяются)

Обратная ситуация с регрессорами x_2 и x_3 . Коэффициенты a_2 и $a_3 < 0$, поэтому уменьшение значений x_2 или x_3 приведет к увеличению значений y при условии, что остальные регрессоры не изменяются.

10. Модель множественной регрессии, построенная для показателя «Индекс потребительских цен/годовой»

$$y = 199,05 + 0,01 * x_1 - 0,008 * x_2 - 0,00001 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Денежные доходы населения»

x_3 - «Численность постоянного населения, на начало периода».

Для показателя «Индекс потребительских цен/годовой» положительная динамика изменения его падение.

Коэффициент при $x_1 > 0$, поэтому наблюдается прямо пропорциональная связь между этим фактором и y . Если значения регрессора будут уменьшаться, то и значения зависимой переменной тоже будут уменьшаться, что и является целью (при условии, что остальные регрессоры не изменяются)

Обратная ситуация с регрессорами x_2 и x_3 . Коэффициенты a_2 и $a_3 < 0$, поэтому увеличение значений x_2 или x_3 приведет к снижению значений y при условии, что остальные регрессоры не изменяются.

3.2 Интервалы инвестирования

В предыдущем параграфе были построены модели множественной линейной регрессии для каждого показателя, и найдены зависимости между изменениями значений зависимой переменной y и независимыми переменными x_i . Для некоторых показателей целью является их рост, для некоторых их падение с течением времени. Необходимо подобрать такие интервалы для каждого регрессора, которые будут изменять значения зависимой переменной в соответствии с заданной целью (либо увеличение значений, либо уменьшение значений).

Количественно рассчитаем интервал для каждой независимой переменной в модели множественной регрессии такой интервал, построим прогноз с использованием данной модели и проанализируем, на сколько такие изменения приведут к достижению цели в виде изменения показателя y . Программа, рассчитывающая эти величины, написана на языке программирования Python. Для более подробного объяснения алгоритма выполнения процедуры, разберем ее на примере двух показателей, для остальных, она выполняется аналогично.

1. Прогнозирование показателя «Оборот розничной торговли»

Найдем прогнозное значение на 2019 год для рассматриваемого показателя с помощью модели множественной регрессии, для которой прогнозные значения регрессоров равны значениям за последний отчетный год.

$$y = 1054497,8 + 197,96 * x_1 + 0,25 * x_2 - 222,06 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Денежные доходы населения»

x_3 - «Численность постоянного населения, на начало периода».

При подстановки значений регрессора в формулу получим прогнозное значение $y_{pred} = 1420929,3$.

Увеличению значений y соответствует увеличение значений x_1 и x_2 , а также снижение значений x_3 . В связи с этим, необходимо максимально увеличить прогнозное значение для x_1 и x_2 и максимально уменьшить прогнозное значение для x_3 . Но нужно соблюдать правило, которое гласит, что изменение значений регрессоров возможно только в пределах области устойчивости.

Согласно таблице 4, максимальное значение, на которое можно увеличить $x_1 = 0.8\%$, для x_2 максимальное значение = 7% ., для x_3 максимальное значение на которое можно его уменьшить = $0,1\%$. То есть,

$$\max(\widehat{x}_1) = 1,008 * x_1$$

$$\max(\widehat{x}_2) = 1,07 * x_2$$

$$\min(\widehat{x}_3) = 0,999 * x_3$$

Диапазон изменений значений для зависимой переменной y лежит в интервале от $0,88 * y_{pred}$ до $1,04 * y_{pred}$, то есть $\widehat{y}_{pred} \in [0,88y_{pred}; 1,04y_{pred}]$

Так как целью является увеличение оборота розничной торговли, то можно утверждать, что выше описанные преобразования, проведенные для прогнозных значений регрессоров, позволят увеличить прогнозное значение оборота розничной торговли на $4,4\%$, не дестабилизируя экономику региона. Это означает, что правильное перераспределение инвестиций между отраслями, которые больше всего влияют на оборот розничной торговли, позволит увеличить значение показателя на $4,4\%$, что в свою очередь повлечет к экономическому росту той отрасли, к которой относится данный показатель.

2. Прогнозирование показателя «Инвестиции в основной капитал на душу населения нарастающим итогом с начала года»

Найдем прогнозное значение на 2019 год для рассматриваемого показателя с помощью модели множественной регрессии, для которой прогнозные значения регрессоров равны значениям за последний отчетный год.

$$y = -227867,3 - 116,8 * x_1 + 0,77 * x_2 + 2,41 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Оборот розничной торговли»

x_3 - «Общая площадь жилищного фонда ».

