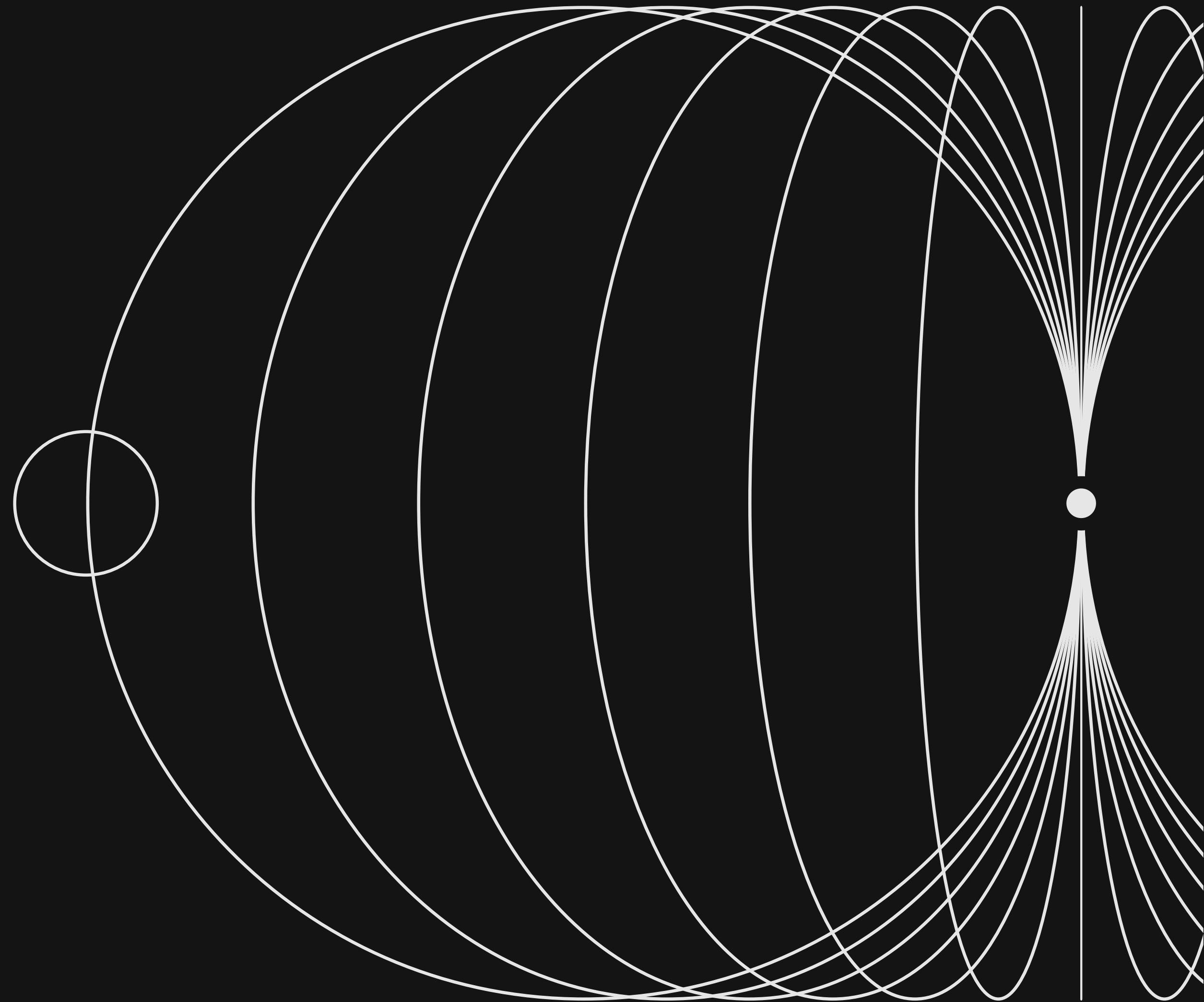


Causal Inference

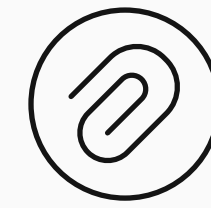
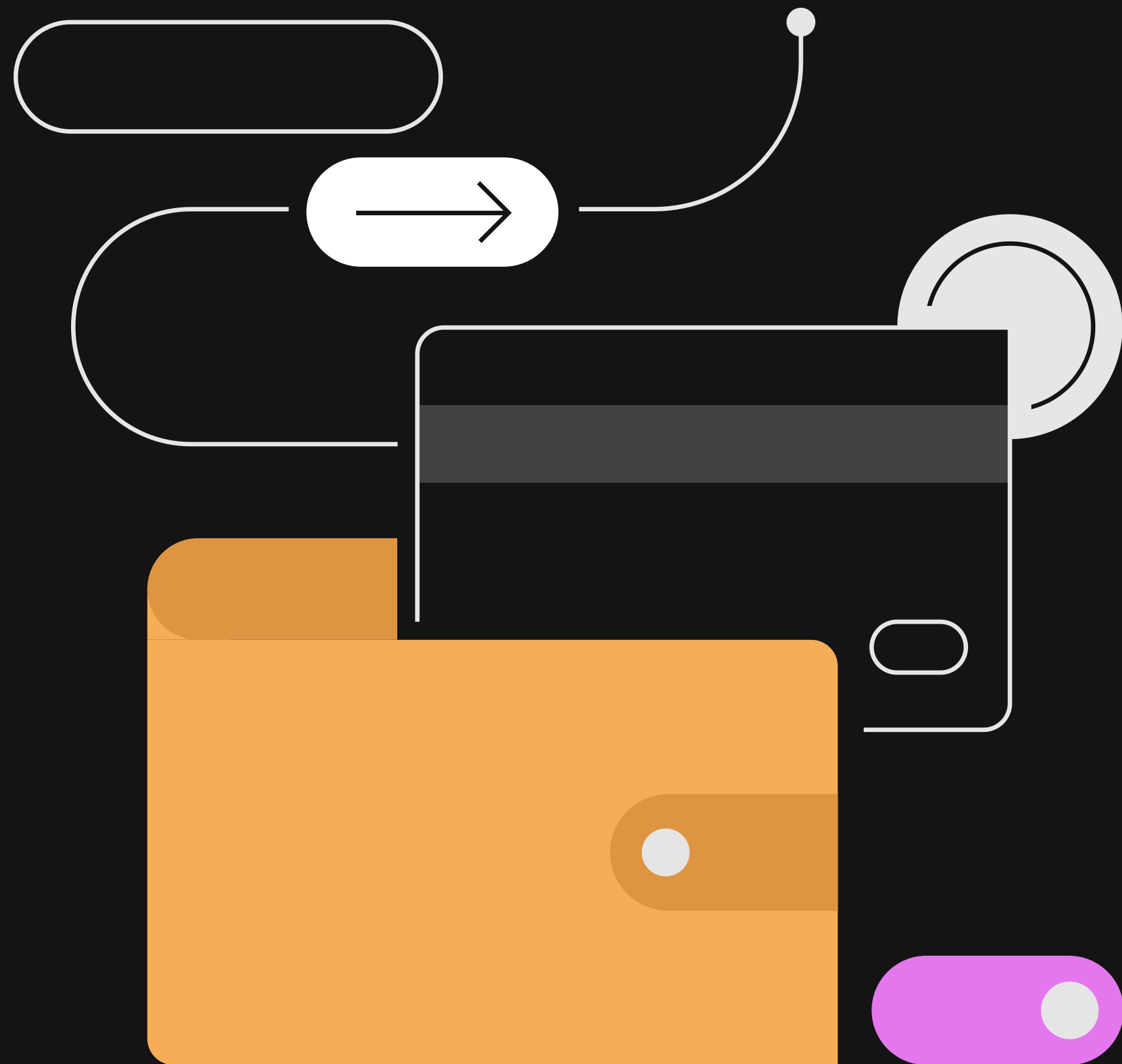
RDD, SCM, Diff in Diff

ПРОДВИНУТЫЕ МЕТОДЫ А/В-ТЕСТИРОВАНИЯ

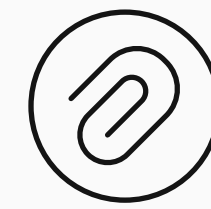
НЕДЕЛЯ 7



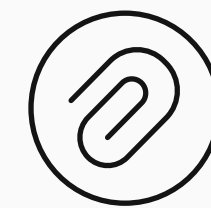
План



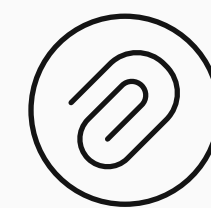
Почему не стандартный A/B тест?



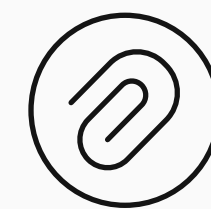
Regression Discontinuity Design



Synthetic Control Method

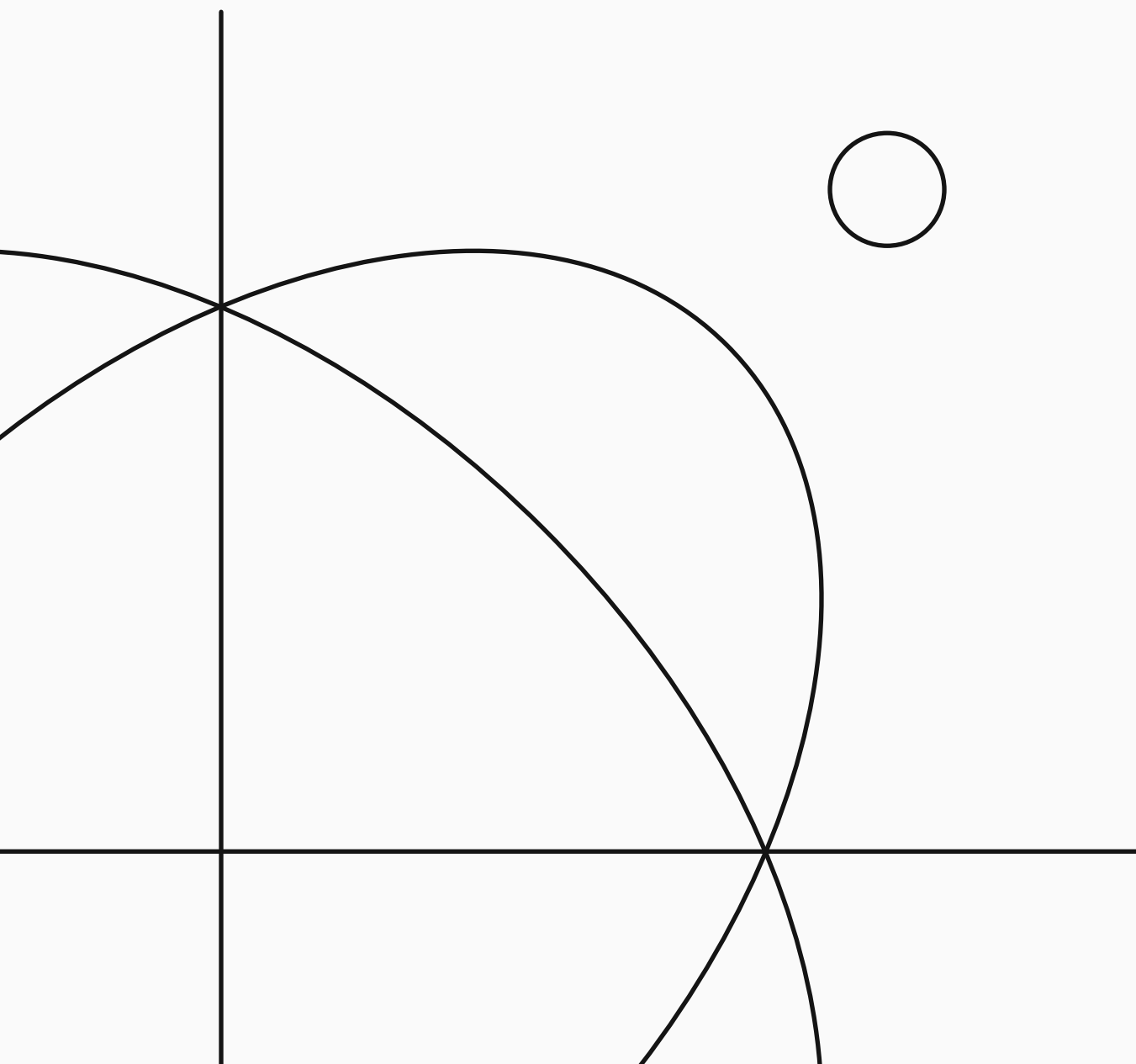


Diff-in-Diff



Сравнение методов

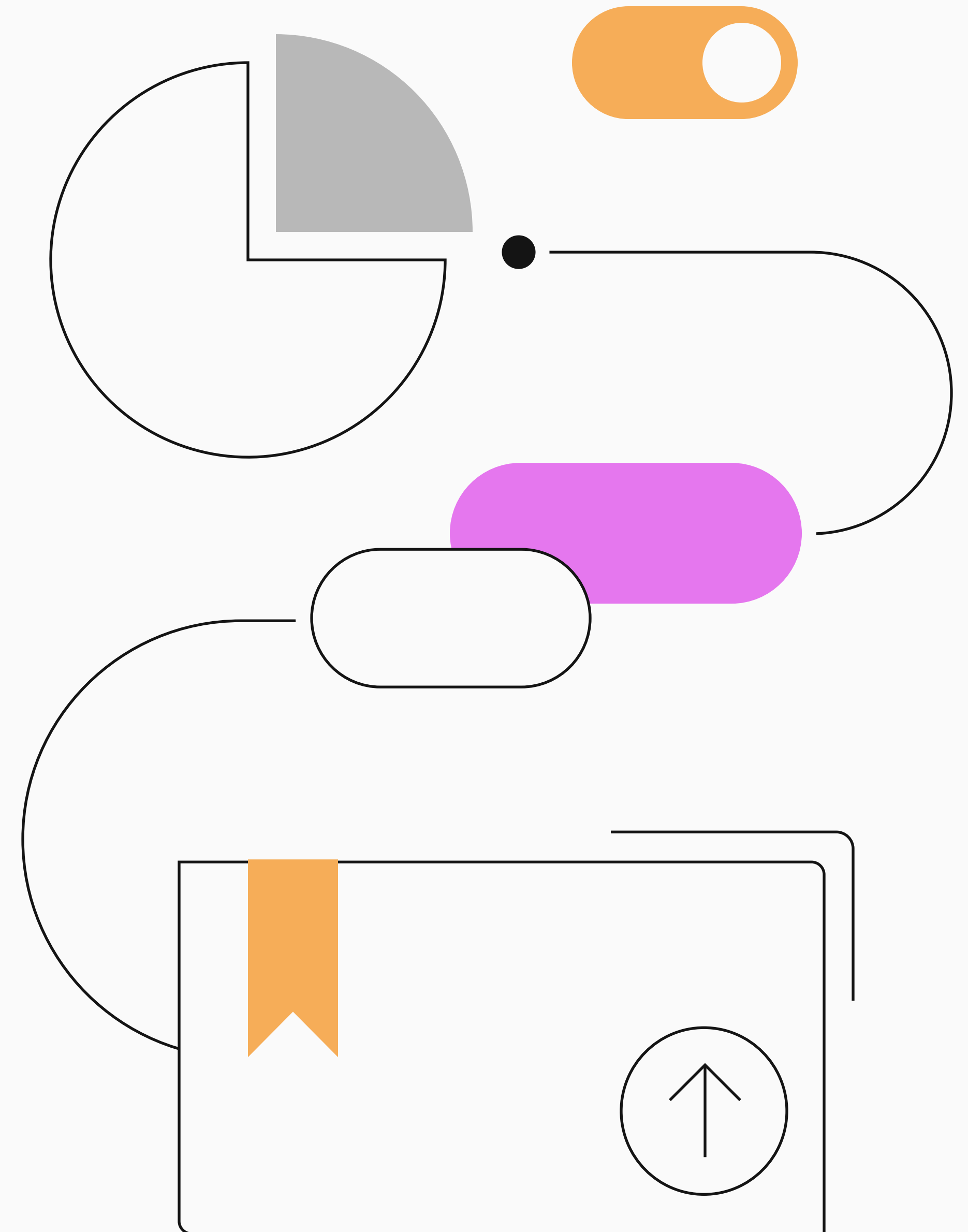
Почему не стандартный A/B тест?



