

Алла Михайлівна Ткаченко,

д-р екон. наук, професор,

ORCID 0000-0002-1843-2579

e-mail: alla0676128584@gmail.com

Національний університет «Запорізька Політехніка», м. Запоріжжя

Роман Степанович Чорний,

д-р екон. наук, професор, директор,

ORCID 0000-0001-8614-9495

Нововолинський-навчально-науковий інститут економіки та менеджменту, м. Нововолинськ

ДОРОЖНЯ КАРТА ДЛЯ ТОП-МЕНЕДЖЕРА ВІД ОПИСОВОЇ СТАТИСТИКИ ДО КАУЗАЛЬНОГО МЕНЕДЖМЕНТУ

Метою статті є розробка та емпіричне обґрунтування цілісного підходу каузального менеджменту для топ-рівня управління, що поєднує: формулювання контрфактичних питань і DAG-мислення; дерево вибору ідентифікаційних стратегій (RCT, DiD із багатоперіодним таймінгом, синтетичний контроль, регресійна розривність, інструментальні змінні, doubly-robust/targeted learning, causal ML для CATE); стандартизований паспорт ефекту (припущення → діагностики → результати ATE/CATE); портфельне ухвалення рішень через очікувану прирідну цінність (EPV) і вартість додаткової інформації (VOI); інституційний дизайн (Експериментальний офіс, комітет з експериментів, каталог втручань, матриця зрілості) та *norms* прозорості/відтворюваності [6, 8, 16, 17].

Наукова новизна полягає у зшиванні методів каузальної ідентифікації з процесною архітектурою рішень, що дозволяє транслювати статистичні оцінки у чіткі політики Go-Scale/Go-Pilot/No-Go, інтегровані в OKR і бюджетний цикл. Додатковий внесок — практичні віньетки (ціноутворення, CAPEX, HR, постачання, маркетинг/ESG), які демонструють, як одні й ті самі принципи працюють у різних функціях та на різних часових горизонтах.

Стаття спирається на сучасну парадигму потенційних наслідків і каузальних діаграм (DAG) як основу для явного формулювання припущень про механізми впливу. Введено відмінності між ATE/ATT/ATC та CATE (умовні ефекти для сегментів), що принципово важливо для таргетованого масштабування — у практичному менеджменті саме гетерогенність часто визначає, кому і де варто розгорнути політику.

Методологія та дизайн дослідження. Пропонується дерево вибору ідентифікації, яке починається з питання про можливість рандомізації. Якщо рандомізація можлива й етично прийнятна — оптимальні польові RCT/A/B-експерименти з письмовими pre-analysis планами, правилами зупинки, баланс-перевірками та контролем spillover. Якщо політика впроваджується поетапно між одиницями спостереження — застосовуються сучасні варіанти DiD з подієвими дослідженнями та робастними вагами, що усувають зміщення класичних TWFE-оцінювачів. Для одиничних втручань (наприклад, встановлення роботи на одному заводі) доцільний синтетичний контроль (SCM) з прозорим добром донорського пулу й placebo-тестами в часі та просторі. Якщо рішення визначаються порогами (бали рейтингу, обсяги, бонуси) — використовується RDD із тестами щільності та робастністю

до вибору bandwidth. За наявності релевантного екзогенного поштовху, який впливає на втручання, але не впливає безпосередньо на результат, застосовують IV із перевіркою partial RI/F та коректною LATE-інтерпретацією. Коли «шоку» немає, але дані багаті — поєднання doubly-robust/TMLE і causal ML забезпечує стійке оцінювання CATE для таргетування.

Усім дизайнам передують DAG-схеми, що фіксують припущення, і супроводжують діагностики: pre-trend/placebo для DiD/SCM; тести безперервності коваріат і щільності для RDD; релевантність/екзогенність для IV; honest-splitting і крос-валідація для CATE. На рівні даних — єдина модель (UnitЧPeriod), чітка семантика KPI, журнал перетворень (*audit-trail*), політики аномалій, правила доступу й анонімізація. Результати подаються у вигляді «паспорта ефекту»: каузальне питання → DAG → дизайн → припущення → план аналізу → діагностики → результати (ATE/CATE із 95% ДІ) → рішення → межі узагальнення → посилання на дані/код (де це можливо).

