

Санкт-Петербургский государственный университет

БЕВЗ Роман Юрьевич

Выпускная квалификационная работа

***Прогнозирование объема спроса на рынке
железнодорожных контейнерных
грузоперевозок в России***

Уровень образования: магистратура

Направление *01.04.02 «Прикладная математика и информатика»*

Основная образовательная программа *ВМ.5505.2020 «Математическое
и информационное обеспечение экономической деятельности»*

Научный руководитель:
доцент, кафедра математического
моделирования энергетических си-
стем, кандидат физ.-мат. наук,
Балыкина Юлия Ефимовна

Рецензент:
Волкова Тамара Александровна

Санкт-Петербург
2022

Содержание

Введение	3
Постановка задачи	5
Обзор литературы	7
Глава 1. Анализ рынка контейнерных перевозок в России	17
1.1 Формализация предметной области	17
1.2 Принцип соотношения спроса и предложения	19
1.3 Выделение сущностей и построение информационно-логической модели рынка	22
Глава 2. Анализ данных и построение математической модели формирования спроса	25
2.1 Сбор и обработка данных о факторах	25
2.2 Агрегирование и анализ данных	26
2.3 Построение математической модели	28
Глава 3. Математическое моделирование и прогнозирование спроса	31
3.1 Определение источников данных	35
3.2 Интеллектуальный анализ данных и прогнозирование	37
3.3 Анализ полученных результатов моделирования	52
Выводы	54
Заключение	56
Приложение А	61

Введение

В современном мире ни одно промышленное или торговое предприятие не обходится без услуг железнодорожных транспортных компаний. Любое производство или розничная торговля в масштабах города, региона, страны или международного масштаба включает в себя транспортировку грузов и различных товаров. В последнее время особенную популярность набирают контейнерные железнодорожные грузоперевозки, за счет универсальности тары, удобства погрузки контейнеров на железнодорожные платформы и преимущества железнодорожного вида транспорта в целом. Из множества макроэкономических факторов складывается потребность в тех или иных услугах перевозок, а от их эффективности, скорости и актуальности зависит как свежесть продуктов питания в магазинах, так и своевременность начала производственного цикла на фабрике. Также важно отметить, что несмотря на достаточно молодой рынок железнодорожных контейнерных перевозок, в России он стремительно развивается конкурентно. Поэтому руководству транспортных компаний следует соответствовать стремительным течениям в области управления бизнесом грузоперевозок для укрепления своих позиций на рынке. История развития данного рынка в России берет свое начало с относительно недавнего времени.

Начиная с 2000-х годов в России железные дороги эксплуатируются условно тремя группами компаний. В первую очередь это группа компаний ОАО «РЖД», владеющих всей сетью железных дорог. Вторая группа — независимые или формально независимые от ОАО «РЖД» железнодорожные операторы. Примечательно, что вся инфраструктура железных дорог России монополизирована данными государственными или около государственными компаниями. Наконец, третья группа компаний состоит из перевозчиков, которые владеют только своим или арендованным подвижным составом и пользуются государственной собственностью — сетью железных дорог и услугами локомотивной тяги. К таким компаниям можно отнести, например,

следующие: транспортная группа FESCO, ПАО «ТрансКонтейнер», АО «Евроросиб СПб-транспортные системы» и другие. В рамках программы структурной реформы 2011–2017 гг. на железнодорожном транспорте в Российской Федерации в целом были достигнуты положительные изменения. В частности, произошло положительное изменение конъюнктуры рынка грузовых железнодорожных операторов, рынок стал более конкурентным:

- 1) завершено формирование конкурентного сегмента рынка оперирования грузовыми вагонами и их предоставления под перевозку;
- 2) привлечены существенные инвестиции в обновление парка грузовых вагонов, значительно уменьшен износ подвижного состава (вагонов), работающего на сети;
- 3) ускорены темпы обновления и развития железнодорожной инфраструктуры, а также другие трансформации [1].

Благодаря проведенным изменениям было достигнуто увеличение конкуренции в железнодорожной отрасли, которое поспособствовало возникновению и развитию множества частных грузовых перевозчиков. Помимо структурных реформ немаловажное значение играют естественная непрерывная трансформация рынка (создание альянсов и коалиций), изменение его конъюнктуры, политические и экономические потрясения, пандемии и другие фундаментальные факторы, которые воздействуют на логистические связи в частности и форму ведения бизнеса в целом.

Таким образом, вышеперечисленные факторы непосредственным образом влияют на будущий спрос на услуги контейнерных перевозок. Все это стимулирует руководства транспортных компаний-перевозчиков применять новейшие техники в управлении своим бизнесом, особенно включая математический и экономический анализы. В свою очередь использование данных приемов позволяет компаниям строить прогнозные модели потребности в перевозках в будущем и на основе этих данных принимать оптимальные решения в менеджменте.

Постановка задачи

Современные тенденции развития рынка логистических услуг диктуют компаниям свои правила ведения бизнеса. Для повышения своей конкурентоспособности руководству транспортной компании необходимо разрабатывать и внедрять новейшие методы и приемы из различных областей научного знания: экономики, математики, теории игр и многих других. При принятии решений наравне с грамотным менеджментом, от специалистов в данной предметной области требуются передовые знания в перечисленных дисциплинах. Соответственно, данная выпускная квалификационная работа направлена на изучение и применение математических и экономических методов в построении прогнозных моделей спроса на рынке контейнерных железнодорожных перевозок, что способствует достижению конкурентного успеха компаний в соответствующей отрасли.

Сформулирована цель работы: составить прогноз объема спроса в отрасли железнодорожных контейнерных перевозок в России путем выявления значимых макроэкономических показателей за счет применения фундаментального и технического анализа.

Для достижения этой цели, были поставлены следующие задачи:

1. Обзор литературы по теме исследования, изучение предметной области и ее экономический анализ. Выявление сущностей предметной области и построение информационно-логической модели.
2. Сбор и обработка данных. Качественный и количественный анализ макроэкономических показателей. Построение математической модели.
3. Проведение вероятностно-статистического анализа. Прогнозирование и сценарное имитационное моделирование.

В качестве основных инструментов для решения выдвинутых задач, автором настоящей работы были использованы следующие математико-статистические методы:

- ABC/XYZ-анализ, метод главных компонент, экспертные оценки для определения значимости/незначимости параметров, характеризующих рынок железнодорожных контейнерных перевозок в России;

- методы дескриптивной статистики для оценки качественного характера влияния выявленных ранее факторов, а именно, корреляционный анализ, анализ временных рядов, а также статистический инфлюентный анализ;

- метод Монте-Карло для проведения статистического сценарного имитационного моделирования.

Численные эксперименты настоящей научно-исследовательской работы были проведены на примере АО «Евросиб СПб - ТС» — компании, занимающейся контейнерными железнодорожными перевозками в России.

Обзор литературы

Российская Федерация территориально является самым большим государством в мире и занимает площадь 17 125 191 км². По данным Росстата за 2020 год по территории страны проложено 126 тысяч километров железных дорог, из которых 87 тысяч общего и 39 тысяч необщего пользования [2]

Отрасль железнодорожной промышленности с самого своего зарождения в 1830-х годах оказывала колоссальное влияние на экономическую, социальную и политическую жизни общества. Железнодорожный транспорт обеспечивает бесперебойную доставку совершенно разнообразных грузов от продуктов питания и одежды до полезных ископаемых и военной техники. Строительство Транссибирской и Байкало-Амурской магистралей являлось стимулом развития экономики и промышленности Российской Империи и СССР. Параллельно с увеличением числа километров новых железных дорог появлялись новые и развивались старые города, основывались предприятия и целые промышленные отрасли. Кроме того, данный вид транспорта является стратегически важным для государства. Еще в середине XIX века В. А. Кокорев в своей статье «Экономические провалы» так отзывался о причинах поражения России в Крымской войне: «Если бы дорога от Москвы к Черному морю была начата постройкою в 1841 г., то Россия не почувствовала бы невозможности с миллионом лучшего в мире своего войска отразить высадившегося около Севастополя неприятеля в количестве 70 тыс. Впрочем, и самой высадки не могло бы быть, когда бы Европа знала, что наши войска по железной дороге, без всякого утомления, могут через несколько дней явиться на берегах Черного моря.» [3].

У железнодорожной транспортной системы также существуют и другие преимущества: железные дороги общего пользования обладают более высокой пропускной способностью, они менее подвержены погодным условиям и позволяют совершать перевозки из одного пункта до другого намного быстрее

вследствие более оптимальных и прямых маршрутов по сравнению с теми же автомобильными дорогами.

Из-за большой территории нашей страны железнодорожный транспорт крайне необходим для различных областей производства и торговли. Он позволяет перевозить разнообразные и большие массы грузов на дальние расстояния значительно выгоднее альтернативным вариантам, и, порой, данный вид транспорта является единственно возможным. По данным компании ОАО «РЖД» (рис. 1) динамика грузовых контейнерных перевозок (%) по сети РЖД 2019–2020 гг. в целом имеет положительную тенденцию. Можно заключить, что на внутреннем рынке России в целом имеется положительная динамика грузеных контейнеров и объема перевозок (обычно выражающиеся в тоннах, тонно-километрах или ед. TUE).

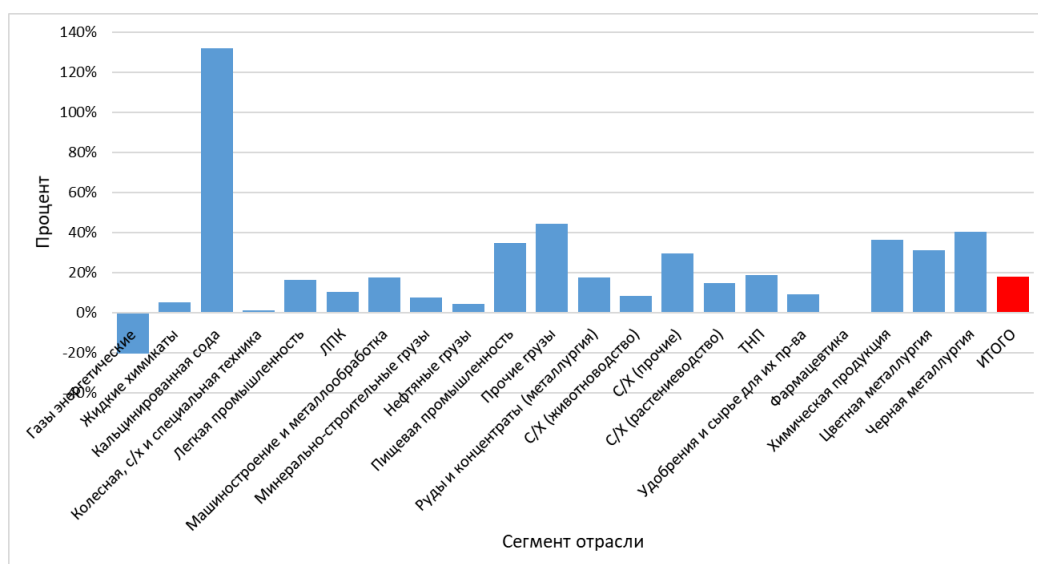


Рис. 1. Динамика перевозок грузеных контейнеров по сети РЖД 2019–2020 гг.

За последние 15 лет наблюдается тенденция роста перевозки грузов железнодорожным транспортом. На рис. 2 приведены данные Росстата о перевозках грузов железнодорожным транспортом с 2005 по 2019 года [2].

Конъюнктура рынка железнодорожных перевозок в Российской Федерации складывается из множества различных показателей. Их подробное исследование посредством проведения математического анализа и построения прогнозных моделей позволяет определить существующие проблемы для частных транспортных компаний в данной сфере. Изучение специальной литературы

является неотъемлемым этапом в достижении поставленных целей. В ходе изучения темы исследования был выделен ряд работ как отечественных, так и зарубежных специалистов.

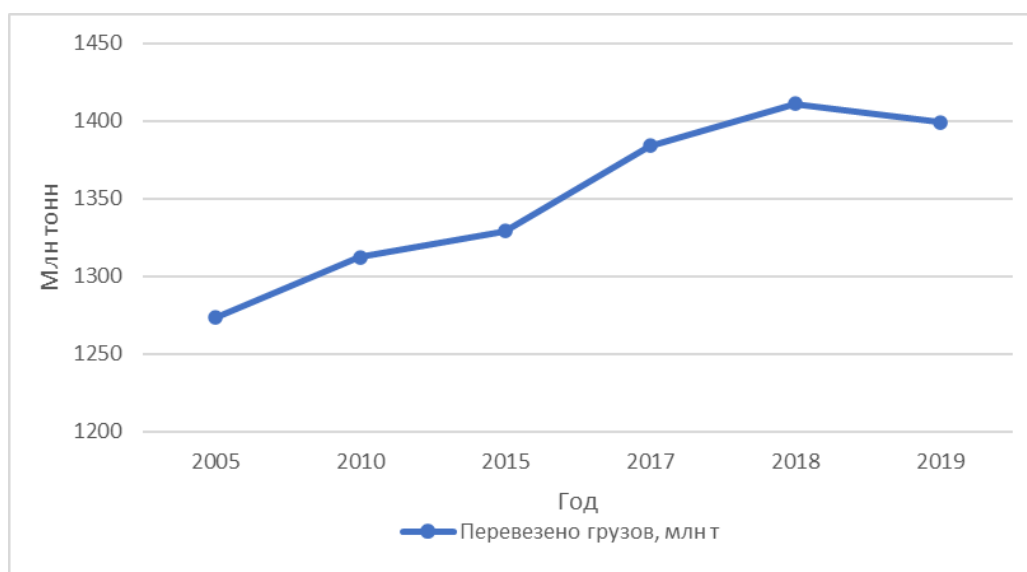


Рис. 2. Перевозки грузов по ж/д транспорту, млн тонн

В работах [4-6] содержатся основные определения макроэкономических показателей и экономические особенности железнодорожного транспорта. Также разобраны вопросы конкурентоспособности грузовых перевозок, их планирование и инвестиции в отрасль. С помощью данных пособий можно выделить ключевые сущности при составлении инфологической схемы данных, построения математических моделей и фундаментального анализа железнодорожной отрасли.

Одним из главных этапов научно-исследовательской работы является выделение макроэкономических индикаторов, характеризующих поведение рынка железнодорожной отрасли. Такими индикаторами могут быть объемы потребления, спроса, импорта, экспорта, транзита, внутренних перевозок и другие показатели динамики объема перевозок в разных разрезах времени. В качестве значимого индикатора макроэкономической конъюнктуры в России можно рассматривать прежде всего объемы грузовых (контейнерных) перевозок на железнодорожном транспорте [7]. Обоснованность использования именно этого показателя заключается в его абсолютной, единой величине, в отличие от параметров, характеризующихся денежными единицами: цены на

погрузку, стоимость добычи полезных ископаемых или производства товаров. Следует отметить, что показатель ВВП не всегда может достоверно оценивать ситуацию макроэкономического уровня по той же причине [7]. Объем грузовых перевозок характеризуется двумя ключевыми показателями: это погрузка грузов (тонны) и грузооборот (тонно-километры). В результате того, что показатель грузооборота оценивает дальность перевозки, то исследователю необходимо в дальнейшем также проанализировать и этот фактор – динамику средней дальности грузовых (контейнерных) перевозок [8]. В другой работе того же автора приводится еще один значимый индикатор – грузонапряженность – параметр, характеризующий интенсивность использования железнодорожной линии, измеряемый количеством т-км (брутто), приходящихся на 1 км эксплуатационной длины линии (главного пути) за определенный период времени (год) [9].

Таким образом, наравне с выделенными, наиболее показательными характеристиками в планировании перевозок грузов, в том числе контейнерных, могут служить следующие:

- объем перевозок (отправление) грузов в тоннах всего, в том числе по установленной номенклатуре грузов;
- грузооборот, в тарифных т*км;
- среднесуточная погрузка, в вагонах;
- средняя статическая нагрузка вагона, в т;
- средняя дальность перевозок грузов, в км;
- грузонапряженность в млн т*км на 1 км, средняя, в том числе по отдельным направлениям.

Дополнительно важно отметить показатели объемов производства в различных отраслях экономики России, которые являются «первообразными» выделенных характеристик грузового (контейнерного) железнодорожного транспорта. Например, основываясь на ключевых внутренних отраслях экономики России, с которыми взаимодействуют транспортные операторы,

необходимо выделить объемы производства лесной, целлюлозно-бумажной (ЦБ), автомобильной, сельскохозяйственной, химической промышленности, а также сфере розничной торговли. Разработка прогнозов на потребность в перевозках ведется в соответствии своей отраслевой номенклатуре, а также типам грузовых сообщений: внутренние, экспорт, импорт и транзитные перевозки [5].

Перечень факторов, которые могут существенно влиять на величину результирующих показателей – спроса на транспортные услуги, выражающийся, например, в объеме экспорта/импорта, может быть довольно велик. Таким образом, имеет место проблема выбора определенного набора таких индикаторов, которые главным образом отражали бы величину эндогенных показателей. Решение данной задачи позволяет аналитику составить адекватные математические модели, а также упростить их за счет «отсеивания» незначимых показателей. Особое внимание к значимым показателям макроэкономической системы позволяет сделать более точные выводы о поведении рынка. Соответственно перед тем, как приступать к непосредственному построению математических моделей, при обработке данных применяются инструменты вероятностно-статистического аппарата для выделения наиболее ключевых факторов. В качестве примеров были рассмотрены работы зарубежных авторов по данной проблематике.

