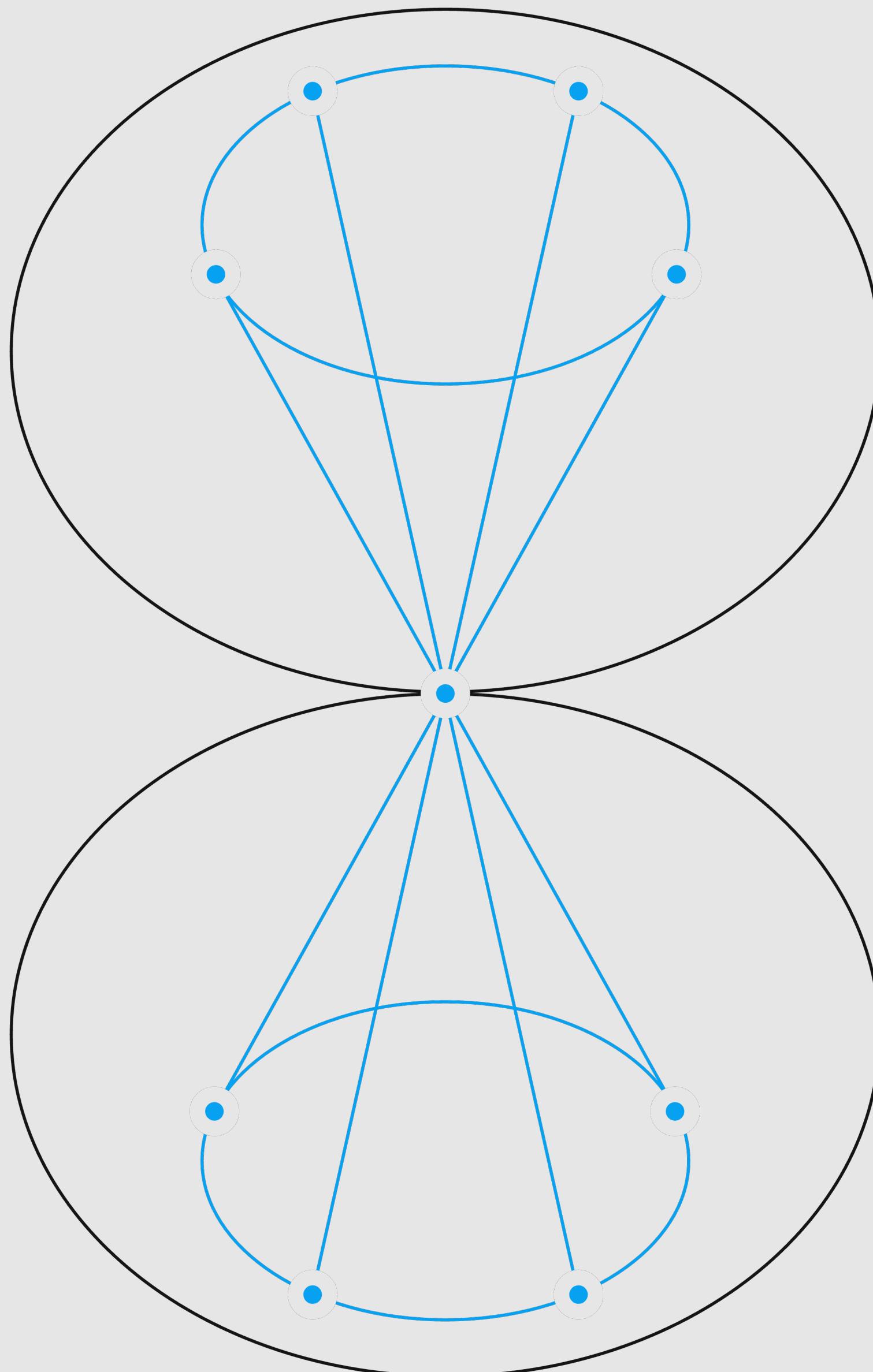


Методы поиска причинности

День 5



- **Корреляция** — это просто связь между двумя переменными.
Она не говорит о том, что одна переменная вызывает изменения в другой.
- **Причинность** — это более глубокое понимание взаимосвязей между переменными.
- **Корреляция не равна причинности.**

ПРИМЕРЫ

На Юпитере год длится
около 12 земных лет.



spurious correlations

correlation is not causation

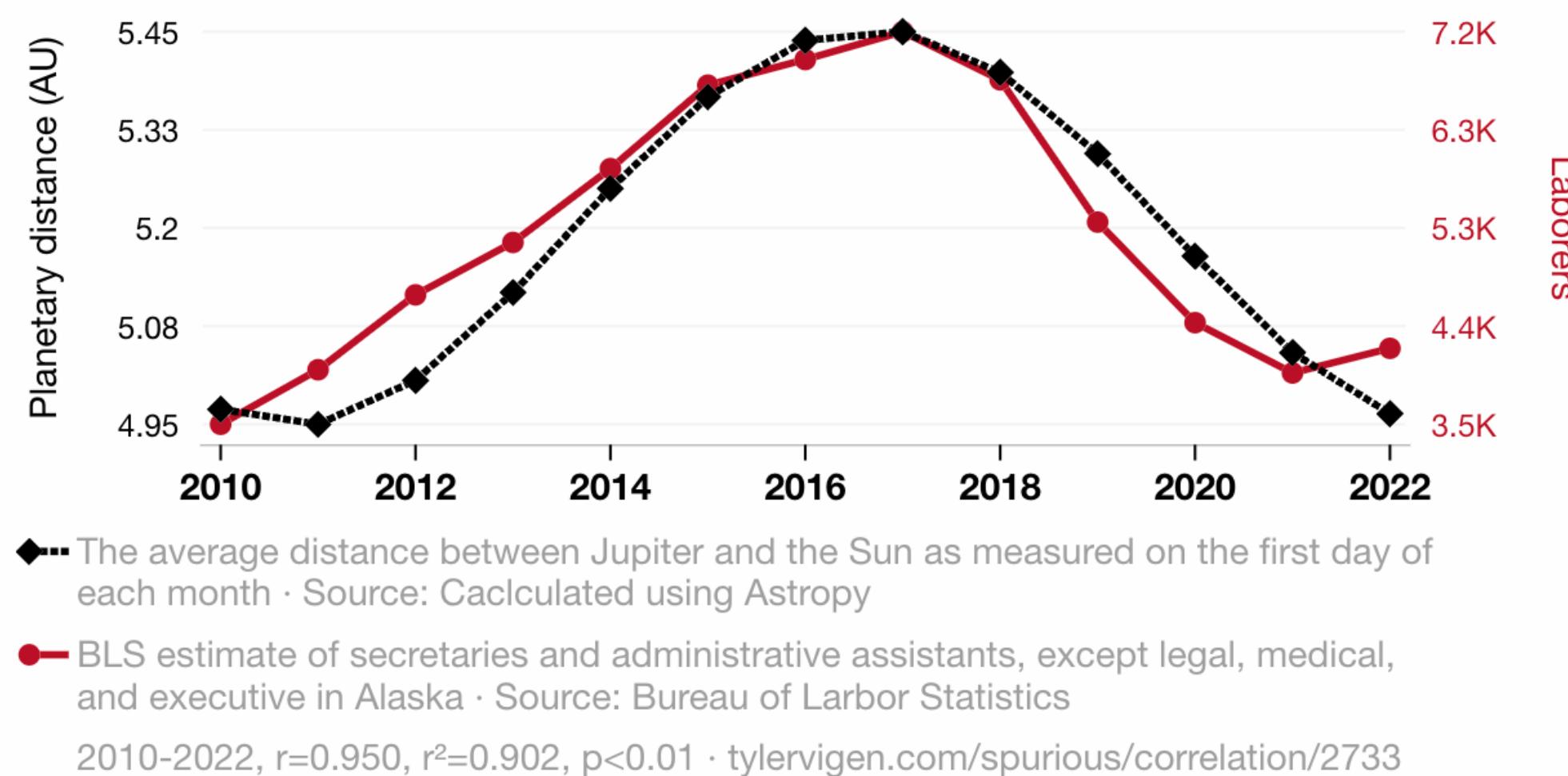
random · discover · [next page →](#)

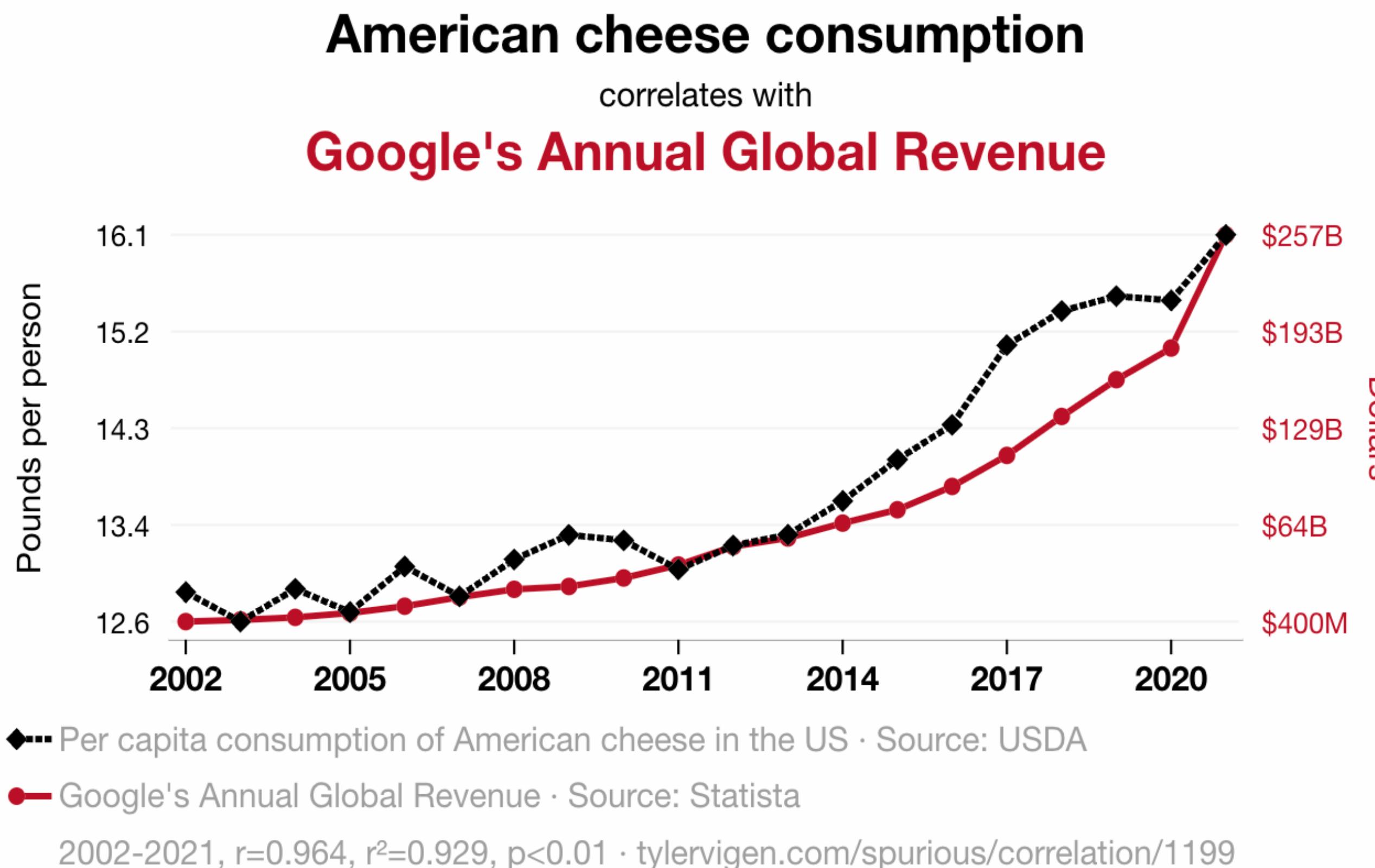
don't miss [spurious scholar](#),
where each of these is an academic paper

