

Задача об эффективности инновации на основе данных ВкусВилл

Студент: Рогоза Ярослав Евгеньевич

Группа: Э408

Научный руководитель: Артамонов Дмитрий Вячеславович

Московский государственный университет имени М.В.Ломоносова
Экономический факультет

26 мая 2025 г.



ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

МГУ имени М. В. Ломоносова

В офлайн-ритейле традиционные методы А/В-тестирования сталкиваются с ограничениями из-за высокой дисперсии метрик. В статьях X5 Retail Group отмечается, что ошибка первого рода может достигать 20%, что значительно выше традиционных 5% [X5Tech, 2023].

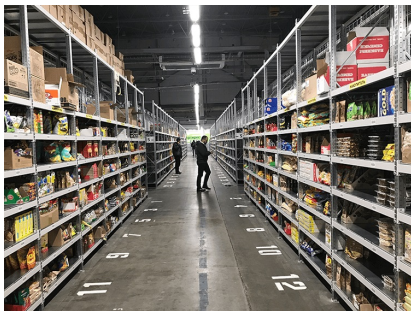
В ответ на это компании применяют методы синтетического контроля для точной оценки воздействия инноваций при ограниченных данных, как например в компании Лента [Васькин, 2022].

Кроме того, синтетический контроль активно обсуждается в научном сообществе, в том числе на воркшопх Princeton DataX [DataX, 2023].

- **Цель работы.** Цель настоящего исследования заключается в разработке и проверке методологических подходов для оценки эффективности внедрения инноваций в условиях малого объёма данных и ограниченности информации.
- **Объект исследования.** Процесс внедрения инновационных решений в деятельность предприятий.
- **Предмет исследования.** Методологические подходы и инструменты применения метода синтетического контроля для оценки *нефинансовой эффективности* инноваций при ограниченности данных.

Описание проблемы тестирования в офлайн ритейле.
Дарксторы (склады) - точки, недоступные покупателям для покупок в офлайне. На точке или части таких точек проводится интервенция. Необходимо оценить эффект от подобных нововведений и учесть факторы колебаний метрик.

Рис.: Пример даркстора



Список ограничений:

- Нельзя выбирать точку, на которой будет проводиться эксперимент.
- Данные по точкам представлены за короткий временной ряд до эксперимента.
- Эксперимент возможно проводить только в короткий промежуток времени 14 - 30 дней.
- В метриках, которые замеряются в процессе эксперимента, присутствуют выбросы и аномалии.

Выбранный подход

В данном исследовании для решения проблемы ограниченности данных будут использованы байесовский и фриквентистский подходы к методу **синтетического контроля**, позволяющие учитывать изменения в объектах с схожей целевой метрикой (донорские точки) и корректно моделировать резкие скачки спроса и влияние иных факторов.

Альтернативные методы (не выбраны)

Метод	Причина отказа
Propensity Score Matching	Недостаток обучающих примеров
Прогнозирование только на основе исторических данных	Игнорирование внешних шоков

- ➊ Размер эффекта и дисперсия
- ➋ Соответствующая донорская группа
- ➌ Отсутствие преждевременного воздействия
- ➍ Отсутствие взаимодействия между единицами (SUTVA)
- ➎ Принцип выпуклой оболочки и аппроксимируемость