В статье [10] для прогнозирования грузовых перевозок на железных дорогах Китая, авторы используют нейронную сеть обратного распространения, обученную на данных основных макроэкономических индикаторов в разрезе 11 лет. В основе предлагаемой авторами модели лежит метод главных компонент (Principal component analysis, PCA). Суть данного алгоритма заключается в уменьшении размерности пула данных за счет выделения сильно коррелирующих между собой объясняющих переменных и объединении их в группы – признаки-комбинации. Результаты вычислений показали, что максимальная доля общей дисперсии приходится на первые два показателя из

рассматриваемых изначально тринадцати – это ВВП и стоимость валовой промышленной продукции.

В другом исследовании подобную задачу в области определения спектра услуг для логистической компании предлагается решать методом классификации на основе ABC-анализа [11]. Данный метод основывается на принципе Парето, который говорит, что 20% от всех производимых товаров дают 80% выручки, в то время как остальные 80% товаров дают лишь 20% выручки. Процент наиболее ценных, промежуточных и наименее ценных товаров можно варьировать. Распределяя услуги по степени их влияния на себестоимость продукции, авторы дифференцируют виды таких услуг. В работе предлагается в категорию «А» относить услуги, приносящие 76% от общей выручки, в категорию «В» - 18% и в категорию «С» - 6% соответственно. По результатам анализа формируется стратегия обслуживания различных групп потребителей и в результате предлагается комплекс наиболее востребованных видов услуг. Например, наибольшие доли в общей выручке занимают услуги распространения продукции и хранения паллетных мест (13% и 10% соответственно). Полные результаты исследования приведены в [11]. По аналогии имеет смысл классифицировать основные макроэкономические параметры, характеризующие рынок контейнерных грузоперевозок в России.

Решая задачу предварительной обработки данных за счет применения, например методов главных компонент и ABC-анализа, можно значительно увеличить точность прогнозирования. Данные инструменты предоставляют мощный инструмент для статистического анализа прогнозирования спроса на грузовые перевозки железных дорог. При осуществлении такого анализа руководство транспортной компании может сделать наиболее общие выводы о дальнейших оптимальных действиях.

Помимо рассмотренных вероятностно-статистических методов отбора ключевых показателей, порой имеет смысл прибегать к классическим, устным или экспертным оценкам. В условиях недоступности данных или их крайней ограниченности на этапе составления математической модели такие способы

отбора показателей являются единственно возможными. Профессиональное мнение о значимости/незначимости факторов, характеризующих предметную область, также имеет решающую роль.

Проведение количественного и качественного анализа является ключевым этапом в работе. Данной темой активно занимаются как отечественные, так и зарубежные коллеги. В работе [12] представлен общий обзор различных методов прогнозирования спроса на перевозки: однофакторная регрессионная модель, метод скользящего среднего, экстраполяция трендов, метод экспоненциального сглаживания, а также метод Хольта. Упор в данной работе сделан на разработку метода анализа временных рядов на основе рассмотренных классических инструментов с учетом сезонности производства и потребления контейнерных грузов [12].

В исследовании [13] дается обзор качественных и количественных методов прогнозирования спроса на железнодорожную отрасль. Важно подчеркнуть, что качественные методы прогнозирования часто субъективны по своей природе и требуют суждения экспертов. Эти методы часто используют в ситуациях, когда исторических данных мало или они совсем отсутствуют (рис. 3). Количественные методы прогнозирования будущего спроса на железнодорожные перевозки используются в том случае, если компания располагает данными о продажах и обороте в прошлом и знает факторы, влияющие на эти показатели. Количественные методы можно классифицировать на две категории (рис. 4).

Ключевыми методами, используемыми аналитиками при количественном анализе можно считать регрессионный анализ и анализ временных рядов. Первый вводится в эконометрических моделях, которые используют взаимосвязь между одним рассматриваемым ключевым фактором и множеством переменных-предсказателей. Здесь имеет место «причинное прогнозирование», поскольку можно считать, что переменные-предсказатели воздействуют некоторой силой на интересующий ключевой фактор. В другом исследовании, решая задачу, аналогичной той, что ставится в данной работе, авторы строят

одномерную регрессионную модель на одной зависимой переменной – ежегодной посадке пассажиров в г. Мельбурн и десяти независимых переменных (среднегодовой процент общей процентной выплаты к доходу домохозяйства, численность населения, число занятых/трудоустроенных и др.) [13].



Рис. 3. Качественные методы анализа [13]



Рис. 4. Количественные методы анализа [13]

В качестве примера, основанного на регрессионных моделях, стоит рассмотреть методы статистического инфлюентного анализа, где главная идея заключается в использовании корреляционного и регрессионного анализа для нахождения статистических оценок и восстановления аналитической зависимости значений результирующего показателя от факторов (плановых и фактических) по заданной выборке [14]. Данные методы основываются на нахождении оценок влияния факторов из выбранной совокупности, оценок влияния неучтенных факторов и девиации (отклонения). Например, в работе [14] Р. И. Трухаева для нахождения аналитической зависимости результирующего показателя y от факторов x_1, \dots, x_n представлены методы простого регрессионного

и корреляционного анализа, методы однофакторных линейных и полиномиальных моделей регрессии, а также многофакторные линейные модели. Наиболее значимым вкладом в разрешение рассматриваемых задач, является композиционная многофакторная модель регрессии. Ее преимущество заключается в том, что при решении основной задачи регрессионного факторного анализа для полиномиальных многофакторных моделей имеются трудности с выбором вида нелинейных однородных форм, нахождением неизвестных коэффициентов, сложности вычислений статистических характеристик и оценке ошибки [14]. Главное достоинство такой модели в том, что нахождение n однофакторных полиномиальных моделей регрессии много проще, чем одной многофакторной. Для примера, простейшей композиционной многофакторной моделью является выражение следующего вида:

$$y = f(x_1, \dots, x_n) = \frac{1}{n} \sum_{j=1}^n f_j(x_j),$$

где $y = f_j(x_j) = a_{0j} + \sum_{m=1}^{M_j} a_{mj}x_j^m$ — однофакторные кривые регрессии y по x_j с коэффициентами a_{0j} и a_{mj} , определяемыми как решение системы линейных уравнений. Общей моделью будет являться параметрическая композиционная многофакторная модель регрессии:

$$y = f(x_1, \dots, x_n | \alpha) = \frac{1}{n} \sum_{j=1}^n \alpha_j f_j(x_j),$$

где $\alpha = (\alpha_1, \dots, \alpha_n)$ — это вектор параметров $\alpha_j \geq 0$, $\sum_{j=1}^n \alpha_j = 1$, который находится из условия минимизации остаточной дисперсии (обусловленной влиянием неучтенных факторов).

Общие модели временных рядов используют статистические свойства исторических данных для составления формальной модели, а затем оцениваются неизвестные параметры, например, методом наименьших квадратов [13]. Примером исследования данных такого типа методами анализа временных рядов служит работа [15].

В исследовании [15] находятся долгосрочные структурные взаимосвязи между спросом на грузовые перевозки и влияющими на него экономическими переменными, за счет использования данных годовых временных рядов за период 1960–1995 гг. Анализ был проведен с помощью многомерной коинтеграционной структуры векторной авторегрессии — VAR (vector autoregression). Для всех выборок были выявлены взаимосвязи между показателями ВВП, тонно-километров и показателем Р. Долгосрочная структурная взаимосвязь между этими переменными позволяет предположить, что экономический рост (ВВП), а также низкая эластичность цены на перевозки являются основными факторами, определяющими спрос на грузовые перевозки. Также в работе были обнаружены свидетельства сильной корреляционной зависимости между показателями ВВП и величиной тонно-километров.

Для специалиста в соответствующей предметной области немалую значимость играет информация об истории реформирования железнодорожного транспорта в других государствах: странах Евросоюза, США, Китае, Японии и других. Доклады ООН, посвященные аналитическим исследованиям, помогают изучить опыт других стран и применить передовые исследования на Родине. Доклад 2018 года по теме «Реформы на железнодорожном транспорте в регионе ЕЭК», подготовленный консультантом Давидом Рангетти показывает, что реформы, проводимые в регионе ЕЭК, в особенности в Российской Федерации дали в целом успешные результаты. Подчеркивается, что странам в регионе ЕЭК «важно как можно полнее учитывать – в процессе оценки «вертикального разделения» – необходимость согласования функции управления инфраструктурой и функции железнодорожных перевозок». Кроме того, акцентируется внимание на том, что каждому руководству страны следует учитывать свои собственные интересы и возможности, «чтобы можно было заметить какие-либо долгосрочные изменения», но перенимать опыт успешных реформ в странах ЕС, оказывающих положительное воздействие на железнодорожную отрасль [16].

Глава 1. Анализ рынка контейнерных перевозок в России

В течение 2020 года был замечен существенный рост деятельности отрасли контейнерных перевозок в России за счет увеличения использования железнодорожного транспорта (подвижного состава), в особенности на экспорт в Китай. В условиях политических и социальных ограничений между странами использование автомобильных фур и контейнеровозов становится невыгодным из-за таможенных оформлений и малом объеме перевозок. Вследствие этого в течение минувших двух лет (2020 и 2021 гг.) значительно увеличилась доля оборота портов Дальнего Востока из-за повышения ставок фрахта на импорт в Россию из стран АТР и КНР. Очевидно, что при таком же развитии в будущем требуется развитие железнодорожной инфраструктуры. Также важно отметить, что между такими странами как Россия и Китай существуют транснациональные проекты в модернизации и технологически новом уровне железнодорожной инфраструктуры.

Например, между странами участниками был заключен глобальный государственный проект: «Один пояс – один путь» - создание транспортных коридоров от западных границ Китая через страны Центральной Азии и Иран в Европу. Россия может стать главным транзитным «мостом» между странами Азии (Китаем) и Европы. Новый «Шёлковый путь» планируется прокладывать также из Китая через страны СНГ: Казахстан, Узбекистан, Туркменистан и Россию и далее в Европу: Болгария, Чехия, Румыния и Германия.

1.1 Формализация предметной области

Были рассмотрены основные принципы работы рынка контейнерных перевозок в Российской Федерации. Для России фундаментальной является следующая схема триангуляции (рис. 5). Данный термин описывает схему взаимодействия участников взаимоотношений на рынке с использованием железных дорог.

Страны Азии являются главным экспортером материальных товаров общего потребления по всему миру, в частности в Россию. Из узла «Азия» стрелка «импорт» направлена в сторону узла «Москва», как к главному распределительному центру России [17]. Основными экспортируемыми грузами для России являются природные: лес, газ, руды и т. п. Соответственно экспорт производится из узлов «Сибирь и ДФО» в узел «Азия». Ключевым направлением в данной схеме является «Внутрироссийское плечо» между узлами «Москва» и «Сибирь и ДФО». Транспортные компании перегоняют контейнерные локомотивы по данному направлению и далее цикл продолжается.

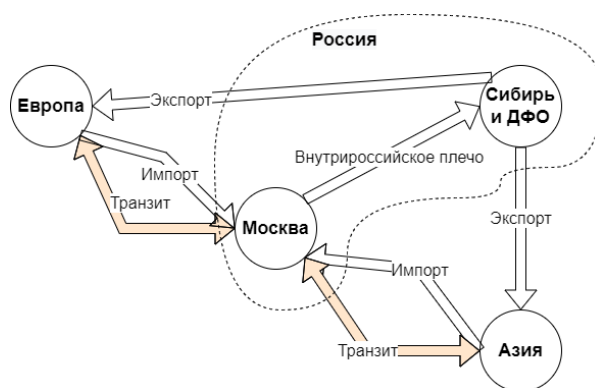


Рис. 5. Диаграмма железнодорожных контейнерных перевозок

Данная схема учитывает еще один, отдельный формат торговых отношений – транзитные перевозки (например, из стран Азии через Россию в страны Европы и обратно). Несмотря на лидирующие позиции в товарообороте между КНР и Европой посредством водного транспорта, контейнерные перевозки железной дорогой через Россию по данным направлениям составили более миллиона единиц TEU (1 млн 76 тыс.) за 2021 год [18]. В целом общая динамика объемов перевезенных грузов контейнерами за последние года положительна по всей номенклатуре, за исключением нефтепродуктов, металлоконструкций и продуктов перемола (–10,9%, –5,6% и –1% соответственно) [19].

Особенность работы транспортно-логистических компаний в таких рамках заключается в том, чтобы получить прибыль не только за счет экспорта, импорта и транзита контейнеров, но также за счет использования внутрироссийского плеча. Этого можно достигнуть за счет эффективного перемещения составов внутри страны посредством так называемых, «сервисов».

Сервисы.

Транспортные перевозчики в Россию работают с двумя видами сервисов: публичными и непубличными. Непубличные сервисы подразумевают доставку одного целого контейнерного поезда одному клиенту (схема В2В – business to business). Публичные сервисы работают по схеме В2С (business to customer), то есть загрузка подвижного состава контейнерами происходит не для одного клиента, а для множества. На рис. 6 продемонстрирована схема загрузки фитинговых платформ контейнерами условно для пяти клиентов. Основная сложность заключается в следующем.

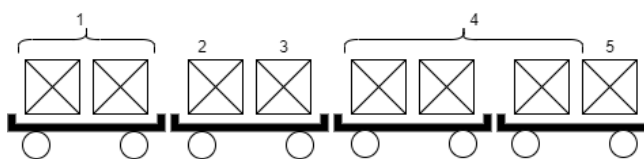


Рис. 6. Схема оформления загрузки контейнерного поезда

При такой загрузке подвижного состава необходимо четко придерживаться расписания и анализировать наиболее выгодные ставки на транспортировку. Один контейнерный поезд может быть загружен не более чем на 60% прямой доставки и не менее 40% экспедиторской. Прямая доставка подразумевает под собой то, что 60% перевозимых грузеных контейнеров обязаны доставляться одному клиенту по схеме В2В. Схема экспедиторской доставки работает иначе: перевозчик заключает договор об экспедиторских услугах с соответствующими организациями. Экспедитор — это контрагент, который принимает груз в свое ведение, обязуясь сделать все возможное для того, чтобы тот доехал в пункт назначения в целостности и сохранности. Соотношение 60% к 40% соблюдается потому, что часто приходится отправлять поезда загруженными только на прямую доставку.

1.2 Принцип соотношения спроса и предложения

При изучении данного аспекта предметной области предложено рассмотреть работы зарубежных ученых и провести аналогию о соотношении спроса и предложения по использованию и предоставлению услуг транспортировки грузов по железнодорожным сетям внутри России. Принимая во

внимание схему взаимодействия участников рынка (рис. 5), можно сделать следующие заключения.

Во-первых, на участке «Внутрироссийского плеча» спрос на услуги транспортных операторов по цене в целом является эластичным. Особенно это заметно в том, что цены и оборот производства полезных ископаемых значительно увеличились за последние года (рис. 1). В работе [20] берется метод двухрегиональной модели пространственного экономического равновесия, где основными переменными в исследовании являлись объемы перевозки зерна по железным дорогам США, а также цена за перевозку (в тоннах), объемы производства зерна и ставки на перевозку. Данные брались в период 1980–2010 гг.

Основные факторы, влияющие на формирование внутрирегионального спроса внутри России, можно сгруппировать в следующие категории [21]:

а. Экономические факторы, определяющие типы и количество производимых грузов, а также места производства и потребления этих грузов.

б. Логистические факторы, определяющие как отрасли и отдельные фирмы получают, производят, направляют и продают свою продукцию.

в. Транспортные факторы, влияющие на то, как удовлетворяется спрос на грузоперевозки, учитывая инфраструктуру, технологии и топливо, которые используются для перемещения товаров.

г. Политические факторы, регулирующие уровни экономики, логистики и транспорта.

В качестве иллюстрации, на рис. 7 приведены данные Росстата, отражающие доли отраслей, формировавших величину внутреннего регионального продукта в 2020 году. Лидирующие позиции приходились на предприятия обрабатывающего производства, торговую сферу, деятельность по операциям с недвижимостью и добычу полезных ископаемых – 17%, 14,1%, 10,6%, 10,5% соответственно. Сфера сельского хозяйства занимает 4,7%, уступая отрасли строительства 1,1% ВРП.



Рис. 7. Структура ВРП Российской Федерации в 2020 г. [19]

Касаясь показателей транзитных перевозок следует учитывать дополнительные параметры внешней среды. В первую очередь это политические и экономические факторы. Например, опыт социального потрясения 2020 года показал, что вследствие введения торгового эмбарго между США и Китаем и введением странами режима локдауна на фоне коронавирусной инфекции, образовался профицит порожних контейнеров в странах Европы и США. В то же время, так как Китай является главным поставщиком товаров общего потребления, то в стране образовался дефицит контейнеров под погрузку, а следовательно, ставка фрахта на перевозку одного контейнера 40 ДЭФ была увеличена почти в 9 раз. В случае анализа объемов экспорта и импорта следует учитывать данные внешние параметры совместно.

Во-вторых, на плече «экспорт» из узлов «Сибирь» в «Китай» имеется положительная тенденция увеличения спроса со стороны КНР. Об этом говорят политико-экономические новости. В ходе разговора между заместителем директора Департамента развития интеграционных проектов и стран СНГ Минэкономразвития России и Ли Цзинюань, Полномочным Министром Посольства КНР в РФ, было объявлено об увеличении несырьевого экспорта продукции в Китай, в особенности сельскохозяйственной продукции, продуктов питания [22]. Но лидирующие позиции на экспорт (с января по май 2021 г.) по-прежнему занимают поставки минерального топлива, нефти и полезные

ископаемые (объемы экспорта в целом в Китай за первый квартал 2021 года выросли на 15,4%). Наконец, в-третьих, на плече «импорта» из узлов «Китай» в «Москва» объемы товаров выросли на 38,7% и достигли \$18,2 млрд за январь-апрель 2021 г. [23].

