

DataBarrels

Кейс

"СПАРК: деятельность компаний"

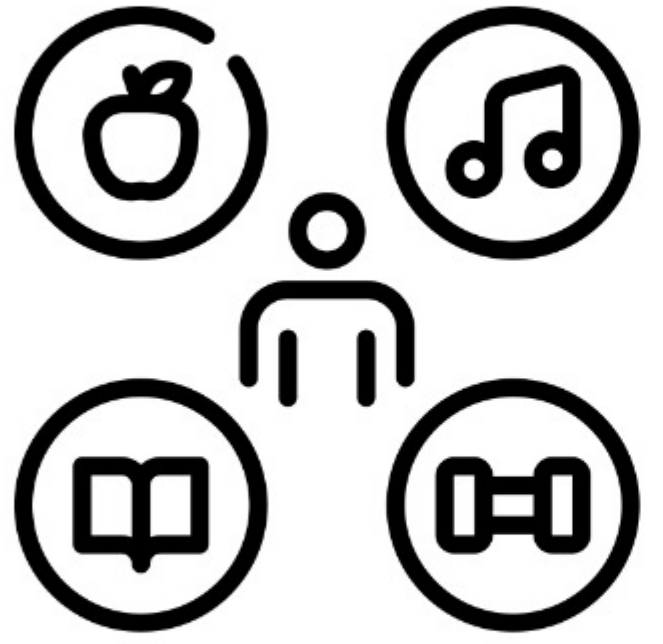
2023

Предобработка данных	Гипотеза + механизм	Мат. модель	Наш индекс!	Проверка гипотезы	Вывод	Применение и Policy implication	Перспективы и ограничения
----------------------	---------------------	-------------	-------------	-------------------	-------	---------------------------------	---------------------------

Наш датасет



Инновационная деятельность



Вид деятельности



Сотрудники

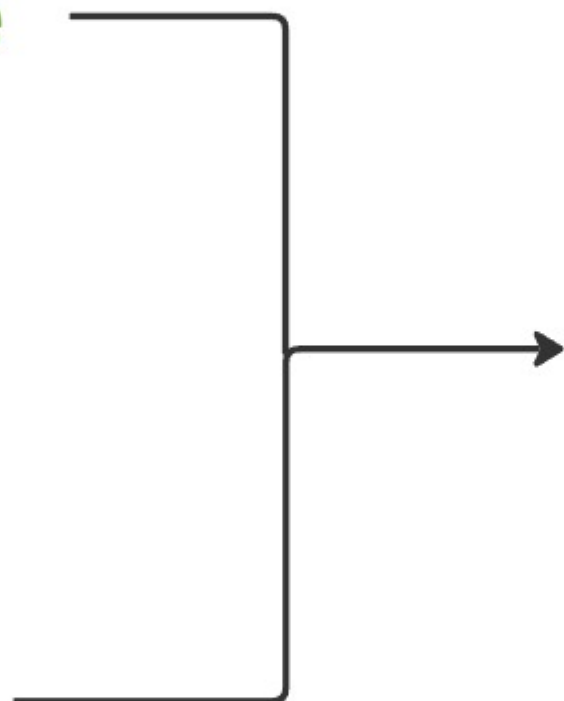


Финансовые результаты

Расширение датасета

ИНН в
предоставленной базе

Единый
Государственный
Реестр Юридических
Лиц



Полная информация из отчетов
сервиса "Прозрачный бизнес"

ИНН — первичный
ключ

Соблюдение логики в данных

> 0

Сотрудники

> 9489*

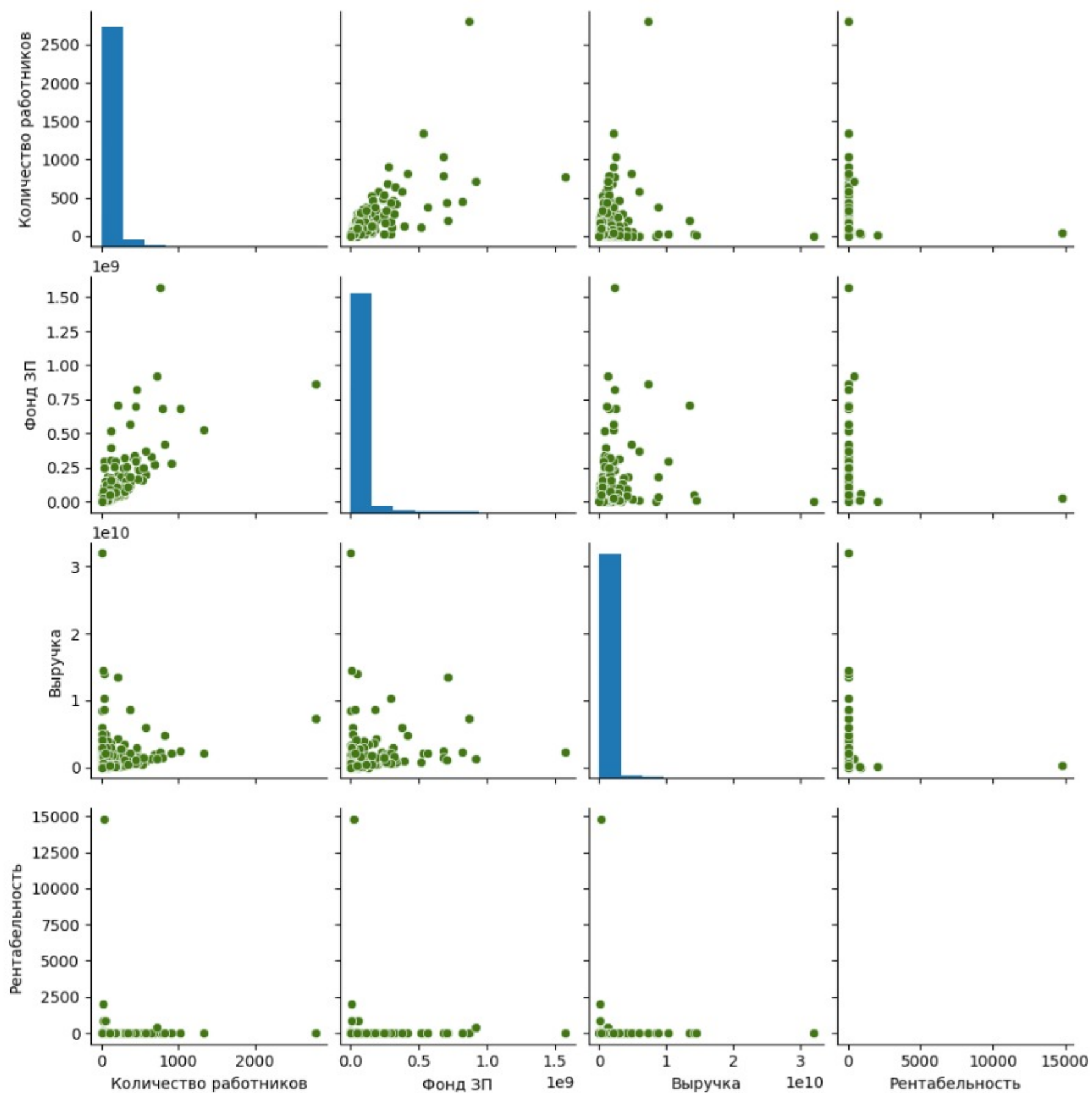
Фонд ЗП

> 0

Выручка

*МРОТ на 2018 составил 9489 рублей

Описательная статистика



Описательная статистика

	Сотрудники	Фонд ЗП	Выручка	Рентабельность
Минимум	1	9 489	0	-36%
Среднее	23	35 563 580	173 586 800	56%
Медиана	4	33 200 000	19 600 000	2%
Максимум	16100	37 909 870 000	96 950 000 000	310%

Как были очищены выбросы?

При классическом подходе к удалению выбросов теряются большие компании.



Группируем похожие



Ищем медианные показатели



Внутри отрасли убираем самые отклоняющиеся компании

Проблема

Решение

Критерии схожести



Похожая
деятельность



Близкая по
размеру выручка



Сходное
количество
персонала

Введём классные термины!

Линейный персонал — это сотрудники, которые выполняют базовые функции

Линейная компания — компания с преобладающим числом линейного персонала

Исследовательский вопрос

В каких отраслях следует внедрять инновации, чтобы повысить эффективность труда в России; в частности, какие отрасли эффективнее поддерживать в условиях рекордно низкой безработицы?

Гипотеза

Есть положительная зависимость между линейностью отрасли и эффективностью инноваций в ней.

Механизм индекса линейности труда

x10



Увеличение выручки



x20



x10



Непонятное
изменение выручки



x20



Механизм индекса линейности труда

1
Линейный персонал в основном использует внешние, в основном материальные инструменты производства

2
Использование примерно одинаковых инструментов похожим линейным персоналом приводит к примерно одинаковому вкладу в выручку

3
Примерно одинаковый вклад в выручку приводит к линейной зависимости выручки от числа линейных работников

4
В линейных компаниях доля линейных работников больше и они в целом схожи, а значит линейный тренд лучше заметен

Группировка компаний

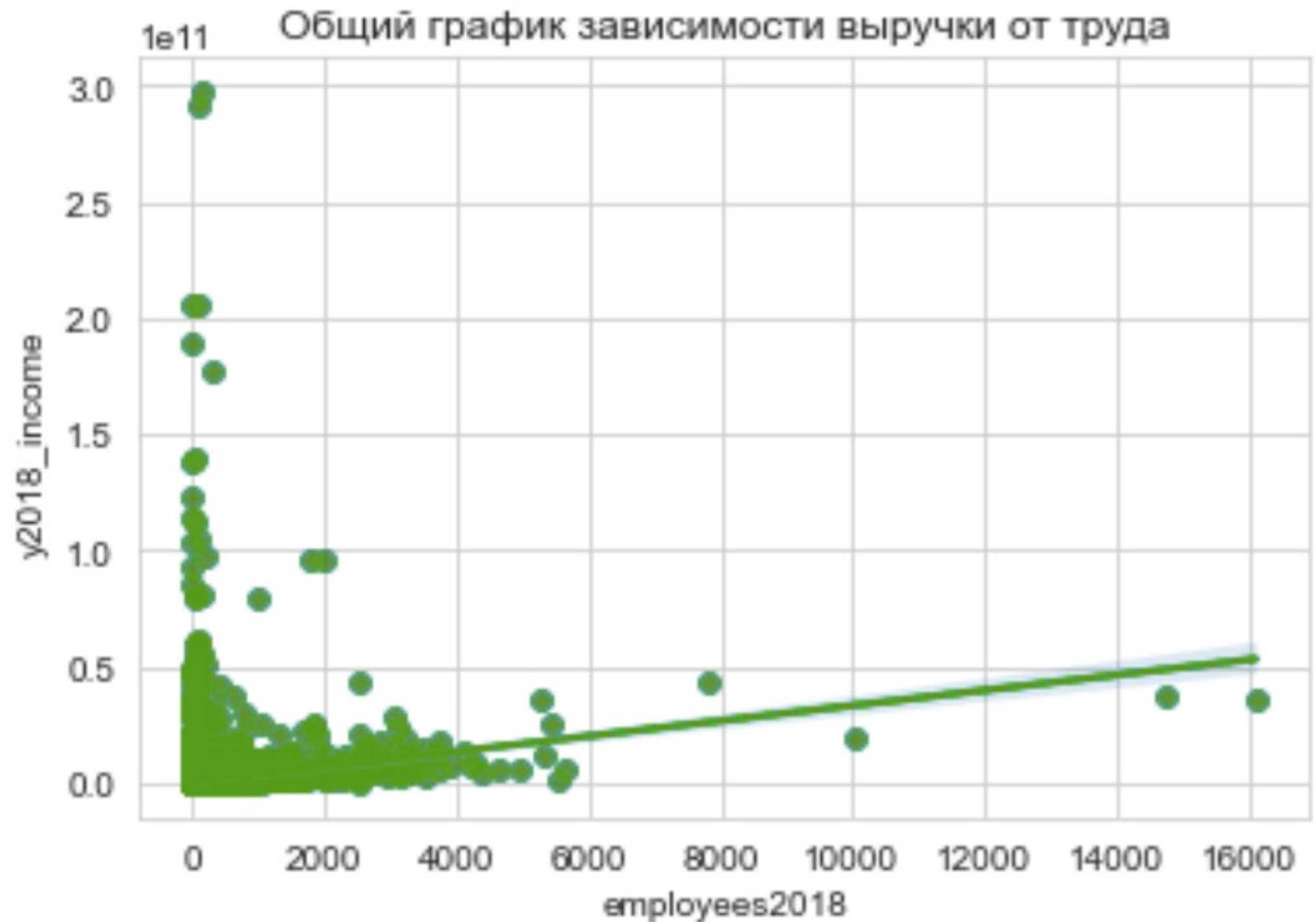
ФНС разделила все ОКВЭДы (2000 шт.)
на 20 разделов по видам экономической
деятельности.