- Необходимость оценки причинно-следственных связей
- Нет возможности распределения в группы теста/контроля
- Риск наличия искажений оценки причинного эффекта (смешанное влияние факторов)
- Требование выполнения заданных предположений
- Временные ряды

Playstation Store предлагает подписчикам, у которых стандартная подписка истекает в течение 30 календарных дней, скидку на новую версию премиум.

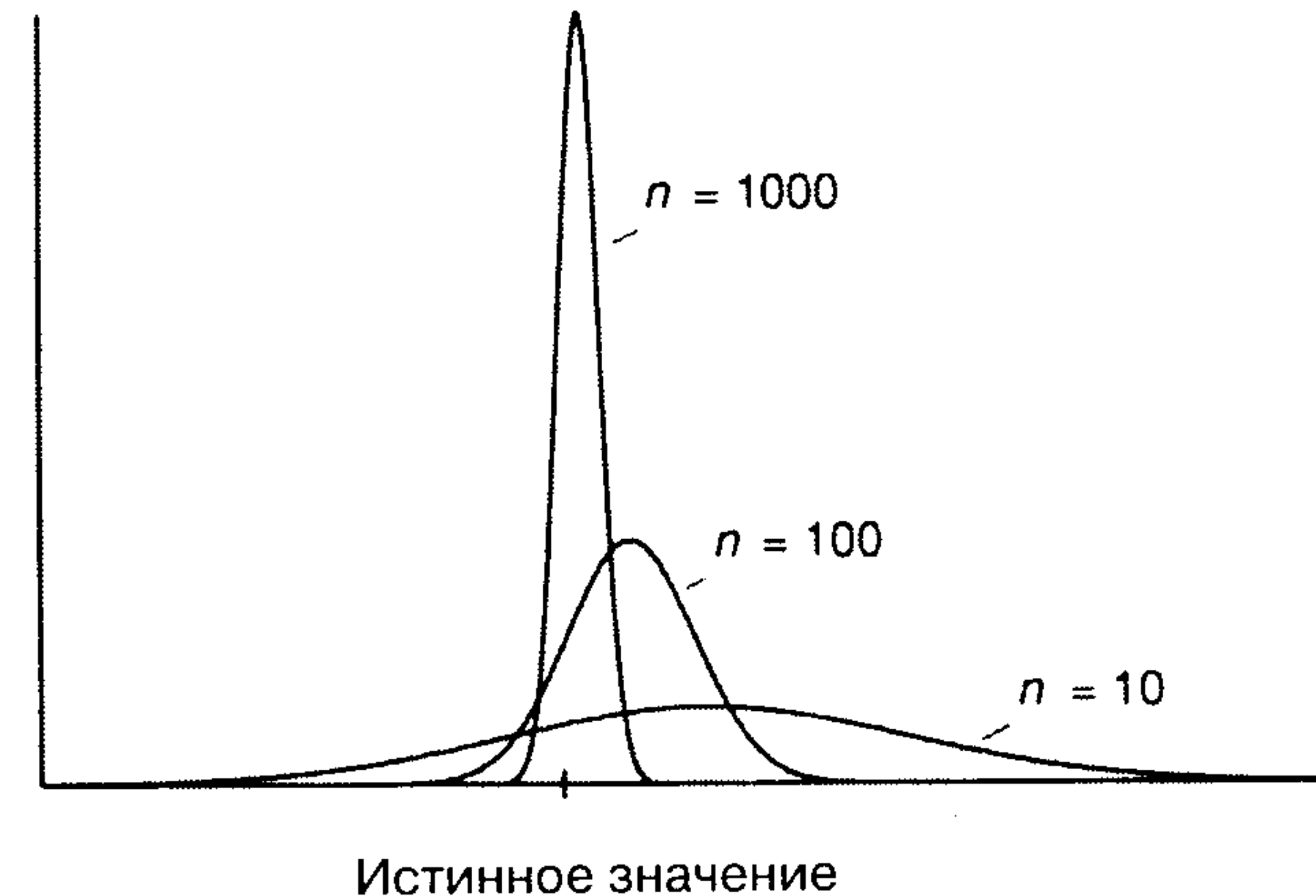
- **Цель акции:** увеличить продажи новой версии премиум подписки
- **Искомый эффект:** оценить влияние предложения скидки на вероятность обновления подписки



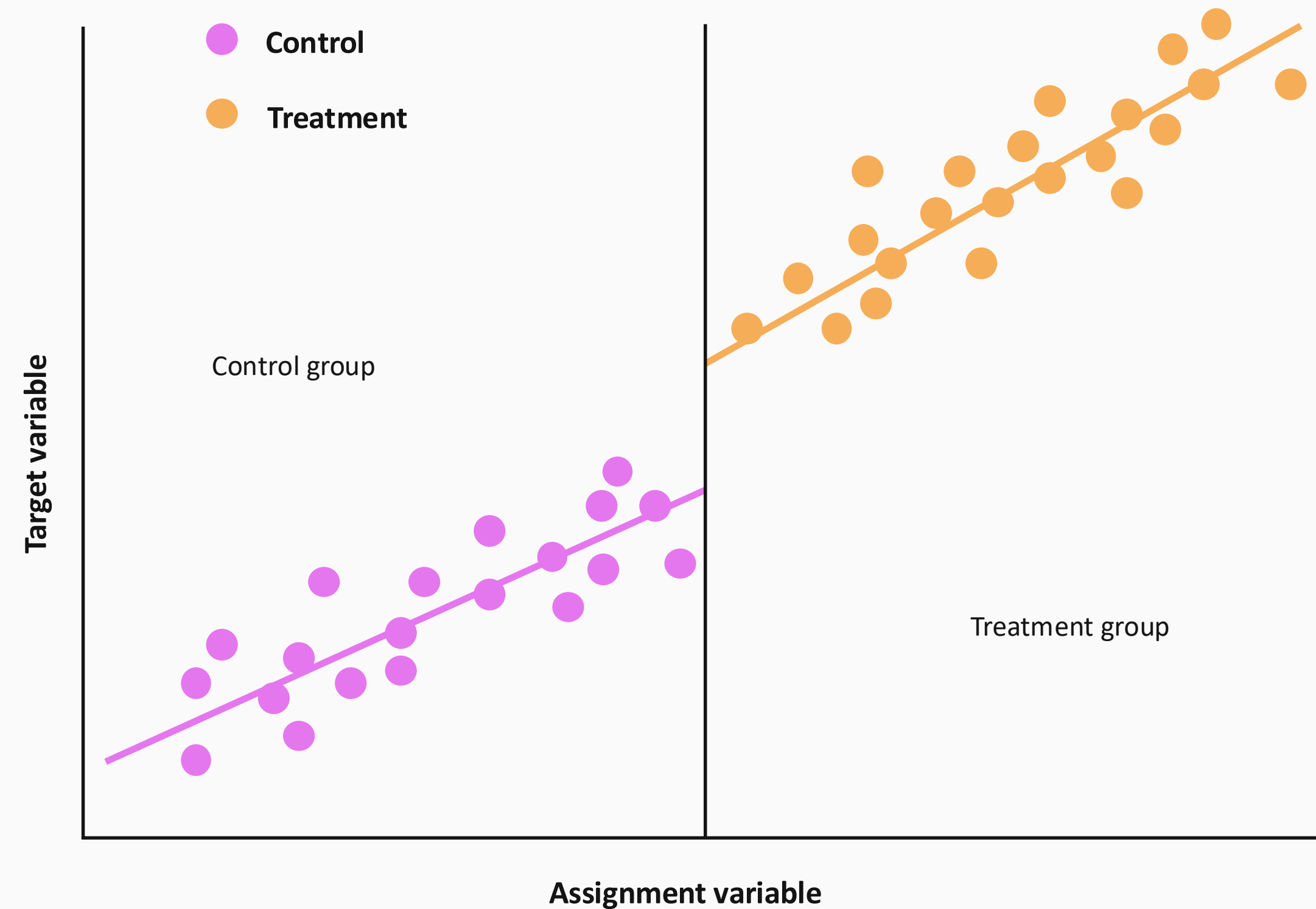
Проблема

- Нет дизайна эксперимента: имеем дело с реальным поведением пользователей без искусственного вмешательства.
- Хотим избежать необходимости в рандомизации: нет необходимости случайным образом распределять пользователей на группы.
- Нарушение несмещенности: пользователи с 29 днями до истечения подписки похожи на пользователей с 31 днями до истечения подписки, при этом они принадлежат к разным группам, т.е.
 $E[Y_0 \mid D_i = 1] - E[Y_0 \mid D_i = 0] = 0$ ожидаемый потенциальный исход без вмешательства (Y_0) одинаков для обеих групп.

Функция плотности вероятности



Regression Discontinuity Design (RDD)



Разрыв в исходах между двумя очень похожими группами может быть приписан ATE, если мы предполагаем, что остальные факторы изменяются плавно через порог

Предположение $E[Y_0 | D_i = 1] - E[Y_0 | D_i = 0] = 0$ глобально нарушается, но может выполняться локально вблизи порога

Алгоритм

1. Выбираем порог — c

Пусть переменная X — определяет относится ли наблюдение к тестовой или контрольной группе. При этом наблюдения, где $X \geq c$ являются тестовой группой, а $X < c$ — контрольной.

2. Оцениваем основное уравнение — $Y_i = \alpha + \tau D_i + f(X_i) + \epsilon_i$

Y_i — бизнес-метрика для наблюдения i .

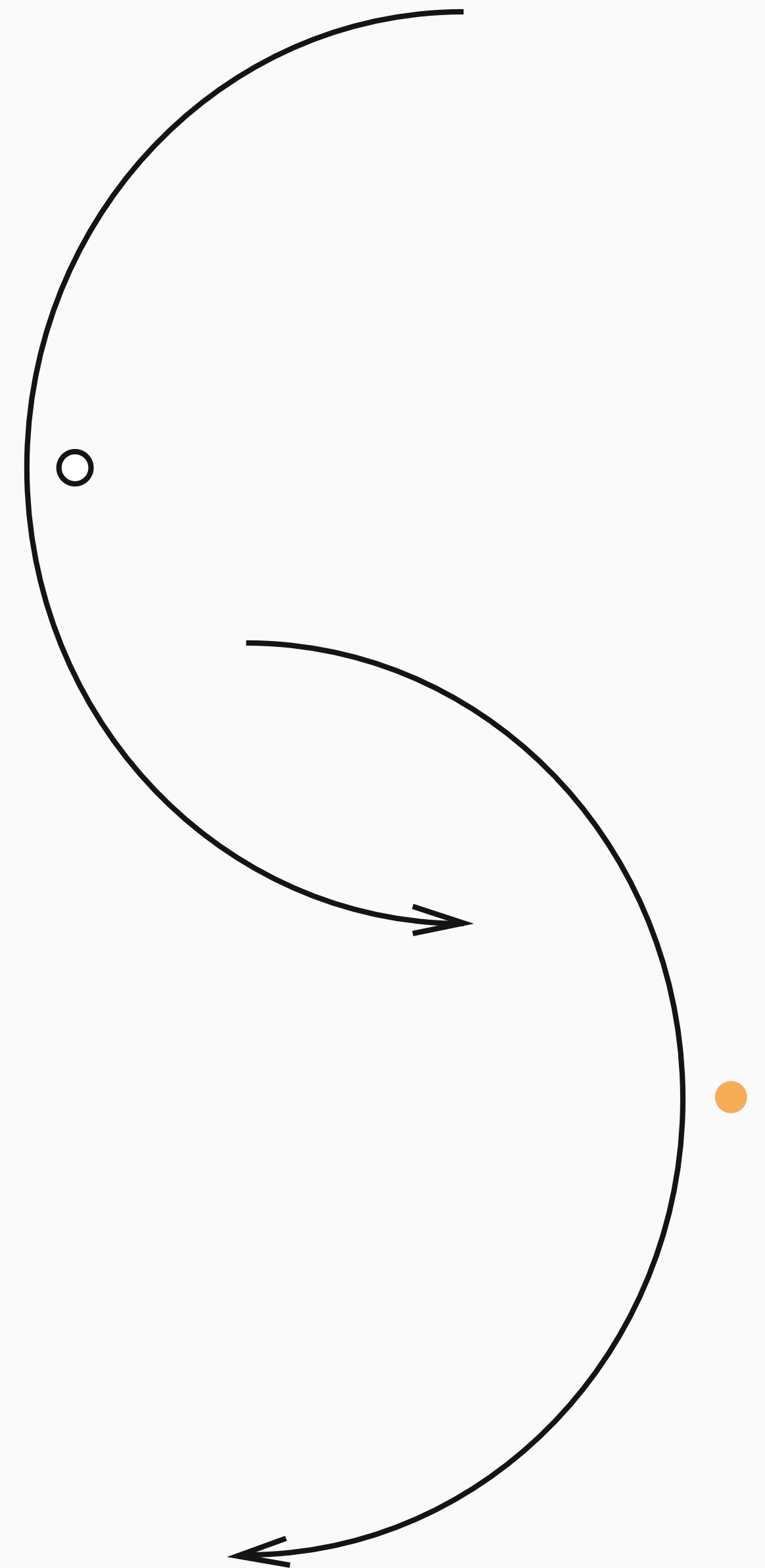
D_i — дамми-переменная, равная **1**, если $X_i \geq c$, и **0** в противном случае.

τ — параметр, который представляет собой оценку **ATE**.

$f(X_i)$ — функция, которая описывает зависимость бизнес-метрики от переменной X . Может быть линейной, квадратичной или более сложной функцией.

ϵ_i — ошибка модели.

Необходимо иметь достаточное количество пользователей вблизи порога для получения статистически значимых результатов!



Локальная оценка вблизи порога

Локальное предположение о непрерывности

$$\lim(x \rightarrow c^-) E[Y_0 \mid X_i = x] = \lim(x \rightarrow c^+) E[Y_0 \mid X_i = x]$$

Вычисление локального среднего эффекта

$$ATE = \lim(x \rightarrow c^+) E[Y \mid X_i = x] - \lim(x \rightarrow c^-) E[Y \mid X_i = x]$$

$$Y_i = Y_0 + D_i \cdot (Y_1 - Y_0), \text{ где } (Y_1 - Y_0) \text{ — индивидуальный эффект}$$

При условии непрерывности $E[Y_0 \mid X_i = x]$ в точке $x = c$, средний эффект:

$$ATE = [E[Y \mid X_i = c^+] - E[Y \mid X_i = c^-]]$$

Bandwidth selection

$$|X_i - c| \leq h$$

где h — ширина полосы

Почему выбор полосы пропускания важен?