Также важно отметить еще один формат торговых отношений – транзитные перевозки грузов из Китая через Россию в страны Европы. Несмотря на лидирующие позиции в товарообороте между КНР и Европой посредством водного транспорта, контейнерные перевозки железной дорогой по данным направлениям составили более миллиона единиц TEU за 2021 год [22].

В итоге, можно с уверенностью сказать, что имеется благоприятная почва для развития инфраструктуры рынка железнодорожных контейнерных грузоперевозок в России. В особенности на границе с партнерами по торговле и в портах Дальнего Востока. Эластичность спроса на услуги пользования подвижным составом по цене на эти услуги растет.

1.3 Выделение сущностей и построение информационно-логической модели рынка

Грузовые перевозки на железнодорожном транспорте характеризуются несколькими факторами. Основными показателями, характеризующими грузовые перевозки железнодорожным транспортом, являются: количество отправок; количество перевезенных тонн груза; грузооборот; густота перевозок; средняя дальность перевозки; неравномерность перевозок, коэффициент перевозимости, коэффициент транспортёмкости. Разберем кратко каждый из показателей в отдельности.

Количество отправок – Под отправкой понимается одна партия груза, оформленная одним перевозочным документом. Такой отправкой может быть как целый состав, так и малогабаритный небольшой груз товара, занимающий один мешок/коробку. Отправки подразделяются на различные типы: подвагонные, мелкие, групповые, контейнерные и другие. В настоящей работе будут рассматриваться последние.

Объем перевозок, количество перевезенных тонн грузов – P – данный параметр характеризует продукцию (или, другими словами, непосредственное перемещение груза из точки А в точку Б) за определенный промежуток времени (месяц, квартал, год). Общий объем перевозок по всей сети ж/д дорог складывается из четырех видов перевозок: внутрироссийские, экспортные, импортные и транзит [4]:

$$\sum P = \sum P_{\text{вн}} + \sum P_{\text{э}} + \sum P_{\text{и}} + \sum P_{\text{тр}}$$

Объем перевозок можно рассматривать также с точки зрения распределения по видам сообщений: ввоз, вывоз, транзит.

Грузооборот – Pl – этот параметр определяет объем работы транспорта по перевозкам грузов. Единицей измерения является тонно-километр. Исчисляется суммированием произведений массы перевезенных грузов в тоннах на расстояние перевозки в километрах (милях). Грузооборот транспорта группируется по видам транспорта, сообщения, ширине колеи, роду грузов и другим признакам.

На показатели грузооборота и объем перевозок в первую очередь влияют объем производства промышленной и сельскохозяйственной продукции; размещение производственных сил на территории страны; степень специализации производства; организации сбыта и другие [4].

Исходя из этого можно выделить следующие коэффициенты.

Коэффициент перевозимости – $K_{\text{перевоз}}$ – отношение объема перевозок i -го груза (P) к объему его производства в целом (Q):

$$K_{\text{перевоз}} = \frac{\sum P_i}{\sum Q_i}$$

Коэффициент транспортности – $K_{\text{транс-ти}}$ – отношение грузооборота i -го груза (Pl) к объему его производства в целом (Q):

$$K_{\text{транс-ти}} = \frac{\sum Pl_i}{\sum Q_i}$$

Густота грузовых перевозок (грузонапряженность) – Γ – количество тонн груза, перевезенного через один километр транспортной сети за единицу

времени. Данный параметр характеризует интенсивность грузового потока отдельных участков и подразделений ж/д дорог. Определяется как сумма перевозок по направлениям «туда» и «обратно»:

$$\Gamma_{AB} = \Gamma_{ab} + \Gamma_{ba}.$$

Данные перевезенных грузов и грузооборота непосредственно влияют на финансовое положение любой транспортной компании. Также и наоборот, инвестиции вовнутрь, в развитие своего потенциала, провоцируют изменение денежных показателей в лучшую сторону.

На основе вышеизложенного было проведено информационно-логическое моделирование: составлена информационно-логическая модель, отражающая структуру рынка железнодорожных контейнерных перевозок в России (рис. 8). Данное описание позволяет рассмотреть рынок как систему взаимосвязанных факторов.

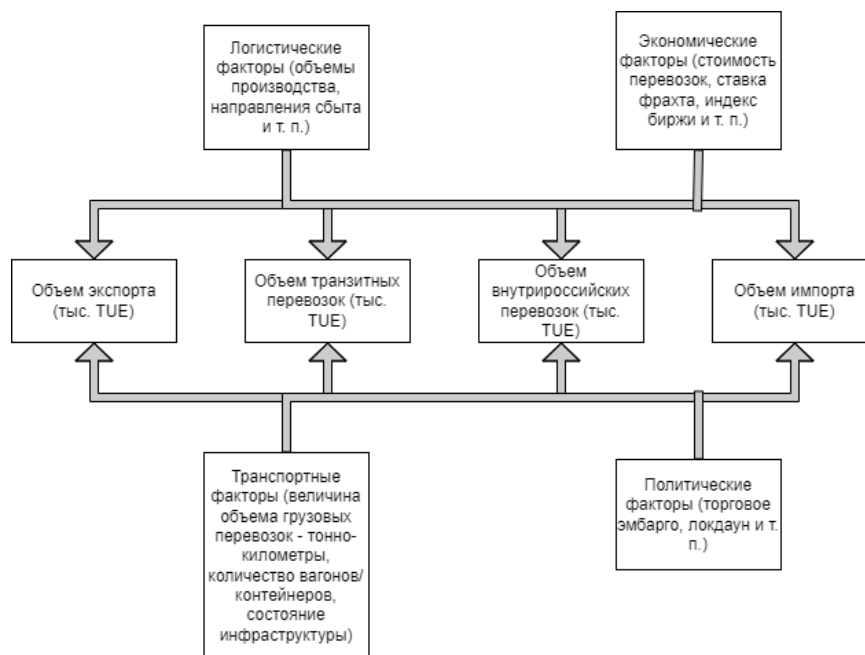


Рис. 8. Информационно-логическая модель рынка ж/д контейнерных перевозок в России

После определения сущностей рынка и их взаимосвязей ставится следующая основная проблема математического моделирования – сбор и обработка данных. Необходимо определить источники данных о факторах и сами данные в различных срезах, а также суметь их правильно интерпретировать для дальнейшего математического моделирования.

Глава 2. Анализ данных и построение математической модели формирования спроса

Сбор, обработка и анализ выявленных данных об имеющихся факторах является одной из ключевых задач перед составлением математических моделей. Методы информатизации процессов обработки таких данных позволяют понять, какие именно данные необходимо рассматривать в дальнейшем для построения наиболее адекватных математико-экономических моделей. В данной главе магистерской диссертации приводятся приемы в решении данных вопросов, а также математическая модель зависимости различных параметров, характеризующих поведение рынка железнодорожного контейнерного транспорта.

2.1 Сбор и обработка данных о факторах

В настоящее время большинство мест, где агрегируется статистика о деятельности транспортных компаний и отрасли в целом — это неструктурированные базы данных или даже просто массивы данных (например, OLAP-кубы или простые сводные таблицы). Такие хранилища имеют ряд недостатков, главными из которых можно назвать дублирование информации, а также неполнота данных. Поэтому, появляется необходимость в использовании методов интеллектуального анализа таких данных – Data mining. Далее, после этапа преобразования, следует воспользоваться методами представления информации с помощью технологии Business Intelligence (BI). Существует большое количество программных продуктов и инструментов для реализации таких задач.

Параллельно с описанной задачей необходимо определиться с выбором наиболее значимых параметров для будущей математической модели. На примере изученных научных работ были выделены следующие важные методики. В основе модели [10] лежит математический метод главных компонент (Principal component analysis, PCA), суть которого заключается в уменьшении размерности пула данных за счет выделения сильно коррелирующих между

собой объясняющих переменных и объединении их в группы – признаки-комбинации. В исследовании [11] подобную задачу в области определения спектра услуг для логистической компании предлагается решать методом классификации на основе ABC-анализа. Данный метод основывается на принципе Парето: 20% от всех производимых товаров дают 80% выручки, в то время как остальные 80% товаров дают лишь 20% выручки. Распределяя услуги по степени их влияния на себестоимость продукции, авторы дифференцируют виды таких услуг. В настоящей работе, соответственно, имеет смысл классифицировать основные макроэкономические параметры, характеризующие рынок контейнерных грузоперевозок в России. Третьим классическим методом является экспертная оценка. В условиях недоступности данных или их крайней ограниченности на этапе составления математической модели такие способы отбора показателей являются единственно возможными. Профессиональное мнение о значимости/незначимости факторов, характеризующих предметную область, также может быть весьма решающим.

2.2 Агрегирование и анализ данных

Эндогенные показатели будут разбиваться на несколько групп в зависимости от типа перевозки: экспорт, импорт, транзит и внутрироссийские и в зависимости от типа грузов по отраслям [24]:

- Ритейл (импорт из Китая в Москву и Новосибирск, а также внутрироссийские направления, как например, решение с торговой компанией Leroy Merlin).
- Перевозка автокомпонентов и автозапчастей (импорт из Европы в Россию, а также транзит из Китая в Европу через порты Дальнего Востока из Японии). Пример – решение с ПАО «КАМАЗ».
- Сельское хозяйство, насыпные и наливные грузы. Погрузка осуществляется в ТЛЦ «Евросиб – Терминал – Новосибирск» с использованием специального оборудования и тары – флекситанков. Наливные грузы — это пищевые масла. Далее данные виды товара могут отправляться в любых направлениях.

- Химическая промышленность (например, гранулированный полиэтилен и полипропилен). Внутривоссийские перевозки и экспорт в Китай и Азию. Пример – сотрудничество с ПАО «СИБУР Холдинг». Экспорт продукции из Тобольска и Томска в Китай через порты Дальнего Востока, в Бельгию через порты Санкт-Петербурга и Турцию через порты Новороссийска.

- Пиломатериалы и целлюлозно-бумажная промышленность (ЦБП). Экспорт товаров в Китай из центров Сибири через погранпереходы в Монголии.

Помимо приведенной группировки показателей, при составлении экономических моделей будут учитываться сезонность спроса на контейнерные перевозки в разрезе месяцев с января 2019 года по февраль 2021 года. Таким образом, на основе проведенного фундаментального анализа, в основу математической модели взяты следующие факторы x . (таблица 1). Объясняемой переменной y будет объем контейнерных перевозок, который непосредственно характеризует спрос на железнодорожные контейнерные перевозки в России (таблица 2).

Таблица 1. Объясняющие переменные

x_1	Объем отрасли торговли (млн. руб., Р)
x_2	Объем отрасли автомобилестроения (млн. руб., Р)
x_3	Объем производства отрасли сельского хозяйства (млн тонн)
x_4	Объем производства отрасли химической промышленности (млн тонн)
x_5	Объем производства пиломатериалов (млн тонн)
x_6	Величина ВВП (млн. руб., Р)
x_7	Грузооборот (т*км)
x_8	Таможенные пошлины (млн. руб., Р)

Таблица 2. Объясняемые переменные

y_1	Годовой объем контейнерных перевозок (тыс. TUE)
y_2	Общий объем импорта из стран Азии и Европы, (тыс. TUE)
y_3	Общий объем экспорта из России (тыс. TUE)
y_4	Общий объем транзитных перевозок (тыс. TUE)
y_5	Общий объем внутривоссийских перевозок (тыс. TUE)

Представим основные зависимости показателей следующим образом:

$$\begin{aligned}
 y_2 = x_1 + x_2; y_3 = \sum_{i=3}^5 x_i; y_4 = x_7 + x_8; \\
 y_5 = \sum_{i=1}^8 x_i; y_1 = \sum_{i=2}^5 y_i.
 \end{aligned}
 \tag{1}$$

При составлении математической модели количественные параметры рассматриваются как случайные величины. Соответственно, возникает задача в определении оценки степени влияния изменения экзогенных факторов на эндогенные. Для решения данной проблемы применительно к предметной области, автором настоящей работы предлагается методология инфлюентного анализа, основанная на регрессионном факторном анализе.

2.3 Построение математической модели

Имеется заданная выборка плановых и фактических значений результирующего показателя y и выбранной совокупности факторов x_1, \dots, x_n , то есть y^0, x^0, y^1, x^1 , где $x^0 = (x_1^0, \dots, x_n^0)$ и $x^1 = (x_1^1, \dots, x_n^1)$ — это плановые и фактические значения факторов x_1, \dots, x_n . Построим, соответственно, две регрессионные модели $y_0 = f_0(x_1, \dots, x_n)$ и $y_1 = f_1(x_1, \dots, x_n)$. Таким образом необходимо найти оценки влияния изменений значений факторов x_i на отклонение значений результирующего показателя y , а именно найти оценки $A_{x_i}^f$ влияния факторов x_i на отклонение $\Delta f = f_1^1 - f_0^0$ регрессионных значений f_1^1 и f_0^0 результирующего показателя y .

То есть необходимо, чтобы выполнялось равенство:

$$\Delta f = f_1(x_1^1, \dots, x_n^1) - f_0(x_1^0, \dots, x_n^0) = \sum_{i=1}^n A_{x_i}^f + A_{\text{ост}}^f + \delta y,$$

где $A_{\text{ост}}^f$ — регрессионная оценка влияния неучтенных факторов, $A_{x_i}^f$ — регрессионные оценки влияния факторов x_i на отклонение $\Delta y = y_1^1 - y_0^0$ значений результирующего показателя y ; δy — дополнительная характеристика — девиация значений Δy [14]. Оценки $A_{\text{ост}}^f$ и $A_{x_i}^f$ называются инфлюентами [25].

Для нахождения аналитической зависимости результирующего показателя от факторов, а также нахождения оценок влияния факторов, неучтенных факторов и девиации применительно к данной задаче, используется композиционная факторная модель регрессии. Она имеет следующий вид:

$$y = f(x_1, \dots, x_n | \alpha) = \frac{1}{n} \sum_{j=1}^n \alpha_j f_j(x_j),$$

где $f_j(x_j) = \alpha_{0j} + \sum_{m=1}^n \alpha_{mj} x_j^m$ – однофакторные кривые регрессии y по x_j с соответствующими коэффициентами, $\alpha = (\alpha_1, \dots, \alpha_n)$ – вектор неотрицательных параметров, находящиеся из условия минимизации остаточной дисперсии. Следовательно, инфлюенты находятся в следующем виде:

$$A_{x_j}^{f_j} = [f_j^1(x_j^1) - f_j^0(x_j^0)] - (a_{0j}^1 - a_{0j}^0) \quad (2)$$

Причем для изменения Δy (отклонения фактического значения y^1 от планового y^0) должно выполняться следующее равенство [25]:

$$\Delta y = \sum_{i=1}^n A_{x_i}^f$$

После нахождения инфлюент, для выражения $y_5 = \sum_{i=1}^8 x_i$ из системы уравнений (2) получим следующее:

$$\Delta y = \sum_{j=1}^8 \Delta y(x_j),$$

где $\Delta y(x_j) = \Delta x_j = x_j^1 - x_j^0 = A_{x_j}^y + (A_{x_j}^{f_j} - A_{x_j}^y) + (a_{0j}^1 - a_{0j}^0) + [\Delta y(x_j) - \Delta f_j(x_j)]$.

Принятие решений осуществляется путем составления множеств X^+ и X^- , состоящих из найденных инфлюент, если плановое значение y^0 по показателю y^1 не было достигнуто. А также находятся их упорядочения π^+ и π^- .

$$X^+ = \{x_i: A_{x_i}^f \geq 0\}; \quad X^- = \{x_i: A_{x_i}^f < 0\};$$

$$\pi^+ = \{A_{x_{i_1}}^f \geq A_{x_{i_2}}^f \geq \dots \geq A_{x_{i_l}}^f \geq 0\};$$

$$\pi^- = \{0 > A_{x_{i_{l+1}}}^f \geq A_{x_{i_{l+2}}}^f \geq \dots \geq A_{x_{i_n}}^f\}.$$

Следовательно, например, фактор x_{i_1} оказывает наилучшее влияние на условия достижения планового значения y^0 , а фактор x_{i_n} – наихудшее. Поэтому необходимо произвести компенсацию или улучшение характера влияния выявленных факторов x_{i_l} .

Также имеет особую важность вычисление ключевых статистических характеристик, которые показывают корреляционную зависимость между плановыми и фактическими значениями факторов и показателей. Например, можно рассмотреть следующие:

Среднее композиционных регрессионных значений результирующего показателя y :

$$\bar{f}(x|\alpha) = \frac{1}{N} \sum_{l=1}^N f(x^l|\alpha) = \frac{1}{N} \sum_{l=1}^N \sum_{j=1}^n \alpha_j f_j(x_j^l)$$

Дисперсия композиционных регрессионных значений результирующего показателя:

$$\sigma_{f(x|\alpha)}^2 = \frac{1}{N} \sum_{l=1}^N [f(x^l|\alpha) - \bar{f}(x|\alpha)]^2$$

Дисперсия композиционных регрессивных значений результирующего показателя y от среднего выборочного значения \bar{y} :

$$\delta_{f(x|\alpha)}^2 = \frac{1}{N} \sum_{l=1}^N [f(x^l|\alpha) - \bar{y}]^2$$

А также другие характеристики, представленные в работе Р. И. Трухаева [14].