Оценка эффекта вмешательства с помощью синтетического контроля

Определение эффекта вмешательства

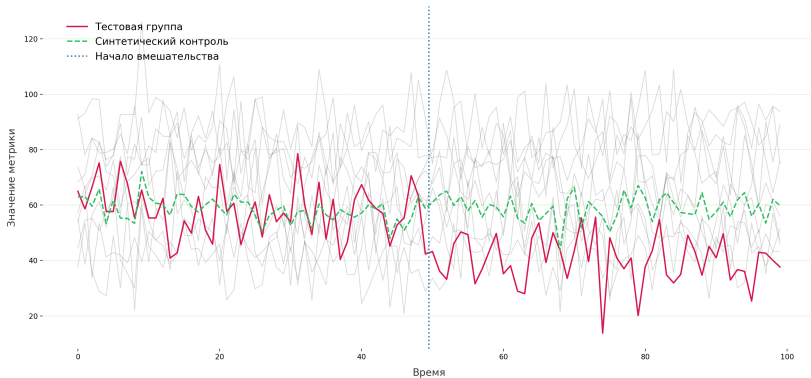
Истинный эффект вмешательства для тритмент-единицы определяется как разность между наблюдаемым исходом при наличии вмешательства и контрфактическим исходом в отсутствии вмешательства:

$$\tau_{1t} = Y_{1t}^I - Y_{1t}^N, \quad t > T_0$$

Поскольку для $t > T_0$ мы не можем наблюдать Y_{1t}^N (контрфактический исход), необходимо построить его оценку [Facure, 2022]

Пример использования синтетического контроля

Пример работы синтетического контроля



Фриквентистские модели синтетического контроля

Синтетический контроль [Abadie and Gardeazabal, 2003]:

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt}, \quad t = 1, \dots, T, \quad w_j \geq 0, \quad \sum_{j=2}^{J+1} w_j = 1$$

Синтетический difference-in-difference

[Arkhangelsky et al., 2019]:

$$\hat{\tau}^{sdid} = \operatorname{argmin}_{\mu, \alpha, \beta, \tau} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - (\mu + \alpha_i + \beta_t + \tau D_{it}))^2 \hat{w}_i^{sdid} \hat{\lambda}_t^{sdid} \right\}.$$

Байесовские модели синтетического контроля

Байесовский синтетический контроль

[Martinez and Vives-i Bastida, 2022]:

$$\hat{Y}_{1t}^N \sim \mathcal{N}\left(\sum_{j=2}^{J+1} w_j \cdot Y_{jt}, \sigma^2\right)$$

$$w_j \sim \text{Dirichlet}(1), \sigma \sim \text{HalfNormal}(1)$$

Байесовский синтетический контроль с применением гауссовского процесса [Arbour et al., 2021]:

$$\hat{Y}_{1t}^N \sim \mathcal{N}\left(\sum_{j=2}^{J+1} w_j Y_{jt} + f(t), \sigma^2\right)$$

$$f(t) \sim \mathcal{GP}(m(t), k(t, t')), w_j \sim \text{Dirichlet}(1), \sigma \sim \text{HalfNormal}(1)$$

Оценка статистической значимости эффекта полученного моделями

Фриквентистский подход:

Оценка доверительного интервала с помощью *плацебо тестов*:

$$\tau \in \hat{\tau} \pm z_{\alpha/2} \sqrt{\frac{1}{B} \sum_{b=1}^B (\hat{\tau}^{(b)} - \bar{\hat{\tau}})^2},$$

Байесовский подход:

Байесовский подход оценивает значимость эффекта через *апостериорные распределения*, используя плотность вероятности, что позволяет гибко учитывать неопределенность.

Описание данных:

- Данные за 180 дней для 90 дарксторов из одного региона.
- Метрики: среднее время сборки заказа, среднее время доставки, средняя дистанция на заказ, количество заказов на курьера, процент опозданий.

Предобработка данных:

- Проверяется наличие данных за весь необходимый период.
- Аномальные скачки в метриках устраняются с использованием скользящего среднего.
- Для снижения дисперсии применяется метод CUPED.

Результаты методов на примере метрики «среднее время доставки»

На примере метрики «среднее время доставки» можно проанализировать оценку ошибок I и II рода для четырех методов синтетического контроля. Аналогичные расчеты были произведены для остальных метрик.

