

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

Инженерная школа ядерных технологий

Направление подготовки: Прикладная математика и информатика

Отделение экспериментальной физики

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Модернизация и применение современного математического подхода к оценке коррупционного риска в государственных закупках (на примере России)
УДК 519.233:328.185:339.33(47+57)

Студент

Группа	ФИО	Подпись	Дата
0BM71	Шибанов Максим Геннадьевич		

Руководитель

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Программист-аналитик ООО «ЭкоТомск»	А.Е. Барышева			

КОНСУЛЬТАНТЫ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Меньшикова Е.В.	Кандидат философских наук		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ст. преподаватель	Исаева Е.С.			

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Семенов М.Е.	Кандидат ф-м. наук		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ»

Студенту:

Группа	ФИО
0BM71	Шибанову Максиму Геннадьевичу

Школа	ИЯТШ	Отделение школы (НОЦ)	ОЭФ
Уровень образования	Магистр	Направление/специальность	01.04.02 Прикладная математика и информатика

Тема ВКР: «Модернизация и применение современного математического подхода к оценке коррупционного риска в государственных закупках (на примере России)»

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	1. Стоимость расходных материалов - 3076.5 руб. 2. Стоимость специального оборудования для научных (экспериментальных) работ – 62000 руб. 3. Затраты по основной заработной плате – 88808 руб. 4. Затраты по дополнительной заработной плате – 8881 руб. 5. Отчисления на социальные нужды – 26473 руб. 6. Затраты на электроэнергию – 1424 руб. 7. Накладные расходы – 29307 руб. 8. Бюджет затрат НИИ – 219970 руб.
2. Нормы и нормативы расходования ресурсов	1. Коэффициенты для расчета заработной платы
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	1. Отчисления во внебюджетные фонды (27,1%) 2. Расчет дополнительной заработной платы (10%) 3. Накладные расходы (30%)

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого и инновационного потенциала НИИ	1. Потенциальные потребители результатов исследования; 2. Анализ конкурентных технических решений; 3. SWOT – анализ 4. Оценка готовности проекта к коммерциализации
2. Разработка устава научно-технического проекта	1. Постановка цели, ожидаемых результатов проекта 2. Определение внутренних и внешних заинтересованных сторон проекта 3. Определение ограничений/допущений проекта
3. Планирование процесса управления НИИ: структура и график проведения, бюджет, риски и организация закупок	1. Структура работ в рамках научного исследования; 2. Определение трудоемкости выполнения работ и разработка графика проведения научного исследования; 3. Бюджет научно - технического исследования (НИИ)
4. Определение ресурсной, финансовой, экономической эффективности	1. Определение интегрального финансового показателя разработки; 2. Определение интегрального показателя ресурсоэффективности разработки; 3. Определение интегрального показателя эффективности

Перечень графического материала (с точным указанием обязательных чертежей):
1. «Портрет» потребителя результатов НТИ 2. Сегментирование рынка 3. Оценка конкурентоспособности технических решений 4. Матрица SWOT 5. График проведения и бюджет НТИ 6. Оценка ресурсной, финансовой и экономической эффективности НТИ 7. Потенциальные риски

Дата выдачи задания для раздела по линейному графику	
---	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН	Меньшикова Е.В.	Кандидат философских наук		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
0BM71	Шибанов Максим Геннадьевич		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
0BM71	Шибанову Максиму Геннадьевичу

Школа	ИЯТШ	Отделение школы (НОЦ)	ОЭФ
Уровень образования	Магистр	Направление/специальность	01.04.02 Прикладная математика и информатика

Тема ВКР:

«Модернизация и применение современного математического подхода к оценке коррупционного риска в государственных закупках (на примере России)»

Исходные данные к разделу «Социальная ответственность»:

1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	Объект исследования – коррупционный риск в сфере государственных закупок. Для его оценки исследован и реализован новый объективный подход, основанный на применении методов математической статистики.
--	--

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Правовые и организационные вопросы обеспечения безопасности: – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны.	– Нормативные документы: Трудовой кодекс РФ от 30.12.2001 N 197-ФЗ (ред. от 01.04.2019), ГК РФ (части I и II), Федеральный закон от 17 июля 1999 г. № 181-ФЗ, ГОСТ 12.2.032-78 ССБТ, ГОСТ 21889-76, ГОСТ 22269-76, ГОСТ Р 50923-96, ГОСТ Р ИСО 9241-1-2007, ГОСТ Р ИСО 9241-2-2009, ГОСТ Р ИСО 9241-4-2009, ГОСТ Р ИСО 9241-5-2009, ГОСТ Р ИСО 9241-7-2007, ГОСТ Р ИСО 9355-1-2009, ГОСТ Р ИСО 9355-2-2009, ГОСТ Р ИСО 9355-3-2010, СанПиН 2.2.2/2.4.1340-03.
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	– Анализ выявленных вредных факторов проектируемой производственной среды в следующей последовательности: • Отклонение показателей микроклимата; • Недостаточная освещенность рабочей зоны; • Опасное воздействие электромагнитного излучения; • Пожароопасность и вред здоровью от статического электричества; • Опасность поражения электрическим током; • Превышение допустимого уровня шума.
3. Экологическая безопасность:	– Анализ воздействия при работе на ПЭВМ на атмосферу, гидросферу, литосферу; – наличие отходов (бумага, картриджи, компьютеры и т. д.);

	– методы утилизации отходов.
4. Безопасность в чрезвычайных ситуациях:	– Приводятся возможные для Сибири ЧС; Возможные ЧС: морозы, диверсия, пожары – разрабатываются превентивные меры по предупреждению ЧС; – разработка действий в результате возникшей ЧС и мер по ликвидации её последствий.

Дата выдачи задания для раздела по линейному графику	
--	--

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ст. преподаватель	Исаева Е.С.			

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
ОВМ71	Шибанов Максим Геннадьевич		

РЕФЕРАТ

Пояснительная записка к магистерской диссертации выполнена на 103 страницах машинописного текста, содержит 31 таблицу, 13 рисунков, 34 формулы, 28 источников, 5 приложений.

Ключевые слова: Единая информационная система в сфере закупок, взвешенный индекс коррупционного риска, логистическая регрессия.

Объект исследования: коррупционный риск в сфере государственных закупок.

Целью исследования является исследование и реализация нового объективного подхода к оценке коррупционного риска в государственных закупках, основанного на применении методов математической статистики.

Методы проведения исследования: теоретические и практические.

Полученные результаты: составлена и реализована математическая модель коррупционного риска на языке программирования Python, проведена валидация модели с использованием статистических методов и на конкретных контрактах, предложены способы улучшения качества коэффициентов модели.

Оглавление

ВВЕДЕНИЕ	9
ОБЗОР ЛИТЕРАТУРЫ	11
1 ОБЪЕКТ И МЕТОДЫ ИССЛЕДОВАНИЯ.....	14
1.1 <i>Существующие подходы к определению коррупционного риска. Индексы коррупции.</i>	<i>14</i>
1.1.1 Индекс восприятия коррупции Transparency International.....	14
1.1.2 The Worldwide Governance Indicators	14
1.2 <i>Описание методологии для объективной оценка коррупционного риска.....</i>	<i>16</i>
1.2.1 Методика расчёта Corruption Risk Index (CRI)	16
1.2.2 Формулировка гипотезы о составляющих коррупционного риска	17
1.2.3 Определение порогов для непрерывных переменных.....	19
1.2.4 Методика расчёта весов элементарных показателей коррупции (ЭПК).....	21
1.3 <i>Применение логистической регрессии для оценки коррупционного риска</i>	<i>21</i>
1.4 <i>Статистические методы для валидации логистической регрессии</i>	<i>23</i>
1.4.1 Проверка значимости коэффициентов регрессии (тест Вальда).....	24
1.4.2 Проверка значимости уравнения регрессии. Критерий хи-квадрат отношения правдоподобия	24
1.4.3 Допущения логистической регрессии	25
1.4.3.1 Стационарность переменных и остатков	25
1.4.3.2 Уровень данных	31
1.4.3.3 Величина выборки.....	31
1.4.3.4 Автокорреляция в остатках.....	31
1.4.3.5 Обнаружение влиятельных наблюдений	34
1.4.3.6 Мультиколлинеарность независимых переменных	35
1.4.3.7 Тесты стабильности коэффициентов регрессии	36
2 ПРИМЕНЕНИЕ МЕТОДОЛОГИИ НА ПРИМЕРЕ РЫНКА ГОСУДАРСТВЕННЫХ ЗАКУПОК В РОССИИ	37
2.1 <i>Подготовка исходных данных для использования в модели</i>	<i>37</i>
2.2 <i>Расчёт порогов для непрерывных переменных.....</i>	<i>38</i>
2.3 <i>Оценка коэффициентов модели.....</i>	<i>39</i>
2.4 <i>Расчёт весов ЭПК, расчёт CRI</i>	<i>41</i>
2.5 <i>Агрегация CRI на различные уровни (контракт, поставщик, регион)</i>	<i>43</i>
2.6 <i>Валидация модели.....</i>	<i>45</i>
2.6.1 Проверка значимости коэффициентов регрессии (тест Вальда).....	46
2.6.2 Проверка значимости уравнения регрессии. Критерий хи-квадрат отношения правдоподобия	46
2.6.3 Проверка допущений логистической регрессии.....	46
2.6.3.1 Результаты тестов на стационарность	46
2.6.3.2 Уровень данных	47
2.6.3.3 Величина выборки.....	47
2.6.3.4 Автокорреляция в остатках.....	47
2.6.3.5 Обнаружение влиятельных наблюдений	48
2.6.3.6 Результаты теста на мультиколлинеарность	51
2.6.3.7 Тесты на стабильность	51
2.6.4 Валидация на конкретных данных.....	52

3	ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И РЕСУРСОСБЕРЕЖЕНИЕ	53
3.1	Предпроектный анализ.....	53
3.1.1	Потенциальные потребители результатов исследования.....	53
3.1.2	Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения.....	54
3.1.3	SWOT-анализ.....	55
3.1.4	Оценка готовности проекта к коммерциализации.....	57
3.1.5	Методы коммерциализации результатов научно-технического исследования	58
3.2	Инициация проекта.....	58
3.2.1	Цели и результат проекта	58
3.2.2	Организационная структура проекта.....	59
3.2.3	Ограничения и допущения проекта.....	60
3.3	Планирование управления научно-техническим проектом	60
3.3.1	План проекта.....	60
3.3.2	Бюджет научного исследования.....	63
3.3.3	Матрица ответственности.....	65
3.3.4	Реестр рисков проекта.....	66
3.4	Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования	67
3.5	Выводы	68
4	СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ	69
4.1	Введение	69
4.2	Правовые и организационные вопросы обеспечения безопасности	70
4.2.1	Специальные правовые нормы трудового законодательства	70
4.2.2	Перечень нормативных документов законодательства охраны труда.....	72
4.2.3	Организационные мероприятия при компоновке рабочей зоны	73
4.3	Производственная безопасность	74
4.3.1	Отклонение показателей микроклимата	76
4.3.2	Недостаточная освещенность рабочей зоны	77
4.3.3	Опасное воздействие электромагнитного излучения	78
4.3.4	Пожароопасность и вред здоровью от статического электричества.....	79
4.3.5	Опасность поражения электрическим током	79
4.3.6	Превышение допустимого уровня шума	81
4.4	Экологическая безопасность	81
4.5	Безопасность в чрезвычайных ситуациях	82
4.6	Выводы	85
	ЗАКЛЮЧЕНИЕ.....	86
	СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ	87
	ПРИЛОЖЕНИЕ А	90
	ПРИЛОЖЕНИЕ Б.....	91
	ПРИЛОЖЕНИЕ В.....	92
	ПРИЛОЖЕНИЕ Г.....	93
	ПРИЛОЖЕНИЕ Д	94

Введение

Одной из ключевых угроз национальной безопасности Российской Федерации является высокий уровень криминализации и коррупции в экономической сфере [1]. В частности, огромной проблемой является коррупция в государственных закупках. По информации Счётной палаты Российской Федерации объём госзакупок только в 2015 году составил 25 трлн. рублей. При этом только за 9 месяцев было выявлено нарушений более чем на 34 млрд. рублей [2].

Эмпирические исследования коррупции в настоящее время имеют глобальный характер. Свои индексы коррупции по большому массиву стран на регулярной основе рассчитывают Transparency International, The World Bank, Transparency International, Heritage Foundation, Freedom House. Характерной особенностью индексов этих и множества других организаций и исследовательских групп является то, что они используют «два основных способа измерения коррупции в обществе: измерение через восприятие и измерение через наличие коррупционного опыта» [3]. Очевидно, что оба этих способа достаточно субъективны.

Текущим трендом в исследованиях коррупции являются попытки преодоления субъективного характера ее измерений в соответствии с вышеназванными подходами и прежде всего методиками «World Bank's Control of Corruption» и «Transparency International's Corruption Perceptions Index» и создания более объективных измерителей коррупции [4]; [5]. Наиболее удачными в этом направлении являются работы М. Фазекаса и его коллег [6]; [7] по CEU Government Transparency Institute и связанному с ним Corruption Research Center (Будапешт), опирающихся на методы статистического анализа, Big Data и сетевой науки (Network Science).

Целью данной работы является исследование и реализация нового объективного подхода к оценке коррупционного риска в государственных

закупках, основанного на применении методов математической статистики. Для достижения данной цели были сформулированы следующие задачи:

1. Изучить существующие подходы к определению коррупционного риска в государственных закупках, изучить методику составления индексов коррупции.
2. Выбрать актуальный и методологически обоснованный подход к определению коррупционного риска. Модернизировать методологию для возможности ее применения в России.
3. Собрать, обработать, визуализировать входные данные.
4. Составить и реализовать математическую модель коррупционного риска на языке программирования Python.
5. Провести валидацию модели.

Обзор литературы

Большой вклад в изучение механизмов коррупции в госзакупках внесла работа Cole, S., & Tran, A [8], в частности глава «Evidence from the Firm: A New Approach to Understanding Corruption». В ней уделяется большое внимание распространённым схемам по отмыванию денег, описываются ключевые различия в государственных и частных закупках, освещается ненадежность социальных опросов, в случае, когда даже тщательные исследования не позволяют количественно определить наличие сговора. В итоге представлен подход, позволяющий выявлять взаимосвязь организаций.

Особый интерес в рамках данной работы представляет статья M. Fazekas [6], где представлен индикатор, полученный исключительно из объективных данных, определенный на микроуровне, например, отдельные транзакции, и имеющий возможность для последовательных временных сопоставлений внутри и между странами. Этот индикатор включает большое количество элементарных показателей, описанных в статье [9], таких как: подача единственной заявки, тип процедуры закупки, длина критериев отбора, длительность периода рассмотрения заявок, стоимость подачи документов на тендер, изменение условий конкурса и т. д. (Подробнее в Разделе 1.2.2).

В статье [10] рассмотрено применение данной методологии в Венгрии для измерения коррупционного риска государственных закупок в масштабе страны. В источнике [11] так же применен объективный метод измерения коррупции на данных в Италии.

Некоторые более общеизвестные и широко используемые индексы, ориентированные на измерение коррупции являются индекс восприятия коррупции (ИВК), который составляется международной организацией Transparency International [12], и Worldwide Governance Indicators (WGI) Всемирного банка [13]. Множество работ направлено на то, чтобы показать их несостоятельность и обосновать необходимость составления объективных индексов. Например, в [14] проведено сравнение этих двух индексов, и показано,

что ИВК не дает полной информации о коррупции в стране, потому что она отражает восприятие коррупции только в государственном секторе с точки зрения деловых людей и экспертов страны. Так же сравнение ИПЦ между годами может быть затруднено, поскольку рейтинг страны может легко измениться из-за изменений в списке анализируемых стран.

В статье [15] критически анализируется методология расчета ИВК. Обосновываются рекомендации по формированию концептуальных и методических основ национальной системы оценки уровня коррупции как универсального стандартизированного инструментария для проведения систематического мониторинга антикоррупционной деятельности в Российской Федерации.

В то же время в статье [16] рассматривается проблема измерения коррупции, отношение к коррупционным рейтингам. Основное внимание уделено анализу недостатков ИВК. Автор приходит к выводу, что очевидная политизированность проблемы коррупции (как в национальном, так и в глобальном контекстах) существенно затрудняет ее объективный научный анализ.

В работе [17] предложены некоторые объективные индексы коррупции, некоторые из которых опираются на политические составляющие. Проведен сравнительный анализ Corruption Risk Index (CRI) – индекс риска коррупции, который отражает коррупцию при формировании и распределении ренты, Political Influence Indicator (PII) - индикатор политического влияния, измеряющий влияние политической группы во власти на успешность государственных закупок компаний, и Political Control Indicator (PCI) - индикатор политического контроля, обозначающий прямой политический контроль подрядчиков по государственным закупкам.

Таким образом, исследователи часто связывают коррупцию с социальными аспектами общества, такими как сегрегация [18], межличностное доверие [19], гражданственность [20] и вовлечение сообщества [21]. Эти

подходы основаны на понимании того, что коррупция является коллективным результатом сообщества, сформированного взаимодействием между людьми [22], что позволяет предположить, что различия в социальном капитале, могут помочь объяснить сохранение коррупции и наблюдаемые различия в его уровнях.

Данное предположение является основополагающим в представляющей один из новых и современных подходов к оценке коррупции работе Johannes Watch [23]. Здесь для анализа связи между социальной структурой государства (на примере Венгрии) и распространенностью коррупции используются данные социальных сетей популярной онлайн-платформы. Рассматриваются социальные, экономические и политические факторы, в результате чего делается вывод, что поселения с фрагментированными социальными сетями, указывающие на избыток связующего социального капитала, имеют более высокий риск коррупции, а населенные пункты с более разнообразной внешней связью, предполагающие избыток связующего социального капитала, менее подвержены коррупции. Предполагается, что данный подход может дополнить показатели восприятия ИБК и WGI более объективными данными на микроуровне.

1 Объект и методы исследования

1.1 Существующие подходы к определению коррупционного риска. Индексы коррупции.

1.1.1 Индекс восприятия коррупции Transparency International

Transparency International's Corruption Perceptions Index или Индекс восприятия коррупции (ИВК) был разработан в 1995 г. в качестве составного показателя для измерения воспринимаемого уровня коррупции в государственном секторе в различных странах мира. В течение последних 20 лет велась работа по корректированию и усовершенствованию как источников, так и методологии для составления Индекса. В 2012 г. в методологию были внесены важные изменения, позволяющие сопоставить показатели во временной динамике, что до 2012 г. не представлялось возможным.

ИВК обладает рядом преимуществ: он имеет глобальный географический охват; использует среднее значение показателей от трёх (минимум) и до 13 (потенциальный максимум) различных источников; шкала ИВК диапазоном от 0 до 100 баллов может указывать на воспринимаемые уровни коррупции с большей неоднородностью, чем, к примеру, шкалы диапазоном от 1 до 7 или от 1 до 10; в то время как источники, используемые в ИВК, оценивают разные направления или проявления коррупции в государственном секторе, ИВК успешно объединяет различные аспекты коррупции в рамках одного индекса.

Источники и исследования, на основе которых составляется ИВК, предлагают своим респондентам ответить на вопросы из тщательно составленных, продуманных и отобранных анкет, что отражает лишь восприятие коррупции и является субъективной оценкой.

1.1.2 The Worldwide Governance Indicators

The Worldwide Governance Indicators (WGI) – шесть индикаторов, отражающих качество государственного управления в различных странах. Данный проект основан на исследованиях Института Всемирного банка (The

World Bank Institute) и Исследовательского отдела Всемирного банка (The Research Department of the World Bank), начатых в конце 1990-х гг. Даниэлем Кауфманом и Аартом Крээм. Большой вклад в развитие и усовершенствование WGI также внесли Пабло Зоидо и Массимо Мастрucci [24].

Критерии анализа вытекают из определения, предлагаемого Даниэлем Кауфманом и его коллегами, где государственное управление – это традиции и институты, при помощи которых осуществляется власть в стране, что включает: а) то, как формируются, контролируются и сменяются правительства; б) способность правительства формулировать и проводить рациональную политику; и в) уважение граждан и государственных служащих к институтам, регулирующим социально-экономические отношения между ними [25].

Итак, 6 индикаторов, соответствующих трём выделенным аспектам понятия государственного управления, образуют WGI.

1. Право голоса и подотчетность (Voice and Accountability) измеряет различные аспекты политических процессов, гражданских свобод и политических прав.
2. Политическая стабильность и отсутствие насилия (Political Stability and Absence of Violence/Terrorism) оценивает вероятность дестабилизации правительства и его свержения неконституционными и насильственными средствами;
3. Эффективность правительства (Government Effectiveness) отражает качество государственных услуг, уровень квалификации госслужащих и степень их независимости от политического давления, эффективность государственной политики;
4. Качество госрегулирования (Regulatory Quality) характеризует способность правительства формулировать и осуществлять рациональную политику, вести регулятивную деятельность, что положительно влияет на развитие частного сектора;

5. Верховенство закона (Rule of Law) измеряет степень уверенности граждан в эффективности отечественного законодательства и то, насколько они следуют установленным правилам.

6. Контроль коррупции (Control of Corruption) отражает отношение общества к коррупции, т. е. возможности использования служебного положения с целью извлечения частной выгоды [24].