Глава 3. Математическое моделирование и прогнозирование спроса

Анализируемое в настоящей работе предприятие является важным звеном в формировании конкурентной экономики в железнодорожной отрасли России. По объему грузооборота и количеству парка в собственности данный оператор на 2021 год входит в ТОП-30 Infoline ведущих частных железнодорожных перевозчиков России. Штаб-квартира находится в городе Санкт-Петербург. Компания предоставляет комплексное транспортно-логистическое обслуживание, включая организацию железнодорожных перевозок, услуги мультимодальной логистики, терминальные операции.

1) Железнодорожные перевозки грузов в собственном и арендованном подвижном составе (13,4 тыс. вагонов в собственности и управлении).

2) Организация ускоренных перевозок грузов в контейнерных поездах по расписанию.

3) Терминальные операции, складская логистика на базе собственного терминально-логистического центра в Новосибирске.

4) Услуги локомотивной тяги для промышленных предприятий [24].

Также компания входит в крупнейшие российские рейтинги по количеству фитинговых платформ в управлении, по показателю среднего возраста подвижного состава, по количеству вагонов в собственности и лизинге – «ТОП-400» и «ТОП-250 Северо-Запад» (РА «Эксперт»), INFOLine Rail Russia Top и др. [24].

Таким образом, можно заключить, что имеется положительная тенденция к конкурентному развитию рассматриваемого транспортно-логистического предприятия в отрасли грузовых контейнерных перевозок в России. Рассмотрим деятельность и оценку финансовой составляющей компании.

Автором настоящей работы была изучена сеть сервисов (таблица 3, рис. 9), по которой работает данный железнодорожный оператор: используемые

терминалы (сухие и морские), направления движения, интенсивность движения поездов.

Таблица 3. Список сервисов [24]

Номер	Города и станции отправления – назначения
1.	Москва, ст. Орехово-Зуево, ст. Тучково – Новосибирск, ст. Иня-Восточная (ООО «Евросиб – Терминал – Новосибирск»);
2.	Москва, ст. Белый Раст – Красноярск, ст. Базаиха (БЛТК);
3.	Москва, ст. Белый Раст – Уссурийск, ст. Уссурийск (Восток-А);
4.	Москва, ст. Белый Раст – Калининград, ст. Черняховская;
5.	Москва – Иркутск (на паузе);
6.	Москва – Хабаровск (на паузе);
7.	Новосибирск, ст. Иня-Восточная (ООО «Евросиб – Терминал – Новосибирск») – Находка, ст. Находка-Восточная (ООО «ВСК»);
8.	Новосибирск, ст. Иня-Восточная (ООО «Евросиб – Терминал – Новосибирск») – Владивосток, ВМПП (ВМП Первомайский, Мыс-Чуркин-Экс.);
9.	Красноярск, ст. Базаиха (БЛТК) – Находка, ст. Находка-Восточная (ООО «ВСК»);
10.	Уссурийск, ст. Уссурийск (Восток-А) – Москва, ст. Белый Раст;
11.	Владивосток, ВМПП (ВМП Первомайский, Мыс-Чуркин-Экс.) – Москва, ст. Белый Раст;
12.	Владивосток, ВМПП (ВМП Первомайский, Мыс-Чуркин-Экс.) – Новосибирск, ст. Иня-Восточная (ООО «Евросиб – Терминал – Новосибирск»);
13.	Владивосток, ВМПП (ВМП Первомайский, Мыс-Чуркин-Экс.) – Санкт-Петербург (на паузе);
14.	Санкт-Петербург, ст. Заневский Пост (Логистический Парк (ЛП) «Якино») – Новосибирск, ст. Иня-Восточная (ООО «Евросиб – Терминал – Новосибирск»);
15.	Санкт-Петербург, ст. Заневский Пост (Логистический Парк (ЛП) «Якино») – Владивосток, ВМПП (ВМП Первомайский, Мыс-Чуркин-Экс.)

По данным информационно-аналитических агентств с 2012 по 2020 гг. была проанализирована общая выручка и чистая прибыль рассматриваемой транспортной компании (таблица 4). Исходя из полученных данных можно сделать некоторые предположения о причинах роста и падения и работе компании в целом.

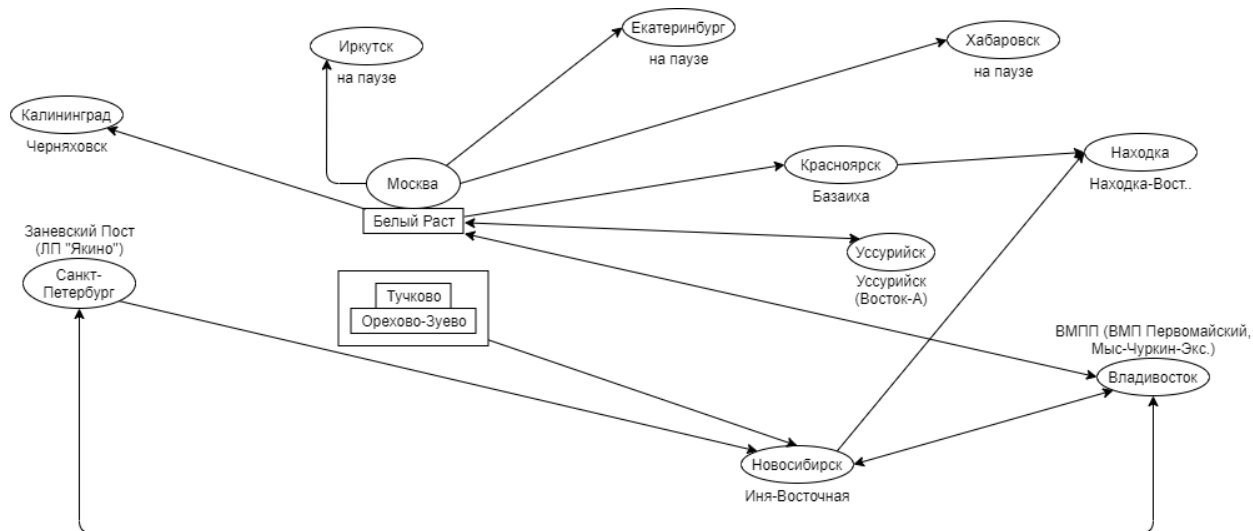


Рис. 9. Схема сервисов

Таблица 4. Данные выручки и прибыли ОАО «Евросиб СПб – ТС»

год	общая выручка (млрд)	чистая прибыль (млн)	год	общая выручка (млрд)	чистая прибыль (млн)
2012	7,654	709,9	2017	9,2	402
2013	8,5439	-468,2	2018	12,014	450
2014	7,2183	-1273,5	2019	15,1403	75,272
2015	6,936	6,5	2020	17,16	22,4
2016	6,7731	515,1			

На рис. 10 и рис. 11 изображены графики изменения общей выручки компании начиная с 2012 года.

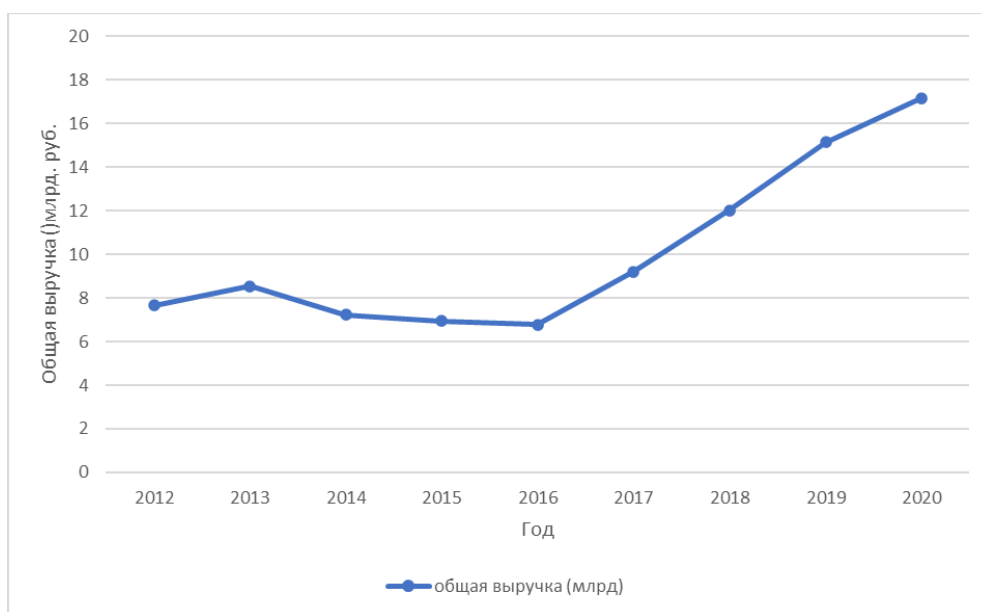


Рис. 10. Общая выручка ОАО «Евросиб СПб – ТС»

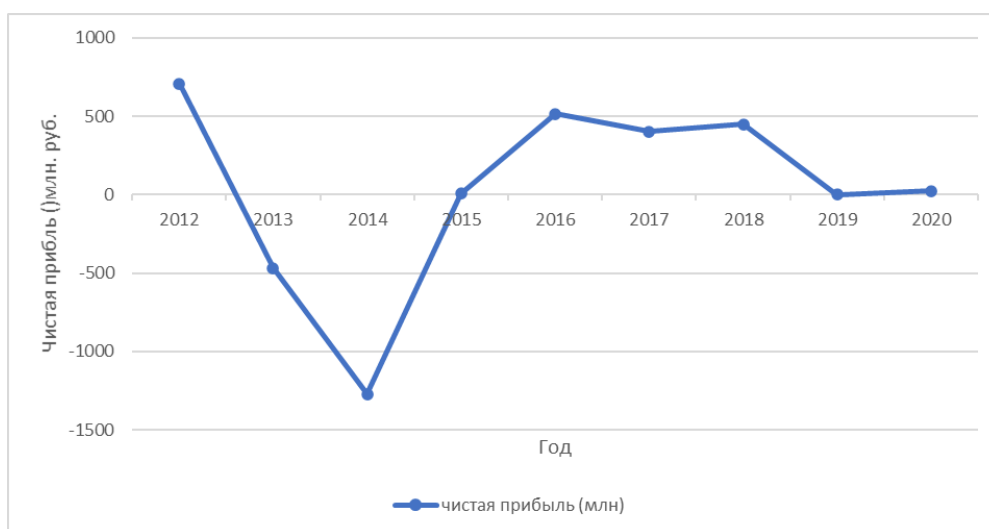


Рис. 11. Чистая прибыль ОАО «Евросиб СПб – ТС»

Можно предположить, что большой спад в объеме чистой прибыли в период с 2012 по 2014 года был вызван тем, что компания совершила крупные инвестиции в свою деятельность (закупку подвижного состава, вагонов, различно оборудование). Параллельно сравнивая с объемом общей выручки «Евросиб СПб – ТС» в тот же период времени совершил в среднем одинаковое количество перевозок (грузооборот и перевезенные грузы).

Резкий рост чистой прибыли начиная с 2014 года как раз обуславливается тем, что после возможных инвестиций в отрасль оправдали себя. Среднее положение общей выручки остается практически неизменным.

Начиная с 2016 года наблюдается подъем общей выручки компании практически в 2,5 раза до 2020 года. Но несмотря на это, объем чистой прибыли падает практически до нуля (до 2020 года – последним данным). Данное изменение скорее всего говорит о том, что руководство компании делает неправильные действия в управлении перевозками, а также возможно изменение внешних макроэкономических показателей. Но с учетом роста общей выручки второе предположение уходит на второй план.

Полученные «скачущие» результаты говорят о том, что компании следует внимательно изучить свои данные и данные макроэкономических показателей. Подробный анализ позволит «сгладить» кривые, отвечающие за финансовый рост «Евросиб СПб – ТС» в будущем.

Составим схему данных (рис. 12). На данной модели представлены ключевые объекты данных, обозначены дата отправки/получения, станции назначения/получения груза, грузоотправители и грузополучатели, подвижной состав и вагоны, находящиеся в собственности или в правах лизинга у компании и, собственно, грузы.

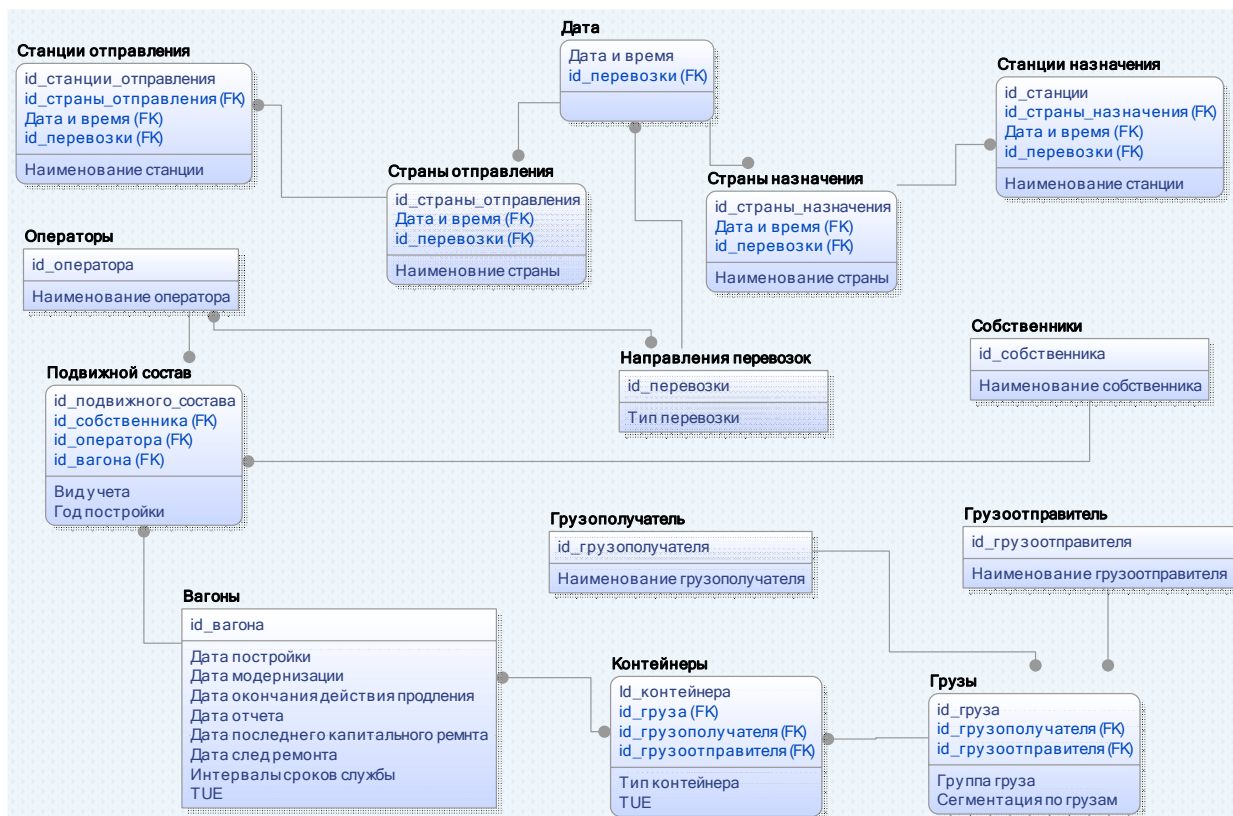


Рис. 12. Схема данных

3.1 Определение источников данных

В ходе выполнения работы были выявлены основные источники данных, характеризующие состояние компании во временном промежутке за несколько лет. Ниже приведен перечень ресурсов, содержащих данные макроэкономического и микроэкономического характера в соответствии с системой взаимозависимых факторов внешней и внутренней сред.

1. Росстат и региональные базы данных Росстата. Объемы производства по отраслям (в тоннах). Этот источник позволяет проанализировать данные о продукции и товарах, производимых внутри России, и которые отправляются на экспорт в разных разрезах времени. Формат данных — это таблицы с ТН ВЭД-кодом, который характеризует тип производимой внутри России

продукции. Цель использования данного ресурса будет заключаться в агрегировании данных видов продукции и вычислении стоимости перевозки – фрахт.

2. Главный вычислительный центр (ГВЦ) РЖД. Данный ресурс представляет собой хранящийся на сервере РЖД файл формата Cube – OLAP-cube (On-Line Analytical Processing-Cube). Другими словами, это база данных РЖД в виде сводной таблицы Excel, отображающая информацию обо всех объемах перевозок во времени, а также станциях, городах и странах назначения/отправления, наименованиях перевозчиков и т. п.

3. Данные бирж – индексы контейнерных перевозок.

а. Доходная ставка показывает, какой доход от перевозки соответствующего груза получает перевозчик (РЖД) в расчете на тонно-километровый измеритель (копеек на 10 т-км).

б. Данные документов НОТИС (англ. notice) - уведомление о готовности судна/авто/локомотива/самолета к отправлению, погрузке и т.д. Сообщается соответственно грузоотправителю или грузополучателю непосредственно капитаном судна или через морского агента.

с. Ресурсы, содержащие информацию об индексах на контейнерные перевозки. Это в первую очередь сайты бирж, где ставятся индексы на контейнерные перевозки. Они бывают двух видов. Шанхайский индекс (SCFI) - показывает изменение фрахта на перевозки контейнеров из Китая по основным мировым направлениям. Балтийский индекс (FBX) - отражает стоимость перевозок сухого груза (уголь, руда, зерно и т. п.) морем по двадцати основным торговым маршрутам и охватывает перевозки, производимые сухогрузами классов Handymax, Panamax и Capesize. Впервые он был опубликован в январе 1985 года.