The distance between Jupiter and the Sun

correlates with

The number of secretaries in Alaska

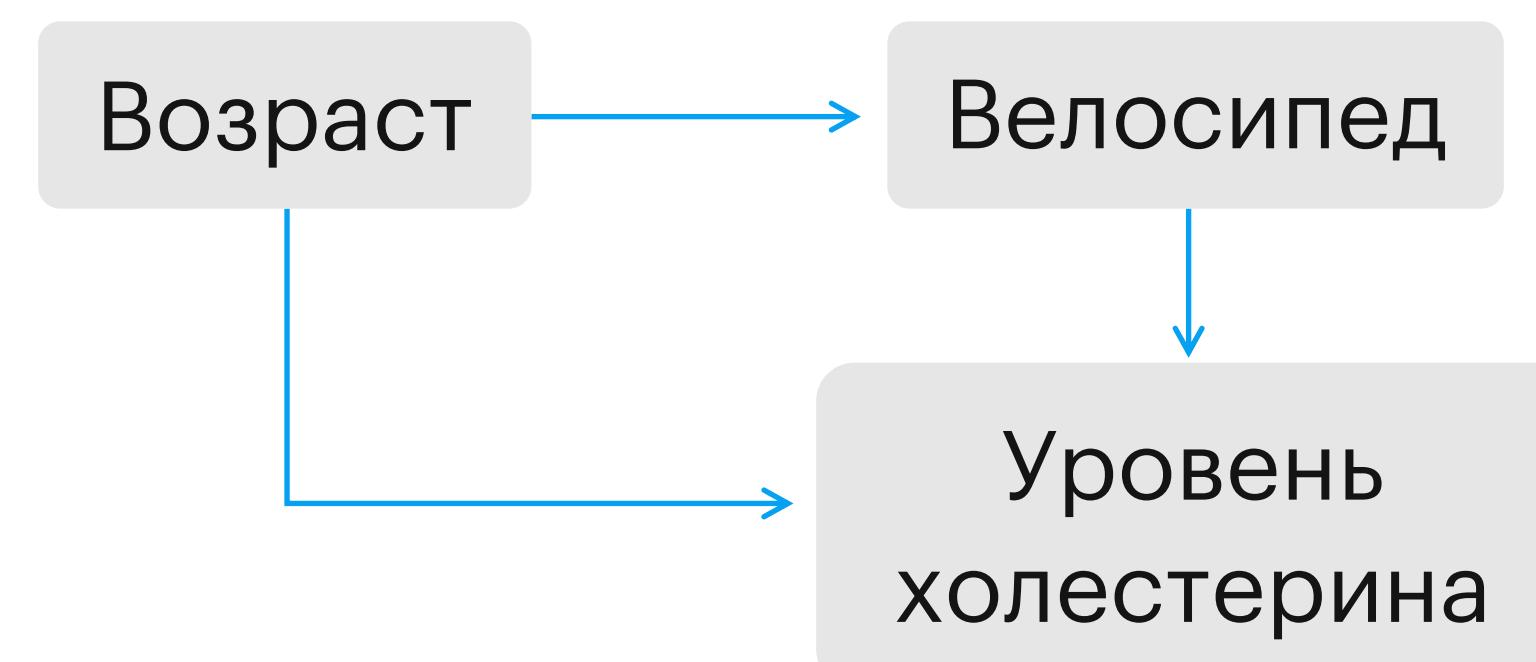




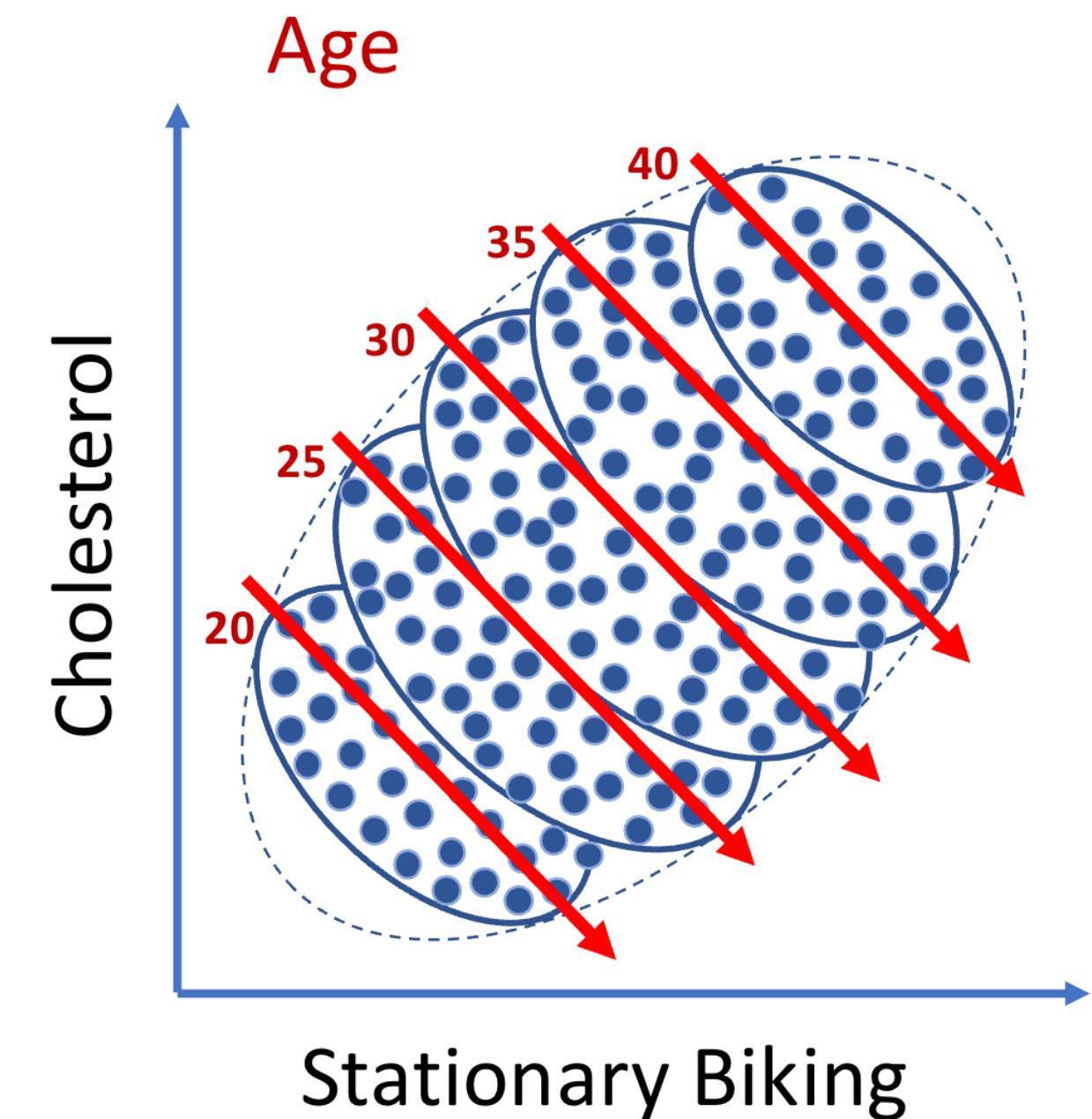
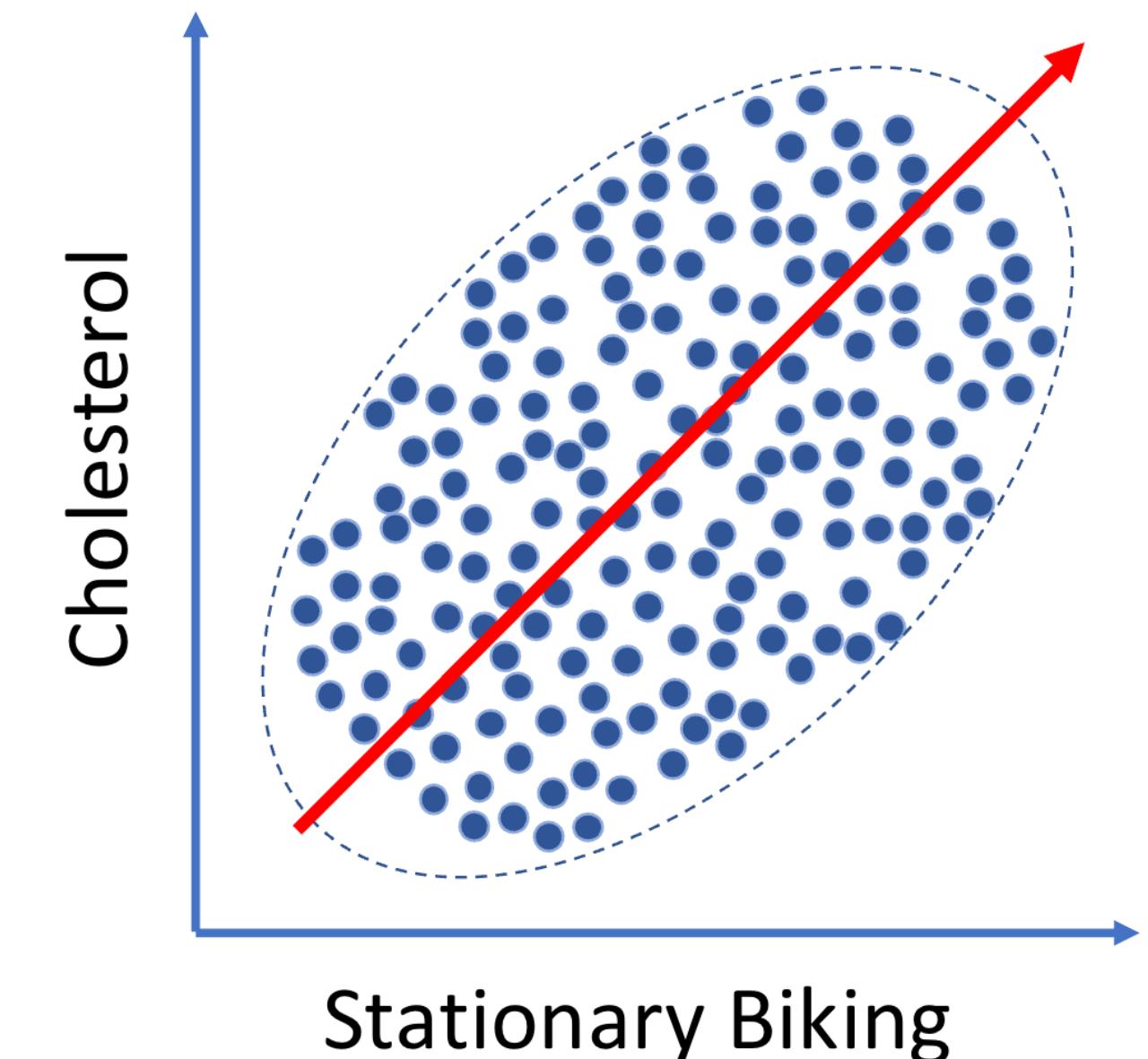
Велосипед →
Уровень холестерина

Возраст → Уровень холестерина

Возраст → Велосипед



Такая переменная, как возраст,
называется спутывающей
переменной — **confounder**.