- Bias: малое $h \rightarrow$ уменьшает смещение в оценке эффекта
- Variance: большое $h \rightarrow$ снижает дисперсию

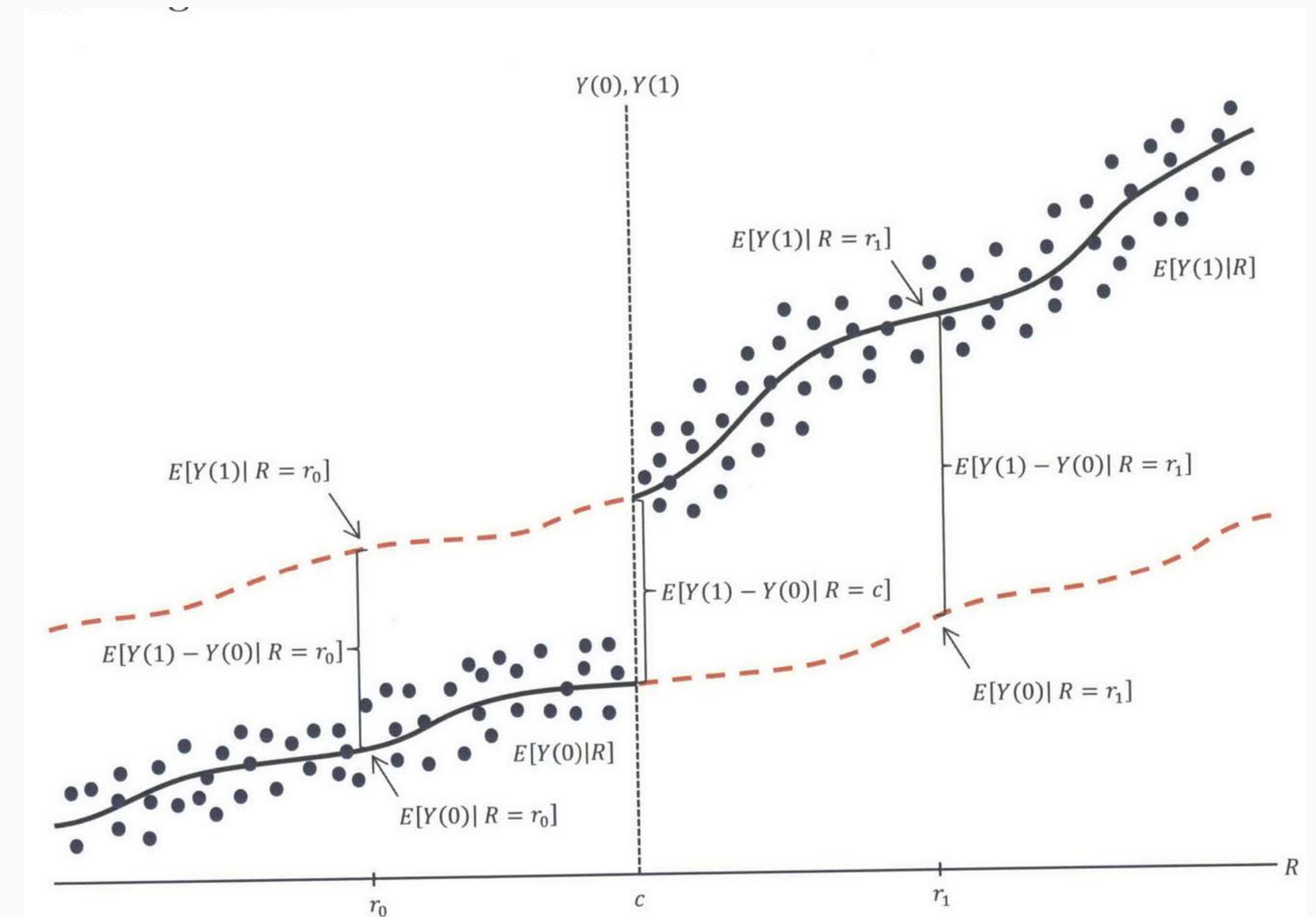
Методики выбора полосы:

Кросс-валидация

Silverman's Rule of Thumb

Sheather Jones Method (AMISE \rightarrow min)

Optimal Bandwidth Choice for the Regression Discontinuity Estimator. 2019 (Imbens & Kalyanaraman)



Bandwidth selection - overfitting

Один из методов уменьшения вероятности ложных эффектов - это сужение h .

Основная идея: чем ближе "приближаемся" к этой точке s , тем меньше вероятность обнаружения тенденции.

1. Использование полиномов высокой степени

- Могут чрезмерно подгонять модель под данные

2. Зависимость оценок от малого количества данных и смещение весов

- При использовании полиномов высокой степени – зависим от небольшого числа наблюдений. Модель может придавать большой вес наблюдениям, удалённым от порога

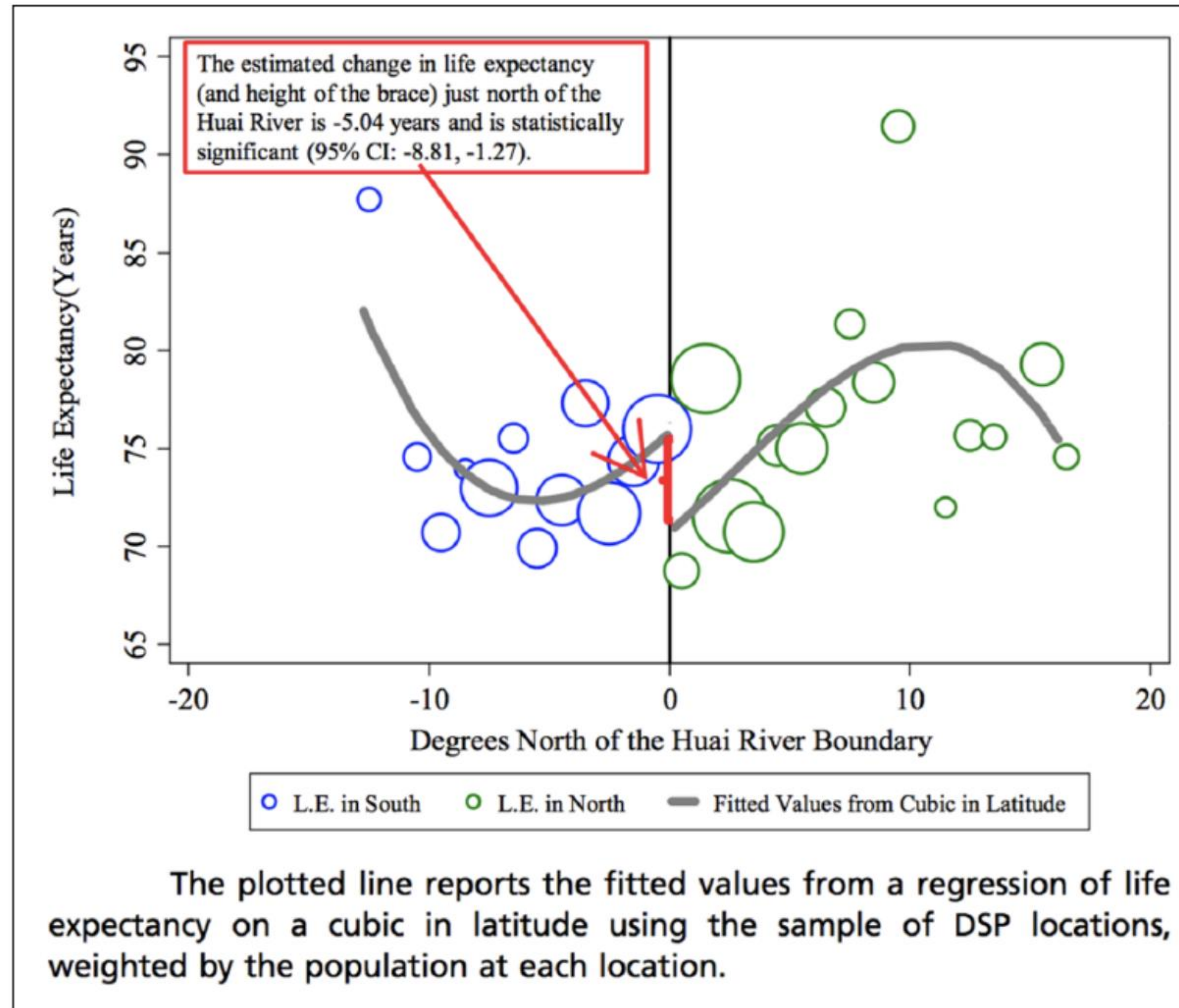
4. Недостаточное отражение неопределённости в стандартных ошибках

Стандартные ошибки могут не корректно отражать неопределённость, связанную с выбором модели и степенью полинома.

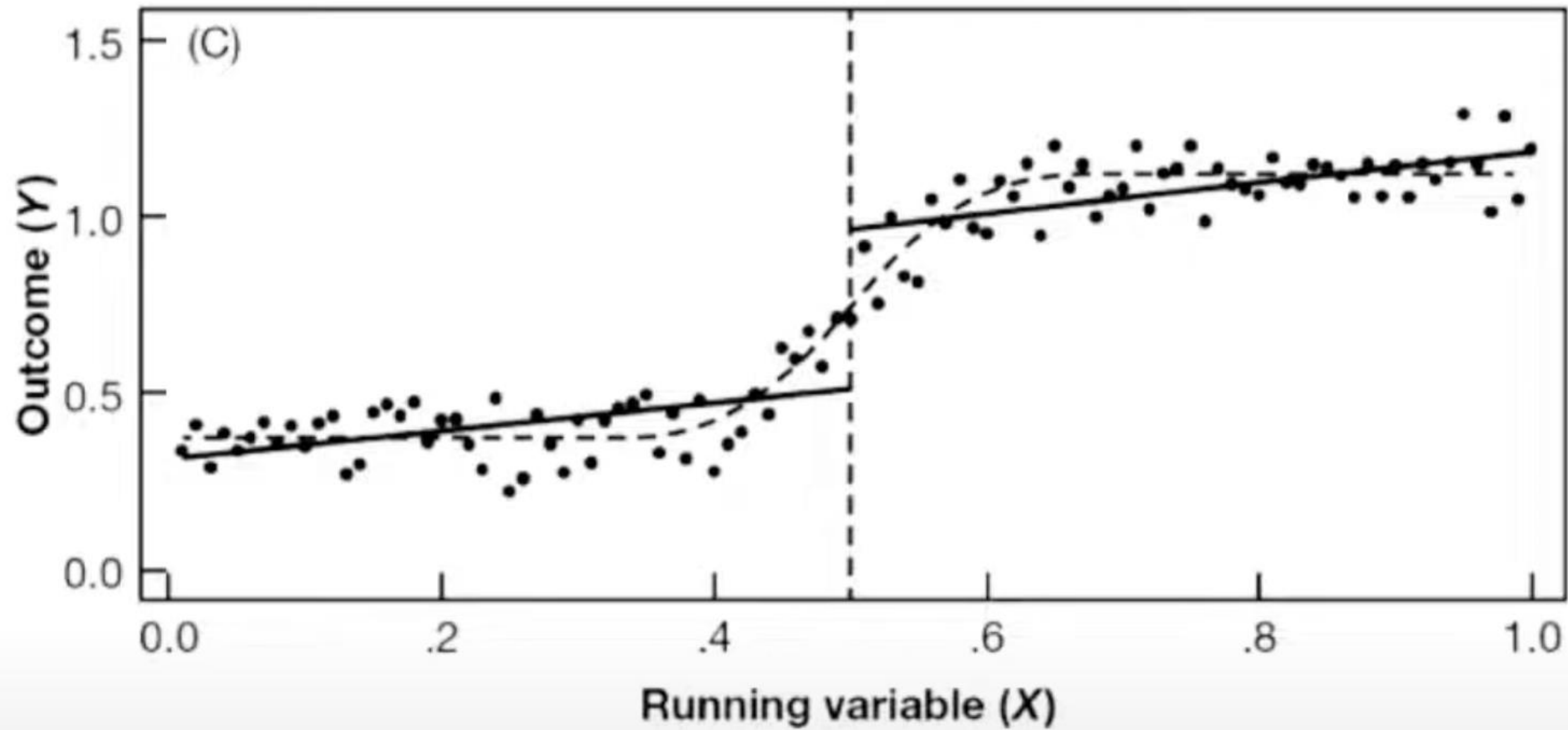
Это означает, что доверительные интервалы и тесты значимости могут быть ненадёжными.

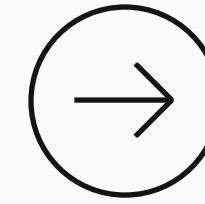
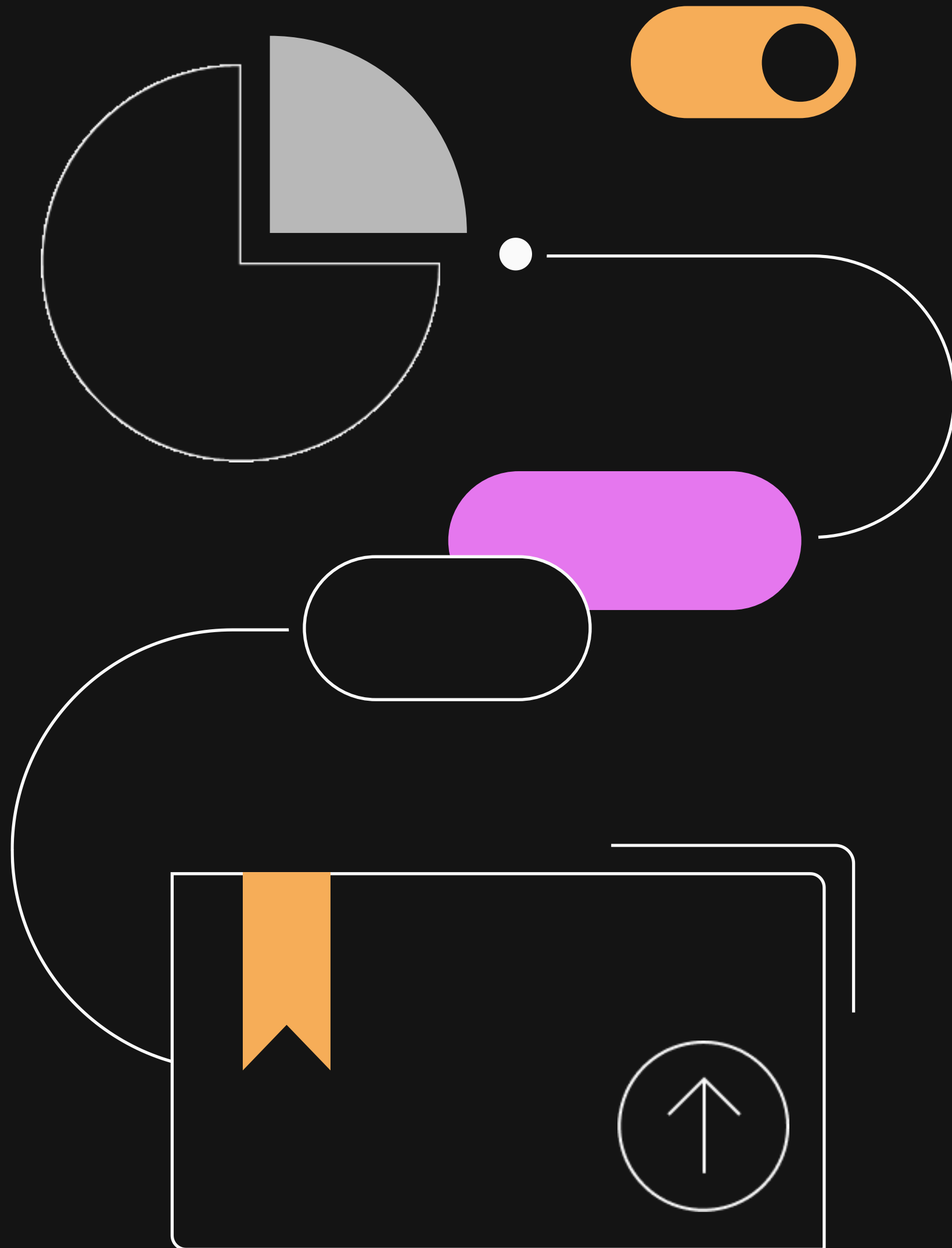
Green et al. (2009) Gelman and Imbens (2019)

Gelman Zelizer (2015)



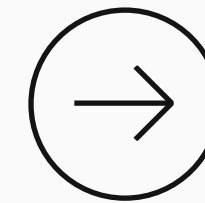
Robustness





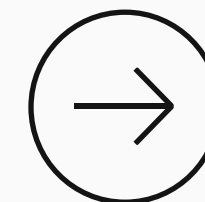
Выбираем пороговое значение:

В данном случае порогом является срок окончания текущей подписки. Клиенты, у которых подписка истекает в ближайшие 30 дней, получают предложение со скидкой, а те, у кого подписка истекает более чем через 30 дней, не получают такую скидку



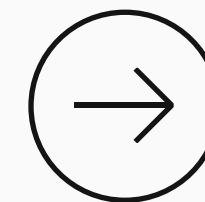
Разбиение на группы

Тестовая группа включает клиентов, чьи подписки истекают в ближайшие 30 дней, а контрольная группа — тех, у кого контракты истекают через 31-60 дней.



Сравнение групп

Сравним вероятность обновления подписки между группами. Предполагаем, что эти группы схожи по всем параметрам, кроме доступа к скидке.