При подстановки значений регрессора в формулу получим прогнозное значение $y_{pred} = 1398905,9$.

Увеличению значений y соответствует увеличение значений x_2 и x_3 , а также снижение значений x_1 . В связи с этим, необходимо максимально увеличить прогнозное значение для x_2 и x_3 и максимально уменьшить прогнозное значение для x_1 . Но нужно соблюдать правило, которое гласит, что изменение значений регрессоров возможно только в пределах области устойчивости.

Согласно таблице 4, максимальное значение, на которое можно уменьшить $x_1 = 10\%$. Для x_2 максимальное значение на которое можно его увеличить = 10% . Для x_3 максимальное значение = 10% . То есть,

$$\min(\widehat{x}_1) = 0,9 * x_1$$

$$\max(\widehat{x}_2) = 1,1 * x_2$$

$$\max(\widehat{x}_3) = 1,1 * x_3$$

Диапазон изменений значений для зависимой переменной y лежит в интервале от $0,88 * y_{pred}$ до $1,27 * y_{pred}$, то есть $\widehat{y}_{pred} \in [0,88y_{pred}; 1,27y_{pred}]$

Так как целью является увеличение инвестиций в основной капитал, то можно утверждать, что выше описанные преобразования, проведенные для прогнозных значений регрессоров, позволят увеличить прогнозное значение инвестиций в основной капитал на $27,02\%$, не дестабилизируя экономику

региона. Это означает, что правильное перераспределение инвестиций между отраслями, которые больше всего влияют на инвестиции в основной капитал, позволит увеличить значение показателя на 27,02%, что в свою очередь повлечет к экономическому росту той отрасли, к которой относится данный показатель.

Построим таблицу для всех 10 показателей, содержащую:

- а. прогнозное значение для y , полученное путем взятия регрессорами значений за последний отчетный период;
- б. нижнюю и верхнюю границу интервала, в который попадет прогнозное значение после изменений значений регрессоров в пределах области чувствительности
- с. максимальное значение на которое может увеличиться прогнозное значение зависимого показателя y

Полученные результаты представлены в Таблице 5.

Название показателя	Прогнозное значение	Нижняя граница интервала	Верхняя граница интервала	Максимальное значение изменения показателя (%)	Новое прогнозное значение
Оборот розничной торговли / всего(млн. руб)	1420929,3	0,88	1,04	4,4%	1483450,2
Численность рабочей силы (экономически активного населения) / всего(тыс.чел.)	3141,71	0,96	1,04	4,05%	3269
Денежные доходы населения / всего(млн.руб)	2997312,4	0,99	1,14	14,04%	4208226,6
Общая площадь жилищного фонда(тыс м2)	1417106,4	0,98	1,08	8,3%	1534726,2
Валовой региональный продукт в основных ценах /	4135,7	0,85	1,21	21,16	5010,8

всего(млрд руб)					
Ввод жилых домов нарастающим итогом с начала года / тыс. кв. м(тыс м2)	3806,4	0,82	1,2	19,67	4567,7
Инвестиции в основной капитал на душу населения нарастающим итогом с начала года (млн руб)	1398905,9 .	0,88	1,27	27,02%	1776610,49
Общая заболеваемость населения (на 1000 чел.)(чел.)	2458,2	0,73	1,07	7	1794,5
Численность постоянного населения, на начало периода / всего (тыс.чел.)	5337	0,98	1,05	4,8%	5603,85
Индекс потребительских цен годовой / сводный(%)	102,88	0,14	1,14	14%	86,1

В таблице 5 представлены расчеты, которые как можно управлять прогнозом показателя, перераспределяя инвестиций между влияющими отраслями. Таким образом, мы придем к более эффективному управлению отраслью и тем самым к ее экономическому росту.

3.3 Итоговый алгоритм

Утверждение: Алгоритм управления региональной экономикой выглядит следующим образом:

1. Производится отбор показателей, характеризующих социально-экономическое развитие региона
2. Трансформируем показатели, содержащие денежную составляющую в соответствии с инфляцией.
3. Для каждого показателя находим три самых влияющих на него показателя
4. Изменяем количество инвестиций, вкладываемых в экономическую нишу.
5. Находим значения увеличения/уменьшения инвестиций в отрасль, при которых группа из трех самых влияющих факторов остается неизменной.
6. Количественно оцениваем, на какую величину увеличится благосостояние отрасли (изменится в сторону цели прогнозное значение показателя, характеризующего отрасль) после оптимального перераспределения инвестиций между зависящими отраслями.