Парадокс Симпсона: бинарный случай

Лечение камней в почках

| Размер камней | Лечение А | Лечение В |
|-----------------|--------------------------|--------------------------|
| Маленькие камни | Группа 1 – 93% (81/87) | Группа 2 – 87% (234/270) |
| Большие камни | Группа 3 – 73% (192/263) | Группа 4 – 69% (55/80) |
| Оба типа | 78% (273/350) | 83% (289/350) |

$P(\text{success} | A) = 78\% \text{ vs } P(\text{success} | B) = 83\%$

Выбираем лечение В?

Парадокс Симпсона: бинарный случай



$$P(\text{success} \mid A) = 78\% \text{ vs } P(\text{success} \mid B) = 83\%$$

Выбираем лечение В?

Кажется, нет.

Проблема в том, что мы неравномерно распределили пациентов по группам.

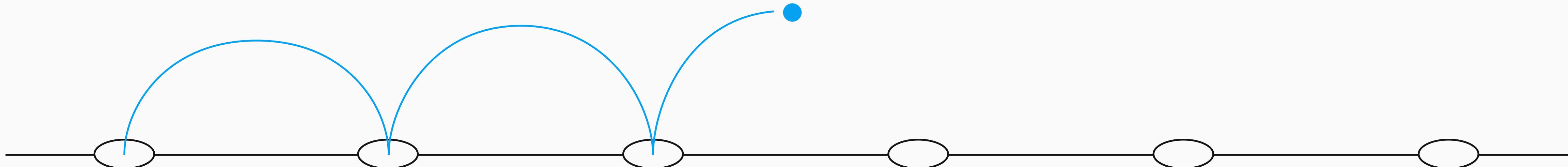
Как можно решить эту проблему?

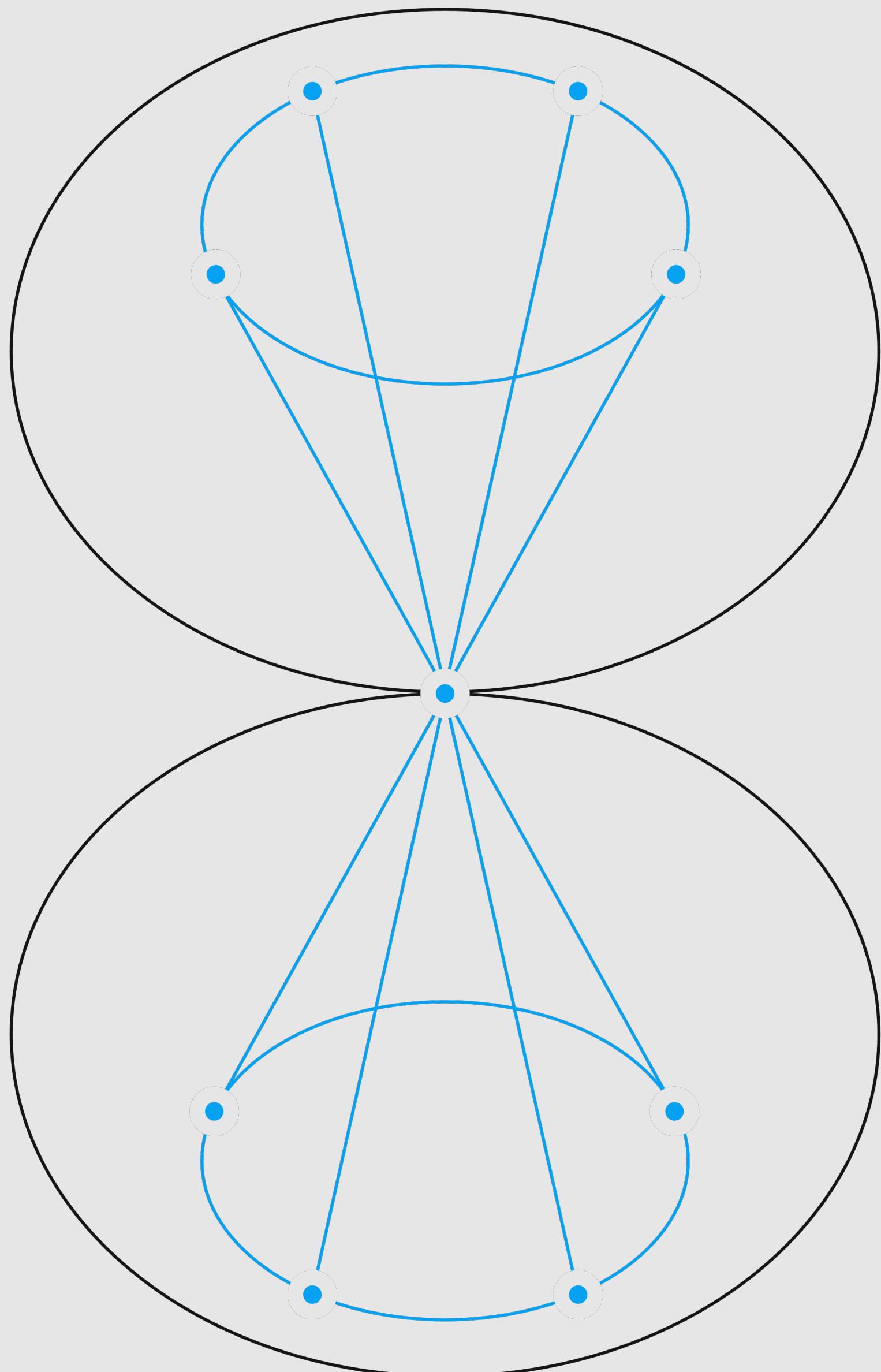
Проверка на А/А-тестах, что мы правильно поставили эксперимент.

Результат А/А-теста

$$P(\text{success} | AA) = \frac{0,93 \times 270 + 0,73 \times 80}{350} = 0,88 \quad , \text{ но в группе A это } 0,78.$$

Одно и то же лечение ведёт к разным результатам?





Учтём размер камней:

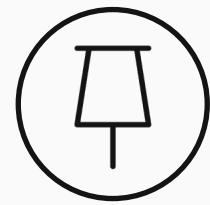
- в группе с маленькими камнями — изменение 6% в пользу A;
- в группе с большими камнями — изменение 4% в пользу A.

Усреднение эффекта по группам — 5% в пользу A.



Вывод: лечение A эффективнее лечения B.

Практическое определение причинности



Определение: Т вызывает Y тогда и только тогда, когда изменение Т приводит к изменению Y при прочих равных условиях.

Причинный эффект — это величина, на которую изменяется Y при изменении Т на единицу. Называется «интервенционистской» интерпретацией причинности.

Н. В. При прочих равных условиях — это вообще возможно?

Лестница причинности от Judea Pearl



| Уровень | Типичная активность | Типичный вопрос | Примеры |
|---|------------------------------|---|---|
| 1. Ассоциация или Предсказание (Association) $P(y x)$ | Наблюдение | Какая связь? Как наблюдение изменения X изменит мой прогноз в Y ? | Что симптом говорит мне о болезни? Что опрос говорит нам о результатах выборов? |
| 2. Действие (Intervention) $P(y do(x), z)$ | Действие, вмешательство | Что если? Что будет, если я сделаю X ? | Что если я приму аспирин, пройдёт ли головная боль? Что если мы запретим сигареты? |
| 3. Контрафакты (Counterfactuals) $P(y_x x', y')$ | Воображение, ретроспекция | Почему? Было ли X причиной Y ? Что, если бы я поступил иначе? | Это аспирин помог от головной боли? Был бы Кеннеди жив, если бы Ос瓦льд не выстрелил? |

Почему недостаточно просто обучить модель?

- Сильная взаимная корреляция между переменными не показывает направление причинности.
- Не все переменные могут быть измерены и учтены в модели. В итоге мы получим смещённую оценку и сделаем неправильные выводы.
- Не понимая причинности, сложнее понимать, что будет происходить в будущем, если процессы изменятся.

ПРИМЕРЫ

Стрелка барометра и давление в кофемашине

- Есть явная корреляция между стрелкой барометра и давлением в кофемашине.
- Но если мы сломаем стрелку барометра, давление в кофемашине не изменится.
- Если мы изменим напряжение и увеличим давление в кофемашине, стрелка барометра поменяет своё положение.

А что показывают методы интерпретации моделей?

ПРИМЕРЫ

- Коэффициенты в линейной регрессии
- Shapley values: SHAP
- LIME

Все они показывают ассоциации или корреляции, но не причинности в общем случае.
Пример с линейной регрессией выше.

SHAP, LIME решают проблему интерпретации моделей в black-box-стиле, когда мы не можем явно выписать параметры модели.



Как оценить причинность?

Идеальный сценарий

Клонировать пациента и дать ему лекарство, а «другому» — нет.

Тогда можно сравнить результаты изменений для конкретного индивида, усреднить результаты по всей выборке и получить причинный эффект или не получить.

К сожалению, это невозможно в реальной жизни.