Как видно, «Контроль коррупции» мало применим к сфере государственных закупок, так как измеряет степень использования публичной власти в личных целях.

1.2 Описание методологии для объективной оценка коррупционного риска

1.2.1 Методика расчёта Corruption Risk Index (CRI)

В данной магистерской работе в качестве критерия оценки коррупционного риска в государственных закупках был выбран индикатор CRI. Его преимуществами является то, что он рассчитывается исключительно на объективных данных, определен на микроуровне, например, отдельные транзакции, и предоставляет возможность последовательных временных сопоставлений внутри и между странами.

Согласно [6], для расчёта CRI, необходимо выполнение следующих задач:

- Получить структурированные данные из базы государственных закупок. Основная проблема здесь заключается в том, что данные не структурированы, содержат много бинарных и категориальных данных и представлены в различных форматах (scan, txt, pdf, excel, docx);
- Выбрать показатели для оценки риска коррупции контрактов. Экспертная оценка / Статистические тесты используются на этом этапе;
- Проанализировать влияние коррупционного риска на один контракт. Статистический анализ используется на этом этапе;
- Рассчитать индекс CRI для контракта и агрегировать его до необходимого уровня (компания, страна, регион).

$$CRI^t = \sum_j w_j * CI_j^t \quad (1)$$

$$\sum_j w_j = 1 \quad (2)$$

$$0 \leq CRI^t, CI_j^t \leq 1, \quad (3)$$

где CRI^t обозначает индекс риска коррупции транзакции t , CI_j^t представляет j – й элементарный индикатор коррупции, наблюдаемый в транзакции t , а w_j представляет вес элементарного индикатора коррупции j .

1.2.2 Формулировка гипотезы о составляющих коррупционного риска

Коррупция в государственных закупках подразумевает выбор определенного поставщика с предварительно завышенной ценой, которая покрывает так называемую коррупционную ренту - сумму денег, возвращаемую организатору конкурса после победы в этой компании.

Коррупция в публичных контрактах обычно включает ограничение конкуренции. Если коррумпированные бюрократы хотят заключить контракт с привилегированной фирмой, они должны каким-то образом исключить другие фирмы из участия в конкурсе на контракт. В данной работе выдвигается гипотеза о том, что это явление количественно определяется на уровне контрактов элементарными показателями коррупции (ЭПК), сигналами, которые можно извлечь из метаданных, свидетельствующих о том, что конкуренции за участие в конкурсе удалось избежать [26].

Обход конкуренции может осуществляться в трех основных формах [6], каждая из которых соответствует фазе процесса государственных закупок:

1. Ограничение набора участников: этап подачи заявок;
2. Несправедливая оценка участников торгов: этап оценки;
3. Изменение условий контракта после его заключения: этап поставки.

В качестве объективных данных, идентифицирующих наличие коррупции в конкретной транзакции, были выбраны элементарные показатели

коррупции CI_j^t , соответствующие определенному этапу реализации конкретной транзакции. В Таблице 1 приведены входные (input) индикаторы, используемые в данной работе для прогнозирования выходных (output) индикаторов, а также методы их расчета.

Таблица 1 – Индикаторы коррупции и методика их расчета [6]

Этап	Названия индикаторов	Расчет
submission (подача заявок)	Контракт с одной поданной заявкой	0=больше, чем одна заявка, 1=одна заявка
	Тип процедуры закупки	Тип процедуры – 1,2,3,4
	Длина критериев отбора	Количество критериев отбора минус среднее количество критериев отбора для данного рынка
	Длительность периода рассмотрения заявок	Дата окончания подачи заявок минус дата начала подачи заявок (в днях)
	Стоимость подачи документов на тендер	Стоимость подачи документов, деленная на стоимость контракта
	Изменение условий конкурса	0=условия конкурса не меняются, 1 = условия конкурса меняются
assessment (рассмотрение заявок)	Исключены все, кроме одного заявителя	0=как минимум два заявителя не исключены, 1=исключены все, кроме одного заявителя
	Вес неценовых критериев	Количество неценовых критериев относительно количества всех критериев
	Отмененная закупка запущена заново	0=контракт оплачен и не аннулирован, 1=оплаченный контракт аннулирован, но возобновлен
	Длительность периода принятия решения по контракту	Дата оглашения результатов конкурса - дата окончания подачи заявок (в рабочих днях)
delivery (осуществление контракта)	Изменение условий контракта (после завершения конкурса)	0=контракт не изменен после оглашения результатов, 1=изменен
	Продление контракта (после завершения конкурса)	Количество дней продления, деленное на общую продолжительность контракта
	Увеличение стоимости контракта (после завершения конкурса)	Изменение в стоимости контракта, деленное на первоначальную стоимость контракта
	Процент побед поставщика у одного заказчика за год	Сумма, полученная победителем от данного поставщика за год, деленная на сумму, отданную данным поставщиком победителям за год

«Контракт с одной поданной заявкой», «Исключены все, кроме одного заявителя», «Процент побед поставщика у одного заказчика за год» - выходные параметры модели. В работе считается, что любой индикатор сам по себе достаточно значимый, чтобы свидетельствовать о вероятности какой-либо сделки в ходе закупки. Однако выходные параметры, кроме того, ещё и зависят от входных, соответствующих их этапу. Для исследования этой зависимости использовалась логистическая регрессия в случае индикаторов «Контракт с одной поданной заявкой», «Исключены все, кроме одного заявителя», так как они бинарные, и линейная регрессия для «Процент побед поставщика у одного заказчика за год».

1.2.3 Определение порогов для непрерывных переменных

Элементарные индикаторы из Таблицы 1 принимают различные значения – некоторые описываются численно, как сумма контракта или относительное изменение в его стоимости, другие описываются бинарными или категориальными переменными, например, когда необходимое условие либо выполняется, либо нет. Третьи индикаторы представляют собой непрерывную функцию от времени, и для сопоставления необходимо их все привести к одному виду. В данной работе все индикаторы приводились к категориям (категориальные переменные). Поскольку для многих индикаторов их количественное увеличение не означало увеличение риска коррупции, для этих индикаторов был автоматизирован метод разделения на категории в зависимости от значений остатков логистической регрессии.

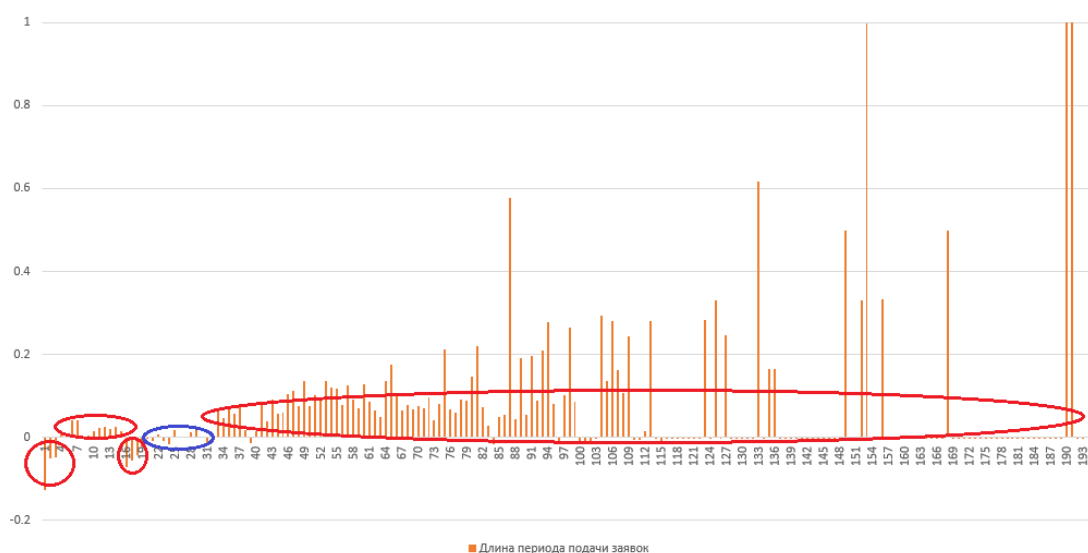


Рисунок 1 – Зависимость средних остатков логистической регрессии от соответствующего значения длины периода подачи заявок (показано разбиение на группы)

После расчёта логистической регрессии, где объясняемая переменная – один из выходных параметров текущей модели, а объясняющие – исследуемый некатегориальный индикатор и контрольные переменные, изучаются остатки этой логистической регрессии.

Если параметр имеет уникальные значения, как, например, для параметра длительности подачи заявок количество дней принимает уникальные значения: 1, 2, 3 дня и т. д., то для каждого уникального значения высчитывается среднее арифметическое соответствующих ему остатков регрессии. И затем можно построить зависимость средних остатков (mean residuals) от увеличения рассматриваемого параметра (пример – Рисунок 1).

Если значения параметра неуникальны, то весь диапазон принимаемых значений от минимального до максимального разбивается на 999 интервалов, и рассчитываются средние остатки уже для каждого интервала.

В зависимости от того, колеблются ли остатки вокруг нуля, либо имеют растущий/падающий тренд выше/ниже нуля, значения параметра разделяются на группы.

1.2.4 Методика расчёта весов элементарных показателей коррупции (ЭПК)

Для расчета весов ЭПК из формулы (2), коэффициенты независимых переменных регрессии проверяются на значимость с помощью статистических тестов, например, тест Вальда. Индикаторам коррупции, соответствующим значимым переменным ($p\text{-value} < 0.05$) присваиваются значения весов равные 1, незначимым - 0. Кроме того, для каждого индикатора разделенные группы перераспределяют веса внутри группы, например, для индикатора с 3-мя значимыми группами веса будут равняться: 1, 0.67, 0.33 по убыванию значимости.

Так же к параметрам, определяющим CRI, согласно формуле (6), относятся и выходные параметры «Контракт с одной поданной заявкой», «Исключены все, кроме одного заявителя», «Процент побед поставщика у одного заказчика за год», их веса равны 1. В итоге конечные значения весов разделяются на количество компонентов модели (считаются только сами параметры, а не их деления на фиктивные переменные).

Затем по формуле (1) рассчитывается CRI для каждого контракта с этими весами. Поскольку в данной работе реализован подход, аналогичный [6], то затем находится максимальный CRI среди всех контрактов, и итоговое значение весов получается делением текущих значений весов на это максимальное значение CRI. Затем CRI для каждого контракта пересчитывается уже с учетом новых значений весов. Таким образом происходит нормирование значений CRI.

1.3 Применение логистической регрессии для оценки коррупционного риска

При анализе данных часто встречаются задачи, где выходная переменная является категориальной, и тогда использование линейной регрессии затруднено. Поэтому в данной работе при поиске связей между набором входных переменных и категориальной выходной переменной (Таблица 1) используется логистическая регрессия, являющаяся методом бинарной классификации. Она

позволяет оценивать вероятность реализации (или не реализации) события в зависимости от значений некоторых независимых переменных. Линия логистической регрессии, в отличие от линейной, не является прямой.

Условное среднее для логической регрессии имеет вид:

$$\rho(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}, \quad (4)$$

где e – основание натурального логарифма; ρ - вероятность того, что произойдет интересующее событие; β_0, β_1 – коэффициенты логистической регрессии; x – значение независимой (входной) переменной.

Эту функцию называют логистической. Значения $\rho(x)$ изменяются в диапазоне от 0 до 1. Если предположить, что значение выходной переменной y , равное 1, рассматривается как успех, а значения 0 — как неуспех, то $\rho(x)$ можно интерпретировать как вероятность успеха, а $1 - \rho(x)$ – неуспеха.

Для оценки коэффициентов логистической регрессии метод наименьших квадратов не применим. Поэтому ее коэффициенты оцениваются на основе метода максимального правдоподобия. Логарифмическая функция правдоподобия имеет вид:

$$L(\beta|x) = \sum_{i=1}^n \{y_i \ln[\rho(x_i)] + (1 - y_i) \ln[1 - \rho(x_i)]\}. \quad (5)$$

Структуру логистической регрессии касательно настоящей работы можно описать следующим образом:

$$\Pr(single\ bidder_i = 1) = \frac{1}{1 + e^{-Z_i}} \quad (6)$$

$$Z_i = \beta_0 + \beta_{1j} S_{ij} + \beta_{2k} A_{ik} + \beta_{3l} D_{il} + \beta_{4m} C_{im} + \varepsilon_i, \quad (7)$$

где параметр $single\ bidder_i$ («подача единственной заявки») равен 1, если на i -ый контракт была подана только одна заявка и 0 – если больше; Z_i - логит от вероятности контракта иметь единственную заявку; β_0 – константа регрессии; S_{ij} – матрица j входных индикаторов на этапе подачи заявки для i -го контракта

(например, длина периода подачи заявки); A_{ik} – матрица k входных индикаторов на этапе рассмотрения заявки для i -го контракта (например, вес неценовых критериев); D_{il} – матрица l входных индикаторов на этапе поставки для i -го контракта (например, продление контракта); C_{im} – матрица m контрольных переменных для i -го контракта (например, количество конкурентов на рынке); ε_i – ошибка регрессии; и β_{1j} , β_{2k} , β_{3l} и β_{4m} являются векторами коэффициентов для объясняющих и контрольных переменных.

Аналогичная модель логистической регрессии применяется относительно выходного параметра «Исключены все, кроме одного заявителя», однако относительно последнего выходного параметра «Процент побед» строится линейная регрессия, так как он не является бинарным, а принимает значения от 0 до 100%:

$$Y_i = \beta_0 + \beta_{1j}S_{ij} + \beta_{2k}A_{ik} + \beta_{3l}D_{il} + \beta_{4m}C_{im} + \varepsilon_i \quad (8)$$

где Y_i представляет winner's share within issuer's contracts («Процент побед»); остальные обозначения те же.

Результаты данных регрессий позволяют перейти к определению весов элементарных индикаторов для расчёта CRI.

1.4 Статистические методы для валидации логистической регрессии

Для проверки работоспособности модели количественно при валидации использовали набор статистических тестов. Эти статистические тесты широко используются в отраслевой практике:

1. Значимость коэффициентов логистической регрессии: критерий Вальда;
2. Значимость уравнения модели логистической регрессии: критерий хи-квадрат отношения правдоподобия;
3. Проверка допущений логистической регрессии;
4. Стабильность коэффициентов логистической регрессии: тесты forward- и backward - стабильности, тест Рой-Чоу.

В следующих подразделах все эти тесты описаны более подробно.

1.4.1 Проверка значимости коэффициентов регрессии (тест Вальда)

Тест Вальда используется для проверки истинного значения параметров на основе оценки выборки. Согласно статистическому критерию Вальда, оценка максимального правдоподобия $\hat{\theta}$ интересующего параметра θ сравнивается с предложенным значением θ_0 (принимается равным 0). Из асимптотической теории оценки максимального правдоподобия (MLE) эта разница, деленная на стандартное отклонение, даст стандартное нормальное распределение со средним 0 и стандартным отклонением 1:

$$W = \frac{\hat{\theta} - \theta_0}{\text{se}(\hat{\theta})} \sim N(0,1) \quad (9)$$

Считается, что входная переменная прошла тест Вальда, если она была значимой на уровне 5%.

1.4.2 Проверка значимости уравнения регрессии. Критерий хи-квадрат отношения правдоподобия

Критерий хи-квадрат отношения правдоподобия используется для оценки качества логистической регрессии. Он рассчитывается путем сопоставления модели, которая не имеет независимых переменных (нулевая модель, которая имеет только константу) с моделью, которая имеет независимые переменные.

Это отношение правдоподобия или, что эквивалентно, его логарифм, может затем использоваться для вычисления p-value или сравниваться с критическим значением, чтобы решить, следует ли отклонить нулевую модель. Уравнение для логарифмической вероятности:

$$\ln L(\beta|y, X) = \sum_{i=1}^N \ln p_i, \quad (10)$$

где p_i - вероятность наблюдения фактического значения y :

$$p_i = \begin{cases} \Pr(y_i = 1|x_i), & \text{если наблюдается } y_i = 1, \\ 1 - \Pr(y_i = 1|x_i), & \text{если наблюдается } y_i = 0. \end{cases} \quad (11)$$

Статистика логарифмического отношения правдоподобия подчиняется распределению хи-квадрат со степенью свободы, равной разнице в размерности двух моделей.

1.4.3 Допущения логистической регрессии

Допущения логистической регрессии следующие:

1. Уровень данных: зависимая переменная должна иметь дихотомический характер для бинарной регрессии.
2. Величина выборки: для оценки логистической регрессии используется метод максимального правдоподобия, поэтому требуется большой размер выборки.
3. Ошибка: ошибки считаются независимым.
4. Линейность: линейная зависимость не предполагается, однако независимые переменные должны быть линейно связаны с логарифмическими коэффициентами.
5. Нет влиятельных значений: предполагается, что в непрерывных предикторах нет экстремальных значений или выбросов.
6. Нет мультиколлинеарности: предполагается, что между независимыми переменными нет мультиколлинеарности, либо оно небольшая.

Кроме того, для моделей временных рядов следует оценить стационарность временных рядов.

Ниже приведены все эти предположения и их результаты описаны более подробно.

1.4.3.1 Стационарность переменных и остатков

Для модели временных рядов все зависимые переменные, независимые переменные и остатки должны быть проверены на стационарность с помощью теста Augmented Dickey-Fuller (ADF), теста Phillips-Perron (PP) и теста Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Каждая переменная должна пройти

как минимум 2 стационарных теста. Для каждого стационарного теста на зависимую переменную и независимые переменные оцениваются:

1. Стационарность нулевого среднего (только для остатков);
2. Стационарность единичного среднего;
3. Стационарность тренда.

Стационарность тренда означает, что переменная содержит член тренда (βt), и этот процесс является стационарным после де-тренда. Для остатков проводится только один тест на стационарность единичного среднего. Для остатков необходимо использовать стационарный тест без константы, потому что остатки должны иметь нулевое среднее значение по определению.

Могут произойти следующие ситуации:

- Если переменная прошла единичное среднее стационарное тестирование, то эта переменная может быть непосредственно использована в модели.
- Если переменная прошла только стационарный тест тренда, то необходимо привести обоснование того, почему эту переменную можно использовать в уравнении.
- Если уравнение содержит нестационарную переменную, но остатки проходят стационарный тест, это указывает на то, что переменные объединены, то есть существует некоторая линейная комбинация набора переменных, которая является стационарной. В этом случае переменные все еще могут использоваться для моделирования при условии, что для использования переменных есть бизнес-обоснование. В противном случае модель требуется реконструировать.
- Если уравнение содержит нестационарную переменную, а остатки также не прошли тест на стационарность, то модель должна быть реконструирована.

Рассмотрим процесс авторегрессии:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t. \quad (12)$$

Для удобства предположим, что $y_0 = 0$. Если $m = 1$ является корнем характеристического уравнения

$$m^p - m^{p-1}a_1 - m^{p-2}a_2 - \dots - a_p = 0, \quad (13)$$

то процесс имеет единичный корень. Например, AR(1) процесс $y_t = a_1 y_{t-1} + \epsilon_t$ имеет единичный корень, когда $a_1 = 1$.

Для теста ADF и PP нулевая гипотеза состоит в том, что ряд данных имеет единичный корень, то есть ряд данных является нестационарным; в то время как нулевая гипотеза для теста KPSS состоит в том, что ряд данных не имеет единичного корня, то есть ряд данных является стационарным. Поэтому желательно иметь небольшие значения p из теста ADF и PP и большие значения p из теста KPSS.

Пороги для ADF, PP и KPSS тестов указаны в Таблице 2.

Таблица 2 – Пороги для стационарных тестов

Тест	Порог
ADF	< 0.05
PP	< 0.05
KPSS	> 0.05
ADF-GLS	Тестовая статистика > 5ct Critical value
BVR-test	<0.05

1.4.3.1.1 Описание теста Augmented Dickey-Fuller (ADF)

ADF-тест - это тест на единичный корень для временных рядов, что эквивалентно проверке необходимого условия стационарности в слабой форме в предположении модели. Для одиночного среднего тест ADF:

$$\Delta y_t = (1 - \phi)y_{t-1} + \mu(1 - \phi) + \sum_{j=1}^N \psi_j \Delta y_{t-j} + \epsilon_t. \quad (14)$$

ADF для тренда:

$$\Delta y_t = \beta t + (1 - \phi)y_{t-1} + \mu(1 - \phi) + \sum_{j=1}^N \psi_j \Delta y_{t-j} + \epsilon_t. \quad (15)$$

N запаздывающих членов, то есть Δy_{t-j} , используются для аппроксимации структуры ARMA (возможной ARMA с неизвестными порядками) ошибок, и значение N устанавливается таким образом, чтобы ошибка ϵ_t была некоррелированной. Отклонения также считаются гомоскедастичными.

Максимальное количество N можно рассчитать с помощью:

$$P_{\max} = (N_{\text{obs}} - 1)^{\frac{1}{3}}. \quad (16)$$

1.4.3.1.2 Описание теста Augmented Dickey-Fuller с применением GLS (ADF-GLS)

Для рядов с детерминированными компонентами в форме постоянного или линейного тренда был разработан асимптотически точечный оптимальный тест для обнаружения единичного корня. Эта процедура тестирования доминирует над другими существующими единичными корневыми тестами с точки зрения мощности. Он локально осуществляет де-тренд рядов данных для эффективной оценки детерминированных параметров ряда и использует преобразованные данные для выполнения обычного единичного корневого теста ADF. Эта процедура помогает удалить средние и линейные тренды для рядов, которые близки по свойствам к нестационарным.