Ключовою інновацією роботи є переклад каузальних оцінок у рішення в гривні/доларах через EPV (Expected Profit Value) і VOI (Value of Information). У стандартній формі $EPV = p_{win} \cdot \Delta\pi - C_{impl}$, де p_{win} — ймовірність позитивного ефекту з урахуванням статистичної невизначеності, $\Delta\pi$ — причинний приріст прибутку/маржі, C_{impl} — витрати впровадження. VOI оцінює доцільність ще одного раунду даних/експериментів перед масштабуванням: масштабувати, якщо $EPV > 0$ та очікувана цінність додаткової інформації (EVSI) менша за вартість затримки; досліджувати далі, якщо EVSI перевищує цю вартість. Таким чином, статистика зшивається з *governance*-циклом: рішення пере-стають бути функцією лише «p-value», натомість зважують ризик, ресурс, час і альтернативну вартість.

Щоб каузальна логіка не залишалася «проектом ентузіастів», обґрунтовано організаційну конфігурацію: експериментальний офіс (ЕХО) як центр методичної підтримки та незалежної валідації; комітет з експериментів (ЕХО + юридичний/етичний офіс + С-рівень) як орган *governance*; каталог втручань із «паспортами ефектів» як корпоративну пам'ять; матриця зрілості (від «дашбордів» до «каузальної культури»). Передбачено *TOP-сумісні* політики прозорості: пре-реєстрація гіпотез, оприлюднення припущень/діагностик, контроль приватності, версійність коду/даних, незалежні реплікації. Усі ключові зміни політик у компанії проходять через правило «жодного масштабування без паспорта ефекту».



У роботі подано п'ять ілюстративних кейсів, що відтворюють запропоновані принципи в різних функціях:

1. Ціноутворення (B2C): короткий А/В-експеримент на рівні «магазин-тиждень» з подальшим CATE-таргетуванням демонструє, що середній позитивний ефект маскує сильні виграші в підмножині точок — отже, масштабування має бути селективним.

2. SAPEX у виробництві: одичне впровадження робота-зварника оцінено через SCM; приріст продуктивності та зменшення вартості браку підтверджено діагностиками; масштабування планується через «staggered adoption» і DiD-моніторинг.

3. HR (зміни): DiD із багатоперіодним таймінгом для мікро-навчання свідчить про скорочення абсентеїзму та приріст випуску; CATE покаже сегменти, де ефект найбільший (новачки/нічні зміни).

4. Ланцюги постачання: IV-оцінка впливу підвищення *safety-stock* на відмови поставок і маржу (інструмент — зовнішні прості). Результат: зниження відмов, але нейтральний ефект на маржу; рішення — таргетоване підвищення запасів у «ризикових» ланцюгах і VOI-оцінка альтернатив.

5. Маркетинг/ESG: польовий RCT із комунікаційними наративами демонструє короткостроковий підйом конверсії на тлі обмеженої стійкості ефекту; раціональна стратегія — «пульсова» комунікація з періодичними реплікаціями, якщо EPV залишається додатним.

Завдяки єдиній процедурі «гіпотеза → дизайн → діагностика → ATE/CATE → EPV/VOI → рішення» всі кейси доведено до управлінського вердикту Go-Scale, Go-Pilot або No-Go/Redesign, синхронізованого з OKR і бюджетним циклом.

Результати і ключові висновки

1. Методологічний результат: створено узагальнене дерево вибору для ідентифікації причинних ефектів у менеджменті, що є незалежним від конкретної галузі та масштабу фірми, але чутливим до етичних/операційних обмежень.

2. Процесний результат: запропоновано паспорт ефекту як стандарт мінімального набору припущень, діагностик і репорту; цей формат підвищує відтворюваність і спрощує незалежний аудит.

3. Інвестиційний результат: розроблено EPV/VOI-рамку для портфельного ранжування втручань; це дозволяє обґрунтовувати масштабування з урахуванням вартості затримки й альтернативної вартості капіталу/ресурсу.

4. Організаційний результат: показано, як ExO + Комітет з експериментів + каталог втручань + матриця зрілості формують каузальну культуру, що переживає зміни керівництва, ринкові цикли й плінність кадрів.

5. Прикладний результат: на віньєтках продемонстровано селективне масштабування через CATE-таргетування та унеможливлення KPI-маніпуляцій через прозорість припущень, пре-реєстрацію і незалежні валідації.