Рассмотрим на примере реализацию алгоритма, смоделировав ситуацию.

Задача: Лицо, принимающее решение, анализируя экономическую ситуацию в регионе, заметил, что денежные доходы населения растут очень медленно. Он хочет увеличить рост показателя к следующему году на 10% путем вложения в отрасль инвестиций. Задача, количественно оценить, как данное вложение повлияет на развитие других отраслей (на какую величину изменятся показатели, выбранные мною, как характеризующие свою отрасль).

Решение:

1. В таблице 3 найдем все факторы, на которые влияет показатель «Денежные доходы населения» (смотрим на все не пустые ячейки в столбце). Это факторы:
 - a. «Оборот розничной торговли»
 - b. «Общая площадь жилищного фонда»
 - c. «Общая заболеваемость населения на 1000 человек»
2. Рассмотрим регрессионные модели, в которых зависимой переменной являются найденные 3 фактора, а показатель «Денежные доходы населения» является регрессором. Рассмотрим подробно каждый из трех показателей.
3. «Оборот розничной торговли»

$$y = 1054497,8 + 197,96 * x_1 + 0,25 * x_2 - 222,06 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Денежные доходы населения»

x_3 - «Численность постоянного населения, на начало периода».

Построим прогноз на 1 год вперед, рассмотрев в качестве прогнозных значений регрессоров, их значения за последний отчетный год (2018 год).

$$y_{1,pred} = 1420929,3$$

Увеличим прогнозное значение регрессора x_2 на 10% в соответствии с условием задачи и построим прогнозное значение на 1 год вперед для оборота розничной торговли.

$$y_{2,pred} = 1493313,2$$

Рассчитаем, на какую величину ε увеличится значение показателя «Оборот розничной торговли»

$$\varepsilon = 100 \left(\frac{y_{2,pred}}{y_{1,pred}} - 1 \right) = 5,1\%$$

4. «Общая площадь жилищного фонда»

$$y = 93443 + 10,93 * x_1 - 0,01 * x_2 + 0,03 * x_3$$

где x_1 – «Валовой региональный продукт в основных ценах»

x_2 – «Денежные доходы населения»

x_3 – «Оборот розничной торговли».

Построим прогноз показателя на 1 год вперед, рассмотрев в качестве прогнозных значений регрессоров, их значения за последний отчетный год (2018 год).

$$y_{1,pred} = 152663,8$$

Показатель «Общая площадь жилищного фонда» зависит не только от показателя «Валовой региональный продукт в основных ценах», который изменится по условию задачи, но и от показателя «Оборот розничной торговли» изменение которого было рассчитано в пункте 3. Поэтому увеличиваем значение x_2 на 10% (в соответствии с условием задачи) и x_3 на 5,1% (в связи с расчетами, полученными в пункте 3.). Построим прогнозное значение на 1 год вперед для общей площади жилищного фонда.

$$y_{2,pred} = 151926,2$$

Рассчитаем, на какую величину ε уменьшится значение показателя «Общая площадь жилищного фонда»

$$\varepsilon = 100 \left(\frac{y_{2,pred}}{y_{1,pred}} - 1 \right) = 0,5\%$$

5. «Общая заболеваемость населения на 1000 человек»

$$y = 8893,13 + 0,03 * x_1 + 0,0007 * x_2 - 1,65 * x_3$$

где x_1 – «Ввод жилых домов нарастающим итогом с начала года»

x_2 – «Денежные доходы населения»

x_3 - «Численность постоянного населения, на начало периода».

Построим прогноз на 1 год вперед, рассмотрев в качестве прогнозных значений регрессоров, их значения за последний отчетный год (2018 год).

$$y_{1,pred} = 2210$$

Увеличим прогнозное значение регрессора x_2 на 10% в соответствии с условием задачи и построим прогнозное значение на 1 год вперед для оборота розничной торговли.

$$y_{2,pred} = 2412,92$$

Рассчитаем, на какую величину ε увеличится значение показателя «Общая заболеваемость населения»

$$\varepsilon = 100 \left(\frac{y_{2,pred}}{y_{1,pred}} - 1 \right) = 9,1\%$$

Ответ: При увеличении показателя «Денежные доходы населения» на 10% показатель «Оборот розничной торговли» увеличится на 5,1%, показатель «Общая площадь жилищного фонда» уменьшится на 0,5%, показатель «Общая заболеваемость населения на 1000 человек» увеличится на 9,1%.