Рассмотрим простую модель временного ряда

$$y_t = d_t + u_t \quad (17)$$

с

$$u_t = \rho u_{t-1} + e_t, \quad (18)$$

где d_t является детерминированной частью, а u_t является стохастической частью y_t .

Когда истинное значение ρ близко к 1, оценка модели, то есть d_t , создаст проблемы эффективности, поскольку ряд y_t будет близким к нестационарному. В этом случае тестирование признаков стационарности данного временного ряда также будет сопряжено с общими статистическими проблемами. Для преодоления таких проблем предлагается локально разделять временные ряды.

Рассмотрим случай, когда близость к 1 для параметра авторегрессии моделируется как

$$\rho = 1 - \frac{c}{T}, \quad (19)$$

где T - количество наблюдений. Теперь рассмотрим фильтрацию ряда с помощью

$$1 - \frac{\bar{c}}{T} L, \quad (20)$$

где L является стандартным оператором запаздывания, т. е.

$$\bar{y}_t = y_t - \left(\frac{\bar{c}}{T} \right) y_{t-1}. \quad (21)$$

Работа с $\bar{y}_t$ приведет к увеличению мощности при тестировании стационарных функций y_t с использованием расширенного теста Дики-Фуллера. Это точечный оптимальный тест, для которого $\bar{c}$ настроен таким образом, чтобы тест имел 50-процентную мощность, когда альтернатива характеризуется (19) при $c = \bar{c}$.

В зависимости от спецификации d_t , $\bar{c}$ будет принимать разные значения.

1.4.3.1.3 Описание теста Phillips-Perron (PP)

PP-тест - это тест для единичного корня для временных рядов, что эквивалентно проверке необходимого условия стационарности в слабой форме в предположении модели

$$y_t = \alpha + \beta t + \phi y_{t-1} + \epsilon_t. \quad (22)$$

В отличие от теста ADF, тест PP использует непараметрический подход для корректировки автокорреляции в тестируемых временных рядах. При

определенных условиях тест РР способен справляться с гетероскедастичностью и особенно эффективен при наличии ошибки скользящего среднего. Нулевая гипотеза теста РР состоит в том, что временной ряд имеет единичный корень (нестационарный), т. е. $\phi = 1$. βt используется только в тесте тренда на стационарность. Расчет тестовой статистики зависит от спецификации параметра усечения N , который используется при оценке долгосрочной дисперсии. N устанавливается равным $N = (4 * (n/100)^{0.25})$ и округляется до целых.

1.4.3.1.4 Описание теста Kwiatkowski-Phillips-Schmidt-Shin Test (KPSS)

KPSS-тест - это тест для единичного корня для временных рядов. Это эквивалентно проверке необходимого условия стационарности в слабой форме в предположении модели

$$y_t = \alpha + \beta t + \mu_t + u_t, \mu_t = \mu_{t-1} + \epsilon_t, \epsilon_t \sim N(0, \sigma_\epsilon^2), \quad (23)$$

где u_t является стационарным, а μ_t - чисто случайное блуждание с инновацией ϵ_t и дисперсией σ_ϵ^2 . Следовательно, нулевая гипотеза о том, что ряд y_t стационарный может быть сформулирована как $H_0: \sigma_\epsilon^2 = 0$. βt используется только в тесте тренда на стационарность. Статистика KPSS является статистикой множителя Лагранжа для $\sigma_\epsilon^2 = 0$ против $\sigma_\epsilon^2 > 0$, как указано в следующей формуле:

$$KPSS = \frac{[T^{-2} \sum_{t=1}^T (\sum_{i=1}^t u_i)^2]}{\hat{\lambda}^2}, \quad (24)$$

$\hat{\lambda}^2$ - долгосрочная дисперсия u_t с использованием $\hat{u}_t$. Статистика KPSS сходится к функции стандартного броуновского движения, которая зависит от формы детерминированных трендов (одной константы или временного тренда). Как и в тесте РР, оценка долгосрочного параметра дисперсии $\hat{\lambda}^2$ зависит от параметра усечения N .

N устанавливается равным $N = (4 * (n/100)^{0.25})$ и округляется до целых.

1.4.3.2 Уровень данных

Зависимая переменная должна иметь дихотомический характер для логистической регрессии. Для бинарной логистической регрессии зависимые переменные делятся на две категории. Обычно прогнозируют более высокую категорию (предполагаемую как 1), беря более низкую контрольную категорию (предполагаемую как 0).

1.4.3.3 Величина выборки

Для множественной логистической регрессии Peduzzi, Concato, Kemper, Holford, & Feinstein (1996) в работе [27] предложили рекомендации по минимальному размеру выборки. Пусть p будет наименьшей из пропорций отрицательных или положительных случаев в выборке, а k будет количеством независимых переменных, тогда минимальное количество наблюдений:

$$N = 10 \frac{k}{p}. \quad (25)$$

1.4.3.4 Автокорреляция в остатках

Логистическая регрессия предполагает, что остатки не имеют автокорреляций. Если допущение нарушается, то заключение о значении независимых переменных может быть ненадежным. Существует несколько способов оценки автокорреляции.

1.4.3.4.1 Графики ACF и PACF

Графики ACF и PACF обычно используются для визуализации автокорреляций и частичных автокорреляций при заданных лагах. Он также может служить полезным инструментом для определения параметров используемой в будущем модели (AR, MA или ARMA) и подходящего лага для остатков модели. Графики также демонстрируют результат теста, согласно которому (частичная) автокорреляция равна нулю при данном запаздывании. Критическая область теста определяется как $\pm \frac{2}{\sqrt{n}}$, где n - количество точек

данных. Если выборочная (частичная) автокорреляция находится в пределах критической области, то делается вывод о том, что при данном лаге (частичной) автокорреляции нет.

Если в графике PACF разностного ряда есть резкий скачок, или автокорреляция lag-1 положительна, то возможно добавление члена AR в модель. Лаг, на котором PACF обрывается, является параметром AR.

Если в графике ACF разностного ряда есть резкий скачок, или автокорреляция lag-1 отрицательна, то возможно добавление в модель члена MA. Лаг, на котором ACF обрывается, является параметром MA.

1.4.3.4.2 Описание теста Durbin-Watson

Тест Дарбина—Уотсона (DW-test) – это тест, используемый для обнаружения наличия автокорреляции остатков первого порядка в регрессионной модели. Подробная формула для расчета статистики испытаний DW выглядит следующим образом:

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}, \quad (26)$$

где T – количество наблюдений, а e_t остатки регрессии.

Статистика, близкая к 2, указывает на то, что в остатках нет автокорреляции.

Тест DW выполняется как для положительного, так и для отрицательного автокорреляционного анализа. В качестве результата теста принимается и самое низкое значение p , и статистику DW.

Нулевая гипотеза состоит в том, что в серии испытаний нет автокорреляции, поэтому критерий прохождения равен $p > 0,05$.

1.4.3.4.3 Описание теста Breusch-Godfrey

LM-тест Бройша — Годфри (Breusch-Godfrey serial correlation LM-test) на автокорреляцию - это тест на автокорреляцию ошибок в регрессионной модели.

Для регрессионной модели: $y_t = \sum_{i=1}^N \beta_i * x_{t,i} + \epsilon$, тест проверяет, является ли $\epsilon_t = \sum_{i=1}^N \beta_i * x_{t,i} + \sum_{j=1}^p \rho_j * \epsilon_{t-j} + \epsilon_t$. Нулевая гипотеза: $H_0: \rho_i = 0$ для всех i . Статистика теста — это R^2 модели остаточной регрессии, и $nR^2 \sim \chi_p^2$. Для годовых данных критерий проверяется с лагом = 1, для квартальных данных с лагом = 4, и с лагом 12 для ежемесячных данных. Для этой модели взят лаг = 1, так как данные не являются временным рядом.

Нулевая гипотеза состоит в том, что в серии испытаний нет автокорреляции, поэтому критерий прохождения равен $p > 0,05$.

1.4.3.4.4 Описание теста Ljung-Box

Q-тест Льюнг — Бокса (The Ljung-Box test) применяется для проверки наличия автокорреляции при заданном лаге. Вместо проверки случайности при каждом отдельном лаге, он тестирует «общую» случайность, основанную на общем количестве лагов. В частности, нулевая гипотеза состоит в том, что остатки регрессии не автокоррелируемы до некоторого лага h . Таким образом, критерий прохождения равен $p > 0,05$. Статистика теста:

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{(n-k)}, \quad (27)$$

где N — размер выборки; $\hat{\rho}_k^2$ — автокорреляция выборки при лаге k ; h количество тестируемых лагов.

Статистика теста соответствует распределению хи-квадрат со степенью свободы $h - p - q$. h — количество тестируемых лагов; p и q — параметры AR и MA, используемые в модели ARMA(p, q).

Аналогично тесту Бройша — Годфри, для годовых данных критерий проверяется с лагом = 1, для квартальных данных с лагом = 4, и с лагом 12 для ежемесячных данных. Для этой модели взят лаг = 1, так как данные не являются временным рядом.

1.4.3.5 Обнаружение влиятельных наблюдений

Влиятельные значения — это крайне обособленные точки данных, которые могут изменить качество модели логистической регрессии. Моделирование логистической регрессии предполагает отсутствие влиятельных точек во входных данных.

На данном этапе для отдельных переменных строится диаграмма размаха (Box Plot) - график, использующийся в описательной статистике, компактно изображающий одномерное распределение вероятностей. Он в удобной форме показывает медиану, нижний и верхний квартили, минимальное и максимальное значение выборки и выбросы.

1.4.3.5.1 Рычаг

Рычаг наблюдения можно рассчитать через hat-value из проекционной матрицы. Проекционная матрица рассчитывается по формуле:

$$H = X(X'X)^{-1}X'. \quad (28)$$

Тогда рычаг i -го наблюдения:

$$h_i = h_i^{hi} = \sum_{j=1}^n h_{ij} = h_{ii}, \quad (29)$$

где H симметрична. Если h_i велико, то i -е наблюдение оказывает большее влияние на коэффициент, чем другие.

1.4.3.5.2 DFFITS

DFFIT - мера, показывающая влияние точки на регрессионную модель. Определяется как изменение предсказанного значения для i -точки, когда эта точка исключена из модели:

$$\text{DFFITS} = \frac{\hat{y}_i - \hat{y}_{i(i)}}{s_{(i)}\sqrt{h_{ii}}}. \quad (30)$$

$\hat{y}_i$ и $\hat{y}_{i(i)}$ — это прогноз для точки i с и без включения этой точки в регрессию, $s_{(i)}$ стандартная ошибка, оцененная без учета рассматриваемой

точки, и h_{ii} это рычаг для точки. Большее значение указывает на то, удалив данную точку данных, модель теряет прогнозную силу, и эта точка считается влиятельной.

1.4.3.5.3 Расстояние Кука

Расстояние Кука измеряет эффект удаления данного наблюдения. Точки данных с большими остатками (выбросами или высоким рычагом могут исказить результат и точность регрессии. Он рассчитывается как:

$$D_i = \frac{(\hat{Y}_J - \hat{Y}_{J(I)})^2}{pMSE}, \quad (31)$$

где $\hat{Y}_J$ прогноз полной модели регрессии для наблюдения J ; $\hat{Y}_{J(I)}$ прогноз для наблюдения J из уточненной регрессионной модели, в которой наблюдение I было опущено; p – число параметров в модели; MSE – среднеквадратическая ошибка модели.

1.4.3.6 Мультиколлинеарность независимых переменных

Мультиколлинеарность – это явление, когда две или более независимых переменных сильно коррелированы в множественной регрессии. Когда мультиколлинеарность существует, затронутые входные переменные имеют большую стандартную ошибку. В этом случае проверка значимости коэффициента может привести к отклонению ошибочной нулевой гипотезы (коэффициент существенно не отличается от 0). Кроме того, из-за большой стандартной ошибки небольшое изменение входных данных может привести к значительному изменению модели, даже приводя к изменению знака коэффициента, что может привести к нежелательному результату в тесте стабильности модели.

1.4.3.6.1 Фактор инфляции дисперсии (VIF)

Мультиколлинеарность проверяется для всех входных переменных в модели с помощью теста VIF. Более высокий VIF является показателем более

высокой мультиколлинеарности, где VIF выше 5 считается большим. VIF рассчитывается следующим образом: сначала OLS регрессия применяется для входной переменной x_i , как функции от всех других входных переменных

$$x_i = \beta X_{-i}. \quad (32)$$

Затем рассчитывается коэффициент VIF для входной переменной x_i :

$$\text{VIF} = \frac{1}{1 - R_i^2}. \quad (33)$$

1.4.3.6.2 Индекс состояния

Индекс состояния равен максимальному значению квадратного корня из отношения квадрата наибольшего собственного значения $X'X$ (X – матрица переменных регрессии) к каждому из них. Индекс должен быть меньше 30.

1.4.3.6.3 Мультиколлинеарность категориальных переменных

Чтобы проверить мультиколлинеарность категориальных предикторов, мы вычисляем ранг матрицы предикторов. Если ранг меньше числа категориальных предикторов, мы можем заключить, что внутри категориальных предикторов существует мультиколлинеарность.

1.4.3.7 Тесты стабильности коэффициентов регрессии

Стабильность каждого коэффициента регрессии гарантирует надежность модели. Тест начинается с удаления последних 30% (forward stability) или первых 30% (backward stability) данных из исходного набора данных в качестве набора данных тестирования. На каждой итерации добавляется еще одно наблюдение в набор данных тестирования и переоценивается модель с теми же входными переменными, что и в предлагаемой модели.

В результате оценивается устойчивость модели путем сравнения знака, величины и доверительного интервала коэффициентов модели. Ожидается, что волатильность коэффициентов регрессии будет низкой, и коэффициенты будут лежать в пределах соответствующего доверительного интервала.

2 Применение методологии на примере рынка государственных закупок в России

2.1 Подготовка исходных данных для использования в модели

Исходными данными являются завершенные контракты по 44 ФЗ (Федеральный закон "О контрактной системе в сфере закупок товаров, работ, услуг для обеспечения государственных и муниципальных нужд") за 2017 год, выгруженные с официального сайта Единой информационной системы в сфере закупок, определение поставщика для которых происходило с помощью электронного аукциона.

Данные представляют собой полную базу данных всех процедур государственных закупок, проводимых в соответствии с 44 ФЗ о государственных закупках. Данная база данных включает переменные, содержащиеся в 1) объявлениях о тендерах, 2) уведомлениях о присуждении контракта, 3) уведомлениях об изменении контракта, 4) объявлениях о завершении контрактов и 5) уведомлениях об административных исправлениях.

При обработке исходных данных были отсеяны ошибки набора и неправдоподобные данные, такие как дата начала подачи заявки позже даты конца подачи заявки, и для последующего анализа выбраны контракты со стоимостью 100 тыс. руб. и выше, отсеяны эмитенты, заключившие меньше трех контрактов за год.

Фрагмент образца исходных данных для одного контракта после первичной обработки представлен в Приложении А. На этом этапе каждый контракт можно сопоставить с дополнительной информацией, позволяющей дать максимальное представление об условиях его заключения: название и ИНН поставщика и заказчика, наименование контракта и характер необходимых работ (или тип продукции, которую нужно поставить в соответствии с ОКПД), список участников аукциона и т.д.

На следующем этапе из полученной базы данных извлекаются параметры, необходимые для последующего регрессионного анализа, согласно

реализуемой методологии и сохраняются в отдельный csv файл. На этом заканчивается этап первичной обработки данных, которому соответствует модуль `data_preprocessing` в реализуемой на языке Python архитектуре. Последующие алгоритмы запускаются из других модулей, что позволяет при внесении изменений в скрипт программы не пересчитывать заново обработку исходных данных, а работать с уже подготовленной базой данных.

2.2 Расчёт порогов для непрерывных переменных

В Разделе 1.2.3 описывается обоснование и идея приведения непрерывных переменных к категориальным за счёт нахождения соответствующих порогов, в которых эта переменная меняет характер своего влияния на соответствующий выходной параметр.

Рассмотрим расчёт порогов для непрерывных переменных на примере параметра «увеличение цены контракта после заключения сделки».

Данный параметр принимает значения от -0.963 до 2.65, среднее – 0.0006, стандартное отклонение – 0.029. Затем рассчитывается логистическая регрессия, где объясняемая переменная – выходной параметр текущей модели (допустим, рассматривается первая модель, где выходной параметр – «подача единственной заявки»), а объясняющие – «увеличение цены контракта после заключения сделки» и контрольные переменные. И затем запоминаются остатки этой логистической регрессии.

Значения параметра «увеличение цены контракта после заключения сделки» не уникальны, они могут принимать бесконечное число значений преимущественно в промежутке от -1 до 2, но также есть и уникальные случаи, когда изменение цены выходит за эти рамки. Для такого параметра весь диапазон принимаемых значений от минимального до максимального разбивается на 999 интервалов, и средние остатки уже для каждого интервала (Рисунок 2).

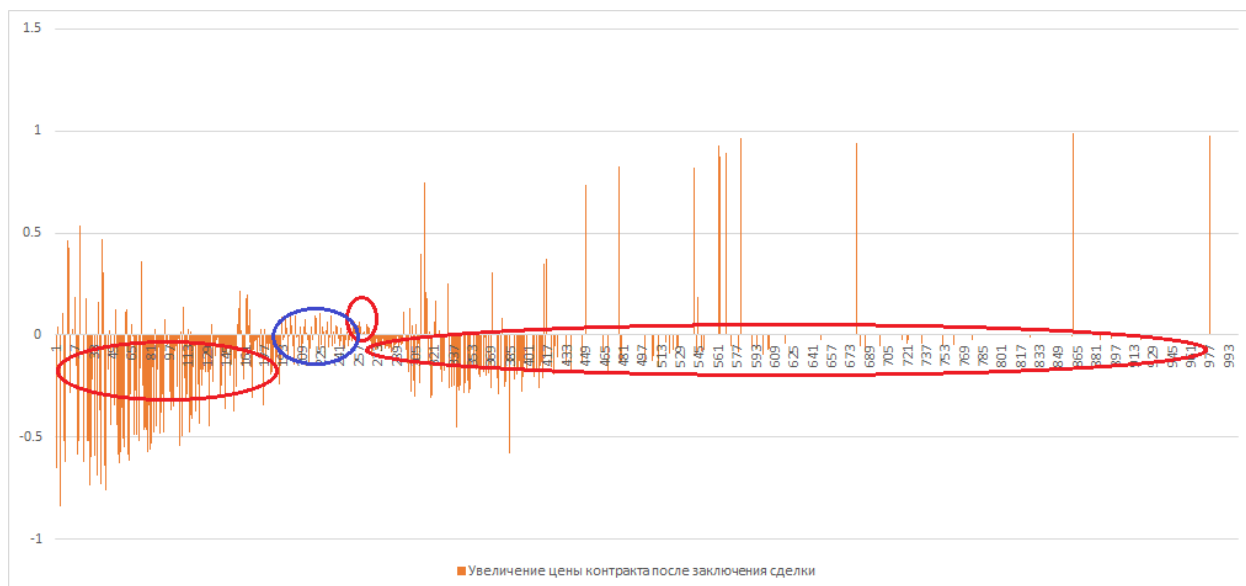


Рисунок 2 – Зависимость средних остатков логистической регрессии от соответствующего значения увеличения цены контракта (показано разбиение на группы)

В зависимости от того, колеблются ли остатки вокруг нуля, либо имеют растущий/падающий тренд выше/ниже нуля, значения параметра разделяются на группы. В данном случае можно видеть, что интервалы, разделяющие группы, являются: 190, 250, 256 и 293, а значения увеличения цены контракта, относительно которого происходит разбиение: -0.275, -0.058, -0.00336, 0.098, соответственно. Можно сделать вывод, что модель в первой и четвертой группе недооценивает вероятность подачи единственной заявки, а в третьей – переоценивает.

2.3 Оценка коэффициентов модели

Для того, чтобы подать набор категориальных переменных на вход логистической регрессии, необходимо преобразовать их в фиктивные переменные, а затем удалить один из столбцов полученной матрицы, чтобы входные параметры регрессии не образовали вырожденную матрицу. Вырожденной она будет по причине того, что один из столбцов обязательно будет линейной комбинацией остальных.

Таблицы 3,4 – Категории, на которые были разделены значения параметра «изменение цены после заключения сделки» (слева) и преобразование их в фиктивные переменные с удалением второго столбца (справа)

Номер	Категория	Номер	Изменение цены_1	Изменение цены_3	Изменение цены_4	Изменение цены_5
1	2	1	0	0	0	0
2	4	2	0	0	1	0
3	4	3	0	0	1	0
4	4	4	0	0	1	0
...	...	...	0	0	1	0
12	4	12	0	0	1	0
13	4	13	0	0	1	0
14	5	14	0	0	0	1
15	4	15	0	0	1	0
...	...	...	...	...	...	...

Рассмотрим оценку коэффициентов модели, где выходной параметр – «подача единственной заявки», а входные параметры – разделенные на фиктивные переменные показатели «длина периода подачи заявок», «длина периода рассмотрения заявок», «продление контракта после заключения сделки», «увеличение стоимости контракта после заключения сделки» и контрольные переменные, такие как: цена контракта, длительность исполнения контракта, количество уникальных заказчиков.