Корреляции

$$r^* = 0.35$$

H_0 - коэффициент
корреляции равен нулю

$|t| < t_{кр}$ - не позволяет
отклонить H_0



Предобработка данных	Гипотеза + механизм	Мат. модель	Наш индекс!	Проверка гипотезы	Вывод	Применение и Policy implication	Перспективы и ограничения
----------------------	---------------------	-------------	-------------	-------------------	-------	---------------------------------	---------------------------

Математическая модель

$$\textit{Labor Linearity Index} = r(TR, L)$$

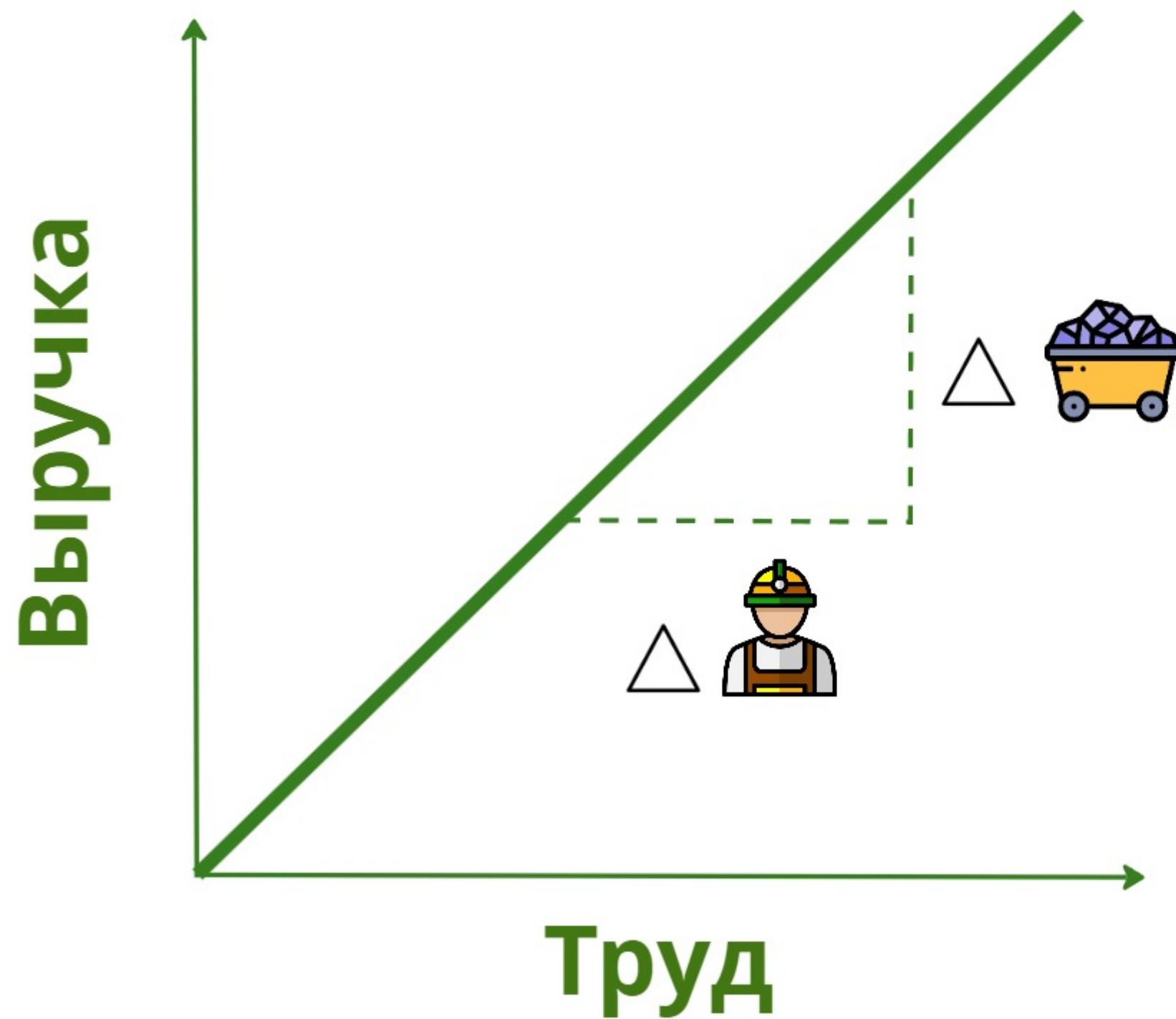
LLI - индекс линейности труда
(индекс репетативности)

$r(TR, L)$ - коэффициент корреляции

TR - полная выручка

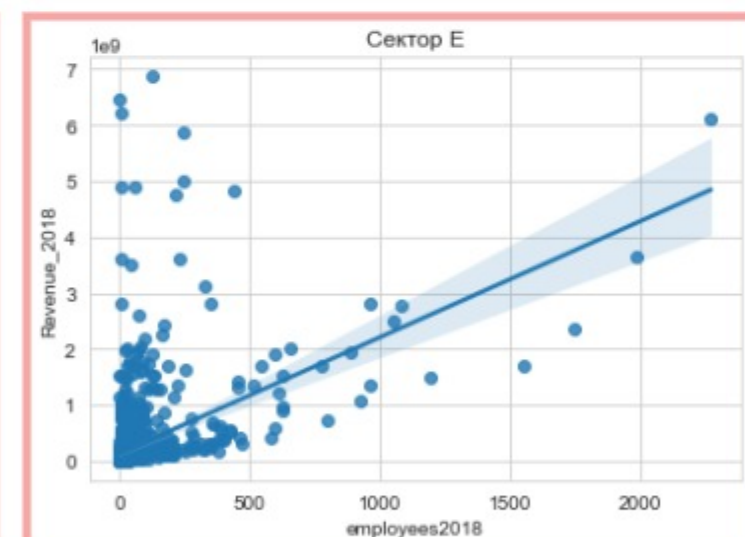
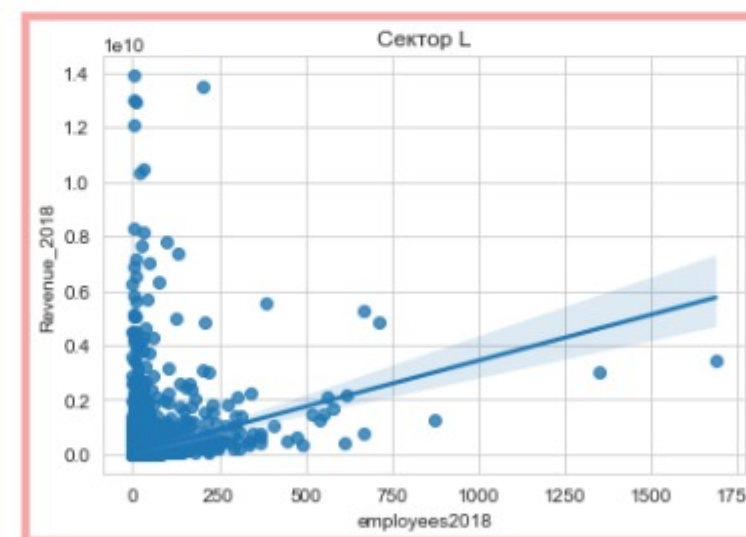
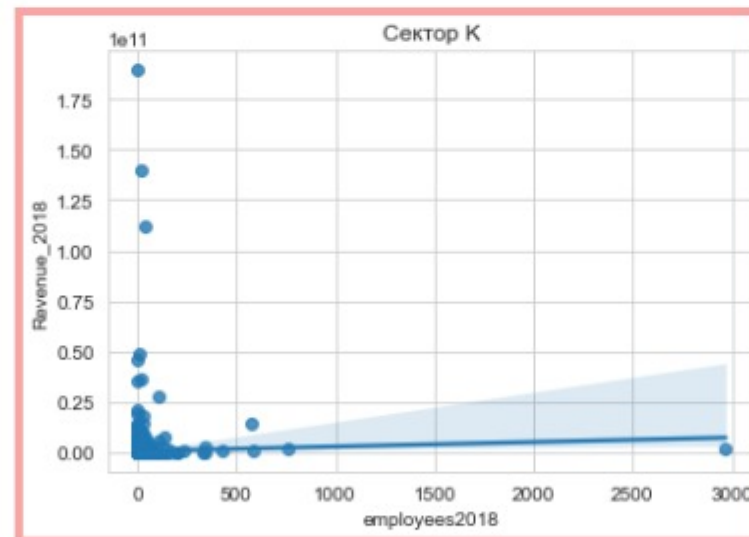
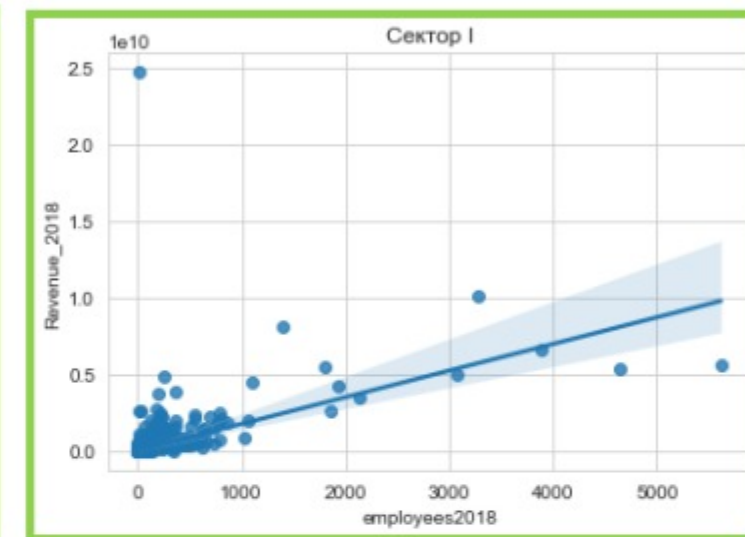
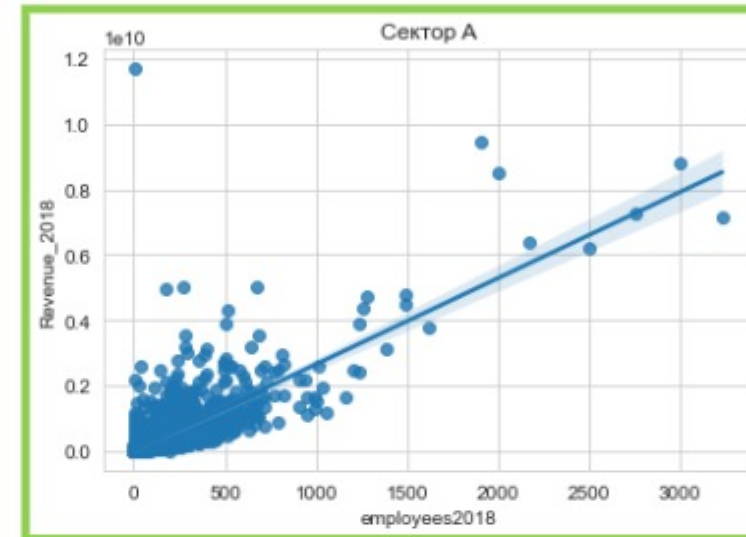
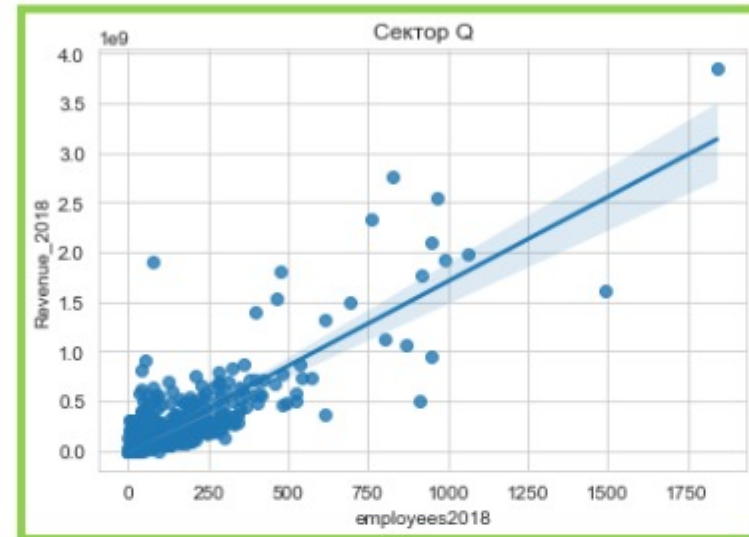
L - трудовые ресурсы