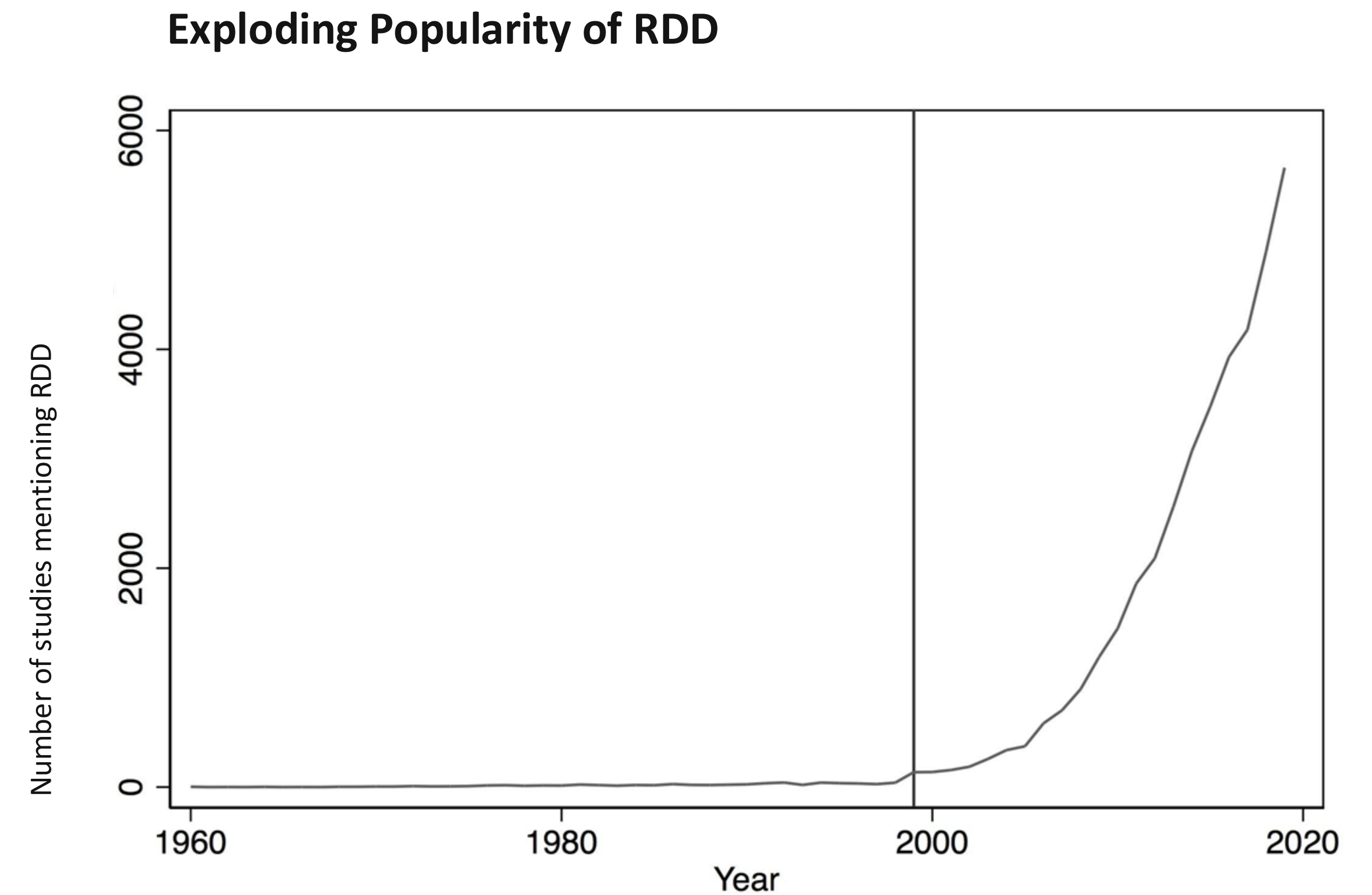


Построение регрессии

Оцениваем разрыв в вероятности обновления подписки между этими двумя группами. Это позволяет оценить причинный эффект предложения скидки на решение клиентов обновить подписку.

Причины использовать RDD

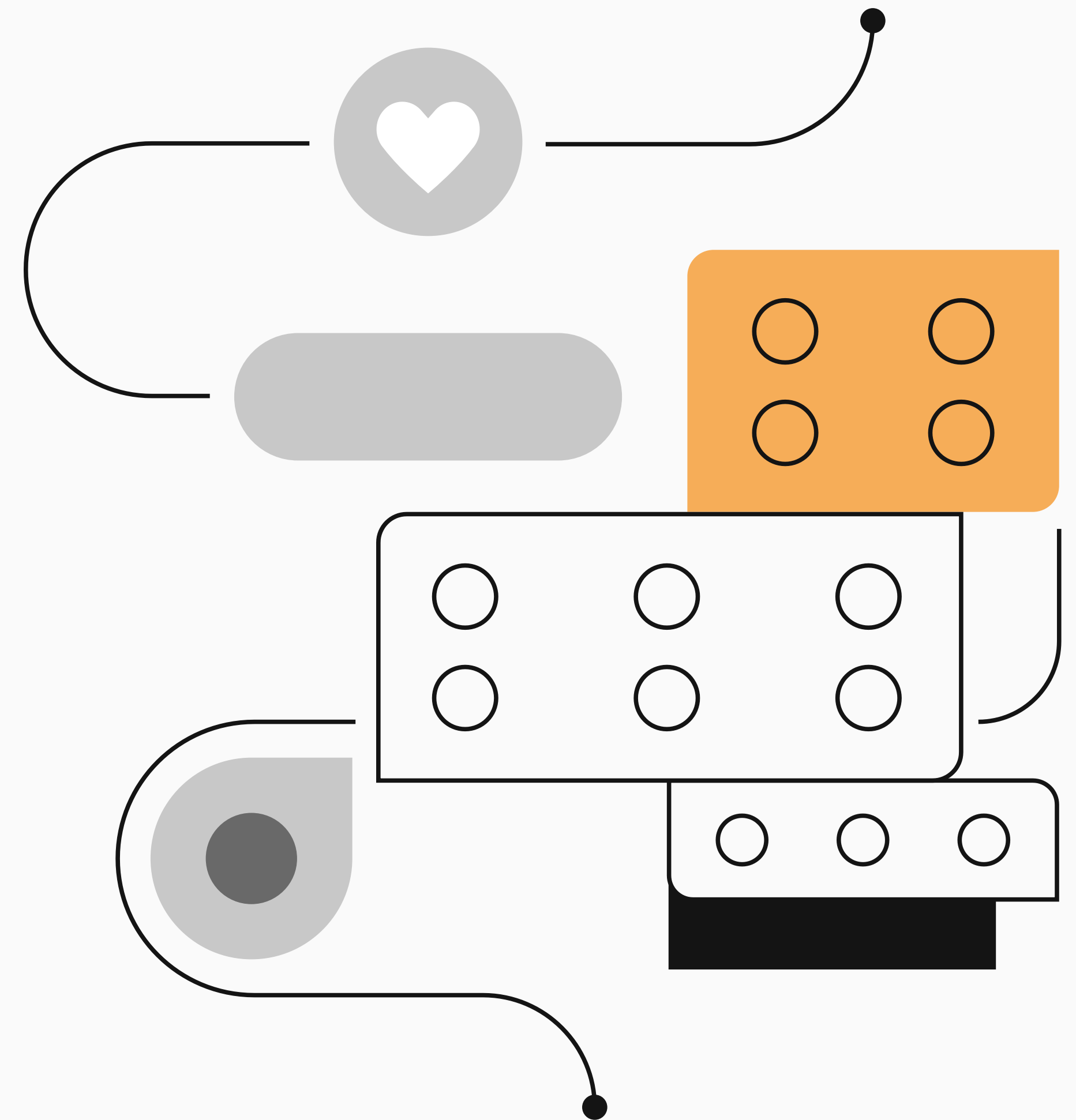
- Снижение ошибки при несопоставимости групп
- Не нужно дизайнить тест
- Простота интерпретации



Vertical bar is Angrist and Pischke (1999) and Black (1999)

Ограничения RDD

- Наблюдения должны находиться вблизи порогового значения
- Необходимость наличия точной информации о пороговом значении и достаточная плотность наблюдений вокруг него
- Требование непрерывности

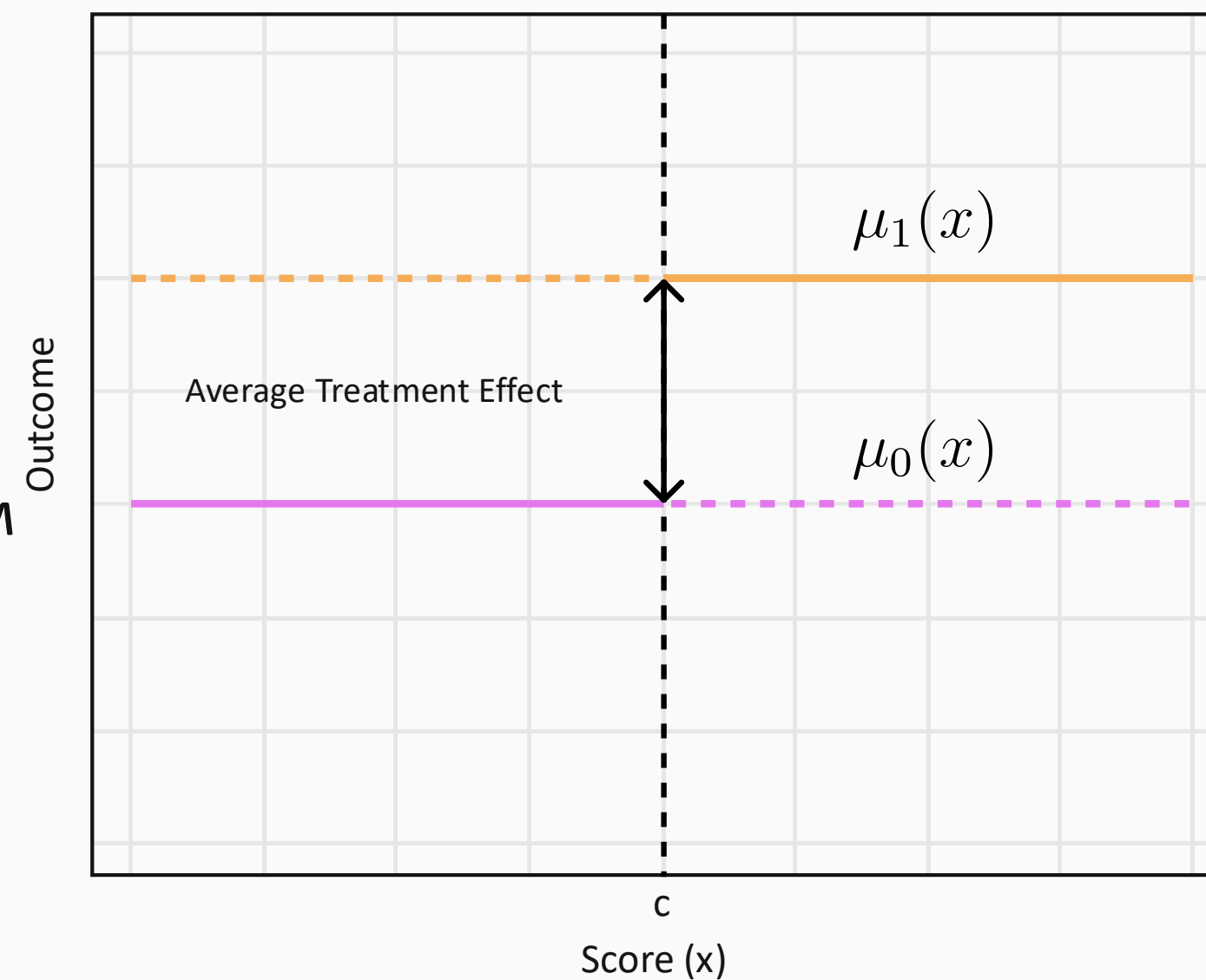


Как можно улучшить RDD?

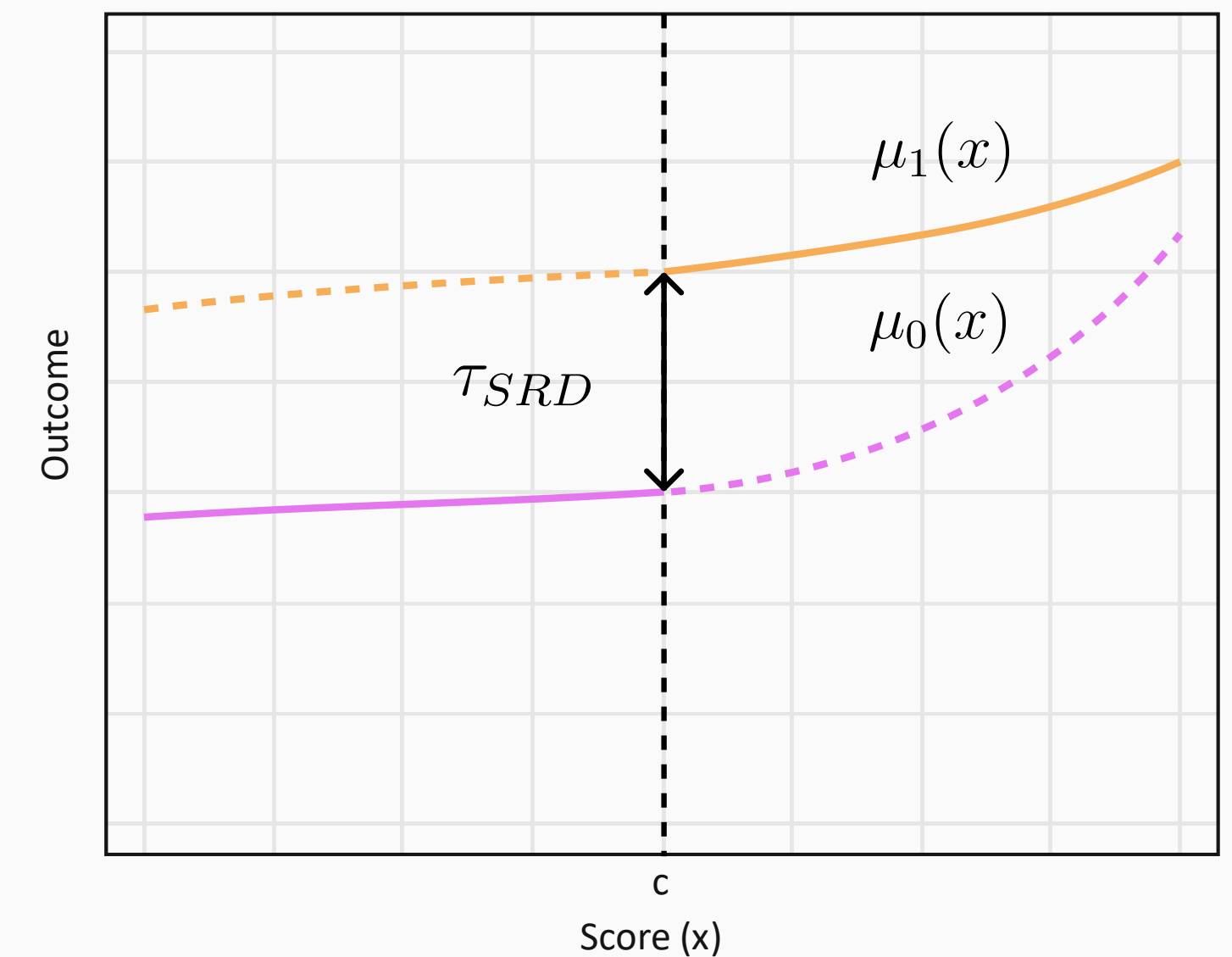
→ Полупараметрические
и непараметрические подходы

→ Вблизи порогового значения рассматриваем
данные
как случайно распределенные
[Local Randomization Approach](#)