Результаты логистической регрессии представлены в Таблице 5, где четвертый столбец соответствует значениям p-value для каждого параметра. Следовательно, незначимыми параметрами являются только ass_length_decision_2 (assessment length decision 2 – длина рассмотрения заявки 2) и del_lengthening_3.0 (delivery lengthening 3 – продление контракта на этапе его реализации 3), так как для них p-value > 0.05.

Таблица 5 – Результаты логистической регрессии для «подачи единственной заявки»

ЭПК	Коэффициент	Станд. откл.	z value	p-value
sub_length_sub_1	0.6591	0.230	2.865	0.004
sub_length_sub_2	0.6118	0.010	58.897	0.000
sub_length_sub_3	-0.0854	0.012	-7.324	0.000
sub_length_sub_5	-0.6021	0.024	-24.980	0.000
ass_length_decision_2	-0.0197	0.013	-1.553	0.120
ass_length_decision_3	-2.7284	0.511	-5.338	0.000
del_lengthening_2.0	-0.2609	0.036	-7.165	0.000
del_lengthening_3.0	0.0719	0.058	1.250	0.211
del_value_incr_1.0	-0.8468	0.062	-13.766	0.000
del_value_incr_3.0	-0.6998	0.029	-24.536	0.000
del_value_incr_4.0	-0.9220	0.021	-43.211	0.000
del_value_incr_5.0	-1.2584	0.029	-43.829	0.000
con_contract_value	0.2334	0.002	123.524	0.000
con_contract_length	0.000	0.000	-5.679	0.000
con_unique_winners	0.000	0.000	46.333	0.000

2.4 Расчёт весов ЭПК, расчёт CRI

Методика расчёта весов подробно описана в Разделе 1.2.4. Распределение весов ЭПК в соответствии с значимостью соответствующих переменных регрессии показано в Таблице 6. Листинг кода для расчёта весов элементарных показателей коррупции на языке Python приведен в Приложении Б.

Таблица 6 – Переменные и их веса, соответствующие изменению значимости по критерию p-value

Номер	Вес	p-value	Имя переменной	Расшифровка
0	0.266667	0.004174	sub_length_sub_1	Длина периода подачи заявки 1
1	0.066667	0	sub_length_sub_2	Длина периода подачи заявки 2
2	0.2	2.41E-13	sub_length_sub_3	Длина периода подачи заявки 3
3	0.133333	1.02E-137	sub_length_sub_5	Длина периода подачи заявки 5
4	0	0.120427	ass_length_decision_2	Длина периода рассмотрения заявок 2
5	0.266667	9.40E-08	ass_length_decision_3	Длина периода рассмотрения заявок 3
6	0.266667	7.77E-13	del_lengthening_2.0	Продление контракта 2

7	0	0.21121	del_lengthening_3.0	Продление контракта 3
8	0.266667	4.11E-43	del_value_incr_1.0	Увеличение стоимости контракта 1
9	0.2	6.07E-133	del_value_incr_3.0	Увеличение стоимости контракта 3
10	0.133333	0	del_value_incr_4.0	Увеличение стоимости контракта 4
11	0.066667	0	del_value_incr_5.0	Увеличение стоимости контракта 5
12	0.266667	-	sub_single_bid	Подача единственной заявки
13	0.266667	-	ass_single_bid	Исключены все, кроме одного

Затем по формуле (1) рассчитан CRI для каждого контракта с этими весами. Найден максимальный CRI среди всех контрактов, и итоговое значение весов получилось путём деления текущих значений весов на это максимальное значение CRI. В расчётах данной модели максимальный CRI оказался примерно равен 1, из-за чего значения весов изменились незначительно. Затем CRI для каждого контракта был пересчитан уже с учетом новых значений весов.

Пример расчёта CRI по формуле (1) для одного контракта приведен в Таблице 7.

Таблица 7 – Пример расчёта CRI для одного контракта

Номер	Имя переменной	Вес параметра	Значение параметра	$w_j * CI_j^t$
0	Длина периода подачи заявки 1	0.266667	0	0
1	Длина периода подачи заявки 2	0.066667	1	0.066667
2	Длина периода подачи заявки 3	0.2	0	0
3	Длина периода подачи заявки 5	0.133333	0	0
4	Длина периода рассмотрения заявок 2	0	0	0
5	Длина периода рассмотрения заявок 3	0.266667	0	0
6	Продление контракта 2	0.266667	0	0
7	Продление контракта 3	0	0	0
8	Увеличение стоимости контракта 1	0.266667	0	0
9	Увеличение стоимости контракта 3	0.2	0	0
10	Увеличение стоимости контракта 4	0.133333	1	0.133333
11	Увеличение стоимости контракта 5	0.066667	0	0
12	Подача единственной заявки	0.266667	1	0.266667
13	Исключены все, кроме одного	0.266667	0	0
Итоговый CRI контракта:				0.46667

2.5 Агрегация CRI на различные уровни (контракт, поставщик, регион)

После расчёта индекса CRI для каждого контракта, наступает этап его агрегации на различные уровни. CRI в регионе считался, как средневзвешенный по стоимости по всем контрактам в регионе за выбранный промежуток времени. Гистограмма распределения индекса CRI по регионам за 2017 год приведена на Рисунке 3.

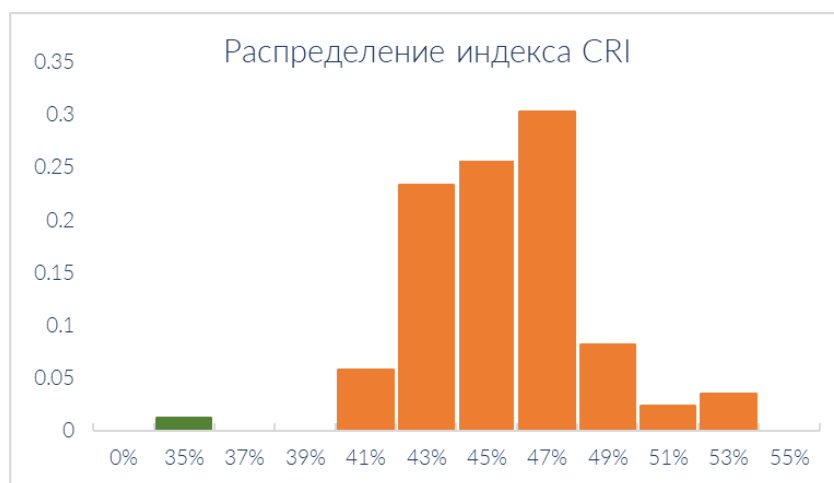


Рисунок 3 – Распределение индекса CRI по регионам

Можно видеть, что коррупционный риск для большинства регионов оценивается в пределах от 43 до 47%, что соответствует среднему уровню риска. Следует подвергнуть тщательному анализу регионы, для которых значение CRI равно 49% и выше – это группа с повышенной риском коррупции.

В Таблице 8 приведены места регионов в рейтинге CRI и национальном рейтинге прозрачности закупок (НППЗ) [28], отражено место регионов по объёму инвестиций в основной капитал, согласно данным Федеральной службы государственной статистики за 2017 год. С помощью этой таблицы можно проверить две гипотезы: первая – регионы, имеющие высокий (низкий) коррупционный риск, должны одновременно занимать высокие (низкие) позиции в рейтингах CRI и НППЗ; вторая – чем выше позиция региона в коррупционных рейтингах, тем ниже его инвестиционная привлекательность.

Таблица 8 – CRI по регионам в сравнении с другими рейтингами

Регион	CRI	Место в рейтинге CRI	Бенчмарк		
			Место в рейтинге НРПЗ	Прозрачность закупок	Место по объему инвестиций в ОК
Чукотский АО	53.0%	85	85	Низкая	81
Белгородская обл.	50.3%	82	81	Низкая	33
Тыва Респ.	49.0%	80	78	Низкая	85
Забайкальский край	47.7%	78	76	Базовая	48
Калмыкия Респ.	47.5%	76	84	Низкая	83
Томская обл.	43.2%	25	25	Высокая	47

Таким образом, место регионов в рейтинге CRI соотносится с местом в национальном рейтинге прозрачности закупок (НРПЗ), а объем инвестиций в основной капитал региона не уменьшается с увеличением коррупционности региона.

На Рисунке 4 приведена динамика CRI по Томской области по месяцам за 2017 год в сравнении со средним CRI за 2017 год.



Рисунок 4 – Динамика CRI по Томской области по месяцам за 2017 год

Можно видеть, что рассмотренные значения CRI колеблются около среднего значения за год. Это порождает возможность дальнейшего более углубленного исследования динамики изменения CRI для различных регионов и её зависимости от политических событий в этом регионе.

При агрегировании полученных значений CRI на уровень конкретного заказчика возможно получить такой же вариант динамики (Рисунок 5), исследовать для него контракты с наибольшим CRI, обнаружить, какие составляющие определяют индекс для конкретного контракта, исследовать зависимость CRI от частоты выигрыша в конкурсе определенного поставщика и т.д. Появляется возможность составления рейтингов заказчиков в рассматриваемом регионе. Аналогично для поставщиков.



Рисунок 5 – Динамика CRI по месяцам за 2017 год для заказчика ФГБНУ «Томский национальный исследовательский медицинский центр Российской академии наук»

2.6 Валидация модели

Для проверки работоспособности модели количественно при валидации был реализован набор статистических тестов, описанных в Разделе 1.4:

1. Значимость коэффициентов логистической регрессии: критерий Вальда;
2. Значимость уравнения модели логистической регрессии: критерий хи-квадрат отношения правдоподобия;
3. Проверка допущений логистической регрессии;
4. Стабильность коэффициентов логистической регрессии: тесты forward- и backward- стабильности, тест Роу-Чоу.

В рамках данной работы особую важность представляют тесты на адекватность коэффициентов регрессии, их стабильность, мультиколлинеарность, значимость и стационарность, поскольку основной идеей реализуемой методологии является обнаружение зависимости вероятности коррупционного риска от потенциальных факторов его

возникновения и определение характера этой зависимости, нежели реализация модели с высокими предсказательными способностями. В следующих подразделах приведены результаты описанных тестов.

2.6.1 Проверка значимости коэффициентов регрессии (тест Вальда)

Результаты теста значимости представлены в Таблице 5 (4-й столбец). Видно, что значимость коэффициентов регрессии `ass_length_decision_2` и `del_lengthening_3.0` недостаточна, так как для них $p\text{-value} > 0.05$.

2.6.2 Проверка значимости уравнения регрессии. Критерий хи-квадрат отношения правдоподобия

Для текущей модели $p\text{-value}$ критерия хи-квадрат отношения правдоподобия стремится к 0 и гораздо меньше порогового значения 0.05. Следовательно, регрессионная модель является статистически значимой.

2.6.3 Проверка допущений логистической регрессии

2.6.3.1 Результаты тестов на стационарность

Результаты проверки переменных регрессии на стационарность представлены в Таблице 9.

Таблица 9 – Результаты тестов на стационарность

Название переменной	Тип теста	ADF p-value	ADF-GLS p-value	PP p-value	KPSS p-value	Результат
<code>con_contract_value</code>	mean	0.01	>0.05	0.01	0.01	Положительный
<code>con_contract_value</code>	trend	0.01	>0.05	0.01	0.01	Положительный
<code>con_contract_length</code>	mean	0.01	>0.05	0.01	0.01	Положительный
<code>con_contract_length</code>	trend	0.01	>0.05	0.01	0.01	Положительный
<code>con_unique_winners</code>	mean	0.01	>0.05	0.01	0.01	Положительный
<code>con_unique_winners</code>	trend	0.01	>0.05	0.01	0.01	Положительный
Residuals	none	0.01	>0.05	0.01	0.01	Положительный

На основании результатов, представленных в Таблице 9, все переменные являются стационарными.

2.6.3.2 Уровень данных

Проверено, что зависимая переменная имеет дихотомический характер. Зависимая переменная разделяется на две категории и подходит для бинарной логистической регрессии.

2.6.3.3 Величина выборки

По формуле (26) рассчитано необходимое количество наблюдений в выборке, согласно [27]. Размер выборки составляет 959173, тогда как минимальный размер выборки для удовлетворения предположения при данном количестве определяющих переменных в регрессии составляет 225. Предположение о размере выборки удовлетворяется.

2.6.3.4 Автокорреляция в остатках

2.6.3.4.1 Графики ACF и PACF

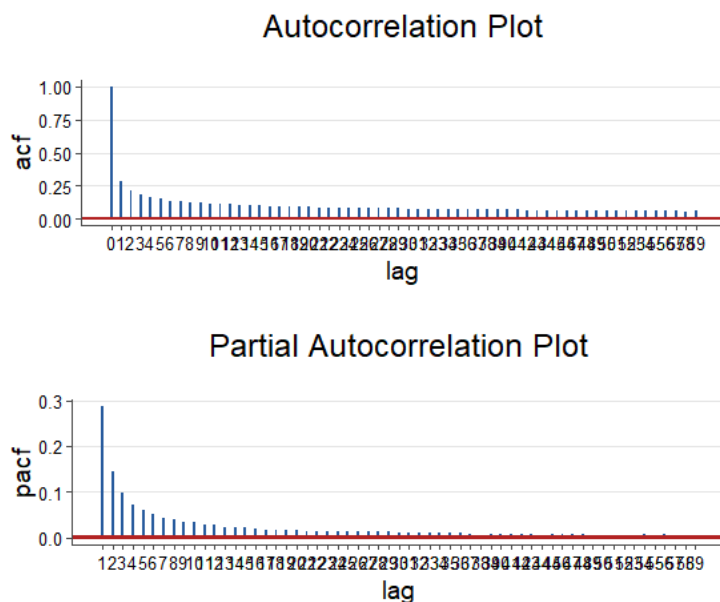


Рисунок 6 – Графики ACF и PACF

Корреляции находятся за пределами пунктирных границ, что свидетельствует о значительной остаточной автокорреляции.

2.6.3.4.2 Результаты тестов на автокорреляцию

Результаты тестов на автокорреляцию представлены в Таблице 10.

Таблица 10 – Результаты тестов на автокорреляцию

Тест	Лег	p-value	Результат
Durbin-Watson	-	0	Отрицательный
Ljung-Box	1	0	Отрицательный
Breusch-Godfrey	1	0	Отрицательный

На основании результатов, представленных в Таблице 10, остатки регрессии автокоррелированы, что совпадает с результатами графиков ACF и PACF. Это свидетельствует о том, что состоятельность и эффективность оценок коэффициентов регрессии может нарушаться, и для большей надежности результатов значимости регрессии необходимо рассмотреть вопрос о перестройке модели так, чтобы остатки не были автокоррелированы.

2.6.3.5 Обнаружение влиятельных наблюдений

Диаграмма размаха, рычаг(leverage), DFFITS и расстояние Кука рассмотрены для определения наиболее влиятельных точек исходной выборки.

2.6.3.5.1 Диаграмма размаха

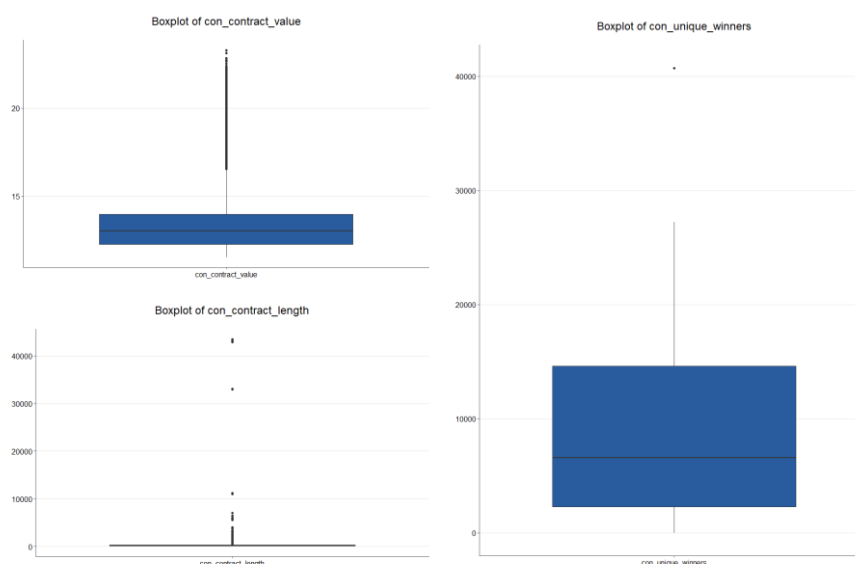


Рисунок 7 – Диаграммы размаха для численных переменных con_contract_value (стоимость контракта), con_contract_length (длительность исполнения контракта) и con_unique_winners (количество уникальных поставщиков у заказчика)

Диаграммы размаха указывают на то, что в данных существует значительное число выбросов.

2.6.3.5.2 Рычаг

Рычаг (Leverage) может быть использован для определения наиболее влиятельных точек.

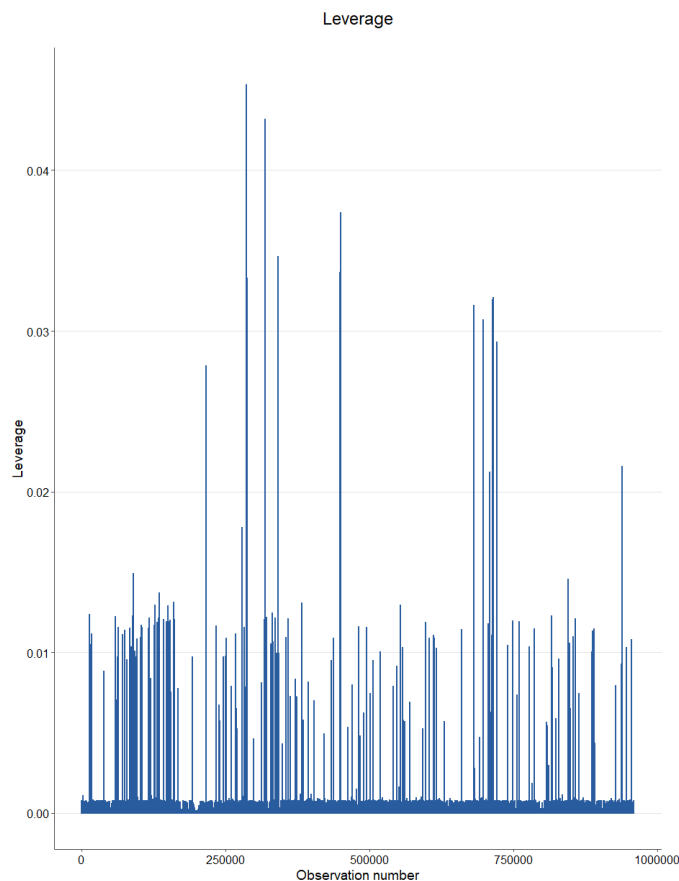


Рисунок 8 – График Рычага (Leverage)

Из графика видно, что в данных существует большое количество измерений, сильно влияющих на коэффициенты регрессии ($h_i > 0.01$). Измерения, для которых критерий выше 0.02: [215889, 285839, 287718, 318465, 340649, 449136, 449388, 681662, 681725, 681810, 697350, 708936, 713318, 715517, 721569, 938343]. Эти наблюдения имеют высокий потенциал воздействия на коэффициенты модели и требуют дополнительного изучения.

2.6.3.5.3 DFFITS

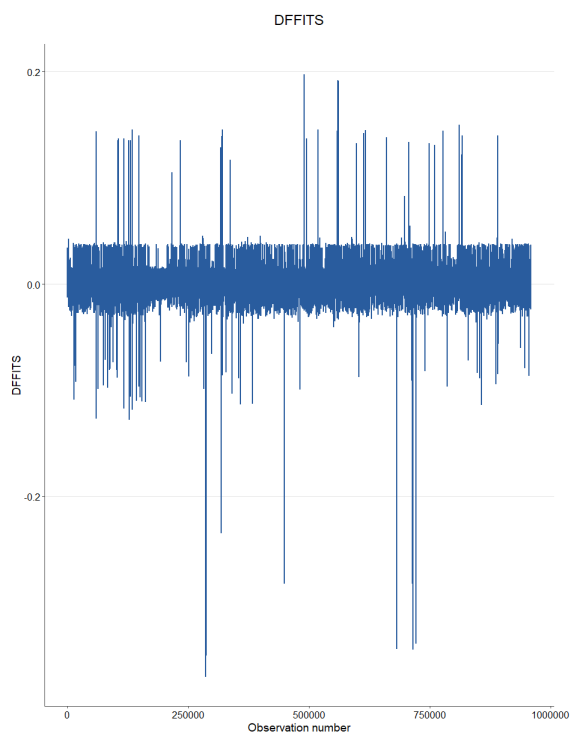


Рисунок 9 – График DIFITS

График DIFITS также свидетельствует о наличии в данных точек со значительным влиянием на регрессионную модель.

