

**Федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА  
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ  
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

**Молчанова Г.О., Рей А.И., Шагаров Д.Ю.,  
Андропова Е.С.**

**Обнаружение сговора на государственных закупках  
в России**

**Москва 2020**

**Аннотация.** При государственных закупках на рынках часто возникают ситуации, когда участники, вместо того, чтобы конкурировать друг с другом за государственные контракты, распределяют между собой доли на рынке государственного заказа и завышают ценовой уровень выше конкурентного. В настоящей работе предлагается использовать методы эконометрики, математической статистики и машинного обучения для прогнозирования вероятности того, что закупка с участием данных контрагентов на стороне поставщика и других кандидатов на победу в конкурсе, будет предметом успешной жалобы в ФАС с точки зрения картельного сговора между поставщиками.

Рей А.И. заведующий научно-исследовательской лабораторией анализа данных и отраслевой динамики ИОРИ. Российской академии народного хозяйства и государственной службы при Президенте РФ

Шагаров Д.Ю. младший научный сотрудник лаборатории анализа данных и отраслевой динамики ИОРИ. Российской академии народного хозяйства и государственной службы при Президенте РФ

Андропова Е.С. младший научный сотрудник лаборатории анализа данных и отраслевой динамики ИОРИ. Российской академии народного хозяйства и государственной службы при Президенте РФ

Молчанова Г.О. младший научный сотрудник лаборатории анализа данных и отраслевой динамики ИОРИ Российской академии народного хозяйства и государственной службы при Президенте РФ

Данная работа подготовлена на основе материалов научно-исследовательской работы, выполненной в соответствии с Государственным заданием РАНХиГС при Президенте Российской Федерации на 2019 год

## Оглавление

<b>ВВЕДЕНИЕ.....</b>	<b>5</b>
<b>1. МЕТОДЫ ВЫЯВЛЕНИЯ КАРТЕЛЕЙ.....</b>	<b>6</b>
1.1 ПОВЕДЕНЧЕСКИЙ И СТРУКТУРНЫЙ ПОДХОДЫ.....	7
1.2 МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ.....	9
1.4 ЭВОЛЮЦИЯ РОССИЙСКОГО АНТИКАРТЕЛЬНОГО ЗАКОНОДАТЕЛЬСТВА.....	10
<b>2. СОСТОЯНИЕ КОНКУРЕНЦИИ НА ЭЛЕКТРОННЫХ АУКЦИОНАХ В РОССИИ И ПРАКТИКА АНТИМОНОПОЛЬНОЙ СЛУЖБЫ ПО ВЫЯВЛЕНИЮ КАРТЕЛЕЙ.....</b>	<b>12</b>
2.1 ОБЩИЕ ТЕНДЕНЦИИ В ПРИМЕНЕНИИ ЭЛЕКТРОННЫХ АУКЦИОНОВ КАК ВИДА ЗАКУПОЧНЫХ ПРОЦЕДУР.....	12
2.2 ПРОЦЕДУРА ПРОВЕДЕНИЯ ЭЛЕКТРОННОГО АУКЦИОНА В РОССИИ.....	14
2.2.1 Анतिकонкурентные действия на электронных аукционах.....	15
2.3 ПРАКТИКА ВЫЯВЛЕНИЯ АНТИМОНОПОЛЬНОЙ СЛУЖБОЙ СГОВОРОВ НА ТОРГАХ.....	16
<b>3. ДАННЫЕ И АЛГОРИТМЫ ИХ СБОРА И ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ.....</b>	<b>21</b>
<b>4. СТРАТЕГИЯ ОЦЕНИВАНИЯ И ГИПОТЕЗЫ.....</b>	<b>26</b>
<b>5. ЖАЛОБЫ В ФАС.....</b>	<b>39</b>
<b>6. МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ.....</b>	<b>41</b>
6.1 ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ.....	42
6.2 МЕТОД БЛИЖАЙШИХ СОСЕДЕЙ.....	44
6.3 СЛУЧАЙНЫЙ ЛЕС.....	45
6.4 ГРАДИЕНТНЫЙ БУСТИНГ.....	46
<b>7. КОРРЕКТИРОВКА КЛАССИФИКАТОРА ЗА СЧЕТ УЧЕТА НЕСБАЛАНСИРОВАННОСТИ ДАННЫХ.....</b>	<b>50</b>
7.1 СПОСОБЫ КОРРЕКТИРОВКИ ДАННЫХ.....	50
7.3 ПЕРЕБАЛАНСИРОВКА ВЫБОРКИ ДЛЯ КЛАССИФИКАТОРА ПО КАРТЕЛЯМ.....	52
<b>8. ДОБАВЛЕНИЕ ФИКТИВНЫХ ПЕРЕМЕННЫХ ПО ВИДАМ ЭКОНОМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ.....</b>	<b>54</b>

8.1 СБОР И АНАЛИЗ ДАННЫХ ОБ ОБЪЕКТЕ ЗАКУПКИ.....	55
8.2 РЕЗУЛЬТАТЫ КЛАССИФИКАЦИИ С УЧЕТОМ ВИДОВ ЭКОНОМИЧЕСКОЙ ДЕЯТЕЛЬНОСТИ.....	62
<b>9. EX-АНТЕ АНАЛИЗ ПРИЗНАКОВ СГОВОРОВ.....</b>	<b>66</b>
9.1 ОБЗОР ОСНОВНЫХ ЭЛЕМЕНТОВ ТЕОРИИ ГРАФОВ.....	67
9.2 ОПИСАНИЕ СТРАТЕГИИ И ПЕРВЫЕ РЕЗУЛЬТАТЫ.....	68
9.3 АПРОБИРОВАНИЕ ГРАФО-ТЕОРЕТИЧЕСКОГО ПОДХОДА К ВЫЯВЛЕНИЮ КАРТЕЛЕЙ МЕЖДУ ПОСТАВЩИКАМИ.....	71
<b>ЗАКЛЮЧЕНИЕ.....</b>	<b>82</b>
<b>СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....</b>	<b>84</b>
<b>ПРИЛОЖЕНИЯ.....</b>	<b>86</b>

## ВВЕДЕНИЕ

При государственных закупках на рынках часто возникают ситуации, когда участники, вместо того, чтобы конкурировать друг с другом за государственные контракты, распределяют между собой доли на рынке государственного заказа и завышают ценовой уровень выше конкурентного. Чтобы обеспечить раскрытие этой противозаконной и наносящей экономический ущерб общественному благосостоянию деятельности, могут применяться различные методы. В настоящей работе предлагается использовать методы эконометрики, математической статистики и машинного обучения для прогнозирования вероятности того, что закупка с участием данных контрагентов на стороне поставщика и других кандидатов на победу в конкурсе, будет предметом успешной жалобы в ФАС с точки зрения картельного сговора между поставщиками.

Предполагалась проверка следующей основной гипотезы: в закупках (или последовательностях таковых), для которых имеется решение ФАС о выявлении сговора, имеются статистически значимые нелинейные отличия в характеристиках цен заявок от прочих закупок (или их последовательностей).

В качестве объекта исследования выступает генеральная совокупность государственных закупок в Российской Федерации в период с 2014 по 2018 год. Для анализа берутся товары с определением победителя по критерию минимальной цены. Применяются эконометрические методы, методы машинного обучения (для анализа *ex post*) и графо-теоретический подход (для анализа *ex ante*).

## 1. Методы выявления картелей

В сговоре при проведении закупок выделяется вертикальный (с участием заказчика и его представителей, исходя из несовпадения интересов принципала - общества, и агента) и горизонтальный (без участия заказчика) сговор. В настоящей работе будет рассматриваться только горизонтальный сговор, именуемый иначе картелем. Необходимо понимать, что в условиях жесткой системы госзакупок в России подавляющее большинство закупок проходит через предварительные (и вытесненные за формальные рамки закона) консультации заказчика и предпочитаемого исполнителя,

В литературе выделяются шесть аспектов деятельности картелей (по Коннору, 2001, 2007 [13, р.27]):

### *Ценообразование:*

- единый прейскурант или единые цены сделок, в том числе по формулам
- стоимость и условия доставки
- единообразные скидки или наценки
- скоординированные заявки на тендере
- расчётные валютные курсы, отличающиеся от рыночных/официальных
- условия «ценовой защиты»

### *Доли на рынке:*

- глобальные доли продаж
- глобальные количественные квоты
- региональные или национальные доли
- закрепление за фирмами покупателей
- распределение победителей на тендерах

### *Производство:*

- глобальное сокращение выпуска
- ограничения на экспорт
- ограничения на географический арбитраж покупателями
- ограничение производственных мощностей
- отказ от пр-ва или продаж

### *Мониторинг:*

- частый/подробный обмен информацией о продажах
- сертификация доли на рынке третьей стороной
- инспекции предприятий или складов

условия «найдёте цену ниже - возместим разницу»

*Санкции:*

фирма-лидер применяет санкции к нарушителям  
механизм регулярных компенсаций  
сбытовое агентство для общих продаж  
объединение и раздел прибылей  
соглашения о пороговых ценах

*Прикрытие:*

уничтожение улики по поездкам / встречам / общению / мониторингу  
создание или использование отраслевой ассоциации  
использование шифров или кодовых обозначений

В рамках данного исследования мы не наблюдаем сговор непосредственно, и вынуждены полагаться на вторичные его признаки из открытых источников. В этом разделе будут рассматриваться существующие в мировой экономической науке подходы обнаружения картелей, а также практические особенности обнаружения горизонтального сговора в госзакупках силами ФАС. Существенным элементом литературы, нашедшим свое применение в ходе исследования, является изучение факторов срока «жизни» картеля. Кроме того, для понимания контекста правоприменения в данный раздел входит также обзор эволюции антикартельного права в России.

Сговоры на торгах представляют серьезный риск для эффективности госзакупок. Борьба со сговором на торгах (bid-rigging) стала одним из важнейших направлений для антимонопольных органов во всем мире; этой проблеме была посвящена рекомендация ОЭСР 2012 года и соответствующий доклад 2016 года [34]. Совершенствование закупочных процедур, их цифровизация помогают предотвращать и выявлять картели, но в то же время приводят к появлению новых антиконкурентных схем поведения. В данном разделе рассматриваются существующие методы выявления картелей, признаки сговора и условия его существования.

### **1.1 Поведенческий и структурный подходы**

Методы обнаружения картелей делятся на структурные и поведенческие. Структурные предполагают выявление рынков с характеристиками, которые способствуют сговору, и проведение сфокусированного расследования работы фирм на этих рынках. Например, в работах [39, 16] показано, что создание картеля вероятнее в отрасли с меньшим количеством фирм, более однородной продукцией и стабильным спросом. Во внимание принимаются три категории факторов:

- структурные факторы (число конкурентов, прозрачность рынка, барьеры входа);
- факторы предложения: степень однородности продукции, издержки фирм, отдача от инноваций;
- факторы спроса: колебания, эластичность и рост спроса, сильная покупательная способность.

Поведенческие методы направлены на выявление картелей путем анализа действий фирм на рынках и их координации между собой. Поведенческий подход также может быть сосредоточен на влиянии такой координации на рынок – на структуру цен, количество фирм, динамику инвестиций и т.д. На подозрение о сговоре может навести параллельное движение цен или необоснованный их рост [18, 30].

Указанные методы используются либо в *ex ante*, либо в *ex post* анализе картелей.

Обычно выявление картелей базируется на одном из следующих вопросов [18]:

- Является ли поведение несовместимым с конкуренцией?
- Существует ли структурный разрыв в поведении фирм?
- Отличается ли поведение фирм, подозреваемых в сговоре, от поведения фирм-конкурентов?

Поведение, характерное для конкуренции

Ответ на первый вопрос базируется на выявлении свойств поведения, которые всегда имеются в условиях конкуренции или, по крайней мере, характерны для широкого класса конкурентных моделей, и проверке гипотезы их наличия для конкретной отрасли. Такой подход к тестированию имеется, например, в исследованиях Портера, Зоны [37, 38] и Баджари и Е [5].

Структурный разрыв в поведении

Второй метод выявления сговора – поиск структурного разрыва в поведении фирмы. Проверяется, изменились ли после предполагаемого момента создания (распада) картеля средняя цена, взаимосвязь между ценами фирм и т.д. [13, 1, 2]. Так, в статьях Абрантес-Метц и соавторов и Болотовой и др. показывается, что средняя цена на продукцию во время сговора возрастает, а стандартное отклонение цен, наоборот, уменьшается [1, 8].

Структурные изменения могут быть и не связаны со сговором. Например, создание ассоциации может привести к усилению корреляции цен компаний, поскольку способствует обмену информацией, но неясно, приведет ли это к росту средних цен.

Сравнение подозреваемых в сговоре с контрольной группой

При сравнительном анализе в качестве бенчмарка могут выступать сопоставимые рынки, на которых фирмы не считаются сговорившимися; в частности, различные географические рынки для данного продукта или услуги. Для панельных данных в качестве бенчмарка может выступать период до создания или после распада картеля. В одной статье российских авторов (Mogozov, Podkolzina, 2013 [31]) выборка разделялась на основе разницы между ценой победителя и резервной ценой. Для каждой подвыборки строилась регрессия относительной цены контракта на количество участников, число позиций (лотов) в закупке, объем текущих контрактов победителя и опыт победителя.

#### Контроль на издержки фирмы

При объяснении величины ставок участников торгов в качестве контрольных переменных чаще всего выделяют расстояние до места заключения контрактов, количество выполняемых фирмами контрактов и специализацию каждой фирмы [37, 36, 4]. Чем ближе фирма располагается к месту выполнения контракта, чем больше у неё свободных мощностей, и чем более точно соответствует специализация фирмы объекту закупки, тем с большей вероятностью фирма выставит наименьшую ставку и выиграет контракт.

Структурный и поведенческие методы дополняют друг друга, снижая тем самым вероятности ошибок I и II рода. Структурный метод часто дает слишком много ложных положительных результатов, например, для отраслей с высокими барьерами входа. С другой стороны, структурные показатели (экраны) дают мало ложных отрицательных результатов, поэтому могут помочь исключить отрасли менее рискованные с точки зрения картелей, и оставить только часть отраслей для использования поведенческих экранов.

### **1.2 Методы машинного обучения**

В последние годы для выявления картелей стали чаще применяться методы машинного обучения [20, 21].

Имхоф и Хубер (2019) [20] используют простые статистические показатели для того, чтобы определить, как сговор влияет на распределение ставок. Они выдвигают несколько предположений и доказывают их на основе данных о крупном картеле в Тичино (Швейцария), существовавшем на протяжении несколько лет. В частности, авторы показывают, что при сговоре на аукционе:

- уменьшается коэффициент вариации ставок;
- распределение ставок становится более асимметричным из-за большей разницы между первой и второй самыми низкими ставками.

Всего рассчитано 6 показателей, к которым в качестве контрольных были добавлены число участников и цена контракта. На основе них применяются методы машинного обучения: лассо для логит-моделей и ансамблевый классификатор. Как показатели эффективности прогнозирования выбраны две меры: средние квадратичные ошибки прогнозируемых вероятностей сговора в тестовой выборке и доля правильных классификаций.

Машинное обучение также вошло в основу подхода в работе Иванова и Нестерова (2019) [21], где анализируются данные о более чем 1,4 млн. запросов котировок на российском рынке в период с 2014 по 2018 год. Авторы пытались выявить «утечку заявок» (bid leakage) – коррупционную схему, при которой заказчик незаконно предоставляет предпочитаемому участнику информацию о других заявках [3]. Результаты исследования показывают, что утечка заявок более вероятна на аукционах с более высокой резервной ценой, меньшим числом участников и меньшим снижением цен, а также в тех случаях, когда выигравшая заявка поступает в последний час до истечения крайнего срока.

В исследовании Иванова и Нестерова вводится важная предпосылка: проигравшие участники считаются честными, а победители – с какой-то вероятностью коррумпированными. Описываемая в статье процедура выявления утечки заявок состоит из двух этапов. На первом этапе обучается бинарный классификатор, чтобы отличать победителей от занявших второе место. С использованием кросс-валидации получают прогнозы классификатора для всех победителей в наборе данных. На втором этапе эти прогнозы преобразуются в вероятности утечки заявок.

#### **1.4 Эволюция российского антикартельного законодательства**

Первый нормативно-правовой акт появился в Российской Федерации при принятии Закона РСФСР «О конкуренции и ограничении монополистической деятельности на товарных рынках» от 22 марта 1991 года. Летом 1991 года был создан государственный орган - Государственный комитет Российской Федерации по антимонопольной политике и поддержке новых экономических структур, сейчас существующий в форме Федеральной антимонопольной службы.

В данный момент действует Федеральный закон №135-ФЗ от 26 июля 2006 года «О защите конкуренции». В этом законе антиконкурентные соглашения представлены в двух статьях – ст. 11, которая посвящена запрету на ограничивающие конкуренцию соглашения и согласованные действия хозяйствующих субъектов и ст. 16, посвященной запрету на антиконкурентные соглашения с участием органов гос.власти.

В 2012 году при принятии «3-го антимонопольного пакета» было закреплено понятие «картеля» как соглашения между хозяйствующими субъектами-конкурентами. Согласно «3-му пакету» за картельный сговор<sup>1</sup> предусматривается уголовная ответственность по ст. 178 УК РФ. При этом согласованные действия отделяются от понятия картеля и соглашений, и не подпадают под статью. Для компаний-участников картеля предусмотрена административная ответственность в виде оборотного штрафа до 15% от годового оборота и уголовная ответственность до семи лет лишения свободы.

Согласно «3-му антимонопольному пакету» кроме письменных и устных документов, фирмы обязаны предоставлять антимонопольному органу свою электронную переписку.

В 2015 году в «4-м антимонопольном пакете» понятие картелей было расширено на картели покупателей, уточнены условия их допустимости и согласования с ФАС.

Была запрещена координация организаторами торгов или заказчиками деятельности их участников и заключение соглашений между организаторами торгов и (или) заказчиками с участниками этих торгов, если такие действия могут привести к ограничению конкуренции.

В феврале 2019 года ФАС России внесла на рассмотрение в правительство поправки к Закону о защите конкуренции и Кодексу об административных правонарушениях, так называемый «5-й антимонопольный пакет». Согласно новым поправкам планируется ввести новые понятия, такие как «сетевые эффекты», «ценовой алгоритм», а также уточнить понятие «соглашение». Также планируется ужесточить наказание за создание картеля через поправки в УК РФ.

---

<sup>1</sup> Антимонопольный орган не может определить картель согласно ч. 7 ст. 11 №135-ФЗ, если в нем участвуют лица, находящиеся под общим управлением, имеющие общего генерального директора, или у одного из субъектов по отношению к другому установлен контроль.

## **2. Состояние конкуренции на электронных аукционах в России и практика антимонопольной службы по выявлению картелей**

В настоящем разделе проводится обзор практики электронных аукционов, типовые антиконкурентные действия участников и стандартные процедуры ФАС РФ по их выявлению и правоприменению.

### **2.1 Общие тенденции в применении электронных аукционов как вида закупочных процедур**

В работе мы концентрируем внимание на электронных аукционах. Открытый аукцион в электронной форме (ОАЭФ) представляет собой торги, которые проводятся в электронном виде на электронных площадках, и в ходе которых победителем становится участник, чье ценовое предложение по выполнению контракта оказалось наиболее низким среди всех ставок участников. ОАЭФ является наиболее популярным способом определения поставщика (исполнителя) для госзакупки. В 2018 году в форме электронного аукциона было опубликовано 1,96 млн. извещений общим объемом 5,66 трлн. рублей, что составило 68,7% от объема всех госзакупок по данным Минфина [43]. Второе место заняли закупки у единственного поставщика, объем которых достиг 1,12 трлн. рублей (13,5% от общего объема госзакупок).

Госзакупки на российском рынке осуществляются разными способами, которые регламентируются федеральным законом 44-ФЗ:

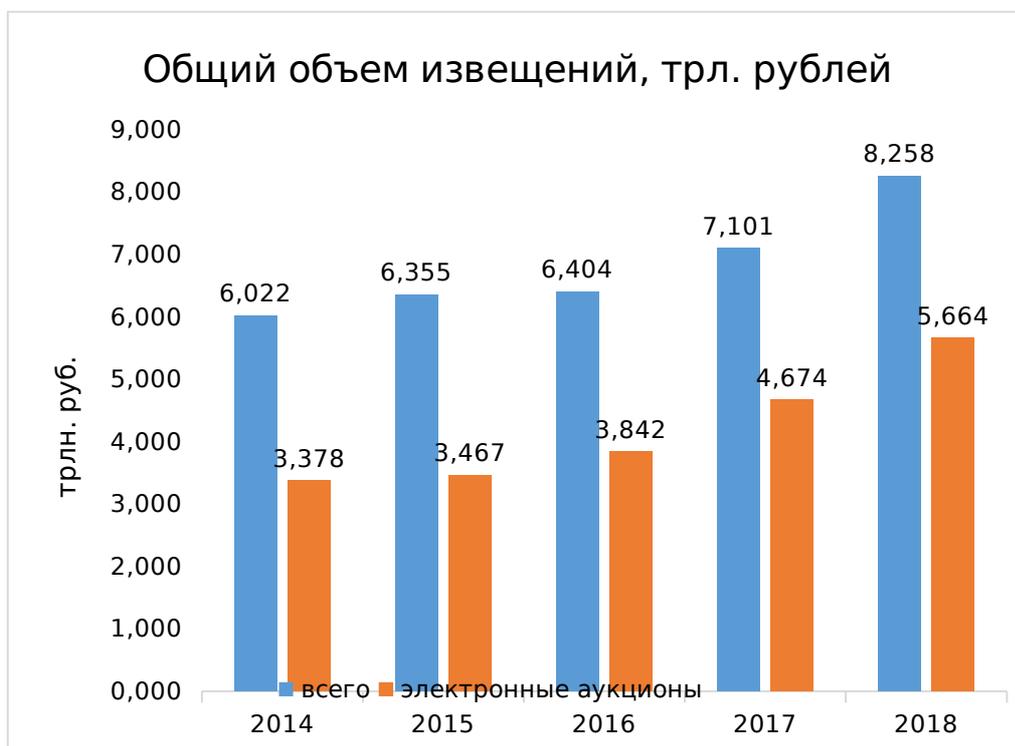
1. Закупка у единственного поставщика
2. Запрос котировок
3. Запрос предложений
4. Электронный аукцион
5. Конкурс.

Особенности разных видов закупок обуславливают возможности применения разных схем и стратегий поведения, как вертикальных, так и горизонтальных [6, 3]. Не все схемы применимы ко всем видам закупок.

### **2.1 Общие тенденции в применении электронных аукционов как вида закупочных процедур**

Мы концентрируем внимание на ОАЭФ (ст. 59 ФЗ-44), которые с каждым годом становятся все более популярным способом определения поставщика для госзакупок. Их доля в общем объеме закупок выросла с 55% в 2015 до 69% в 2018 году, а по количеству извещений – с 57% до 60%, достигнув 1,96 млн. извещений на общую сумму 5,66 трлн.

рублей [45, 43, 44]. Денежный объем ОАЭФ стабильно растет, что видно из ежегодных докладов о мониторинге в сфере закупок товаров и услуг по Ф3-44 (Рисунок 1).



Источники: Минэкономразвития (2016, 2017), Минфин (2018, 2019)

### Рисунок 1. Объем извещений по государственным закупкам, трлн. рублей

Эта динамика во многом объясняется рядом преимуществ электронного аукциона:

- Информация о проводимых электронных аукционах хранится в открытом доступе, и производителям легко отслеживать предложения по закупкам.
- Дистанционное взаимодействие с заказчиком без посредников уменьшает денежные и временные издержки.
- Заявки анонимные, и только при оглашении результатов называется фирма-победитель, это снижает коррупционные риски. Все этапы проведения электронного аукциона фиксируются, информация размещается на сайте электронной площадки, это увеличивает прозрачность торгов.

Тем не менее, возможности для нечестных сделок сохраняются и в рамках электронного аукциона, и встречаются как вертикальные сговоры между заказчиком и поставщиком, так и горизонтальные между поставщиками. То, что электронные аукционы дают доступ к большому количеству крупных госзаказов, усиливает стимулы участников торгов к антиконкурентным соглашениям. Для того, чтобы понять, какие стратегии поведения возможны со стороны оппортунистически настроенных фирм, рассмотрим элементы электронного аукциона.