Выводы

В рамках данной работы был построен алгоритм для лица, принимающего решение, помогающий ему перераспределить инвестиций между отраслями экономики. В качестве метрики, показывающей, как изменяется количество инвестиций, вложенных в отрасль, был взят наклон линейного тренда для каждого временного ряда. Так как каждая отрасль содержит огромное количество показателей, была проведена агрегация факторов и были выбраны факторы, характеризующие основные выделенные мною отрасли: социально-экономическую, социально-политическую, социально-демографическую, производственно-экономическую, финансово-кредитную сферы.

Проведена предварительная обработка данных для приведения дисперсий показателей к одному порядку, также была учтена инфляция для всех показателей, содержащих денежную составляющую. Трансформированные данные готовы для применения теста на причинность.

С помощью теста Грэнджера на причинность были найдены причинные связи между факторами, и для каждого фактора была найдена группа, состоящая из трех самых влияющих на него показателей. Перераспределение инвестиций может сильно повлиять на взаимосвязи между отраслями в экономике и на их дальнейшее развитие. В связи с этим, мною была написана программа на языке Python, находящая области устойчивости для каждого показателя, при которых изменение инвестиций в это показатель, не изменяет ту группу факторов, которые на него влияют. Максимальное изменение инвестиций в отрасль было взято, равным 10%, так как в реальной жизни такие большие вложения (заборы) инвестиций никто не производит из-за больших рисков.

Для количественной оценки того, как перераспределение инвестиций отразится, построены множественные регрессионные модели для каждого

показателя, где в качестве регрессоров выступали 3 показателя, находящиеся в группе самых влияющих. Было построен прогноз на 1 год вперед для всех 10 показателей. В качестве прогнозных значений регрессоров были взяты ретроспективные значения за последний отчетный (2018) год. Это исходная модель.

Далее, в соответствии с таблицей областей устойчивости прогнозные значения каждого регрессора были изменены в пределах значений этой области. После, с использованием новых прогнозных значений регрессоров построен прогнозный интервал для зависимой переменной.

В полученном интервале было найдено оптимальное прогнозное значение, в зависимости от целей каждой отрасли (увеличение или снижение значений). Далее, было количественно оценено, насколько изменится прогнозное значение показателя в целевую сторону путем сравнения нового полученного прогнозного значения и значения, полученное из исходной модели.

Эта оценка показывает, на какую величину лицо, принимающее решение, может улучшить развитие отрасль при эффективном распределении инвестиций между отраслями.

Программа, реализующая алгоритм, была реализована на языке программирования Python.

В дальнейшем, планируется совершенствование работы, включающее в себя усложнение построенных моделей прогнозирования, автоматизация программы для реализации алгоритма для любого числа показателей, нахождения более тесных связей между показателями. Это позволит найти более точные оценки интервалов изменения инвестиций, и облегчить лицам принимающее решение в перераспределении инвестиций задачу управления регионом.

Заключение

В работе был реализован алгоритм, который помогает органам исполнительной власти региона, как лицам, принимающим решение, перераспределить инвестиции в отрасли таким образом, чтобы увеличить инвестиции в те отрасли, который не показывают нужного роста, при этом, не ухудшив ситуацию в тех отраслях, из которых будут взяты инвестиции. Проблема заключается в том, что между некоторыми отраслями очень сложно найти видимые взаимосвязи. Построенный в данной работе универсальный алгоритм позволяет найти оптимальное распределение инвестиций между отраслями, даже не находя видимых экономических взаимосвязей между ними, для достижения эффективного управления регионом.

Список литературы

1. А.В. Прасолов, Анализ нескольких метрик в задаче о замещении сводного показателя, 2007
2. ЕМИСС – государственная статистика. <https://www.fedstat.ru/>
3. Федеральная служба государственной статистики. <https://www.gks.ru/>
4. Прасолов А.В., Математические методы экономической динамики. СПб: "Лань", 2008. 352 с
5. Кистанов В.В., Копылов Н.В. Региональная экономика России, Учебник. — М.: Финансы и статистика, 2003. — 584 с.
6. Буре В.М., Парилина Е.М., Теория вероятностей и математическая статистика. Учебник. — СПб.: Издательство «Лань», 2013
7. Joel Grus, Data Science from Scratch: First Principles with Python (2019). 406 с
8. Wes McKinney, Python for Data Analysis, 2nd edition, 2015
9. Хованов Н.В. Математические модели риска и неопределенности. СПб, СПбГУ, 1998.