(a) Randomized Experiment

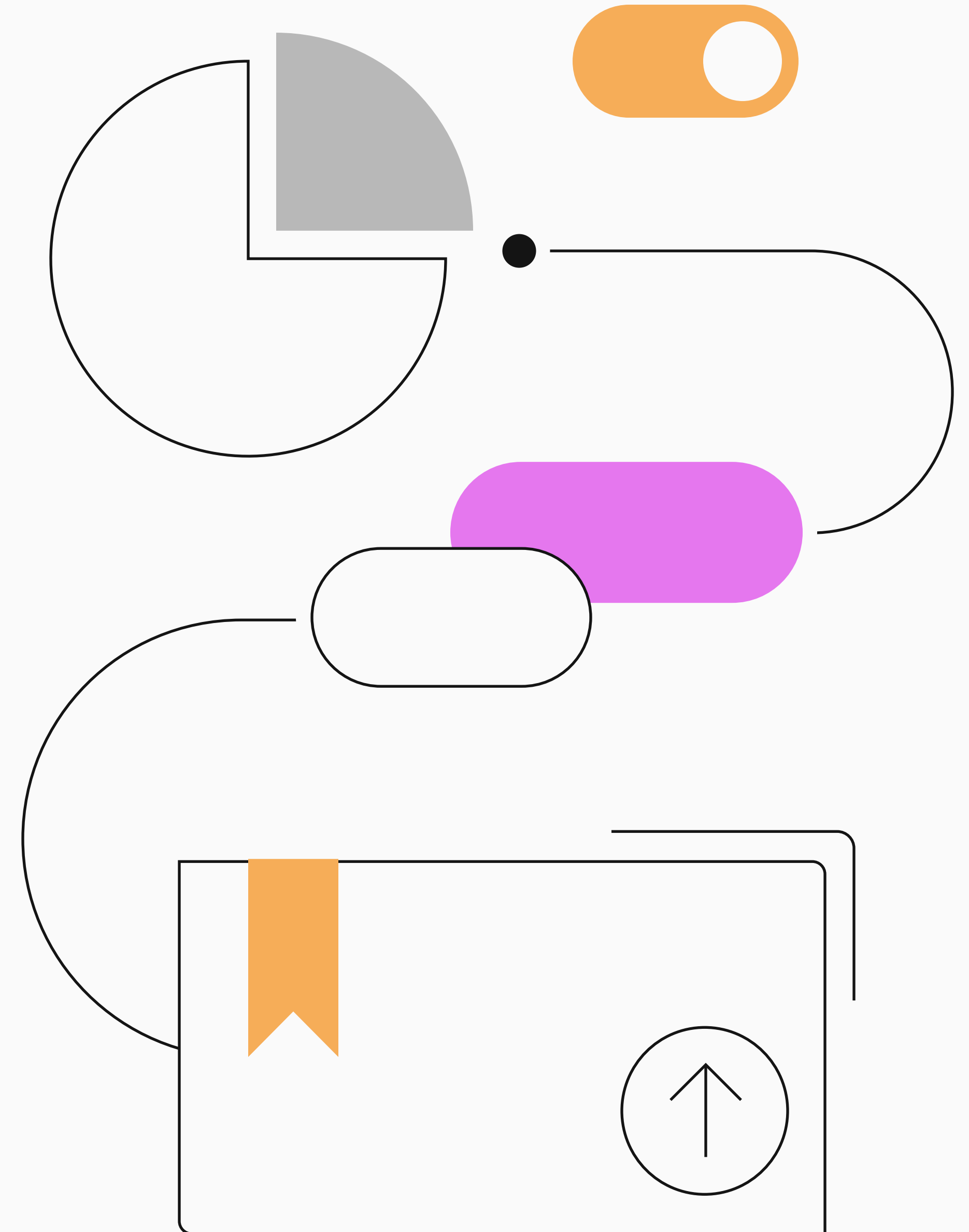


(b) RD Design

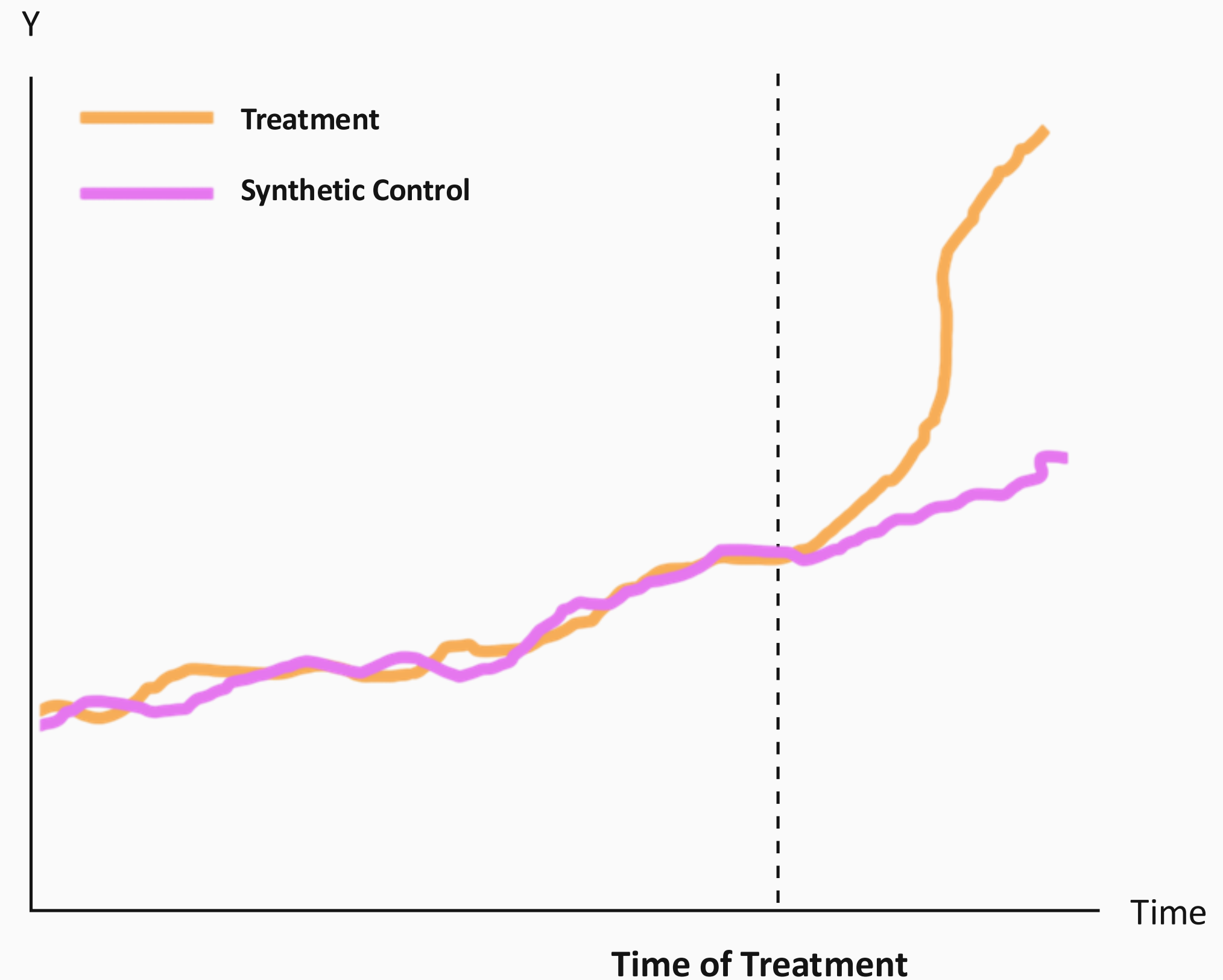


Компания управляющая сетью кофеен решила протестировать новую концепцию меню, включающую больше веганских блюд и блюд без глютена.

- **Цель акции:** увеличить прибыль
- **Искомый эффект:** оценка влияния нового меню на средний чек и число посетителей
- **Проблема:** внешние факторы, которые влияют на посещаемость и выручку кафе



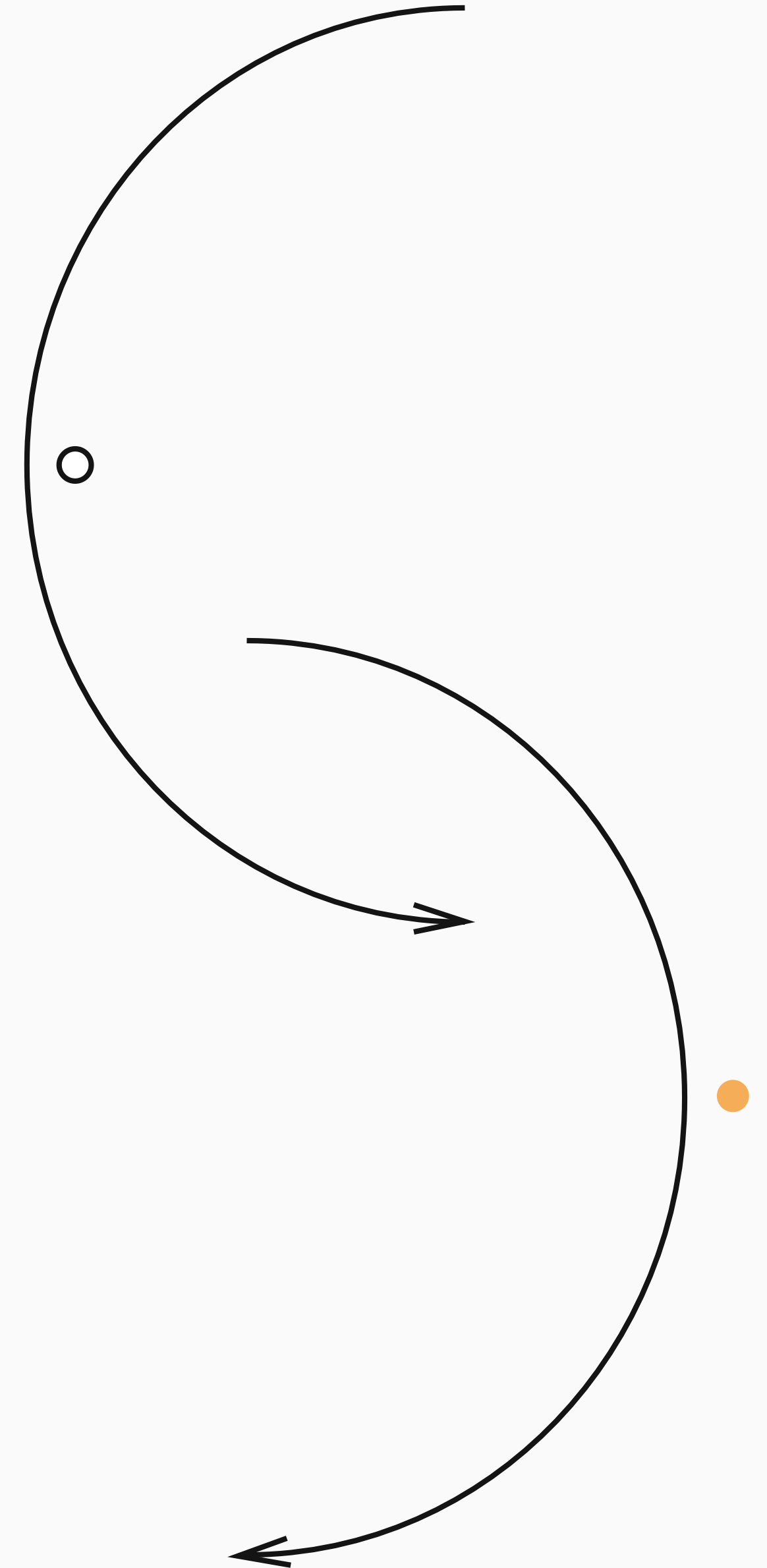
Synthetic Control Method (SCM)



Создаем «синтетическую» версию контрольной группы: комбинацию нескольких тестовых групп, которая максимально соответствует характеристикам группы до вмешательства.

Алгоритм

- Пусть
 X_1 — матрица ковариат для тестовой выборки до начала эксперимента
 X_0 — матрица ковариат для контрольной выборки.
Найдем веса $W = (w_1, w_2, \dots, w_j)$, которые минимизируют: $\|X_1 - X_0 W\|$
Причем для любого $w \in W$: $w_i \geq 0$ и $\sum_{j=1} w_j = 1$
- Пусть
 $Y(T, t)$ — значение бизнес-метрики для тестовой выборки в момент времени t
 $Y^*(C, t) = \sum_{j=1} w_j Y(j, t)$ — соответствующее значение для синтетической контрольной группы.
- Тогда **ATE** в момент времени t оценивается как:
 $\Delta_t = Y(T, t) - Y^*(C, t)$
- Если разница Δ_t значительна и стабильна в течение всего эксперимента, то эффект есть