Управлінське значення

Для С-рівня запропонована рамка дає три відчутні переваги. По-перше, вона зменшує вартість помилок: замість покладатися на «середні» кореляції, менеджери отримують ідентифіковані причинні ефекти з діагностиками. По-друге, вона прискорює цикл навчання: «слабкі» ініціативи припиняються раніше, «сильні» — масштабуються таргетовано. По-третє, вона узгоджує аналітику з бюджетом: рішення формулюються в категоріях EPV/VOI, що робить супе-

речки між підрозділами менш ідеологічними та більш даними-керованими.

В роботі наголошено, що етика та прозорість — частина керування ризику. Тому пропонуються політики пре-реєстрації (pre-analysis plan), публічного формулювання припущень (DAG), мінімальні набори діагностик (pre-trend/placebo, RDD-щільність, IV-релевантність, стабільність CATE), контроль доступу до персональних даних і анонімізація, обмеження на експериментальний вплив, а також відкриті репозиторії артефактів (код/метадані/шаблони) там, де це дозволено організаційними політиками. Усе це знижує ризик р-hacking, «полювання за значущістю» та навмисного/ненавмисного перекручування KPI.

Як і будь-яка каузальна стратегія у реальних середовищах, запропонована рамка має обмеження. По-перше, за спостережних даних зберігається ризик невимірних confounders; його зменшує ретельний дизайн, чутливісні аналізи та реплікації. По-друге, динамічна гетерогенність у часі може ускладнювати інтерпретацію — необхідні події дослідження та стабільні вагові схеми. По-третє, локальність інструментів в IV обмежує узагальненість; це вимагає обережних трактувань і додаткових перевірок. По-четверте, етичні й регуляторні рамки не завжди дозволяють експериментувати; тоді на перший план виходять квазіексперименти з прозорими припущеннями. Нарешті, ресурсні обмеження (кадри, час, бюджет) інколи роблять VOI-підхід ключовим вирішувачем у дилемі «досліджувати чи масштабувати».

Напрями подальших досліджень

Перспективними є: інтеграція causal ML у операційні конвеєри з безперервним моніторингом стабільності CATE; розробка гібридів SCM4DiD для SAPEX-рішень із довгими горизонтами; формалізація портфельних правил EPV/VOI з урахуванням взаємозалежностей між інтервенціями; гармонізація стандартів прозорості між галузями та журналами; емпіричні дослідження освітніх інтервенцій для керівників з оцінкою їхнього каузального впливу на якість стратегічних рішень; верифікація «паспортів ефектів» у мультинаціональних групах і порівняння траєкторій руху за матрицею зрілості.

У статті пропонується системний шлях від опису — до керуваної причинності. Через поєднання DAG-мислення, сучасних ідентифікаційних стратегій, стандартизованого «паспорта ефекту», портфельної логіки EPV/VOI та інституційного дизайну (ExO, комітети, каталоги, зрілість), фірма здатна перетворити аналітику на відтворювану функцію ухвалення рішень. Практичний ефект — зменшення вартості помилок, прискорення циклу «гіпотеза → доказ → масштабування», таргетоване впровадження політик у сегментах з найбільшим причинним впливом та стійка конкурентна перевага в умовах невизначеності. Запропонована рамка не лише зводить міст між статистикою й менеджментом, а й формує культуру причинності, де кожна суттєва політика має свій «паспорт ефекту», а масштабування ґрунтується на перевірених припущеннях, діагностиках і прозорих правилах прийняття рішень.

Вступ. За останнє десятиліття бізнес масово інвестує в дашборди, описову аналітику й прогностичні моделі, але більшість рішень топ-менеджерів і далі спираються на кореляційні зв'язки, які не відповідають на ключове запитання управління: що станеться з показниками, якщо ми втрутимося? Цей розрив між «бачити» та «впливати» породжує три системні ризики: 1) помилкове приписування результатів марке-

тинговим чи операційним ініціативам; 2) масштабування локальних «успіхів», зумовлених сезоном або відбором; 3) хронічне недоінвестування у втручання з високою очікуваною цінністю, але без «яскравої» кореляції в історичних даних. Сучасні методологічні огляди прямо попереджають: традиційні регресії з фіксованими ефектами, широко вживані у практиці, можуть давати оманливі оцінки причинного ефекту, особливо за неоднорідного таймінгу політик і гетерогенних ефектів у часі [1, р. 1; 2, рр. 1–2].