2.6.3.5.4 Расстояние Кука

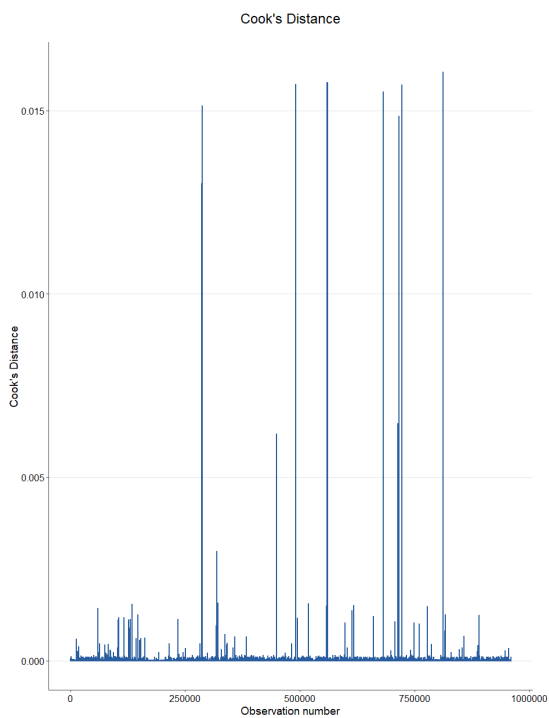


Рисунок 10 – График расстояния Кука (Cook's Distance)

Из графика видно, что в данных существует большое количество измерений, сильно влияющих на коэффициенты регрессии ($D_i > 0.005$). Измерения, для которых критерий выше 0.005: [285839, 287718, 449136, 449388, 490184, 559264, 560643, 681662, 681725, 681810, 713318, 715517, 721569, 811552]

Когда в непрерывном предикторе есть выбросы, потенциальные решения включают в себя: удаление соответствующей записи, преобразование данных в логарифмический масштаб, использование непараметрических методов.

2.6.3.6 Результаты теста на мультиколлинеарность

Результаты тестов на мультиколлинеарность представлены в Таблице 11.

Таблица 11 – Результаты тестов на мультиколлинеарность

Переменная	Название статистики	Значение статистики	Результат
con_contract_value	VIF	1.39	Положительный
con_contract_length	VIF	1.02	Положительный
con_unique_winners	VIF	1.01	Положительный
con_contract_value	Condition Index	2.45	Положительный
con_contract_length	Condition Index	2.99	Положительный
con_unique_winners	Condition Index	2.45	Положительный

Как видно из результатов тестов, для числовых и категориальных предикторов отсутствуют признаки мультиколлинеарности.

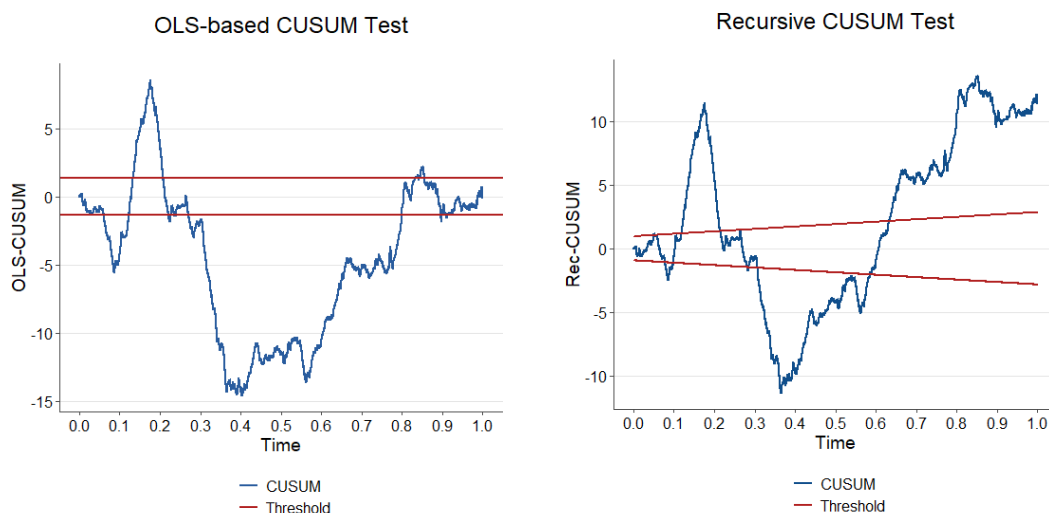
2.6.3.7 Тесты на стабильность

2.6.3.7.1 Стабильность коэффициентов

Для проверки надежности модели были проведены тесты forward и backward стабильности для каждого предиктора (Приложение Б).

На основании этих графиков можно сделать вывод, что не для всех предикторов точки всегда лежат внутри доверительных интервалов, и существует небольшая нестабильность коэффициентов регрессии.

2.6.3.7.2 CUSUM тесты



Рисунки 11, 12 – Графики OLS-based CUSUM и Recursive CUSUM тестов

Видно, что статистики OLS-based CUSUM и Recursive CUSUM тестов выходят за доверительные интервалы в виде ограничивающих линий на графиках, и существует некоторая нестабильность коэффициентов регрессии.

2.6.4 Валидация на конкретных данных

Для проверки индекса CRI на конкретных примерах, рассмотрены три контракта, которые подозреваются в коррупции, однако некоторые из них ещё не завершились, и пока уголовных осуждений ещё не было. Имеет смысл проверить их с помощью полученного индекса.

Посчитанные значения CRI для этих контрактов получились выше среднего: 0.533, 0.66 и 0.67.

Нужно учитывать, что возможна погрешность в расчетах в связи с тем, что модель обучалась на данных 2017 года и типом закупки был электронный аукцион, а в третьем контракте – контракт с единственным поставщиком.

В целом значения получились выше среднего, потому что помимо того, что все потенциальные заказчики, кроме выигравшего, отсеивались (или был заключен контракт с единственным поставщиком без конкурса), заявки рассматривались за один день, а также были изменения в условиях контрактов.

3 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

3.1 Предпроектный анализ

3.1.1 Потенциальные потребители результатов исследования

Потребителями разработанной технологии могут быть:

- Государственные органы Российской Федерации как на уровне страны, так и на уровне регионов – счётная палата, администрация регионов и т.д.;
- Владельцы бизнеса;
- Заказчики;
- Инвесторы.

		Виды использования методологии			
		Повышение эффективности и прозрачности проводимых закупок		Работа только с надёжными поставщиками	
Вид потребителя	Гос. органы РФ				
	Владельцы бизнеса				
	Заказчики				
	Инвесторы				

Фирма А

Фирма Б

Рисунок 13 – Карта сегментирования рынка услуг по виду использования разрабатываемой методологии

3.1.2 Анализ конкурентных технических решений с позиции ресурсоэффективности и ресурсосбережения

Основными конкурентами являются организации, использующие другие существующие индексы коррупции [8]. Данный научно-исследовательский проект рассматривает адаптацию и реализацию Corruption Risk Index (CRI) – индекс риска коррупции, который отражает коррупцию при формировании и распределении ренты. Одними из альтернативных подходов являются рейтинг НРПЗ - “Национальный рейтинг прозрачности закупок” является независимым негосударственным исследовательским аналитическим центром, специализирующимся в области экономического и правового анализа российского рынка государственных и корпоративных закупок и индекс восприятия коррупции – составляемый международной неправительственной организацией Transparency International показатель с целью отражения оценки уровня восприятия коррупции аналитиками и предпринимателями по стобальной шкале.

Бф – Индекс Corruption Risk Index (CRI)

Бк1 – Рейтинг НРПЗ

Бк2 – Индекс восприятия коррупции

Таблица 12 - Оценочная карта для сравнения индекса CRI с рейтингом НРПЗ и индексом восприятия коррупции

Критерии оценки	Вес критерия	Баллы			Конкурентоспособность		
		Б _ф	Б _{к1}	Б _{к2}	К _ф	К _{к1}	К _{к2}
Технические критерии оценки ресурсоэффективности							
1. Скорость расчёта	0,05	4	4	1	0,2	0,2	0,05
2. Удобство использования выбранной модели	0,05	3	3	3	0,15	0,15	0,15
3. Легкость интерпретации результатов	0,04	3	3	3	0,12	0,12	0,12
4. Простота понимания модели	0,04	3	4	5	0,12	0,16	0,2
5. Потребность в вычислительных ресурсах	0,06	5	5	5	0,3	0,3	0,3
6. Адаптивность метода	0,04	3	3	3	0,12	0,12	0,12
7. Возможность оценки коррупционного риска для на заданном уровне	0,1	5	2	2	0,5	0,2	0,2

(регион, город, отдельный контракт)							
8. Объективность расчетов	0,1	5	5	0	0,5	0,5	0
9. Точность расчетов	0,1	3	4	2	0,3	0,4	0,2
Экономические критерии оценки эффективности							
1. Конкурентоспособность продукта	0,07	3	4	2	0,21	0,28	0,14
2. Уровень проникновения на рынок	0,07	2	3	5	0,14	0,21	0,35
3. Цена	0,06	5	4	3	0,3	0,24	0,18
4. Послепродажное обслуживание	0,06	5	2	2	0,3	0,12	0,12
5. Финансирование научной разработки	0,07	5	5	5	0,35	0,35	0,35
6. Срок выхода на рынок	0,06	3	5	5	0,18	0,30	0,30
7. Наличие сертификации разработки	0,03	3	3	5	0,09	0,09	0,15
Итого	1	60	59	51	3.88	3.74	2.93

Анализ конкурентных технических решений определяется по формуле:

$$K = \sum B_i * B_i, \quad (34)$$

где K – конкурентоспособность научной разработки или конкурента;

B_i – вес показателя (в долях единицы);

B_i – балл i -го показателя.

Конкурентоспособность по сравнению с Бк1 = $3.88/3.74 = 1.04$.
 Конкурентоспособность по сравнению с Бк2 = $3.88/2.93 = 1.32$. Конкуренты уязвимы перед нашей разработкой, потому что они используют субъективные методы для расчета своего индекса и проводят исследование на одном выбранном уровне. В свою очередь объективность рассматриваемого индекса CRI является преимуществом данного исследования.

3.1.3 SWOT-анализ

SWOT – Strengths (сильные стороны), Weaknesses (слабые стороны), Opportunities (возможности) и Threats (угрозы) – представляет собой комплексный анализ научно-исследовательского проекта. SWOT-анализ применяют для исследования внешней и внутренней среды проекта.

Таблица 13 - Матрица SWOT

	Сильные стороны научно-исследовательского проекта: С1. Применение в любых государственных или коммерческих организациях С2. Низкая конкуренция на российском рынке по сравнению с европейским рынком С3. Простота использования программы	Слабые стороны научно-исследовательского проекта: Сл1. Большая теоретическая база применяемых математических методов Сл2. Необходимость работы с большим количеством данных Сл3. Прототип рассмотрен только на данных 2017 года Сл4. Долгое время расчетов
Возможности: В1. Выбор наиболее точной оценки коррупционного риска В2. Возможность оценки коррупционного риска для на заданном уровне (регион, область, город, отдельный контракт) В3. Работа как с российскими, так и с иностранными организациями В4. Применимость модели для контрактов с разными условиями исполнения	Клиент из любой государственной или коммерческой организации может получить наиболее точную оценку коррупционности риска рассматриваемого контракта на интересующем его уровне. Клиент может рассматривать надежность поставщиков вне зависимости от того, принадлежат ли они к иностранным компаниям. Повышение эффективности и прозрачности проводимых закупок.	Добавление возможности прочитать инструкцию использования предложенных методов. Разделение алгоритма на разные модули для контрактов различного типа. Улучшение интерфейса программы.
Угрозы: У1. Появление ошибок при определенных введенных параметрах У2. Эффективность существующих решений У3. Неточность внутренних алгоритмов в уникальных случаях	Добавить тесты в код программы, позволяющие выявлять неблагоприятные результаты и оперативно дорабатывать программу. Вести отладку ошибок.	Предотвращение ввода некорректных параметров. Улучшение алгоритма работы. Апробирование на дополнительном объеме данных. Введение алгоритмов оптимизации для уменьшения времени расчётов.

Результаты SWOT-анализа:

- необходимо ввести отладку ошибок, unit-тесты кода;
- необходимо ввести инструкцию по использованию методов;

- расширить диапазон используемых данных;
- ввести алгоритмы оптимизации для уменьшения времени расчётов.

3.1.4 Оценка готовности проекта к коммерциализации

Таблица 14 - Бланк оценки степени готовности научного проекта к коммерциализации

№ п/п	Наименование	Степень проработанности научного проекта	Уровень имеющихся знаний у разработчика
1	Определен имеющийся научно-технический задел	5	5
2	Определены перспективные направления коммерциализации научно-технического задела	5	4
3	Определены отрасли и технологии (товары, услуги) для предложения на рынке	5	4
4	Определена товарная форма научно-технического задела для представления на рынок	5	3
5	Определены авторы и осуществлена охрана их прав	5	4
6	Проведена оценка стоимости интеллектуальной собственности	4	4
7	Проведены маркетинговые исследования рынков сбыта	4	3
8	Разработан бизнес-план коммерциализации научной разработки	1	1
9	Определены пути продвижения научной разработки на рынок	4	3
10	Разработана стратегия (форма) реализации научной разработки	1	1
11	Проработаны вопросы международного сотрудничества и выхода на зарубежный рынок	1	1
12	Проработаны вопросы использования услуг инфраструктуры поддержки, получения льгот	1	1
13	Проработаны вопросы финансирования коммерциализации научной разработки	5	4
14	Имеется команда для коммерциализации научной разработки	5	5
15	Проработан механизм реализации научного проекта	5	5
ИТОГО БАЛЛОВ		56	48

По результатам проведенной оценки степени готовности научного проекта к коммерциализации суммарное количество баллов по степени проработанности научного проекта превышает уровень имеющихся знаний у разработчика. Согласно полученным баллам, можно сказать, что перспективность данной разработки выше среднего.

3.1.5 Методы коммерциализации результатов научно-технического исследования

Для коммерциализации результатов научно-технического исследования будет использоваться «Торговля патентными лицензиями», потому что данная разработка требует получения патента, который в дальнейшем может передаваться клиентам из государственных органов Российской Федерации, заинтересованным в повышении эффективности и прозрачности проводимых закупок, повышении экономии бюджетных средств и работе только с надежными поставщиками, что ведет к повышению уровня жизни населения за счет увеличения конкурентности на рынке поставщиков, а также за счет привлекаемых инвестиций в регион.

Необходима разработка прототипа и его презентация заинтересованным организациям.

3.2 Инициация проекта

3.2.1 Цели и результат проекта

Таблица 15 - Заинтересованные стороны проекта

Заинтересованные стороны проекта	Ожидания заинтересованных сторон
Государственные органы Российской Федерации как на уровне страны, так и на уровне регионов – счётная палата, администрация регионов и т. д.	Эффективный критерий оценки коррупционного риска в государственных закупках (возможно, для отдельного региона)
Владельцы бизнеса, заказчики, инвесторы	Простая в использовании программа, позволяющая отсеять ненадежных поставщиков.

Таблица 16 – Цели и результат проекта

Цели проекта:	Исследование и реализация нового объективного подхода к оценке коррупционного риска в государственных закупках, основанного на применении методов математической статистики.
Ожидаемые результаты проекта:	Программное обеспечение, позволяющие производить оценку коррупционного риска в государственных закупках на разных уровнях в виде значения индекса CRI.
Критерии приемки результата проекта:	Апробированный, эффективный метод, сопоставимость результатов с уже имеющимися индексами и здравым смыслом.
	Требование:
Требования к результату проекта:	Низкое потребление ресурсов.
	Высокая скорость расчёта.
	Высокая эффективность и адекватность индекса CRI.

3.2.2 Организационная структура проекта

Таблица 17 – Организационная структура проекта

№ п/п	ФИО, основное место работы	Роль в проекте	Функции	Трудозатр. час.
1	А.Е. Барышева, программист-аналитик ООО «ЭкоТомск»	Адаптация методологии под Российский рынок, консультирование по вопросам реализации отдельных методов, представление результатов.	Руководитель проекта	80
2	М.Г. Шибанов, ТПУ, магистр	Обработка исходных данных, разработка и реализация алгоритмов на языке Python	Программист, исследователь-стажёр, основной разработчик	558
ИТОГО:				638

3.2.3 Ограничения и допущения проекта

Таблица 18 – Ограничения проекта

Фактор	Ограничения/допущения
Бюджет проекта	219970 руб.
Источник финансирования	НИ ТПУ
Сроки проекта	6 месяцев
Дата утверждения плана управления проектом	21.12.2018
Дата завершения проекта	31.05.2019
Прочие ограничения и допущения	Отсутствуют

3.3 Планирование управления научно-техническим проектом

3.3.1 План проекта

Таблица 19 – Ограничения проекта

Код работы (из ИСР)	Название	Длительность, рабочие дни	Дата начала работ	Дата окончания работ	Состав участников (ФИО ответственных исполнителей)
1	Изучение существующих подходов к определению коррупционного риска в государственных закупках, изучение методик составления индексов коррупции	6	28.01.2019	3.02.2019	М.Г. Шибанов, А.Е. Барышева
2	Выбор актуального и методологически обоснованного подхода к определению коррупционного риска. Модернизация методологии для применения в России	6	4.02.2019	10.02.2019	М.Г. Шибанов, А.Е. Барышева
3	Сбор и первичная обработка данных, анализ данных, визуализация данных	6	11.02.2019	17.02.2019	М.Г. Шибанов
4	Составление математической модели коррупционного риска, реализация модели на языке R и Python	11	18.02.2019	3.03.2019	М.Г. Шибанов

5	Анализ полученных результатов и оформление промежуточного отчёта	4	4.03.2019	10.03.2019	М.Г. Шибанов, А.Е. Барышева
6	Применение технологии машинного обучения для автоматического извлечения дополнительных данных из текстовых документов для повышения качества модели	12	11.03.2019	24.03.2019	М.Г. Шибанов
7	Применение теории графов для расширения функционала модели	12	25.03.2019	7.04.2019	М.Г. Шибанов
8	Апробирование и валидация модели	12	8.04.2019	21.04.2019	М.Г. Шибанов
9	Анализ полученных результатов и оформление магистерской работы	24	22.04.2019	27.05.2019	М.Г. Шибанов, А.Е. Барышева
И т о г о:		93			

В Таблице 20 представлен календарный план-график проведения НИОКР по теме, где закрашенными прямоугольниками обозначен временной промежуток работы инженера, а пунктирными – руководителя.

Таблица 20 – Календарный план-график проведения НИОКР по теме

Код работ ы (из ИСР)	Вид работ	Исполни тели	Т _к , кал, дн.													
				январь		февраль		март			апрель			май		
				3	1	2	3	1	2	3	1	2	3	1	2	3
1	Изучение существующих подходов к определению коррупционного риска в государственных закупках, изучение методик составления индексов коррупции	Руководитель, инженер	6													
2	Выбор актуального и методологически обоснованного подхода к определению коррупционного риска. Модернизация методологии для применения в России	Руководитель, инженер	6													
3	Сбор и первичная обработка данных, анализ данных, визуализация данных	Инженер (дипломник)	6													
4	Составление математической модели коррупционного риска, реализация модели на языке R и Python	Инженер (дипломник)	13													
5	Анализ полученных результатов и оформление промежуточного отчёта	Руководитель, инженер	6													
6	Применение технологии машинного обучения для автоматического извлечения дополнительных данных из текстовых документов для повышения качества модели	Инженер (дипломник)	13													
7	Применение теории графов для расширения функционала модели	Инженер (дипломник)	13													
8	Апробирование и валидация модели	Инженер (дипломник)	13													
9	Анализ полученных результатов и оформление магистерской работы	Руководитель, инженер	35													

3.3.2 Бюджет научного исследования

Таблица 21 - Сырье, материалы, комплектующие изделия и покупные полуфабрикаты

Наименование	Марка, размер	Кол-во	Цена за единицу, руб.	Сумма, руб.
Бумага	Пачка	1	250	250
Картридж для принтера	Шт	1	2500	2500
Ручка	Шт	2	30	60
Блокнот	Шт	1	120	120
Всего за материалы				2930
Транспортно-заготовительные расходы (3-5%)				146.5
Итого по статье С _м				3076.5

Таблица 22 - Расчет затрат по статье «Спецоборудование для научных (экспериментальных) работ»

№ п/п	Наименование оборудования	Кол-во единиц оборудования	Цена единицы оборудования, тыс.руб.	Общая стоимость оборудования, тыс.руб.
1.	Компьютер	1	42000	42000
2.	МФУ	1	8000	8000
3.	Монитор	1	10000	10000
4.	Комплектующие	1	2000	2000
Итого:				62000

Таблица 23 - Баланс рабочего времени

Показатели рабочего времени	Руководитель	Инженер
Календарное число дней	365	365
Количество нерабочих дней		
- выходные дни	48	48
- праздничные дни	14	14
Потери рабочего времени		
- отпуск	52	52
- невыходы по болезни		
Действительный годовой фонд рабочего времени	251	251

Таблица 24 - Расчёт основной заработной платы

Исполнители	З _б , руб.	k _р	З _м , руб.	З _{дн} , руб.	T _р , раб. дн.	З _{осн} , руб.
Руководитель	33664	1.3	43763	1813	14	25382
Инженер	12664	1.3	16462	682	93	63426

Дополнительная заработная плата

Дополнительная заработная плата для руководителя: $10\% \cdot (25382) = 2538$
(р)

Дополнительная заработная плата для инженера: $10\% \cdot (63426) = 6342.6$
(р)

Отчисления на социальные нужды

Отчисления во внебюджетные фонды для руководителя:
 $27,1\% \cdot (25382 + 2538) = 7566$ (р)

Отчисления во внебюджетные фонды для инженера:
 $27,1\% \cdot (63426 + 6342.6) = 18907$ (р)

Накладные расходы

Накладные расходы для руководителя: $30\% \cdot (25382 + 2538) = 8376$ (р)

Накладные расходы для руководителя для инженера: $30\% \cdot (63426 + 6342.6) = 20931$ (р)

Расчёт электроэнергии

Потребление:

- Системный блок ~ 400 Вт/ч;
- Монитор ~ 40 Вт/ч;

Затраты на электричество = $0.440 \text{ кВт} \cdot 93 \text{ дней} \cdot 6 \text{ часов/день} \cdot 5,8 \text{ р/кВтч} = 1424 \text{ р.}$

Таблица 25 – Расчет бюджета затрат НТИ

Наименование статьи	Сумма, руб.
1. Сырье, материалы, комплектующие изделия и покупные полуфабрикаты	3076.5
2. Специальное оборудование для научных (экспериментальных) работ	62000
3. Затраты по основной заработной плате	88808
4. Затраты по дополнительной заработной плате	8881

5. Отчисления на социальные нужды	26473
6. Электроэнергия	1424
7. Накладные расходы	29307
8. Бюджет затрат НТИ	219970

3.3.3 Матрица ответственности

Таблица 26 – Матрица ответственности

Этапы проекта	Инженер	Руководитель
Изучение существующих подходов к определению коррупционного риска в государственных закупках, изучение методик составления индексов коррупции	И	С
Выбор актуального и методологически обоснованного подхода к определению коррупционного риска. Модернизация методологии для применения в России	И	С
Сбор и первичная обработка данных, анализ данных, визуализация данных	И	С
Составление математической модели коррупционного риска, реализация модели на языке R и Python	И	С
Анализ полученных результатов и оформление промежуточного отчёта	И	С
Применение технологии машинного обучения для автоматического извлечения дополнительных данных из текстовых документов для повышения качества модели	И	С
Применение теории графов для расширения функционала модели	И	С
Апробирование и валидация модели	И	С
Анализ полученных результатов и оформление магистерской работы	И	С

Исполнитель (И) – лицо (лица), выполняющие работы в рамках этапа проекта.