## 2.2 Процедура проведения электронного аукциона в России

Процедура проведения ОАЭФ подчинена закону №44-ФЗ и состоит из нескольких этапов (Рисунок 2):

1. В единой информационной системе (ЕИС) в открытом доступе размещается извещение о проведении закупки и документация.

2. На сбор заявок для аукционов с начальной максимальной ценой (НМЦК) меньше 3 млн. рублей участникам дается минимум 7 дней, для аукционов с большей НМЦК – минимум 15 дней.

3. Заказчик рассматривает первые части заявок и готовит протокол рассмотрения заявок (Протокол-1), где указано, какие поставщики допускаются до аукциона, а какие – нет.

4. Аукцион проходит в 1-й рабочий день после 2 дней с момента завершения рассмотрения первых частей заявок. На подачу ценовых предложений отводится 10 минут, после каждого предложения время продлевается на 10 минут. Если в течение 10 минут после подачи последнего ценового предложения никто из участников больше не сделал более низкую ставку, то аукцион завершается. Участники делают ценовые предложения, снижая текущую цену в пределах «шага аукциона», который составляет 0,5%-5% от НМЦК (ст.68 44-ФЗ). После завершения аукциона в течение 30 минут публикуется протокол о проведении аукциона (Протокол-2).

5. После проведения аукциона в течение 3 рабочих дней рассматриваются вторые части заявок, после чего принимается решение о соответствии\несоответствии фирм требованиям аукциона и публикуется протокол рассмотрения вторых частей заявок (Протокол-3). Только на этом этапе называются фирмы-участники.

6. После рассмотрения вторых частей заявок подводятся итоги аукциона, определяется победитель и подписывается контракт.



обеспечено путем дополнительного соглашения с заказчиком. Например, заказчик может установить такие требования к продукции, что техническому заданию будут соответствовать только заявки участников картеля.

Последнее время стали применяться «аукционные роботы», которых программируют на минимальное снижение цены.

2. «**Таран**». Ещё одной популярной стратегией является «Таран», при котором добросовестные участники вытесняются с торгов. Идея схемы состоит в том, что два или несколько членов картеля активно торгуются, демпингуют и этим поведением отпугивают добросовестных участников, которые понимают, что не могут так сильно снижать цену и в какой-то момент останавливаются. Ещё один член картеля делает ставку чуть ниже добросовестных участников, но выше «таранов». По итогам рассмотрения вторых частей заявок оказывается, что фирмы, которые играли роль «таранов», не проходят по требованиям к документации, и в итоге выигрывает участник картеля, сделавший более высокую ставку.

Схемы при вертикальном сговоре (соглашение между участником и заказчиком) были подробно описаны в предыдущей главе. Из них применимо к электронному аукциону и включают сдерживание подачи заявок, отклонение заявок и отмена аукциона.

### **2.3 Практика выявления антимонопольной службой сговоров на торгах**

Картели, или сговоры, являются наиболее опасной формой ограничения конкуренции, поэтому по отношению к участникам сговоров применяются жесткие санкции. Согласно антимонопольной практике к нелегальным по букве закона картельным соглашениям относятся три типа сговоров – сговор о ценах, раздел рынка и сговор участников торгов [42]. Другие виды соглашений рассматриваются на основе взвешенного подхода.

По мнению представителей ФАС, оценка состояния конкуренции в 2018 году говорит о сохранении негативных тенденций 2017 года, а также о увеличении опасности антиконкурентных соглашений, наиболее опасными из которых являются картели.

Согласно данным ФАС, ежегодный объем госзакупок равен 25 трлн. рублей, что составляет около 30% ВВП, из них в электронной форме: около 7 трлн. рублей (по 44-ФЗ) и почти 2 трлн. рублей (по 223-ФЗ) – закупки у малого и среднего бизнеса. Ущерб от картелей оценивается в 1,5 – 2% ВВП.

В 2018 году было возбуждено 384 дела о картелях, что на 6% больше, чем в 2017 году (360 дел). Преобладающую долю из возбужденных ФАС дел о картелях являются сговоры на торгах: в 2017 и 2018 гг. – 86%, в 2016 г показатель достиг 90% (Рисунок 3). Сам по себе этот факт может указывать на наличие систематической погрешности в механизмах поиска картелей и/или принятия решения по возбуждению дел по ним, однако мы пока не располагаем статистической информацией, способной пролить свет на это потенциальное обстоятельство.



Источник: ФАС России (<https://fas.gov.ru>)

**Рисунок 3. Динамика количества дел по выявлению картелей и сговоров на торгах за период 2015-2018 гг.**

Картели широко распространены в госзакупках, закупках компаний с гос.участием и закупках в сфере гособоронзаказа, в большинстве сфер экономики (Рисунок 4).



Источник: Доклад о состоянии конкуренции в РФ за 2018 год (ФАС России)

**Рисунок 4. Отраслевые особенности нарушений по ст. 11, 16  
Закона о защите конкуренции за 2018 г.**

Отличительной особенностью картельных соглашений в России является участие в них органов власти, организаций и должностных лиц. Сокрытие и/или обеспечение заключения и исполнения антиконкурентных соглашений происходит из-за злоупотребления должностными полномочиями или получения взятки. В 2018 г. ФАС России было возбуждено 300 дел о заключении антиконкурентных соглашений с органами власти, что на 32% больше, чем в предыдущем году (204 дела).

С целью заключения и реализации антиконкурентных соглашений участники торгов начали использовать современные технологии: специальные программные модули, интегрированные в интерфейс электронных торговых площадок («аукционные роботы», «режим автоторга»). Участились случаи противоправного применения специального программного обеспечения (коммерческие продукты/коммерческие «аукционные роботы»).

Методы, применяемые ФАС для выявления сговоров, основаны на мониторинге и анализе конкурентных процедур. Единый информационный портал госзакупок дает доступ к большим массивам данных, на основе которых можно выявлять признаки возможных картельных соглашений.

На основании опыта разбирательств, начатых ФАС, можно выделить следующие доказательства наличия сговора между хозяйствующими субъектами:

- Минимальное снижение НМЦК;
- Одинаковые:
  - IP-адреса, с которых фирмы отправляют заявки;
  - электронные адреса, почты, телефоны фирм;
  - учетные записи, под которыми созданы заявки;
  - размеры файлов;
  - дата/время создания, внесения изменений заявок;
- Сходство текстов заявок; Аффилированность участников:
  - Идентичность лицензионных документов;
  - Идентичность торговых наименований;
  - Общие учредители;
  - Совпадение или близкое расположение фактических адресов компаний;
- Связи со сторонней организацией, которая, возможно, координирует участие в торгах;
- Договоренности о получении квалифицированного сертификата ключа проверки электронной подписи;
- Заключение договоров субподряда между участниками;
- Устойчивые финансовые связи между участниками.

Обобщая опыт антимонопольных органов России по правоприменению в сфере борьбы с картелями, можно утверждать, что принятие законодательства о госзакупках и передача контрольных функций в ФАС России были в определенной степени аналогом того, что произошло с антимонопольными органами в США после принятия закона Харта-Скотта-Родайно<sup>2</sup> в 1976 году: они позволили перейти к постоянному потоку правоприменительных действий, которые могут быть сравнительно безопасными и выгодными с политической точки зрения, и стабильно обеспечивают высокие цифры количественных показателей эффективности в годовой отчетности. Недостатком этого потока является отвлечение сил и средств от обнаружения сговора в других сферах экономики.

Вместе с тем в ряде случаев между анализом госзакупок и анализом коммерческого рынка в целом на наличие картеля возникает синергия. Так, сотрудничество ФАС и торговых площадок позволяет ФАС получать необработанную первичную информацию по заявкам и их авторству. Во многом этому способствует низкая компьютерная грамотность участников картелей, которые явно не задумываются о том, что они своей деятельностью генерируют компьютерно-криминалистические доказательства для дел ФАС.

---

<sup>2</sup> Hart-Scott-Rodino Antitrust Improvements Act of 1976.

Можно рационально предполагать, что это временное явление, и через 2-3 года участники картелей будут тщательнее маскировать следы сговора, и ФАС будет вынуждена полагаться на те же самые косвенные индикаторы, которые используются в настоящей работе. Отдельную обеспокоенность может вызвать гипотеза о том, что те картели в госзакупках, которые используют более серьезные меры маскировки своей деятельности, систематически отличаются от картелей, выявляемых ФАС. Здесь и далее предполагаем, что эта систематическая смещенность либо не существует, либо ей можно пренебречь.

### 3. Данные и алгоритмы их сбора и предварительной обработки

В этом разделе представлен метод обнаружения картелей, основанный на ex-post анализе сговоров на закупках: модель строится с учетом данных о том, на каких из аукционов был обнаружен сговор. В качестве источников данных для модели были использованы Единая информационная система госзакупок (ЕИС) и сайт ФАС.

Стратегия выявления картелей включает два шага. На первом шаге на основе данных о закупках и о решениях ФАС определяется, какие есть характерные признаки у аукционов, на которых были выявлены картели. На втором шаге с помощью машинного обучения тренируется классификатор, который для каждого аукциона определяет вероятность участия в нем картеля.

Сбор и обработка данных состояли из нескольких этапов:

1. получение исходной выборки;
2. предварительная обработка;
3. построение матрицы документов и фильтрация по шаблону;
4. извлечение данных из объединенного XML документа.

#### 1 этап. Получение исходной выборки

Информация о закупках, в которых участвовал картель, были получены путем скрэпинга сайта ФАС. Анализировались тексты решений по делам об антиконкурентных соглашениях за период с 2014 по 2016 и с помощью машинной лингвистики извлекались номера закупок. Исходная выборка включила 214 дел, по которым признавалось наличие антиконкурентных соглашений на 5055 закупочных процедурах (Таблица 1).

В большинстве дел признавались нарушения со стороны участников картеля по 1-10 закупкам. Однако было обнаружено несколько крупных дел, которые велись по нескольким сотням закупочных процедур (Таблица 2).

**Таблица 1. Распределение дел ФАС по количеству закупок, представленных в деле**

Количество «нечестных» закупок по одному делу	количество дел
1-10 закупок	136
11-40 закупок	52
41-100 закупок	12
Больше 100 закупок	14
всего дел	214

Источник: расчеты авторов.

**Таблица 2. Дела ФАС с максимальным количеством закупок, в которых обнаружено действие картеля**

Номер решения по делу	Количество аукционов
№ 30/51430/16	741
№ 22/11377/17	700
№ 22/90388/16	465
№ 22/93010/17	419
№ 22/86212/17	360
№ ММ 06-06/2015-371	291
№ 22/79888/16	219

Источник: расчеты авторов по данным ФАС РФ.

Для контрольной группы закупок было случайным образом отобрано 20000 аукционов из всего массива закупок в 2014-2016 гг.

#### 2 этап. Предварительная обработка

Далее была проведена предварительная обработка данных, в ходе которой были отброшены

- закупки с номерами, не соответствующими формату электронного аукциона,
- и дублирующиеся номера закупок.

В результате осталось 18015 (90,1%) аукционов случайной выборки и 4982 (98,6%) выборки по делам ФАС.

#### 3 этап. построение матрицы документов и фильтрация по шаблону

Каждый этап электронного аукциона должен сопровождаться публикацией на сайте электронной площадки соответствующих документов: извещение, три протокола и итоги аукциона с указанием выбранного поставщика. Но в реальности некоторые документы часто отсутствуют или являются пустыми. В некоторых случаях, например, есть только извещение. Это может говорить о том, что аукцион был отменен из-за того, что не набралось минимальное необходимое число участников, либо заказчик передумал совершать закупку, либо были выявлены какие-то нарушения со стороны заказчика и так далее. В исходную выборку включились аукционы, для которых был опубликован хотя бы один из необходимых документов.

Нам интересны аукционы, которые успешно завершились, и в результате которых был определен поставщик, поэтому в итоговый набор данных были включены закупки содержащие:

1. Извещение о проведении электронного аукциона (fcsNotificationEF);
2. Протокол рассмотрения заявок на участие в электронном аукционе (fcsProtocolEF1);
3. Протокол проведения электронного аукцион (fcsProtocolEF2);
4. Протокол подведения итогов электронного аукциона (fcsProtocolEF3);
5. Результат проведения процедуры определения поставщика (fcsPlacementResult);

В результате фильтрации данных осталось 10381 (51,9%) закупок из случайной выборки и 3452 (68,3%) – из выборки ФАС.

Если аукцион успешно завершился, то по нему должны быть опубликованы минимум пять основных документов, указанных выше. Но по разным причинам число их может оказаться больше. Чаще всего публикуются дополнительные документы с разъяснениями положений о закупке. Каждый участник аукциона вправе подать не более трех запросов на разъяснения, и заказчик должен будет ответить на них в течение определенного периода времени (Ст.65 закона № 44-ФЗ). С одной стороны, большое количество разъяснений может говорить о том, что участвовали игроки, которые не были в сговоре с заказчиком, не знали его требований и пытались их уточнить, чтобы оценить свои шансы на победу. С другой стороны, непонятно сформулированное извещение, с большим количеством неоднозначных моментов может свидетельствовать о том, что заказчик пытался воспрепятствовать участию фирм в аукционе с целью отдать заказ поставщику, с которым находился в сговоре. Помимо разъяснений часто заново публикуется извещение о закупке. Это происходит, как правило, когда заказчик обнаруживает ошибку или важное упущение в извещении. Намного реже переписываются протоколы рассмотрения заявок, проведения аукциона и итоговые.

#### 4 этап. извлечение данных из объединенного XML документа

На последнем, четвертом, этапе из документов, опубликованных на сайте ЕИС извлекались отдельные показатели: количество участников, подавшие заявки и допущенные, НМЦК, время подачи заявок, количество и величина ставок, результаты аукциона. В ходе извлечения данных оказалось, что в некоторых закупках Протоколы являются пустыми, такие наблюдения были исключены из выборки. Также были исключены закупки, протоколы которых были составлены с явными ошибками. Например, были убраны наблюдения, для которых относительное снижение цены меньше 0, близко к 1 или больше 1; разница между минимальной и второй ценой больше 1; разница между временем подачи заявки и публикацией извещения меньше 0. В итоге осталось 5101 наблюдение из Случайной выборки и 2789 наблюдений из выборки ФАС.

В таблице (Таблица 3) приводятся описательная статистика по закупкам, где обнаружен картель (выборка ФАС), и где не обнаружен (Случайная выборка). Мы видим, что средние значения и стандартные отклонения для некоторых переменным значительно отличаются по выборкам.

**Таблица 3. Описательная статистика процедур проведения электронных аукционов**

ПЕРЕМЕННЫЕ	N	mean	sd	min	max
<b>NO CARTELS (Горизонтальный сговор ФАС не зафиксирован)</b>					
Наличие картеля	5101	0	0	0	0
Количество заявок по 1 Протоколу	5101	3,022	1,487	2	17
то же, по 2 Протоколу	5101	2,599	1,035	2	14
Разница в количестве заявок между 1 и 2 Протоколами	5101	0,423	1,096	0	15
то же, между 2 и 3 Протоколами	5101	0,0392	0,384	0	11
День недели публикации Извещения	5101	1,994	1,504	0	6
Медиана разницы во времени подачи заявки и началом приема заявок	5101	860495	749239	7623	1,983e+07
Медиана разницы во времени подачи заявки между i и (i-1) участниками	5101	87157	153056	1	2,377e+06
НМЦК в тыс. руб.	5101	5643	40911	0,0745	1,616e+06
Длительность проведения электронного аукциона	5101	996,1	1822	0	24385
Максимальное количество ставок	5101	2,934	6,690	1	80
Среднее снижение цены за один шаг (в % от НМЦК)	5101	0,0239	0,0396	0	0,382
Разница между минимальной ценой и НМЦК (в % от НМЦК)	5101	0,151	0,175	0,00500	0,799
Разница между минимальной и второй ценами (в % от НМЦК)	5101	0,00720	0,00886	0	0,0500
Количество документов с разъяснениями	5101	0.3562	0.920547	0	13
	N	mean	sd	min	max
<b>CARTELS (Имеется решение ФАС о горизонтальном сговоре)</b>					
Наличие картеля	2789	1	0	1	1
Количество заявок по 1 Протоколу	2789	2,333	0,863	2	11
то же, по 2 Протоколу	2789	2,087	0,382	2	7
Разница в количестве заявок между 1 и 2 Протоколами	2789	0,246	0,747	0	7
то же, между 2 и 3 Протоколами	2789	0,00251	0,0628	0	2
День недели публикации Извещения	2789	2,046	1,494	0	6
Медиана разницы во времени подачи заявки и началом приема заявок	2789	841671	474250	9407	5,270e+06
Медиана разницы во времени подачи заявки между i и (i-1) участниками	2789	35264	94283	0	1,055e+06
НМЦК в тыс. руб.	2789	6482	27813	1,055	734363
Длительность проведения электронного аукциона	2789	278,1	722,3	0	8949
Максимальное количество ставок	2789	1,180	2,130	1	49

ПЕРЕМЕННЫЕ	N	mean	sd	min	max
Среднее снижение цены за один шаг (в % от НМЦК)	2789	0,00916	0,0230	0	0,267
Разница между минимальной ценой и НМЦК (в % от НМЦК)	2789	0,0357	0,0974	0,00500	0,738
Разница между минимальной и второй ценами (в % от НМЦК)	2789	0,00509	0,00411	0	0,0717
Количество документов с разьяснениями	2789	0.145572	0.542623	0	10
	N	mean	sd	min	max

Источник: рассчитано авторами

#### 4. Стратегия оценивания и гипотезы

Для выявления картельных сговоров мы применяем статистические показатели процесса закупки в сочетании с машинным обучением. Для проведения различия между сговором и конкурентным поведением участников аукциона рассматриваются несколько переменных, построенных на основе распределения заявок и ставок в каждой закупке. Для каждого из показателей выдвигается гипотеза по взаимосвязи его величины и факта наличия картеля на аукционе. В частности, делаются предположения относительно ставок участников, времени подачи заявок, числа участников и их активности на торгах, длительности этапов аукциона, а также максимальной начальной цены аукциона. Поскольку каждый показатель отражает различные аспекты распределения заявок, их совместное использование потенциально позволяет учитывать различные типы стратегий неконкурентного поведения фирм. Ниже представлено подробное описание показателей, используемых нами в эмпирическом анализе.

##### 1. Разница во времени подачи заявок.

Возможно, если фирмы находятся в сговоре, то разница во времени подачи ими заявок будет минимальной. Если фирмы являются членами картеля, то их заявки на участие в аукционе часто подаются с одного IP-адреса, а документы к заявкам создаются с использованием одной и той же учетной записи. Это может означать, что одно и то же лицо (или группа лиц) занимается подготовкой и отправкой заявок нескольких компаний, и поэтому документы фирм-участников картеля с большей вероятностью будут отправлены в одно и то же время. Даже если фирмы готовят и отправляют документы самостоятельно, согласованность их действий, скорее всего, повлияет на то, что разница во времени подачи заявок будет небольшой. Для того, чтобы учесть коррелированность во времени между временем подачи заявки и наличием картеля, были посчитаны несколько показателей.

1.1 Первая переменная –  $\frac{app_j}{mean_i}$  – средняя разница во времени подачи заявки между  $i$ -м и  $i-1$  участниками аукциона ( $i-1$  участник – фирма, которая подала заявку последней перед  $i$ -м участником).

*Гипотеза 1.* Вероятность картеля выше в аукционе, в котором меньше средняя разница во времени подачи заявки между  $i$ -м и  $i-1$  участниками.

1.2 Вторая переменная –  $\overset{app}{j}$   
 $median_i$  – медиана разницы во времени подачи заявки между  $i$ -м участником и  $i-1$  участниками аукциона ( $i-1$  участник – фирма, которая подала заявку последней перед  $i$ -м участником).

*Гипотеза 2.* Вероятность картеля выше в аукционе, в котором меньше медиана разницы во времени подачи заявки между  $i$ -м и  $i-1$  участниками. В выборке с картелями распределение разницы во времени между заявками участников больше смещено к 0, чем в случайной выборке (Рисунок 5). Для картелей доля закупок, в которых

$\overset{app}{j}$   
 $median_i$  лежит в пределах одного часа, в два раза больше, чем в случайной выборке.

Это подтверждает поставленные гипотезы.



Источник: рассчитано авторами.

**Рисунок 5. Распределение медианы и средней разницы во времени подачи заявок**

Разница между временем подачи заявки  $i$ -м участником и началом сбора заявок, скорее всего, не будет связана с вероятностью победы  $i$ -го участника и наличием картеля. В закрытом аукционе такая связь есть, потому что вместе с подачей заявки фирма одновременно делает ценовое предложение, и если есть сговор с заказчиком, то фирма  $i$  может узнать ценовые предложения всех фирм от 1 до  $i-1$ . В электронном аукционе, ставки делаются в отдельный день, поэтому не важно, в какой момент ты подал заявку. На следующем рисунке (Рисунок 6) представлено распределение во времени подачи заявок относительно момента начала их сбора. Значительных отличий между двумя выборками нет, как и предполагалось. А наличие двух пиков в районе 7 дня (144-168 часов) и 15го дня (336-360 часов) свидетельствует о том, что многие участники подают заявки ближе к концу срока их сбора.



Источник: рассчитано авторами.

**Рисунок 6. Распределение разницы во времени между подачей заявки и началом сбора заявок**

1.3 Если в торгах участвует картель, то с большой вероятностью выиграет один из его участников. Также мы предположили, что если фирмы находятся в сговоре, то с большой вероятностью их заявки будут отправлены на электронную торговую площадку почти одновременно. Поэтому на вероятность наличия

картеля в аукционе  $j$  может влиять фиктивная переменная  $\underset{i}{\overset{mint}{win}_i^j}$ , такая

что

— Если  $\underset{mint}{\min}(t_{appi} - t_{appi-1}) = t_{appk} - t_{appk-1}$ , то

—  $\underset{i}{win}_i^j \begin{cases} 1, \text{ если участник } k \text{ или } k-1 \text{ стал победителем} \\ 0, \text{ если участник } k \text{ или } k-1 \text{ не стал победителем} \end{cases}$

*Гипотеза 3.* Если в закупке стал победителем один из участников, для которого разница во времени с предыдущей или последующей заявкой была минимальной, то вероятность картеля в таком аукционе больше.

По нашей выборке получилось, что доля наблюдений, для которых  $\underset{i}{\overset{mint}{win}_i^j} = 1$ , в выборке ФАС – 95%, а в случайной выборке – 84%. Но такие показатели, скорее всего, будут завышены, потому что во многих закупках участвовали всего две компании, а,

значит, вероятность  $\underset{i}{\overset{mint}{win}_i^j} = 1$  равняется 100% в таких наблюдениях. Показатель был пересчитан только для аукционов с количеством участников больше 2. Получилось, что

вероятность  $\underset{i}{\overset{mint}{win}_i^j} = 1$  в картелях – 85%, в случайной выборке 68%.

## 2. Длительность сбора заявок.

Также предполагалось, что длительность сбора заявок может отличаться. Но графики распределения длительности приема заявок не показали значительных отличий между выборками по торгам, где выявлен картель, и где не выявлен.

Следующие несколько показателей характеризуют сам аукцион.

Если аукцион проходит в условиях честной конкуренции, то участники будут торговаться активнее, это отразится на количестве ставок (оно будет больше, чем в менее конкурентной обстановке), времени проведения аукциона (увеличится) и величине снижения цены (также возрастет).

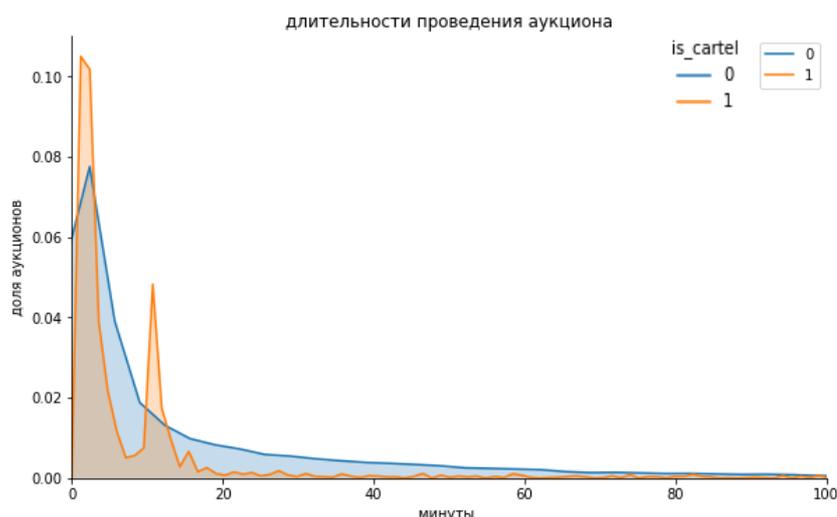
### 3. Длительность аукциона

Как было отмечено в предыдущем разделе, после начала аукциона участникам дается 10 минут на то, чтобы сделать ставки, и после каждого ценового предложения время автоматически продлевается. Поэтому, если фирмы активно конкурируют за контракт, то продолжительность аукциона увеличивается. В качестве показателя длительности проведения аукциона берем разницу между первым ценовым предложением и последним, т.е. минимальной ставкой.

$$Dt_{\text{auction}} = t_{\text{winner}}(\text{last}_{\text{date}}) - t_{\text{firstbid}}(\text{first}_{\text{date}})$$

*Гипотеза 4.* Чем дольше длился аукцион, тем меньше вероятность наличия в нем картеля.