Линейность труда



Корреляции

Sector	Labour_Revenue_corr
O	0.058840
K	0.186930
L	0.364829
E	0.421627
N	0.441684
S	0.453114
P	0.466552
G	0.471995
M	0.493786
R	0.504174
H	0.542176
J	0.558407
F	0.569159
B	0.573425
D	0.598364
C	0.667128
I	0.694405
A	0.730304
Q	0.825648



Проверка индекса линейности труда

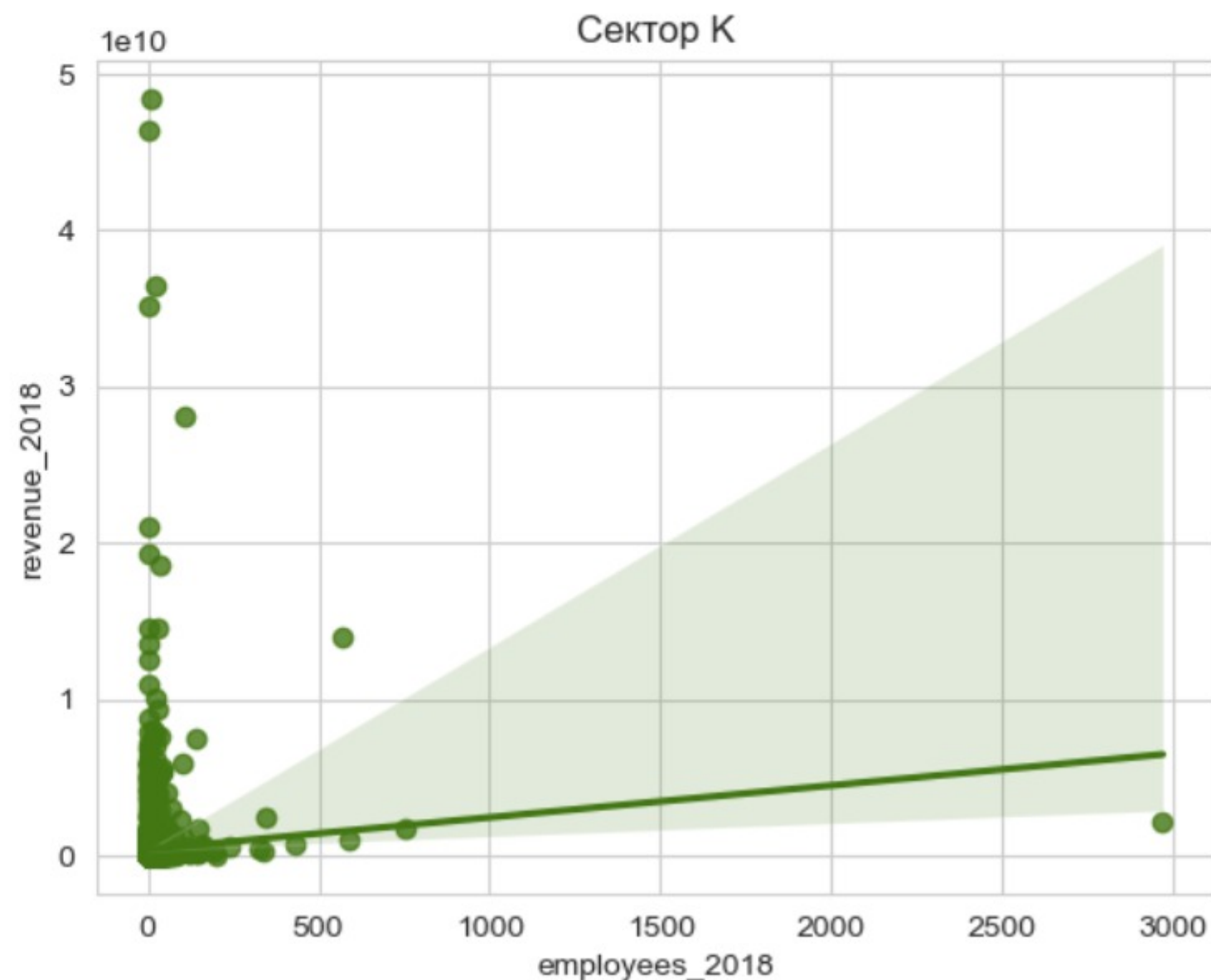
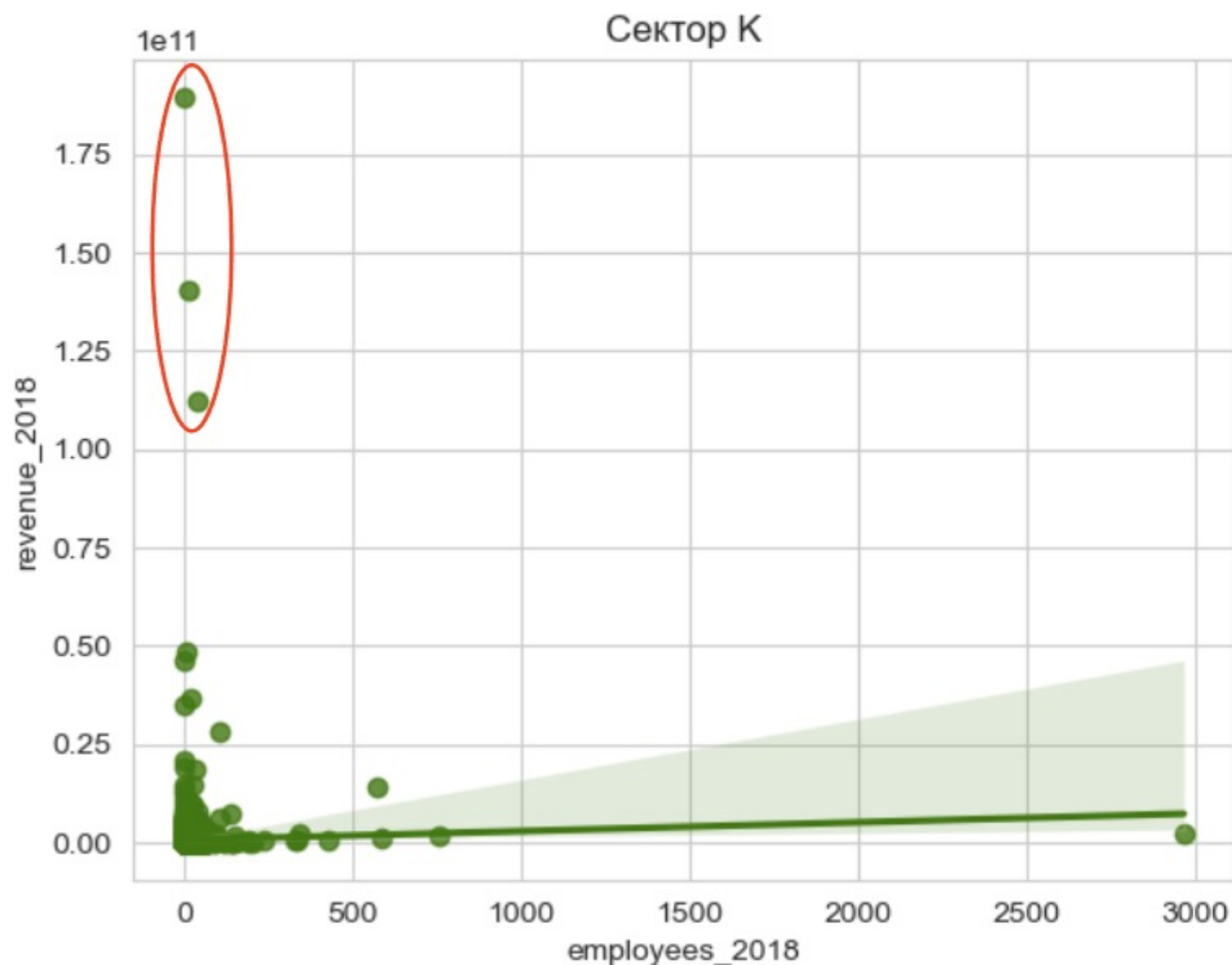
Линейные сектора

Нелинейные сектора



Устойчивость матмодели

Крайние точки не влияют на общий результат ($R1=0.34$, $R2=0.33$)



Взвешенный индекс инноваций

$$\text{Weighted Innovation Index} = \hat{\beta}_1 * N_{employees}$$

$$\hat{\beta}_1 = \frac{\sum (I_i - \bar{I})(TR_i - \bar{TR})}{\sum (I_i - \bar{I})^2}$$

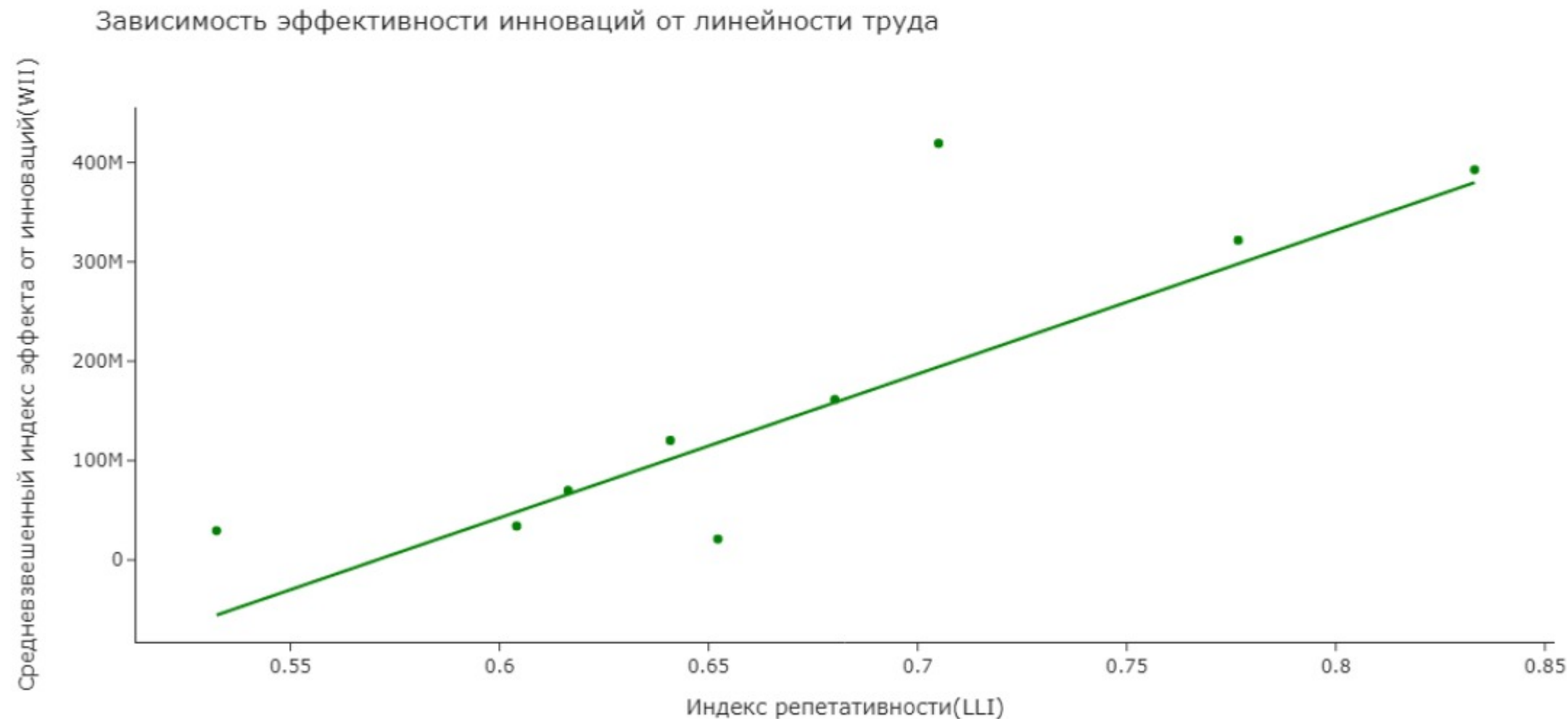
$N_{employees}$
- число работников, чел.

TR - выручка компании, руб.
 I - инновационная активность компании, у.е.