Согласующее лицо (С) – лицо, осуществляющее анализ результатов проекта и участвующее в принятии решения о соответствии результатов этапа требованиям.

3.3.4 Реестр рисков проекта

Таблица 27 – Реестр рисков

№	Риск	Потенциальное воздействие	Вероятность наступления (1-5)	Влияние риска (1-5)	Уровень риска	Способы смягчения риска	Условия наступления
1	DOS атака	Внедрение вредоносного кода в проект	1	4	Средний	Повышение уровня безопасности	Атака конкурентов
2	Отсутствие соединения с сетью интернет	Отсутствие доступа к требуемым данным	2	5	Высокий	Наличие запасного источника сети с другим провайдером	Проблемы на линии передач, например, в связи с природными катаклизмами
3	Несовместимость ПО	Элементы программы не будут работать правильно	1	4	Низкий	Составление требований к пользовательскому ПО	Не установлены необходимые версии программ для правильной работы алгоритма

3.4 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Интегральный финансовый показатель разработки: 1.

Таблица 28 – Сравнительная оценка характеристик вариантов исполнения проекта

Критерии \ ПО	Весовой коэффициент параметра	Текущий проект (Индекс CRI)	Аналог 1 (Рейтинг НРПЗ)	Аналог 2 (Индекс восприятия коррупции)
Адекватность (статическая значимость)	0,2	4	5	2
Возможность применения любым предприятием	0,15	5	4	2
Требуется наличия исторических данных	0,25	3	4	2
Простота применения	0,15	4	4	4
Конкурентоспособность модели	0,25	4	5	4
ИТОГО	1			

$$I_{тп} = 4 \cdot 0,2 + 5 \cdot 0,15 + 3 \cdot 0,25 + 4 \cdot 0,15 + 4 \cdot 0,25 = 3,9;$$

$$\text{Аналог 1} = 5 \cdot 0,2 + 4 \cdot 0,15 + 4 \cdot 0,25 + 4 \cdot 0,15 + 5 \cdot 0,25 = 4,45;$$

$$\text{Аналог 2} = 2 \cdot 0,2 + 2 \cdot 0,15 + 2 \cdot 0,25 + 4 \cdot 0,15 + 4 \cdot 0,25 = 2,8;$$

Сравнительная эффективность проекта с аналогом 1 = $3.9/4.45 = 0.88$, а с аналогом 2 = $3.9/2.8 = 1.39$.

Таблица 29 – Сравнительная эффективность разработки

№ п/п	Показатели	Рейтинг НРПЗ	Индекс восприятия коррупции	Индекс CRI
1	Интегральный финансовый показатель разработки	1	1	1
2	Интегральный показатель ресурсоэффективности разработки	4.45	2.8	3.9
3	Интегральный показатель эффективности	4.45	2.8	3.9
4	Сравнительная эффективность вариантов исполнения	0.88	1.39	

3.5 Выводы

В настоящем разделе рассмотрена перспективность и успешность разработанного метода оценки коррупционного риска в государственных закупках, создан механизм управления и сопровождения конкретных проектных решений на этапе реализации.

Во-первых, оценен коммерческий потенциал и перспективность проведения исследования. Полученные результаты говорят о потенциале и перспективности на уровне выше среднего.

Во-вторых, проведено планирование НИР, а именно: определена структура и календарный план работы, трудоемкость и бюджет НТИ. Результаты соответствуют требованиям к ВКР по срокам и иным параметрам.

В-третьих, определена эффективность исследования в разрезах ресурсной, финансовой, бюджетной, социальной и экономической эффективности.

4 Социальная ответственность

4.1 Введение

Целью данной работы является исследование и реализация нового объективного подхода к оценке коррупционного риска в государственных закупках, основанного на применении методов математической статистики.

Данное исследование актуально, поскольку высокий уровень криминализации и коррупции в экономической сфере (в частности, в государственных закупках) является одной из ключевых угроз национальной безопасности Российской Федерации [1]. Оценка коррупционного риска в государственных закупках и расчёт объективного индекса коррупционного риска Corruption Risk Index (CRI), который отражает коррупцию при формировании и распределении ренты, позволят повысить эффективность и прозрачность проводимых закупок, повысить экономию бюджетных средств государства, повысить конкуренцию на рынке поставщиков.

Данное исследование проводилось с использованием персональной электронно-вычислительной машины (ПЭВМ), а также его эксплуатация конечным пользователем тоже подразумевается на ПЭВМ. Известно, что компьютер является источником вредного воздействия на организм человека, а, следовательно, и источником профессиональных заболеваний. Вследствие этого каждый пользователь персонального компьютера должен быть осведомлен о вредном воздействии ПЭВМ на организм человека и необходимых мерах защиты от этих воздействий.

Целью настоящего раздела является: выявление и изучение вредных и опасных производственных факторов при работе с ПЭВМ и определение способов их устранения; оценка условий труда; рассмотрение вопросов техники безопасности, безопасности в чрезвычайных ситуациях и охраны окружающей среды.

4.2 Правовые и организационные вопросы обеспечения безопасности

4.2.1 Специальные правовые нормы трудового законодательства

В данном подразделе рассматриваются особенности трудового законодательства применительно к конкретным условиям проекта.

Во время прохождения научно-исследовательской и преддипломной практик в компании ООО «ЭкоТомск» был действителен трудовой договор, согласно которому труд работника осуществляется в нормальных условиях. Задачи не связаны с выполнением тяжелых работ, работ в местностях с особыми климатическими условиями, работ с вредными, опасными и иными особыми условиями труда.

Режим рабочего времени и времени отдыха регламентировались следующими пунктами:

- Работник по своей работе (должности) либо по состоянию здоровья относится к категории работников с нормированным рабочим днем.
- Работнику устанавливается 40-часовая пятидневная рабочая неделя с 8-часовым рабочим днём с выходными днями субботой и воскресеньем.
- В течение рабочего дня Работнику устанавливается перерыв продолжительностью 1 час для отдыха и питания, который в рабочее время не включается.
- Работнику предоставляется ежегодный оплачиваемый отпуск 28 (двадцать восемь) календарных дней, а также может предоставляться дополнительный отпуск. Предоставление дополнительного отпуска определяется «Положение об отпусках». Порядок предоставления ежегодного оплачиваемого отпуска, его перенос или продление регулируются законодательством РФ.
- Работа в выходные и праздничные дни проводится с соблюдением требований трудового законодательства Российской Федерации и с оплатой труда в размере не ниже установленного законодательством. Работник выходит на работу в выходные или праздничные дни по письменному распоряжению Работодателя. В случае, если выход Работника на работу в выходные или праздничные дни осуществляется по его самостоятельному решению, без

письменного распоряжения Работодателя, то оплата производится по общим правилам.

– С согласия Работодателя Работнику может быть предоставлен отпуск без сохранения заработной платы (в случаях, предусмотренных ТК РФ).

Гарантии и льготы регламентировались следующими пунктами:

– На Работника полностью распространяются льготы и гарантии, установленные для остальных работников, предусмотренные действующим законодательством РФ.

– Работник подлежит социальному страхованию в порядке и на условиях, установленных действующим законодательством Российской Федерации. Виды и условия социального страхования: обязательное медицинское страхование; обязательное социальное страхование от несчастных случаев на производстве и профессиональных заболеваний; обязательное пенсионное страхование.

– Имущественный ущерб, причиненный Работнику Работодателем, подлежит возмещению в порядке, в объеме и в сроки, установленные законодательством РФ.

– Ущерб, нанесенный Работнику увечьем либо иным повреждением здоровья, связанным с исполнением им своих трудовых обязанностей, подлежит возмещению Работодателем в соответствии с законодательством РФ.

Кроме того, было необходимо соблюдать требования не разглашать полученные во время работы сведения, составляющие коммерческую тайну, а так же иную конфиденциальную информацию о деятельности Работодателя и его клиентов: не известную третьим лицам информацию, включая информацию обо всех клиентах, а также информацию о проектах (продолжительность, количество участников и т.д.); техническую информацию, включая математические формулы, алгоритмы, идеи, концепции, исходные коды, дизайны и т.д.; секреты, связанные с управлением компанией, включая структуру компании, иерархию, личную информацию о сотрудниках (имена, адреса почтовые и электронные), информацию о зарплатах, льготах, а также иную финансовую информацию.

4.2.2 Перечень нормативных документов законодательства охраны труда

Работа над настоящей магистерской диссертацией регламентировалась следующие нормативно-правовыми актами:

1. Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 01.04.2019).
2. Гражданский кодекс РФ (ГК РФ) (части I и II).
3. Федеральный закон от 17 июля 1999 г. № 181-ФЗ "Об основах охраны труда в Российской Федерации".
4. ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.
5. ГОСТ 21889-76. Система «человек-машина». Кресло человека-оператора. Общие эргономические требования.
6. ГОСТ 22269-76. Система «человек-машина». Рабочее место оператора. Взаимное расположение элементов рабочего места. Общие эргономические требования.
7. ГОСТ Р 50923-96. Дисплей. Рабочее место оператора. Общие эргономические требования и требования к производственной среде. Методы измерения.
8. ГОСТ Р ИСО 9241-1-2007. Эргономические требования к проведению офисных работ с использованием видеодисплейных терминалов (VDTs). Часть 1. Общее введение.
9. ГОСТ Р ИСО 9241-2-2009. Эргономические требования к проведению офисных работ с использованием видеодисплейных терминалов (VDT). Часть 2. Требования к производственному заданию.
10. ГОСТ Р ИСО 9241-4-2009. Эргономические требования к проведению офисных работ с использованием видеодисплейных терминалов (VDT). Часть 4. Требования к клавиатуре.
11. ГОСТ Р ИСО 9241-5-2009. Эргономические требования к проведению офисных работ с использованием видеодисплейных терминалов (VDT). Часть 5. Требования к расположению рабочей станции и осанке оператора.

12. ГОСТ Р ИСО 9241-7-2007. Эргономические требования при выполнении офисных работ с использованием видеодисплейных терминалов (ВДТ). Часть 7. Требования к дисплеям при наличии отражений.
13. ГОСТ Р ИСО 9355-1-2009. Эргономические требования к проектированию дисплеев и механизмов управления. Часть 1. Взаимодействие с человеком.
14. ГОСТ Р ИСО 9355-2-2009. Эргономические требования к проектированию дисплеев и механизмов управления. Часть 2. Дисплеи.
15. ГОСТ Р ИСО 9355-3-2010. Эргономические требования к проектированию дисплеев и механизмов управления. Часть 3. Механизмы управления.
16. СанПиН 2.2.2/2.4.1340-03. Гигиенические требования к персональным электронно-вычислительным машинам и организации работы.

4.2.3 Организационные мероприятия при компоновке рабочей зоны

Согласно санитарным нормам СанПин 2.2.2.542-96, на каждого работника должно быть выделено не менее 6м² площади и не менее 20м³ объема, при высоте помещения – 2,8м, расстояние от глаз до клавиатуры, экрана и документов было 500 – 600мм, клавиатура находилась на высоте 740 – 790мм от уровня пола, верхний край дисплея находился на уровне прямого взгляда, а нижний – не ниже 40° от прямого взгляда.

Основными элементами рабочего места являются стол и стул, т.к. рабочим положением является положение сидя. Рациональная планировка рабочего места определяет порядок и местоположение предметов, в особенности тех, которые для работ необходимы чаще.

При проектировании письменного стола должны быть учтены следующие требования.

Высота рабочей поверхности стола рекомендуется в пределах 680–800 мм. Высота рабочей поверхности, на которую устанавливается клавиатура, должна быть 650 мм. Рабочий стол должен быть шириной не менее 700 мм и длиной не менее 1400 мм. Должно иметься пространство для ног высотой не

менее 600 мм, шириной — не менее 500 мм, глубиной на уровне колен — не менее 450 мм и на уровне вытянутых ног — не менее 650 мм.

Рабочее кресло должно быть подъёмно-поворотным и регулируемым по высоте и углам наклона сиденья и спинки, а также расстоянию спинки до переднего края сиденья. Рекомендуется высота сиденья над уровнем пола 420–550 мм. Конструкция рабочего кресла должна обеспечивать: ширину и глубину поверхности сиденья не менее 400 мм.

Монитор должен быть расположен на уровне глаз оператора на расстоянии 500–600 мм. Согласно нормам, угол наблюдения в горизонтальной плоскости должен быть не более 45° к нормали экрана. Лучше если угол обзора будет составлять 30°. Кроме того должна быть возможность выбирать уровень контрастности и яркости изображения на экране. Должна предусматриваться возможность регулирования экрана.

Для внутренней отделки интерьера помещений, должны использоваться диффузно отражающие материалы с коэффициентом отражения для потолка - 0,7 - 0,8; для стен - 0,5 - 0,6; для пола - 0,3 - 0,5.

Таким образом, рассмотрены правовые и организационные вопросы обеспечения безопасности: перечислены специальные правовые нормы трудового законодательства, особенности организации работы в условиях разработки данной магистерской диссертации и необходимые организационные мероприятия по компоновке рабочей зоны. Работа осуществлялась в нормальных условиях - задачи не были связаны с выполнением тяжелых работ, работ в местностях с особыми климатическими условиями, работ с вредными, опасными и иными особыми условиями труда. Установленные законодательством охраны труда нормы не нарушались, все требования к рабочему месту выполнены.

4.3 Производственная безопасность

В данном подразделе анализируются вредные и опасные факторы, которые могут возникать при разработке или эксплуатации проектируемого метода.

Вредным называется производственный фактор, воздействие которого на работающего в определенных условиях приводит к заболеванию или снижению работоспособности. При изменении уровня и времени воздействия вредные производственные факторы могут стать опасными.

Опасными считаются производственные факторы, воздействие которых на работающего в конкретных условиях может привести к травмам, а также другим внезапным резким ухудшениям здоровья.

В Таблице 30 представлены опасные и вредные производственные факторы, которым подвергается пользователь при работе с ПЭВМ, согласно ГОСТ 12.0.003-2015 «Опасные и вредные производственные факторы. Классификация».

Таблица 30 – Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003-2015)	Этапы работ			Нормативные документы
	Разрабо тка	Изготов ление	Эксплуа тация	
1.Отклонение показателей микроклимата			+	СанПиН 2.2.2/2.4.1340-03 "Гигиенические требования к персональным электронно-вычислительным машинам и организации работы", ГОСТ 29335—92 и 29338—92 "Костюмы мужские и женские для защиты от пониженных температур. Технические условия".
2.Недостаточная освещенность рабочей зоны	+	+	+	СанПиН 2.2.2/2.4.1340-03 "Гигиенические требования к персональным электронно-вычислительным машинам и организации работы".
3.Опасное воздействие электромагнитного излучения	+	+	+	СанПиН 2.2.2.542-96 "Гигиенические требования к видеодисплейным терминалам, персональным электронно-вычислительным машинам и организации работ", ГОСТ 12.4.154 "Система стандартов безопасности труда (ССБТ). Устройства экранирующие для защиты от электрических полей промышленной частоты. Общие технические требования, основные параметры и размеры".
4.Пожароопасность и вред здоровью от статического электричества		+		ГОСТ 12.4.124-83 "Система стандартов безопасности труда (ССБТ). Средства защиты от статического электричества. Общие технические требования".
5. Опасность поражения		+		ГОСТ 12.1.030-81 "ССБТ. Электробезопасность. Защитное заземление, зануление".

электрическим током				
6.Превышение допустимого уровня шума	+	+	+	ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности. ГОСТ 12.1.029-80 ССБТ. Средства и методы защиты от шума. Классификация.

4.3.1 Отклонение показателей микроклимата

Микроклимат производственных помещений – это климат внутренней среды помещений, который определяется действующими на организм человека сочетаниями температур воздуха и поверхностей, относительной влажности воздуха, скорости движения воздуха и интенсивности теплового излучения. Показатели микроклимата должны обеспечивать сохранение теплового баланса человека с окружающей средой и поддержание оптимального или допустимого теплового состояния организма, не приводящего к отклонению здоровья.

Нормы оптимальных и допустимых показателей микроклимата при работе с ЭВМ устанавливает СанПиН 2.2.2/2.4.1340-03. Все категории работ разграничиваются на основе интенсивности энергозатрат организма в ккал/ч (Вт). Работа, производимая сидя и сопровождаемая незначительным физическим напряжением, относится к категории Ia – работа с интенсивностью энергозатрат до 120 ккал/ч (до 139 Вт). Для данной категории допустимые нормы микроклимата представлены в Таблице 31.

Таблица 31 – Допустимые нормы микроклимата в рабочей зоне производственных помещений

Сезон года	Категория тяжести выполняемых работ	Температура, С°		Относительная влажность, %		Скорость движения воздуха, м/сек	
		Фактическое значение	Допустимое значение	Фактическое значение	Допустимое значение	Фактическое значение	Допустимое значение
Холодный	Ia	(22÷24)	(19÷24)	55	(15÷75)	0,1	≤0,1
Теплый	Ia	(23÷25)	(20÷28)	55	(15÷75)	0,1	≤0,2

Анализируя Таблицу 31, можно сделать вывод, что в рассматриваемом помещении параметры микроклимата соответствуют нормам СанПиН.

Допустимый уровень микроклимата помещения обеспечивается системой водяного центрального отопления и естественной вентиляцией.

Если помещение не удовлетворяет вышеперечисленным нормам, применяются следующие меры: использование систем местного кондиционирования воздуха, индивидуальных средств защиты от повышенной или пониженной температуры; регламентация периодов работы в неблагоприятном микроклимате; отдых работников в помещении с микроклиматом, нормализующим тепловое состояние; сокращение рабочей смены и др. В случае переохлаждения защита осуществляется посредством одежды, изготовленной в соответствии с требованиями ГОСТ 29335—92 и 29338—92 "Костюмы мужские и женские для защиты от пониженных температур. Технические условия".

4.3.2 Недостаточная освещенность рабочей зоны

Свет является естественным условием жизни человека. Правильно спроектированное и выполненное освещение обеспечивает высокий уровень работоспособности, оказывает положительное психологическое действие на человека и способствует повышению производительности труда. На рабочей поверхности должны отсутствовать резкие тени, которые создают неравномерное распределение поверхностей с различной яркостью в поле зрения, искажает размеры и формы объектов различия, в результате повышается утомляемость и снижается производительность труда.

Существует три вида освещения: естественное – за счёт солнечного излучения, искусственное – за счёт источников искусственного света и совмещенное – освещение, включающее в себя как естественное, так и искусственное освещения.

В данном рабочем помещении используется комбинированное освещение: искусственное и естественное. Искусственное освещение создается люминесцентными лампами типа ЛД.

Оценка освещенности рабочей зоны в соответствии с СанПиН 2.2.2/2.4.1.1340-03 показала, что необходимый световой поток светильника не выходит за пределы требуемого диапазона. При отклонении показателей освещенности от нормы, необходимо в срочном порядке исправить проблему, например, добавить дополнительные источники света или пересмотреть планировку помещения.