График распределения длительности аукциона (Рисунок 7) подтверждает наше предположение. Всплеск доли аукционов в выборке по картелям на отметке в 10 минут говорит о том, что часто последняя ставка делалась в последний момент перед окончанием времени, отведенного на новое ценовое предложение.



Источник: рассчитано авторами

**Рисунок 7. График распределения длительности проведения электронного аукциона**

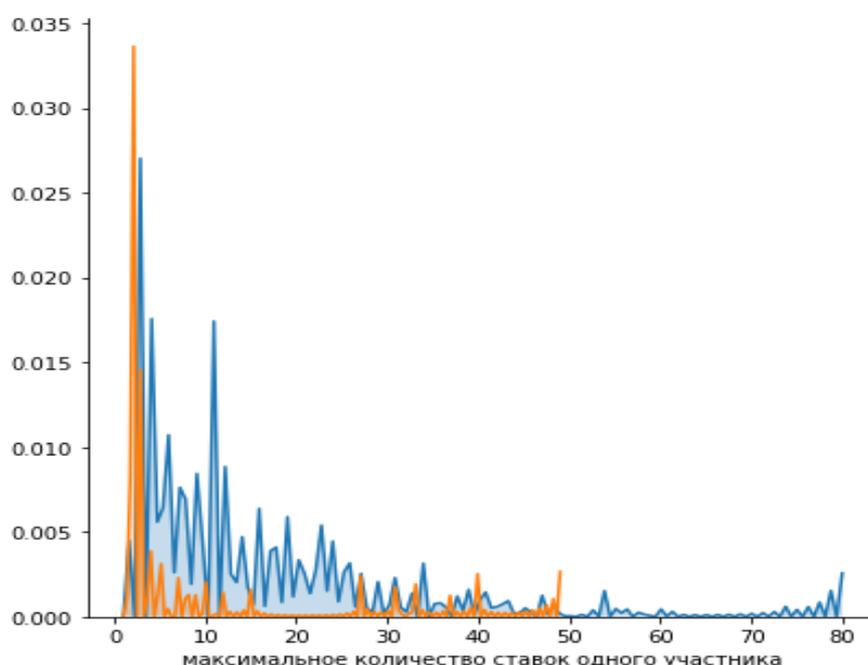
### 4. Активность торгов

4.1. Первый показатель активности торгов – максимальное количество ставок, которое сделал участник в аукционе  $j$ .

$$\max_{countbids} = \max_i (offers_{quantityi})$$

*Гипотеза 5.* Чем больше максимальное количество ставок, которое сделал участник в аукционе, тем меньше вероятность наличия картеля.

Из графика распределения максимального количества ставок видно (Рисунок 8), что в наблюдениях, где был выявлен картель, делается меньше ставок, в большинстве аукционов с картелем было сделано максимум 2-5 ценовых предложений. Такие результаты логичны. Если на торгах есть сговор, то победитель, как правило, заранее определен, и нужно делать ставки только, чтобы аукцион не был отменен, и чтобы создать иллюзию некоторой конкуренции.



Источник: рассчитано авторами

**Рисунок 8. Плотность распределения максимального количества ставок, сделанных участником**

## 5. Разница в ценах.

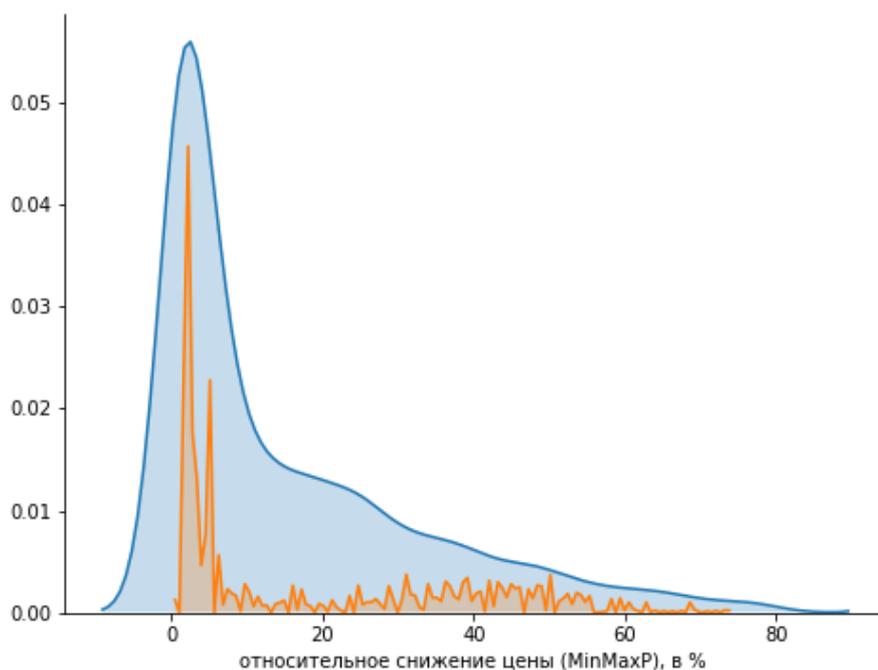
5.1. При наличии картельного сговора цена в ходе торгов снижается намного меньше при прочих равных условиях. Для того, чтобы обеспечить сопоставимость закупок при анализе используем относительное снижение цены, взвешиваем разницу по максимальной начальной цене.

$$MaxMinP_j = \frac{\max_{pricej} - \min_{priceij}}{\max_{pricej}}$$

где  $max_{price j}$  – резервная цена на аукционе  $j$ ,  $min_{price j}$  – минимальная цена на аукционе  $j$ .

*Гипотеза 6.* Чем меньше разница между минимальной ценой и резервной (НМЦК), тем больше вероятность наличия картеля.

В выборке аукционов с картелями распределение разницы между минимальной и резервной ценами является очень сжатым, и в большинстве наблюдений разница составляет менее 5% от НМЦК (Рисунок 9).



Источник: рассчитано авторами

**Рисунок 9. Плотность распределения разницы между максимальной и минимальной ценой в процентах от максимальной цены**

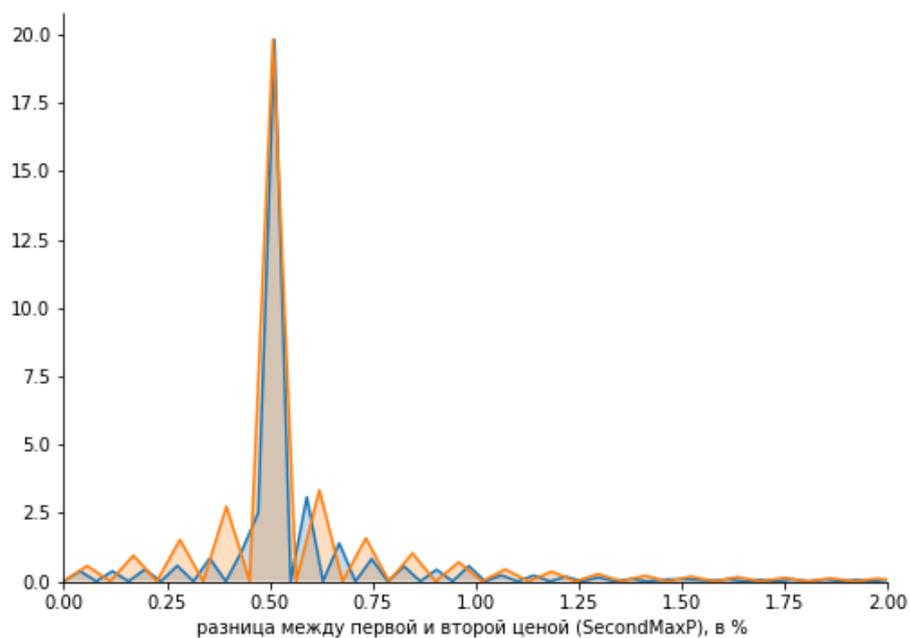
5.2. Если вторая самая низкая ставка близка к минимальной цене, то закупочное агентство может заключить контракт с фирмой, сделавшей второе-наилучшее ценовое предложение, если другие критерии, помимо цены, представляются более благоприятными (компенсируют разницу в цене). Даже в случае стандартных работ или товаров, где главным критерием является цены, фирмы могут предпочесть сохранить определенную минимальную разницу между первой и второй самыми низкими ставками, чтобы гарантировать определенный исход, желательный для картеля. Поэтому разница  $(secondbest_{price j} - win_{price j})$ , предположительно, будет больше в случае сговора. Для того, чтобы обеспечить

сопоставимость аукционов будем использовать разницу, взвешенную по максимальной цене аукциона

$$SecondMinP_j = \frac{secondbest_{price_j} - win_{price_j}}{max_{price_j}} .$$

*Гипотеза 7.* Чем больше разница между минимальной и второй по величине ценой, тем больше вероятность наличия картеля.

График плотности распределения разницы  $SecondMinP_j$  не выявляет значительных отличий между картелями и закупками из случайной выборки (Рисунок 10).



Источник: рассчитано авторами

**Рисунок 10. Плотность распределения разницы между второй и минимальной ценами**

На цены влияет множество различных факторов, и на них нужно контролировать. Важен объект закупки и его отрасль<sup>3</sup>. На возможности снижения цены могут оказывать воздействие макроэкономические условия, государственная политика в отношении отраслей и т.д. Кроме того, разные отрасли характеризуются разным уровнем конкуренции в целом.

Число участников и снижение цены также могут зависеть от количества позиций (лотов) в закупке. Если контракт включает много товарных позиций, то, вероятно,

---

<sup>3</sup> Так, неурожаем каких-то видов с/х культур может привести к более высоким ценам на эту продукцию.

меньшее число компаний будут готовы выполнить его из-за ограничений специализации или мощностей фирмы.

6. Стоит ввести контроль на НМЦК аукциона. Предполагаем, что фирмы сильнее конкурируют за контракты с более высокой стоимостью, т.к. потенциально получают больший доход от таких сделок.
7. Количество участников.

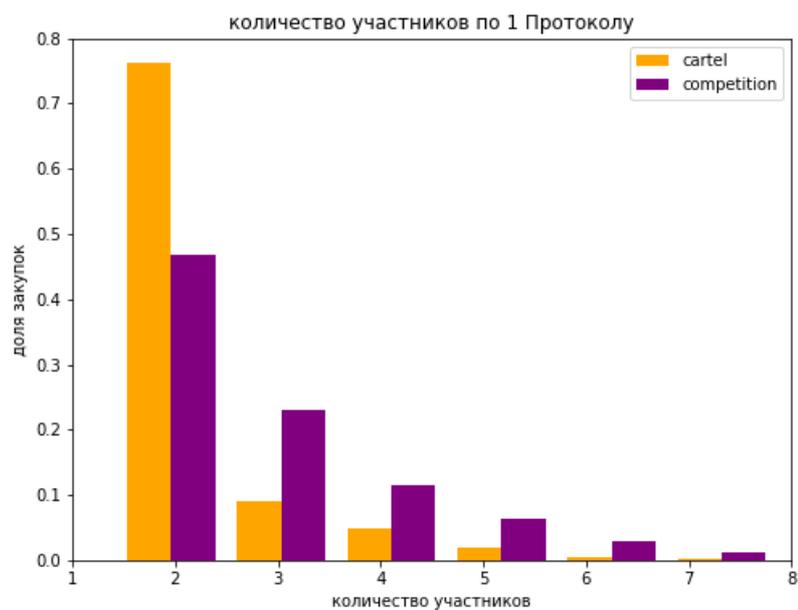
Чем больше число участников в аукционе, тем тяжелее осуществить стратегию картеля по поддержанию цены. С ростом числа участников, скорее всего, будет расти активность конкуренции, и цена будет больше снижаться в ходе торгов.

*Гипотеза 8.* С ростом числа участников аукциона снижается вероятность наличия картеля.

Для учета числа участников были использованы три показателя

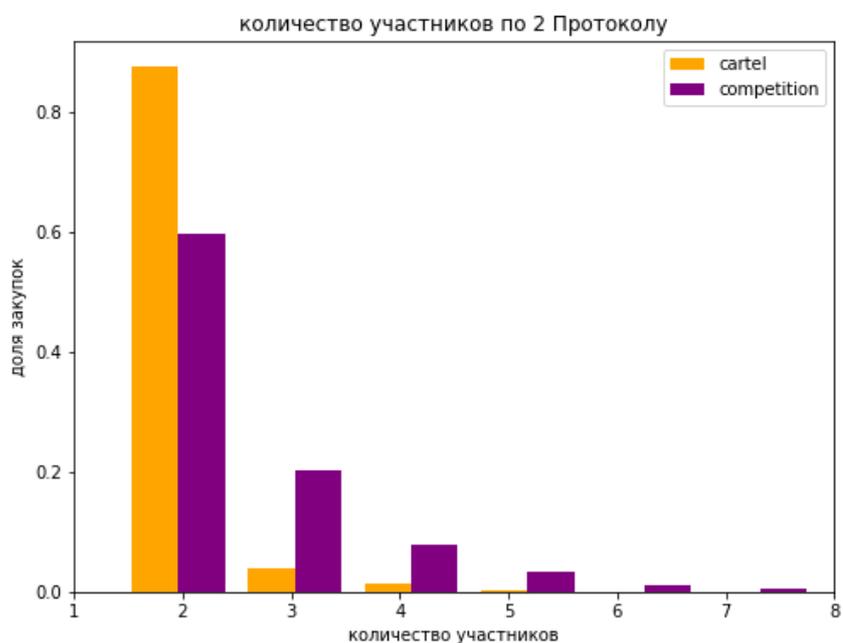
- 6.1.  $particip_{i,1} Protocol_j$  – количество тех, кто подал заявку на аукцион  $j$ .
- 6.2.  $particip_{i,2} Protocol_j$  – количество тех, кого допустили до электронного аукциона  $j$ .
- 6.3.  $activ_{i,2} particip_2 Protocol_j$  – количество тех, кто делал ставку в аукционе  $j$  (т.е. из 2го протокола исключаются участники, не выставившие ценовых предложений).

В 77% наблюдений для выборки по картелям заявки на участие в аукционе подавали всего две фирмы. В случайной выборке аукционов доля закупок, состоящих из двух участников – намного меньше, около 46% (Рисунок 11). Картелю выгодно, когда на торгах нет реальных конкурентов. Если характеристики отрасли и состояния рынка предполагают наличие таковых, то картель может сдержать их появление на аукционе путем сговора с заказчиком. Если же сторонние компании, всё-таки, решили участвовать, заказчик может не допустить их к торгам по первой части заявок. Член картеля может способствовать отклонению заявок конкурентов, пример такого поведения содержится в деле ФАС № 1-11-123/00-22-17: поставщик, находившийся в сговоре с заказчиком, отправлял заказчику письма, где указывал основания, по которым нужно отклонить некоторые заявки. Ниже (Рисунок 12) видим, что, по нашим данным, в случае наличия картеля доля аукционов с двумя участниками – около 90%, в случае отсутствия выявленного картеля – 60%.



Источник: рассчитано авторами

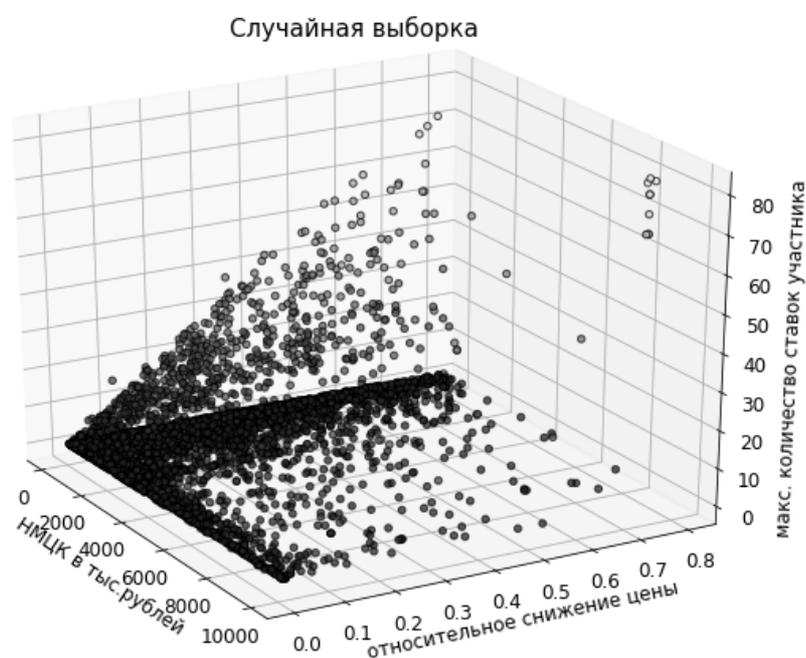
**Рисунок 11. Плотность распределения числа участников аукциона по 1 протоколу (протоколу рассмотрения заявок)**



Источник: рассчитано авторами

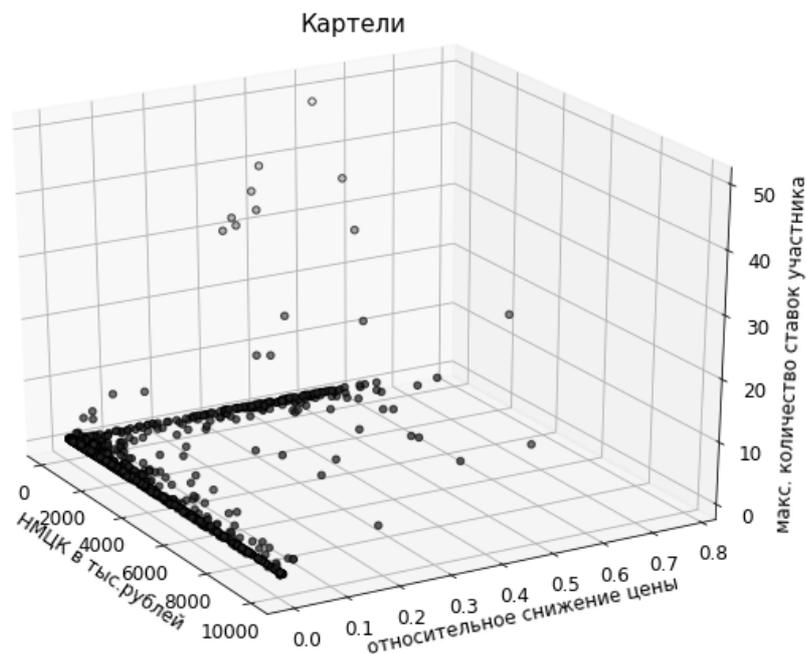
**Рисунок 12. Плотность распределения числа участников аукциона по 2 протоколу (протоколу рассмотрения заявок)**

Значительные различия можно увидеть также в совместном распределении показателей в двух выборках. Например, в случайной выборке наблюдается положительная связь между количеством ставок, сделанных одним участником, и относительным снижением цены. Это означает, что участники аукциона активно торгуются, вследствие чего цена сильно уменьшается. При этом цена значительно падает как для крупных, так и для небольших контрактов (Рисунок 13, Рисунок 15). В выборке по картелям почти во всех аукционах участники делают только одно ценовое предложение (Рисунок 14), при этом цена значительно снижается, только когда НМЦК контракта небольшая, меньше 200 тыс. рублей (Рисунок 15).



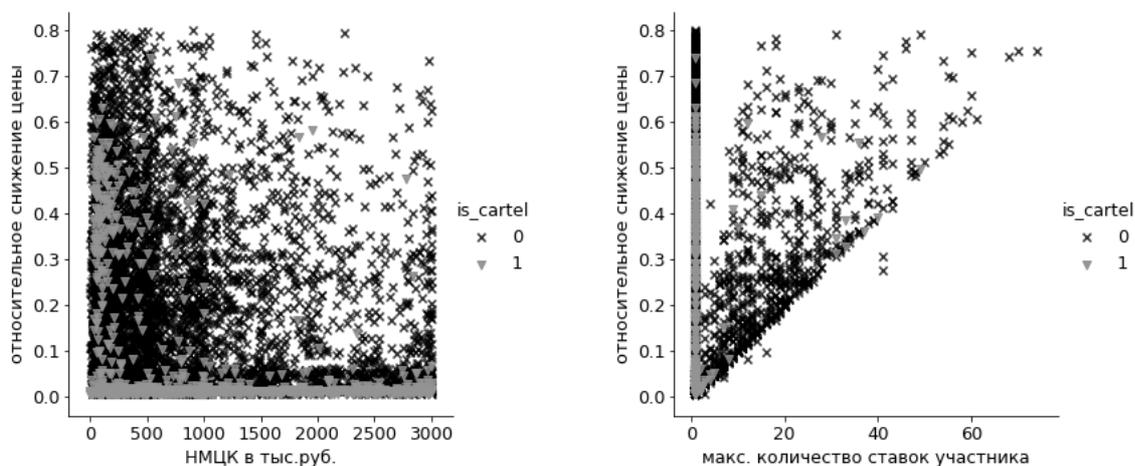
Источник: построено авторами

**Рисунок 13. Совместное распределение НМЦК, снижения цены в ходе аукциона и максимального количества ставок. Случайная выборка**



Источник: построено авторами

**Рисунок 14. Совместное распределение НМЦК, снижения цены в ходе аукциона и максимального количества ставок. Выборка по картелям**

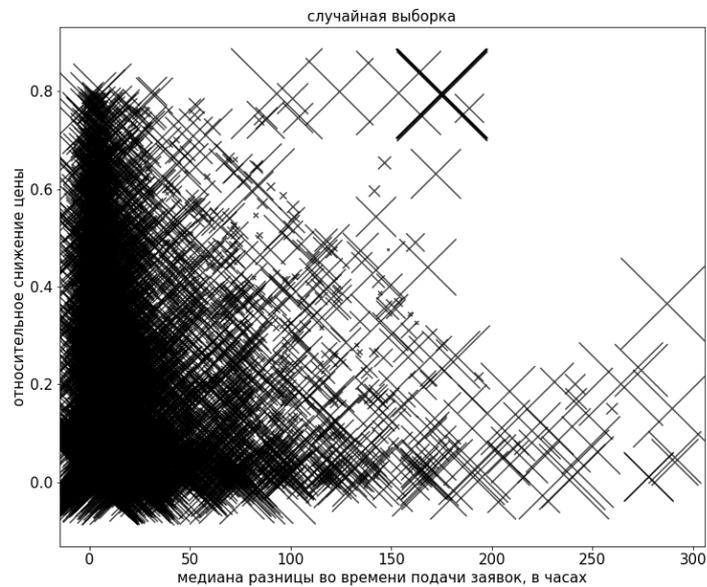


**Рисунок 15 Совместные распределения НМЦК и снижения цены в ходе аукциона; максимального количества ставок и снижения цены в ходе аукциона.**

Источник: построено авторами

Из графиков совместного распределения разницы во времени подачи заявок и относительного снижения цены (Рисунок 16, Рисунок 17) можно видеть, что в выборке по картелям почти во всех крупных аукционах и медиана разницы во времени, и относительное снижение цены были минимальными. В случаях, когда эти показатели были одновременно высокими, НМЦК была очень маленькой. Для случайных аукционов

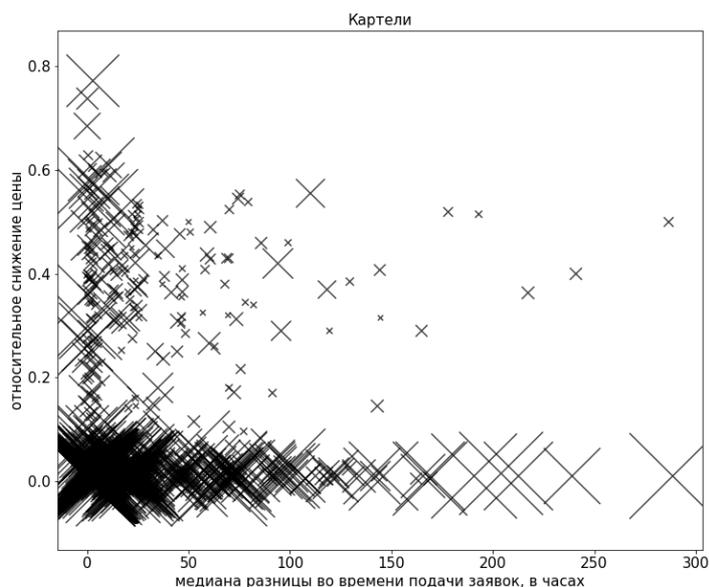
наблюдается иная ситуация. Хотя для этой выборки распределение тоже смещено к началу координат, но существует достаточно много закупок с большим НМЦК, для которых медиана разницы во времени была больше 50 часов, а снижение цены - больше 10%.