Эффект от инноваций в компаниях

H0 - коэффициент корреляции равен нулю (отклоняем)

Эффект от инноваций зависит от линейности труда
($r=0.91$, $p\text{-value}=0.0045$)



Эффект от инноваций в компаниях

1 H_0 : коэффициент корреляции равен нулю

2 H_1 коэффициент корреляции не равен нулю

$$3 \quad t = \frac{r * \sqrt{n - 2}}{\sqrt{1 - r^2}} = \frac{0,91 * 2,65}{0.415} = 5,81$$

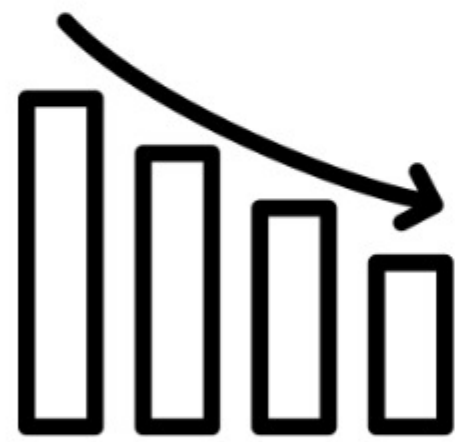
4 $t_{кр}$ - для 0,95 доверительного интервала и $n_{выборки} = 9$ составляет 3.25

5 $t > |t_{кр}| \rightarrow$ отклоняем H_0 (т.к. вероятность ошибки первого рода меньше 1го процента)

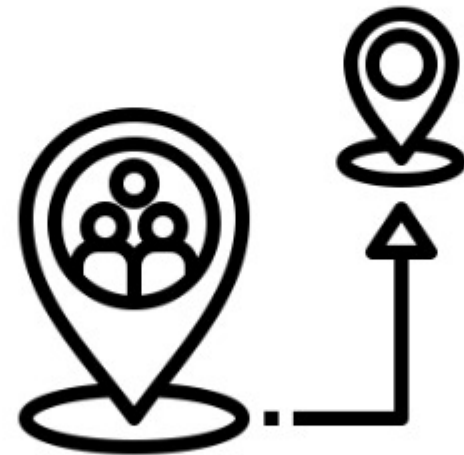
Исходная гипотеза подтвердилась

Предобработка данных	Гипотеза + механизм	Мат. модель	Наш индекс!	Проверка гипотезы	Вывод	Применение и Policy implication	Перспективы и ограничения
----------------------	---------------------	-------------	-------------	-------------------	-------	---------------------------------	---------------------------

Краткий экскурс в экономическую ситуацию



Внешний шок

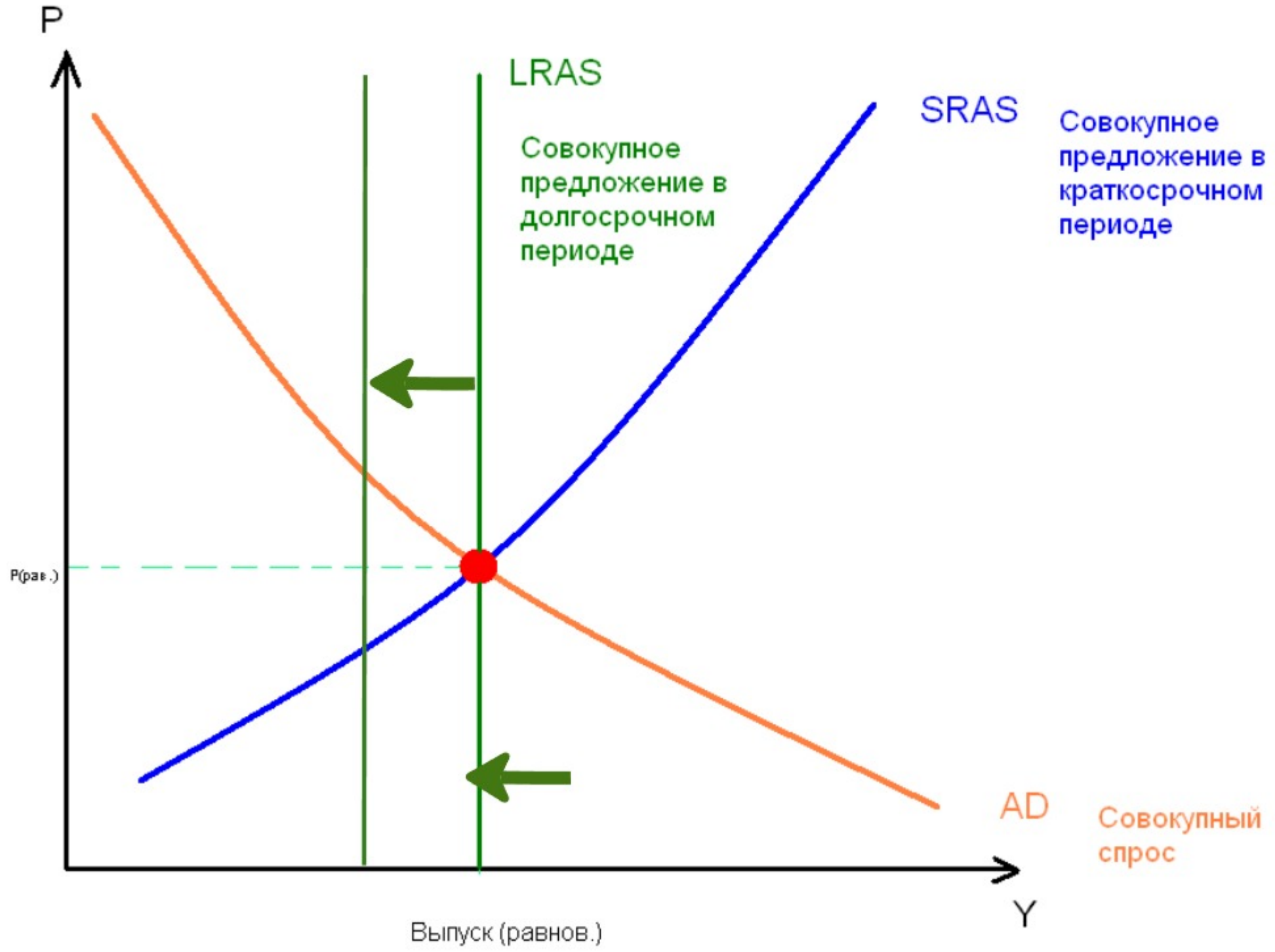


Массовая эмиграция



Неиспользуемый капитал

Модель AD AS



Закон Оукена

$$\frac{Y - \bar{Y}}{\bar{Y}} = -\beta * (U - U_{nat.})$$

Y - фактический ВВП
 \bar{Y} - потенциальный ВВП
 β - эмпирический коэф.

U - фактическая
безработица
 $U_{nat.}$ - естественная
безработица

Слишком низкая безработица — плохо

Дефицит кадров



**Издержки компаний
больше,
разгоняется инфляция**

**Меньше конкуренция на
рынке труда,
меньше
производительность**

*"Аномально низкий уровень безработицы тормозит экономику страны", — заявил 6 сентября министр экономического развития Максим Решетников.

Вывод

*вероятность ошибки
первого рода меньше 1-го
процента*

Sector	Labour_Revenue_corr
O	0.058840
K	0.186930
L	0.364829
E	0.421627
N	0.441684
S	0.453114
P	0.466552
G	0.471995
M	0.493786
R	0.504174
H	0.542176
J	0.558407
F	0.569159
B	0.573425
D	0.598364
C	0.667128
I	0.694405
A	0.730304
Q	0.825648

фин сектор → K
недвиж → L
водоснабж → E

гостиные → I
с/х → A
соц. услуги → Q

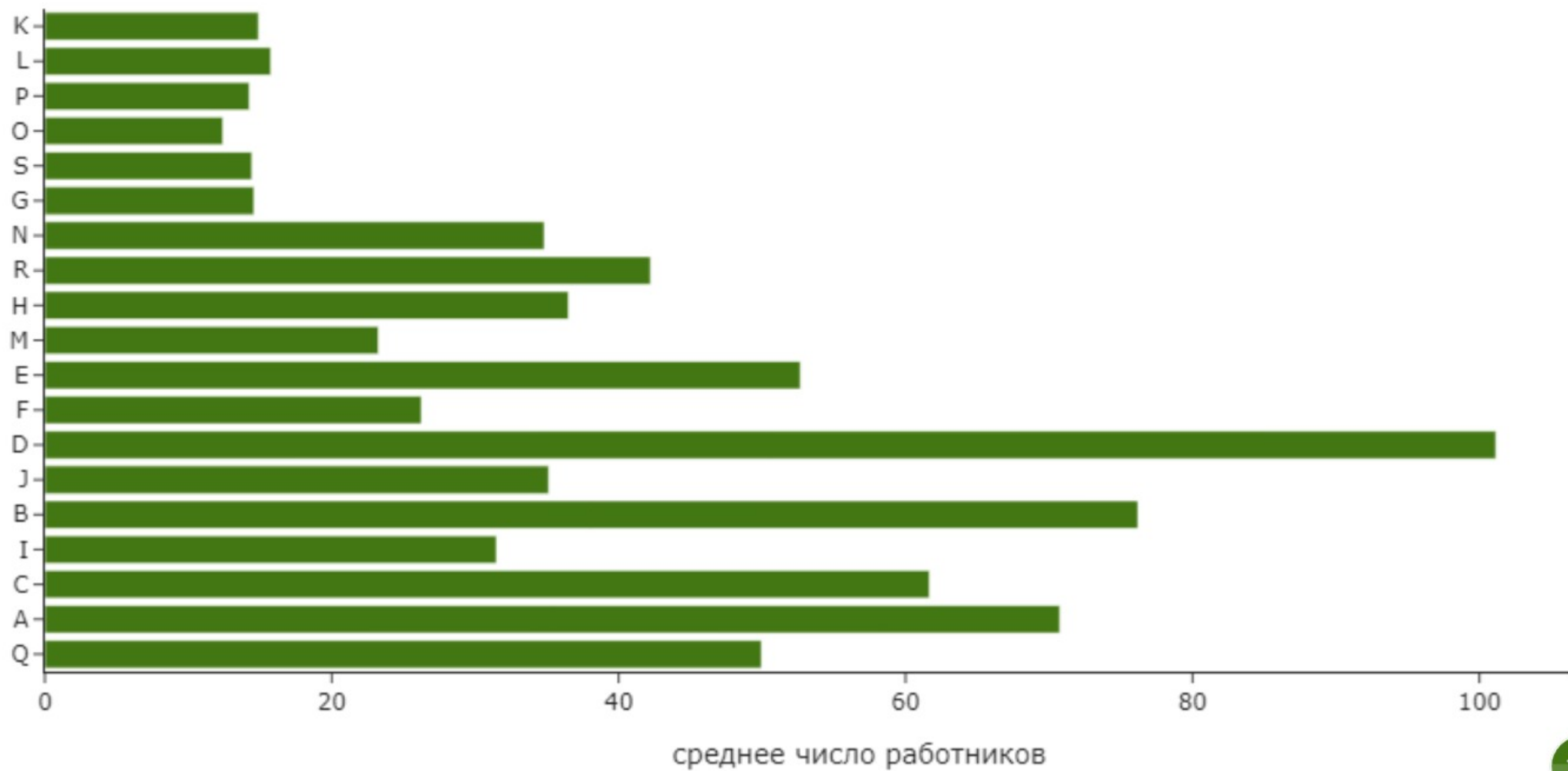
Policy Implication



Линейность труда



отрасль



Механизм применения

Стимулируем
инновациями
линейные
отрасли



Производительность
труда растет



Требуется
меньше труда



Эффективно
улучшается
ситуация на
рынке труда



Конкуренция
на рынке труда возрастает,
и повышается доступность
трудовых ресурсов в
экономике