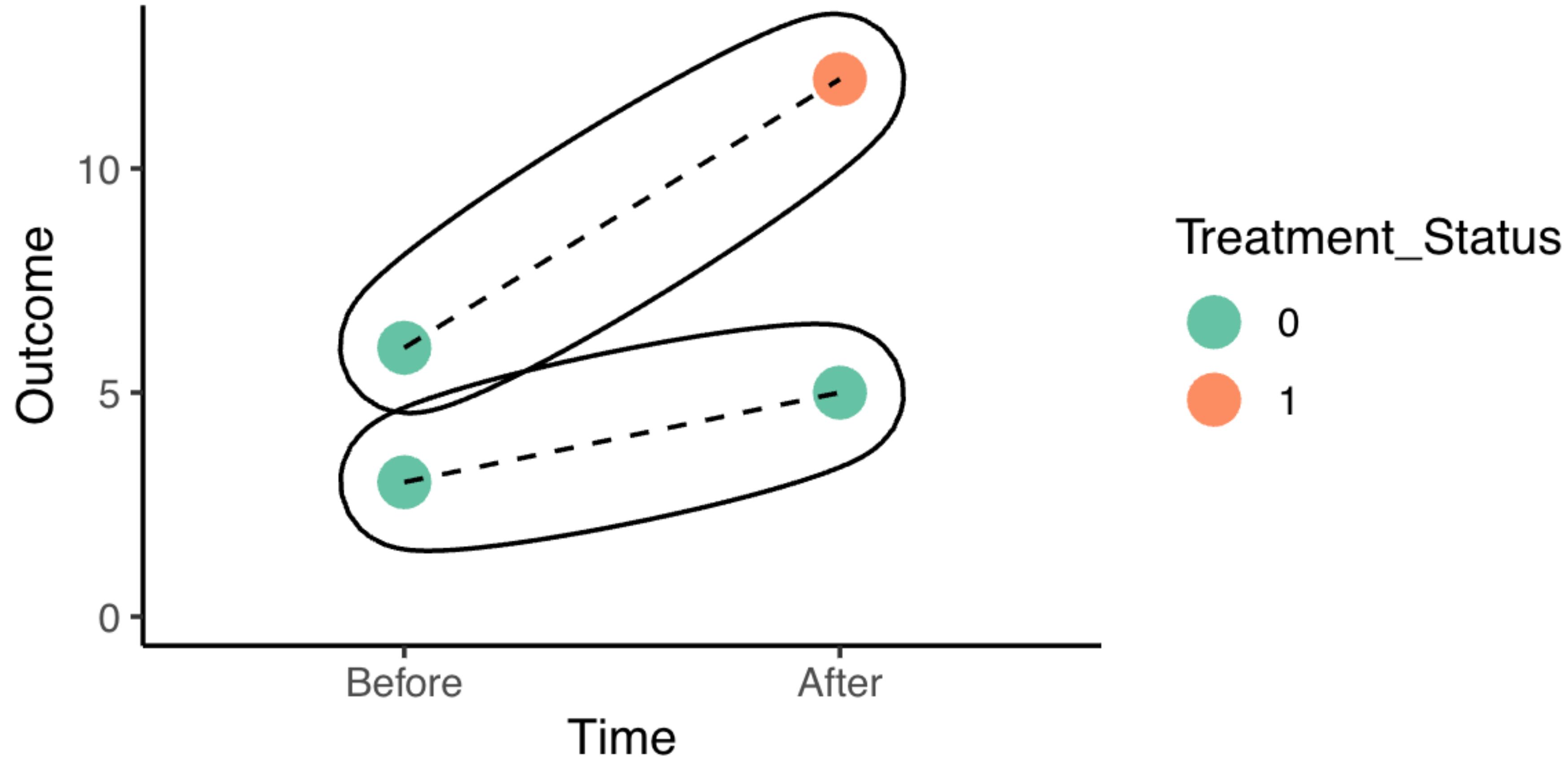
Мы всегда будем иметь дело с неполными данными и неизвестными факторами.

Как оценить причинность?

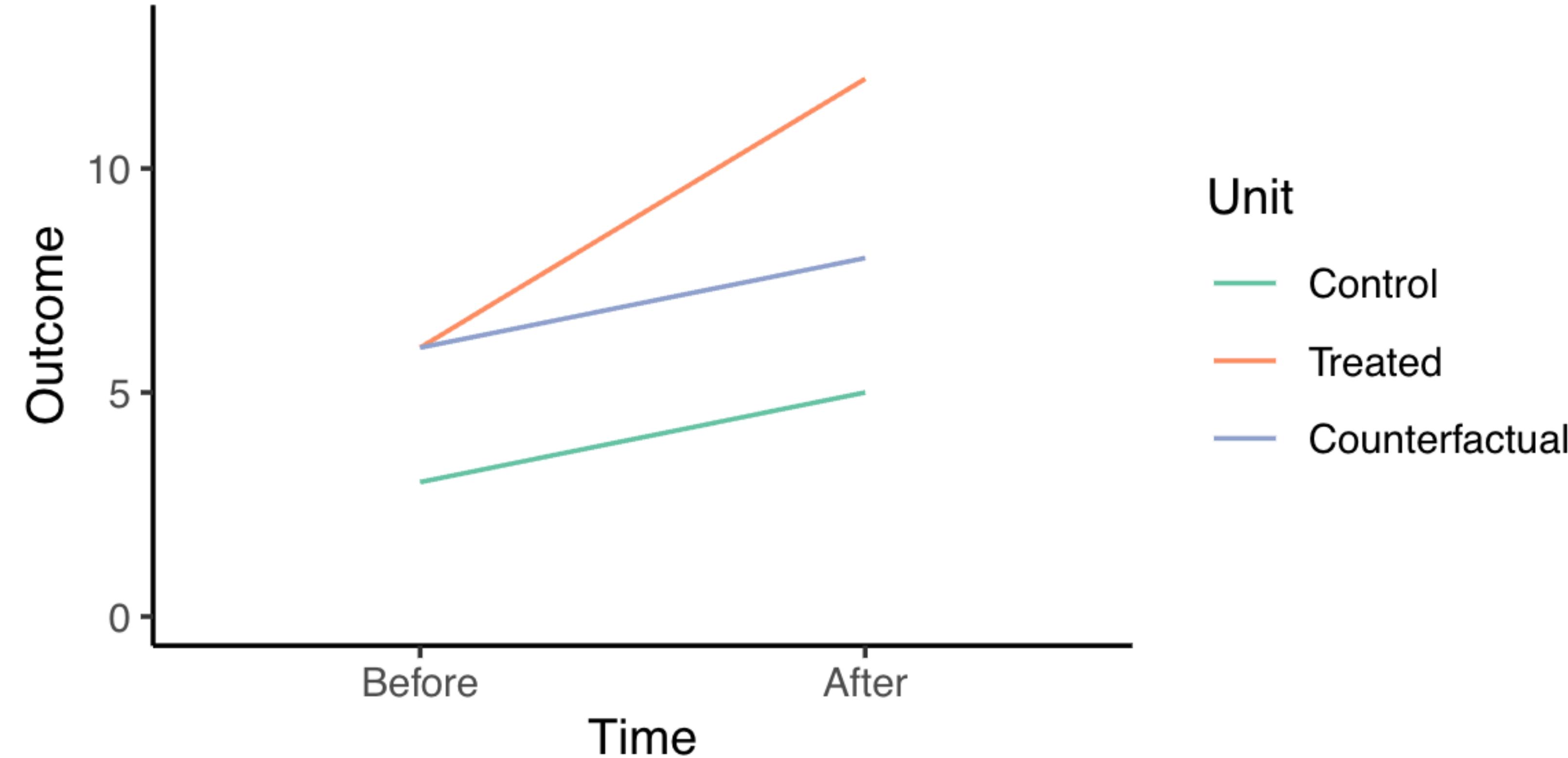
- **Рандомизированный контролируемый эксперимент** — это золотой стандарт оценки причинности. Например, через А/В-тесты.
Под рандомизацией понимается случайное распределение объектов между группами таким образом, что никакие другие факторы не влияют на результат. **Но он не всегда возможен (например, по этическим соображениям).**

- **Наблюдательные исследования** — это данные, которые собраны без специального эксперимента.

Difference-in-differences



Difference-in-differences



w:800

Difference-in-differences

Алгоритм

01

Учим зависимость
на контрольной группе.

02

Переносим обученную
модель на тестовую группу.

03

Строим тренд для тестовой
группы по модели
контрольной группы.

04

Находим разницу между предсказанным
и реальным значениями.

05

Это и будет размером эффекта.

Synthetic difference-in-differences

Проблема: не всегда есть контрольная группа.

Решение: создать синтетическую контрольную группу.

Наш процесс как-то зависит от других факторов.

Значит, на данных факторах можно построить модель и предсказать, что было бы, если бы не было воздействия.

Главное: воздействие не должно влиять на эти факторы.

Difference-in-differences

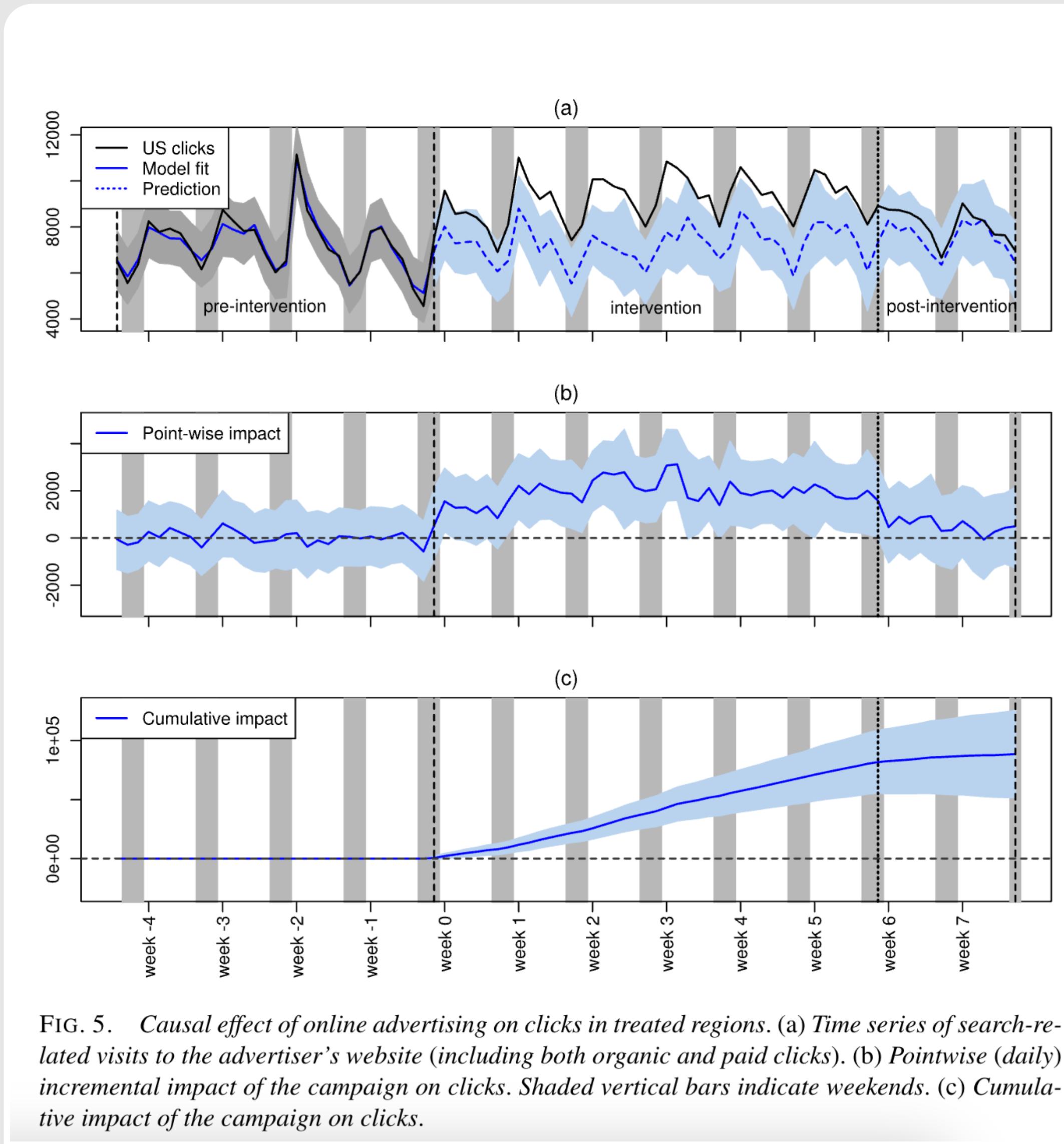


FIG. 5. *Causal effect of online advertising on clicks in treated regions.* (a) Time series of search-related visits to the advertiser's website (including both organic and paid clicks). (b) Pointwise (daily) incremental impact of the campaign on clicks. Shaded vertical bars indicate weekends. (c) Cumulative impact of the campaign on clicks.