**Рисунок 16. Совместное распределение медианы разницы во времени подачи заявок и относительного снижения цены с учетом размера контракта в случайной выборке**

Примечание: Размер маркера, обозначающего наблюдение, зависит от размера аукциона. Чем больше НМЦК, тем больше маркер «x».

Источник: построено авторами



**Рисунок 17. Совместное распределение медианы разницы во времени подачи заявок и относительного снижения цены с учетом размера контракта в выборке по картелям**

Примечание: Размер маркера, обозначающего наблюдение, зависит от размера аукциона. Чем больше НМЦК, тем больше маркер «x».

Источник: построено авторами

## 5. Жалобы в ФАС

С одной стороны, дела по горизонтальным соглашениям фирм-поставщиков ведутся на данных о прошедших закупках, в рамках одного дела может одновременно рассматриваться до нескольких сотен закупок. Из-за того, что расследование по картелю ведется после закупки, оно не может повлиять на длительность проведения аукциона или иные показатели закупки. С другой стороны, если какие-то фирмы выступают в качестве ответчиков по антимонопольному делу, то они могут поменять свое поведение и начать вести себя более аккуратно в следующих закупках. Это никак не может повлиять на результаты моделей, разбираемых в этой главе, так как если члены картеля перестают применять антиконкурентные схемы на время рассмотрения дела, то можно считать, что картель на время прекращает свое существование и соответствующие закупки можно считать честными. Если антимонопольная служба признает фирмы виновными по статье Закона о защите конкуренции, то потом открывается соответствующее дело по административному правонарушению по статье 14.32 КоАП РФ или по статье 178 УК РФ.

И уже на основе этих дел назначаются штрафы или иные наказания для участников картеля.

## 6. Методы машинного обучения

Мы применяем методы машинного обучения для тренировки и тестирования модели, которая предсказывает вероятность существования картеля на электронном аукционе. Используются несколько различных моделей, в том числе два простых классификатора: Логистическая регрессия и метод ближайших соседей – и два ансамблевых: случайный лес и градиентный бустинг.

Каждый из классификаторов распределяет аукционы на «честные» и «нечестные», основываясь на данных о восьми переменных:

- $pr_{count_i}$  – количество участников, допущенных к аукциону  $i$ .
- $dif_{countapp}^{12_i}$  – разница между количеством фирм, подавших заявки, и количеством фирм, допущенных к аукциону  $i$ .
- $median_{dif}^{t_i}$  – медиана разницы во времени подачи заявок попарно взятых участников, идущих друг за другом по времени в  $i$ -м аукционе.
- $maxP_{thousi}$  – максимальная начальная цена контракта (НМЦК) на  $i$ -м аукционе, в тыс. рублей.
- $maxPrice_i$  – разница между НМЦК и минимальной ставкой на  $i$ -м аукционе.
- $Min_{wD_i}^{Price_i}$  – разница между второй-наилучшей ставкой и минимальной ставкой на  $i$ -м аукционе.
- $max_{countbidsi}$  – максимальное число ценовых предложений, сделанных участником на  $i$ -м аукционе.
- $clarification_{count_i}^i$  – количество опубликованных поясняющих документов для аукциона  $i$ .

Предполагаем, что большое количество разъяснений может говорить о том, что участвовали игроки, которые не были в сговоре и пытались уточнить требования, чтобы оценить свои шансы на победу.

Выбор модели, которая предсказывает вероятность существования картеля на электронном аукционе, осуществлялся в несколько этапов.

Перед построением классификаторов мы разбили данные на обучающую и тестовую выборки. Обучающая выборка включила 70% от общего числа наблюдений, и использовалась для оценки параметров модели. Тестовая выборка, которая состоит из 30% наблюдений, использовалась для прогнозирования и оценки эффективности моделей.

Далее данные были нормализованы, чтобы снизить влияние масштаба данных на оценки классификаторов.

### 6.1 Логистическая регрессия

Логистическая регрессия – это модель для бинарной классификации, но может быть расширена для многоклассового анализа. Логистическая функция позволяет спрогнозировать вероятность отнесения наблюдения к конкретному классу. Ниже представлена спецификация модели для нашей задачи классификации.

$$p_i = P(y_i = 1) = \frac{1}{1 + e^{-z}}$$

где

$$z = \omega_0 + \omega_1 \text{count}_{i} + \omega_2 \text{dif}_{\text{countapp}} + \omega_3 \text{median}_{\text{diff}} + \omega_4 \text{maxP}_{\text{thousi}} + \omega_5 \text{clarification}_{i} + \omega_6 \text{Price}_{i} + \omega_7 \text{max}_{\text{countbidsi}} + \omega_8 \text{pr}_{\text{maxPrice}}_{i}$$

—  $p_i = P(y_i = 1)$  – вероятность того, что на аукционе присутствует картель.  $y_i$  – бинарная переменная, равная 1, если в закупке  $i$  ФАС обнаружил картельный сговор, 0 – иначе.

Оптимальные параметры для логистической модели, в том числе уровень регуляризации  $C$ , были подобраны с использованием кросс-валидацию по 10 блокам на всей выборке данных. Подробное описание подбора оптимальных параметров представлено в Приложении 1.

Правильность логистической модели на обучающей выборке составила 0,7758, на тестовой – 0,7752. Показатели для тестовой и обучающей выборки почти совпадают, это говорит о том, что нет проблемы переобучения, но 77% – относительно невысокая правильность прогнозирования. В нашем случае правильность (ассурасу) будет не очень корректно показывать качество модели, потому что у нас очень неравномерно распределено число наблюдений с картелем и без него. В нашей выборке много наблюдений, где антимонопольная служба не выявила картель, и этот класс модель будет определять более точно. При этом может быть так, что наличие картеля будет определять очень плохо, с большим количеством ошибок. Для того, чтобы разделить эффективность

модели для определения класса cartel (есть картель) и no\_cartel (нет картеля) стоит использовать показатели точности<sup>4</sup> и полноты<sup>5</sup> модели.

Ниже (Таблица 4) видно, что намного более правильно наша логистическая модель определяет отсутствие картеля, чем его наличие. Точность обнаружения картеля находится на уровне 69%.

**Таблица 4. Полнота и точность логистической регрессии при выявлении картелей на электронных аукционах**

	precision	recall	f1-score	наблюдения
no_cartel	0.83	0.82	0.83	1544
cartel	0.67	0.69	0.68	823
micro avg	0.78	0.78	0.78	2367
macro avg	0.75	0.75	0.75	2367
weighted avg	0.78	0.78	0.78	2367

Источник: рассчитано авторами.

Для того, чтобы нагляднее показать Recall и Precision, нарисуем матрицу ошибок классификации (Таблица 5).

**Таблица 5. Общий вид матрицы ошибок классификации для задачи выявления картеля**

	$y=0$ (нет картеля)	$y=1$ (есть картель)
$\hat{y}=0$ (нет картеля)	True-positive (TP) Правильно показывает отсутствие картеля	False-negative (FN) Неверно показывает наличие картеля ( <b>ошибка 1 рода</b> )
$\hat{y}=1$ (есть картель)	False-positive (FP) Неверно показывает отсутствие картеля ( <b>ошибка 2 рода</b> )	True-negative (TN) Верно указывает на наличие картеля

$$Recall = \frac{TP}{TP + FN}, \quad Precision = \frac{TP}{TP + FP}$$

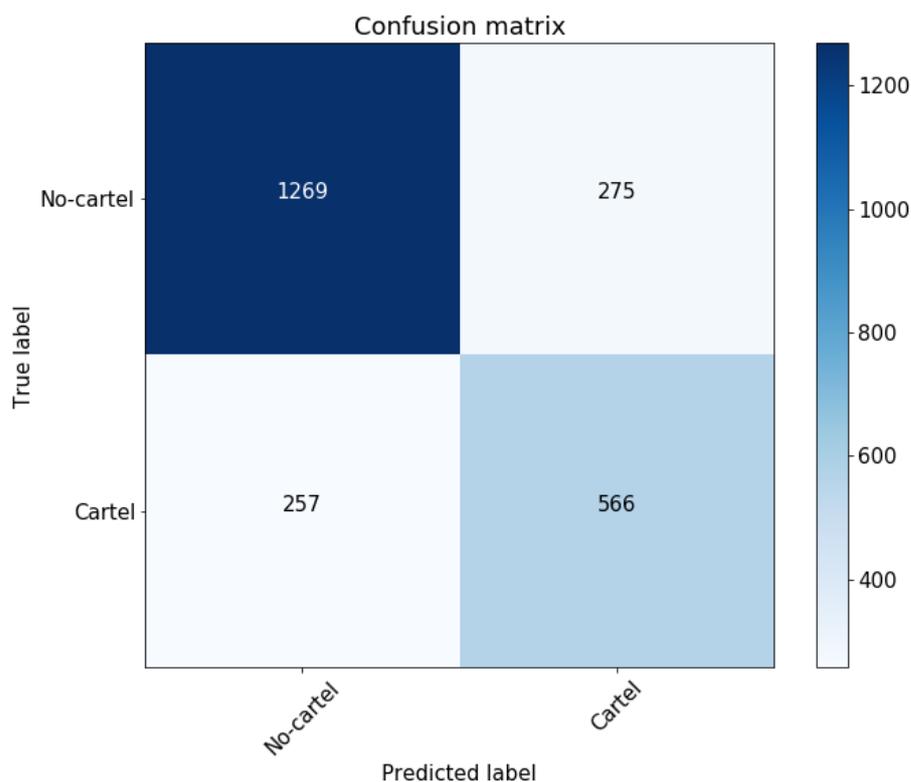
4 Точность (precision) показывает, какая доля из тех наблюдений, которые помечены как положительные (т.е. принимается гипотеза о принадлежности к определенному классу), действительно принадлежат этому классу.

5 Полнота (recall) показывает, какую долю наблюдений, относящихся к определенному классу, классификатор отнес к этому классу.

Строим матрицу ошибок по тестовой выборке (Рисунок 18) и получаем, что процент наблюдений, в которых неверно предсказано наличие картеля равняется

$$11,62 = \frac{275}{2367} * 100$$

, а процент наблюдений, в которых неверно предсказано отсутствие картеля равняется  $10,86 = \frac{257}{2367} * 100$ .



Источник: рассчитано авторами

**Рисунок 18. Матрица ошибок классификации для логистической регрессии**

## 6.2 Метод ближайших соседей

Ещё один часто используемый классификатор – это метод k ближайших соседей. Согласно этому методу наблюдение относится к тому классу, к которому относится большинство из k ближайших к нему по характеристикам наблюдений в обучающей выборке.

Оптимальное значение для k было определено путем проведения тестов на эффективность модели. Была посчитана правильность предсказаний моделей, построенных по нескольким блокам кросс-валидации для k от 1 до 20 (Приложение 1). Оптимальным оказалось k=15.

Для модели с 15 соседями в тестовой выборке 74% закупок, в которых был обнаружен картель, были правильно классифицированы. Точность распознавания картеля составила 69% (Таблица 6).

**Таблица 6. Точность и полнота модели k ближайших соседей (k=15)**

	precision	recall	f1-score	наблюдения
no_cartel	0.85	0.82	0.84	1544
cartel	0.69	0.74	0.71	823
avg / total	0.80	0.79	0.79	2367

Источник: рассчитано авторами

Правильность (accuracy) итоговой модели с учетом кросс-валидации равняется 78,55%. Для нашей начальной модели с k=5 правильность предсказаний с учетом кросс-валидации равнялась 77,95%.

### 6.3 Случайный лес

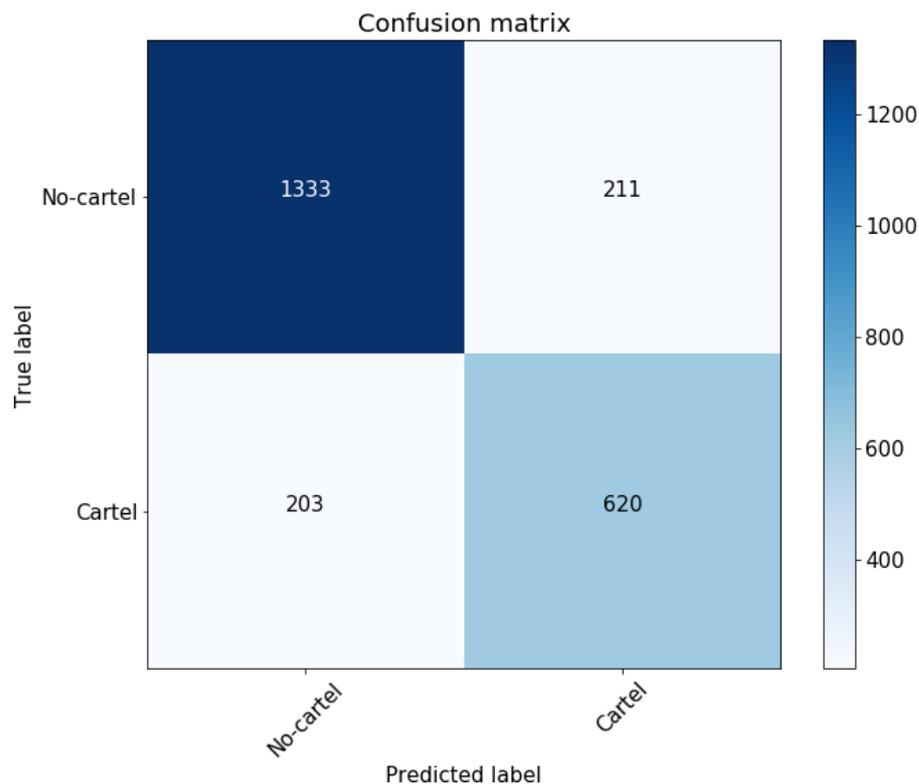
Следующий классификатор для выявления картелей на торгах был построен на основе случайного леса. Мы обучили случайный лес на основе 100 деревьев принятия решений и получили 99% правильность предсказания на обучающей выборке и 82,5% правильность на тестовой выборке. Показатели полноты и точности предсказания наличия картеля равняются 0,75 (Таблица 7). Всё это свидетельствует о том, что случайный лес является более эффективным, чем логистическая регрессия, и больше подходит для решения задачи выявления картелей.

**Таблица 7. Точность и полнота классификатора, построенного на основе случайного леса.**

	precision	recall	f1-score	support
no_cartel	0.87	0.86	0.87	1544
cartel	0.75	0.75	0.75	823
micro avg	0.83	0.83	0.83	2367
macro avg	0.81	0.81	0.81	2367
weighted avg	0.83	0.83	0.83	2367

Источник: рассчитано авторами

Для модели случайного леса матрица ошибок классификации представлена ниже (Рисунок 19).



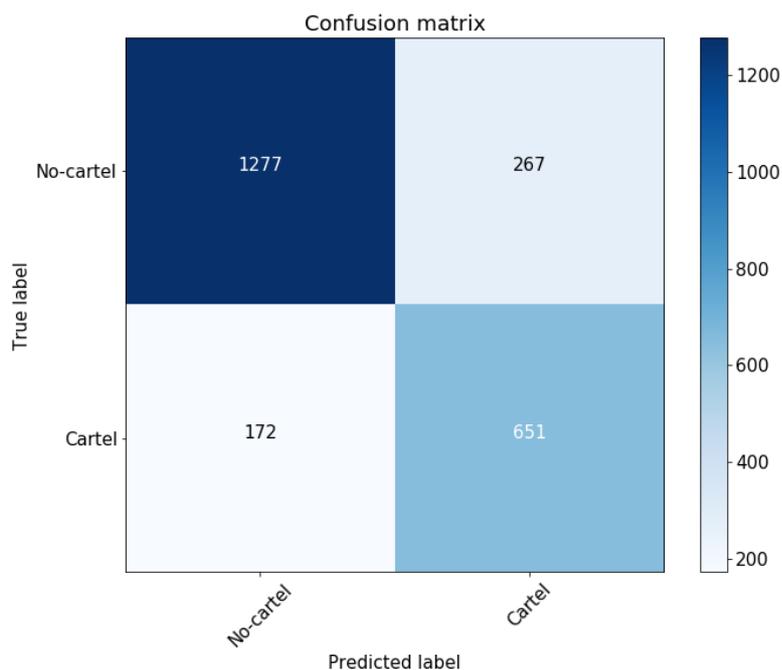
Источник: рассчитано авторами

**Рисунок 19. Матрица ошибок классификации  
для случайного леса**

#### 6.4 Градиентный бустинг

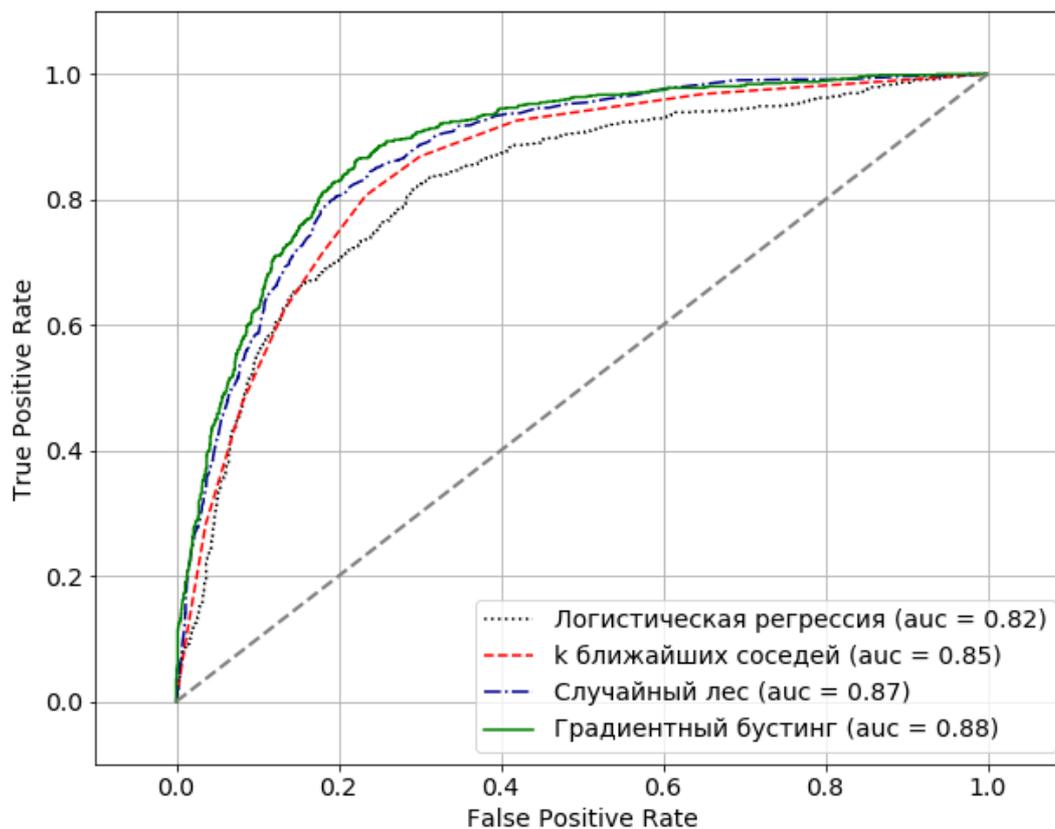
Мы использовали также ещё один ансамблевый метод, основанный на построении деревьев – Градиентный бустинг. Этот метод показал правильность предсказаний на тренировочной выборке – на уровне 90,6%, а на тестовой – 84,1 %.

Для модели Градиентного бустинга была построена матрица ошибок классификации

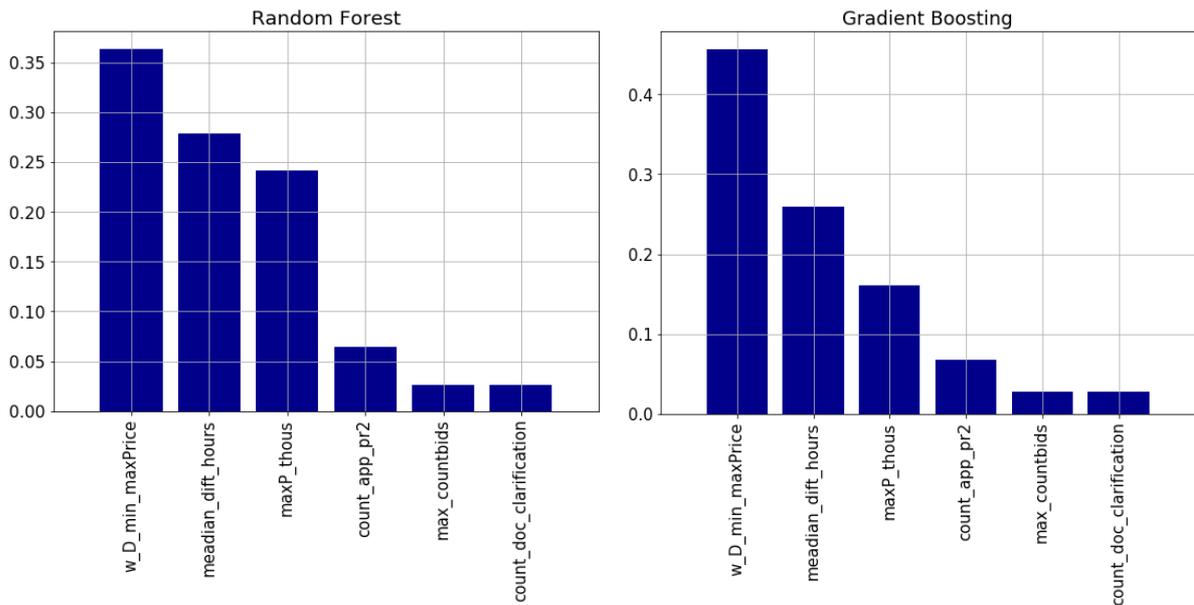


**Рисунок 20. Матрица ошибок классификации для градиентного бустинга**

Результаты моделей показывают, что ансамблевые методы (Случайный лес и Градиентный бустинг) наиболее точно классифицируют закупки на те, в которых ФАС выявила картель, и в которых не выявила. Эти результаты подтверждаются и в ходе применения ещё одного инструмента – ROC-кривой. ROC-кривые наших моделей располагаются далеко от 45-градусной диагонали и близко к левому верхнему углу, что говорит об их хорошей предсказательной силе.



Для Случайного леса и Градиентного бустинга можно вычислить коэффициенты важности переменных (importances) в определении классов. По обоим классификаторам были построены диаграммы относительной важности признаков (Рисунок 21). Коэффициенты переменных в классификаторах различаются, но перечень основных признаков остается одинаковым. На основе графика можно сделать вывод, что наибольший вклад в вероятность наличия картеля на аукционе вносит уровень снижения цены. Также стоит обращать особое внимание на разницу во времени подачи заявок, НМЦК и разницу между минимальной и второй-наилучшей ставками.



**Рисунок 21. Важность переменных в определении класса (существования картеля)**

Примечание:

- $pr$  — количество участников, допущенных к аукциону  $i$ .
- $dif_{countapp}^{12i}$  — разница между количеством фирм, подавших заявки, и количеством фирм, допущенных к аукциону  $i$ .
- $median_i$  — медиана разницы во времени подачи заявок попарно взятых участников, идущих друг за другом по времени в  $i$ -м аукционе.
- $maxP_{thousi}$  — максимальная начальная цена контракта (НМЦК) на  $i$ -м аукционе, в тыс. рублей.
- $wD_i$  — разница между НМЦК и минимальной ставкой на  $i$ -м аукционе.
- $wD_i Price_i$  — разница между второй-наилучшей ставкой и минимальной ставкой на  $i$ -м аукционе.
- $max_{countbidsi}$  — максимальное число ценовых предложений, сделанных участником на  $i$ -м аукционе.
- $count_i$  — количество опубликованных поясняющих документов для аукциона  $i$ .