Трудовой ресурс
высвобождается

Итоговый эффект

Нормализация
безработицы

Снижение
инфляции

Повышение
эффективности
труда

Ограничения

Маленький размер выборки

Есть данные только за 1-2 года

У каждой компании даны не все показатели

Перспективы исследования

Увеличить число компаний и улучшить качество данных о них

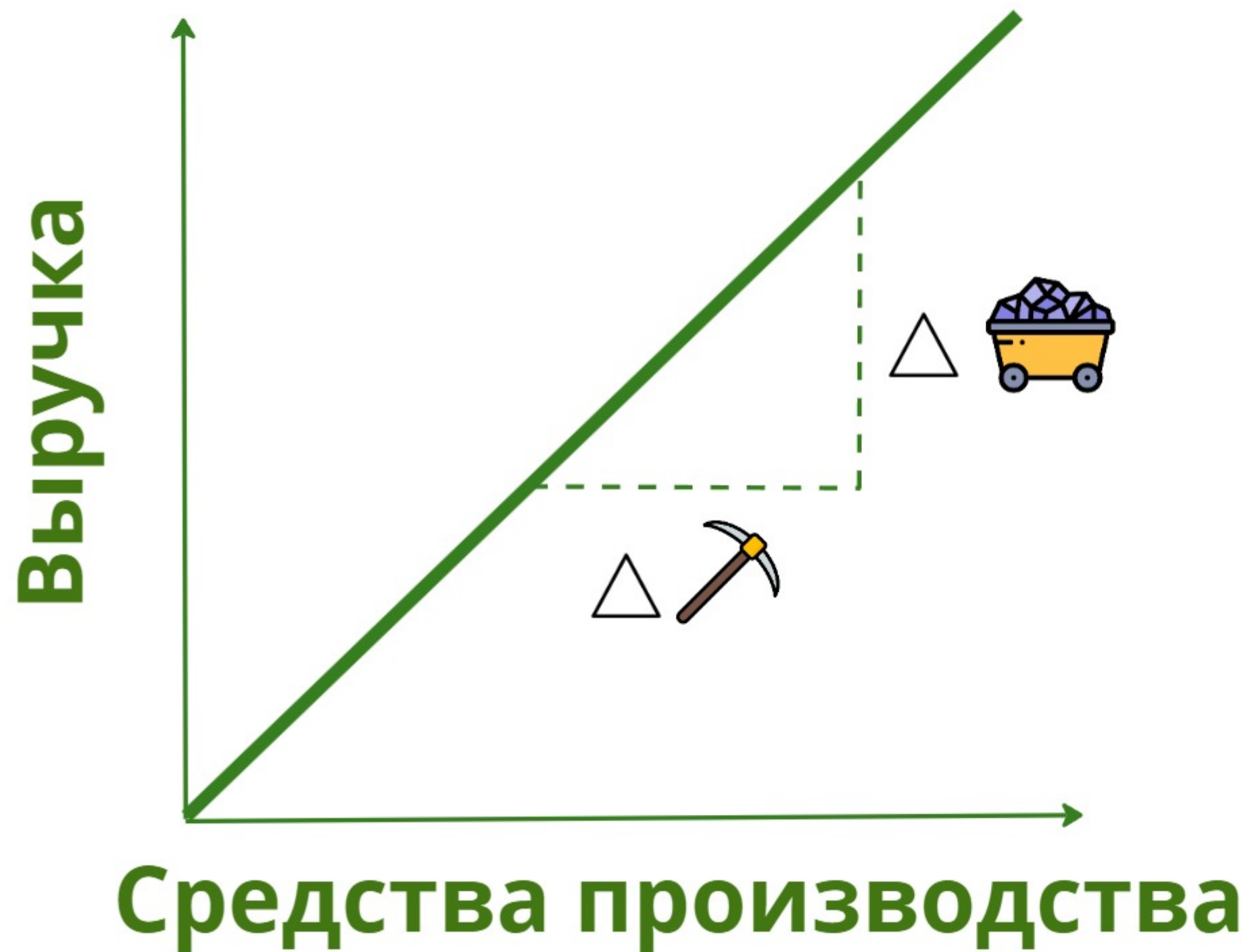
Получить данные за более длительный период

Получить дополнительные показатели о деятельности компании

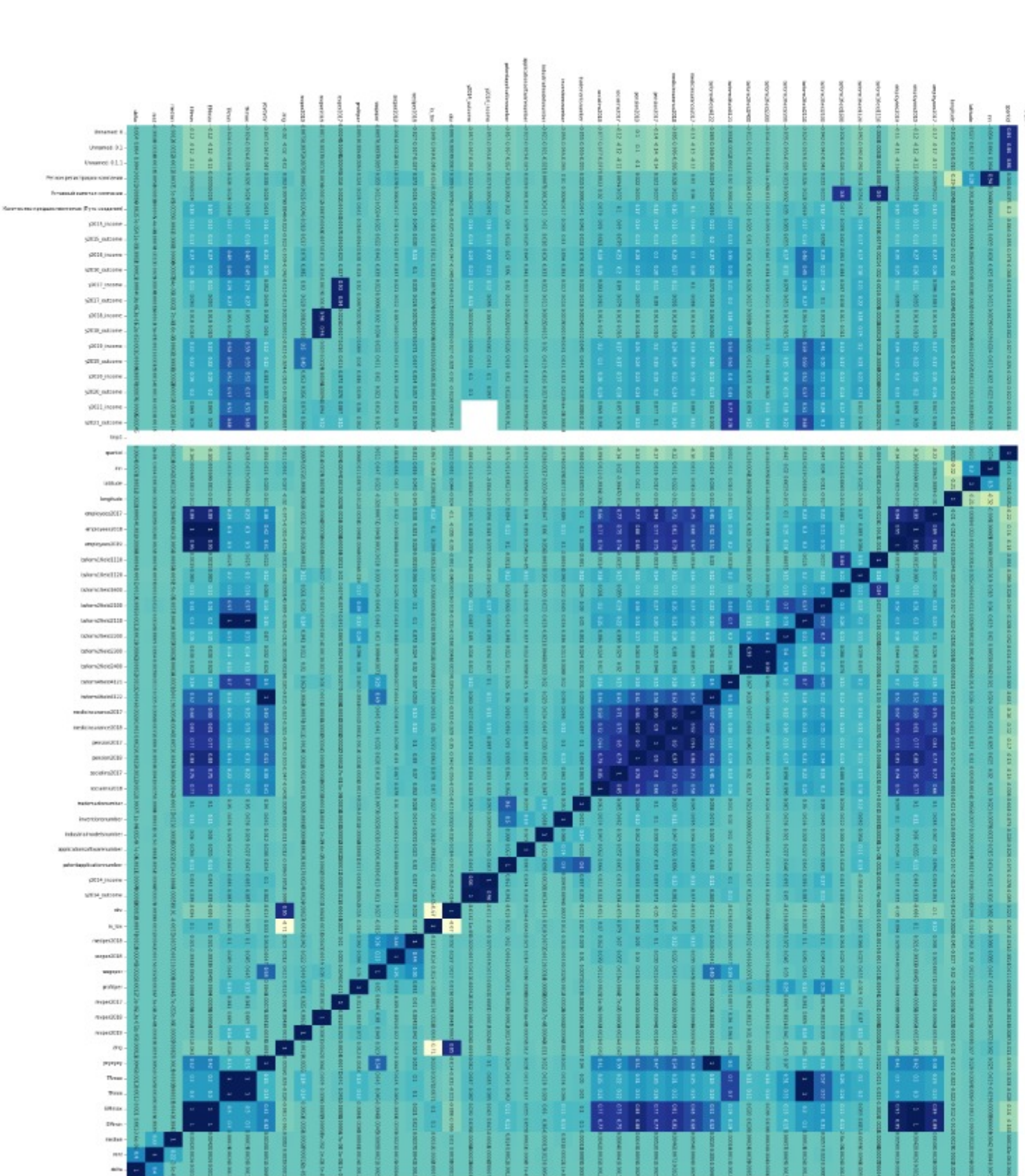
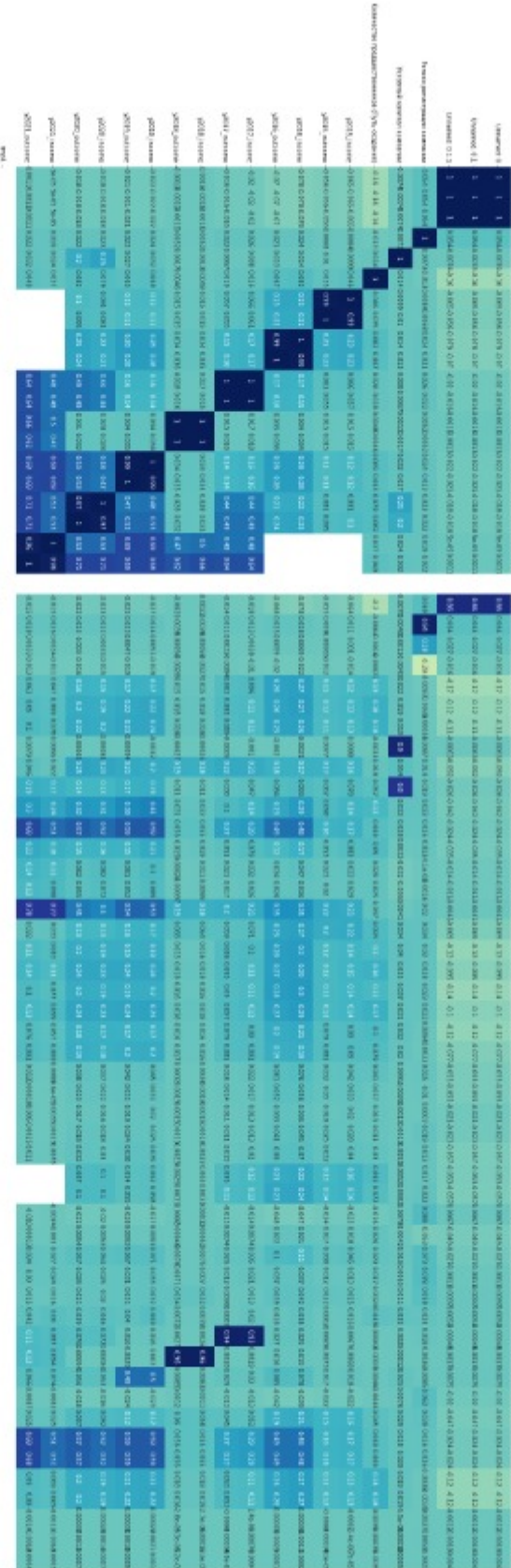
Возможность использовать расширенный индекс линейности труда для предсказания поведения компании в разных экономических ситуациях

ПРИЛОЖЕНИЯ

Линейность капитала



Корреляционный анализ



Очистка выбросов

Представление
компаний в
виде векторов

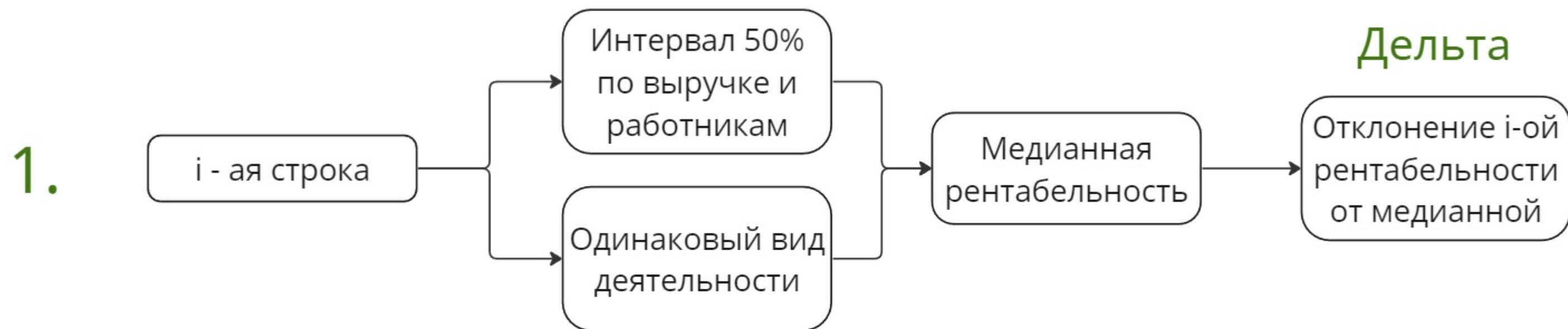


Представление
компаний по
группам на
основании схожести

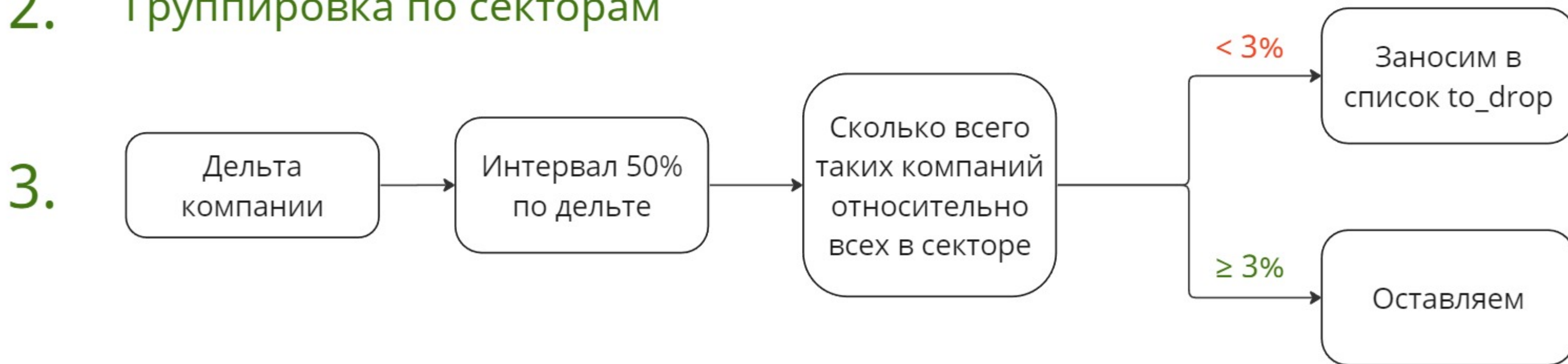
- Если компания работает в том же регионе, занимается похожей деятельностью, получают ~равную выручку, имеют ~равное к-во персонала, то компании похожи