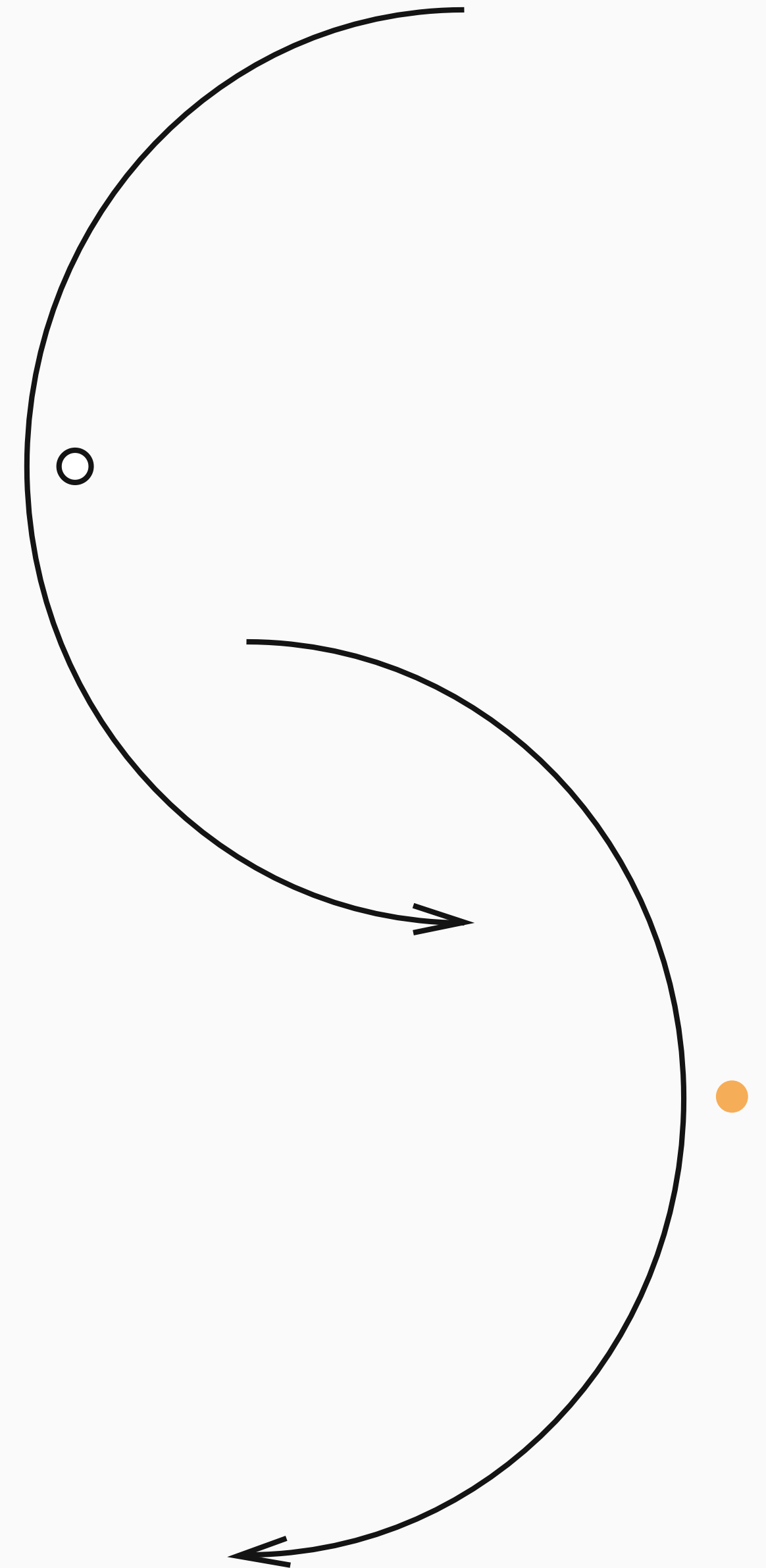
Взаимосвязь с регрессией

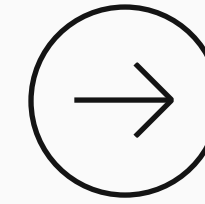
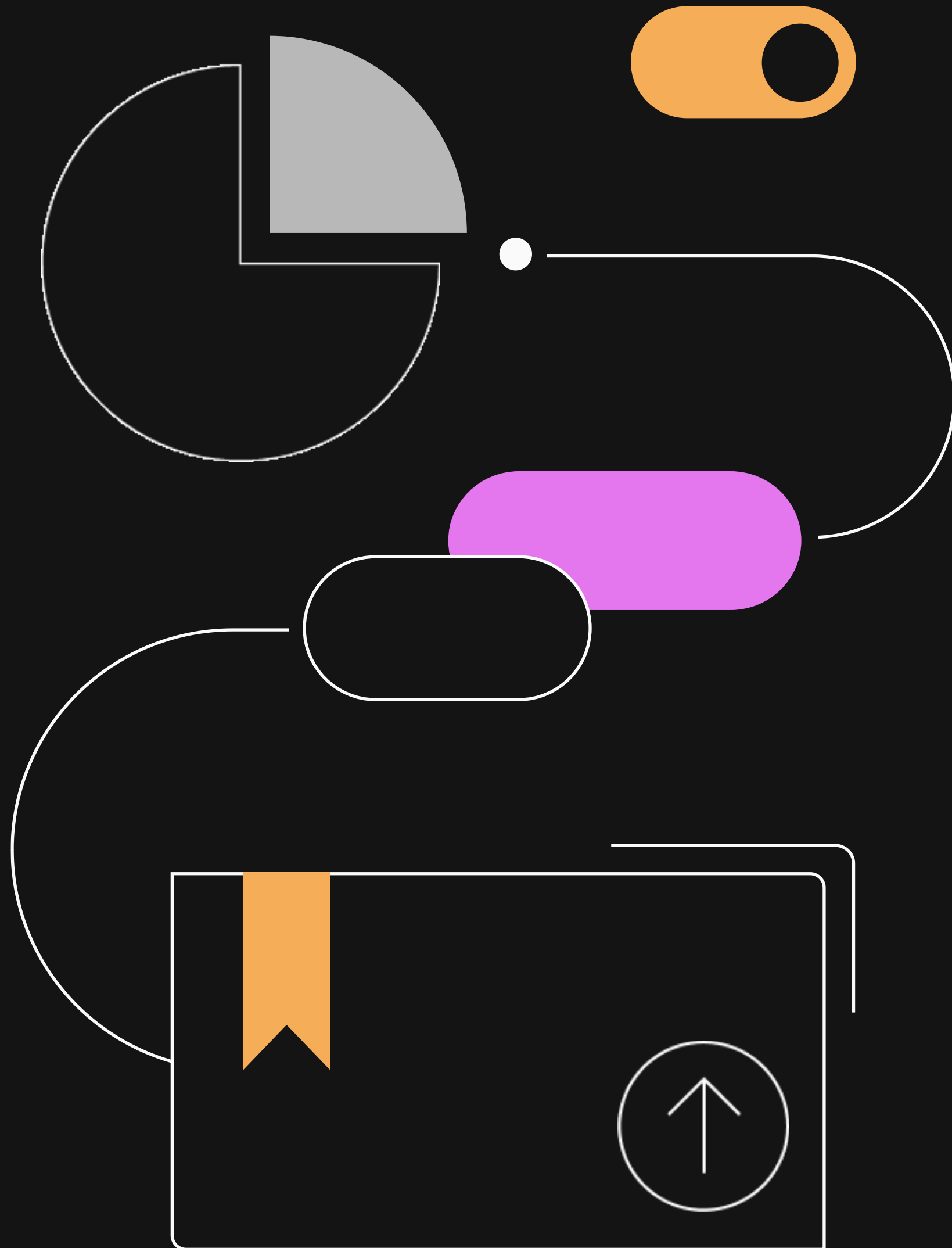
При следующих ограничениях:

1. Нулевая константа
2. Положительные веса: $w_i \geq 0$ для всех $i = 1, \dots, N-1$
3. Сумма весов равна единице: $\sum_{i=1} w_i = 1$

Дополнительные условия:

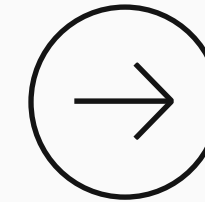
- Нет постоянной разницы во времени между тестовой и контрольной группой.
- Каждая единица анализа находится в выпуклой оболочке контрольных единиц.
- Регуляризация необходима, когда N велико по сравнению с T .





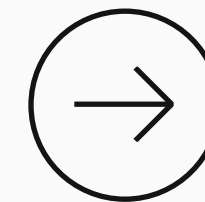
Определим тестовую группу:

Отберем несколько кафе для внедрения новой концепции меню. Эти кафе станут экспериментальной группой.



Создадим синтетическую контрольную группу:

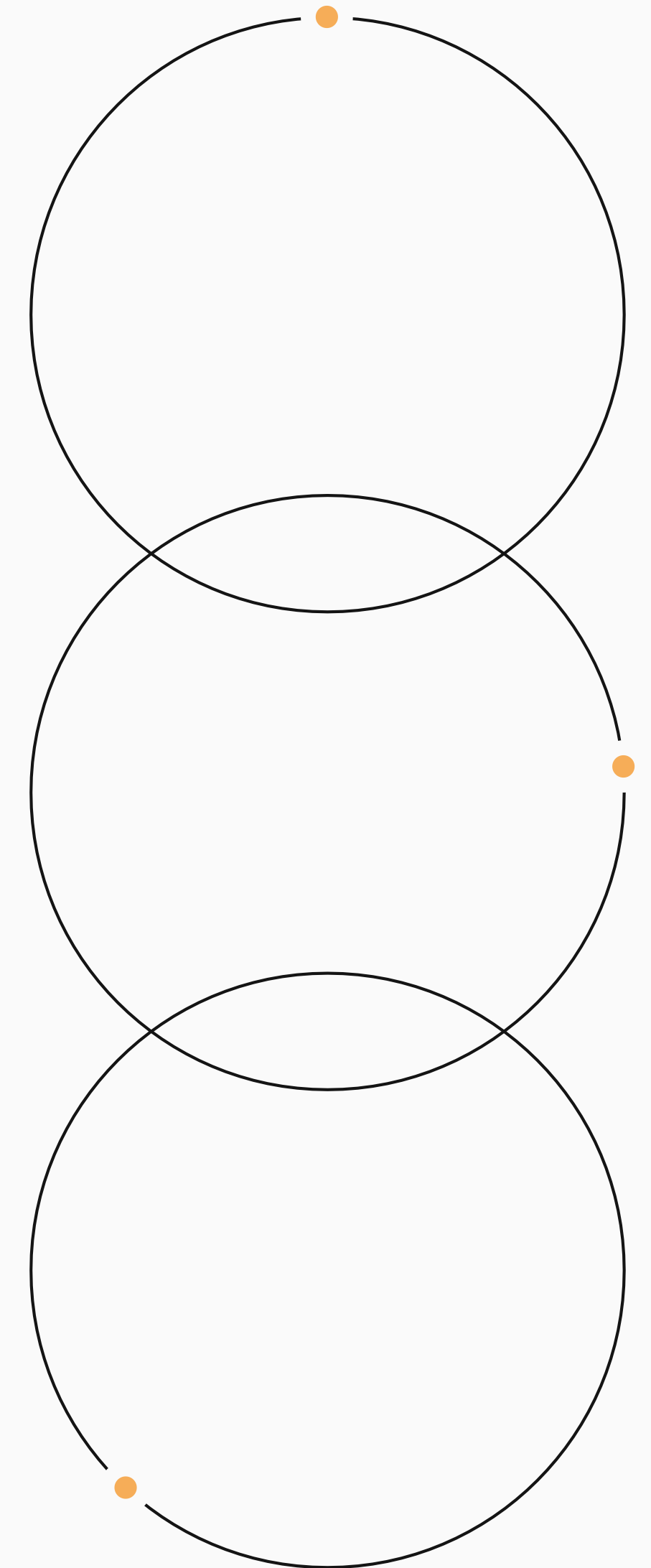
Характеристики (средняя посещаемость/средний чек/расположение) должны максимально соответствовать характеристикам тестовых кафе до внедрения изменений.



Изменения в посещаемости и среднем чеке в тестовых кафе сравниваются с изменениями в синтетической контрольной группе

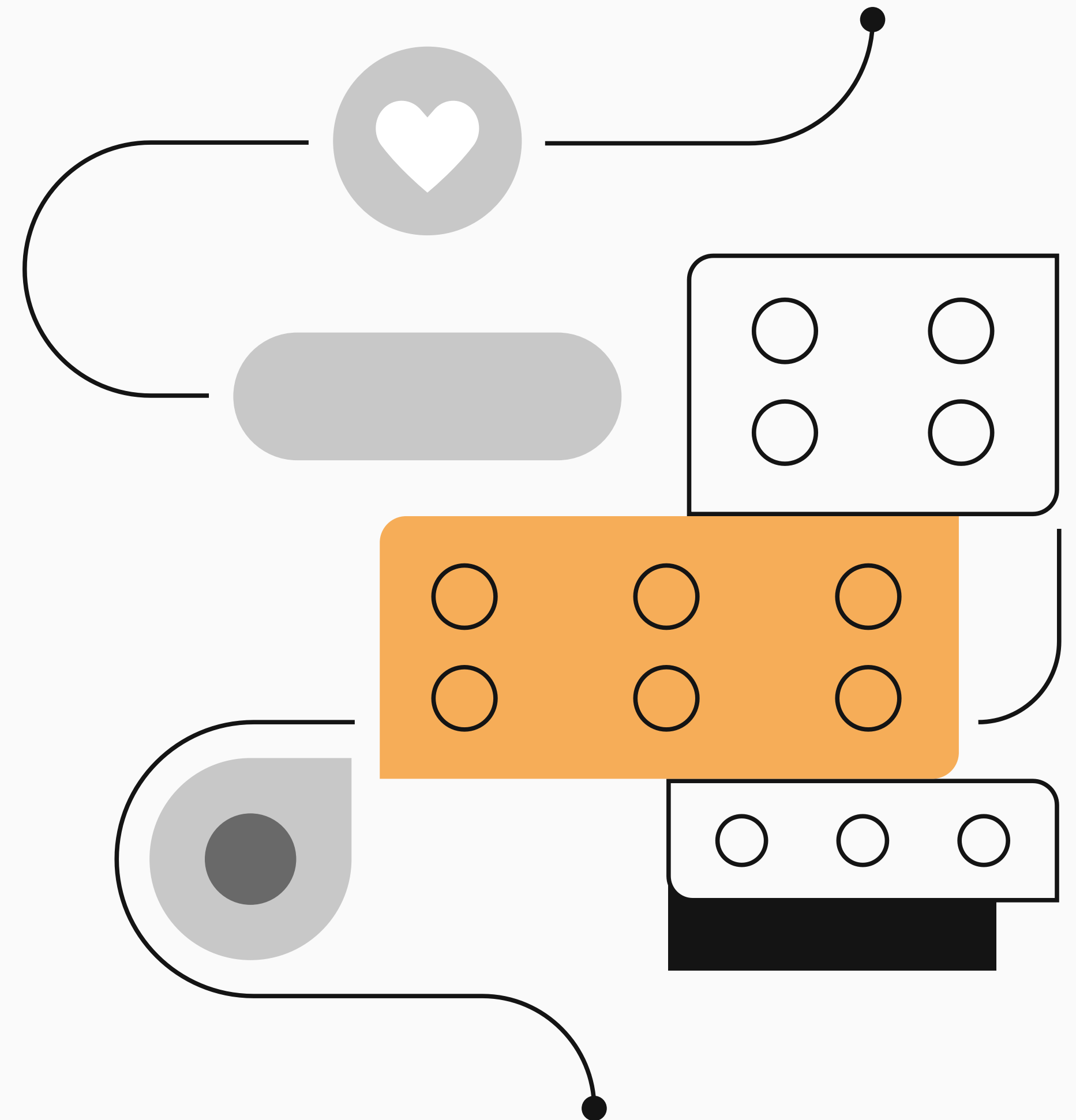
Причины использовать SCM

- Работа с агрегированными данными
- Отсутствие параллельных трендов
- Не предполагается однородность эффекта



Ограничения SCM

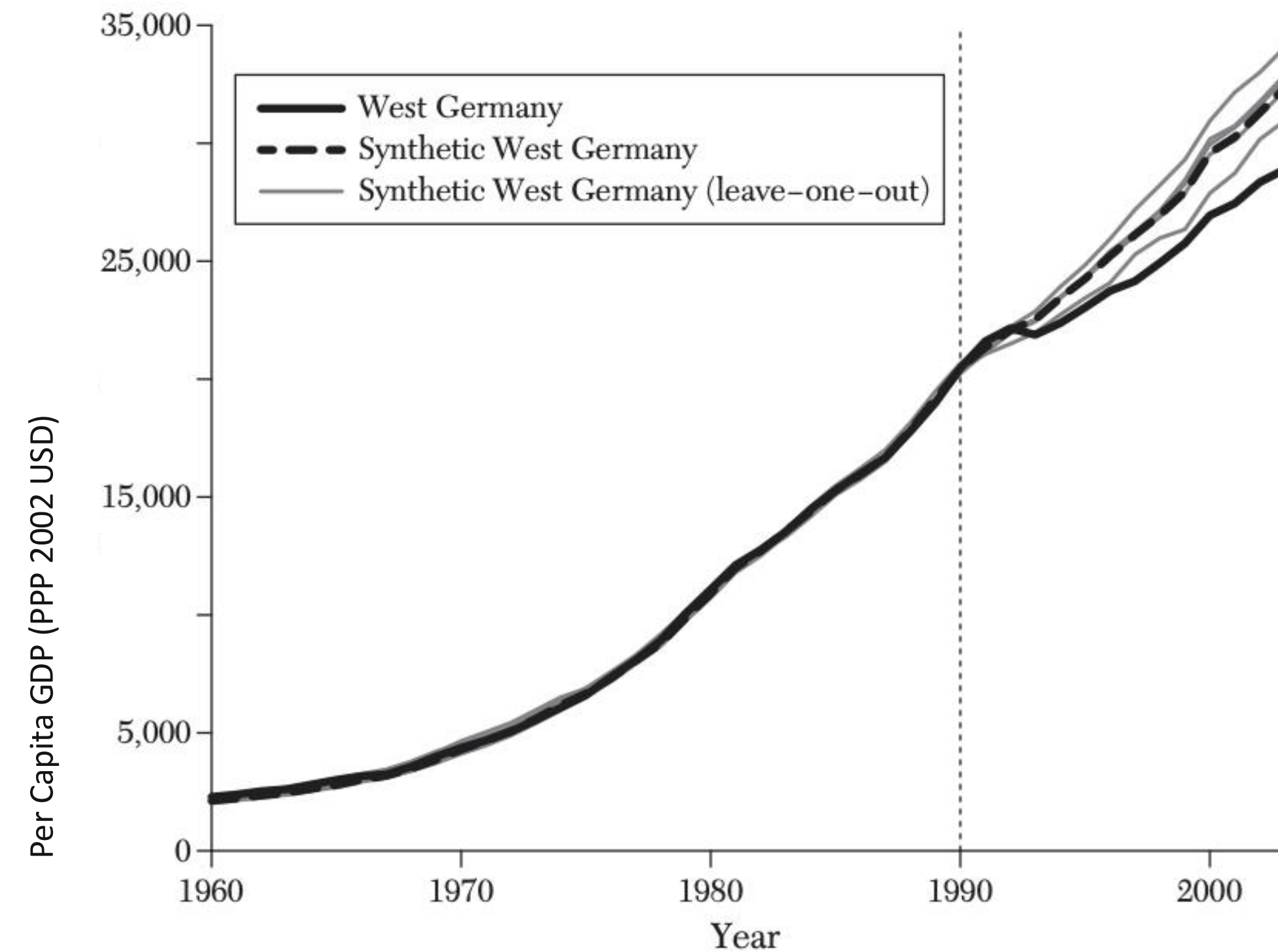
- Чувствительность к выбору пула наблюдений, из которого создается синтетический контроль.
- Сложность в интерпретации весов.
- Предполагает, что синтетический контроль может точно воспроизвести тренд исследуемой переменной до вмешательств.



Как можно улучшить SCM?

→ Использовать байесовские методы для учета неопределенности в выборе весов **Bayesian Synthetic Control**.