Источник: рассчитано авторами

Через многие переменные, важные для задачи классификации, проходит красной нитью общая идея о том, что участники картеля не ощущают особой необходимости создавать видимость энтропии.

Результат, согласно которому НМЦК является третьим по важности индикатором картельного сговора на конкретной госзакупке, может трактоваться по-разному.

С одной стороны, это может объясняться наличием квази-фиксированных транзакционных издержек при организации картелей - чем меньше цена контракта, тем (при прочих равных условиях) большей доли прибыли участники должны будут лишиться.

С другой стороны, есть ещё одно потенциальное и не менее логичное объяснение. Более высокая НМЦК привлекает большее внимание ФАС, что может создать систематическую смещенность в изучаемых выборках, как вероятное следствие того, что рациональный антимонопольный орган ограничен в ресурсах. Сейчас мы не можем различить два подвида данной гипотезы - максимизирует ли антимонопольный орган свои политические дивиденды или же он всерьез озабочен максимизацией потенциала устрашения своего правоприменительного аппарата. Кроме того, эта зависимость может оказаться обратной U-образной, если ситуации с очень высокими ценами контрактов создают для антимонопольного органа неприемлемые риски противодействия со стороны заинтересованных чиновников из других органов власти.

## **7. Корректировка классификатора за счет учета несбалансированности данных.**

Наши данные довольно сильно не сбалансированы, случайная выборка почти в 2 раза больше, чем выборка по делам ФАС. Это может влиять на результаты классификатора. Если дисбаланс высокий, то в стандартных классификаторах класс большинства предсказывается хорошо, а класс меньшинства классифицируется неправильно [12]. В нашем случае класс большинства – это электронные аукционы, где не обнаружен сговор. Существуют разные методы, помогающие уменьшить падение качества предсказаний, которое появляется из-за несбалансированности данных, однако, полностью нивелировать этот эффект трудно. Тем не менее, можно выделить два основных способа бороться с проблемой неравномерности разных классовых групп в данных: первый – работа с выборкой, и второй – выбор моделей, которые менее чувствительны к объему классов или, например, больше штрафуют за ошибки классификации в группе, по которой имеется меньше наблюдений.

## 7.1 Способы корректировки данных

Если реальные данные почти сбалансированы, но в рассматриваемой выборке есть смещение пропорций (например, из-за метода сбора), то оценки будут смещены, поэтому нужно попытаться собрать более репрезентативные данные. Но обычно не сбалансирован сам исходный набор данных, так это – и в нашем случае с картелями. Есть несколько способов скорректировать выборку.

- Убрать часть наблюдений из большого класса;
- Добавить наблюдения из меньшего класса (дублировать некоторые наблюдения);
- Добавить синтетические наблюдения в минорный (меньший) класс. Здесь основными моделями являются: метод синтетического меньшинства (SMOTE) [11], borderline SMOTE [17] и адаптивная синтетическая выборка (ADASYN) Skalidis, 2016<sup>6</sup> и [19].

- 7.2 Взвешивание классификатора по классам

Второй подход к повышению эффективности решения задачи классификации состоит в том, чтобы корректировать не данные, а модель.

Некоторые модели могут учитывать несбалансированность данных и придавать больший вес классу меньшинства и меньший вес классу большинства. Например, взвесить предсказания на основе частот классов можно в модели Случайного леса, настроив соответствующий дополнительный параметр.

Также часто при распределении на классы важнее минимизировать либо ложноположительные, либо ложноотрицательные исходы, т.е. ошибки I и II рода имеют разный вес для исследователя. В ситуации с закупками можно предположить, например, что не обнаружение картеля в аукционе, где он присутствует, приведет к большим финансовым потерям для государства, чем ложное признание того, что в закупке участвует картель. Если картель не будет выявлен, то его участники смогут вместе присутствовать и на последующих аукционах, где будут применять схему сговора. Поэтому на будущих закупках, скорее всего, снижение цены будет очень маленьким, т.е. выделенные бюджетные средства будут расходоваться неэффективно на этапе проведения аукциона. С другой стороны, высокая вероятность ошибок первого рода, т.е. ситуаций, когда на аукционе был выявлен сговор, а в реальности его не было, может усилить стимулы поставщиков участвовать в сговоре. В зависимости от того, ошибки какого рода мы хотим минимизировать, можно корректировать классификатор.

---

<sup>6</sup> URL: <https://github.com/stavskal/ADASYN>.

### 7.3 Перебалансировка выборки для классификатора по картелям

Для того, чтобы сгладить несбалансированность выборки, мы используем SMOTE и его модификации. Выбор в пользу одного из методов будет осуществляться на основании показателей качества предсказаний моделей.

С помощью SMOTE алгоритма [11] были построены синтетические наблюдения для миноритарного класса (закупок с картелями). После применения SMOTE каждый класс стал занимать 50% от выборки, а общее число наблюдений увеличилось до 10202. Затем 70% наблюдений от получившегося набора данных были отнесены к обучающей выборке, которая использовалась для оценки параметров модели. Тестовая выборка из 30% наблюдений применялась для прогнозирования и оценки эффективности. Далее данные были стандартизированы, чтобы снизить влияние масштаба данных на оценки классификаторов.

Для новой выборки были снова построены логистическая регрессия, метод ближайших соседей, случайный лес и градиентный бустинг.

Можно видеть, что добавление новых наблюдений, относящихся к аукционам с картелями, увеличивает правильность предсказания классификаторов. После применения SMOTE правильность снизилась только в логистической модели. Для остальных моделей: Случайного леса, метода k ближайших соседей и Градиентного бустинга – это показатель вырос на несколько п.п. (Таблица 8). Причем, наиболее хорошие результаты продолжают показывать ансамблевые методы, для Леса и градиентного бустинга Accuracy равняется 0,84.

Аналогичные преобразования выборки были проделаны с помощью алгоритмов Border SMOTE и ADASYN. Но правильность предсказания моделей меньше улучшилось, чем при применении обычного SMOTE.

**Таблица 8. Правильность предсказания классификаторов, построенных на первоначальной выборке и выборке с синтетическими наблюдениями**

	Accuracy. Первоначальная выборка	Accuracy. Выборка с учетом SMOTE	Accuracy. Выборка с учетом Border SMOTE	Accuracy. Выборка с учетом ADASYN
Логистическая модель	0,761	0,716	0,681	0,679
k ближайших соседей	0,783	0,792	0,779	0,755

Случайный лес (100 деревьев)	0,772	0,836	0,803	0,803
Градиентный бустинг	0,792	0,840	0,818	0,811

Источник: рассчитано авторами

В нашей задаче классификации особенно важно, с какой точностью определяется каждый из классов. Применение SMOTE алгоритма изменило соотношение количества наблюдений классов, увеличив объем наблюдений с выявленными сговорами, поэтому классификаторы стали точнее определять признаки наличия картеля. Сравнение таблиц позволяет понять, как изменяются показатели точности и полноты вследствие перебалансировки данных. После добавления синтетических наблюдений точность и полнота для класса «картель» возросли, а для группы, где «нет картеля», – упали. Во всех моделях произошли резкие положительные изменения для картелей, а степень ухудшения показателей для группы без картелей сильно варьирует в зависимости от модели. f1-score (среднее гармоническое между precision и recall) для группы без картелей сильно снизилось в Логистической модели, с 0,81 до 0,68, и в модели k ближайших соседей – с 0,84 до 0,78. Для ансамблевых классификаторов f1-score для группы без картелей мало изменилось: для Случайного леса – с 0,86 до 0,84, для Градиентного бустинга – с 0,85 до 0,84

**Таблица 9. Точность и полнота моделей, показывающих наличие или отсутствие картеля на электронном аукционе, по данным с синтетическими наблюдениями SMOTE**

		precisio n	recall	f1- score	N (тестовая выборка)
Логистическая модель	нет картеля	0,83	0,58	0,68	1523
	картель	0,68	0,88	0,76	1538
k ближайших соседей	нет картеля	0,80	0,77	0,78	1523
	картель	0,78	0,81	0,79	1538
Случайный лес (100 деревьев)	нет картеля	0,86	0,81	0,84	1523
	картель	0,83	0,87	0,85	1538
Градиентный бустинг	нет картеля	0,88	0,80	0,84	1523
	картель	0,82	0,89	0,85	1538

Источник: рассчитано авторами

По показателям точности и полноты модификации SMOTE тоже проигрывают своей базовой модели, таблица с результатами представлены в Приложении 2.

Мы попытались улучшить качество предсказаний путем корректировки самой модели, а не выборки. Были учтены веса классов в модели случайного леса. Но результаты оказались противоречивыми и не показали явных положительных изменений эффективности. После того, как наша модель случайного леса была взвешена по объему классов, правильность предсказаний на тестовой выборке увеличилась с 0,77 до 0,80 (`class_weight= balanced`) и 0,81 (`class_weight= balanced_subsample`), а f1-score, наоборот, снизился для обоих классов.

В итоге, на основании показателей эффективности классификаторов был выбран SMOTE в качестве алгоритма для сглаживания эффекта несбалансированности данных для моделей по горизонтальным картелям на электронных аукционах. С учетом синтетических данных, правильность классификации на тестовой выборке равняется примерно 84% в Градиентном бустинге и Случайном лесе. В двух других моделях правильность предсказания ниже.

## **8. Добавление фиктивных переменных по видам экономической деятельности**

В описанной в предыдущих разделах модели классификатора никак не учитывался объект закупки. А он может сильно влиять как на поведение участников рынка, так и на поведение антимонопольной службы.

Во-первых, есть много ненаблюдаемых характеристик отраслей, от которых могут зависеть количество участников электронного аукциона, объемы закупки, среднее снижение НМЦК в ходе аукциона и т.д. Например, если отрасль характеризуется достаточно конкурентным рынком с большим числом производителей, то количество участников в закупках соответствующих товаров будет больше. Снижение цены закупки тоже будет больше, скорее всего. Поставщики будут понимать, что если они не снизят цену достаточно цену, то это, с большой вероятностью, сделают другие участники. Может быть ситуация, когда большое число производителей активно конкурируют даже за мелкие сделки. С другой стороны, есть категории товаров, которые по всей России производятся только несколькими крупными фирмами. Эти фирмы могут активно соперничать, но в то же время, им легче договориться между собой о распределении закупок и о величине снижения цен. Есть разные структурные особенности рынков, влияющие на вероятность наличия картеля. По мере увеличения однородности продукции, снижении барьеров входа

на рынок и снижении концентрации отрасли уменьшается вероятность образования и поддержания картельного сговора.

Во-вторых, различные отрасли могут иметь разную социальную и экономическую значимость для государства. При совершении преступления на госзакупках в одной отрасли потери благосостояния общества могут быть намного выше, чем при совершении того же преступления в другой. Одно дело, когда идет закупка каких-нибудь канцелярских товаров или мебели для госучреждения, другое дело – покупка лекарств или технически сложных аппаратов для медицинских учреждений. Есть производства, для которых требуются сильно специфические материалы и сырье, и если традиционные поставщики материалов будут завышать цены и отказываться продавать по более низким ценам, то производство некоторого товара может просто остановиться, и сбои в производстве могут привести к огромным потерям для государства. Контролирующие органы, вероятно, будут чаще и усерднее проверять закупки, например, по социально значимым сферам экономики, таким как: образование, здравоохранение, транспорт, энергетика.

Всё это говорит о том, что в модель классификатора нужно вводить контроль на объект закупки. В следующем разделе описано, какие есть данные о закупаемом товаре, и какие трудности возникают при определении объектов закупок и формировании переменных для модели.

### **8.1 Сбор и анализ данных об объекте закупки**

В документах электронного аукциона есть несколько параметров, которые характеризуют объект закупки. Часть из них – это текстовые описания, а часть – кодовые представления товара. В xml-документе извещения содержится:

1. код ОКПД или ОКПД2;
2. общая информация о закупке – «purchaseObjectInfo» (Это короткое описание закупки, в одно предложение);
3. точное наименование кода ОКПД или ОКПД2;
4. название товара/услуги.

Первое, что можно использовать для выделения групп товаров, это коды по видам экономической деятельности. Но здесь возникает несколько трудностей. Первая проблема заключается в том, что в некоторых закупках использован ОКПД, а в других – ОКПД2 (утвержден в 2014 году, но окончательно на него перешли только с начала 2017 года). В нашей выборке по электронным аукционам за 2014-2016 гг. оказалось достаточно много закупок, распределенных по старому классификатору ОКПД. Два указанных классификатора довольно сильно отличаются. Исходя из этого, первая задача, которую

нужно решить, это привести все коды в нашей выборке к одному виду, например, к ОКПД2. Ключи соответствия ОКПД и ОКПД2 были взяты с сайта Минэкономразвития<sup>7</sup>. В закупках, в которых использовался ОКПД все коды заменили на ОКПД2, и затем создали общий столбец с приведенными кодами по всем закупкам (то же самое сделали с названиями кодов).

Вторая трудность заключается в том, что многие закупки состоят из нескольких лотов, для каждого из которых указаны название товара или услуги и их коды ОКПД (ОКПД2). В нашей выборке количество кодов ОКПД (ОКПД2) для одного электронного аукциона варьировалось от 1 до 80. Исходя из этого, вторая задача – это определение основного товара закупки.

Для каждого лота была взята его стоимость, и затем стоимости лотов взвешены по НМЦК. В качестве основного лота был выбран товар, стоимость которого занимала наибольшую долю в заказе. Если было много лотов одного или примерно одного размера, то они принадлежали похожим ОКПД, т.е. близким категориям товаров, поэтому такой способ выбора объекта закупки не сильно сместил распределения категорий товаров.

Третья трудность - выделение относительно крупных групп видов экономической деятельности. Категория товаров – качественный признак, который можно включить в модель в виде бинарных переменных. В документации электронного аукциона указываются точные коды закупаемых товаров, т.е. указываются коды до 9 знака включительно ХХ.ХХ.ХХ.ХХХ (если брать такую детализацию, то может быть около 24000 разных кодов). Включение слишком большого числа переменных по узким товарным группам приведет, скорее всего, к сильной мультиколлинеарности. Если говорить о машинном обучении, то включение слишком большого числа переменных при не очень большой выборке увеличивает риск переобучить модель. Классификатор будет слишком сильно зависеть от конкретных данных в обучающем наборе.

Сначала в коде основного объекта закупки были оставлены только первые два знака – классы товаров. Затем некоторые близкие классы были объединены в более крупные категории. В итоге получилось 46 укрупненных групп товаров по видам экономической деятельности. Таблица с соотношением кодов ОКПД2 и групп товаров, по которым далее строились бинарные переменные для модели представлена в Приложении 3.

Ещё один подход к выделению товарных групп – это анализ текста, относящегося к общей информации о закупке, или наименовании ОКПД2, или названию товара.

<sup>7</sup> <http://economy.gov.ru/minec/activity/sections/classificators/> Переходные ключи для общероссийских классификаторов

наличия проблем с кластеризации векторных представлений этих текстов этот способ в данной работе не используется.

В итоге разбиение на группы по видам экономической деятельности при помощи кодов ОКПД2 оказывается наиболее объективным и удобным подходом. Поэтому в модель для выявления картелей на электронных аукционах мы включаем фиктивные переменные, построенные по группам из таблицы (см. приложение 2).

Для товарных групп были посчитаны основные показатели (Таблица 10). В таблице ниже отрасли проранжированы по убыванию количества сделок, попавших в выборку.

На первом месте и по количеству наблюдений, и по количеству зафиксированных картелей располагается 21 класс ОКПД2 – средства лекарственные и материалы для медицины. На втором месте – 32 класс ОКПД2 – прочие готовые изделия. 3-е место по количеству закупок в выборке занимает сфера строительства. По абсолютному числу картелей строительство занимает 5 место, здесь выявлено 64 горизонтальных сговора. Но если смотреть на долю картелей в общем числе рассмотренных закупок в отрасли, то этот показатель для строительной отрасли равен всего 7,5%, что относительно низко.

Как было сказано, в области медицины чаще всего выявляются картели. Вероятно, доля картелей в других отраслях может быть завышена из-за того, что закупки медицинского оборудования могут находиться также в «прочих готовых изделиях» и в «компьютерном оборудовании». Рассмотрим подробнее эти классы.

**Таблица 10. Количественные характеристики выборки электронных аукционов в разрезе видов экономической деятельности**

коды ОКПД2	группа по видам экономической деятельности	количество сделок	обнаружено картелей	среднее снижение цены	средняя НМЦК в тыс.руб.	std НМЦК	общий объем НМЦК тыс.руб,	количество участников в среднем	std количество участников
21	средства лекарственные и материалы для медицины	1746	993	0,07096	1696,33	9450,443	2 961 792	2,504	0,992
32	прочие готовые изделия	1543	839	0,06158	6900,467	18367,65	10 647 421	2,279	0,776
41-43	строительство	843	64	0,0838	11588,05	45608,28	9 768 726	3,123	1,371
26	оборудование компьютерное, электронное и оптическое	570	164	0,07999	12069,41	58169,77	6 879 566	2,605	1,239
20	химические вещества и продукты	513	248	0,07357	3739,68	12870,98	1 918 456	2,692	1,291
27-28	прочее оборудование	361	31	0,17622	4254,601	17368,59	1 535 911	3,568	2,135
14	одежда	184	119	0,24662	2563,253	22293,49	471 638,6	3,288	1,116
17	бумага	181	8	0,22423	900,2582	1319,548	162 946,7	3,624	1,746
58-63	услуги в области информации и связи	175	39	0,1583	4957,833	16722,04	867 620,7	2,634	1,261
22	резина и пластик	174	56	0,17867	1466,879	4389,424	255 236,9	3,08	1,464
10	пищевая продукция	165	53	0,14561	2849,207	13359,79	470 119,1	2,927	1,382
33	ремонт и монтаж машин и оборудования	161	20	0,1795	5571,566	11061,66	897 022,1	3,205	1,972
29-30	транспортные средства	139	33	0,08739	26635,57	89302,04	3 702 344	2,712	1,187
24-25	металлы и металлическая продукция	117	22	0,14049	2698,068	6369,021	315 673,9	3,393	1,843
31	мебель	110	17	0,12298	3348,927	14823,1	368 381,9	3,064	1,699

77-82	услуги административные и вспомогательные	102	8	0,16468	20938,1	159841,1	2 135 686	3,157	1,806
46-47	торговля оптовая и розничная	70	9	0,17222	2702,093	9798,593	1 89 146,5	3,4	1,527
70-75	научная и инженерно-техническая деятельность	70	1	0,35284	972,6708	1896,65	68 086,95	3,057	1,214
23	проч. продукты минеральные неметаллич.	62	6	0,0718	1997,168	2913,916	123 824,4	2,855	1,252
19	кокс и нефтепродукты	58	0	0,10657	1180,112	2575,457	68 446,52	3,259	1,97
94-99	прочие услуги	55	3	0,27468	4454,544	15455,76	244 999,9	3,164	1,512
05-09	горнодобывающая пр-ть	52	38	0,0461	29242,01	186970,7	1 520 585	3,077	1,296
84	гос.управление и обеспечение военной безопасности	50	7	0,08537	3700,008	5465,988	185 000,4	3,68	1,659
85	образование	49	1	0,35111	739,0405	1107,026	36 212,99	3	1,258
1	сельское хозяйство	48	18	0,19784	3961,935	21765,81	190 172,9	3,063	1,405
38-39	рекультивация и утилизация отходов	48	14	0,15526	2848,74	6939,751	136 739,5	2,625	0,89
13	текстиль	42	3	0,13346	1286,192	2914,196	54020,06	3,119	1,641
86	здравоохранение	37	10	0,27382	797,1498	907,4694	29 494,54	2,973	1,213
45	торговля автотранспортом	32	3	0,25202	5261,307	11756,26	168 361,8	3,125	1,519
93	спорт и отдых	32	0	0,13943	1195,255	1677,172	38 248,17	2,625	1,476
64-66	услуги финансовые и страховые	30	0	0,22888	7297,752	8005,858	218 932,5	3,6	1,38
55-56	гостиницы и общественное питание	29	5	0,12341	4448,691	6692,597	129 012	2,724	1,131
18	услуги печатные и услуги по копированию	27	0	0,31184	853,4906	945,3081	23 044,25	3,963	1,87

90-91	искусство и культура	20	1	0,22461	2442,051	4778,919	48 841,03	2,9	1,483
49-53	транспортные услуги	19	1	0,16793	25197,68	102190	478 756	2,895	1,41
2	лесоводство	15	0	0,27569	650,0587	503,2625	9 750,881	3,533	1,457
68	услуги по недвижимости	14	0	0,26205	1521,245	1448,92	21 297,43	2,857	0,864
15	изделия из кожи	11	0	0,1114	2099,39	2657,228	23 093,29	2,636	1,027
36-37	водоснабжение	11	0	0,45697	859,4951	1976,136	9 454,446	4	1,483
16	древесина	10	0	0,18872	1292,16	2127,531	12 921,6	2,8	0,632
35	электроэнергия газ и кондиционирование воздуха	7	3	0,16059	1711,857	2227,758	11 983	2,429	0,787
11	напитки	2	0	0,55889	102,116	35,04987	204,232	5	1,414
69	юридические и консультационные услуги	1	0	0,38	65		65	2	
87-88	социальные услуги	1	0	0,01	1390,908		1 390,908	2	
3	рыболовство и рыбоводство	0	0	0	0	0	0	0	0
12	табачная продукция	0	0	0	0	0	0	0	0

Источник: построено авторами

К прочим готовым изделиям относятся товары из разных отраслей, поэтому были рассмотрены более узкие категории (код до 4 знака). Оказалось, что в «прочих готовых изделиях» 99% всех обнаруженных картелей (828 из 839) и 92% всех рассмотренных электронных аукционов (1416 из 1543) занимают медицинские инструменты и оборудование, код 32.50. А в классе «оборудование компьютерное, электронное и оптическое» 111 из 164 картелей относятся к подгруппе 26.60 – «Оборудование для облучения, электрическое диагностическое и терапевтическое, применяемые в медицинских целях».

Объединим группы с ОКПД2 32.50 и 26.60 в отдельную категорию «медицинское оборудование».

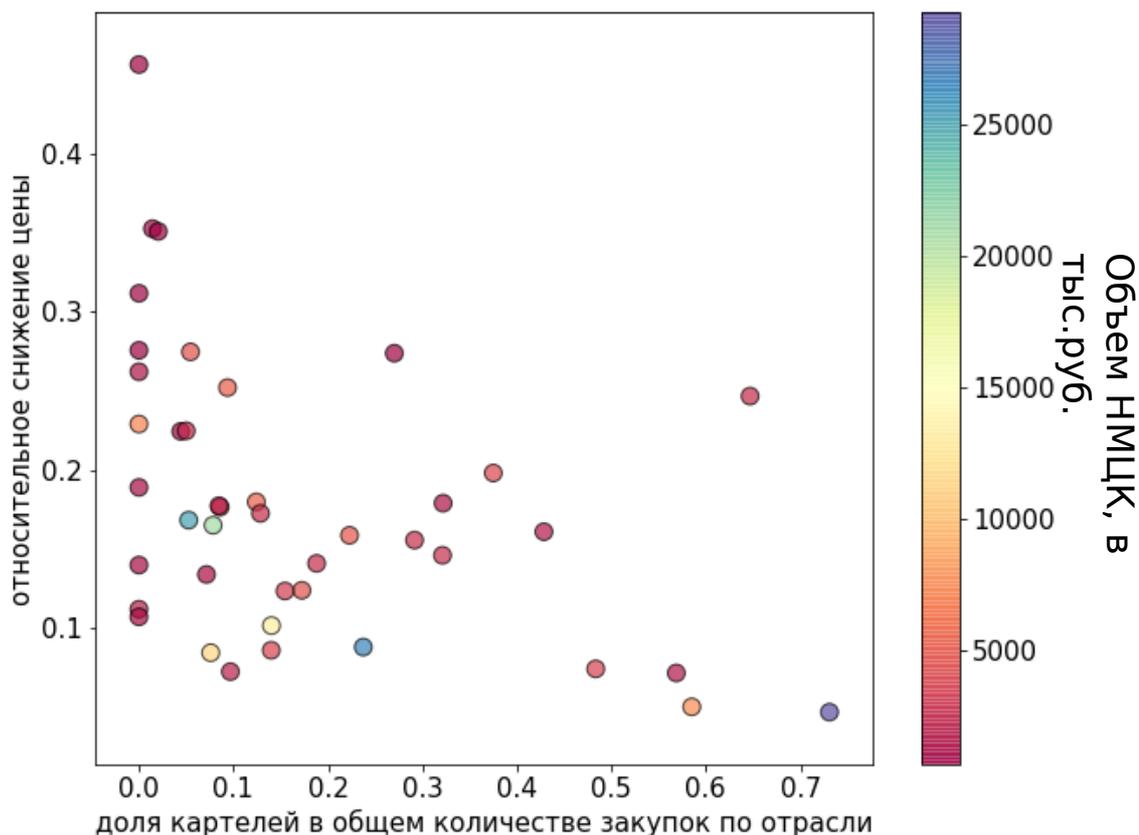
Результаты количеству картелей в наших категориях товаров во многом совпадают с выводами ФАС относительно того, в каких отраслях больше всего картелей [46]. Это говорит в пользу того, каким способом мы выделяли основной вид экономической деятельности в ЭА. Но в то же время, сама наша выборка оказалась не сбалансированной по отраслям. Доли ЭА, относящихся к разным видам экономической деятельности,

отличается от распределений, представленных в статистических отчетах Минфина об осуществлении госзакупок по 44-ФЗ. Поэтому результаты о долях картелей в отраслях могут быть смещены.