Как были очищены выбросы? (Алгоритм)

Фильтр на компании

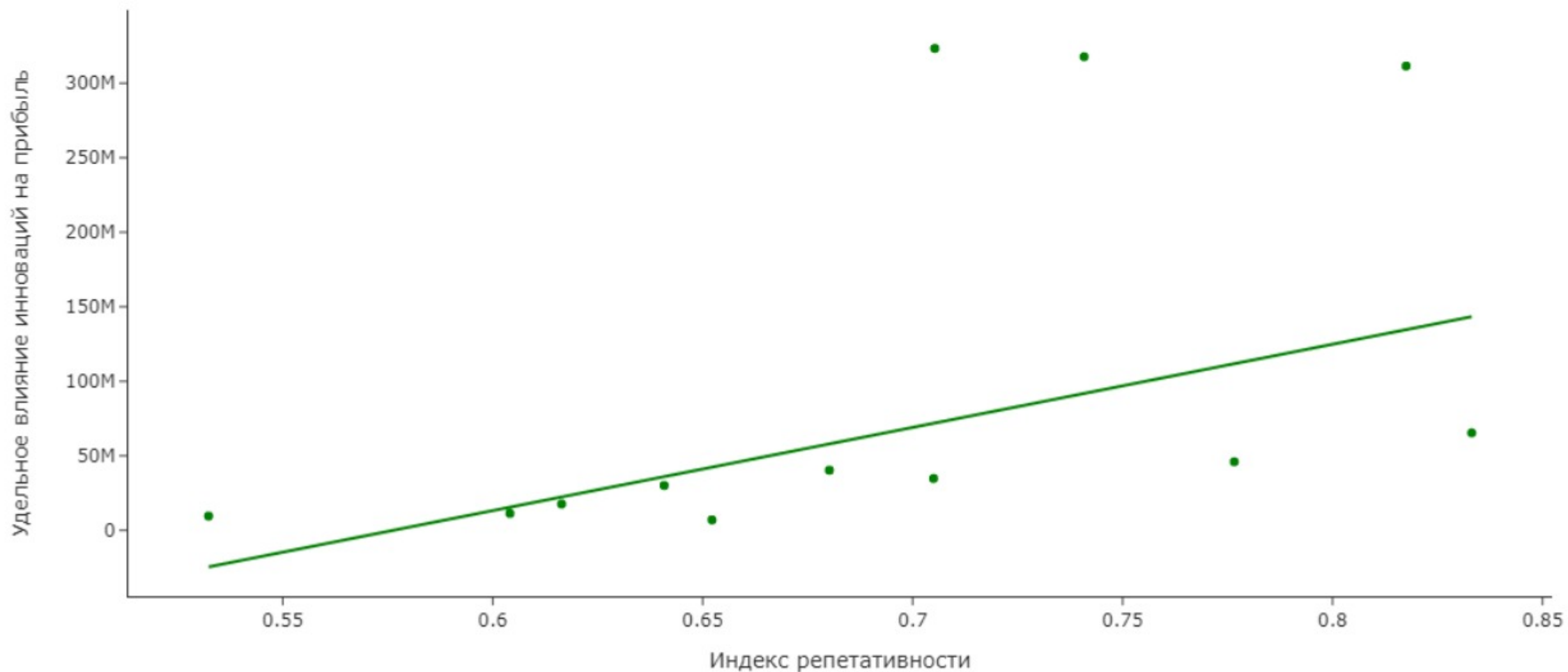


2. Группировка по секторам

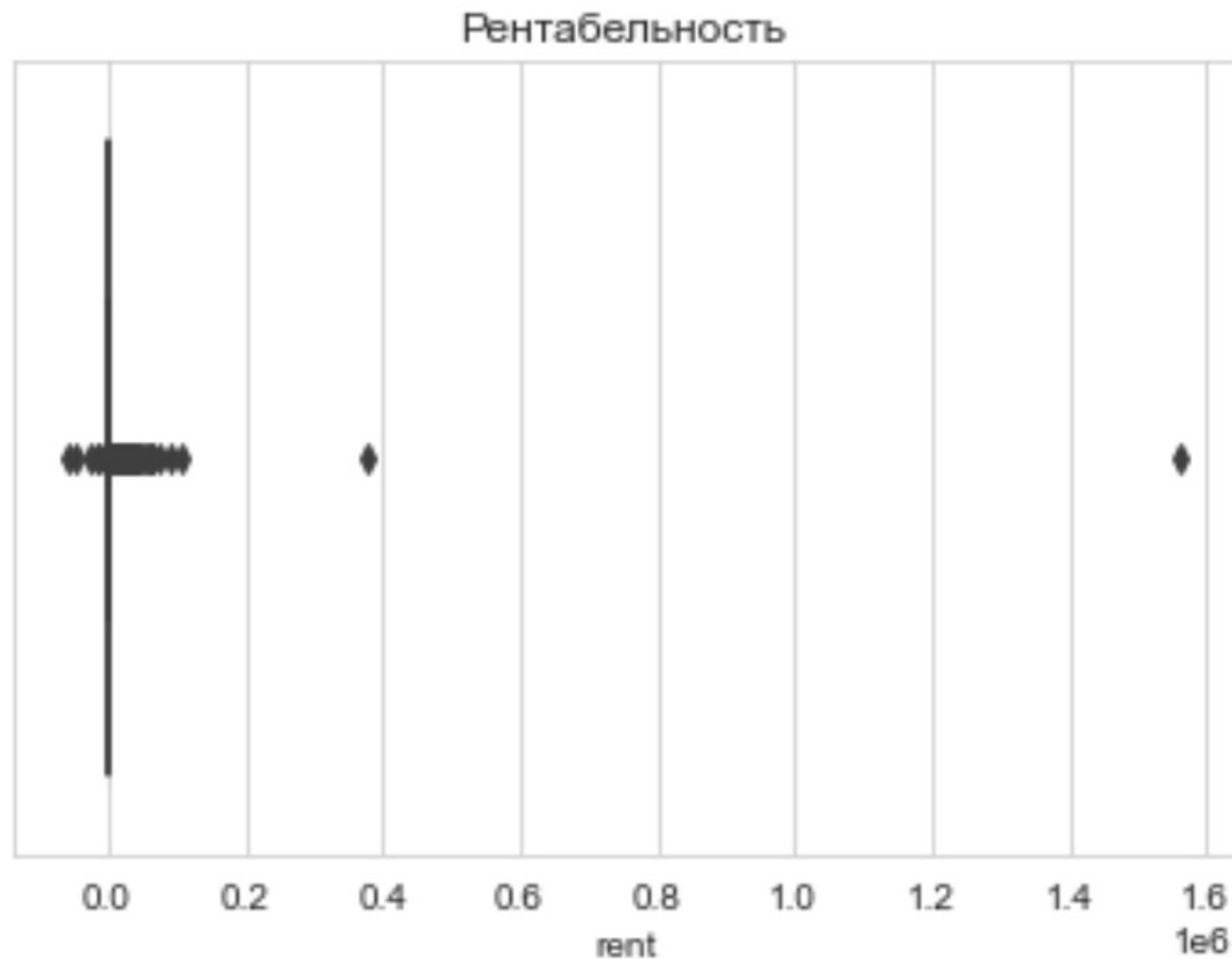
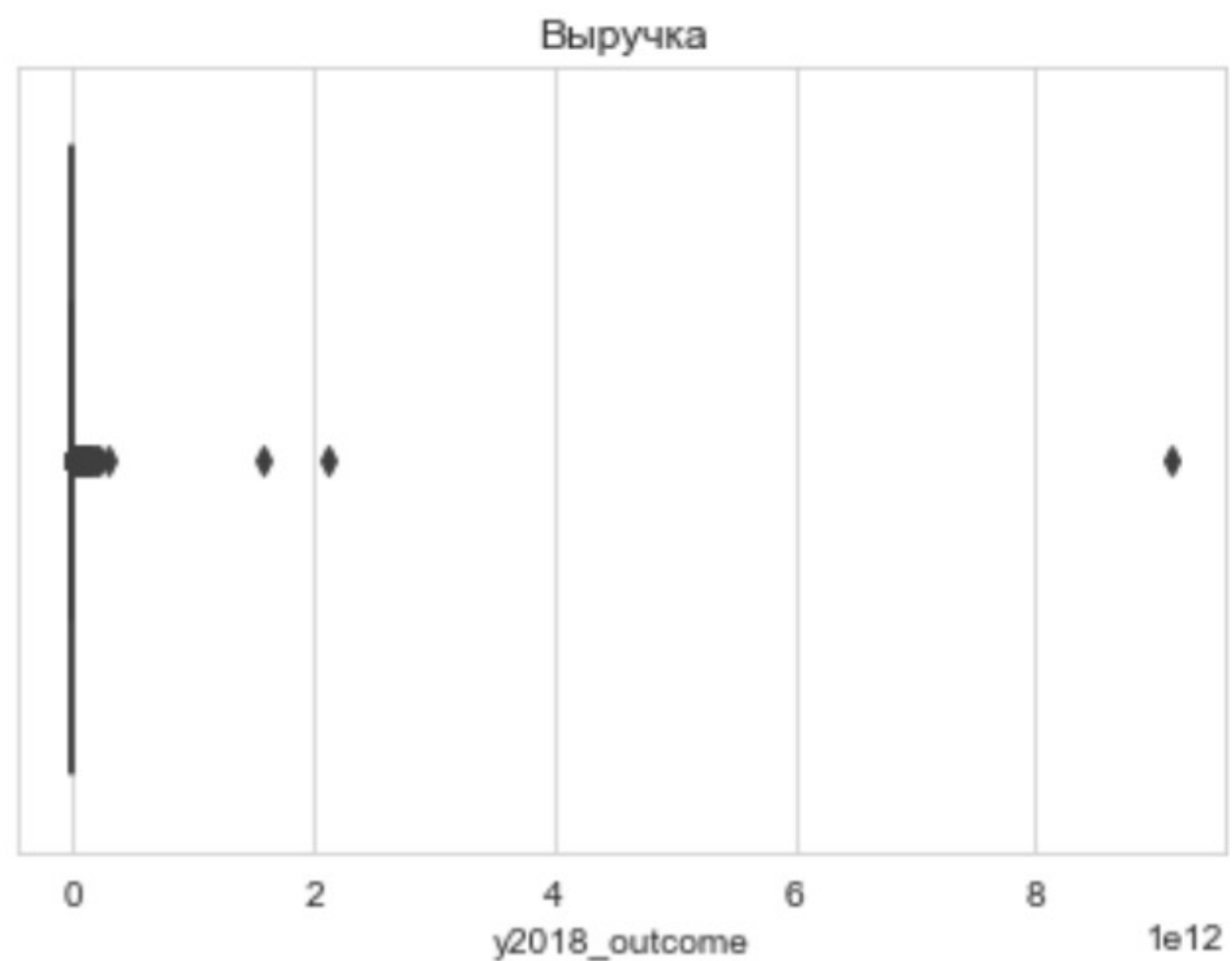


Инновации в нелинейных компаниях

Эффект от инноваций для линейных компаний возрастает (даже несколько наоборот, $r=0.55$)