Менеджерам потрібна дорожня карта переходу від описових панелей до системи прийняття рішень, де кожна суттєва політика (ціни, промо, CAPEX, навчання персоналу, зміни технології) має чітко ідентифікований причинний ефект та «паспорт» відтворюваності. Передумови такої системи добре описані у сучасній економетриці: формулювання контрфактичного питання (який результуючий показник оцінюємо за втручанням проти без нього); побудова каузальної діаграми (DAG) для виявлення змішувальних факторів; ідентифікаційна стратегія, сумісна з даними й етикою: випадкові експерименти, різниця-в-різницях (DiD) зі «стагерованим» впровадженням, регресійна розривність, інструментальні змінні чи синтетичний контроль; процедури надійності (pre-trend, placebo, чутливість) [1, 5].

Емпірична база на користь експериментального стилю менеджменту швидко зростає. У масштабній реплікації чотирьох рандомізованих досліджень серед 759 фірм доведено, що «науковий» підхід до ухвалення рішень (чіткі гіпотези, протоколи тестування, відсікання невалідних ідей) змінює поведінку підприємців і покращує якість стратегічних виборів — зокрема, зростає частка своєчасного «закриття» нежиттєздатних проєктів і зменшується хаотичне «перепристроювання» стратегії [3]. У сфері комунікацій з інвесторами рандомізований польовий експеримент показав, що інформаційні інтервенції (ESG-звіти та соціальні сигнали) здатні — хай і помірно та короткостроково — змінювати реальну купівельну поведінку, що принципово важливо для маркетингових інвестицій із вимогами до ROI [4]. Отримані результати підсилюють управлінську тезу, що без експериментів або їхніх квазіекспериментальних аналогів неможливо надійно розрізнити «спрацювало» й «збіглося».

Водночас, коли рандомізація недосяжна, менеджмент має спиратися на сучасні квазіекспериментальні дизайни. Для політик із поетапним впровадженням (каналні ціни, навчання на різних майданчиках, послідовні хвилі CAPEX) класичні TWFE-оцінки чутливі до зміщень і неоднорідностей, тоді як нові підходи до DiD та синтетичного контролю з «частково агрегованими» вагами дають більш стійкі результати та прозорі діагностики [1, 5]. Практичний висновок для топ-менеджера: спочатку дизайн — потім регресія; пріоритет — наочним припущенням і перевіркाम ідентифікації, а не «зворотній інженерії» з готового рівняння [1].

Таким чином, управлінська проблема, яку розв'язує ця стаття, полягає у переході від кореляційно-описових інсайтів до каузального менеджменту — інституційно закріпленого процесу формулювання, тестування та масштабування рішень на основі ідентифікованих причинних ефектів. Конкретні запитання, що виводяться з цієї проблеми:

1. Які бізнес-ефекти (дохід, маржа, продуктивність, ризики, задоволеність клієнтів) є каузально релевантними для пріоритизації?

2. Які ідентифікаційні стратегії є найбільш доречними під обмеження ресурсів і етики (RCT, DiD з багатоперіодним таймінгом, синтетичний контроль при одиничних втручаннях тощо) [5]?

3. Як інтегрувати оцінені ефекти у портфель рішень через очікувану прирістну цінність (EPV) та вартість додаткової інформації (VOI) — тобто, коли масштабувати, а коли продовжувати збір доказів [1]?

4. Які мінімальні стандарти даних, протоколів та аудиту надійності потрібні, щоб результати були відтворювані у різних підрозділах і на різних ринках [2]?

Відповідаючи на ці запитання, стаття пропонує структуровану «дорожню карту» для С-рівня: від постановки контрфактичних питань та побудови DAG — до вибору дизайну, перевірок стійкості, оцінювання гетерогенності ефектів і впровадження рішень із прозорою комунікацією невизначеності. У центрі — менеджерська керованість ризику: ми мінімізуємо імовірність масштабування «хибних перемог» та робимо інвестиції у зміни керованими за допомогою статистично ідентифікованих ефектів [3, 4].

Виходячи із мети статті, були поставлені такі завдання:

1. Синтез сучасних підходів до ідентифікації (RCT/DiD/SCM/RDD/IV/DR) і формулювання дерева вибору дизайну з прив'язкою до типових управлінських рішень [1, 5, 10].

2. Опис і впровадження стандарту «паспорт ефекту» (DAG → припущення → діагностики → ATE/CATE) як базового документа управління ризиком і відтворюваністю [6, 8].

3. Моделювання EPV/VOI для трансляції причинних оцінок у портфельні рішення (Go-Scale/Go-Pilot/No-Go) та розробка порогових правил з урахуванням «вартості затримки» [16].