После преобразований в пятерку отраслей с наибольшей долей картелей вошли:

1. горнодобывающая промышленность – 73,1% (38 ЭА с картелем из 52 ЭА);
2. одежда – 64,7% (119 из 184 ЭА);
3. медицинское оборудование – 58,5% (939 из 1605 ЭА);
4. лекарственные средства и медицинские материалы – 56,9% (993 из 1605 ЭА);
5. химические вещества и продукты – 48,3% (248 из 513 ЭА).

Тем не менее, результаты расчетов показывают, чем больше доля картелей в отрасли, тем меньше в среднем снижается цена на ЭА. Это наглядно отражено на диаграмме рассеяния, показывающей взаимосвязь между относительным снижением цены в ходе ЭА и долей картелей в общем количестве рассмотренных закупок в отрасли (Рисунок 22). К таким отраслям относятся горнодобывающая промышленность, медицинское оборудование и препараты, химические вещества, транспортные средства, услуги госуправления и т.д. Кроме того, из графика видно, что если средняя величина НМЦК высокая, то относительное снижение цены будет небольшим (Цвет маркеров на графике зависит от НМЦК). В горнодобывающей промышленности, транспортных средствах и транспортных услугах средняя НМЦК была больше 25 млн. рублей, а снижение цены – в среднем меньше 17%.



**Рисунок 22. Взаимосвязь между снижением цены и количеством картелей, с учетом среднего НМЦК по отрасли**

Источник: построено авторами

На рис. выше (Рисунок 22) представляет интерес не только форма статистической зависимости, но и то обстоятельство, что не обнаружено отраслей с высокой долей картелей в закупках отрасли, имеющих высокое относительное снижение цены. Кроме того, в контексте предшествующего обсуждения того, как происходит конкуренция в разных отраслях и как она отражается в ежегодном докладе ФАС, обращает на себя внимание и относительно высокий уровень снижения цены на услуги в сфере ИТ.

## 8.2 Результаты классификации с учетом видов экономической деятельности

Построим классификатор, включив новые бинарные переменные, показывающие принадлежность товара к отрасли. Исключим только отрасли, где было менее пяти наблюдений.

Так же, как и в Разделе 6, мы строим синтетические наблюдения класса меньшинства (закупок с горизонтальным сговором), затем делим выборку на

тренировочную (70%) и тестовую (30%). На одной части выборки тренируем классификатор, и на другой – проверяем качество предсказаний модели.

Как и в предыдущих спецификациях, наилучшие результаты показывают ансамблевые модели, самая низкая эффективность – у Логистической модели.

Правильность предсказания классификатора на тестовой выборке оказалась равной

- 0,792 – в Логистической модели,
- 0,839 – в модели к ближайших соседей,
- 0,887 – в модели Случайного леса,
- 0,884 – в модели Градиентного бустинга.

После включения фиктивных переменных по ОКПД2 эффективность моделей в определении наличия/отсутствия картеля возросла во всех моделях на несколько проц.пунктов. Мера f1 (среднее гармоническое между precision и recall) в Случайном лесе и Градиентном бустинге приблизилась к 0,9, как для картелей, так и для честных сделок. Полнота модели для класса «картель» в Случайном лесе равняется 0,90. Это значит, что классификатор верно отнес к нечестным 90% из всех выявленных антимонопольной службой горизонтальных сговоров. Точность (precision) равняется 0,88. Значит, из всех наблюдений, отмеченных классификатором как нечестные, 88% определены верно, и только 12% не содержат выявленных картелей. Это говорит о том, что наши модели могут довольно точно предсказать, будет ли выявлен горизонтальный сговор на том или ином аукционе.

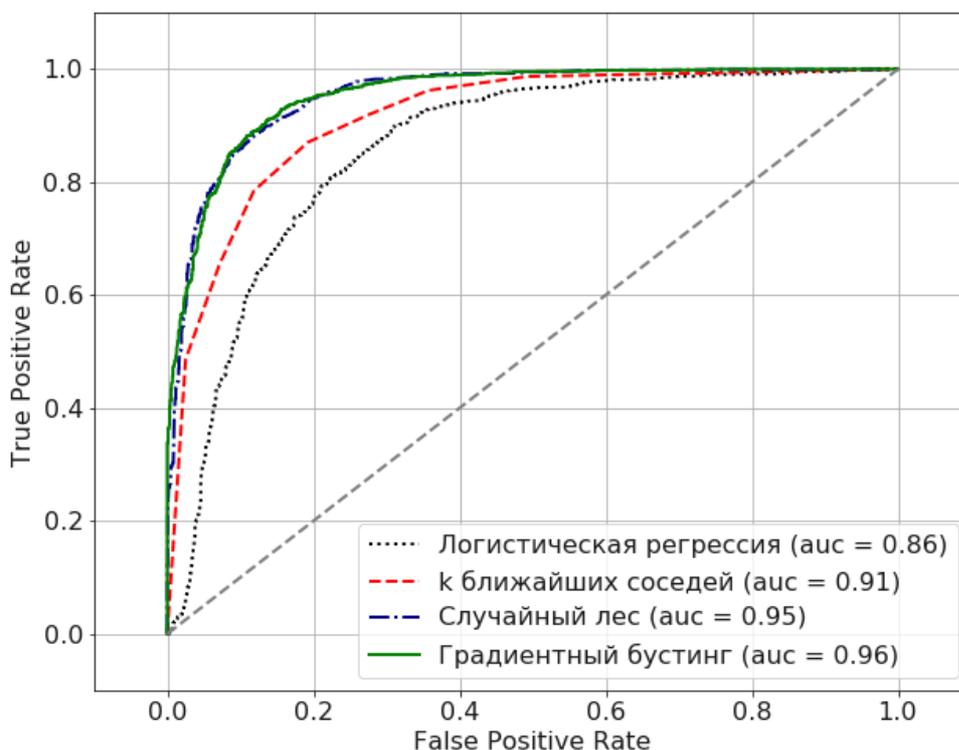
**Таблица 11. Точность и полнота моделей, показывающих наличие или отсутствие картеля на электронном аукционе, с учетом SMOTE и фиктивных переменных по видам экономической деятельности**

		precisio n	recall	f1- score	N (тестовая выборка)
Логистическая модель	нет картеля	0,84	0,72	0,78	1550
	картель	0,75	0,87	0,81	1543
к ближайших соседей	нет картеля	0,86	0,81	0,83	1550
	картель	0,82	0,87	0,84	1543
Случайный лес (100 деревьев)	нет картеля	0,89	0,88	0,89	1550
	картель	0,88	0,90	0,89	1543
Градиентный бустинг	нет	0,91	0,85	0,88	1550

	картеля				
	картель	0,86	0,92	0,89	1543

Источник: рассчитано авторами

Ещё одним показателем высокой предсказательной силы наших моделей служат ROC-кривые, которые показывают соотношение между долей верных предсказаний некоторого класса к доле ложноположительных предсказаний этого же класса. ROC-кривые наших моделей располагаются далеко от 45-градусной диагонали и близко к левому верхнему углу, что говорит об их хорошей предсказательной силе. Площадь под кривыми ROC, обозначается аббревиатурой AUC. Чем больше значение AUC, тем лучше модель в определении 0-класса как 0, а 1-класса как 1. В наших моделях AUC близок к 1, это значит, что классификаторы хорошо отличают аукционы, в которых выявлен горизонтальный сговор, от аукционов, где такое нарушение не выявлено.



Источник: построено авторами

**Рисунок 23. ROC-кривые для классификатора по картелям, составленного с учетом отрасли объекта закупок**

Некоторые переменные имеют больший вес в прогнозировании наличия картеля. Например, один из самых важных показателей – это доля снижения цены в ходе аукциона. Если, например, цена снизилась меньше, чем на 5%, то вероятность того, что закупка

проходила нечестно, с участием картеля, сильно увеличивается при прочих равных. Для Случайного леса и Градиентного бустинга можно вычислить коэффициенты важности переменных (importances) в определении классов. В таблице ниже (Таблица 12) указаны переменные, важность которых для определения класса не меньше 1%. Величина коэффициентов и место переменных в таблице отличаются в моделях. Но многие из переменных содержатся и в том, и в другом списках.

**Таблица 12. Важность переменных в классификаторах, с учетом фиктивных переменных по видам экономической деятельности**

Случайный лес			Градиентный бустинг		
	переменная	важность		переменная	важность
1)	относительное снижение цены	0,263	1)	относительное снижение цены	0,274
2)	Медиана разницы во времени подачи заявок	0,142	2)	Количество участников в протоколе 2	0,129
3)	НМЦК в тыс. руб.	0,114	3)	Медиана разницы во времени подачи заявок	0,089
4)	Относительная разница между минимальной и второй ценами	0,085	4)	Медицинские препараты (код 21)	0,073
5)	Количество участников в протоколе 2	0,066	5)	Медицинское оборудование (коды 26.60 и 32.50)	0,063
6)	Медицинские препараты (код 21)	0,058	6)	Строительство (коды 41-43)	0,055
7)	Медицинское оборудование (коды 26.60 и 32.50)	0,035	7)	НМЦК в тыс. руб.	0,051
8)	Строительство (коды 41-43)	0,029	8)	Относит. разница между миним. и второй ценами	0,048
9)	Одежда (код 14)	0,029	9)	Одежда (код 14)	0,041
10)	Разница в количестве участников в 1 и 2 протоколах	0,024	10)	Количество разъяснений	0,029
11)	Количество разъяснений	0,022	11)	Максимальное количество ставок одного участника	0,028
12)	Максимальное количество ставок одного участника	0,018	12)	Химические вещества (код 20)	0,017
13)	Химические вещества (код 20)	0,012	13)	Разница в количестве участников в 1 и 2 протоколах	0,014
			14)	прочее оборудование	0,013

				(коды 27-28)	
			15)	Компьютерная техника (код 26), кроме медицинской	0,012
			16)	Прочие товары (коды 32-33)	0,012
			17)	резина и пластик (код 22)	0,012

Источник: рассчитано авторами

Подводя итоги рассмотрения подхода, основанного на Ex-post анализе признаков горизонтального сговора, хотелось бы отметить несколько вещей. Во-первых, полученные модели позволяют с вероятностью близкой к 90% определить, выявит ли антимонопольная служба на конкретном электронном аукционе горизонтальный сговор между поставщиками. При этом, с помощью классификаторов можно посчитать для каждого аукциона вероятность обнаружения в нём картеля. Во-вторых, графики распределения показателей аукционов демонстрируют, при каких значениях переменных вероятность наличия картеля выше. В-третьих, результаты модели могут служить основанием для проведения более детальной проверки аукциона сотрудниками ФАС. Здесь важно отметить, что если классификатор не отнес закупку к нечестным, то это не значит, что в ней точно нет картеля, это говорит лишь о том, что по характеристикам она меньше похожа на те закупки, в которых антимонопольная служба обычно выявляет картели. Если же классификатор относит сделку к нечестным, то это означает, что с большой вероятностью в закупке участвует картель, и он будет выявлен ФАС. Мы не можем говорить о точном наличии или отсутствии картеля, потому что истинное количество картелей нам неизвестно, и эффективность классификаторов оценивается относительно совпадений с решениями ФАС о наличии или отсутствии горизонтального сговора на аукционе.

## 9. Ex-ante анализ признаков сговоров

Теперь мы переходим к подходу выявления сговоров, когда у нас нет данных об их наличии, по которым мы могли бы обучить классификатор. Идея подхода заключается в том, чтобы выделить кластеры поставщиков (возможные сговоры) на основе показателей их взаимодействия при участии в госзакупках. Эта стратегия будет основываться на построении и анализе графов.

Первым плюсом этого подхода является то, что мы можем использовать всю совокупность закупок. В предыдущем подходе у нас возникало ограничение по объему

выборки, так как если увеличивать случайную выборку и оставлять фиксированной выборку по делам ФАС, то данные будут становиться всё более несбалансированными. При тренировке классификатора на сильно несбалансированной выборке хорошо будет предсказываться только наиболее крупный класс, в нашем случае – это электронные аукционы, где не обнаружен сговор. И несмотря на то, что есть разные методы, которые помогают уменьшить падение качества предсказаний, возникающее из-за несбалансированности данных, но полностью нивелировать этот эффект трудно, и по мере увеличения несбалансированности снижается точность классификатора.

Ещё одной слабой стороной предыдущего подхода является то, что выводы о наличии картеля на закупке очень сильно зависят от качества работы ФАС и особенностей её деятельности. Важно, что истинное количество картелей неизвестно, и эффективность наших моделей оценивается относительно совпадений с решениями ФАС о наличии или отсутствии сговора на аукционе. Данные ФАС, скорее всего, не показывают реальное число нарушений, потому что в решениях ФАС могут присутствовать ошибки как I, так и II рода. Во-первых, у антимонопольной службы могло не найтись достаточного количества доказательств вины некоторых компаний. Во-вторых, по некоторым рассмотренным аукционам могут ещё вестись дела, т.е. по ним нет заключения относительно наличия картеля. Кроме того, крупные сделки, как правило, привлекают больше внимания антимонопольной службы, потому что приводят к большим потерям со стороны потребителей и производителей. В связи с этим, небольшие аукционы, на которых присутствовали картели, могут быть не отмечены как нечестные. Также, к примеру, антимонопольная служба может уделять большее внимание закупкам по социально значимым отраслям.

### **9.1 Обзор основных элементов теории графов**

В сетевом подходе взаимодействующие элементы изображаются как вершины сети, а их взаимодействия – как ребра, соединяющие вершины. Проводимые исследования выявили множество различных структурных характеристик сетей, которые могут многое сказать об особенностях системы и взаимодействия её составляющих. К таким структурным показателям относятся, например, распределение степеней, длины путей между узлами, модульность и уровень кластеризации. Очень подробный обзор структурных свойств, мер и моделей сетей, а также процессов, происходящих в сетях, можно найти, например, в работе Ньюмана 2003 года [33].

Теория графов имеет важное значение для анализа социальных сетей, которые показывают взаимоотношения между социальными субъектами, в частности: общение и

обмен информацией между членами групп, развитие социальных интернет сетей, осуществление торговых сделок между фирмами или странами и т.д. [14]. В последнее время особый интерес уделяется изучению сложных сетей, т. е. сетей, структура которых нерегулярна, сложна и динамично развивается во времени [7, 41].

Определение центральности узлов, или выявление наиболее “центральных”, было долго одной из основных задач при анализе сетей Freeman, 1978 [15], Borgatti et al., 2006 [9]. В статье Фримена 1978 года [15] были сформулированы три различных показателя централизации узлов: степень, близость и центральность по посредничеству.

Во многих сетях связи между узлами неодинаковы и различаются по своей мощности и интенсивности. Такие различия в связях характерны для многих реальных ситуаций, и проявляются, например, в наличии сильных и слабых связей между людьми в социальных системах, в различных возможностях передачи сигналов в нейронных сетях, неравномерном интернет-трафике и т.д. Модели, в которых учитываются эти отличия, называются взвешенными сетями и сейчас активно применяются [35, 32, 10]. В случае анализа взаимоотношений между участниками электронных аукционов важно учитывать то, как часто компании взаимодействуют друг с другом. В контексте взвешенных сетей возник вопрос о том, считается ли большое число связей с малым весом важнее, чем общая сумма весов связей, или связи с большими весами оказывают гораздо более значимое воздействие, чем много связей, но только с маленькими весами. Opsahl et al. [35] постарались учесть оба показателя (и степень, и силу связей) при определении роли узла.

В работе [40] проведен подробный обзор литературы по проблеме обнаружения сообществ во времени с применением сетевого и графового анализа. Наиболее популярны следующие модели: Случайные графовые модели, Экспоненциальные случайные графовые модели, Блочные модели, Сетевые пространственные модели, Скрытые марковские модели и т.д

## **9.2 Описание стратегии и первые результаты**

Графо-теоретический подход к выявлению кластеров поставщиков, вероятных картелей, предполагается использовать на всей генеральной совокупности госзакупок, проведенных с помощью электронного аукциона в последние несколько лет. Использование больших данных накладывает ограничения на подход. В частности, представляется невозможным выгрузить и составить таблицу, а также обрабатывать её сразу по всем годам и по всем отраслям из-за слишком большого объема выборки. Например, в 2018 году было опубликовано около двух миллионов извещений о проведении электронных аукционов [44]. Кроме того, получаются панельные данные очень большой

размерности: т.е. нужно учитывать и отрасль, и время, и регион. При этом по каждой отрасли закупки проводятся с разной периодичностью, что также усложняет рассмотрение данных как панели с единой временной шкалой. С учетом этого, для целей кластеризации удобнее рассматривать отдельно разные категории товаров и услуг.

#### I этап

На первом этапе выделяем из генеральной совокупности последовательности закупок по отраслям. Затем перегруппировываем данные по поставщикам и формируем таблицу узлов для будущих графов. При формировании таблицы с узлами, необходимо удалить дублирующиеся наблюдения, т.е. повторное упоминание тех поставщиков, которые участвовали в нескольких аукционах. Для формирования ребер графов из каждой закупки вытаскиваются уникальные пары поставщиков. Число таких пар равняется числу сочетаний из  $n$  объектов (где  $n$  – это число участников закупки) по  $k$  (где  $k=2$ , так как мы рассматриваем пары поставщиков).

После того, как сформированы таблицы с узлами и ребрами, составляется ненаправленный граф, каждый узел которого – это фирма-поставщик, ребро между двумя узлами  $A$  и  $B$  – наличие закупки/закупок, в которых вместе участвовали фирмы  $A$  и  $B$ .

#### II этап

На втором этапе, на основании графов мы выделяем группы поставщиков, которые участвовали в  $k\%$  закупок за последний год вместе. Порог для  $k$  устанавливается с учетом взвешивания по уровню концентрации в отрасли. Если отрасль характеризуется большим числом фирм, и все эти фирмы активно участвуют в госзакупках. То вероятность того, что две случайные фирмы будут попадать в всегда или почти всегда только на одни и те же закупки, будет очень низкой. По крайней мере, данная вероятность будет ниже, чем в случае олигополистической отрасли. Если на рынке какого-то товара присутствует всего несколько поставщиков, то они будут вместе участвовать почти в 100% закупок этого товара. Но это не значит, что они обязательно будут участвовать в сговоре и завышать цены. Маленькое количество фирм в отрасли способствует увеличению вероятности формирования картеля, но не говорит об обязательном его наличии. В ходе отраслевого жизненного цикла есть ситуации, когда несколько крупных фирм активно конкурируют между собой, и уровень цен намного ниже, чем он мог бы быть при конкуренции между большим количеством мелких компаний. Исходя из вышесказанного, возникает необходимость корректировать в зависимости от концентрации в отрасли порог отсечения кластеров  $k$ .

Значение количества закупок, в которых поставщики вместе участвовали, равняется весу ребра в нашем графе. Для каждой пары узлов вес ребра между этими узлами известен, что позволяет выделить группы поставщиков, для которых вес ребра является высоким. На этом шаге мы сужаем выборку «подозрительных» поставщиков.

### III этап

Фирмы могут участвовать вместе в большом количестве закупок, но не находиться в сговоре друг с другом. Важным показателем возможного наличия картеля является доля побед каждого из игроков кластера, определенного на II этапе, а также суммарная доля их побед.

Пусть есть  $b$  закупок и  $p$  участников, которые участвовали вместе в этих закупках. Если какая-то доля из  $p$  – участники картеля, то им выгодно делить между собой сделки. Т.е. члены картеля договариваются между собой, контракты по каким закупкам достанутся каждому участнику сговора. В таком случае количество побед каждого из участников должно быть примерно одинаковым. При этом суммарная доля побед участников картеля в указанных  $b$  закупках должна быть высокой. Если мы видим, что какие-то два участника вместе участвовали в большом количестве сделок, но ни разу в них не побеждали, то такая ситуация будет, скорее, свидетельствовать об отсутствии сговора между этими фирмами, так как противоречит основной цели создания картеля – победе на закупках с минимальным снижением цены. Однако такая ситуация может быть, если обе эти фирмы – подставные участники, то есть участвуют только для того, чтобы увеличить число участников, создать видимость конкуренции или просто обеспечить проведение электронного аукциона. Это могут быть либо фиктивные фирмы, которые не ведут реальную производственную деятельность, либо реальные фирмы, которым платят за то, чтобы они участвовали, но не делали конкурентные ставки. Если имеет место ситуация с подставными участниками, то должна быть ещё одна фирма из  $p$ , которая участвует в аукционах вместе с подставными и побеждает во всех этих закупках. Отбросить фиктивные компании можно, например, за счет контроля на годовую выручку фирмы.

Итак, следующим показателем наличия картеля является ситуация, когда фирмы не только вместе участвуют в большом числе аукционов, но и когда суммарная доля их побед в этих аукционах очень высокая. Поэтому на III этапе выборка «подозреваемых» сужается за счет выделения из кластеров, полученных на II этапе, групп компаний, которые в сумме

побеждали в  $n\%$  закупок. Порог для  $n$  тоже устанавливается с учетом взвешивания по уровню концентрации в отрасли.

#### IV этап

На последнем, четвертом этапе, фильтрация компаний, поведение которых похоже на поведение членов картеля, осуществляется на основании различных показателей аукционов. В первую очередь, важно обратить внимание на среднее относительное снижение НМЦК в аукционах, в которых принимала участие фирма. Также можно учитывать максимальное и среднее число ставок в аукционах.

Благодаря тому, что в ходе предыдущих этапов выборка должна сильно сжаться, потребуются меньшие компьютерные мощности и меньшее время для сбора указанных переменных из данных об электронных аукционах.

### **9.3 Апробирование графо-теоретического подхода к выявлению картелей между поставщиками**

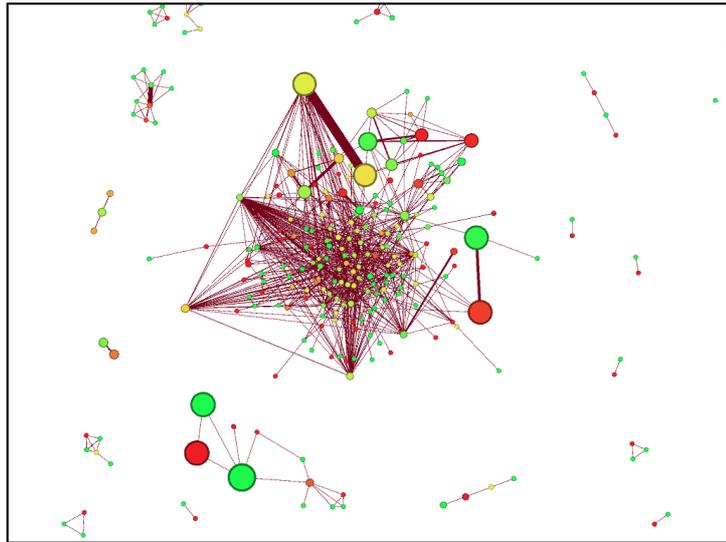
Графо-теоретический подход к выявлению картелей между поставщиками был протестирован на выборке электронных аукционов, используемой в предыдущем подходе.