Synthetic difference-in-differences

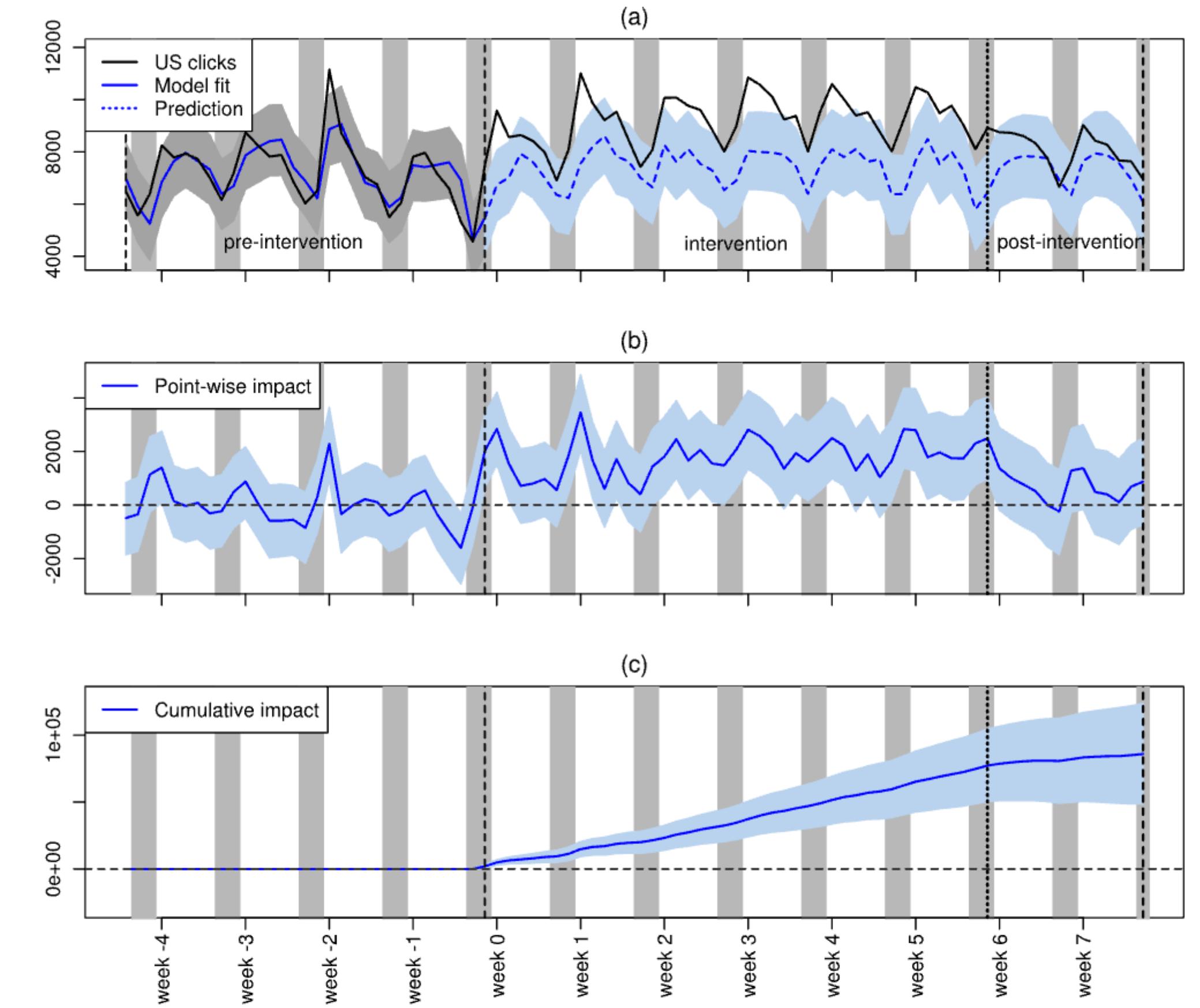


FIG. 6. *Causal effect of online advertising on clicks, using only searches for keywords related to the advertiser's industry as controls, discarding the original control regions as would be the case in studies where a randomised experiment was not carried out.* (a) Time series of clicks on to the advertiser's website. (b) Pointwise (daily) incremental impact of the campaign on clicks. (c) Cumulative impact of the campaign on clicks. The plots show that this analysis, which was based on observational covariates only, provided almost exactly the same inferences as the first analysis (Figure 5) that had been based on a randomised design.



Нет воздействия

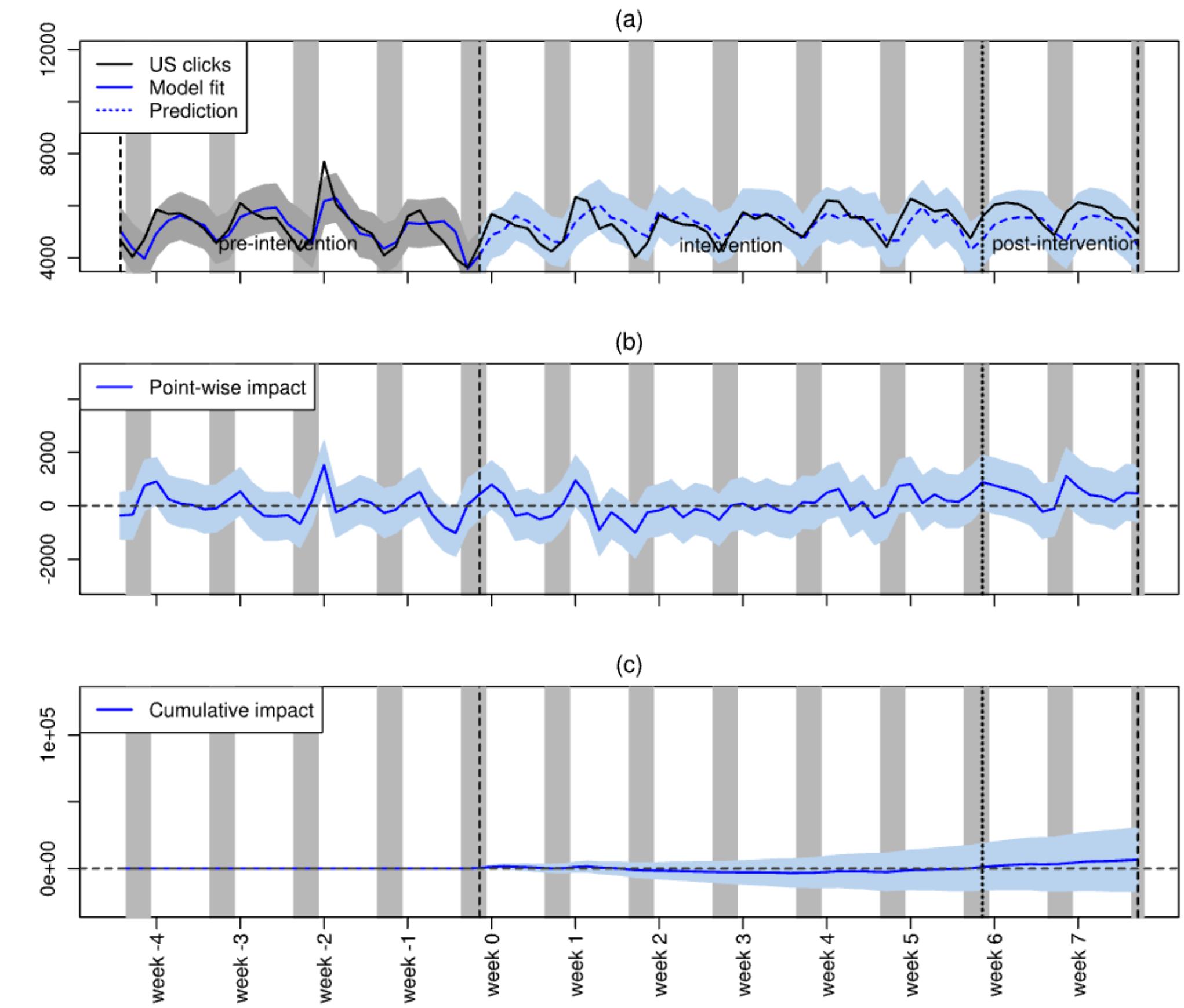


FIG. 7. *Causal effect of online advertising on clicks in nontreated regions, which should not show an effect. Searches for keywords related to the advertiser's industry are used as controls. Plots show inferences in analogy with Figure 5. (a) Time series of clicks to the advertiser's website. (b) Pointwise (daily) incremental impact of the campaign on clicks. (c) Cumulative impact of the campaign on clicks.*

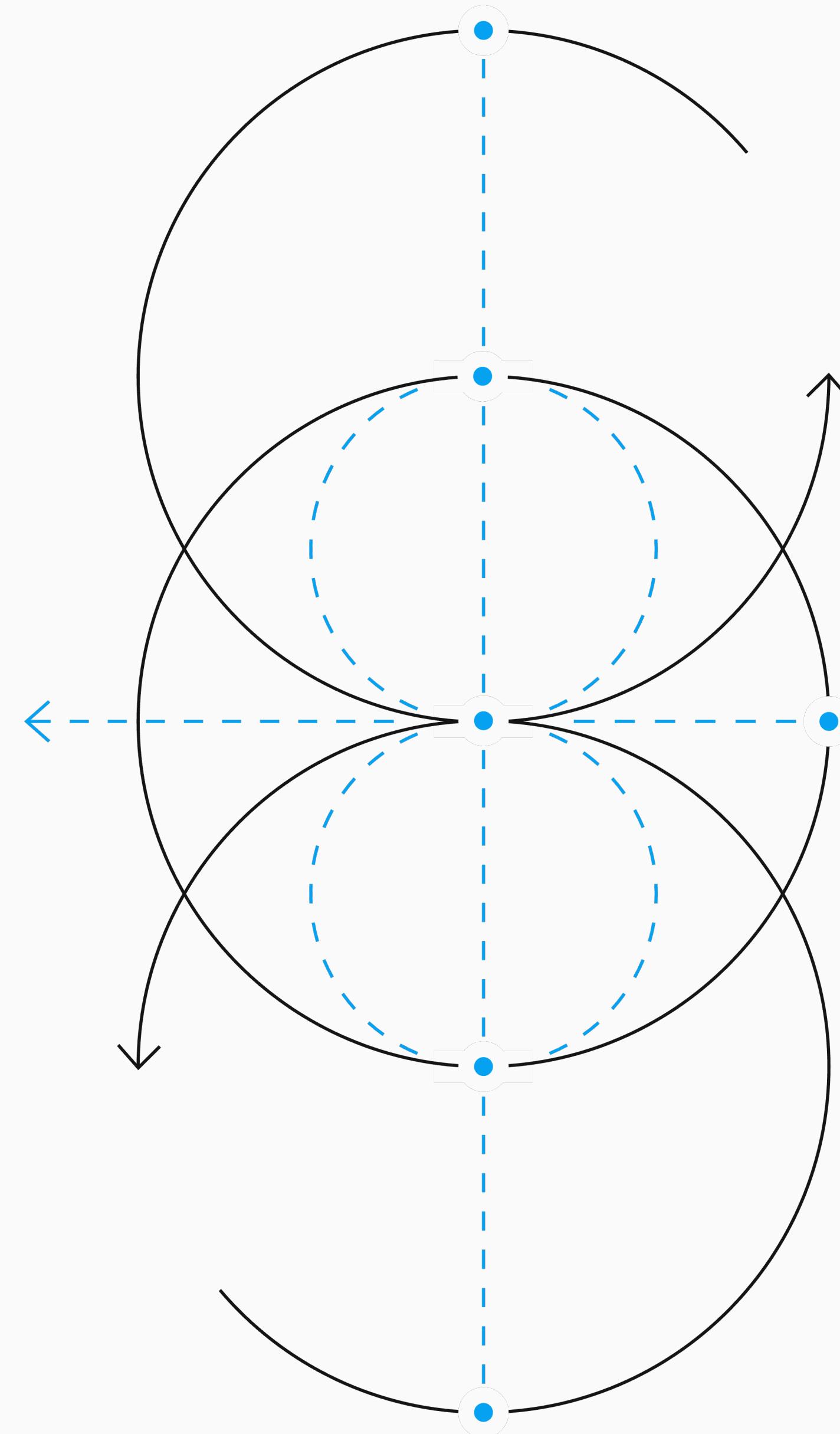
Инструментальные переменные

Немного про Нобелевскую премию 2024 года

Асемоглу и Робинсон получили Нобелевскую премию по экономике за исследования в области экономической истории.