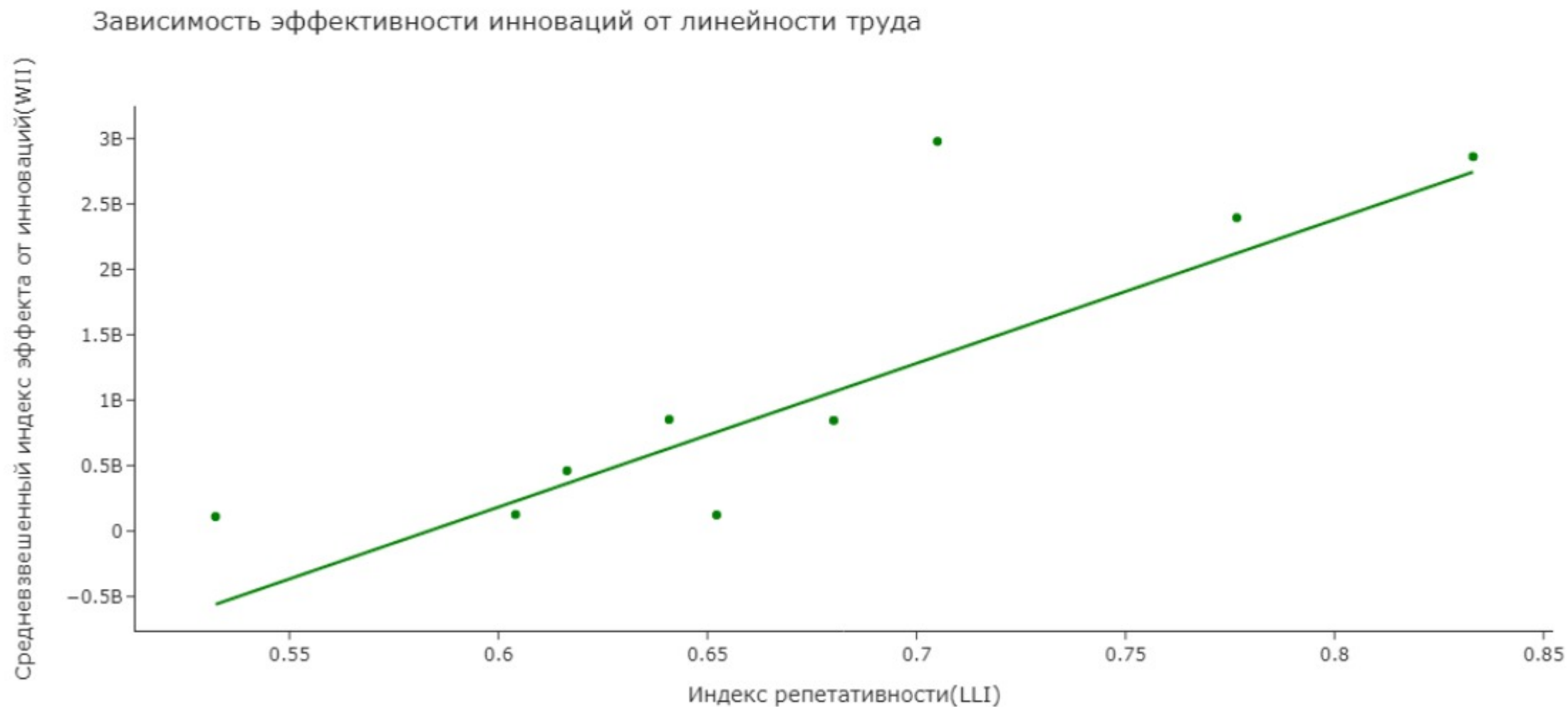


Описательная статистика



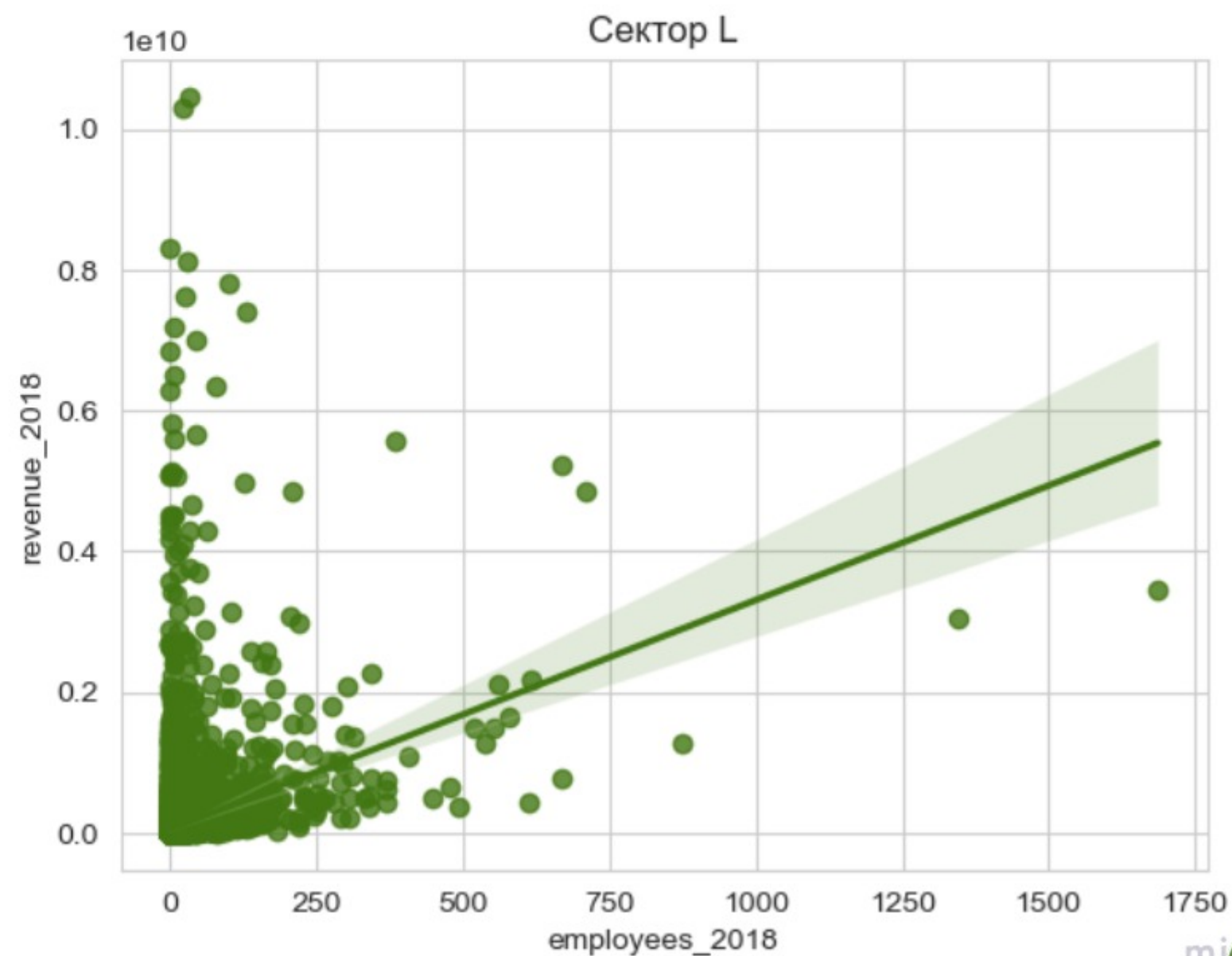
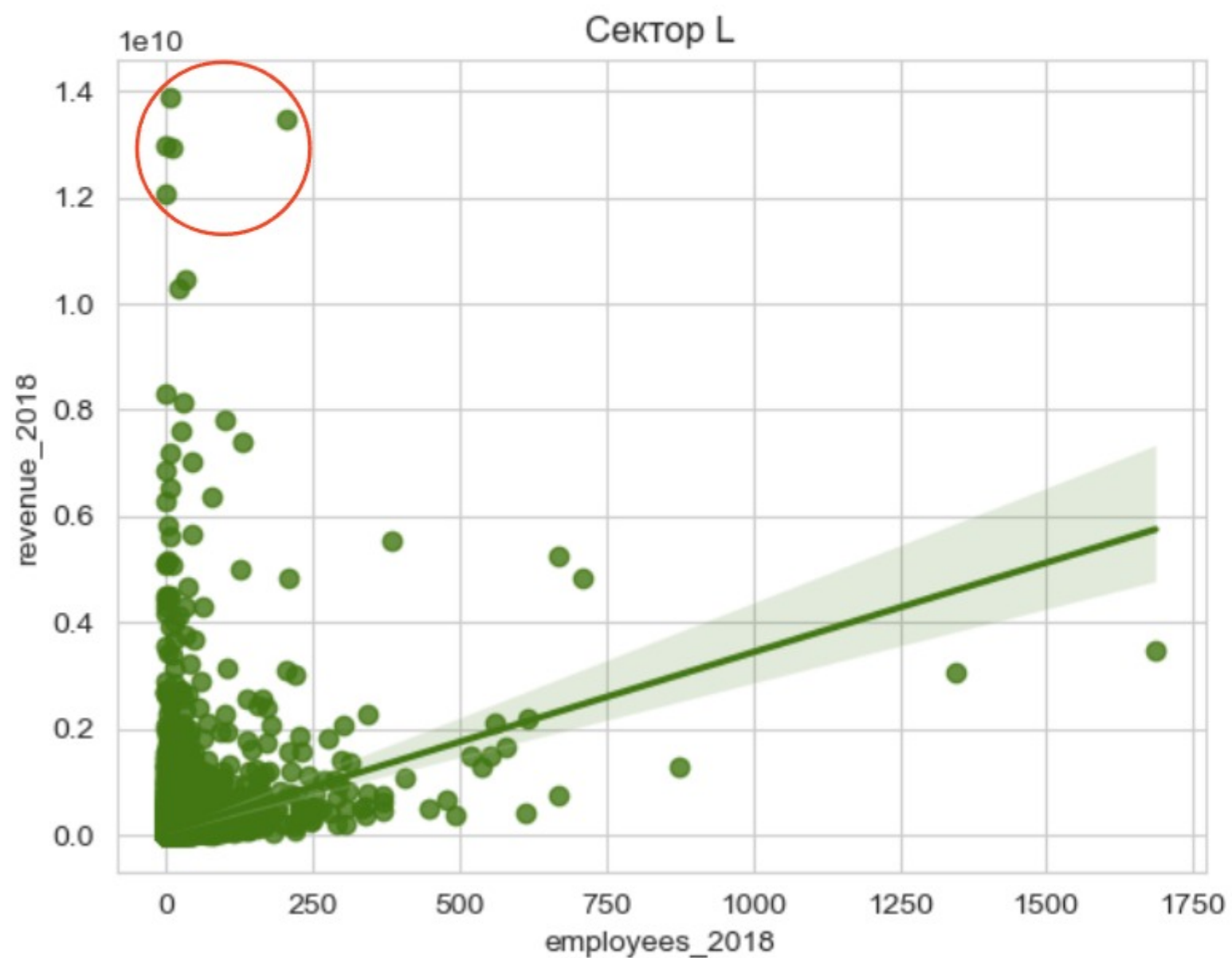
Эффект от инноваций в компаниях

Эффект от инноваций зависит от линейности труда
(Тоже самое, но число работников по отрасли среднее, а не медианное)