Цель использования данных ресурсов – анализ цен на фрахт, то есть обусловленная договором или законом плата за перевозку груза. Уплачивается перевозчику отправителем груза или фрахтователем.

4. База данных Федеральной таможенной службы - содержится информация о перевозках (в тоннах) между странами. Данный ресурс позволяет получить информацию об импорте/экспорте/транзите товаров.

5. Базы данных собственников подвижного состава у РЖД. Можно узнать информацию о списании подвижного состава в будущем.

3.2 Интеллектуальный анализ данных и прогнозирование

Проведем ABC-анализ. В основе метода лежит принцип Парето: 20% усилий обеспечивают 80% результата. В настоящей работе, соответственно, классифицированы основные макроэкономические параметры, характеризующие отрасли экономики по объемам перевозок по данным, предоставленным компанией Евросиб (рис. 13). Всего имеется данных о двадцати двух сферах промышленности на протяжении двух лет: с января 2019 по февраль 2021 годов (Приложение А).

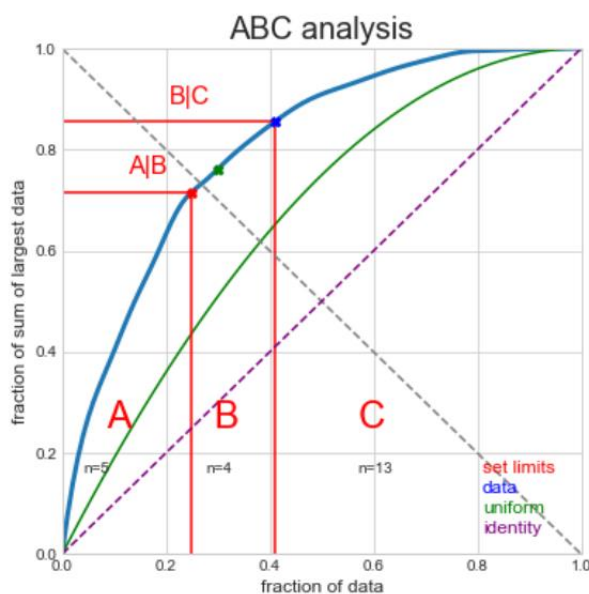


Рис. 13. Результаты ABC-анализа

Результаты выполненного анализа говорят о том, что приблизительно 25% наименований всех секторов экономики, а именно пять пунктов (см. таблица 5) обеспечивают 72% объема полезных данных (группа А). Далее, в группу В входят четыре наименования секторов (25% + 16% = 41%) экономики, дающие в сумме 85% полезности данных и в группе С содержатся все остальные.

Таблица 5. Результаты ABC-анализа

№	'A_ind': (25% - 72%)	'B_ind': (41% - 85%)
1	Лесопромышленный комплекс (ЛПК)	Цветная металлургия
2	Химическая продукция	Пищевая промышленность
3	Машиностроение и металлообработка	Минерально-строительные грузы
4	Черная металлургия	Товары народного потребления (ТНП)
5	Жидкие химикаты	

На основе имеющихся данных построим график регрессионной модели. В соответствии с полученными результатами ABC-анализа были взяты показатели перевозок грузов из группы А. В качестве объясняющих переменных выбраны величина ВВП в млрд руб. и величина грузооборота в млрд т-км. по месяцам начиная с января 2019 года и заканчивая февралем 2021 года.

Проиллюстрируем временные ряды обоих показателей (рис. 14 и рис. 15). Построим корреляционное поле значений перевозок грузов ЛПК в млн т. и грузооборота (рис. 16). Коэффициент корреляции Пирсона между данными показателями равняется 32%. Следовательно, можно предположить о наличии зависимости между данными.



Рис. 14. Временной ряд для показателя ЛПК

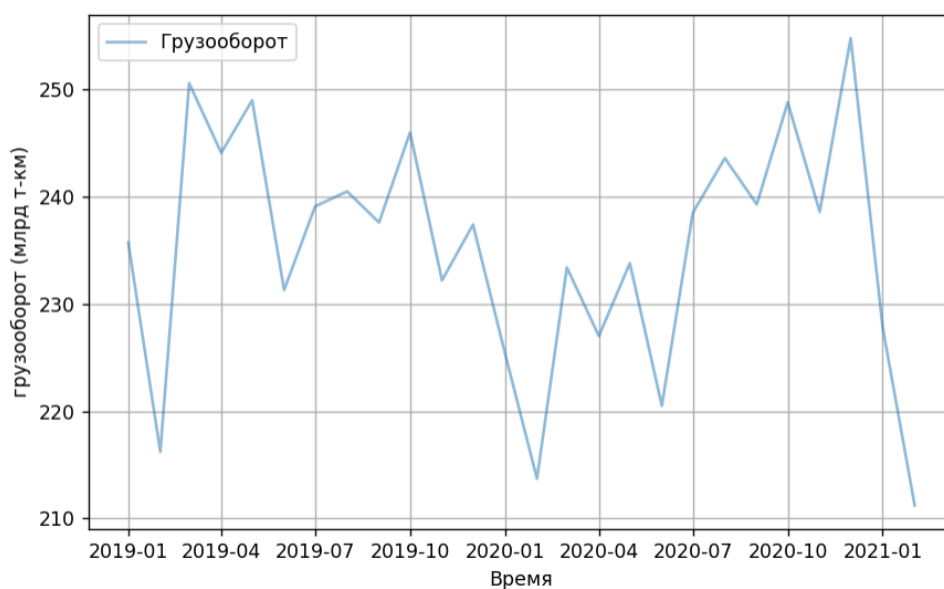


Рис. 15. Временной ряд для показателя грузооборота

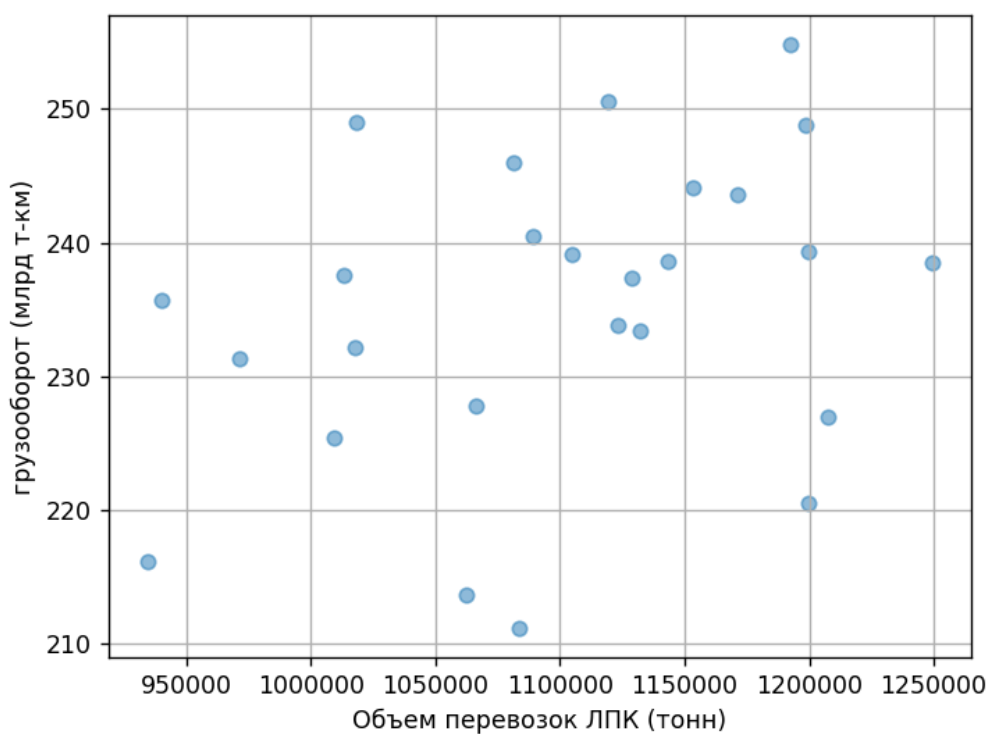


Рис. 16. Зависимость ЛПК от грузооборота

Построим уравнение линейной регрессии зависимости y от X (зависимость объема перевозок ЛПК в тоннах и набора остальных показателей).

дата	ЛПК	Химическая продукция	Машиностроение и металлообработка	Черная металлургия	Жидкие химикаты	грузооборот (млрд т-км)	ВВП (млрд руб)
2019-01-01	940054.0	358995.0	359975.0	324093.0	396797.0	235.7	8661.645785
2019-02-01	934602.0	373480.0	394487.0	293906.0	369341.0	216.2	8045.220640

Коэффициенты уравнения: $\beta_1 = 0,353$, $\beta_2 = 0,666$, $\beta_3 = -0,125$, $\beta_4 = 0,396$, $\beta_5 = 2860,184$, $\beta_6 = -48,9$. Смещение: $\beta_0 = 232451.17$. R^2 – статистика: 0.713.

Сначала проверим модель на нелинейность. Чтобы обнаружить нелинейность, рассмотрим графики наблюдаемых/прогнозируемых значений и остатков/прогнозируемых значений показателя ЛПК. Линейность заключается в том, что точки будут симметрично распределены вокруг диагональной линии на первом графике или вокруг горизонтальной линии на последнем. В обоих случаях с примерно постоянной дисперсией. Как можно заметить, график прогнозируемых и наблюдаемых значений y в принципе имеет линейный характер (рис. 17). Соответственно построим уравнение многофакторной линейной регрессии. Данные результаты также подтверждаются следующими вычислениями статистик средней квадратичной ошибки регрессии и статистикой R^2 . Для уравнения регрессии второй степени данные статистики резко возрастают: $RMSE = 707587.854$ и $R^2 = -165.630$. Для третьей: $RMSE = 923375.661$ и $R^2 = -282.758$.

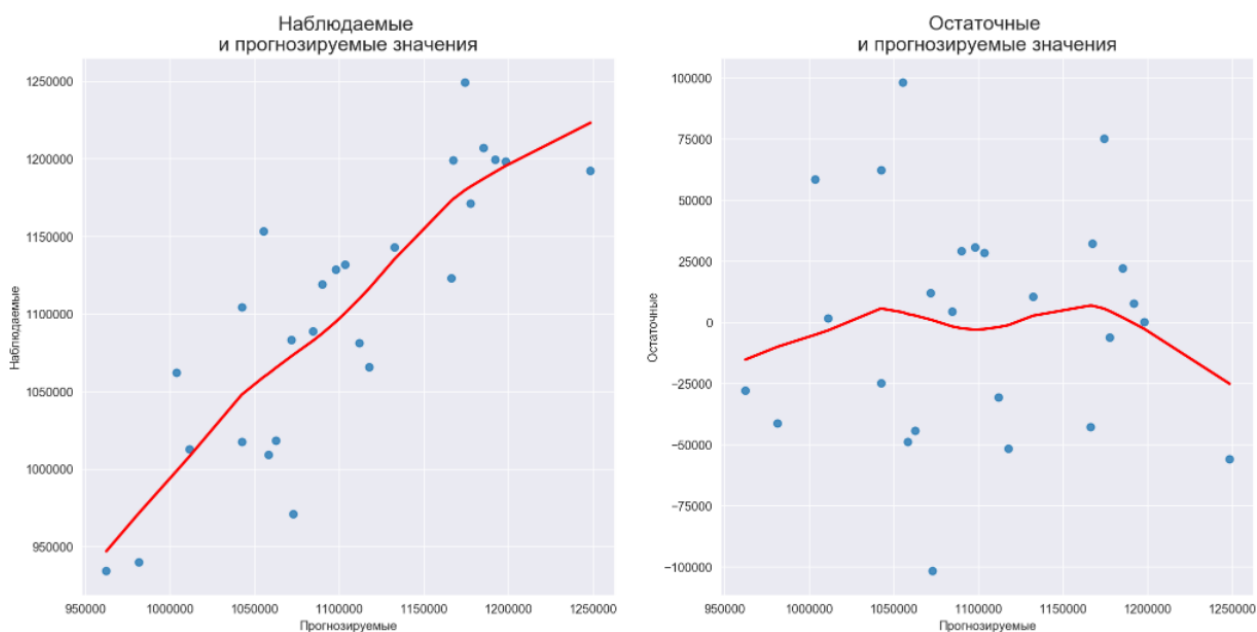


Рис. 17. Проверка на линейность

Проверим модель на предположения о регрессии: автокорреляция, мультиколлинеарность и гомоскедастичность.

Автокорреляция.

Для установления автокорреляции временного ряда в данных воспользуемся критерием Дарбина-Уотсона (рис. 18). Данная статистика равна $DB = 2,133$, что говорит о том, что имеет место незначительная положительная автокорреляция во ременных рядах (рис. 19).

OLS Regression Results

```
=====
Dep. Variable:          y      R-squared:                0.713
Model:                 OLS    Adj. R-squared:           0.642
Method:                Least Squares  F-statistic:              9.961
Date:                  Tue, 17 May 2022  Prob (F-statistic):       6.73e-05
Time:                  17:46:14  Log-Likelihood:          -315.66
No. Observations:      26      AIC:                     643.3
Df Residuals:          20      BIC:                     650.9
Df Model:              5
Covariance Type:      nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	1.95e+05	2.49e+05	0.784	0.442	-3.24e+05	7.14e+05
x1	0.7385	0.319	2.316	0.031	0.073	1.404
x2	0.1710	0.251	0.682	0.503	-0.352	0.694
x3	0.7320	0.402	1.822	0.083	-0.106	1.570
x4	2496.1530	1157.835	2.156	0.043	80.952	4911.353
x5	-47.7623	17.331	-2.756	0.012	-83.915	-11.610