4.3.3 Опасное воздействие электромагнитного излучения

ЭМП обладает способностью биологического, специфического и теплового воздействия на организм человека, что может повлечь следующие последствия: биохимические изменения в клетках и тканях; нарушения условно - рефлекторной деятельности, снижение биоэлектрической активности мозга, изменения межнейронных связей, отклонения в эндокринной системе; вследствие перехода ЭМП в тепловую энергию может наблюдаться повышение температуры тела, локальный избирательный нагрев тканей и так далее.

Согласно СанПиН 2.2.2.542-96, напряженность ЭМП на расстоянии 50 см вокруг ВДТ по электрической составляющей должна быть не более 25В/м в диапазоне частот 5Гц-2кГц и 2,5В/м в диапазоне частот 2кГц/400кГц; плотность магнитного потока должна быть не более 250нТл в диапазоне частот 5Гц-2кГц и 25нТл в диапазоне частот 2кГц/400кГц.

Защита человека от опасного воздействия электромагнитного излучения осуществляется посредством СКЗ (защита временем; защита расстоянием; снижение интенсивности излучения непосредственно в самом источнике излучения; экранирование источника; защита рабочего места от излучения) и СИЗ (очки и специальная одежда, выполненная из металлизированной ткани (кольчуга)).

Экранирование источника излучения и рабочего места осуществляется специальными экранами по ГОСТ 12.4.154.

4.3.4 Пожароопасность и вред здоровью от статического электричества

Электризация заключается в следующем: нейтральные тела, в нормальном состоянии не проявляющие электрических свойств, при условии отрицательных контактов или взаимодействий становятся электрозаряженными. Опасность возникновения статического электричества проявляется в возможности образования электрической искры и вредном воздействии его на человеческий организм, и не только в случае непосредственного контакта с зарядом, но и за счет действий электрического поля, которое возникает при заряде. При включенном питании компьютера на экране дисплея накапливается статическое электричество. Электрический ток искрового разряда статического электричества мал и не может вызвать поражение человека. Тем не менее, вблизи экрана электризуется пыль и оседает на нем. В результате чего искажается резкость восприятия информации на экране. Кроме того, пыль попадает на лицо работающего и в его дыхательные пути.

Основные способы защиты от статического электричества следующие: заземление оборудования, увлажнение окружающего воздуха. Также целесообразно применение полов из антистатического материала.

4.3.5 Опасность поражения электрическим током

Электрический ток оказывает на человека термическое, электролитическое, механическое и биологическое воздействие.

1. Термическое воздействие тока проявляется в ожогах, нагреве кровеносных сосудов и других органов, в результате чего в них возникают функциональные расстройства.
2. Электролитическое действие тока характеризуется разложением крови и других органических жидкостей, что вызывает нарушения их физико-химического состава.
3. Механическое действие тока проявляется в повреждениях (разрыве, расслоении и др.) различных тканей организма в результате электродинамического эффекта.

4. Биологическое действие тока на живую ткань выражается в опасном возбуждении клеток и тканей организма, сопровождающемся непроизвольными судорожными сокращениями мышц. В результате такого возбуждения может возникнуть нарушение и даже полное прекращение деятельности органов дыхания и кровообращения.

К основным электрозащитным средствам в электроустановках напряжением до 1000 В относятся изолирующие штанги, изолирующие и электроизмерительные клещи, диэлектрические перчатки; изолированный инструмент.

Электрозащитные средства коллективной защиты в зависимости от назначения подразделяют на классы:

1. от повышенного уровня электромагнитных излучений (оградительные устройства, защитные покрытия, герметизирующие устройства, устройства автоматического контроля и сигнализации, устройства дистанционного управления, знаки безопасности);
2. от повышенной напряженности магнитных и электрических полей (оградительные устройства, защитные заземления, изолирующие устройства и покрытия, знаки безопасности);
3. от поражения электрическим током (оградительные устройства, устройства автоматического контроля и сигнализации, изолирующие устройства и покрытия, устройства защитного заземления и зануления, устройства автоматического отключения, устройства выравнивания потенциалов и понижения напряжения, устройства дистанционного управления, предохранительные устройства, молниеотводы и разрядники, знаки безопасности);
4. от повышенного уровня статического электричества (заземляющие устройства, нейтрализаторы, увлажняющие устройства, антиэлектростатические вещества, экранирующие устройства).

4.3.6 Превышение допустимого уровня шума

Источниками шума являются практически все технологическое оборудование: насосы, вентиляционные установки, компрессоры, транспортеры, разливочные автоматы, тестомесильные машины, электродвигатели, компьютеры, принтеры и т. п.

Допустимый уровень шума ограничен ГОСТ 12.1.003-83 и СанПиН 2.2.4/2.1.8.10-32-2002. Уровень шума на рабочем месте математиков-программистов и операторов видеоматериалов не должен превышать 50дБА, а в залах обработки информации на вычислительных машинах - 65дБА.

Для снижения шума применяют различные методы коллективной защиты: уменьшение уровня шума в источнике его возникновения; рациональное размещение оборудования; борьба с шумом на путях его распространения, в том числе изменение направленности излучения шума, использование средств звукоизоляции, звукопоглощение и установка глушителей шума, в том числе акустическая обработка поверхностей помещения. СИЗ включают в себя применение спецодежды, спецобуви и защитных средств органов слуха: наушники, беруши, антифоны.

4.4 Экологическая безопасность

Функционирование любого производства (промышленного, сельскохозяйственного и т. п.) сопровождается образованием отходов. Они поступают в окружающую среду в виде выбросов в атмосферу, сбросов в водоемы, твердых промышленных и бытовых отходов и мусора на поверхность и в недра Земли. В наше время обеспечение безопасности жизнедеятельности человека неразрывно связано с решением задач по охране природной среды.

Для уменьшения количества выбросов и отходов предприятия внедряют новейшие технологии, электрооборудование, ПЭВМ в процесс эксплуатации, отчистки и утилизации отходов производства. Однако все большее внедрение и применение ПЭВМ приводит к увеличению затрат электроэнергии, количества электростанций и их мощностей. Это может привести к различным

экологическим нарушениям: глобальное потепление, загрязнение атмосферы, воды и почвы вредными веществами и т.д.

В ходе работы были использованы такие ресурсы, как бумага и оргтехника.

Макулатура – отходы производства, переработки и потребления всех видов бумаги, картона, пригодных для вторичного использования в качестве волокнистого сырья.

Макулатура всех видов проходит несколько этапов обработки, причем их число зависит от качества исходного материала и назначения полученного сырья. Для выпуска низкосортного картона достаточно только первичного роспуска (отделение от целлюлозы грязи, песка, жира) и грубой очистки, а дальнейшая обработка для изготовления качественной бумаги различной фактуры требует полного цикла переработки (вторичный роспуск, тонкая очистка от примесей, химическая и механическая обработка).

Вся оргтехника включает в свой состав как органические составляющие (пластик различных видов, материалы на основе поливинилхлорида, фенолформальдегида), так и почти полный набор металлов. Благодаря комплексной системе утилизации оргтехники сводятся к минимуму не перерабатываемые отходы, а основные материалы (пластмассы, цветные и черные металлы) и ценные компоненты (редкие металлы, люминофор, ферриты и др.) возвращаются в производство.

Для перехода к безотходным производствам необходимо позаботиться о отдельных контейнерах для отходов бытового характера: отдельные мусорные баки для бумаги, стекла, металлических частей, пластика. Необходимо заключить договор с компанией, вывозящей мусор, чтобы она обеспечивала доставку разделенных отходов фирмам, занимающимся переработкой отходов.

4.5 Безопасность в чрезвычайных ситуациях

Чрезвычайные ситуации могут быть техногенного, природного, биологического, социального или экологического характера.

Техногенные ЧС связаны с техническими объектами, например, взрывы, пожары, транспортные катастрофы и пр. Природные – со стихийными силами природы, например, землетрясения, цунами, наводнения, извержения вулканов, оползни, сели, ураганы, смерчи, бури, природные пожары и др. Биологические ЧС связаны с заболеваниями, например, эпидемии, эпизоотии, эпифитотии. К социальным ЧС относятся такие межличностные конфликты в мире, как терроризм, грабежи, насилия, войны, голод и др. Экологические ЧС связаны с изменениями состояния природной среды: загрязнения биосферы, разрушение озонового слоя, опустынивание, кислотные дожди и т. д.

Возможными ЧС в Томске могут быть сильные морозы, диверсия, пожары, авария, оползни. Наиболее вероятный вид ЧС – пожар, следовательно, необходимо подробнее разобрать особенности пожарной безопасности.

Под пожарной безопасностью понимают состояние объекта народного хозяйства или иного назначения, при котором путем выполнения правовых норм, противопожарных и инженерно-технических мероприятий исключается или снижается вероятность возникновения и развития пожара, воздействия на людей опасных факторов пожара, а также обеспечивается защита материальных ценностей.

Источниками пожара в офисе являются: неисправности электрооборудования, в проводке и розетках, легковоспламеняющиеся объекты, короткое замыкание.

Обеспечение пожарной безопасности достигается комплексом следующих мероприятий: прогнозирование возможной пожарной опасности; проведение инженерно-технических и пожарно-профилактических мероприятий; соответствующая подготовка пожарных служб населения; создание постов из числа работающих; оценка пожарной обстановки и наблюдение за ней; обеспечение необходимого количества средств пожаротушения; разведка очагов пожаров; функционирование средств сигнализации и оповещения; локализация и тушение пожаров.

При возникновении пожара, особенностью которого является образование дыма, главной задачей, как и в любой ЧС, является спасение людей, которые впадают в панику при возникновении огня и могут в состоянии аффекта спрыгнуть с высокого этажа, в воду и т.д.

К числу простых, но надежных мероприятий, относится обеспечение путей эвакуации из зоны пожара, включая заранее разработанный и доведенный до сведения людей план вывода из помещения, изображенный графически на видном месте, обозначение световыми сигналами мест выхода; обеспечение обозначенных проходов и проездов свободными. Пути выхода с верхних этажей должны быть наружными или иметь постоянно открытые выходы дыма.

Чтобы отвести избыточное тепло от компьютера используют системы кондиционирования и вентиляции воздуха. Однако этими системами обеспечивается подача кислорода, который способен быстро распространять огонь, поэтому подобные системы становятся дополнительной пожарной опасностью в машинных залах и других помещениях. В вычислительных центрах установка пожарных кранов в коридорах, на площадках лестничных клеток и у входов способствует защите помещений от нежелательного возгорания. Потушить пожар можно также с помощью ручных углекислотных огнетушителей, установленных в помещениях из расчета один огнетушитель на 40-50 м².

План эвакуации из помещения, где была разработана магистерская диссертация, представлен в Приложении Г.

Оценивая пожарную опасность этого помещения, можно отнести его к пожароопасному помещению (категория В), поскольку в нём расположены деревянные столы и большое количество кабелей. В помещении отсутствуют горючие газы, легковоспламеняющиеся жидкости и т.д., а в процессе обращения с аппаратурой помещения не возникает воспламенения веществ, не используется топливо, огнетушители равномерно распределены по помещению.

4.6 Выводы

В настоящем разделе рассмотрено влияние разработанного метода оценки коррупционного риска в государственных закупках с применением методов математической статистики на окружающую среду. Выявлены и изучены вредные производственные факторы при работе с ПЭВМ (отклонение показателей микроклимата, недостаточная освещенность рабочей зоны, опасное воздействие электромагнитного излучения, пожароопасность и вред здоровью от статического электричества, опасность поражения электрическим током, превышение допустимого уровня шума) и определены способы их устранения.

Рассмотрены социальные и правовые нормы условий труда в процессе написания магистерской диссертации, приведены соответствующие нормативные акты, описаны особенности компоновки рабочей зоны.

Касательно вопроса экологической безопасности можно подметить, что необходимо следить за утилизацией отходов, в особенности за макулатурой.

Работа осуществлялась в нормальных условиях - задачи не были связаны с выполнением тяжелых работ, работ в местностях с особыми климатическими условиями, работ с вредными, опасными и иными особыми условиями труда. Установленные законодательством охраны труда нормы не нарушались, все требования к рабочему месту выполнены.

Заключение

В данной работе рассмотрен, исследован и реализован новый объективный подход к оценке коррупционного риска в государственных закупках, основанный на применении методов математической статистики.

Особенности этого подхода в том, что он:

1. Методологически обоснован и практически апробирован на ряде стран;
2. Рассчитывается на разных уровнях, в том числе на уровне контракта, что позволяет проверять каждую закупку в отдельности;
3. Лишен недостатка субъективности, базируется на математической модели и объективных данных (выбираются только значимые признаки и взвешивание не по экспертной оценке);
4. Поддается валидации.

В ходе работы решены следующие задачи:

1. Изучены существующие подходы к определению коррупционного риска в государственных закупках и методика составления индексов коррупции ИВК и WGI.
2. Выбран актуальный и методологически обоснованный подход к определению коррупционного риска. Методология модернизирована для возможности ее применения в России.
3. Проведена подготовка исходных данных (завершенные контракты по 44 ФЗ за 2017 год, выгруженные с официального сайта Единой информационной системы в сфере закупок) для использования в модели.
4. Составлена и реализована математическая модель коррупционного риска на языке программирования Python, результаты агрегированы на разные уровни.
5. Проведена валидация модели с использованием статистических методов и на конкретных контрактах, предложены способы улучшения качества коэффициентов модели.

Список используемых источников

- [1] С. А. Воронцов, О. В. Локота и А. В. Понеделков, ОСНОВЫ НАЦИОНАЛЬНОЙ БЕЗОПАСНОСТИ РОССИЙСКОЙ ФЕДЕРАЦИИ, Ростов-на-Дону, 2017, pp. 136-137.
- [2] «Миллиарды откатов. Как делают состояния на госзакупках.,» 19 Апрель 2016. [В Интернете]. Available: <https://news.rambler.ru/articles/33393070-milliardy-otkatov/>. [Дата обращения: 22 Март 2019].
- [3] С. Б. Котляров и Е. А. Чичеров, «Индикаторы измерения коррупции: понятие и совершенствование,» в *Актуальные проблемы права: материалы IV Междунар. науч. конф.*, Москва, 2015.
- [4] O. Bandiera, A. Prat и T. Valletti, «Active and Passive Waste in Government Spending: Evidence from a Policy Experiment,» *American Economic Review*, т. 99, № 4, pp. 1278-1308, 2009.
- [5] R. Di Tella и E. Schargrodsky, «The Role of Wages and Auditing during a Crackdown on Corruption in the City of Buenos Aires,» *Journal of Law and Economics*, т. 46, № 1, p. 269–292, 2003.
- [6] M. Fazekas, I. J. Tóth и L. P. King, «Anatomy of Grand Corruption: A Composite Corruption Risk Index Based on Objective Data,» *SSRN Electronic Journal*, 2013.
- [7] N. Charron, L. Dijkstra и V. Lapuente, «Mapping Quality of Government in the European Union: A Study of National and Sub-National Variation,» University of Gothenburg , Gothenburg , 2010.
- [8] S. Cole и A. Train, «Evidence from the Firm: A New Approach to Understanding Corruption,» в *International Handbook on the Economics of Corruption*, т. 2, S. Rose-Ackerman и T. Søreide, Ред., Cheltenham, Edward Elgar Publishing, 2011.
- [9] M. Fazekas, I. J. Tóth и L. P. King, «Corruption manual for beginners: Inventory of elementary “corruption techniques” in public procurement using the case of Hungary,» *SSRN Electronic Journal*, 2013.

- [10] M. Fazekas и I. J. & Tóth, «Public Procurement, Corruption and State Capacity in Hungary – objective measures and new insights,» *Corruption Research Centre, Budapest*, 2012.
- [11] M. A. Golden и L. Picci, «Proposal for a New Measure of Corruption, illustrated with Italian data,» *Economics & Politics*, т. 17, № 1, p. 37–75, 2005.
- [12] «Transparency International,» [В Интернете]. Available: <https://www.transparency.org/cpi2018>. [Дата обращения: 22 Май 2019].
- [13] D. Kaufmann, A. Kraay и P. Zoido, «Governance Matters. World Bank Policy Research Working Paper No. 2196,» World Bank Publications, Washington, 1999.
- [14] A. Rohwer, «Measuring Corruption: A Comparison between the Transparency International's Corruption Perceptions Index and the World Bank's Worldwide Governance Indicators,» *CESifo DICE Report*, т. 7, № 3, pp. 42-52, 2009.
- [15] Д. А. Квон, «Индекс восприятия коррупции: оценка методологии,» *Власть*, т. 23, № 6, pp. 72-76, 2015.
- [16] В. Н. Руденкин, «Проблема измерения коррупции: возможны ли объективные международные коррупционные рейтинги?,» *Вестник Уральского института экономики, управления и права*, т. 2, № 39, 2017.
- [17] M. Fazekas и I. J. Tóth, «Three indicators of institutionalised grand corruption using administrative data,» *Corruption Research Centre, Budapest*, 2014.
- [18] A. Alesina и E. Zhuravskaya, «Segregation and the quality of government in a cross section of countries,» *American Economic Review*, т. 101, № 5, pp. 1972-1911, 2011.
- [19] B. Rothstein и E. Uslaner, «All for all: equality, corruption, and social trust,» *World Pol.*, т. 58, № 1, pp. 41-72, 2005.
- [20] L. Guiso, P. Sapienza и L. Zingales, «Civic capital as the missing link,» в *Handbook of social economics*, San Diego, Elsevier, 2011, pp. 417-480.

- [21] R. Putnam, «Bowling alone: the collapse and revival of American community,» *The Journal of Catholic Education*, т. 6, № 2, 2002.
- [22] A. Persson, B. Rothstein и J. Teorell, «Why anticorruption reforms fail—systemic corruption as a collective action problem,» *Governance*, т. 26, pp. 449-471, 2013.
- [23] J. Wachs, T. Yasseri, B. Lengyel и J. Kerte'sz, «Social capital predicts corruption risk in towns,» *Royal Society Open Science*, т. 6, 2019.
- [24] D. Kaufmann, A. Kraay и M. Mastruzzi, «Governance Matters VIII : Aggregate and Individual Governance Indicators,» *World Bank Publications*, pp. 3-5, 2009.
- [25] D. Kaufmann и A. Kraay, «Governance indicators: Where are we, where should we be going?,» *The World Bank Research Observer*, т. 23, pp. 1-30, 2008.
- [26] M. Fazekas и G. Kocsis, «Uncovering high-level corruption: cross-national objective corruption risk indicators using public procurement data,» *Br. J. Pol. Sci.*, pp. 1-10, 2017.
- [27] P. Peduzzi, J. Concato, E. Kemper, T. Holford и A. Feinstein, «A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis,» *Journal of Clinical Epidemiology*, № 49, pp. 1373-1379.
- [28] «Результаты "Национального рейтинга прозрачности закупок 2017",» [В Интернете]. Available: <https://www.nrpz.ru/rezultati-2017>. [Дата обращения: 22 Март 2019].