Они показали, что качество институтов влияет на экономическое развитие страны.

Для доказательства они использовали инструментальные переменные и с их помощью получили оценки этого влияния.



Термины

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

- x_i — значения объясняющей переменной (регрессора);
- y_i — значения объясняемой (зависимой) переменной;
- ε_i — случайные ошибки модели.
- Если регрессор не коррелирован со случайной ошибкой, $\text{cov}(x_i, \varepsilon_i) = 0$,
то он называется **экзогенным**.
- Если регрессор коррелирован со случайной ошибкой $\text{cov}(x_i, \varepsilon_i) \neq 0$,
то он называется **эндогенным**.

Причинно-следственные связи в эконометрике

$$y_i = \beta_1 + \beta_2 \times x_i + \varepsilon_i \quad i = 1, 2, \dots, n$$

Вопрос о причинно-следственной связи:

является ли изменение переменной x причиной для изменения переменной y ?

- Для ответа на этот вопрос нужно получить несмещённую оценку коэффициента β_2 .
- Для этого часто используют метод наименьших квадратов.
- В каких случаях это действительно хорошая идея?

Оценка коэффициента β_2 в модели: какая есть проблема?

$$\begin{aligned}\widehat{\beta}_2 &= \frac{\text{cov}(x_i, y_i)}{\text{var}(x_i)} = \frac{\text{cov}(x_i, \beta_1 + \beta_2 \cdot x_i + \varepsilon_i)}{\text{var}(x_i)} \\ &= \frac{\text{cov}(x_i, \beta_1) + \beta_2 \cdot \text{cov}(x_i, x_i) + \text{cov}(x_i, \varepsilon_i)}{\text{var}(x_i)} = \beta_2 + \frac{\text{cov}(x_i, \varepsilon_i)}{\text{var}(x_i)}\end{aligned}$$

Если $\text{cov}(x_i, \varepsilon_i) = 0$, то $\widehat{\beta}_2 \xrightarrow{p} \beta_2$.



Вывод: чтобы МНК-оценка коэффициента β_2 была несмешенной, необходима **экзогенность** регрессора.

Типичные причины эндогенности

Если предпосылка об экзогенности регрессора нарушена, то обычный метод наименьших квадратов даёт смещённые. Типичные случаи, в которых следует ожидать нарушения этой предпосылки:

01

двусторонняя причинно-следственная связь;

02

ошибки измерения объясняющей переменной;

03

смещение из-за пропуска существенной переменной. В том числе из-за selection bias. Т. е. наш регрессор мог включить вклад влияния других переменных, которые не учтены в модели.

Эндогенность из-за пропуска существенной переменной: как это выглядит?

- Пусть на зависимую переменную влияют не только регрессор x , но и ещё один фактор w :

$$y_i = \beta_1 + \beta_2 \times x_i + \beta_3 \cdot w_i.$$

- Представим, что мы игнорируем второй фактор и по-прежнему оцениваем парную регрессию переменной y по переменной x .

В этом случае пропущенный фактор «прячется» внутри случайной ошибки:

$$y_i = \beta_1 + \beta_2 \times x_i + \varepsilon_i, \quad \varepsilon_i = \beta_3 \times w_i.$$

- Если переменная x коррелирована с пропущенным фактором, то она коррелирована со случайной ошибкой. То есть эндогенна.

Эндогенность из-за пропуска существенной переменной

РЕШЕНИЯ

Ситуация 1 (простая): пропущенная переменная наблюдаема.

- Это ситуация, в которой у вас есть данные о пропущенной переменной. Тогда для решения проблемы нужно просто добавить пропущенную переменную в модель.
- Переменные, которые вы добавляете в модель, чтобы устранить смещение оценки нужного вам коэффициента, называют **контрольными**.



Эндогенность из-за пропуска существенной переменной

РЕШЕНИЯ

Ситуация 2 (сложная): пропущенная переменная не наблюдаема.

- Используйте контролируемый эксперимент.
- Используйте замещающие переменные или прокси-переменные.

Замещающей переменной называется переменная, которая сильно коррелирована с ненаблюдаемой существенной переменной и которая при этом является наблюдаемой.

- Если ни то ни другое недоступно, используйте инструментальные переменные.

Инструментальные переменные

$$y_i = \beta_1 + \beta_2 \cdot x_i + \varepsilon_i$$

- x — эндогенный регрессор: $\text{cov}(x_i, \varepsilon_i) \neq 0$.
- Пусть в нашем распоряжении есть дополнительная переменная z , которая удовлетворяет двум свойствам.
 - **Экзогенность:** переменная не коррелирована со случайными ошибками: $\text{cov}(z_i, \varepsilon_i) = 0$.
 - **Релевантность:** переменная коррелирована с регрессором, т. е. переменная должна быть полезна и иметь какой-то смысл: $\text{cov}(x_i, z_i) \neq 0$.
- Пусть в нашем распоряжении есть дополнительная переменная z , которая удовлетворяет двум свойствам — экзогенности и релевантности.
- Такая переменная называется **инструментальной**.
- Тогда можно получить несмешенную оценку параметра β_2 , используя **двуэтаповый МНК**.

Идея двухшагового МНК (2МНК)

Two-stage Least Squares (TSLS, 2SLS)

- Первый шаг: оцениваем регрессию: $\hat{x}_i = \hat{\theta}_1 + \hat{\theta}_2 z_i$.
Получаем прогнозные значения \hat{x}_i .
- Второй шаг: оцениваем регрессию: $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 \hat{x}_i$.

Идея двухшагового МНК (2МНК)

$$\text{Cov}(z_i, \varepsilon_i) = 0 \quad \Rightarrow \quad \text{Cov}(\hat{x}_i, \varepsilon_i) = 0$$

\hat{x}_i — линейно выражено через z_i

$$\text{Cov}(\hat{x}_i, \varepsilon_i) = 0$$



Проблема эндогенности регрессора решена.

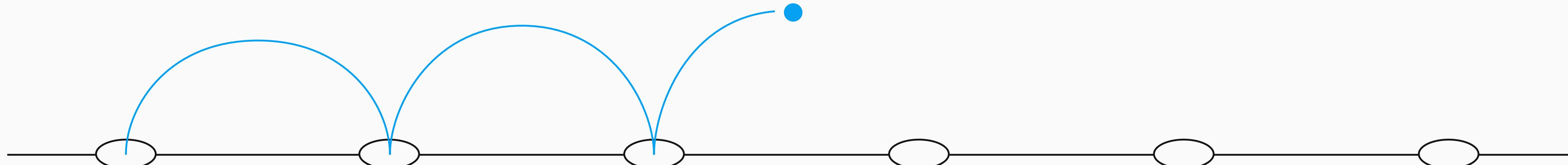
2МНК-оценка и её состоятельность

$$\widehat{\beta_2^{TSLS}} = \frac{\widehat{\text{Cov}(y, z)}}{\widehat{\text{Cov}(x, z)}} \xrightarrow{p} \frac{\text{Cov}(y_i, z_i)}{\text{Cov}(x_i, z_i)}$$
$$= \frac{\text{Cov}(\beta_1 + \beta_2 \cdot x_i + \varepsilon_i, z_i)}{\text{Cov}(x_i, z_i)} = \frac{\beta_2 \text{Cov}(x_i, z_i) + \text{Cov}(\varepsilon_i, z_i)}{\text{Cov}(x_i, z_i)} = \beta_2$$

Где взять подходящие инструменты?



- Это самый сложный вопрос во всех исследованиях, использующих инструменты. Стандартного ответа нет.
- Обычно помогает хорошее содержательное понимание исследуемого вопроса и процессов.
 - Это помогает найти тот самый фактор, который вызывает изменения регрессора, но не связан со случайными ошибками.
 - После того, как вы придумали инструменты из содержательных соображений, не забудьте подвергнуть их формальным тестам на релевантность и экзогенность.
- Насмотренность.



Acemoglu, Robinson (2001), исследовательский вопрос

Влияет ли качество институтов на экономическое развитие?

- Регрессор (x) – Average protection against expropriation risk – переменная, которая характеризует среднее качество институтов за период с 1985 по 1995 год. Переменная измерена от 0 до 10, где более высокие значения характеризуют более высокий уровень защиты прав собственности.

- Объясняемая переменная (y) – логарифм ВВП на душу населения в 1995 году.

Acemoglu, Robinson (2001), в чём сложность?

- Наверняка есть прочие (в том числе ненаблюдаемые) факторы, которые сегодня влияют на экономическое развитие.
- Возможно, что причинно-следственная связь является двусторонней: не только хорошие институциональные условия способствуют росту экономики, но и, наоборот, более богатые экономики могут позволить себе лучшие институты.
- Скорее всего, объясняющая переменная измеряется с ошибкой.

Всё это влияет на состоятельность оценки коэффициента β_2 в регрессии.

Acemoglu, Robinson (2001), данные

- Данные по 110 странам мира (пространственные данные).
- Из них 64 страны являются бывшими колониями, что оказывается важным далее, при выборе инструмента.
- В качестве контрольных переменных использовали переменные регионов мира (например, Азии и Африки) и широту, на которой расположена страна.

Acemoglu, Robinson (2001), инструмент



- **Инструмент:** смертность колонистов во время колонизации данной страны.

- **Идея** состоит в том, что в колониях, где завоеватели часто умирали от малярии и прочих напастей, они не собирались жить сами, а потому не развивали там качественных институтов. Поэтому между смертностью колонистов и будущим качеством институтов должна быть отрицательная корреляция (релевантность инструмента).

- С другой стороны, сегодняшние экономические переменные никак не могут влиять на смертность колонистов несколько веков назад (экзогенность инструмента).

Acemoglu, Robinson (2001), что получилось

- **Высокое качество институтов** (в терминологии авторов — **инклюзивные институты**) способствует экономическому росту.

- **Оценки коэффициента** при переменной интереса в 2МНК-регрессии существенно выше, чем в регрессии, оценённой без использования инструментов.