```
=====
Omnibus:                0.042  Durbin-Watson:           2.133
Prob(Omnibus):          0.979  Jarque-Bera (JB):        0.119
Skew:                   0.070  Prob(JB):                 0.942
Kurtosis:               2.699  Cond. No.                 1.94e+07
=====
```

Notes:

- [1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
- [2] The condition number is large, 1.94e+07. This might indicate that there are strong multicollinearity or other numerical problems.

Среднее остатков: 4.2536367590610796e-10

Рис. 18. Результаты регрессии методом МНК

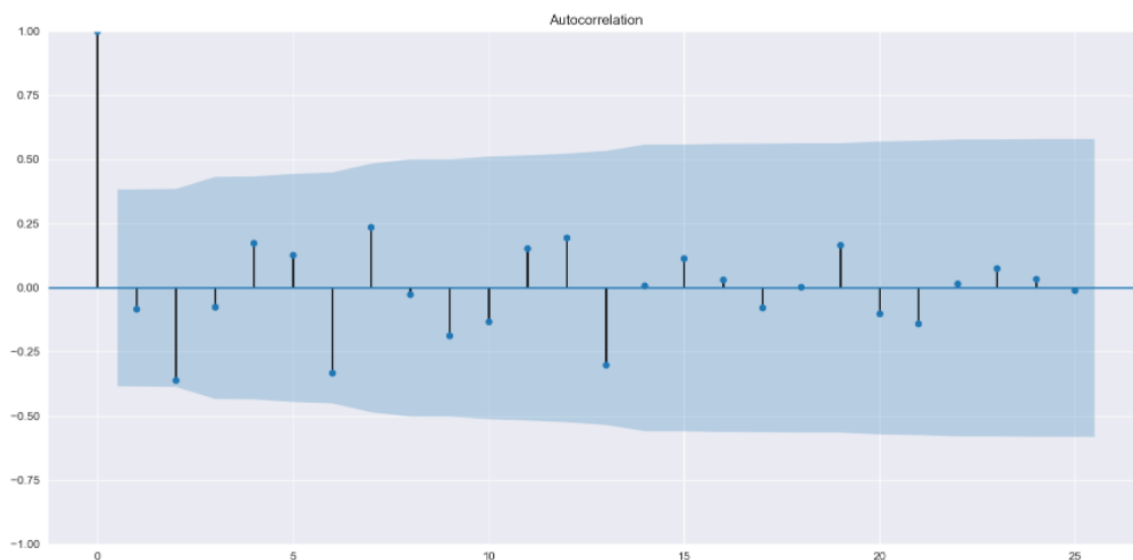


Рис. 19. Автокорреляция остатков

Мультиколлинеарность.

Важность выбора среди нескольких альтернативных моделей «наиболее подходящей», в которой правая часть уравнения не перегружена «лишними» переменными, объясняется, в частности, тем, что включение в модель большого количества объясняющих переменных часто приводит к ситуации, которую называют мультиколлинеарностью. Указанием на то, что p -я объясняющая переменная «почти является» линейной комбинацией остальных объясняющих переменных, может служить большое значение коэффициента возрастания дисперсии (VIF) [26].

$$(VIF)_p = \frac{1}{1 - R_p^2}$$

	Химическая продукция	Машиностроение и металлообработка	Черная металлургия	Жидкие химикаты	грузооборот (млрд т-км)	ВВП (млрд руб)
vif	602.589582	5.775589	4.787289	1.332399	1.615366	2.002298

В соответствии с результатами, изображенными на рис. 18 можно точно сказать о том, что в данных присутствует мультиколлинеарность. Для показателя «Химическая продукция» коэффициент возрастания дисперсии равен 602,589, а также на данное явление показывает число обусловленности матрицы объясняющих переменных: $Cond. No. = 194000000$.

Для борьбы с мультиколлинеарностью проведем анализ ключевых секторов экономики по объемам перевозок методом главных компонент. Идея PCA заключается в уменьшении размерности пула данных за счет выделения сильно коррелирующих между собой объясняющих переменных и объединении их в группы – признаки-комбинации. Главные компоненты между собой не коррелированы. Предварительно данные были нормализованы и стандартизированы. Сначала построим попарные диаграммы рассеяния (рис. 20). Как можно заметить, те показатели, что коррелируют между собой имеют положительную корреляцию.



Рис. 20. Попарные диаграммы рассеяния

Стрелочки, соответствующие исходным координатам, направлены вправо (это соответствует тому, что коэффициенты в первом столбце матрицы вращения все положительные). При этом стрелочки «ЛПК» и «Машиностроение и металлообработка» направлены вверх (и соответствующие коэффициенты во втором столбце (PC2) положительные), а стрелочки «Жидкие химикаты», «Химическая продукция» и «Черная металлургия» – вниз (и их коэффициенты отрицательные). Каждой главной компоненте РС соответствует свой собственный вектор ковариационной матрицы длины 1, который задает направление новой оси - главной компоненты. Каждая главная компонента получается линейной комбинацией старых показателей.

Таким образом, выделено три главных компоненты PC1, PC2 и PC3. Проинтерпретируем полученные результаты: из матрицы на рис. 21 мы видим, например, что в первую главную компоненту все виды перевозок вносят примерно одинаковый вклад, за исключением показателя «Жидкие химикаты». Во вторую главную компоненту наибольший вклад вносит показатель «ЛПК» В третью – «Жидкие химикаты», «Химическая продукция» и «ЛПК». Проиллюстрируем также результаты на рис. 22.

	PC1	PC2	PC3
ЛПК	0.44063169	0.84214547	0.2790529
Химическая. продукция	0.59273628	-0.41292103	0.3420043
Машиностроение. и. металлообработка	0.41253062	0.11576578	-0.7744870
Черная. металлургия	0.52981547	-0.32241185	-0.0626363
Жидкие. химикаты	0.06017873	-0.05418603	0.4487863
	PC4	PC5	
ЛПК	0.1306319	0.04122545	
Химическая. продукция	-0.2416102	0.55028857	
Машиностроение. и. металлообработка	-0.4588712	-0.07761351	
Черная. металлургия	0.5878917	-0.51556400	
Жидкие. химикаты	-0.6069435	-0.65088621	

Рис. 21. Матрица результатов метода главных компонент

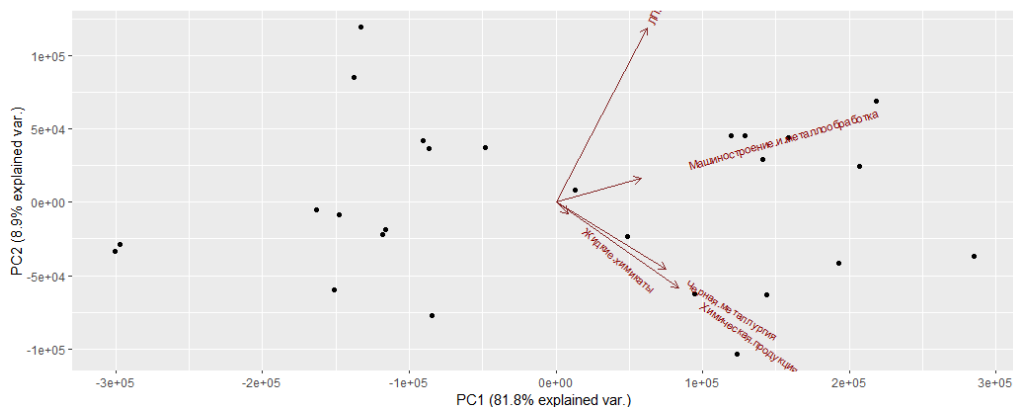


Рис. 22. Зависимость первой и второй главных компонент

Посредством языка программирования Python на рис. 23 и рис. 24 составлены трехмерный и двумерный графики, иллюстрирующие значения результирующего показателя Y в новом пространстве главных компонент.

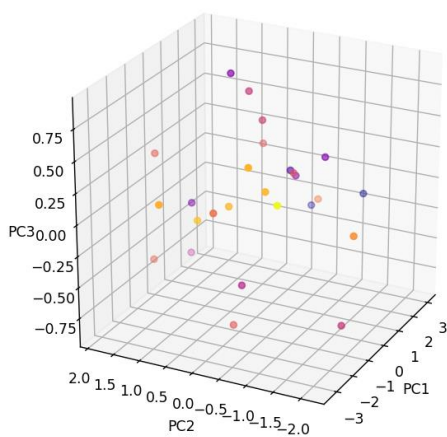


Рис. 23. Метод главных компонент 3D

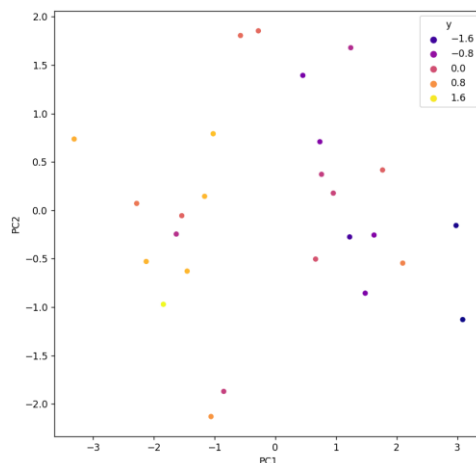


Рис. 24. Метод главных компонент 2D

Вычислим объясненную дисперсию:

Importance of components:	PC1	PC2	PC3	PC4
Standard deviation	1.639e+05	5.393e+04	4.685e+04	2.603e+04
Proportion of Variance	8.179e-01	8.858e-02	6.684e-02	2.063e-02
Cumulative Proportion	8.179e-01	9.064e-01	9.733e-01	9.939e-01
	PC5			
Standard deviation	1.412e+04			
Proportion of Variance	6.070e-03			
Cumulative Proportion	1.000e+00			

Можно увидеть, что PC1 содержит 81,79% дисперсии, PC2 содержит 8,858% дисперсии. Вместе эти компоненты содержат 90,64% информации. Таким образом можно исключить из рассмотрения показатель «Жидкие химикаты».

Гомоскедастичность

Гомоскедастичность, (homoscedasticity) — одинаковость (однородность) дисперсий случайных ошибок в линейной эконометрической модели [26]. Когда остатки не имеют постоянной дисперсии (они проявляют гетероскедастичность), трудно определить истинное стандартное отклонение ошибок прогноза, что обычно приводит к слишком широким/узким доверительным интервалам. Чтобы выяснить, являются ли остатки гомоскедастическими, мы можем посмотреть на график остатков (или стандартизированных остатков) против прогнозируемых значений (рис. 25).

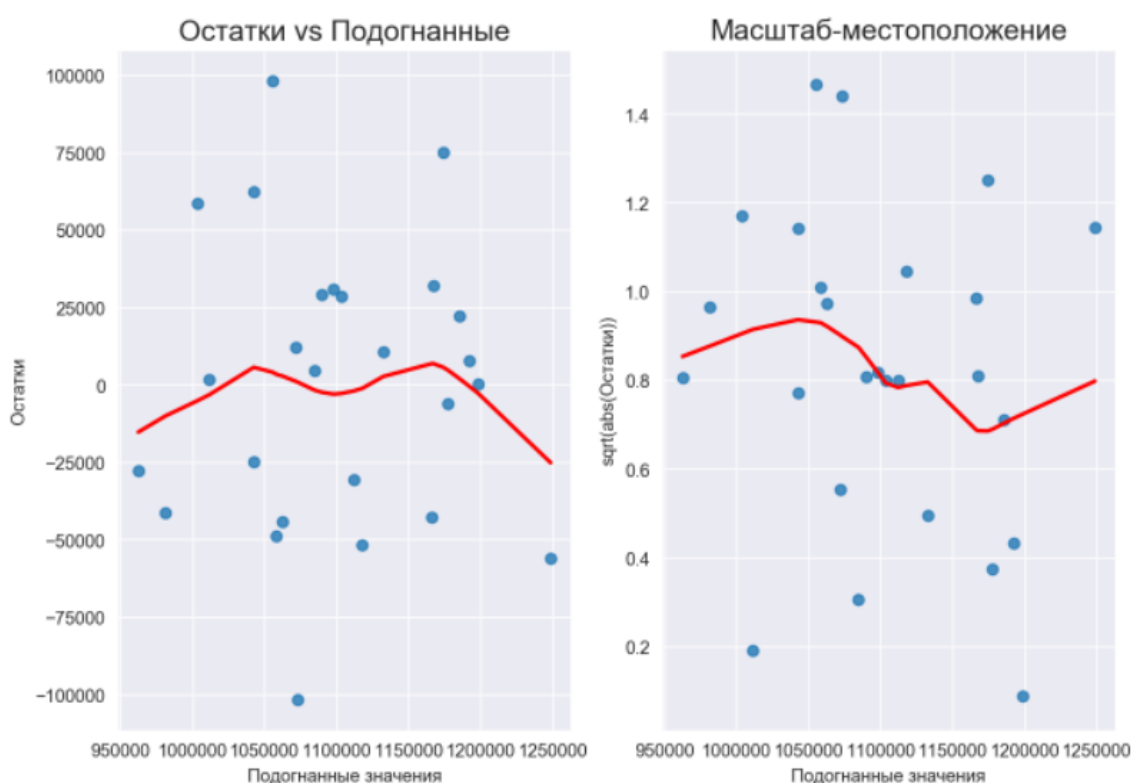


Рис. 25. Проверка на гомоскедастичность остатков

Как можно заметить, размещение точек на графиках случайно и не видна их закономерность (увеличение/уменьшение значений) — красная линия на графиках является практически плоской.

Также воспользуемся двумя статистическими тестами: Breusch-Pagan и Goldfeld-Quandt.

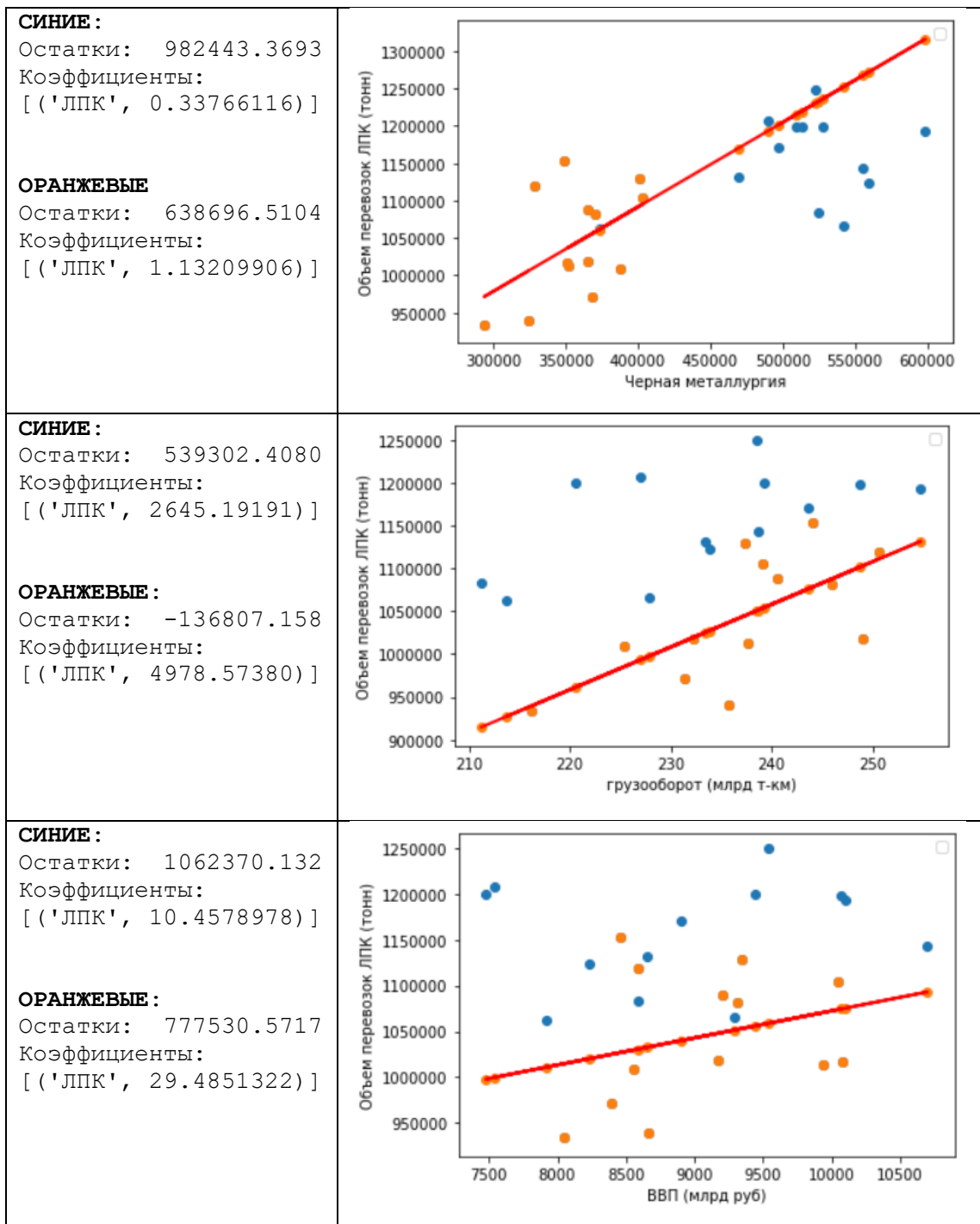
Тест Бройша – Пагана ----		Тест Голдфелда – Куандта ----	
	value		value
Lagrange multiplier statistic	2.394096	F statistic	0.257678
p-value	0.792353	p-value	0.952854
f-value	0.405678		
f p-value	0.839125		

При уровне значимости $\alpha = 0,05$ мы принимаем гипотезу о гомоскедастичности. Следовательно, выполняется еще одно свойство Гаусса-Маркова.

Изобразим графики зависимости показателя «ЛПК» от всех остальных. Синими точками обозначены фактические значения, оранжевыми – плановые. Красной линией обозначена линия регрессии (таблица 6).

Таблица 6. Коэффициенты и графики корреляции и регрессии

<p>СИНИЕ : Остатки: 914992.2369 Коэффициенты: [('ЛПК', 0.38044318)]</p> <p>ОРАНЖЕВЫЕ : Остатки: 822792.145 Коэффициенты: [('ЛПК', 0.4741697)]</p>	
<p>СИНИЕ : Остатки: 908796.5795 Коэффициенты: [('ЛПК', 0.45875254)]</p> <p>ОРАНЖЕВЫЕ : Остатки: 728037.8082 Коэффициенты: [('ЛПК', 0.70809126)]</p>	



Посчитаем оценки степени влияния изменения набора показателей из X на эндогенный показатель y – показатель «ЛПК», то есть величины $A_{x_i}^f$. Согласно формуле (2) получим следующие равенства.

$$\begin{aligned}
 &\text{Оценка влияния показателя «Химическая продукция» на «ЛПК»: } A_{x_{\text{ХИМ}}}^f = \\
 &a_{0_{\text{ХИМ}}}^1 + a_{1_{\text{ХИМ}}}^1 x_{\text{ХИМ}} - a_{0_{\text{ХИМ}}}^0 + a_{1_{\text{ХИМ}}}^0 x_{\text{ХИМ}} = 914992.2369 + 0.38044318x_{\text{ХИМ}} - \\
 &822792.145 + 0.4741697x_{\text{ХИМ}} = 92200.091 + 0.85461288x_{\text{ХИМ}} = 562902.06.
 \end{aligned}$$

Оценка влияния показателя «Машиностроение и металлообработка»:

$$A_{x_{\text{маш}}}^f = 756086.36.$$

Оценка влияния показателя «Черная металлургия»: $A_{x_{\text{чермет}}}^f =$

$$985096.775.$$

Оценка влияния показателя «Грузооборот»: $A_{x_{\text{груз}}}^f = 2194080.19.$

Оценка влияния показателя «ВВП»: $A_{x_{\text{ввп}}}^f = 644569.37.$

Составим множество π^+ :

$$\pi^+ = \{2194080.19 \geq 985096.775 \geq 756086.36 \geq 644569.37 \geq 562902.06 \geq 0\};$$

Можно заключить, что фактор «Грузооборот» оказывает наилучшее влияние на условия достижения планового значения «ЛПК». Следующим по значимости оказывается фактор «Черная металлургия». Фактор «ВВП» является четвертым по значимости.

Следующим этапом проведем статистическое сценарное имитационное моделирование методом Монте-Карло. Для этого сначала определим закон распределения случайных величин (факторов).

Для малой выборки данных (в данной работе число наблюдений равно 26) оценку на нормальность распределения случайной величины достаточно точно показывает статистика Шапиро-Уилка (рис. 26). Можно заключить, что значение p – *value* оказалось больше значения заданного уровня значимости $\alpha = 0,05$, следовательно показатели имеют распределение нормальное или близко к нормальному, за исключением показателя «Черная металлургия». Данный вывод проиллюстрирован на рис. 27 (количество бинов взято семь).

ЛПК: W-статистика: 0.963826060295105 p-value: 0.4723808765411377	Черная металлургия: W-статистика: 0.9093565344810486 p-value: 0.025483038276433945
Химическая продукция: W-статистика: 0.9430975914001465 p-value: 0.15923215448856354	грузооборот (млрд т-км): W-статистика: 0.9648318290710449 p-value: 0.49541208148002625
Машиностроение и металлообработка: W-статистика: 0.9626293182373047 p-value: 0.44591638445854187	ВВП (млрд руб): W-статистика: 0.9727712273597717 p-value: 0.6959710717201233

Рис. 26. Тест Шапиро-Уилка

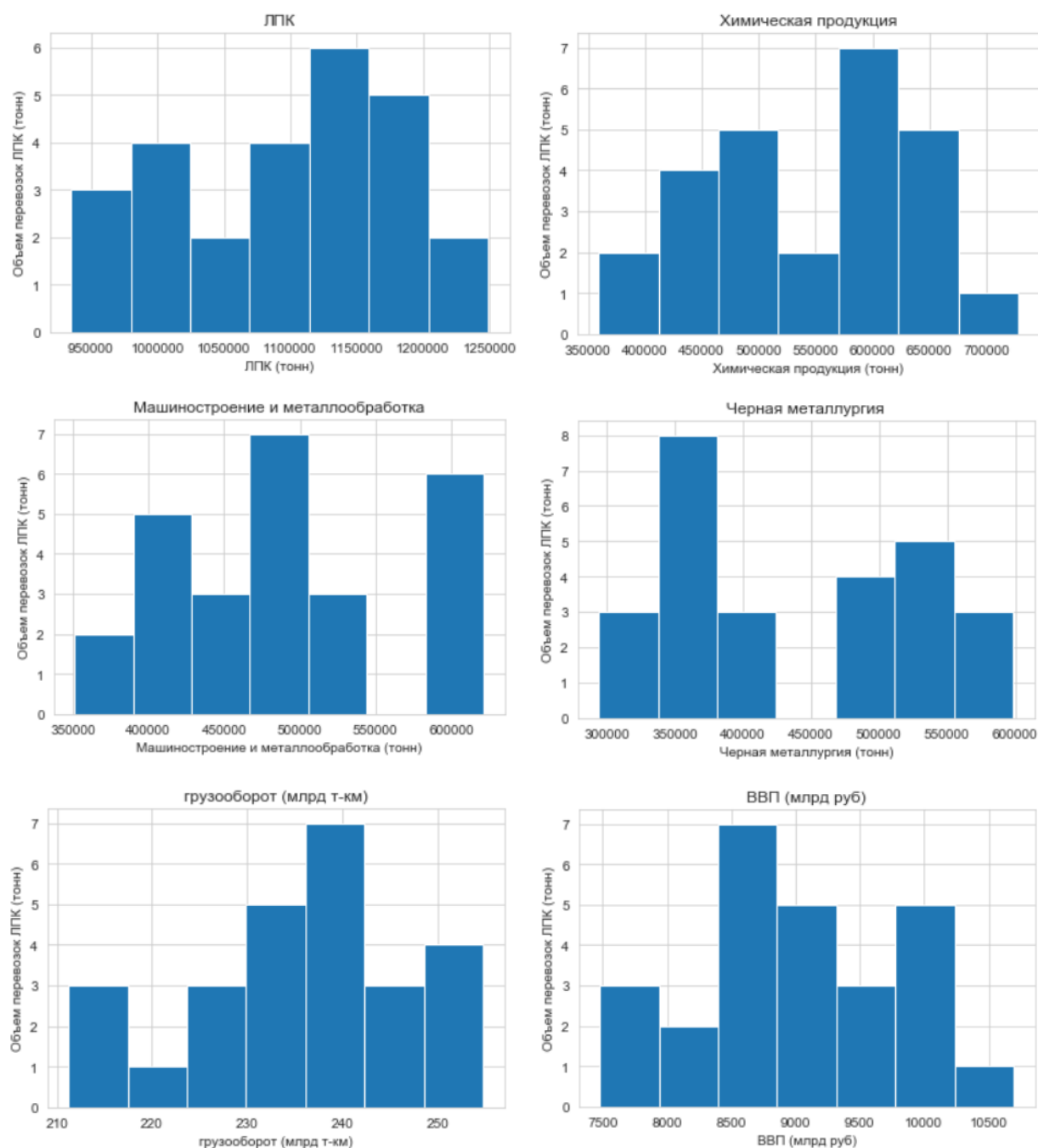


Рис. 27. Гистограммы показателей

На основе вышеизложенного проверим качество предлагаемой модели и построим прогнозную модель до конца 2021 года. В результате применения инфлюентного анализа мы получили, что наиболее благоприятно на достижение плановых значений «ЛПК» влияет показатель «Грузооборот». Так как оба этих показателя имеют различные единицы измерения, следует их стандартизировать и нормализовать с тем, чтобы привести к единой размерности. Методом машинного обучения обучим модель регрессии на 70% выборки данных

«Грузооборота». На рис. 28 представлен график фактических и плановых объемов перевозок грузов лесопромышленного комплекса. Коэффициент детерминации для данной модели равен $R^2 = 0.635$.

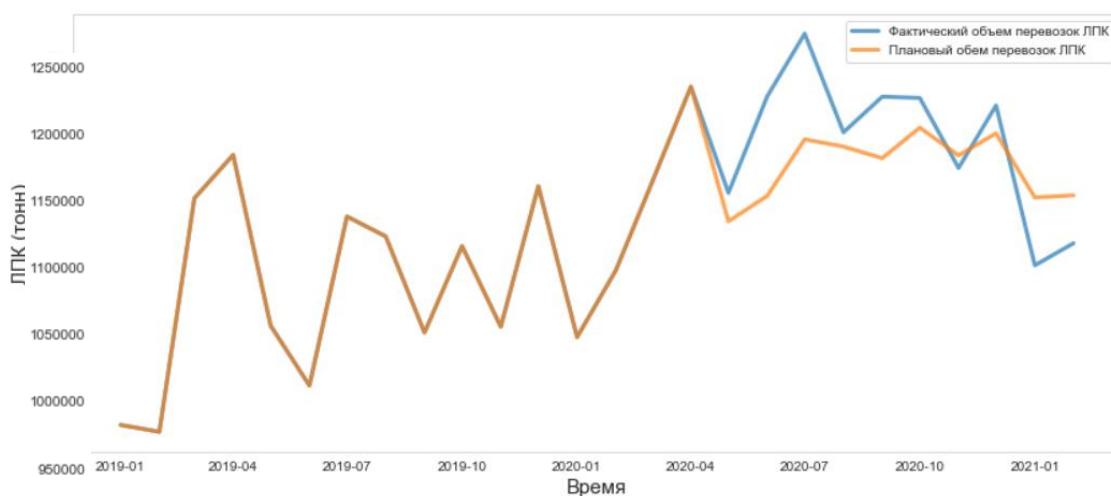


Рис. 28. Оценка качества модели

Описательная статистика многофакторной модели регрессии от всех показателей представлена на рис. 29. Коэффициент детерминации теперь равен $R^2 = 0.998$ против $R^2 = 0.713$ на рис. 18. Качество модели значительно увеличилось.

OLS Regression Results						
Dep. Variable:	y	R-squared (uncentered):	0.998			
Model:	OLS	Adj. R-squared (uncentered):	0.998			
Method:	Least Squares	F-statistic:	2306.			
Date:	Thu, 19 May 2022	Prob (F-statistic):	5.09e-28			
Time:	19:23:47	Log-Likelihood:	-316.63			
No. Observations:	26	AIC:	643.3			
Df Residuals:	21	BIC:	649.6			
Df Model:	5					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
x1	0.6528	0.312	2.093	0.049	0.004	1.301
x2	0.4643	0.284	1.633	0.117	-0.127	1.055
x3	-0.2493	0.420	-0.593	0.559	-1.123	0.625
x4	4479.6748	588.039	7.618	0.000	3256.781	5702.569