4. Організаційний дизайн: опис ExO, ролей/процедур, каталогу «паспортів ефектів» і матриці зрілості з метриками прогресу [8, 17].

5. Емпірична валідація на віньєтках (ціноутворення, CAPEX, HR, ланцюги постачання, маркетинг/ESG) із перевітками pre-trends/placebo, RDD-щільності, IV-релевантності та CATE-стабільності [3–5, 12].

Аналіз літератури. За останні роки у менеджменті закріпився консенсус: прогностичні моделі без ідентифікації не відповідають на контрфактичне запитання «що буде, якщо ми втрутимось?» — вони описують тренд «за відсутності дії» [7, рр. 123–126; 8, рр. 1846–1849]. Тому сучасні рекомендації вимагають явних припущень (через DAG) і відповідного дизайну (RCT/DiD/SCM/RDD/IV), перш ніж інтерпретувати оцінки як причинні [6, 8].

Сучасні дизайни: як обирати й на що дивитися: DiD переживає методологічне оновлення: критика TWFE за гетерогенності ефектів і «стагерованого» таймінгу та перехід до wag і подієвих досліджень (event-study) із прозорими діагностиками [1, 2].

SCM — стандарт для одиничних втручань; ключ — якість «донорського пулу» і placebo-тести в часі й просторі [5, 9].

IV із гетерогенністю потребує не лише F-статистики/partial RI, а й чіткої LATE-інтерпретації та обґрунтування виключення [10, 15].

DR/TMLE і causal ML дозволяють стабільно оцінювати CATE для таргетованих політик, за умови honest-splitting і крос-валідації [8, 12].

Великомасштабні рандомізовані дослідження на рівні фірм показують: «науковий» стиль прийняття рішень (чіткі гіпотези, протоколи, відсікання невалідних ідей) покращує якість виборів і прискорює закрит-

тя слабких проєктів [3]. У маркетингу польові RCT демонструють короткострокові, але реальні поведінкові зрушення від інформаційних інтервенцій — важливий сигнал для ROI-мислення у комунікаціях [4].

Редакційні стандарти вимагають пре-реєстрації, явних припущень і публікації діагностик; це перетворює аналітику з «чорної скриньки» на керовану практику (зменшує р-hacking, захищає від «хибних перемог») [8, 17]. Для менеджменту це означає «паспорт ефекту»: DAG → дизайн → припущення → діагностика → результати (ATE/CATE) → рішення (EPV/VOI).

Рамка Expected Profit Value (EPV) і Value of Information (VOI) зшиває статистичну невизначеність із бюджетними рішеннями: масштабувати, якщо EPV > 0 і EVSI нижча за «вартість затримки»; інакше — спершу добрати дані [16]. Це дисциплінує CAPEX/операційні ініціативи й робить портфель втручань керованим за ризиком.

Попри методичний прогрес, менеджерська література рідко дає цілісну операційну рамку, що одночасно: поєднує вибір дизайну з DAG-мисленням і CATE-таргетуванням; вбудовує EPV/VOI у цикл OKR/бюджетів; інституціоналізує процес через ExO, каталоги «паспортів ефектів» і зрілість практики. Саме цю прогалину адресує наша робота, розвиваючи запропоновану в рукописі дорожню карту для топ-менеджера

Основна частина. Управлінські дашборди та прогнози відповідають на «що сталося» та «що ймовірно станеться», але не на головне запитання політики: «що буде, якщо ми втрутимось?». Причинна аналітика додає саме цю ланку — через явні припущення ідентифікації, які дозволяють інтерпретувати оцінений ефект як контрфактичний [6, 7]. Поза спеціальним дизайном (експерименти, квазіексперименти) або формалізованими припущеннями регресії залишаються описовими, навіть якщо вони дуже точні за RI [7]. Для менеджера це означає: прогноз — не причинність, і рішення про ціни, CAPEX чи навчання персоналу мають базуватись на ідентифікованих ефектах, а не лише на кореляціях.

У підході потенційних наслідків ефект втручання a_u — різниця між результатами зі втручанням і без нього: $a_u = E[Y(1) - Y(0)]$. Для управління важлива не лише ATE, а й ATT/ATC (ефекти для вже оброблених/необроблених) та CATE — умовні ефекти для цільових сегментів (клієнтів, цехів, команд), що напряму підживлює таргетовані політики [7]. У сучасних оглядах наголошується, що стратегічна цінність криється у виявленні гетерогенності ефектів та у виборі політик призначення з урахуванням цієї гетерогенності [7].