Метод	Ошибка I рода	Ошибка II рода			
		5	10	15	20
Замешиваемый эффект					
BayesianSyntheticControl	0.700	24.4	12.2	7.8	3.3
BayesianSyntheticControlGP	0.500	25.0	12.0	6.0	3.0
SyntheticControl	0.067	92.2	62.2	16.7	3.3
Synthetic difference-in-differences	0.044	88.9	53.3	17.8	3.3

Результаты методов на примере метрики «среднее время доставки»

В рамках анализа метрики «среднее время доставки» выделяются значения, которые наиболее близки к стандартам тестирования: ошибка I рода, составляющая 5%, и ошибка II рода, достигающая 20%.

Метод	Ошибка I рода	Ошибка II рода			
Замешиваемый эффект		5	10	15	20
BayesianSyntheticControl	0.700	24.4	12.2	7.8	3.3
BayesianSyntheticControlGP	0.500	25.0	12.0	6.0	3.0
SyntheticControl	0.067	92.2	62.2	16.7	3.3
Synthetic difference-in-differences	0.044	88.9	53.3	17.8	3.3

Рекомендации к применению метода синтетического контроля в компании

Предпочтительны частотные методы:

- *Synthetic Difference-in-Differences*
- *Synthetic Control*

Рекомендуемые метрики для тестирования:

- среднее время сборки заказа
- среднее время доставки
- средняя дистанция на заказ
- количество заказов на курьера

Особенности байесовских моделей:

- требуют более информативных априорных оценок
- по возможности увеличивать объём исторических данных
- см. [Martinez and Vives-i Bastida, 2022] для примеров влияния априоров и размера выборки

Поскольку фриквентистские модели показали себя лучше всего, то они были реализованы в отдельной библиотеке для версий Python > 3.9, так как существующие на данный момент решения являются устаревшими



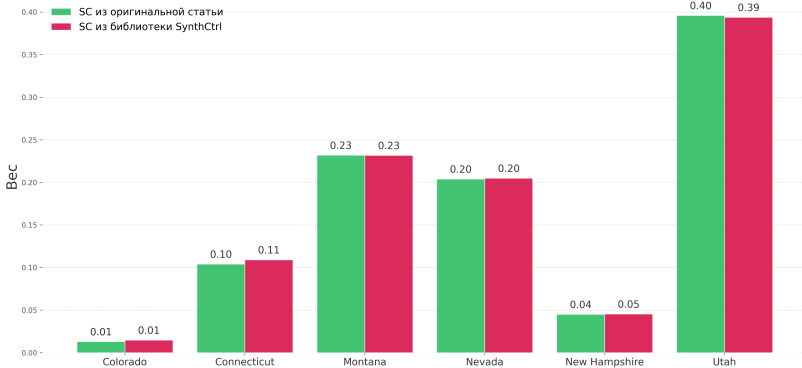
Качество реализаций фриквентистских методов реализованных в библиотеке были сравнены с оригинальными статьями [Arkhangelsky et al., 2019], [Abadie et al., 2010], посвященными проблеме Proposition 99.

Метод	Оценка эффекта	Стандартная ошибка
SDID (статья)	-15.6	8.4
SDID (SynthCtrl)	-15.6	9.02
SC (статья)	-19.6	9.9
SC (SynthCtrl)	-19.5	10.98