x5	-48.1382	17.576	-2.739	0.012	-84.689	-11.587
Omnibus:	0.985	Durbin-Watson:	2.443			
Prob(Omnibus):	0.611	Jarque-Bera (JB):	0.807			
Skew:	-0.095	Prob(JB):	0.668			
Kurtosis:	2.158	Cond. No.	4.99e+04			

Рис. 29. Описательная статистика

Проведем статистическое имитационное моделирование методом Монте-Карло. Разыграем десять случайных величин, которые будут интерпретировать временной период с марта 2021 года по декабрь 2021 года. На рис. 30

представлены диаграммы корреляции и линии регрессии зависимого показателя от объясняющих. Как и прежде, синие точки — фактические значения, оранжевые — плановые. Линия регрессии достаточно точно описывает фактические значения в зависимости между ЛПК и Черной металлургией.

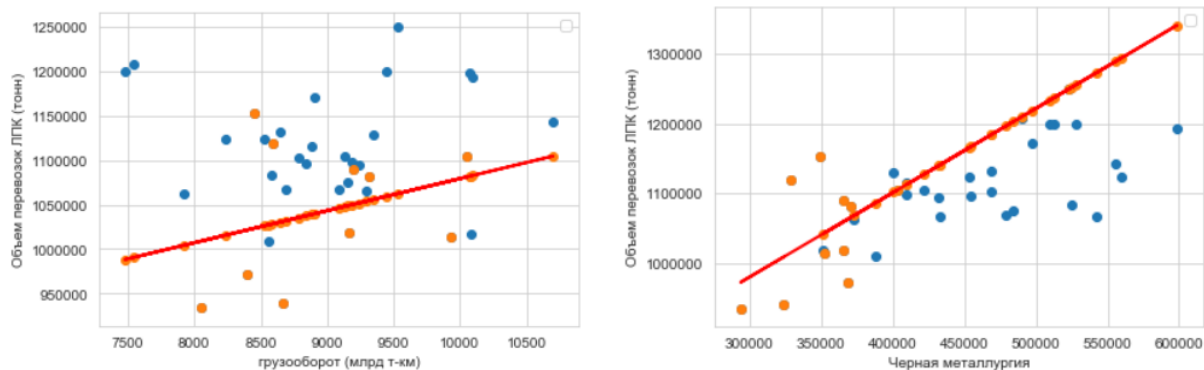


Рис. 30. Метод Монте-Карло

Построим прогноз объема перевозок товаров лесопромышленного комплекса (ЛПК) до конца 2021 года на основе спрогнозированных значений модели линейной регрессии по данным грузооборота (рис. 31). Синяя линия обозначает прогнозные значения.

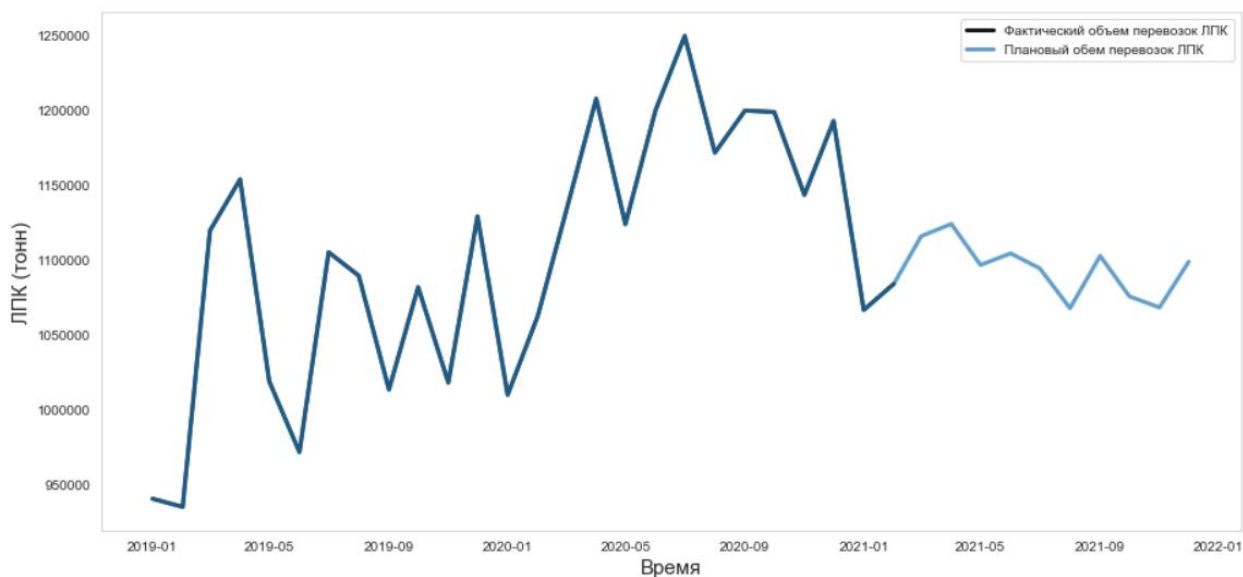


Рис. 31. Прогноз объема перевозок ЛПК

Можно заключить, что спрос на железнодорожные перевозки грузов лесопромышленного комплекса до конца 2021 года является достаточно стабильным и полученная модель качественно соответствует имеющимся данным.

3.3 Анализ полученных результатов моделирования

Исследование компании показало, что ключевыми перевозками грузов контейнерами она занимается в отраслях розничной торговли, автомобилестроения, сельского хозяйства, химической промышленности и отрасли пиломатериалов. Обладая широким спектром сервисов, компания обладает достаточным влиянием на внутрироссийском рынке. На основе выявленных показателей составлена схема данных (информационно-логическая модель данных), позволяющая агрегировать и структурировать выявленные факторы вместе с тем, чтобы управлять данными централизованно через СУБД.

Интеллектуальный анализ предоставленных данных (см. приложение А) показал следующее. В имеющемся наборе информации отражены объемы перевозок грузов внутри России контейнерами в тоннах по отраслям. Проведенный ABC-анализ показал, что на лесопромышленный комплекс (ЛПК), химическую продукцию, машиностроение и металлообработку, черную металлургию и жидкие химикаты приходится 72% всех контейнерных перевозок в целом по России за 2019–2021 гг. Остальные четыре типа перевозок из группы В (таблица 5) дополняют общий грузооборот до 85%.

Предполагаемая модель линейной многофакторной регрессии проверялась на гипотезы об автокорреляции, мультиколлинеарности и гомоскедастичности. Автокорреляция не выявлена. Обнаружена сильная мультиколлинеарность между показателями, для борьбы с которой применен метод главных компонент. Можно уменьшить размерность исходного пула данных (рис. 20–24). Показано, что объясненная дисперсия трех компонент покрывает 97,33% информации обо всех перевозках.

Проверка гипотезы о гомоскедастичности проводилась посредством тестов Бройша-Пагана и Голдфелда-Куандта. Выявлено, что в показателях присутствует одинаковость (однородность) дисперсий случайных ошибок.

Далее, при решении задачи прогнозирования построены зависимости экзогенного показателя y – «ЛПК» от эндогенных, объясняющих факторов X .

Синими точками отображены фактические значения показателей, а оранжевыми – плановые. Красной линией обозначена прямая регрессии между показателями. Для нахождения оценок изменения набора объясняющих показателей X на объясняемый y были найдены инфлюенты. Показано, что факторы «Грузооборот», «Черная металлургия» и «Машиностроение и металлообработка» оказывают наиболее благоприятное влияние на условия достижения планового значения «ЛПК». Показатели «ВВП» и «Химическая продукция» оказывают менее значимое влияние.

Прогнозирование будущих значений для объема перевозок лесопромышленного комплекса проводилось с помощью метода Монте-Карло. С помощью теста Шапиро-Уилка (рис. 26) определено, что распределения всех случайных величин (кроме «Черная металлургия») подчиняются нормальному закону. Разыграны десять случайных величин посредством проведения 100 экспериментов, которые интерпретируются десятью месяцами (с марта 2021 по декабрь 2021 г.). На рис. 31 изображен график прогноза объема перевозок ЛПК до конца 2021 года. По полученному результату можно заключить, что объем перевозок лесоматериалов в тоннах контейнерами в 2021 году достаточно стабилен.

Выводы

Качественный анализ предметной области показал, что в целом в России имеется положительная динамика грузевых контейнеров и спрос на них внутри страны является эластичным. В частности, лидирующие позиции в 2020 г. приходились на предприятия обрабатывающего производства, торговую сферу, деятельность по операциям с недвижимостью и добычу полезных ископаемых – 17%, 14,1%, 10,6%, 10,5% соответственно. Сфера сельского хозяйства занимает 4,7%, уступая отрасли строительства 1,1% от ВРП. В это же время динамика внутреннего грузооборота была положительной. Построена информационно-логическая модель, отражающая рынок железнодорожных контейнерных перевозок как систему взаимосвязанных факторов. Выявленные внешние макроэкономические индикаторы непосредственным образом влияют на качество работы всей отрасли перевозок. Для дальнейшего описания рассматриваемого участка предметной области применены методы математического моделирования.

Ключевыми задачами в данной работе являются сбор и обработка данных. Они разрешимы путем использования передовых инструментов интеллектуального анализа данных, таких как Data mining и Business Intelligence. В ходе работы проанализированы различные источники данных и построена схема данных. В качестве основного инструмента для построения математической модели взят метод статистического инфлюентного анализа. В нем оценки степени влияния изменения экзогенных факторов на эндогенные (инфлюенты) отражают то, насколько тот или иной показатель благоприятно влияет на достижение плановых значений результирующего показателя. Инфлюенты (оценки) находятся через композитную факторную модель регрессии. Таким образом, при дальнейшем прогнозировании значений Y минимизируются отклонения в соответствии с системой взаимосвязанных факторов. Далее мы подходим к непосредственному применению полученных аналитических результатов на практике.

Для понимания положения рассматриваемой компании на рынке проанализирована ее деятельность и составлена схема сервисов, по которым работает данная транспортная компания в России. Также приведен ее краткий анализ экономического состояния. К имеющемуся пулу данных, состоящему из 22 видов грузов применен ABC-анализ для выявления наиболее значимых показателей, вкладывающих наибольшее значение в общий объем перевозок за указанный период времени. Во время математического моделирования проведена предобработка данных с использованием математических и статистических инструментов, в частности проверены три предположения о линейной регрессии: автокорреляция, мультиколлинеарность и гомоскедастичность. Для борьбы с высокой мультиколлинеарностью использован метод главных компонент, позволивший уменьшить размерность данных на один показатель – «Жидкие химикаты». Далее, в таблице 6 показаны корреляционные зависимости и линии регрессии объясняющих и объясняемой переменных. Посчитан коэффициент детерминации для оценки качества полученной регрессионной модели: $R^2 = 0.998$ против $R^2 = 0.713$, которое значение получено на первоначальной исходной модели.

Перед проведением статистического имитационного моделирования методом Монте-Карло проверены тесты Шапиро-Уилка на нормальность распределения выборок. Имеющими нормальный закон распределения или близкий к нормальному имеют все показатели, за исключением «Черная металлургия».

На рис. 31 изображен прогноз объема перевозок ЛПК на основе полученной модели регрессии от значения показателя «Грузооборота». Можно заключить, что до конца 2021 года объем перевозок лесопромышленного комплекса является стабильным и полученная модель качественно соответствует имеющимся данным.

Заключение

Фундаментальный и технический анализы вместе являются мощным инструментом, который помогает оценить ситуацию на рынке железнодорожной отрасли. Большая конкуренция в этой сфере вынуждает компании – частных железнодорожных грузовых перевозчиков применять передовые технологии и приемы в управлении своим бизнесом для увеличения своей конкурентоспособности. Ответственное лицо, принимающее решение в данных вопросах должно владеть глубинными знаниями экономики железнодорожной транспортной системы и понимать взаимосвязи влияния макроэкономических показателей на данную отрасль.

В ходе выполнения выпускной квалификационной работы достигнута цель: составлен прогноз объема спроса в отрасли железнодорожных контейнерных перевозок в России путем выявления значимых макроэкономических показателей за счет применения фундаментального и технического анализа.

Для достижения этой цели, выполнены все поставленные задачи:

1. Выполнен обзор литературы по теме исследования, изучена предметная область и ее экономический анализ. Выявлены сущности предметной области и построена информационно-логической модель.

2. Проведен сбор и обработка данных. Проведен качественный и количественный анализ макроэкономических показателей. Построена математическая модель.

3. Проведен вероятностно-статистического анализ. Произведено прогнозирование и сценарное имитационное моделирование.

Сбор данных осуществлен из открытых источников и проведена их обработка. Большую значимость имеют предоставленные данные компанией об объемах перевозок грузов. Количественный анализ данных проводился с использованием языка программирования Python с библиотеками pandas, scikit-learn, numpy, statsmodels и другими, языка программирования R для

статистической обработки данных и работы с графикой, а также посредством программных продуктов MS Excel и Erwin Data Modeler.

Посредством интеллектуального анализа данных и применения математической модели составлен прогноз объема спроса на перевозки грузов лесопромышленного комплекса. Построенная регрессионная модель является достаточно качественной, и она соответствует данным. Беря во внимание анализ транспортной компании, приведенной в данной работе, можно заключить, что она имеет достаточное преимущество к удовлетворению спроса на такие перевозки, покрыв его своим предложением. С учетом внешних макроэкономических факторов аналогичные прогнозы можно составить для других видов грузов.

Полученная в данной работе модель может быть использована руководством транспортных компаний для анализа своей деятельности и прогнозирования объема спроса на контейнерные перевозки. Принимая во внимание полученные результаты, ЛПР может принимать решения для удовлетворения спроса предоставлением соответствующего предложения на перевозки. В дальнейшем, выполненный технический анализ может быть автоматизирован в виде программного продукта.

Список литературы

1. Институт исследования проблем железнодорожного транспорта. Целевая модель рынка грузовых железнодорожных перевозок на период до 2025 года. Москва: 2018.