Приложение А

Образец исходных данных (на примере одного контракта)

Номер	Регион	Организация-заказчик	Регистрационный номер контракта	Наименование контракта	Стадия контракта	Контракт с 1 заявкой	Исключены все, кроме одного заявителя	Продление периода рассмотрения	Продление контракта	Увеличение стоимости контракта
29	Адыгея Респ	УПРАВЛЕНИЕ ФЕДЕРАЛЬНОЙ НАЛОГОВОЙ СЛУЖБЫ ПО РЕСПУБЛИКЕ АДЫГЕЯ	1010504380517000007	Поставка моторного топлива	4	0	1	0	1	0

Старая сумма контракта	Новая сумма контракта	Сумма контракта	Дата начала подачи заявок	Дата окончания подачи заявок	Дата подведения итогов конкурса	Дата заключения договора	Дата подведения итогов закупки	Дата начала исполнения договора	Дата окончания исполнения договора
717826	657010	657010	06.02.2017	15.02.2017	15.02.2017	06.03.2017	20.02.2017	06.03.2017	31.12.2017

Дата начала контракта ПОСЛЕ измен	Дата начала контракта ДО изменения	Дата окончания контракта ПОСЛЕ из	Дата окончания контракта ДО измен	Тип процедуры закупки	Признак отмены закупки	ИНН заказчика	Статус организации заказчика	Код статуса организации заказчика
				Электронный аукцион	0	0105043805	Федеральные государственные казенные учреждения	75104

Наименование поставщика	ИНН поставщика	Количество деталей лота	список ОКПД2	Идентификационный код закупки	Федеральный округ	ОКПД новый	Маркет
ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ "КУБАНЬ-СЕРВИС"	2309085910	2	["47.30.10.000"]	171010504380501050100100060014730244	ЮФО	47	ЮФО47

Приложение Б

Листинг кода для расчёта весов элементарных показателей коррупции на языке Python

```
# Dataframe for weight assigning
regdat_explan_vars = pd.concat([sub_length_sub_dummy, ass_length_decision_dummy,
del_lengthening_dummy, del_value_incr_dummy, workdf['sub_single_bid'],
workdf['ass_single_bid']], axis=1)
number_of_outcome_variables=2
# Get p-values from the model results
Pvalues = model_result.pvalues[:regdat_explan_vars.shape[1]-
number_of_outcome_variables]

# Dataframe for weights
weights_df = pd.DataFrame({'Pvalues': Pvalues})
# Define weights array
weights = [0] * (regdat_explan_vars.shape[1] - number_of_outcome_variables)
weights_df['weights'] = weights

# Get p-values<=0.05
sign_Pvalues_values = Pvalues[Pvalues <= 0.05]
# Names of variables with p-values<=0.5
sign_Pvalues_names = sign_Pvalues_values.index
# Dataframe for significant p-values and corresponding names
sign_pvalues_df = pd.DataFrame({'Pvalues': sign_Pvalues_values, 'full_names':
sign_Pvalues_names, 'only_names': sign_Pvalues_names})

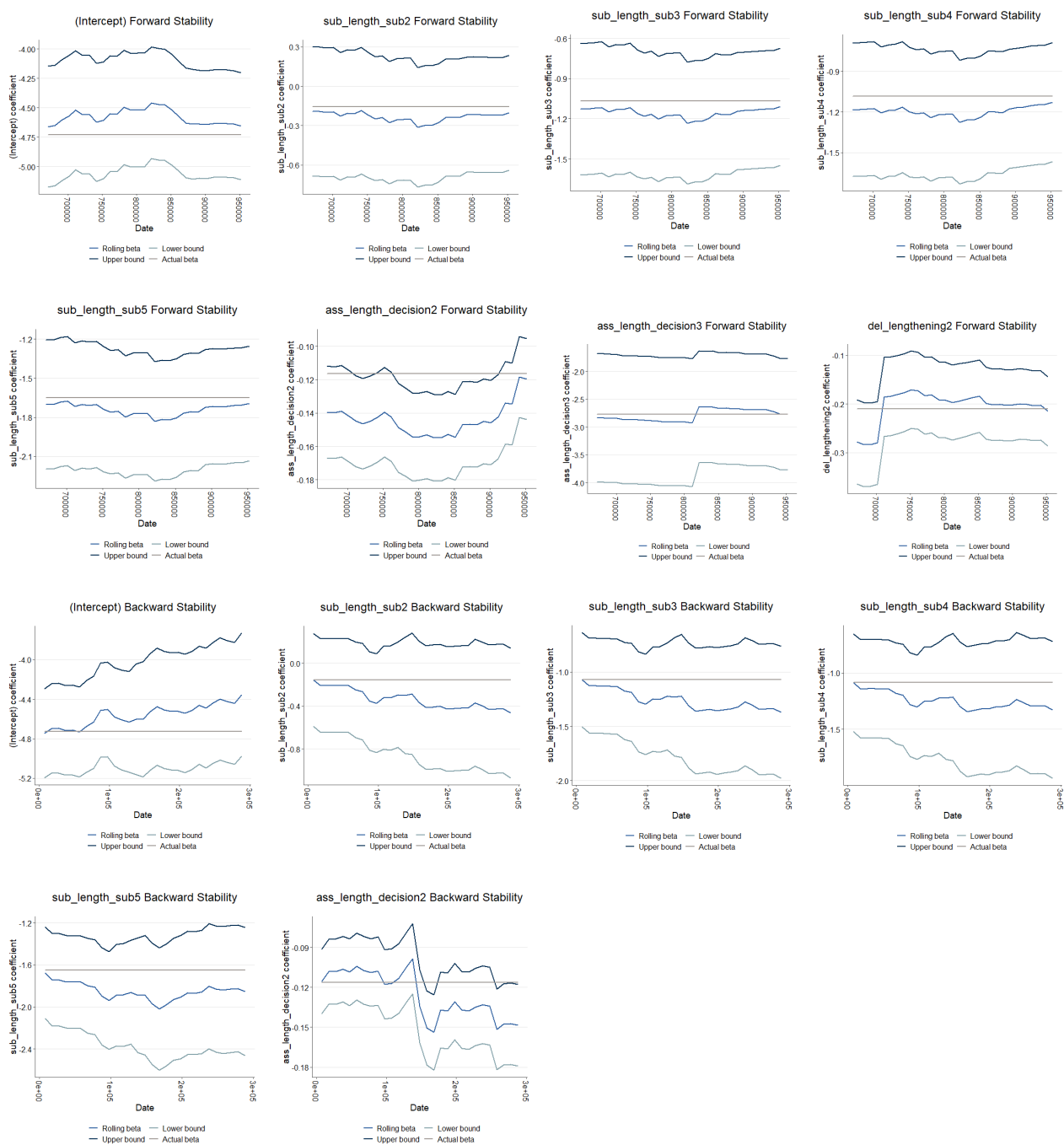
# Make 'only_names' column from sign_pvalues_df by excluding numbers from strings
for i in range(sign_Pvalues_names.shape[0]):
    sign_pvalues_df.iloc[i, 2] = re.sub(r'\d+', '', sign_Pvalues_names[i])
# Fix indexing of sign_pvalues_df
sign_pvalues_df.index = range(sign_pvalues_df.shape[0])

# Define variable for checking if category in loop was already given weights
I_no_repeat = list([-1])
for i in range(sign_pvalues_df.shape[0]):
    # Indices of variables from the same category
    I = sign_pvalues_df[sign_pvalues_df.iloc[:, 2] == sign_pvalues_df.iloc[i, 2]].index
    # Checking if this category was already given weights
    if I[0]!=I_no_repeat[0]:
        # Indices of variables in current category but in weights_df
        real_indices = sign_pvalues_df.iloc[I,1]
        # If without number some string matches with another (means that there are
several significant variables in current category)
        if I.shape[0] > 1:
            Isort = np.argsort(-sign_pvalues_df.iloc[I,0]) # Sorted indices
            # Assign weights according to decreasing p-value in current category
            for j in range(I.shape[0]):
                weights_df.loc[real_indices.iloc[Isort.iloc[j]], 'weights'] =
(I.shape[0] - j) / (I.shape[0])
            I_no_repeat = I # Remember indices
        else:
            # Assign weights to single significant variable
            weights_df.loc[real_indices, 'weights'] = 1

Weights = weights_df.iloc[:,1] # Take weights values form dataframe column
Weights['sub_single_bid'] = 1 # Give weight 1 to outcome variables
Weights['ass_single_bid'] = 1
total_number_of_explan_vars = 6
Weights = [i / total_number_of_explan_vars for i in Weights]
```

Приложение В

Результаты тестов forward и backward стабильности для переменных логистической регрессии

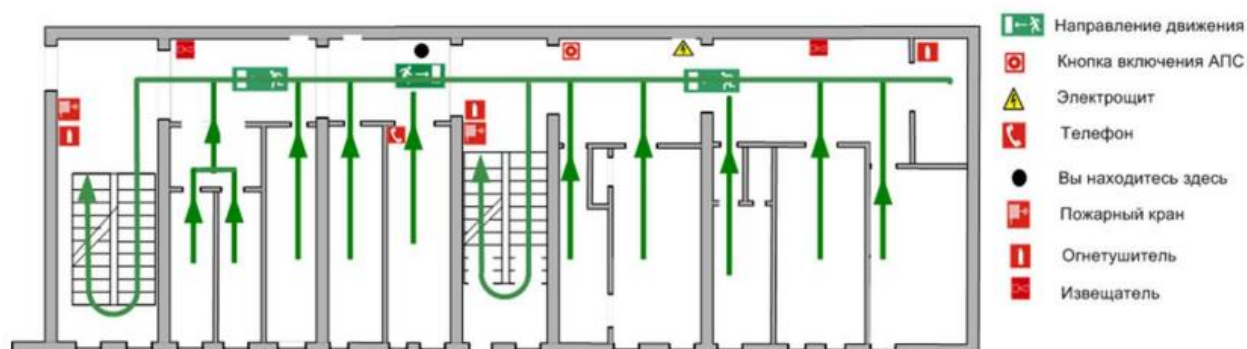


Приложение Г

План эвакуации в случае пожара

ПЛАН ЭВАКУАЦИИ

2-го этажа



Ответственный за эвакуацию и включение системы оповещения

Действия при пожаре Сохранять спокойствие			
1	Сообщить по телефону		• Адрес объекта • Место возникновения пожара • Свою фамилию
2	Эвакуировать людей		• Ориентироваться по знакам направления движения • Взять с собой пострадавших
3	По возможности принять меры по тушению пожара		• Использовать средства противопожарной защиты • При необходимости обесточить помещение

Приложение Д

Раздел 5

Modernization and application of the modern mathematical approach to assessing the risk of corruption in public procurement (applying to the Russian market)

Студент:

Группа	ФИО	Подпись	Дата
ОВМ71	Шибанов Максим Геннадьевич		

Консультант школы отделения (ОЭФ ИЯТШ):

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент	Семенов М.Е.	к. ф-м. н., доцент		

Консультант – лингвист отделения (ОИЯ ШБИП):

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Ст. преподаватель	Кабрышева О.П.			

Introduction

A high level of criminalization and corruption in the economic sphere [1] is one of the key threats to the Russian Federation national security. Particularly, public procurement corruption is a big problem nowadays. According to the information from the Accounts Chamber of the Russian Federation, the volume of public procurement was equal to 25 trillion rubles only in 2015. Meanwhile, violations for more than 34 billion rubbles were identified after for 9 months [2].

In our time, the empirical corruption research is global. Corruption indices for a large array of countries are calculated by Transparency International, Heritage Foundation, Freedom House on regular basis. And the main feature of the indices of these and many other organizations and research groups is the usage of “two main ways to measure corruption in society: the measurement through the perception and the measurement through the presence of corruption experience” [3]. It is obvious that both of these methods are too subjective.

The current trend in the corruption research is efforts to overcome subjectivity of its measurement in accordance with the aforesaid approaches and, first of all, methods of «World Bank’s Control of Corruption» and «Transparency International’s Corruption Perceptions Index» and the establishment of much more objective measurers of corruption. [4]; [5]. The most successful works in this sphere are the works of M. Fazekas and his colleagues [6]; [7] in CEU Government Transparency Institute and, connected with it, Corruption Research Center (Budapest) based on statistical analysis, Big Data, and Network Science.

The goal of the present work is to research and implement a new objective approach to assessing the risk of public procurement corruption which is based on the usage of mathematical statistics methods. To achieve this goal, the following tasks were formulated:

1. Study the existing approaches of corruption risk measurement in public procurement and the methodology of corruption indices compilation.

2. Choose the most relevant and methodologically supported approach in measurement the corruption risk. Modernize the methodology for the possibility of its usage in Russia.
3. Get, process and visualize the input data.
4. Create and implement a mathematical model of the corruption risk in the programming language Python.
5. Carry out the validation of this model.

Literature review

The work of Cole, S., & Tran, A [3], in particular the chapter «Evidence from the Firm: A New Approach to Understanding Corruption», made the great contribution to learn the mechanism of the public pronouncement corruption. In this chapter great attention is paid to common money laundering schemes, describing the key differences between public and private procurements, emphasizing unreliability of social surveys in the situations when even the careful research does not allow quantifying the existence of conspiracy.

Of particular interest in this work is the article of M. Fazekas [6], which presents the indicator obtained exclusively from objective data, defined at the micro level (for example, separate transactions) and having the possibility for successive time comparisons within and across countries. This indicator includes the huge amount of elementary indexes, described in the article [9], such as whether a single bid was submitted to the tender, the type of the procurement process, the length of eligibility criteria, the length of decision period, the cost of submitting documents for tender, the change of competition conditions etc. (See section 1.2.2 for details).

The usage of this methodology in Hungary for measuring the risk of public procurement across the whole country was considered in the article [10]. Also the objective method of measuring corruption based on data in Italy was tried on in the source [11].

The analysis of some of the existing corruption indices was introduced in [17]. In this work there is the comparative analysis of Corruption Risk Index (CRI) – which captures corruption in the generation and allocation of rents, Political Influence Indicator (PII) - the indicator of the political influence measuring the influence of the political group in the government on the success of public procurement companies and Political Control Indicator (PCI) - the indicator of the political control designating direct political control of public procurement contractors.

1 Object and methods of research

1.1 Calculation of Corruption Risk Index (CRI)

The CRI indicator was chosen as a criterion for assessing the risk of corruption in public procurement. It is obtained exclusively from objective data, defined at the micro level (for example, separate transactions) and having the possibility for successive time comparisons within and across countries

According to [6], for the calculation of CRI, the following tasks must be performed:

1. Get structured data from the public procurement database. The main problem of this issue is that data is unstructured, contains a lot of binary and categorical data and is introduced in different formats (scan, txt, pdf, excel, docx);
2. Choose indexes for assessing the risk of corruption in contracts. Expert assessment / statistical tests are used at this phase;
3. Analyze the corruption risk influence on one contract. Statistical analysis is used at this phase;
4. Calculate the index CRI for the contract and aggregate it for the necessary level (company, country, region).

$$CRI^t = \sum_j w_j * CI_j^t \quad (1)$$

$$\sum_j w_j = 1 \quad (2)$$

$$0 \leq CRI^t, CI_j^t \leq 1, \quad (3)$$

where CRI^t stands for the corruption risk index of transaction t , CI_j^t represents the j -th elementary corruption indicator observed in the transaction t , and w_j represents the weight of the elementary corruption indicator j .

1.2 Formulation of hypothesis about the components of corruption risk.

Public procurement corruption denotes the choice of the certain contractor with pre-inflated price that covers the so-called corruption rent - the amount of money returned to the organizer of the competition after winning this company. Bypassing the

competition can be carried out in three main forms each of which corresponds to the phase of the public procurement process:

1. Limiting the set of bidders: submission phase.
2. Unfairly assessing bidders: assessment phase.
3. Ex-post modifying conditions of performance: delivery phase.

Elementary corruption indicators CI_j^t corresponding to a certain phase of the realization of a certain transaction were chosen as objective data identifying the existence of corruption in a certain transaction. The input indicators used in this work to predict the output indicators and also the methods of their calculation are introduced in Table 1.

Table 1 – Indicators of corruption and methods of their calculation [6]

Phase	Indicator name	Definition
submission (submission of applications)	Single bidder contract	1=1 bid received, 0=more than 1 bid received
	Procedure type	procedure type – 1,2,3,4
	Length of eligibility criteria	number of characters of the eligibility criteria minus average number of characters of the given market's eligibility criteria
	Length of submission period	number of days between publication of call for tenders and submission deadline
	Relative price of tender documentation	price of tender documentation divided by contract value
	Call for tenders modification	0=call for tenders not modified 1=call for tenders modified
assessment (assessment of applications)	Exclusion of all but one bid	1=1 bid not excluded, 0=more than 1 bid not excluded
	Weight of non-price evaluation criteria	proportion of non-price related evaluation criteria within all criteria
	Annulled procedure re-launched subsequently	0=contract awarded in a NON-annulled procedure, 1=contract awarded in procedure annulled, but relaunched
	Length of decision period	number of working days between submission deadline and announcing contract award
delivery (implementation of a contract)	Contract modification	0=contract not modified during delivery 1=contract modified during delivery
	Contract lengthening	relative contract extension (days of extension/days of contract length)

	Contract value increase	relative contract price increase (change in contract value/original, contracted contract value)
	Winner's share within issuer's contracts	The amount received by the winner from this supplier for the year divided by the amount this supplier gave to the winners for the year

"Single bidder contract", "exclusion of all but one bid", "winner's share within issuer's contracts" are the output variables of the model. In this work it is considered that any indicator is sufficient to prove the possibility of any deal during the procurement. However, the output variables also depend on the input variables corresponding to their phase. To examine this dependence logistic regression was used in the case of indicators "single bidder contract" and "exclusion of all but one bid" because they are binary and the linear regression was used for "winner's share within issuer's contracts".

1.3 Definition of thresholds for continuous variables

The elementary indicators from Table 1 have different values - some of them are numeric as the cost of the contract or the relative change in its cost, others are binary or categorical variables when, for example, the necessary condition is either met or not. The third indicators are represented as a continuous function of time and it is necessary to unify them for further comparison. In this work all indicators were converted into categorical variables. For many indicators their quantitative increase did not mean the increased corruption risk, that is why, the method of division into different categories depending on the values of logistic regression residuals was automated for these indicators (Fig. 1).

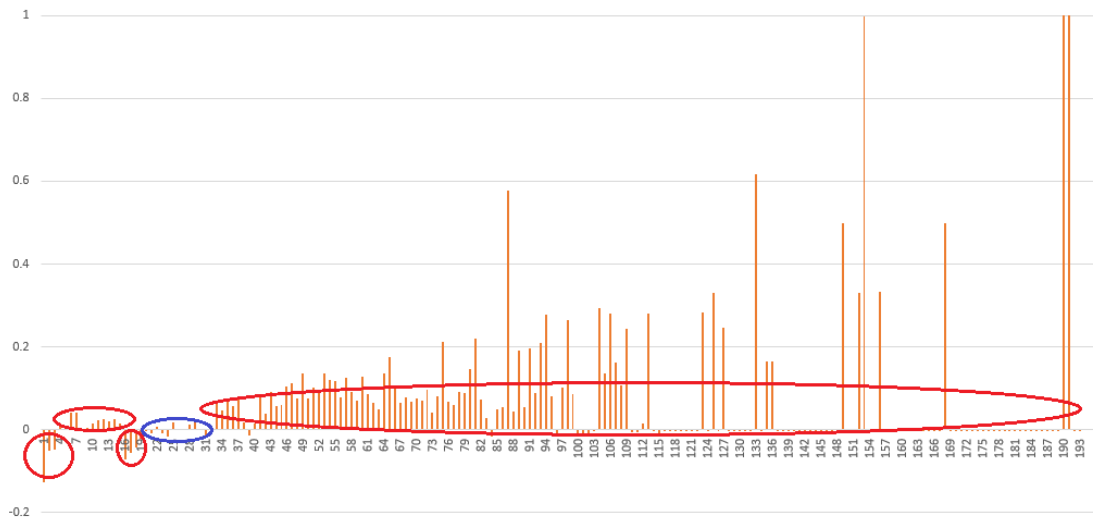


Figure 1 – Dependence of the average logistic regression residuals on the corresponding value of length of submission period (division into the groups is shown)

Thus, the variables are divided into groups for which the residuals fluctuate around zero or have a rising/falling trend above/below zero.

1.4 The usage of logistic regression for assessment of the corruption risk

When analyzing data, there are frequent issues where the output variable is categorical and then the usage of linear regression is difficult. Therefore, in this work when searching for the relationship between the set of input variables and the categorical output variable (Table 1), logistic regression is used, which is a method of binary classification. It allows us to assess the possibility of event realization (or not realization) according to the values of some independent variables. The logistic regression line, in contrast to the linear regression, is not straight.

Conditional average for logical regression:

$$\rho(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}, \quad (4)$$

Where e – the base of the natural logarithm; ρ - the possibility of occurring of the certain event; β_0, β_1 – the coefficients of the logistic regression; x – the value of the independent (input) variable.

This function is called “logistic”. The values of $\rho(x)$ change in the range from 0 to 1. If we assume that the value of the output variable y which is equal to 1 is considered to be the success and the value which is equal to 0 is considered to be the failure, $\rho(x)$ can be interpreted as the possibility of the success and $1-\rho(x)$ can be interpreted as the possibility of the failure.

The method of the least squares cannot be used for assessing the coefficients of the logistic regression. That is why its coefficients are assessed based on the maximum plausibility method. The logarithmic plausibility function:

$$L(\beta|x) = \sum_{i=1}^n \{y_i \ln[\rho(x_i)] + (1 - y_i) \ln[1 - \rho(x_i)]\}. \quad (5)$$

The structure of the logistic regression according to this work can be described as follows:

$$\Pr(\text{single bidder}_i = 1) = \frac{1}{1 + e^{-Z_i}} \quad (6)$$

$$Z_i = \beta_0 + \beta_{1j} S_{ij} + \beta_{2k} A_{ik} + \beta_{3l} D_{il} + \beta_{4m} C_{im} + \varepsilon_i, \quad (7)$$

Where single bidder_i («only one submission») is equal to 1 if the i th contract awarded had only one bidder и 0 if it has more; Z_i - the logit of a contract being a single bidder contract; β_0 – the constant of the regression; S_{ij} – the matrix of j corruption inputs of the submission phase for the i th contract (for example, the length of submission period); A_{ik} – the matrix of k corruption inputs of the assessment phase for the i th contract (for example, the weight of non-price evaluation criteria); D_{il} – the matrix of l corruption inputs of the delivery phase for the i th contract (for example, the contract lengthening); C_{im} – the matrix of m control variables for the i th contract (for example, the number of competitors on the market); ε_i – the error term; and β_{1j} , β_{2k} , β_{3l} , и β_{4m} are the vectors of coefficients for explanatory and control variables.

The similar model of the logistic regression is applied to the output index "exclusion of all but one bid", however the linear regression is constructed according

to the last output index “winner’s share within issuer’s contracts” because it is not binary and takes the value from 0 to 100%:

$$Y_i = \beta_0 + \beta_{1j} S_{ij} + \beta_{2k} A_{ik} + \beta_{3l} D_{il} + \beta_{4m} C_{im} + \varepsilon_i \quad (8)$$

where Y_i represents winner’s share within issuer’s contracts; other designations remain the same.

The results of these regressions allow us to proceed to the determination of the weights of elementary indicators for calculating CRI.

Significant indicators (p-value <0.05) are assigned weights - 1, non-significant - 0. Moreover, for each indicator, divided groups redistribute weights inside the group, for example, for an indicator with 3 significant groups weights will be equal to: 1, 0.67, 0.33 descending significance.