Таблица: Сравнение эффектов и стандартных ошибок методов SDID и SC

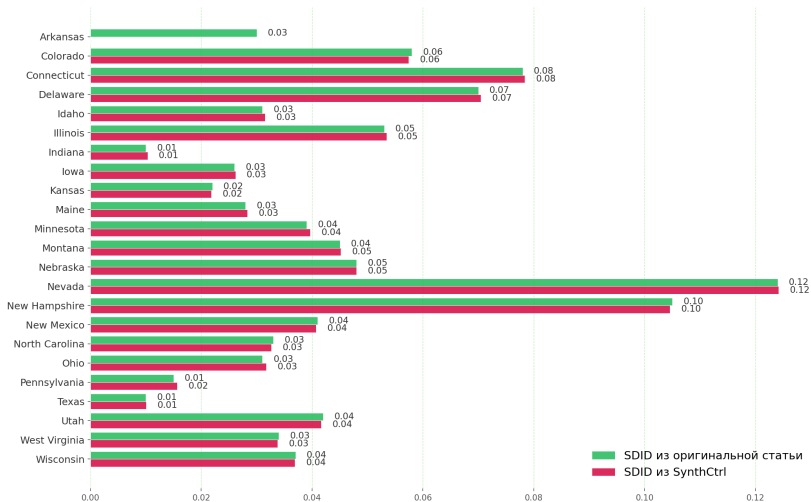
Разработанная библиотека. Сравнение весов синтетический контроль

Сравнение весов по штатам: SC из SynthCtrl и оригинальные данные



Разработанная библиотека. Сравнение весов синтетическая разница разниц

Сравнение весов по штатам: SDID из SynthCtrl и оригинальные данные



- В рамках эксперимента было зафиксировано, что модели синтетического контроля на основе фриквентистского подхода справились с нахождением эффекта лучше, чем байесовские, что может объясняться ограниченным количеством данных.
- На основе исследования были подготовлены рекомендации по использованию метода синтетического контроля для основных метрик доставки компании ВкусВилл.
- Для удобства использования в прикладных исследованиях и тестах была реализована библиотека SynthCtrl и были реплицированы результаты авторов фриквентистских методов синтетического контроля [Arkhangelsky et al., 2019], [Abadie et al., 2010].

Спасибо за внимание!



ЭКОНОМИЧЕСКИЙ ФАКУЛЬТЕТ

МГУ имени М. В. Ломоносова

Классический синтетический контроль

Идея синтетического контроля состоит в том, чтобы аппроксимировать Y_{1t}^N взвешенной суммой исходов донорских единиц (единиц, не подвергшихся вмешательству). Таким образом, синтетический контрфактический исход определяется как

$$\hat{Y}_{1t}^N = \sum_{j=2}^{J+1} w_j Y_{jt}, \quad t = 1, \dots, T, \quad w_j \geq 0, \quad \sum_{j=2}^{J+1} w_j = 1$$

Для выбора оптимального вектора весов w используются пред-интервенционные данные. Обозначим:

$$\mathbf{y}_{\text{pre}}^{\text{tr}} = (Y_{11}, Y_{12}, \dots, Y_{1T_0})^\top, \quad \mathbf{Y}_{\text{pre}}^{\text{co}} = \begin{pmatrix} Y_{21} & \dots & Y_{(J+1)1} \\ Y_{22} & \dots & Y_{(J+1)2} \\ \vdots & \ddots & \vdots \\ Y_{2T_0} & \dots & Y_{(J+1)T_0} \end{pmatrix}$$

Классический синтетический контроль. Продолжение

Задачу можно записать в следующем виде:

$$\min_{\mathbf{w} \in \Delta_J} \left\| \mathbf{y}_{\text{pre}}^{\text{tr}} - \mathbf{Y}_{\text{pre}}^{\text{co}} \mathbf{w} \right\|_2^2,$$

где Δ_J обозначает J -мерное симплексное множество, то есть

$$\Delta_J = \left\{ \mathbf{w} \in \mathbb{R}^J \mid w_j \geq 0, \sum_{j=1}^J w_j = 1 \right\}.$$

Решив данную задачу, получаем оптимальные веса $\hat{\mathbf{w}}$, с помощью которых можно оценить контрфактические исходы для тритмент-единицы [Abadie and Gardeazabal, 2003].