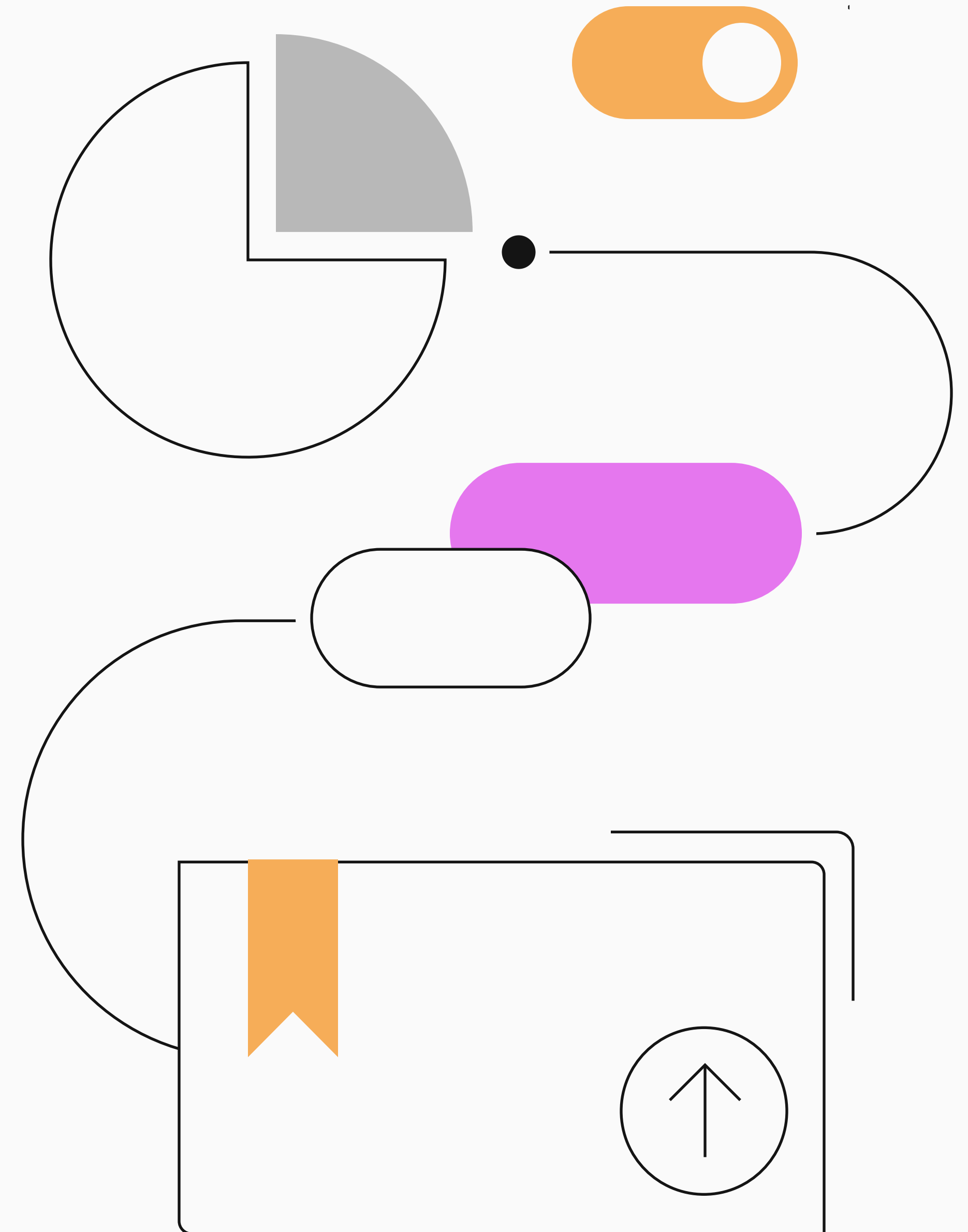
→ **ML + SCM**



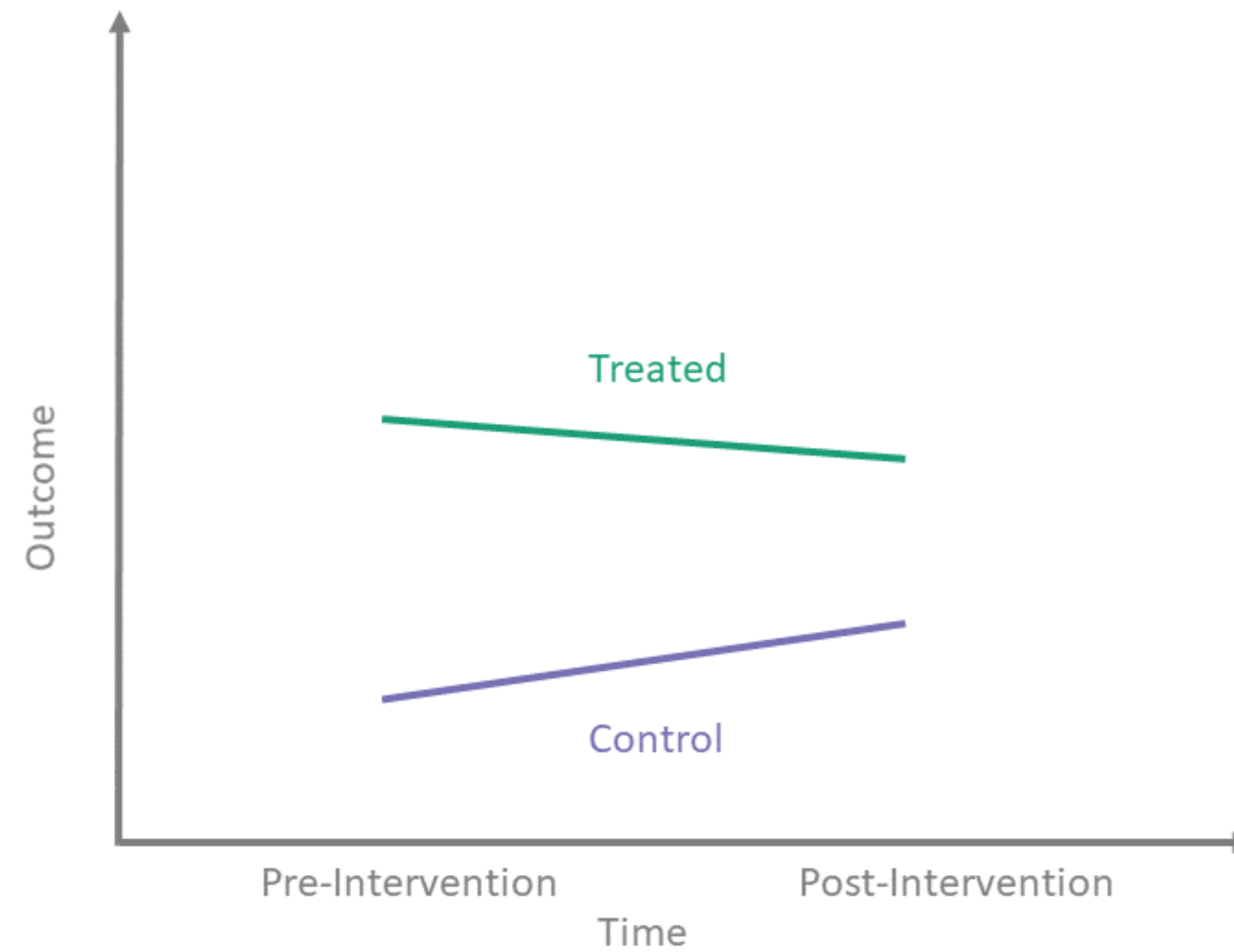
Оценки последствий объединения Германии 1990 года
Abadie, A. (2021). "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects".
Journal of Economic Literature.

Платформа онлайн-продаж хочет оценить влияние новой ценовой политики на объем продаж. В рамках эксперимента компания решила изменить цены на определенные товары в одной группе регионов (тестовая группа), в то время как в другой группе регионов цены остались прежними (контрольная группа).

- **Цель политики:** увеличить выручку
- **Искомый эффект:** оценка влияния изменения цен на объем продаж, учитывая возможные внешние факторы, такие как сезонные колебания спроса



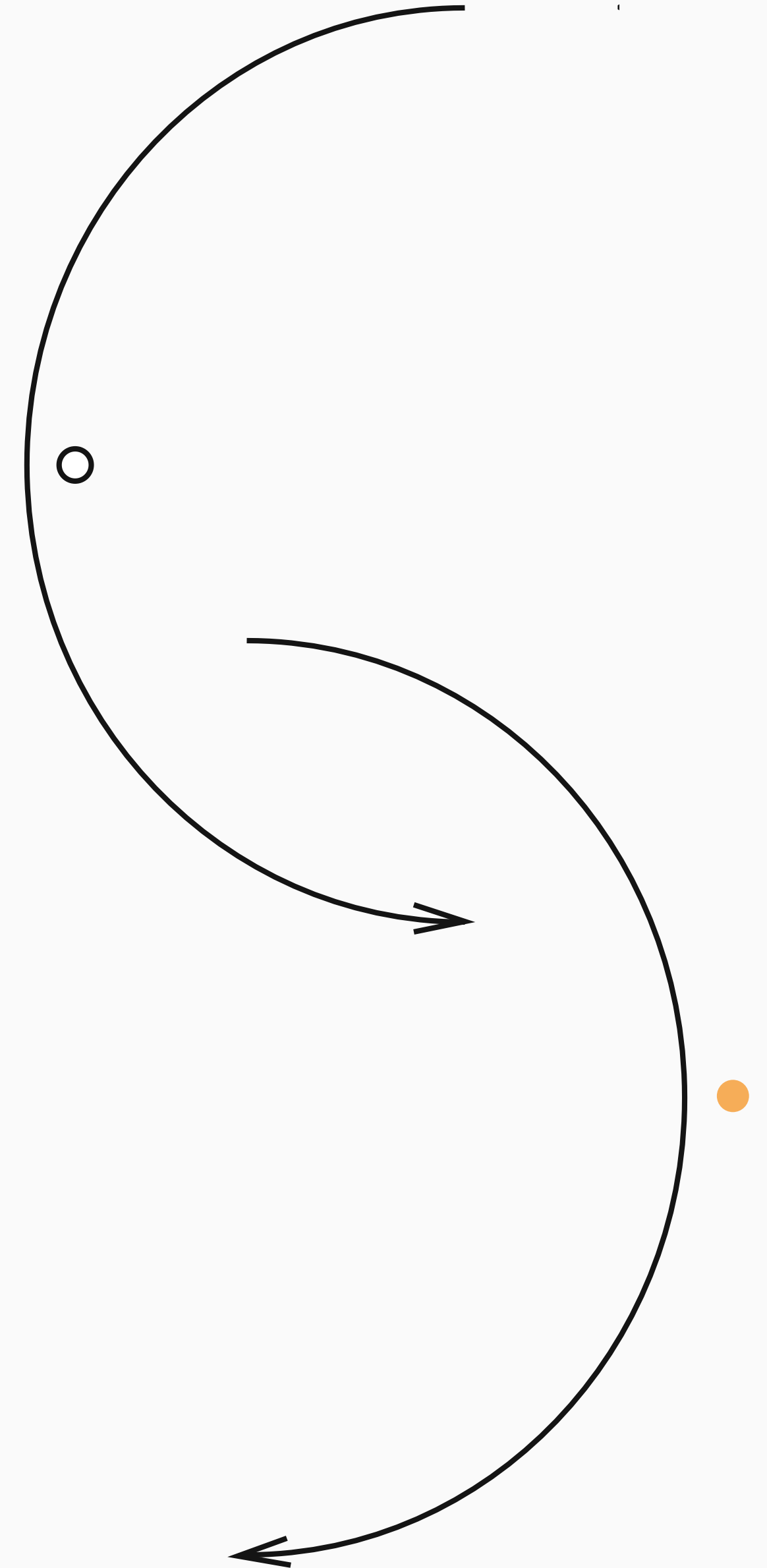
Diff-in-Diff



DiD сравнивает изменения в результатах между тестовой и контрольной группами до и после вмешательства.

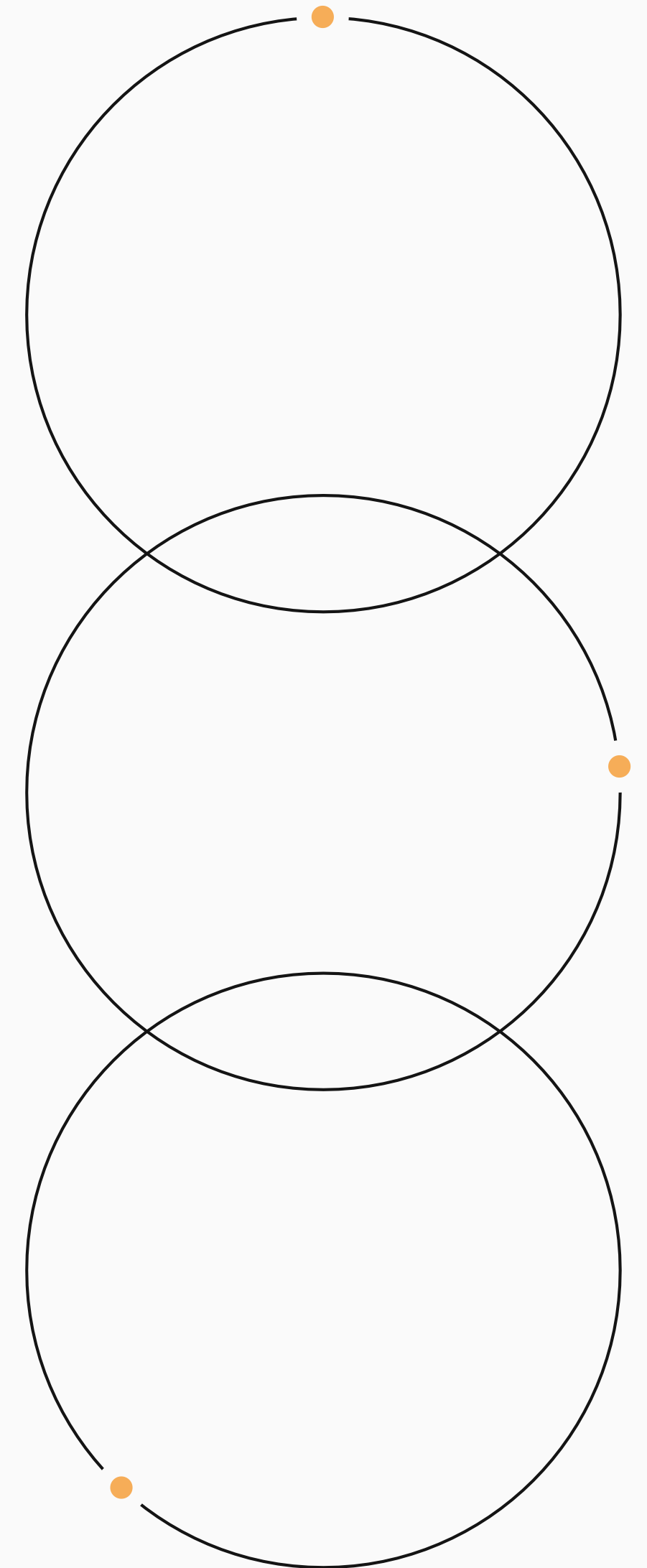
Алгоритм

- Вычисляем средние значения бизнес-метрики для каждой группы до начала эксперимента.
- Вычисляем изменения для каждой группы:
 $\Delta Y(\text{tr}) = Y(\text{tr, after}) - Y(\text{tr, before})$
 $\Delta Y(\text{c}) = Y(\text{c, after}) - Y(\text{c, before})$
- Расчет разности в разностях
 $\Delta(\text{DiD}) = (\Delta Y(\text{tr}) - \Delta Y(\text{c}))$
- Если $\Delta(\text{DiD}) \gg 0$, значит эффект есть