Angrist, 1990, в чём сложность

Lifetime Earnings and the Vietnam Era Draft Lottery:
Evidence from Social Security Administrative Records

- Простое сравнение средних доходов людей, которые служили и не служили в армии, показывает, что доходы ветеранов устойчиво ниже.
- Такой эффект наблюдается для ветеранов всех рас.
- Эти оценки не вызывают доверия, так как они могут быть смещены из-за эндогенности решения о прохождении службы.
- Возможно, дело в том, что в армию идут менее способные к гражданской работе люди.
- Подход с инструментальными переменными позволяет преодолеть эту проблему.

Angrist, 1990, данные

- Ангрист использует данные о ветеранах войны во Вьетнаме. В те времена в США использовался призыв на военную службу.

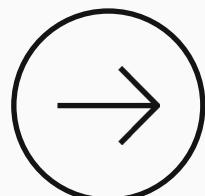
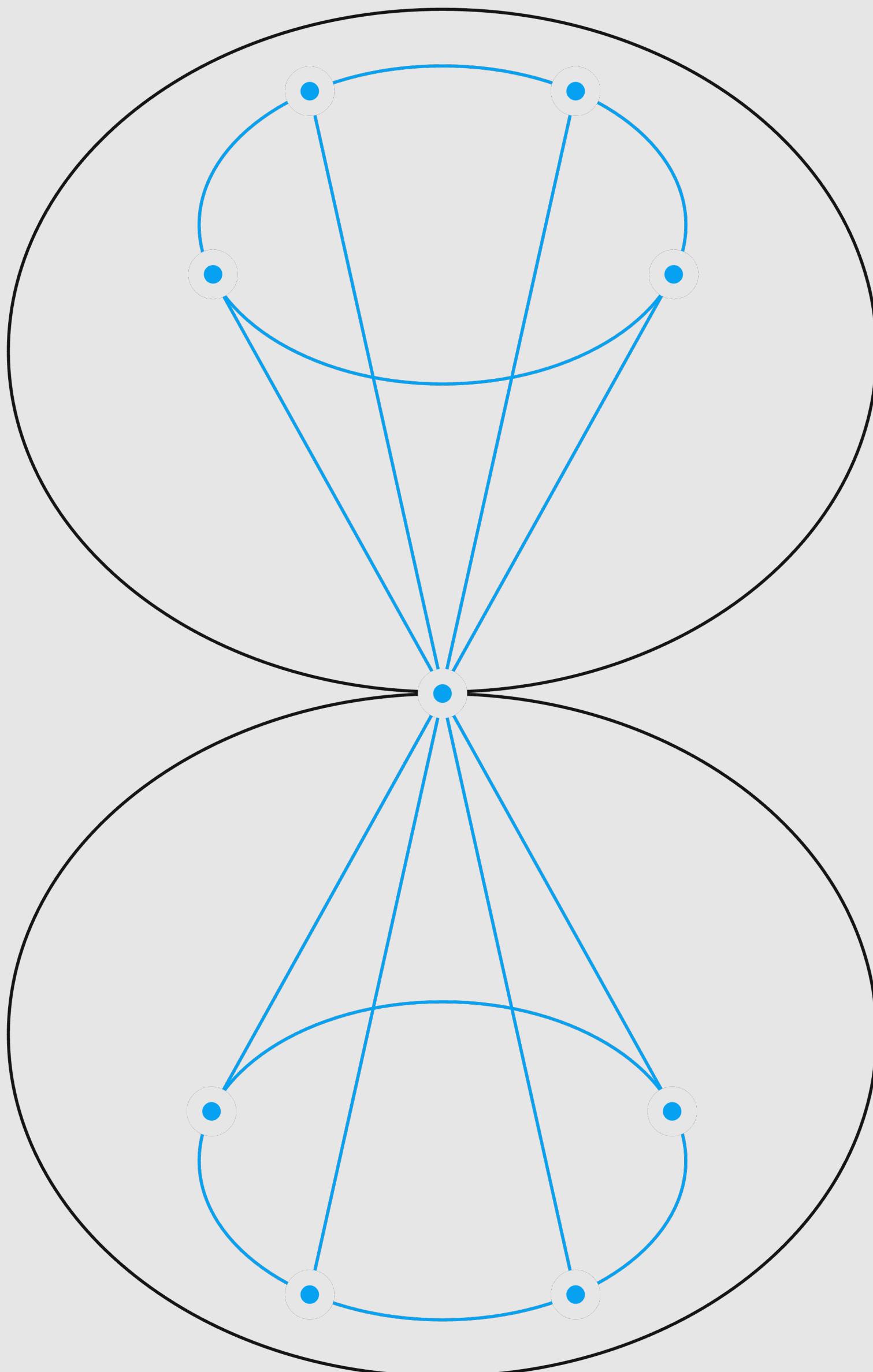
- Приоритетность призыва зависела от так называемого случайного порядкового номера (**Random Sequence Number**, RSN). Этот номер присваивался каждому мужчине в результате розыгрыша лотереи. Номер изменялся от 1 до 365, так как был привязан к дате рождения.

- Министерство обороны каждый год определяло некоторый потолок (пороговое значение) случайного порядкового номера. После этого на службу призывались все мужчины с RSN ниже этого порогового значения.

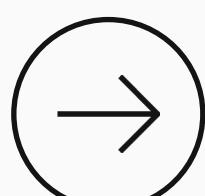
Angrist, 1990, инструмент

- Инструментальная переменная — статус победителя лотереи.

- Этот статус вовсе не равен статусу ветерана войны (это две разные переменные).
 - С одной стороны, вовсе не все ветераны войны были победителями лотереи, ведь кто-то записывался на службу добровольно.
 - С другой стороны, не все победители лотереи стали ветеранами, так как кто-то из победителей избежал службы в силу, например, медицинских ограничений.



Тем не менее статус победителя лотереи и статус ветерана положительно коррелированы друг с другом: победители лотереи в среднем оказывались в армии с большей вероятностью, чем остальные мужчины (**релевантность**).



Статус победителя лотереи не коррелирован с прочими характеристиками индивида, которые могут влиять на его доход (**экзогенность**).

Angrist, 1990, что получилось

- Ответ на вопрос о последствиях службы в армии зависит от расы ветерана.

- По прошествии десяти лет после увольнения со службы белые ветераны зарабатывали значительно меньше, чем не ветераны. Потери доходов белых ветеранов составляют порядка **15%** от годовой заработной платы.
И дело не в способностях, а именно в службе в армии как таковой.

- Напротив, для не белых ветеранов значимых потерь в заработной плате не выявлено.

Причинность по Грейндже

Причинность по Грейндже

Этот метод утверждает, что X **вызывает** Y по Грейндже, если прошлые значения X уменьшают дисперсию в предсказании Y , то есть:

$$\text{Var}[Y_t - P(Y_t \mid H_{$$

где:

- $H_{}$ — история всех релевантных переменных до момента времени t ;
- $H_{ — история до момента времени t , исключая все значения X до момента времени t .$

На практике $H_{}$ включает значения всех известных и релевантных переменных в задаче вплоть до времени $t - 1$.

Интерпретация причинности по Грейнджеру

- Причинность по Грейнджеру предполагает, что улучшение предсказания означает потенциальный причинный эффект.

- Однако утверждается, что это необязательно устанавливает истинные причинно-следственные связи.

Тест Грейнджера лишь показывает, что включение прошлых значений X улучшает предсказание Y по сравнению с исключением X .

Моделирование причинности по Грейнджеру

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_m y_{t-m} + \text{error}_t$$

Затем авторегрессия расширяется включением лагов переменной x .

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_m y_{t-m} + b_p x_{t-p} + \cdots + b_q x_{t-q} + \text{error}_t$$

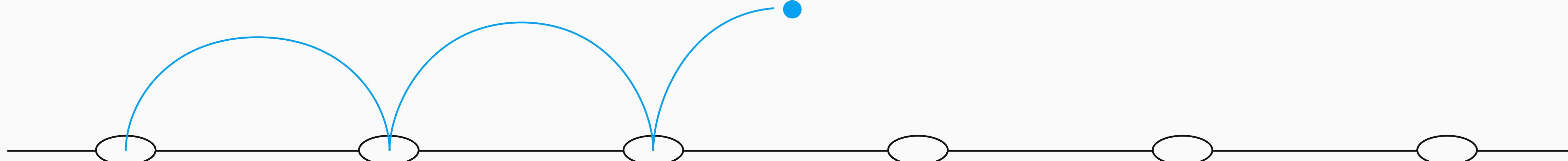
В обозначениях приведённой выше расширенной регрессии:

- p — минимальная длина лага;
- q — максимальная длина лага, при которой лаговое значение x остаётся значимым.

Моделирование причинности по Грейнджеру

Тесты для проверки:

- проверка коэффициентов на значимость при x ;
- если x не объясняет y , то коэффициенты должны быть равны 0;
- F-тест для сравнения дисперсий ошибок двух моделей.



Причинность по Грейнджеру

Основные предположения

- **Стационарность** — временной ряд должен быть стационарным.
- **Линейность** — зависимость между переменными должна быть линейной.
- **Отсутствие ненаблюдаемых переменных** — модель должна учитывать все релевантные переменные; достаточно сложное условие.

Выявление причинно-следственных связей – Causal Discovery и Structural Causal Models