Из выборки были взяты несколько видов экономической деятельности, по которым в выборку попало наибольшее число наблюдений, и для которых было выявлено в прошлом подходе наибольшее число картелей. К видам экономической деятельности относятся: средства лекарственные и материалы для медицины; строительство. Выборка по медицинским препаратам составила 1746 закупок, по строительству – 843 закупки. В них ФАС обнаружено 993, 64 картелей соответственно. Также для сравнения была взята группа с относительно маленькой долей обнаруженных антимонопольной службой картелей, в качестве таковой был выступает это бумажная промышленность (17 код ОКПД2), по ней объем выборки составил 181 наблюдение.

Для каждой из указанных отраслей были построены графы, в которых

- Узлы – участники электронных аукционов;
- Ребро между узлами А и В – факт участия фирм А и В в одной закупке;
- Толщина ребра зависит от числа совместных закупок;
- Цвет узла показывает долю побед фирмы в рассмотренных электронных аукционах (Красный – победа в 100% закупок, зеленый – в 0% закупок, желтый – 50%);
- Размер узла зависит от суммарного объема НМЦК, в которых участвовала фирма.

Для сферы медицинских препаратов и материалов получился следующий граф (Рисунок 24).

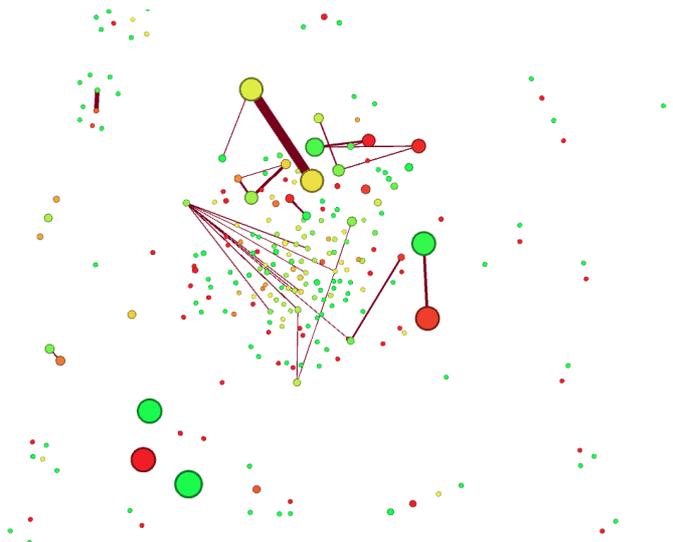


Источник: построено авторами

### **Рисунок 24. Граф закупок медицинских препаратов и материалов для медицины**

Анализ графа закупок медицинских препаратов показывает, что в 1746 закупках принимали участие 397 участников, т.е. в графе есть 397 узлов, при этом – 885 ребер. Обращает на себя внимание то, что есть некоторые поставщики, которые принимали участие в очень большом числе закупок. Узлы, обозначающие такие компании, имеют высокую степень центральности. Однако стоит отметить, что фирмы, обозначенные этими узлами, участвовали обычно в небольших закупках, об этом говорит размер узла: даже с учетом большого числа закупок, суммарный объем НМЦК, в которых участвовала фирма, остается очень маленьким. Компании, которые претендовали на очень крупные контракты, имеют мало связей. Это означает, что они участвуют редко, но зато – в аукционах с высокой начальной максимальной ценой контракта.

Важно, что большая часть ребер тонкие, т.е. поставщики всего один или несколько раз участвовали вместе. Есть всего несколько компаний, которые имеют друг с другом очень плотные связи, выражающиеся в весе ребра (толщина ребра). Минимальный вес ребра в этом графе равняется 1, максимальный – 270, причем только 25 ребер имеют вес больше 10 (Рисунок 25).



Источник: построено авторами

### **Рисунок 25. Граф закупок медицинских препаратов и материалов для медицины -2**

Примечание: Отмечены ребра с весом больше 10

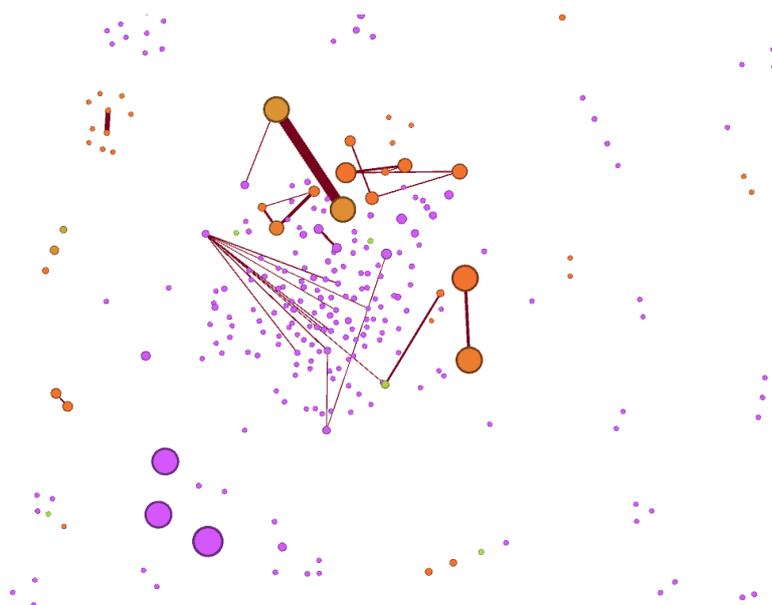
Как было сказано в прошлом разделе, следующее, на что стоит обратить внимание при ответе на вопрос о наличии сговора между участниками, это то, какая суммарная доля побед в закупках для подозреваемых участников. На наших графиках доля побед отображается цветом узла. Из графика (Рисунок 25) видно, например, что есть две компании, которые участвовали почти всегда только вместе, у них было большое количество совместных сделок, и при этом, победы в аукционах распределились примерно поровну между этими двумя компаниями. Указанные фирмы обозначены на рисунке крупными узлами желтого цвета, толщина ребра (вес) – самая большая из всех ребер графа. Первая фирма участвовала в 280 аукционах, а вторая в 273, и в 270 из закупок – вместе. Средний объем НМЦК электронных аукционов, в которых участвовали эти две фирмы, равняется 1,162 и 1,177 млн рублей. Поведение описанных выше фирм очень похоже на поведение картеля. Такое предположение подтверждается также средним относительным снижением цены в аукционах, в которых участвовали обе компании. В аукционах, в которых участвовали две фирмы цена снижалась в среднем на 1,5%.

Также на графике (Рисунок 24, Рисунок 25) видны, например, две фирмы, которые вместе участвовали более чем в 10 аукционах, кроме них в этих закупках не было участников. Одна из этих фирм выиграла в 100% совместных закупок. Такая ситуация тоже похожа на сговор. Вероятно, что компания, которая всегда проигрывала, является подставной и создает только видимость конкуренции на закупке. Между фирмами может быть заключен договор о том, что одна компания участвует и проигрывает, а вторая,

например, платит ей за это деньги. Аналогичная ситуация может быть среди группы фирм из 6 участников, находящихся в верхней части графика и обозначенных двумя ярко красными узлами и четырьмя зелеными.

Если проанализировать графы и применить к данным все фильтры, описанные в предыдущем разделе, то можно выделить несколько кластеров или пар фирм, которые похожи на участников горизонтального картеля между поставщиками.

Для оценки качества графового анализа полезно также посмотреть на то, насколько результаты анализа графов совпадают с решениями ФАС о наличии картелей. Для этого был построен граф, в котором цвет узла зависит от факта выявления картеля в закупке, в которой участвовала фирма (Рисунок 26). Оранжевым обозначены компании (узлы), для которых доля закупок с картелем в общем числе рассмотренных закупок равняется 1. Фиолетовым обозначены узлы, для которых доля картелей в закупках равняется 0. И зеленым, где эта доля равна 0,5. Из рисунка видно, что почти все компании, которые много вместе участвовали в аукционах, признаны входящими в картель.



Источник: построено авторами/

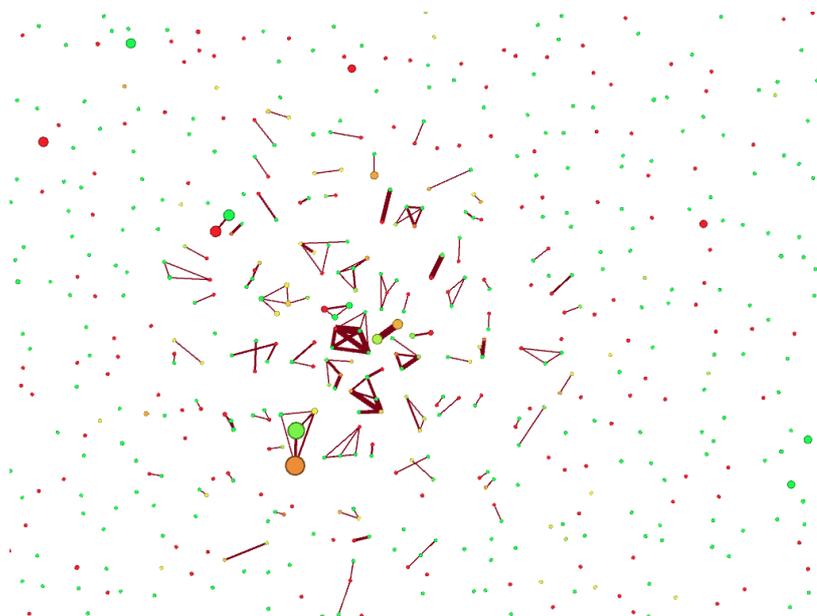
### Рисунок 26. Граф закупок медицинских препаратов и материалов для медицины - 3

Примечания:

1. Цвет узлов зависит от факта наличия выявленного картеля в закупке.
2. Отмечены ребра с весом больше 10.

Далее был построен граф по закупкам из сферы строительства. Этот граф состоит из 1281 узла и 1290 ребер. Вес ребер изменяется от 1 до 10. Это говорит о том, что почти все компании из нашей выборки участвовали только в одной закупке из рассмотренных.

Это можно увидеть на графах, где отмечен только ребра с весом больше 2 (Рисунок 27, Рисунок 28). На первом из графов, построенных по закупкам в строительной сфере, цвет узла зависит от доли побед компании (красный – 100% побед, зеленый – 0%), а размер узла зависит от совокупного объема НМЦК закупок с участием компании. В центре этого графа можно видеть группу из четырех компаний с очень плотными связями между собой. Из этих фирм только одна постоянно выигрывала, у остальных доля побед близка к 0. Аналогичная ситуация наблюдается и для большинства других групп или пар фирм. Есть только две пары, где распределение побед было более равномерным при относительно большом объеме закупок. Суммарная доля побед в таких небольших группах из двух-трех фирм равняется 100% или близка к этому значению, что является признаком сговора. Следующий важный показатель наличия картеля – это относительное снижение НМЦК. Результаты расчетов показывают, что в нашей выборке закупок в сфере строительства для всех ребер с весом больше 2, снижение цены не превышает 5%. Это говорит о низком уровне конкуренции на электронных аукционах, и одной из основных причин такой низкой конкуренции является наличие картеля. Ещё одна характеристика закупок, повышающая вероятность сговора – это большой суммарный объем НМЦК. Если цена каждой закупки небольшая, но в сумме они приносят большой потенциальный выигрыш компаниям, то это тоже увеличивает вероятность участия картеля в такой последовательности электронных аукционов.



Источник: построено авторами

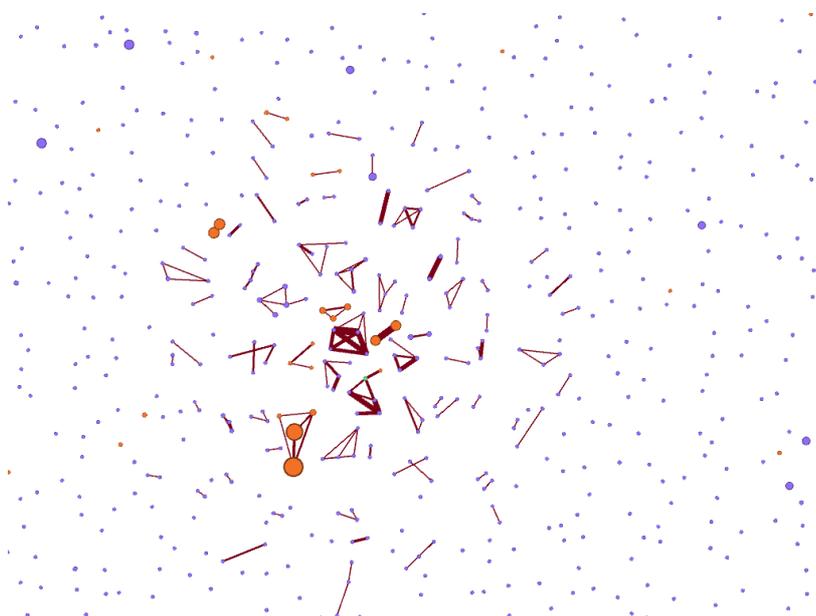
**Рисунок 27. Граф закупок в сфере строительства**

Примечание: Отмечены ребра с весом больше 2

Результаты подхода, основанного на анализе сетей и их графов, во многом совпадают с решениями ФАС по горизонтальным картелям в строительной отрасли (Рисунок 28). Антимонопольной службой выявлены сговоры между фирмами, если

- компании много раз вместе участвовали и не принимали участие по отдельности;
- суммарный объем закупок высокий;
- относительное снижение цены в ходе проведения аукциона низкое.

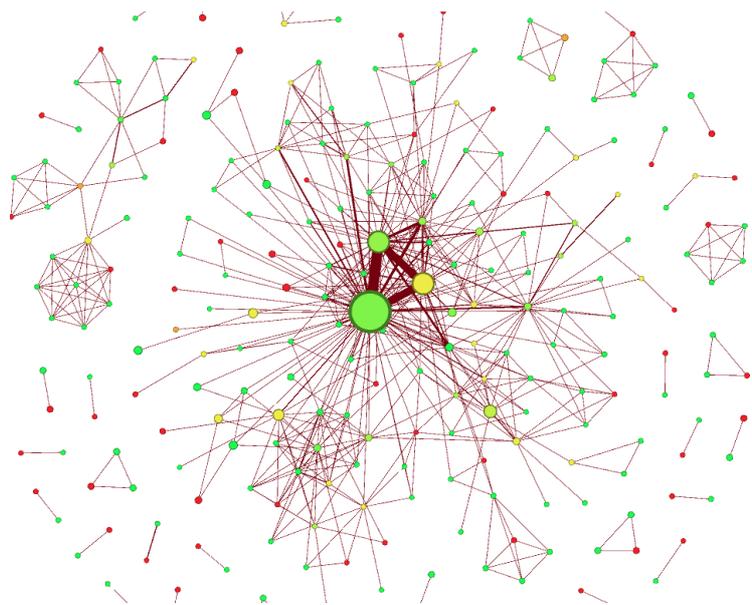
Из графика видно, что антимонопольная служба не выносит решений о наличии картеля для маленьких аукционов, даже если две фирмы участвовали в очень большом числе таких закупок.



Источник: построено авторами

**Рисунок 28. Граф закупок в сфере строительства, с узлами, цвет которых зависит от факта наличия выявленного картеля в закупке. Отмечены ребра с весом больше 2**

Последний рассмотренный вид экономической деятельности – производство бумаги и бумажных изделий. В закупках этой отрасли было выявлено мало картелей, 8 сговоров в 181 закупке. Граф, построенный по этим закупкам, состоит из 281 узла и 569 ребер (Рисунок 29). Вес ребер лежит в промежутке от 1 до 163.



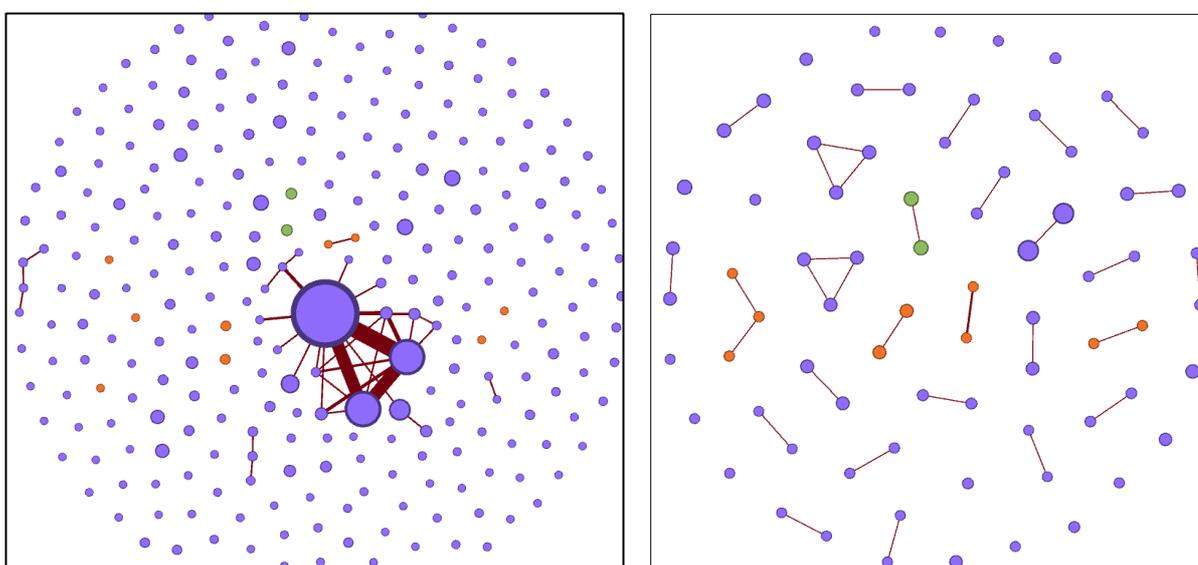
**Рисунок 29. Граф закупок бумаги и бумажных изделий. Отмечены все ребра**

Источник: построено авторами

При изучении графа сразу бросается в глаза наличие крупного центрального узла, имеющего очень высокую степень, т.е. от него отходит очень большое число ребер. Компания, обозначенная этим узлом, имеет самый большой суммарный объем НМЦК аукционов, значение этого параметра составляет 62,212 млн. рублей. Для этой фирмы степень узла, т.е. число электронных аукционов, в которых она принимала участие, равна 55. Максимальный вес ребра, соединяющего этот узел с другим, равен 20. Это может свидетельствовать о том, что узел отображает крупную фирму, выполняющую большое количество госзаказов в отрасли. Также наличие таких объемов поставок и плотных связей между поставщиками может являться признаком картеля. Однако, если посмотреть на решения ФАС, то можно заметить, что ни в одном из аукционов с участием этой фирмы не выявлен картель. Поэтому, если ориентироваться в первую очередь на вес ребер, то графовый анализ даст неправильные<sup>8</sup> выводы. Это отображено на левом графе из следующего рисунка (Рисунок 30). Кроме того, в ходе такой методики отбрасывания менее подозрительных фирм, упускаются сговоры, которые были осуществлены всего на одной закупке. Вероятно, правильнее будет смотреть сначала на среднее снижение цены в аукционах. Если снова обратиться к той фирме, которая участвовала в наибольшем числе закупок бумажных изделий, то можно видеть, что среднее снижение цены составляет 27,4%, а процент побед фирмы – 24%. Такое значительное падение цены в результате

<sup>8</sup> Здесь мы предполагаем, что решения ФАС обладают окончательным весом в решении того, где картель есть, а где он отсутствует. В общем случае такое допущение не является верным, но без обоснованных сведений о картеле из других источников мы вынуждены идти на его принятие.

проведения аукциона, а также невысокий процент побед компании, свидетельствует о честной конкуренции в закупках, и, в частности, добросовестному поведению рассматриваемой фирмы. При построении графа, где отображены только компании, для которых среднее относительное снижение цены во время аукциона оказалось меньше 10%, все ребра между участниками выявленных картелей сохраняются (Правый график, Рисунок 30). Также сохраняются ребра между некоторыми другими участниками закупок. Это может указывать на какое-то другое нарушение конкуренции, например, на вертикальный сговор между заказчиком и поставщиком, который тоже часто приводит к завышению цен на электронном аукционе.



**Рисунок 30. Граф закупок бумаги и бумажных изделий**

Примечания:

1. Цвет узла зависит от доли обнаруженных картелей.
2. На левом графике отображены все ребра с весом больше 2. Вес ребра принадлежит отрезку  $[2,20]$ . На правом графике отображены все ребра, но видны только узлы, для которых среднее относительное снижение цены во время аукциона оказалось меньше 10%

Источник: построено авторами

Для графа из каждой отрасли были рассчитаны основные статистические показатели (Таблица 13).

- Средняя степень графа. Самую большую среднюю степень имеет граф, построенный на основе закупок медицинских препаратов и материалов.

- Наши графы построены таким образом, что вес ребер показывает, сколько раз ребра связывают пару узлов (сколько было общих аукционов между парой участников). Взвешенная степень узла  $Wdeg(v_i)$  основана на количестве ребер для узла, но также учитывает вес каждого ребра. Это означает, что если вершина  $v_i$  имеет четыре ребра с весами 1, 1, 3, 2 соответственно, то  $deg(v_i)=4$ , а  $Wdeg(v_i)=7$ . Средняя взвешенная степень также самая высокая в медицинских препаратах, а самая низкая – в сфере строительства, где большинство поставщиков участвовали всего в одной закупке.
- Диаметр графа: наши графы являются несвязными, поэтому диаметр определен, как максимальный из диаметров его связных компонент.
- Связной компонентой графа называется максимальный связный подграф. В нашем случае с закупками в каждом из графов находится очень много маленьких связных компонент, состоящих из двух или трех вершин. При прочих равных условиях, большое количество связных компонент свидетельствует о том, что многие участники никак не связаны между собой, даже по какой-то длинной цепи ребер, а находятся в изолированных подграфах. В нашей выборке граф закупок бумажных изделий имеет наименьшее число связных компонент, но на такой результат повлияло маленькое число закупок, которые послужили основой для графа. Больше всего связных компонент оказалось в сфере строительства
- Модулярность ( $M$ ) показывает степень разбиения графа на отдельные кластеры, группы.
- Коэффициенты кластеризации измеряют среднюю вероятность того, что два соседа вершины сами являются соседями (мера плотности треугольников в сети). Средний коэффициент кластеризации равняется

$$C_{av}(G) = \frac{1}{n} \sum_{i \in V(G)} C_i$$

где  $n$  – количество вершин в графе  $G$ ;

$C_i$  – локальный коэффициент кластеризации, который определяется как число треугольников, в которых участвует вершина  $i$ , нормированное на максимально возможное число таких треугольников. Самый большой уровень средней кластеризации наблюдается в сфере строительства.

**Таблица 13. Основные статистические показатели графов закупок по выбранным отраслям**

показатель	Медицинские препараты и материалы	Строительство	Бумага
Количество закупок в выборке	1746	843	181
Количество связей (общая для двух поставщиков закупка)	3863	1679	730
Количество узлов (поставщики)	397	1281	282
Количество ребер	885	1290	569
Средняя степень	4,458	2,014	4,035
Средняя взвешенная степень	11,652	2,418	4,681
Диаметр графа	6	14	6
Плотность графа	0,011	0,02	0,014
Модулярность	0,731	0,982	0,633
Связные компоненты	83	383	50
Средний коэффициент кластеризации	0,601	0,792	0,754

Источник: расчеты авторов.

Итак, тестирование подхода к обнаружению горизонтальных картелей, основанного на анализе сетей и их графов, показало, что подход хорошо выявляет устойчивые картели, которые существовали и реализовывались больше чем на одном аукционе. Но метод не позволяет обнаружить сговоры, которые были реализованы только на одной госзакупке. Результаты показали, что важно учитывать долю общих закупок, суммарную долю побед возможных участников картеля, величину НМЦК электронных аукционов, а также относительное снижение цены в ходе проведения аукциона.

## ЗАКЛЮЧЕНИЕ

Борьба со сговором на торгах (bid-rigging) стала одним из важнейших направлений для антимонопольных органов во всем мире, о чем говорится в Докладе ОЭСР 2016 года [34]. Значительная часть экономических сделок, особенно в госсекторе, осуществляется через аукционы, поэтому проблема нечестных аукционов является крайне актуальной и важной. В нашей работе мы концентрируем внимание на горизонтальном сговоре, при котором фирмы-участники картеля выбирают заранее оговоренного победителя до даты аукциона.