Синтетический difference-in-difference

Традиционный метод синтетического контроля подбирает веса контрольных единиц для точного воспроизведения трендов до воздействия, а метод DID оценивает эффект, предполагая параллельность трендов до вмешательства. Синтетический difference-in-difference (SDID) объединяет оба подхода, вводя веса для единиц w_i и временные веса λ_t , и оценивает эффект через минимизацию соответствующей функции потерь. Таким образом, оценка эффекта вмешательства производится посредством минимизации следующей функции потерь:

$$\hat{\tau}^{sdid} = \underset{\mu, \alpha, \beta, \tau}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \sum_{t=1}^T (Y_{it} - (\mu + \alpha_i + \beta_t + \tau D_{it}))^2 \hat{w}_i^{sdid} \hat{\lambda}_t^{sdid} \right\}.$$

Синтетический difference-in-difference. Продолжение

Для оценки временных весов решается следующая задача:

$$\hat{\lambda}^{sdid} = \operatorname{argmin}_{\lambda} \left\| \bar{y}_{post,co} - (\lambda_{pre} Y_{pre,co} + \lambda_0) \right\|_2^2,$$

при условии, что $\sum_t \lambda_t = 1$ и $\lambda_t > 0$ для всех t .

Оптимизационная задача для определения весов для точек решается следующим образом:

$$\hat{w}^{sdid} = \operatorname{argmin}_w \left\{ \left\| \bar{y}_{pre,tr} - (Y_{pre,co} w_{co} + w_0) \right\|_2^2 + \zeta^2 T_{pre} \left\| w_{co} \right\|_2^2 \right\},$$

при условии, что $\sum_i w_i = 1$ и $w_i > 0$ для всех i
[Arkhangelsky et al., 2019], [Facure, 2022].

Оценка статистической значимости результата

Повторяя синтетический контроль для контрольных единиц (доноров) получим плацебо оценку дисперсии эффекта.

$$\hat{V}_{\tau}^{placebo} = \frac{1}{B} \sum_{b=1}^B \left(\hat{\tau}^{(b)} - \bar{\hat{\tau}}^{(b)} \right)^2,$$

где $\hat{\tau}^{(b)}$ — оценка эффекта в b -ом плацебо-тесте, $\bar{\hat{\tau}}^{(b)}$ — среднее значение оценок по всем плацебо-тестам, B — общее число проведённых тестов.

$$\tau \in \hat{\tau}^{sdid} \pm z_{\alpha/2} \sqrt{\hat{V}_{\tau}},$$

где $\hat{\tau}^{sdid}$ — исходная оценка эффекта, $z_{\alpha/2}$ — критическое значение стандартного нормального распределения при уровне значимости α [Facure, 2022], [Abadie et al., 2015].

Байесовская модель синтетического контроля

Пусть, как и ранее, имеется $J + 1$ единиц, причём первая единица подвергается воздействию с момента $T_0 + 1$.

байесовский подход задаёт вероятностную модель:

$$\hat{Y}_{1t}^N \sim \mathcal{N}(\mu_i, \sigma^2)$$

$$\mu_i = \sum_{j=2}^{J+1} w_j \cdot Y_{jt}$$

$$w_j \sim \text{Dirichlet}(1)$$

$$\sigma \sim \text{HalfNormal}(1)$$

где $w = (w_2, \dots, w_{J+1})'$ — вектор весов, который распределен согласно Дирихле, благодаря чему удовлетворяет условиям $w_j \geq 0$ и $\sum_{j=2}^{J+1} w_j = 1$ [Martinez and Vives-i Bastida, 2022]

Байесовская модель синтетического контроля с применением гауссовских процессов

Формально модель на основе гауссовского процесса для оценки контрфактических исходов можно представить следующим образом:

$$\hat{Y}_{1t}^N \sim \mathcal{N}\left(\sum_{j=2}^{J+1} w_j Y_{jt} + f(t), \sigma^2\right)$$

$$f(t) \sim \mathcal{GP}(m(t), k(t, t')), w_j \sim \text{Dirichlet}(1), \sigma \sim \text{HalfNormal}(1)$$

где $m(t)$ – априорная средняя функция, которая, как правило, принимается равной нулю, а $k(t, t')$ – положительно определенная ковариационная функция (ядро), задающая корреляцию значений $f(t)$ в различные моменты времени. Гауссовский процесс в качестве компонента модели вводит гибкую нелинейную временную структуру, что позволяет учитывать сложные тенденции во времени [Arbour et al., 2021].