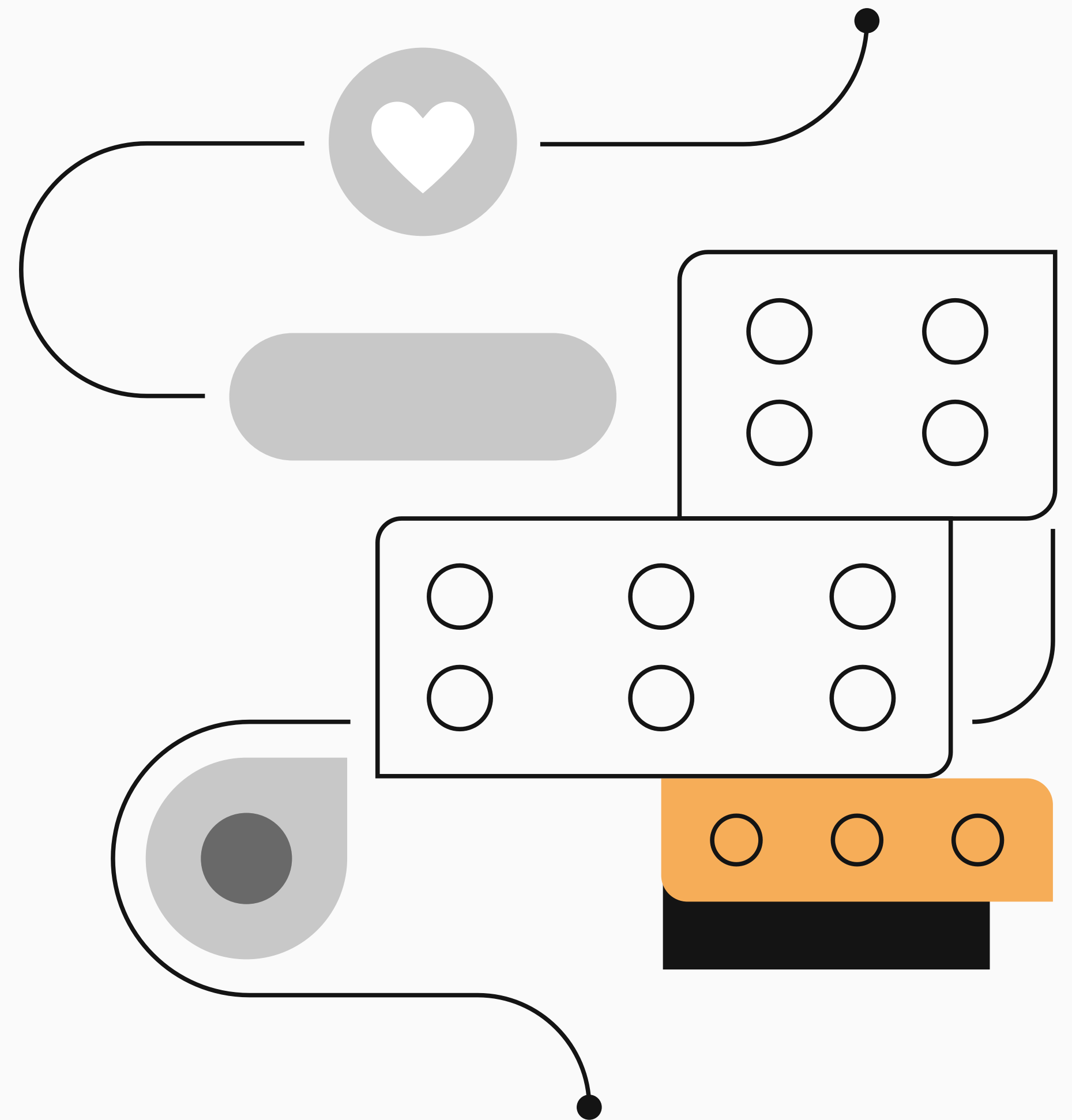
Причины использовать DiD

- У вас есть данные до и после изменения для обеих групп.
- Вы хотите учесть временные тренды или сезонные колебания.
- Необходим контроль за неизменными во времени различиями между группами.



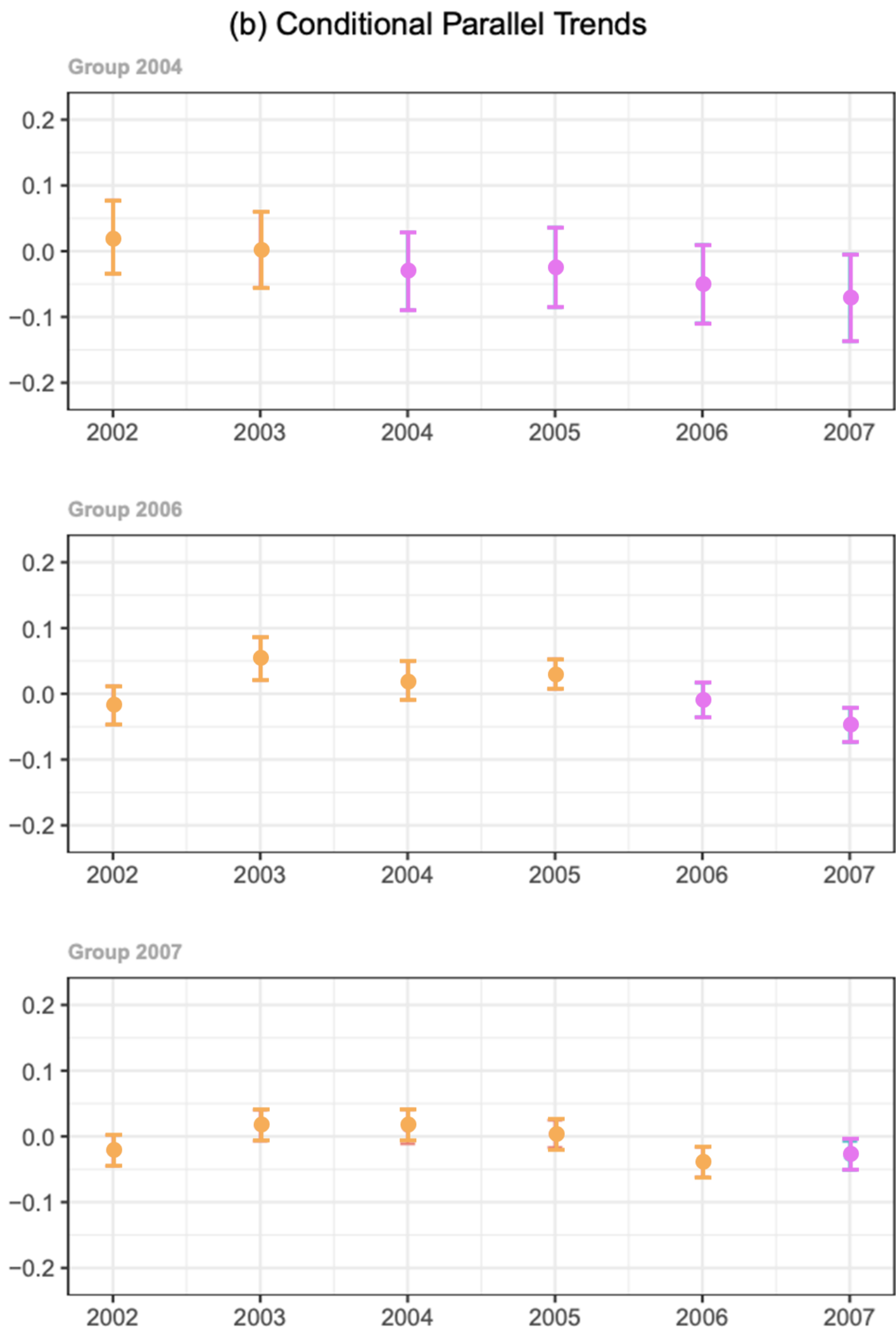
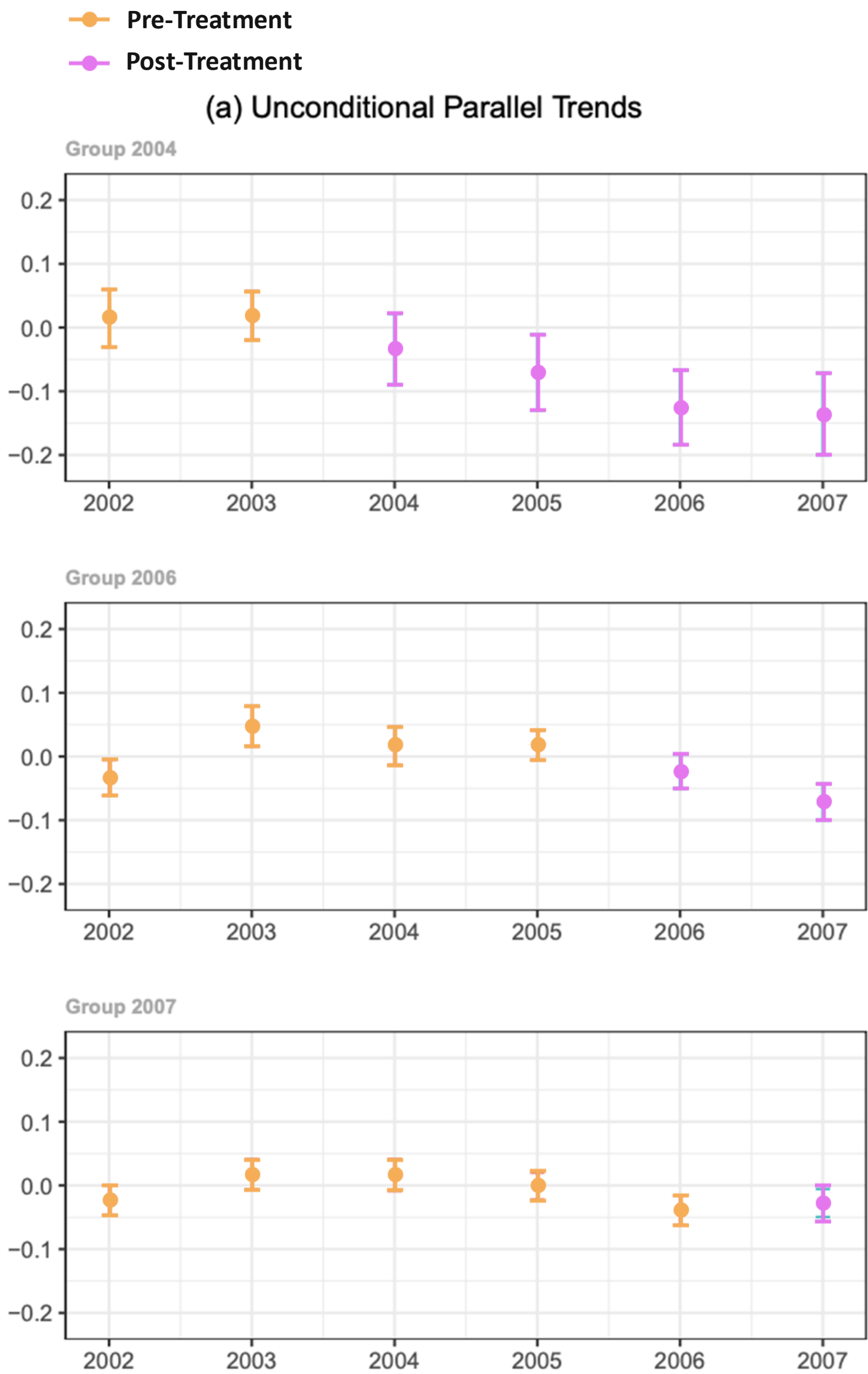
Ограничения DiD

- Предположение о параллельных трендах
- Запрет на изменения в составе групп
- Гетерогенность эффектов



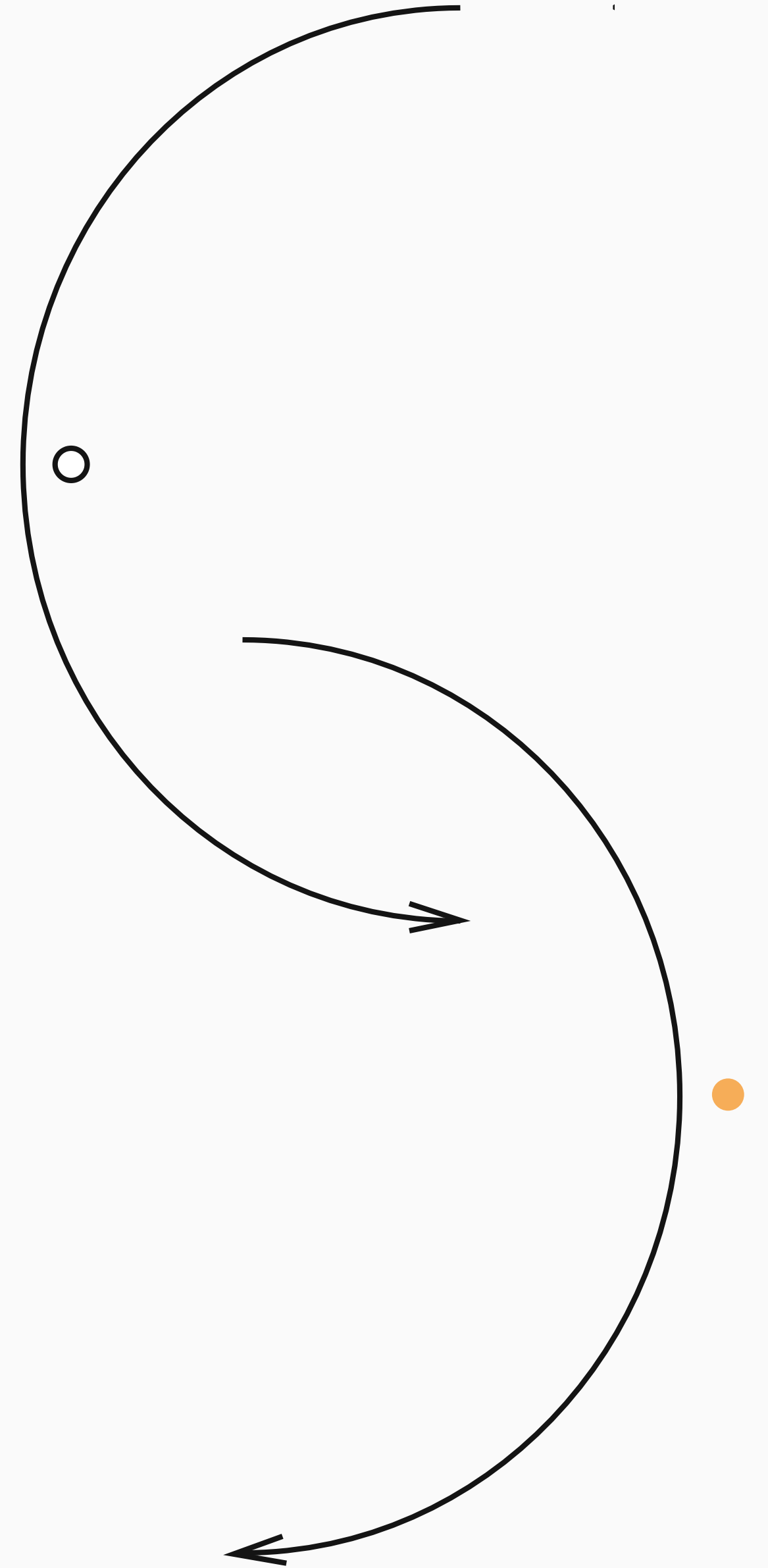
Callaway, B., Sant’Anna, P. H. C. (2021).
"Difference-in-Differences with Multiple
Time Periods". Journal of Econometrics.

Влияние минимальной заработной платы на
занятость подростков



Как можно улучшить?

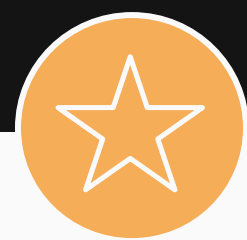
- Добавить нелинейные тренды
- SCM + Diff-in-Diff
Synthetic Difference-in-Differences
- Использовать несколько временных отрезков
Generalized Difference-in-Differences



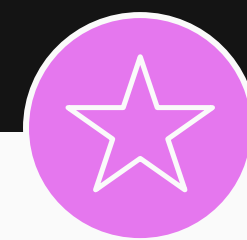
В чем разница между методами?

	RDD	SCM	DiD
Применение	Существует четкая граница, определяющая тестовую и контрольную группы	Необходимо создать контрольную группу из доступных данных.	Имеются данные до и после вмешательства как для тестовой, так и для контрольной группы.
Ограничения	Около порога распределение всех факторов случайное Единственное различие между группами — это наличие/отсутствие эксперимента	Синтетическая контрольная группа отражает данные до начала эксперимента	При отсутствии вмешательства изменения обеих групп были бы одинаковыми
Метод	Сравниваем ATE для наблюдений, которые находятся чуть выше и чуть ниже порога.	Создаем взвешенную комбинацию данных для формирования «синтетической» контрольной группы	Сравнивает изменения во времени для обеих групп

Выводы



Выбор подходящего метода зависит от темы исследования, доступности данных и специфики проведенного эксперимента.



Комбинирование нескольких методов **может** улучшить результаты.