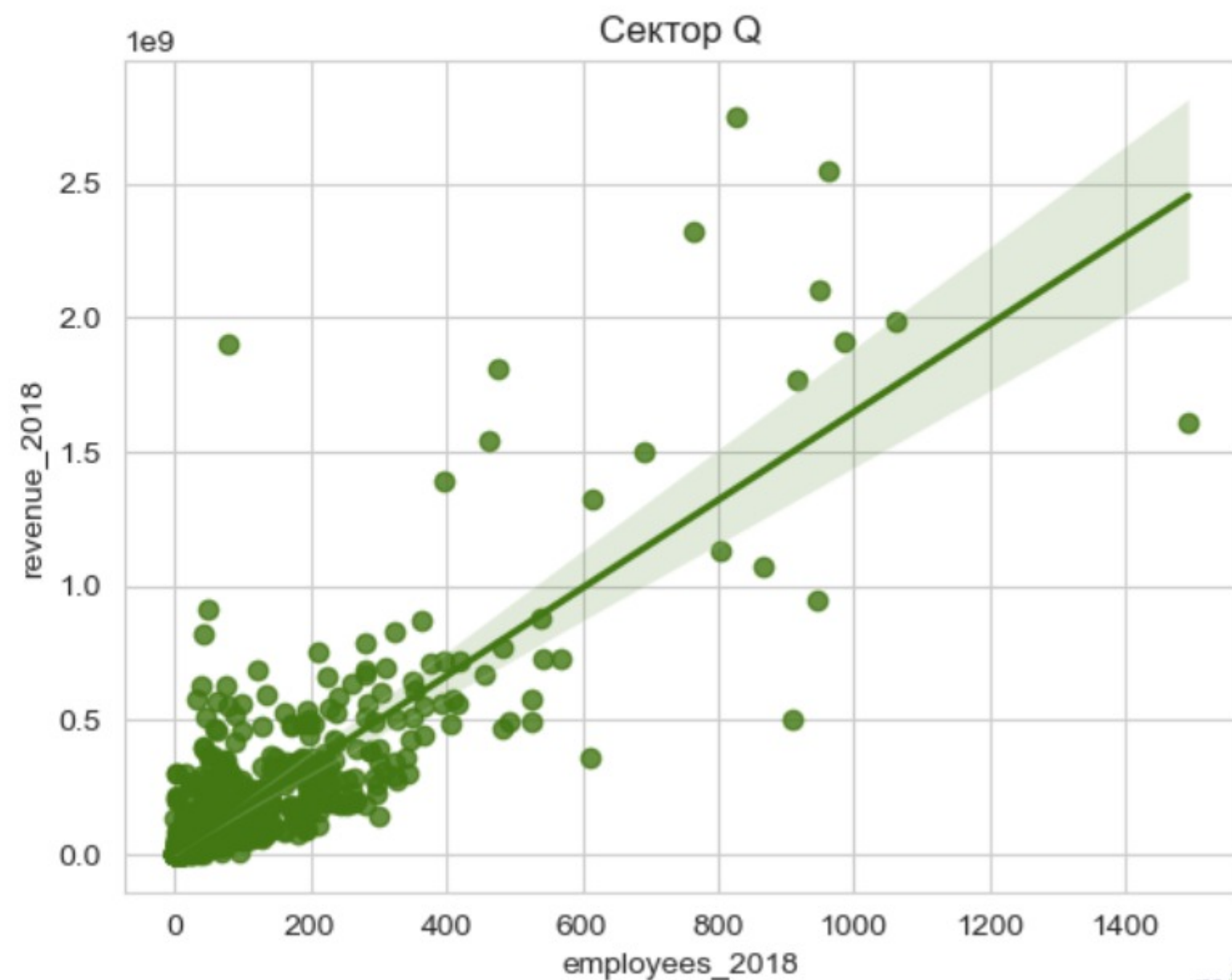
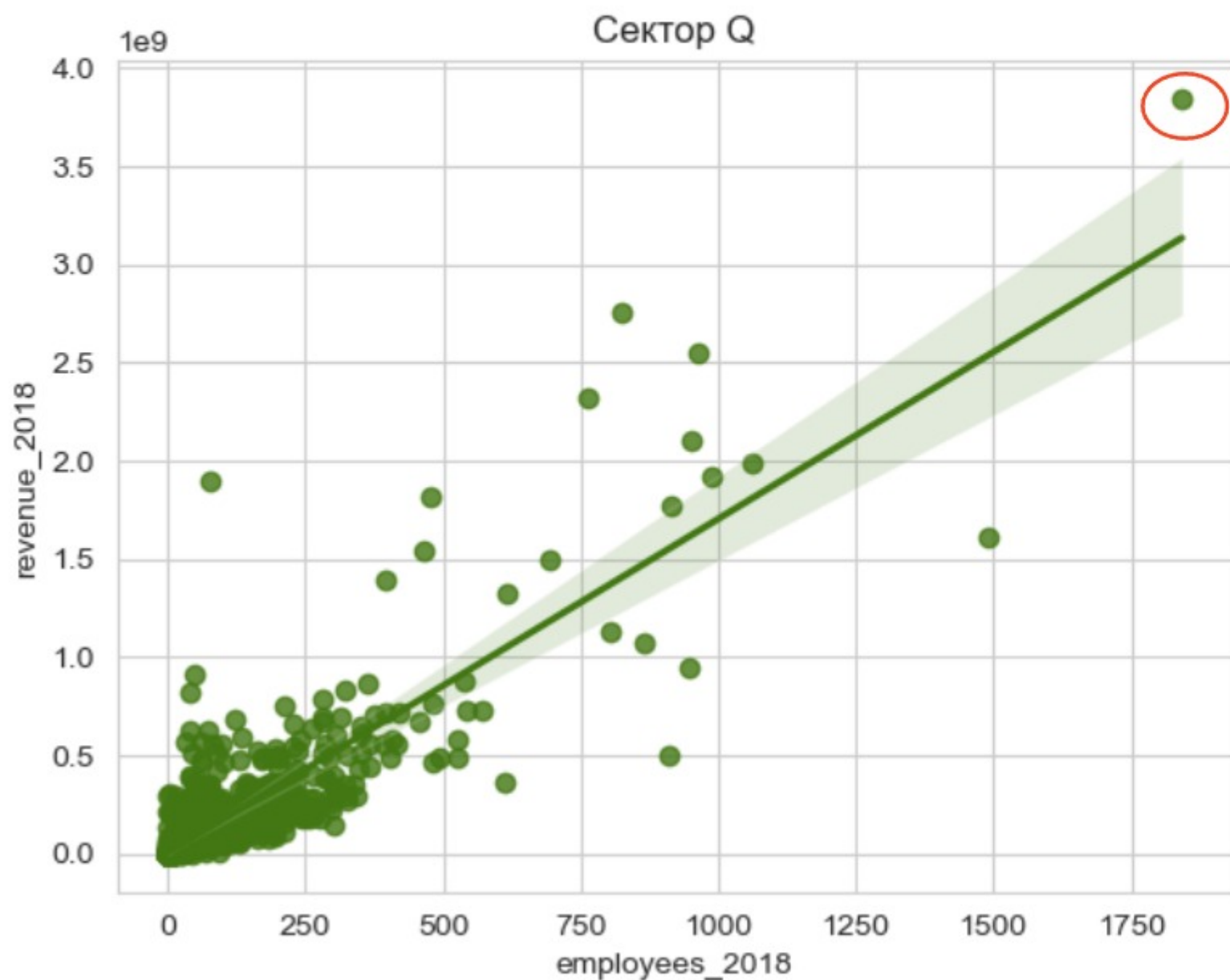
Устойчивость матмодели

Крайние точки не влияют на общий результат ($R1=0.49$, $R2=0.49$)

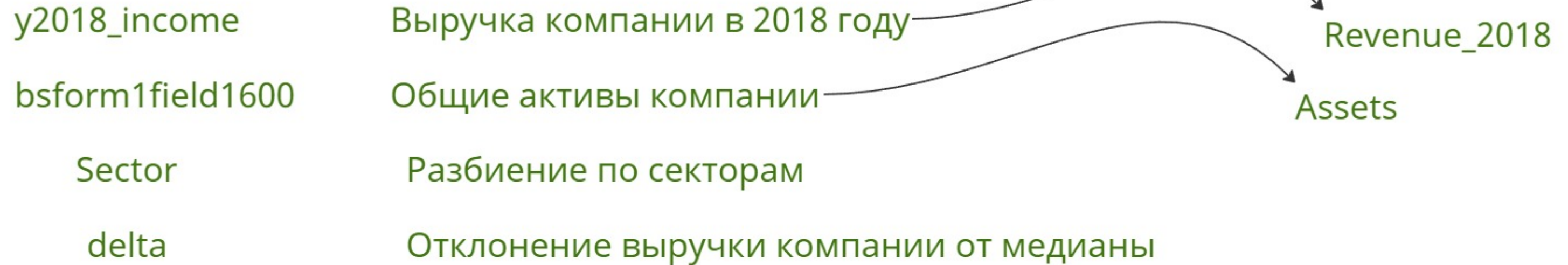


Устойчивость матмодели

Крайние точки не влияют на общий результат ($R1=0.84$, $R2=0.84$)



Предобработка данных



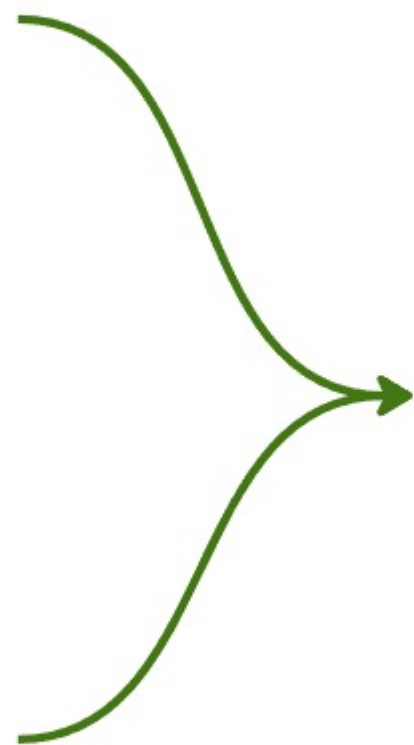
Удаляем выбросы

Смотрим, насколько сильно отклонилась компания от аналогичных в секторах, используя к/ф `delta`

`.drop`
~ у 700 компаний слишком сильно отклоняются выручки

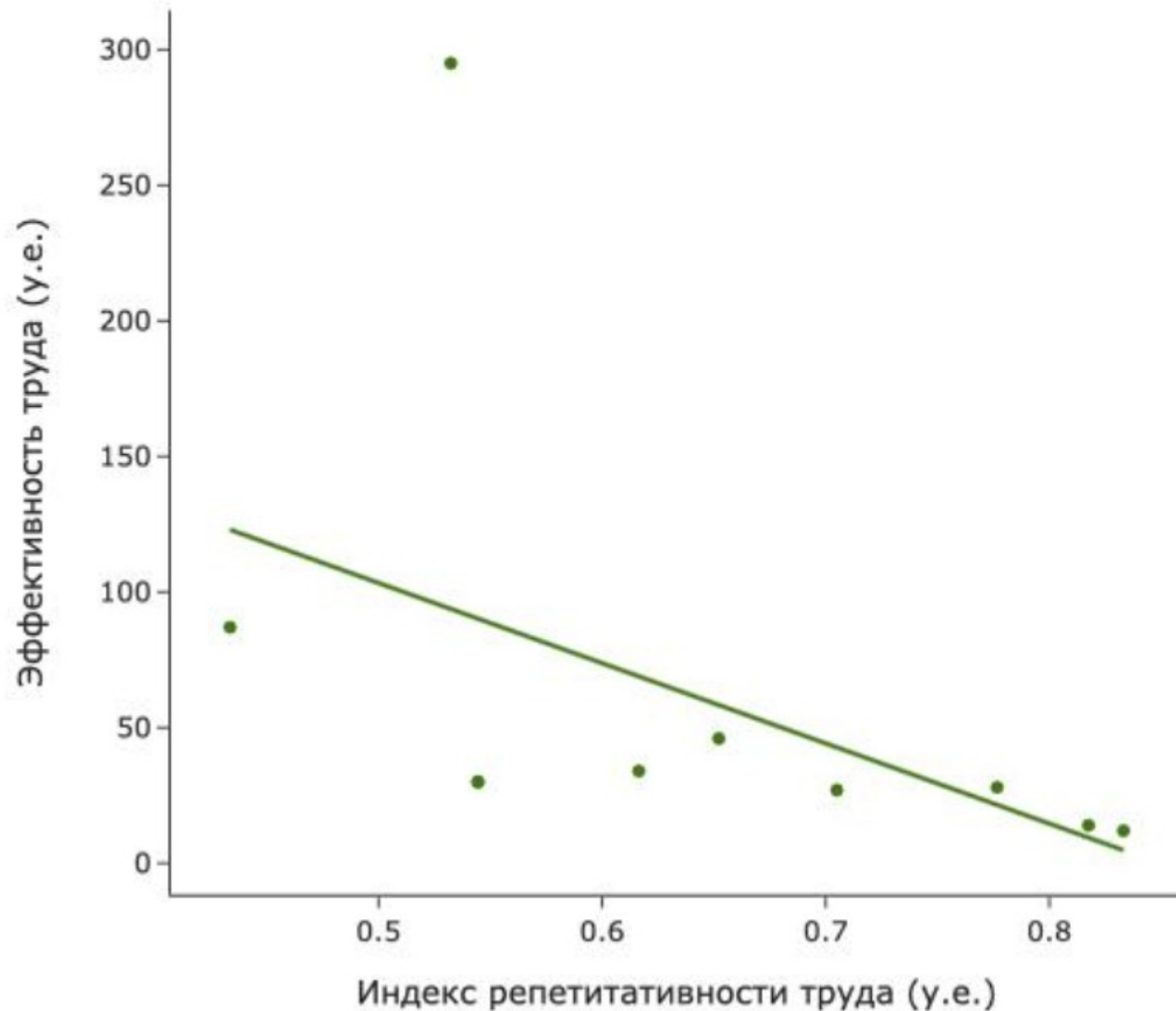
Связь КПД и отрасли

Наш индекс
линейности труда
Данные
исследований по
ВВП/работник/час
в рамках отрасли



В нелинейных
отраслях трудовые
ресурсы
используются
более эффективно

Подтверждение гипотезы



Эффективность труда -
международный индекс
продуктивности,
подсчитанный по
отраслям

Будущее проекта

Наше исследование - MVP



Более точный индекс на основании лучшего набора данных и разбиения

Отличие корреляций Компаний в секторах

более чем

не менее

в 3 раза

3000

отличаются
корреляции

компаний в
каждом секторе

Механизм

Изобретения, создаваемые в линейных компаниях, имеют большую практическую применимость, не пытаются решить фундаментальную задачу, поэтому имеют большее влияние на КПД каждого сотрудника и на рентабельность компании, соответственно.

Гипотеза

Линейные компании меньше заботятся о своих сотрудниках (делают меньшие дополнительные отчисления в ДМС, ПФР, и т.д.)

Механизм

Линейные компании меньше заботятся о своих сотрудниках, так как им проще найти замену; работодателям не нужно создавать уникальные условия, чтобы привлечь лучших специалистов.

Линейность капитала

Sector	Assets_Revenue_c orr
O	0.117127
K	0.514947
L	0.544002
E	0.531781
N	0.589406
S	0.555405
P	0.637891
G	0.678565
M	0.454895
R	0.528160
H	0.642161
J	0.634546
F	0.568518
B	0.526566
D	0.601215
C	0.731491
I	0.504683
A	0.775595
Q	0.615727