На собранных данных были рассчитаны прогностически-классификационные алгоритмы для установления признаков сговора между участниками торгов в ходе госзакупок. Из использованных методов эконометрики и машинного обучения наибольшие характеристики точности и полноты (по 75%) продемонстрировал метод случайного леса Бреймана-Катлер, правильности прогнозирования - алгоритм градиентного бустинга (79,2%). При перебалансировке данных методом SMOTE показатель правильности прогнозирования вырос для градиентного бустинга и случайного леса до 84,0%.

В результате отраслевого анализа закупок была выявлена следующая закономерность: чем больше доля картелей в отрасли, тем меньше в среднем снижается цена на электронном аукционе. При добавлении переменных, характеризующих отрасли, в модель классификации показатель правильности предсказания вырос до 88,7% для метода случайного леса; мера  $f1$  в случайном лесе и градиентном бустинге приблизилась к 90%, точность - 88%, полнота - 88% и 86% соответственно.

Наиболее важными признаками, помогающими предсказывать наличие зафиксированного ФАС картеля, стали относительное снижение цены, медиана разницы во времени подачи заявок и НМЦК (для метода случайного леса) и относительное снижение цены, количество участников в протоколе 2 и медиана разницы во времени подачи заявок (для градиентного бустинга). Из отраслевых признаков самыми важными стали отрасли "медицинские препараты" и "медицинское оборудование".

Результаты модели могут служить основанием для проведения более детальной проверки аукциона сотрудниками ФАС. Подчеркнем: если классификатор не отнес закупку к недобросовестным, то это не значит, что в ней точно нет картеля, это говорит лишь о том, что по характеристикам она меньше похожа на те закупки, в которых антимонопольная служба обычно выявляет картели. Если же классификатор относит сделку к недобросовестным, то это означает, что с большой вероятностью в закупке

участвует картель, и он будет выявлен ФАС. Мы не можем говорить о точном наличии или отсутствии картеля, потому что истинное количество картелей при оценках методами *ex post* на имеющихся данных нам неизвестно, и эффективность классификаторов оценивается относительно совпадений с решениями ФАС о наличии или отсутствии горизонтального сговора на аукционе.

С использованием графо-теоретического подхода в работе были выделены подозрительные группировки фирм с учетом степени интенсивности взаимодействия на последовательности аукционов в конкретной отрасли, а полученные результаты визуализированы и показано их частичное соответствие решениям ФАС, в том числе в закупках медицинских препаратов и материалов для медицины, строительстве, закупках бумаги и бумажных изделий.

Тестирование подхода к обнаружению горизонтальных картелей, основанного на анализе сетей и их графов, показало, что подход хорошо выявляет устойчивые картели, которые существовали и реализовывались больше чем на одном аукционе. Но метод не позволяет обнаружить сговоры, которые были реализованы только на одной госзакупке. Результаты показали, что важно учитывать долю общих закупок, суммарную долю побед возможных участников картеля, величину НМЦК электронных аукционов, а также относительное снижение цены в ходе проведения аукциона.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

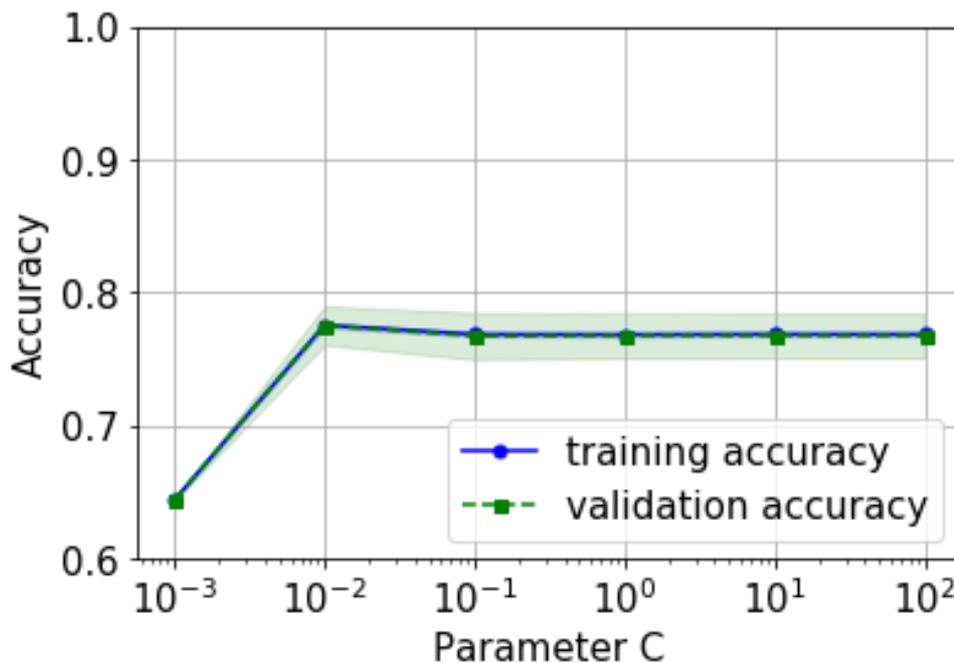
1. Abrantes-Metz R. M., Froeb L. M., Geweke J. F., and Taylor C. T. (2006), A variance screen for collusion, *International Journal of Industrial Organization*, 24, 467–486.
2. Abrantes-Metz R. M., Kraten M., Metz A. D., and Seow G. (2012). Libor manipulation // *Journal of Banking and Finance*, 36, 136–150.
3. Andreyanov P., Davidson A., and Korovkin V. (2016). Corruption vs Collusion: Evidence from Russian Procurement Auctions // Technical Report. mimeo: UCLA
4. Aryal G., and Gabrielli M. F. (2013). Testing for collusion in asymmetric first-price auctions // *International Journal of Industrial Organization*, vol 31, 26–35.
5. Bajari, P., and L. Ye (2003). Deciding between competition and collusion // *The Review of Economics and Statistics*, 85, 971–989.
6. Balsevich, A. and Podkolzina, E. A. (2014). Indicators of Corruption in Public Procurement: The Example of Russian Regions // Higher School of Economics Research Paper No. WP BRP 76/EC/2014. Available at SSRN: <https://ssrn.com/abstract=2530518> or <http://dx.doi.org/10.2139/ssrn.2530518>.
7. Boccaletti S., Latora V., Moreno Y., Chave M., Hwang D.-U. (2006). Complex networks: Structure and dynamics // *Physics Reports*, Volume 424, Issues 4–5, pp. 175-308.
8. Bolotova, Y., Connor, J., & Miller, D. (2008). The Impact of Collusion on Price Behavior: Empirical Results from Two Recent Cases // *International Journal of Industrial Organization*. 26. 1290-1307.
9. Borgatti S.P., Carley K., Krackhardt D. (2006). Robustness of centrality measures under conditions of imperfect data // *Social Networks* 28 (2), pp. 124–136.
10. Brandes U., (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology* 25, pp. 163–177.
11. Chawla N. V., Bowyer K. W., Hall L. O., Kegelmeyer W. P. (2002). SMOTE: Synthetic minority over-sampling technique // *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357.
12. Chawla, N. V., Japkowicz, N., & Drive, P. (2004). Editorial : Special issue on learning from imbalanced data sets // *ACM SIGKDD Explorations Newsletter*, 6 (1), 1–6 .
13. Connor, John M., *Global Price Fixing: Our Customers are the Enemy* // Boston: Kluwer Academic Publishers, 2001.
14. Fortunato S. (2010). Community detection in graphs // *Physics Reports*, Vol. 486, Issues 3–5, Pages 75-174.
15. Freeman L.C. (1978). Centrality in social networks: conceptual clarification // *Social Networks* 1, pp. 215–239.
16. Grout, P., and S. Sonderegger (2005). Predicting cartels, Discussion Paper, OFT.
17. Han, H., Wang, W. Y., & Mao, B. H. (2005, August). Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. // In *International conference on intelligent computing*. Springer, Berlin, Heidelberg. Pp. 878-887.
18. Harrington, J. E., *Handbook of Antitrust*, chap. 6 Detecting Cartel, 2006.
19. He, H.; Bai, Y.; Garcia, E. A.; Li, S. (2008). ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. // In *Proc. Int. Joint Conf. Neural Networks (IJCNN'08)*, pp. 1322-1328.
20. Huber M. and Imhof D. (2019). Machine Learning with Screens for Detecting Bid-Rigging Cartels // *International Journal of Industrial Organization*, Volume 65, Pages 277-301.
21. Ivanov D. I., Nesterov A. S. (2019). Identifying Bid Leakage In Procurement Auctions: Machine Learning Approach// [arXiv:1903.00261 \[econ.GN\]](https://arxiv.org/abs/1903.00261)
22. Imhof D. (2017). Simple screens to detect bid rigging // Working Papers SES 484, Faculty of Economics and Social Sciences, University of Fribourg (Switzerland).
23. Imhof D. (2018). Empirical Methods for Detecting Bid-rigging Cartels. *Economies and finances*. Université Bourgogne Franche-Comté, 2018. English.
24. Imhof D., Y. Karagoek, and S. Rutz (2017). Screening for bid rigging, does it work? // Working Paper UFC.
25. Ishii R. (2009). Collusion in repeated procurement auction: a study of paving market in Japan // *International Journal of Industrial Organization*, 27, 137–144.

26. Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125–1134).
27. Jakobsson M. (2007). Bid rigging in Swedish procurement auctions, Working Paper.
28. Jimenez J. L., and J. Perdiguero (2012). Does rigidity of price hide collusion? // Review of Industrial Organization, Vol. 41, No. 3, pp. 223-248.
29. Kawai K. and Nakabayashi J. (2014). Detecting Large-Scale Collusion in Procurement Auctions // Available at SSRN: <https://ssrn.com/abstract=2467175> or <http://dx.doi.org/10.2139/ssrn.2467175>.
30. Levenstein M. and Suslow V. (2001). Private International Cartels and Their Effect on Developing Countries // University of Massachusetts.
31. Morozov I., Podkolzina E. (2013). Collusion Detection in Procurement Auctions // SSRN Electronic Journal. 10.2139/ssrn.2221809
32. Newman M.E.J. (2001). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality // Physical Review E 64, 016132.
33. Newman M.E.J. (2003). The structure and function of complex networks. // SIAM Review, Vol. 45, pp. 167–256.
34. OECD (2016). Fighting bid rigging in public procurement: Report on implementing the OECD Recommendation // URL: <http://www.oecd.org/daf/competition/Fighting-bid-rigging-in-public-procurement-report-2016.pdf>.
35. Opsahl T., Agneessens F., Skvoretz J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths // Social Networks Volume 32, Issue 3, July 2010, Pages 245-251.
36. Pesendorfer, M. (2000). A study of collusion in first-price auction // The Review of Economic Studies, 67, 381–411.
37. Porter R. H., and J. D. Zona (1993). Detection of bid rigging in procurement auctions // The Journal of Political Economy, 101, 518–538.
38. Porter R. H., and Zona J. D. (1999). Ohio school milk markets: an analysis of bidding // RAND Journal of Economics, 30, 263–288.
39. Symeonidis G. (2003). In which industries is collusion more likely? Evidence from the UK // Journal of Industrial Economics 51, 45–74.
40. Tamimi I., El Kamili M. (2015). Literature survey on dynamic community detection and models of social networks // International Conference on Wireless Networks and Mobile Communications (WINCOM). – IEEE, pp. 1-5.
41. Wilhelm T., Kim J. (2008). What is a complex graph? // Physica A. Vol. 387, pp. 2637-2652.
42. Авдашева С. Б.; Шаститко А. Е.; Калмычкова Е. Н. (2007). Экономические основы антимонопольной политики: российская практика в контексте мирового опыта // Экономический журнал Высшей школы экономики, vol. 11, issue 2, 234-270.
43. Минфин (2018). Сводный аналитический отчет по результатам осуществления мониторинга закупок, товаров, услуг для обеспечения государственных и муниципальных нужд в соответствии с Федеральным законом от 05.04.2013 «О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд» № 44-ФЗ (далее – Закон о контрактной системе) по итогам 2017.
44. Минфин (2019). Сводный аналитический отчет по результатам осуществления мониторинга закупок, товаров, услуг для обеспечения государственных и муниципальных нужд в соответствии с Федеральным законом от 05.04.2013 № 44-ФЗ ... по итогам 2018 года.
45. Минэкономразвития (2016). Доклад о результатах мониторинга применения Федерального закона от 5 апреля 2013 г. № 44-ФЗ «О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд» в 2015 году // URL: <http://economy.gov.ru/minec/about/structure/depfks.old/44/20160217>
46. ФАС (2019 г). Доклад о состоянии конкуренции в Российской Федерации за 2018 год. // URL: <https://fas.gov.ru/documents/685806> 632 с.

## ПРИЛОЖЕНИЯ

Приложение 1. Подбор оптимальных параметров для моделей логистической регрессии и метода ближайших соседей

Для построения модели логистической регрессии необходимо выбрать уровень регуляризации  $C$ . При увеличении  $C$  регуляризация ослабляется, и классификатор начинает учитывать более сложные зависимости (т.е. при  $C \uparrow$  сложность модели  $\uparrow$ ). Чем меньше  $C$ , тем меньше модель штрафует за ошибки классификации, и при очень маленьких  $C$  будет проблема недообучения. Наоборот, при увеличении  $C$  в какой-то момент модель переобучится, т.е. будет слишком точно подгоняться под параметры отдельных наблюдений в обучающей выборке. Поэтому важно найти оптимальный уровень  $C$ . Для этого используем кривые правильности при обучении (Training curve) и валидации (проверке) (Validation curve). Мы применили кросс-валидацию по 10 блокам и построили несколько моделей с разными  $C$ . Результаты показывают, что наиболее оптимальным значением  $C$  является  $C=10^{-2}$  (Рисунок



31).

Примечание:

Accuracy: правильность предсказания.

Training accuracy: кривая правильности предсказания на обучающей выборке.

Validation accuracy: кривая правильности предсказания на тестовой выборке.

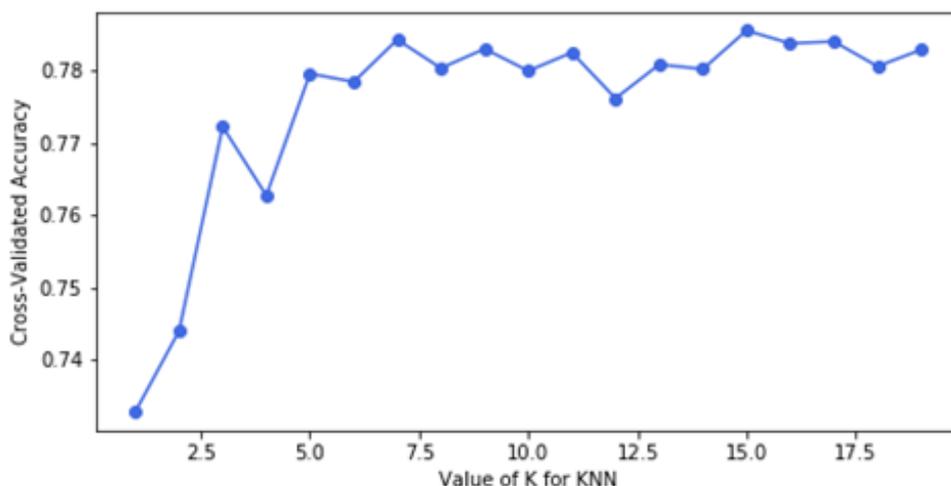
Источник: рассчитано авторами

### **Рисунок 31. Кривая проверки и кривая обучения.**

#### **Логистическая регрессия**

Для подбора оптимального  $k$  в модели ближайших соседей сначала была построена модель для  $k=5$  (значение, автоматически устанавливаемое пакетом). Далее мы предполагаем, что значение  $k$  не оптимально, и, изменив его, мы сможем улучшить качество модели.  $k$  напрямую влияет на эффективность kpp-классификатора. Если, например,  $k$  взять равным 1, то классификатор будет очень чувствителен к выбросам данных, его дисперсия будет очень высокой. И наоборот, если  $k$  будет очень большим, то это приведет к слишком сглаженным границам между классами, т.е.  $k$  должен быть достаточно мал, чтобы включить только близлежащие образцы, а не охватывать всю совокупность, в которой смешаны оба класса.

Если посчитать правильность предсказаний моделей, построенных по нескольким блокам кросс-валидации, то среднее значение от полученных показателей (CV\_accuracy) больше будет соответствовать тому, как будет вести себя модель на случайной выборке электронных аукционов. На рисунке ниже (Рисунок 32) представлен график кросс-валидационной правильности классификации, построенной по 10 блокам кросс-валидации, для диапазона  $k$  от 1 до 20.



Источник: рассчитано авторами

**Рисунок 32. График правильности классификации knn-модели в зависимости от значения k**

Для расчета k используем функцию поиска по решетке GridSearchCV, которая n число раз обучит нашу модель для каждого значения k из некоторого заданного диапазона. Далее на основе результатов GridSearchCV выберем такой k, чтобы точность модели была максимальной. GridSearchCV, построенная по 10 блокам кросс-валидации и для k [1,20] показывает, что оптимальным является k=15.

Приложение 2. Сравнение методов дополнения класса меньшинства, основанных на построении синтетических наблюдений

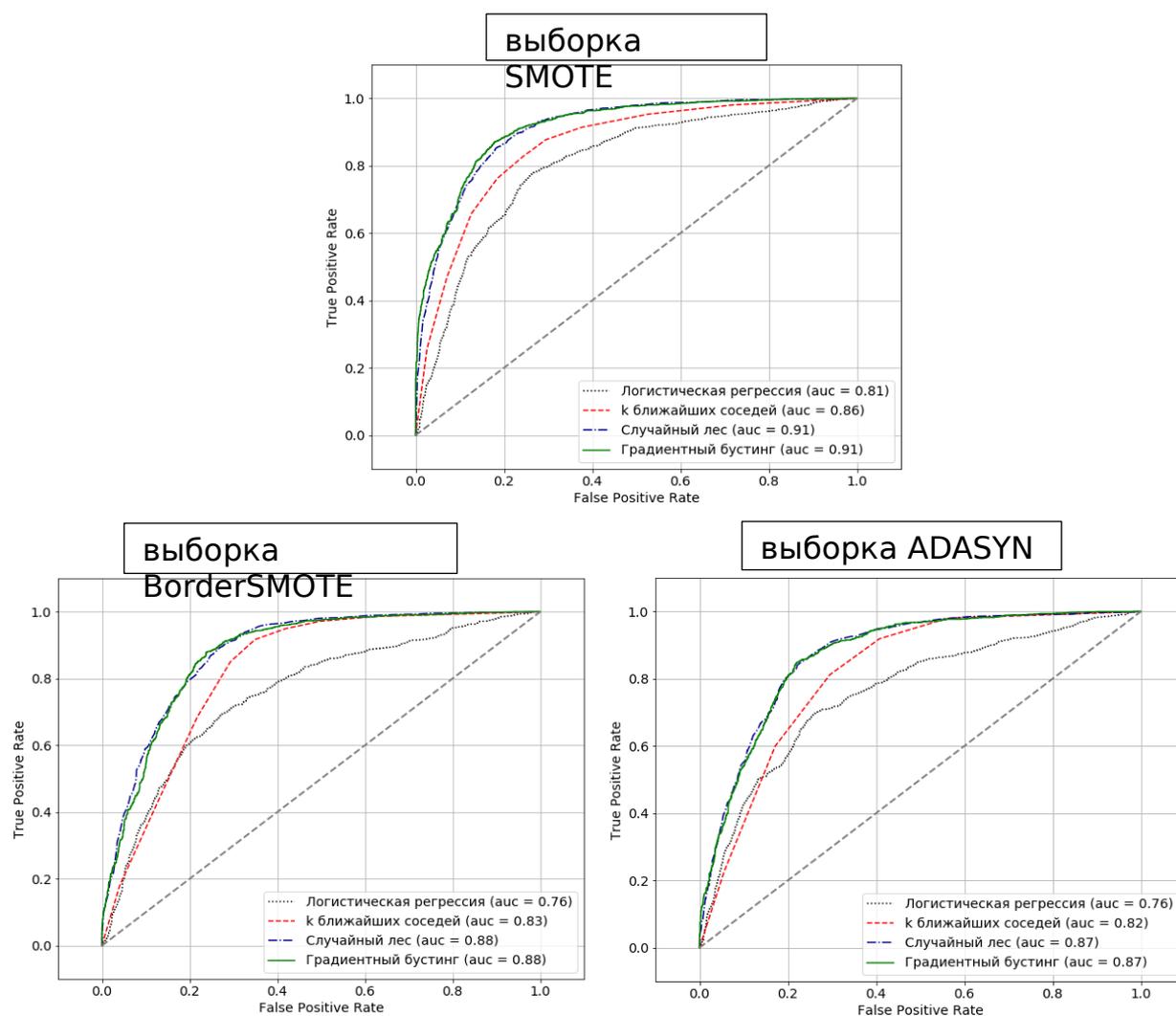
**Таблица 14. Точность и полнота моделей, показывающих наличие или отсутствие картеля на электронном аукционе, по данным, после применения BorderSMOTE и ADASYN**

		BorderSMOTE			ADASYN		
		Precision	recall	f1-score	Precision	recall	f1-score
Логистическая модель	нет картеля	0,77	0,52	0,62	0,77	0,57	0,65
	картель	0,64	0,84	0,72	0,62	0,81	0,70
k ближайших соседей	нет картеля	0,83	0,71	0,76	0,81	0,70	0,75
	картель	0,74	0,85	0,79	0,71	0,81	0,76
Случайный лес (100)	нет картеля	0,82	0,78	0,80	0,82	0,80	0,80

деревьев)								
	картель	0,79	0,83	0,81		0,78	0,80	0,79
Градиентный бустинг	нет картеля	0,86	0,76	0,81		0,85	0,78	0,81
	картель	0,79	0,87	0,83		0,77	0,85	0,81

Источник: рассчитано авторами

ROC-кривые для классификаторов, обученных на всей выборке с учетом 10 блоков кросс-валидации, также указывают на то, что перебалансировка данных с помощью обычного SMOTE способствует большему улучшению качества предсказаний (Таблица 14). Площадь под кривыми – AUC – равняется 0,91 для наших ансамблевых классификаторов в выборке SMOTE. Это означает, что с вероятностью 91% модель сможет различать положительный класс (сговора нет) и отрицательный класс (сговор есть). Для двух других выборок BorderSMOTE и ADASYN, значение AUC меньше 0,9 для всех классификаторов.



**Рисунок 33. Сравнение ROC-кривых для моделей, построенных на разных скорректированных выборках**

Источник: составлено авторами

**Таблица 15. Группы по видам экономической деятельности**

коды ОКП Д2	группа по видам деятельности	коды ОКПД2	группа по видам деятельности
01	сельское хозяйство	32	прочие готовые изделия
02	лесоводство	33	ремонт и монтаж машин и оборудования
03	рыболовство и рыбоводство	35	электроэнергия газ и кондиционирование воздуха
05-09	горнодобывающая	36-37	водоснабжение

	промышленность		
10	пищевая продукция	38-39	рекультивация и утилизация отходов
11	напитки	41-43	строительство
12	табачная продукция	45	торговля автотранспортом
13	текстиль	46-47	торговля оптовая и розничная
14	одежда	49-53	транспортные услуги
15	изделия из кожи	55-56	гостиницы и общественное питание
16	древесина	58-63	услуги в области информации и связи
17	бумага	64-66	услуги финансовые и страховые
18	услуги печатные и услуги по копированию	68	услуги по недвижимости
19	кокс и нефтепродукты	69	юридические и консультационные услуги
20	химические вещества и продукты	70-75	научная и инженерно- техническая деятельность
21	средства лекарственные и материалы для медицины	77-82	услуги административные и вспомогательные
22	резина и пластик	84	гос.управление и обеспечение военной безопасности
23	прочие продукты минеральные неметаллические	85	образование
24-25	металлы и металлическая продукция	86	здравоохранение
26	оборудование компьютерное и электронное оптическое	87-88	социальные услуги
27-28	прочее оборудование	90-91	искусство и культура
29-30	транспортные средства	93	спорт и отдых
31	мебель	94-99	прочие услуги

Источник: составлено авторами на основе кодов ОКПД2