Таблица: Результаты оценки ошибок первого и второго рода для различных методов синтетического контроля и переменных

Метод	Переменная	Ошибка I рода	Ошибка II рода (по доле замешивания)			
			5%	10%	15%	20%
BayesianSyntheticControl	среднее время сборки заказа	0.544	0.322	0.200	0.078	0.056
	среднее время доставки	0.700	0.244	0.122	0.078	0.033
	средняя дистанция на заказ	0.467	0.144	0.100	0.100	0.089
	количество заказов на курьера	0.456	0.444	0.200	0.067	0.011
	процент опозданий	0.456	0.478	0.489	0.422	0.378
BayesianSyntheticControlGP	среднее время сборки заказа	0.400	0.300	0.150	0.050	0.020
	среднее время доставки	0.500	0.250	0.120	0.060	0.030
	средняя дистанция на заказ	0.350	0.200	0.100	0.050	0.025
	количество заказов на курьера	0.420	0.350	0.180	0.080	0.040
	процент опозданий	0.410	0.360	0.300	0.250	0.200
SyntheticControl	среднее время сборки заказа	0.067	0.856	0.556	0.156	0.022
	среднее время доставки	0.067	0.922	0.622	0.167	0.033
	средняя дистанция на заказ	0.100	0.922	0.456	0.033	0.000
	количество заказов на курьера	0.033	0.900	0.611	0.244	0.056
	процент опозданий	0.100	0.911	0.933	0.944	0.922
Synthetic difference-in-differences	среднее время сборки заказа	0.022	0.811	0.522	0.156	0.011
	среднее время доставки	0.044	0.889	0.533	0.178	0.033
	средняя дистанция на заказ	0.100	0.711	0.033	0.000	0.000
	количество заказов на курьера	0.033	0.944	0.589	0.278	0.056
	процент опозданий	0.089	0.922	0.922	0.911	0.900



Abadie, A., Diamond, A., and Hainmueller, J. (2010).

Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program.

Journal of the American statistical Association, 105(490):493–505.



Abadie, A., Diamond, A., and Hainmueller, J. (2015).

Comparative politics and the synthetic control method.

American Journal of Political Science, 59(2):495–510.



Abadie, A. and Gardeazabal, J. (2003).

The economic costs of conflict: A case study of the basque country.

American economic review, 93(1):113–132.



Arbour, D., Ben-Michael, E., Feller, A., Franks, A., and Raphael, S. (2021).

Using multitask gaussian processes to estimate the effect of a targeted effort to remove firearms.

arXiv preprint arXiv:2110.07006.



Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. (2019).

Synthetic difference in differences.

Technical report, National Bureau of Economic Research.



DataX, P. (2023).

Synthetic control methods workshop.

Дата обращения: 21 апреля 2025.



Facure, M. (2022).

Causal inference for the brave and true.

<https://matheusfacure.github.io/python-causality-handbook/landing-page.html> [Дата обращения: 2025-01-08].



Martinez, I. and Vives-i Bastida, J. (2022).

Bayesian and frequentist inference for synthetic controls.

arXiv preprint arXiv:2206.01779.



X5Tech (2023).

От а/б-тестирования к causal inference в офлайн ритейле.

Дата обращения: 21 апреля 2025.



Васькин, (2022).

Synthetic control для а/б тестов на малых размерах выборок.

ODS Reliable ML AB Testing & Causal Inference Meetup, 17 декабря 2022 г. Дата обращения: 21 апреля 2025.