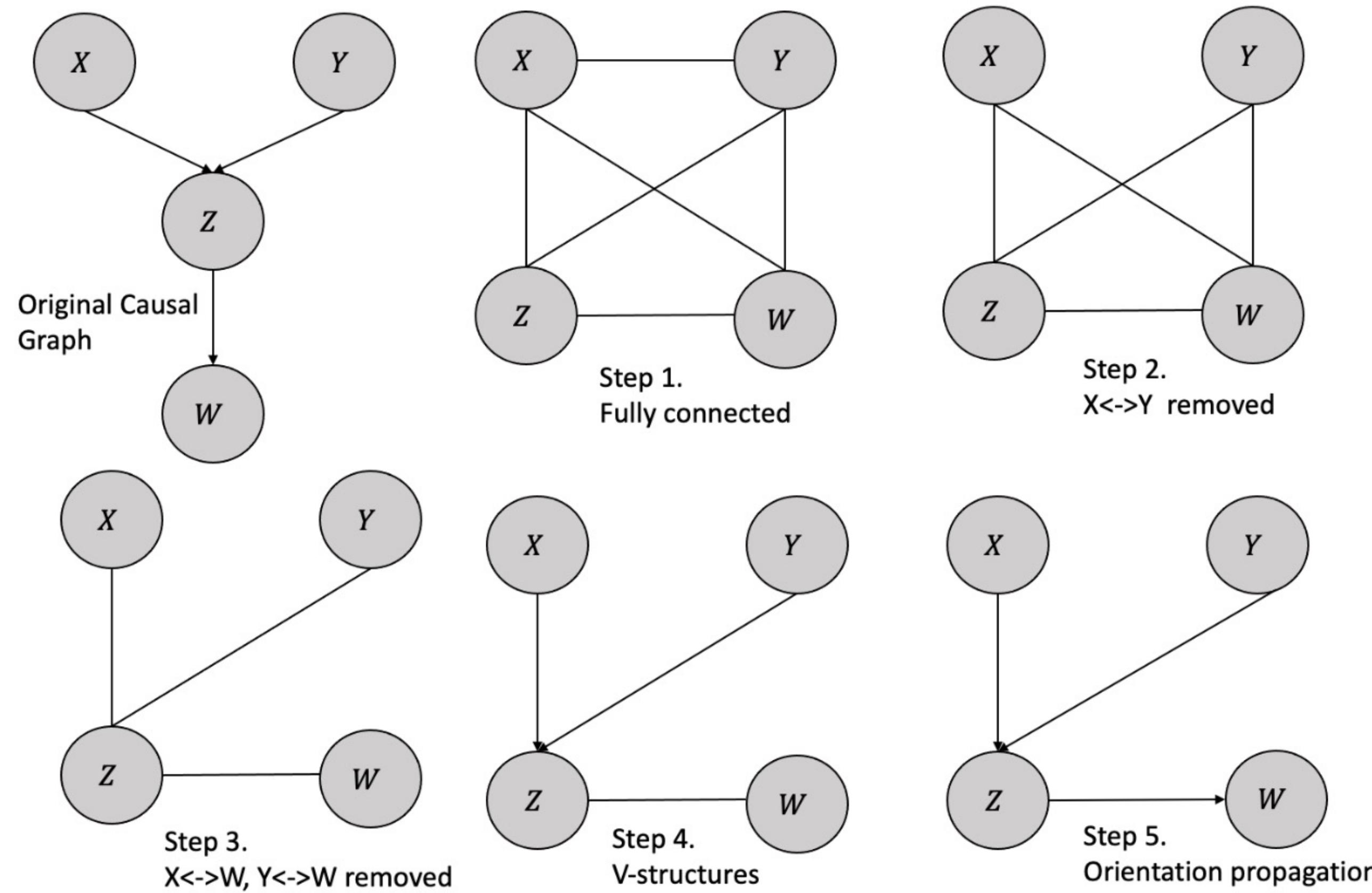


Мы обсудили методы оценки причинности, которые предполагают, что мы задали как-то связи и направления между переменными, сделали это экспертно или на основе наблюдаемых предположений.

Но что делать, если мы не знаем, какие переменные влияют на другие и как качественно они влияют?



Выявление причинно-следственных связей – Causal Discovery и Structural Causal Models



Как на практике

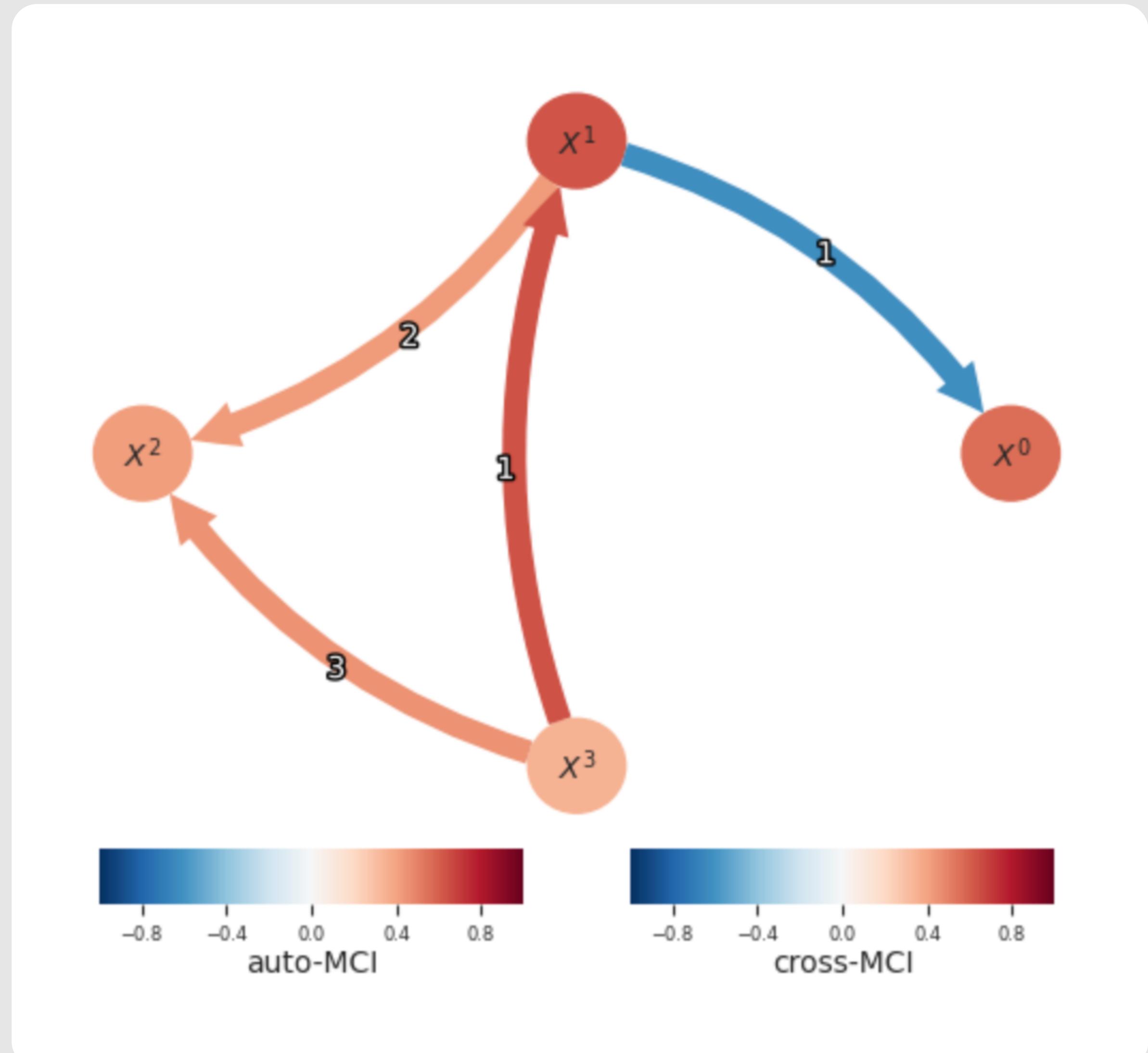
Tigramite – библиотека для работы с графами причинности на временных рядах.

$$X_t^0 = 0,7X_{t-1}^0 - 0,8X_{t-1}^1 + \eta_t^0$$

$$X_t^1 = 0,8X_{t-1}^1 + 0,8X_{t-1}^3 + \eta_t^1$$

$$X_t^2 = 0,5X_{t-1}^2 + 0,5X_{t-2}^1 + 0,6X_{t-3}^3 + \eta_t^2$$

$$X_t^3 = 0,7X_{t-1}^3 + \eta_t^3$$

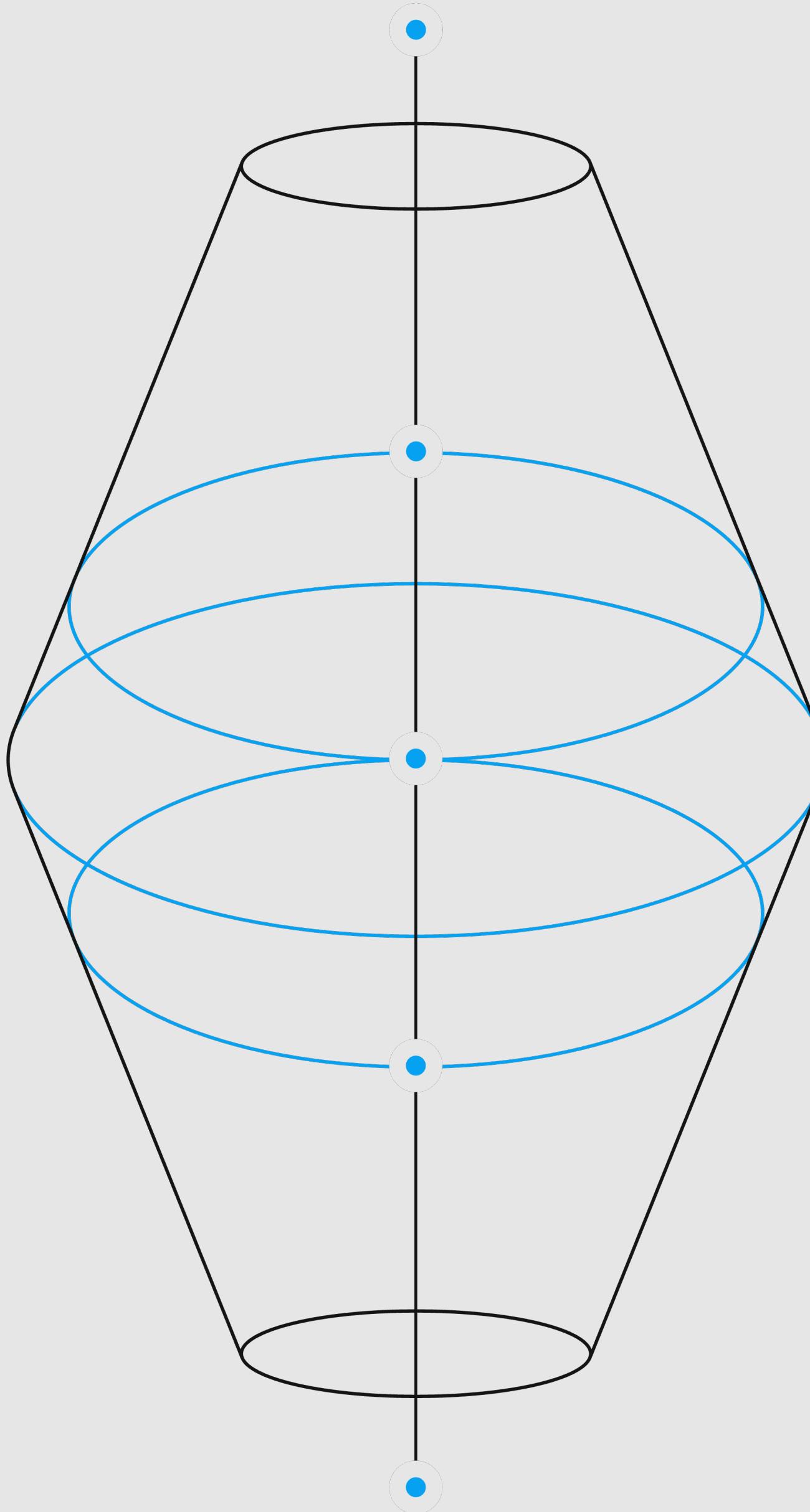


Что нужно запомнить

- Просто обучить модель не всегда достаточно для понимания причинности.
- Для оценки причинности нужно понимать, какие переменные влияют на другие и как выглядит ваш процесс.
- Мы разобрали лишь часть методов оценки причинности — это очень большая и активно развивающаяся область.
- В основном лучше проводить контролируемые эксперименты, в бизнесе это делать проще, чем в медицине или экономике. В этом вам помогут А/В-тесты и А/А-тесты.
- Но если контролируемый эксперимент невозможен, то можно использовать методы оценки причинности, которые мы обсудили.



Материалы



- [KDD 2018 Tutorial: Causal Inference in Machine Learning](#)
- [Quantitative Methods III: Explanation and Causation](#)
- [Applied Causal Inference](#)
- [Inferring causal impact using Bayesian structural time-series models](#)
- [Tigramite](#)
- [Не хотите ли чашечку LATE? За пределами контролируемых экспериментов: инструментальные переменные](#)