2. Транспорт в России. Статистический сборник // Федеральная служба государственной статистики. 2020. URL: https://rosstat.gov.ru/storage/mediabank/UbzIvBZj/Transport_2020.pdf (дата обращения: 28.04.2022).
3. Кокорев В.А. Экономические провалы. Москва: Общество купцов и промышленников России, 2005. 336 с.
4. Терёшина Н.П., Галабурда В.Г., Токарев В.А. Экономика железнодорожного транспорта: Учеб. для вузов ж.-д. транспорта. Москва: УМЦ ЖДТ, 2008. 996 с.
5. Терёшина Н.П., Галабурда В.Г., Трихунков М.Ф. Экономика железнодорожного транспорта: учебник для вузов ж.-д. транспорта. Москва: УМЦ ЖДТ, 2006. 801 с.
6. Быченко О.Г., Сыцко А.Ф. Экономика железнодорожного транспорта: учебное пособие. Гомель: Белорусский государственный университет транспорта, 2017. 223 с.
7. Мачерет Д. Динамика железнодорожных перевозок грузов как макроэкономических индикаторов // Экономическая политика, Т. 10, № 2, 2015. С. 133-150.
8. Мачерет Д. О чем свидетельствует столетняя динамика показателей крупнейших железнодорожных систем // Экономическая политика. 2016. Т. 11. № 6. С. 138–169.
9. АО "Кодекс". Об утверждении Методики классификации и специализации железнодорожных линий ОАО "РЖД" // Электронный фонд правовых и

нормативно-технических документов. 2020. URL: <https://docs.cntd.ru/document/564286419> (дата обращения: 28.04.2022).

10. Jianguo Z., Gang Q. International conference on computer science and service system // Application of BP neural network forecast model based on principal component analysis in railways freight forecas. 2012. pp. 2201-2204.
11. Ziyadin S., Sousa R.D., Suieubayeva S., Yergobek D., Serikbekuly A. The 1st International Conference on Business Technology for a Sustainable Environmental System // Differentiation of logistics services on the basis ABC analysis. 2020. Vol. 159. pp. 1-8.
12. Перминова А.А. Прогнозирование спроса и оценка конкурентоспособности контейнерных перевозок на железных дорогах России. Москва. 2016. 139 с. Диссертация на соискание ученой степени кандидата экономических наук.
13. Milenković M., Vojović N. Railway demand forecasting // IGI Global. Jun 2016. No. 5. pp. 100-129.
14. Трухаев Р.И., Горшков И.С. Факторный анализ в организационных системах. Москва: "Радио и связь", 1985. 184 с.
15. Kulshreshtha M., Nag B., Kulshrestha M. A multivariate cointegrating vector auto regressive model of freight transport demand: evidence from Indian railways // Transportation Research. Jul 2001. Vol. Part A. No. 35. pp. 29-45.
16. Организация объединенных наций. Реформа на железнодорожном транспорте в регионе ЕЭК // UNECE. 2018. URL: https://unece.org/sites/default/files/2020-12/1737000_R_ECE_TRANS_261.pdf (дата обращения: 28.04.2022).
17. Главное контейнерное направление России [Электронный ресурс] // FESCO: [сайт]. [2018]. URL: <https://www.fesco.ru/ru/press-center/blog/glavnoe-konteynernoje-napravlenie-rossii/> (дата обращения: 17.05.2022).

18. Россия увеличила несырьевой экспорт в Китай по итогам 2020 года [Электронный ресурс] // Министерство экономического развития Российской Федерации: [сайт]. [2021]. URL: https://www.economy.gov.ru/material/news/rossiya_uvelichila_nesyreyvoy_eksport_v_kitay_po_itogam_2020_goda.html (дата обращения: 28.04.2022).
19. РЖД в цифрах [Электронный ресурс] // РЖД: [сайт]. [2022]. URL: <https://company.rzd.ru/ru/9377#cargo> (дата обращения: 28.04.2022).
20. Babcock M.W., Gayle P.G. Specifying and estimating a regional agriculture railroad demand model // Journal of the transportation research forum, Vol. 53, No. 1, Spring 2014. pp. 35-48.
21. Grenzeback L.R., Brown A., Fischer M.J. Freight transportation demand: energy-efficient scenarios for a low-carbon future. Washington, DC: 2013.
22. Forbes [Электронный ресурс] // Транзит грузов из Китая в Европу по железной дороге через Россию вырос почти на 50%: [сайт]. [2021]. URL: <https://www.forbes.ru/biznes/443349-tranzit-gruzov-iz-kitaa-v-evropu-po-zeleznoj-doroge-cerez-rossiu-vyros-pochti-na-50> (дата обращения: 28.04.2022).
23. Евразия эксперт [Электронный ресурс] // Ставка на несырьевую торговлю: как меняются экономические отношения России и Китая в 2021 году: [сайт]. [2021]. URL: <https://eurasia.expert/kak-menyayutsya-ekonomicheskie-otnosheniya-rossii-i-kitaya-v-2021/> (дата обращения: 28.04.2022).
24. Железнодорожные перевозки грузов [Электронный ресурс] // Евросиб: [сайт]. URL: <http://www.eurosib.biz/ru/produkty-i-uslugi/uslugi-zheleznodorozhnogo-operatora/zheleznodorozhnye-perevozki> (дата обращения: 28.04.2022).
25. Трухаев Р.И. Инфлюентный анализ и принятие решений. Москва: "Наука", 1984. 236 с.
26. Носко В.П. Эконометрика. Т. 1. Москва: "Дело", 2011. 672 с.

Приложение А

Сегмент	Рынок янв 19	Рынок фев 19	Рынок мар 19	Рынок апр 19	Рынок май 19	Рынок июн 19	Рынок июл 19	Рынок авг 19	Рынок сен 19	Рынок окт 19	Рынок ноя 19	Рынок дек 19
Общий итог по рынку	3 375 615	3 452 401	4 005 765	3 980 470	3 875 158	3 886 804	4 240 869	4 151 698	3 983 855	4 237 792	4 123 051	4 410 267
ЛПК	940 054	934 602	1 119 164	1 153 440	1 018 421	971 184	1 104 764	1 089 044	1 013 004	1 081 420	1 017 650	1 128 713
Химическая продукция	358 995	373 480	448 113	421 184	439 188	476 841	498 014	478 482	457 305	474 592	490 251	600 418
Машиностроение и металлообра- ботка	359 975	394 487	424 394	416 330	438 876	460 116	492 318	480 080	474 989	495 538	493 275	461 825
Черная металлургия	324 093	293 906	328 754	349 420	365 331	368 586	403 504	365 178	352 607	370 667	351 469	400 650
Жидкие химикаты	396 797	369 341	423 316	392 853	406 099	407 960	406 910	425 611	391 574	434 811	443 865	474 366
Цветная металлургия	157 312	156 671	186 291	182 856	184 612	171 656	202 903	177 287	166 426	183 579	174 631	187 363
Пищевая промышленность	84 340	115 701	151 220	156 695	146 297	135 815	147 416	151 487	197 771	231 694	225 892	189 815
Минерально-строительные грузы	112 079	143 513	167 801	170 675	170 053	184 261	198 517	190 731	174 673	174 387	150 705	149 855
ТНП	125 945	125 058	115 280	108 985	123 219	140 332	186 107	178 066	151 950	150 567	134 084	148 276
Колесная, с/х и специальная техника	116 282	120 347	139 580	144 652	120 862	135 538	133 765	135 304	138 983	163 325	155 863	149 641
Удобрения и сырье для их пр-ва	99 943	114 007	117 333	104 213	98 820	78 815	100 695	89 768	85 273	81 122	81 527	85 552
Руды и концентраты (металлургия)	53 291	41 634	66 031	68 811	62 048	56 763	65 062	71 932	67 961	73 962	72 626	88 724
Нефтяные грузы	39 430	51 814	79 048	76 038	74 322	76 105	82 167	88 114	60 504	59 392	67 104	81 831
С/Х (растениеводство)	39 050	54 261	65 776	73 620	66 451	59 054	57 603	55 001	65 761	76 534	74 881	91 720
Газы энергетические	67 281	62 465	67 452	63 800	60 241	59 915	52 808	69 411	68 781	67 291	66 468	62 144
Легкая промышленность	44 713	42 622	37 944	42 508	48 528	45 327	51 202	53 930	50 028	49 294	52 238	49 895
С/Х (животноводство)	36 615	39 408	46 255	37 286	33 002	37 355	42 286	38 738	50 641	43 964	45 098	36 020
Кальцинированная сода	5 365	5 302	8 482	574	5 507	8 096	602	465	2 974	13 046	11 997	10 621
Прочие грузы	9 013	7 084	5 744	8 772	5 491	5 604	6 764	6 506	5 412	5 652	6 113	6 381
Фармацевтика	4 682	6 426	6 745	6 575	7 362	7 210	7 176	6 296	6 971	6 692	7 036	5 886
Каменный уголь	64	27	51	123	5	10	76	29	69	125	98	157
С/Х (прочие)	296	245	991	1 060	423	261	210	238	198	138	180	414

Сегмент	Рынок янв 20	Рынок фев 20	Рынок мар 20	Рынок апр 20	Рынок май 20	Рынок июн 20	Рынок июл 20	Рынок авг 20	Рынок сен 20	Рынок окт 20	Рынок ноя 20	Рынок дек 20
Общий итог по рынку	4 003	3 994	4 462	4 614	4 634	4 644	4 996	4 716	4 942	5 145	5 000	5 263
	569	586	330	893	697	981	909	552	338	319	990	737
ЛПК	1 009	1 062	1 131	1 207	1 123	1 199	1 249	1 171	1 199	1 198	1 143	1 192
Химическая продукция	445	245	920	365	469	645	291	261	454	446	079	498
Машиностроение и металлообра- ботка	566 627	531 999	621 770	616 474	657 468	597 232	653 056	613 753	613 121	654 740	669 751	728 507
Черная металлургия	421 191	350 940	426 385	501 220	505 331	536 954	607 213	598 219	583 734	621 705	587 236	614 610
Жидкие химикаты	387 839	373 291	469 164	490 042	559 388	509 220	522 913	497 167	513 021	528 219	555 740	598 691
Цветная металлургия	453 278	451 253	469 507	450 411	425 832	437 040	411 650	372 871	419 593	428 494	442 914	469 029
Пищевая промышленность	197 115	197 664	210 615	236 685	245 456	211 290	247 053	235 547	244 605	260 297	249 875	260 987
Минерально-строительные грузы	119 054	156 753	208 702	218 766	228 058	217 113	208 923	215 724	265 583	284 358	248 281	240 694
ТНП	111 726	150 813	184 411	163 024	173 740	194 813	216 348	202 414	207 177	188 890	168 051	181 193
Колесная, с/х и специальная техника	157 004	106 759	110 403	131 899	137 604	181 253	215 801	218 023	201 792	193 988	175 163	179 156
Удобрения и сырье для их пр-ва	135 928	126 896	138 564	98 387	109 477	126 495	144 701	121 132	162 570	176 741	172 888	163 209
Руды и концентраты (металлургия)	91 957	92 690	86 692	129 325	113 047	110 761	113 712	90 566	99 795	117 214	106 417	91 121
Нефтяные грузы	77 519	72 184	58 442	61 950	66 915	31 404	85 286	69 584	95 222	109 266	97 238	103 634
С/Х (растениеводство)	56 720	77 765	93 168	72 016	69 136	71 975	75 255	63 196	65 947	65 661	73 223	87 920
Газы энергетические	47 434	68 038	78 711	72 435	63 245	60 540	69 120	57 688	69 653	100 522	94 912	114 914
Легкая промышленность	66 058	67 664	68 661	51 671	41 411	38 539	47 502	47 925	48 531	43 350	45 752	45 626
С/Х (животноводство)	50 068	36 438	33 564	41 378	43 580	52 125	61 635	65 879	65 943	73 767	68 119	68 578
Кальцинированная сода	27 870	39 196	40 361	37 217	44 980	44 318	46 493	50 909	51 964	46 257	45 613	53 388
Прочие грузы	13 204	17 751	15 396	13 435	5 716	2 586	5 947	10 911	19 339	24 509	21 631	18 983
Фармацевтика	8 112	9 088	8 719	11 779	12 145	12 953	8 129	7 393	7 707	8 785	8 913	9 613
Каменный уголь	5 037	4 637	6 248	8 750	7 980	8 395	6 565	6 107	6 707	6 277	6 348	6 262
С/Х (прочие)	31	24	94	15	64	40	5	12	506	13 319	18 877	32 827
	352	498	833	664	655	290	316	271	374	403	369	1 017

Сегмент	Рынок янв 21	Рынок фев 21	Рынок итого
Общий итог по рынку	4 624 585	4 616 333	104 144 646
ЛПК	1 066 159	1 083 676	26 459 578
Химическая продукция	661 411	617 451	13 041 361
Машиностроение и металлообра- ботка	535 048	537 693	11 746 941
Черная металлургия	541 858	524 744	10 278 860
Жидкие химикаты	423 682	372 397	10 205 375
Цветная металлургия	248 848	209 642	4 928 776
Пищевая промышленность	154 202	202 750	4 546 152
Минерально-строительные грузы	140 605	154 978	4 129 850
ТНП	175 846	170 984	3 696 714
Колесная, с/х и специальная техника	143 635	139 974	3 331 130
Удобрения и сырье для их пр-ва	103 514	92 619	2 380 365
Руды и концентраты (металлургия)	86 356	85 367	1 717 489
Нефтяные грузы	56 109	69 386	1 707 851
С/Х (растениеводство)	90 500	104 303	1 676 924
Газы энергетические	42 851	41 383	1 380 747
Легкая промышленность	62 189	57 881	1 229 303
С/Х (животноводство)	27 143	50 808	1 015 234
Кальцинированная сода	16 162	16 331	242 439
Прочие грузы	6 531	10 198	191 872
Фармацевтика	5 738	6 901	158 370
Каменный уголь	35 396	65 815	66 648
С/Х (прочие)	432	542	10 696