Наступним важливим кроком є робота з каузальними діаграмами (DAG), confounding і правила back-door/front-door. DAG роблять припущення дослідника явними: які змінні є спільними причинами (confounders), що є медіаторами та де можливі колайдери. Правило задніх дверей (back-door): для ідентифікації ефекту потрібно заблокувати всі «задні» шляхи між втручанням та результатом через кондиціонування на спільних причинах; front-door працює, коли відомий повний медіатор і відсутній неконтрольований confounding на шляхах «втручання → медіатор» і «медіатор → результат». Для менеджера DAG — це «креслення» політики, яке передуює вибору оцінювача і дисциплінує інтерпретацію [6]. Схематизація припущень для оцінки причинного ефекту зміни ціни на продажі з контролем confounders і медіаторів розглянемо через каузальну діаграму (DAG) для політики ціноутворення (рис. 1).

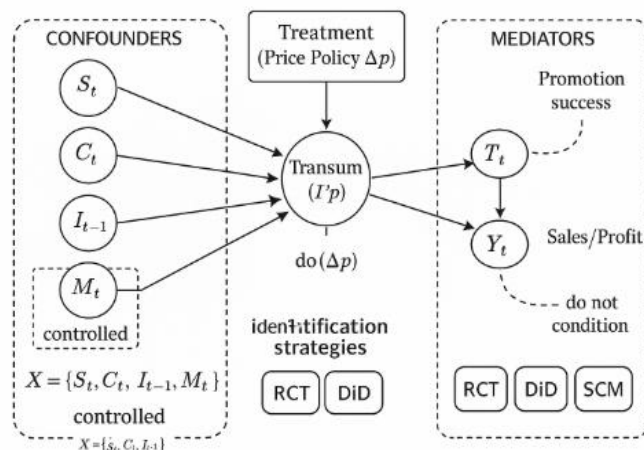


Рис. 1. Каузальна діаграма (DAG) для політики ціноутворення

Схематизація припущень, що була подана на рисунку 1, дозволяє сформулювати меню ідентифікаційних стратегій, а саме коли що застосовувати. Таким чином сформувати RCT (рандомізовані польові експерименти). «Золотий стандарт», якщо допустимі рандомізація й етичні умови (ціни, промо, UX-тести). Навіть невеликі експерименти можуть радикально підвищувати якість управлінських висновків [8].

DiD з багатоперіодним таймінгом. За поетапного впровадження політик класичні TWFE-оцінки чутливі до гетерогенності у часі; сучасні підходи (Callaway–Sant’Anna та похідні) забезпечують прозорі припущення, коректують ваги й надають надійні діагностики [7].

Синтетичний контроль (SCM). Коли втручання стосується одного/небагатьох об’єктів (завод, регіон), SCM дозволяє побудувати «штучного близнюка» з донорського пулу та наочно перевіряти передумови [9].

Регресійна розривність (RDD). Доречна за наявності чітких порогів у правилах прийняття рішень (рейтинги, бонуси, відсічення за обсягом) [7].

Інструментальні змінні (IV) з неоднорідними ефектами. Сучасні огляди підкреслюють важливість перевірки релевантності/екзогенності інструмента та коректної інтерпретації локальних ефектів (LATE) в умовах гетерогенності [10].

Doubly-robust та таргетоване навчання (TMLE). Поєднання моделі результату та моделі схильності до втручання забезпечує стійкі оцінки навіть за частково помилкової специфікації; ці підходи дедалі частіше рекомендовані редакційними настановами для обсерваційних досліджень [8].

Таким чином важливим є перехід від прогнозного ML до каузального ML. ML-моделі без ідентифікації добре передбачають, але не відповідають на «що, якщо». Каузальний ML (causal forests, uplift-моделі, doubly-robust learners) поєднує ідентифікаційні припущення з некерованою складністю моделей, щоб стабільно оцінювати CATE і таргетувати втручання [7]. Свіжі методологічні огляди 2025 року демонструють, як дослідники застосовують causal forests для практичних задач модерації ефектів у бізнесі та політиці [12]. Редакційні політики провідних журналів радять структурувати обсерваційні дослідження навколо шести запитань: каузальне питання, цільова величина, дизайн, припущення, аналітика, правдоподібність каузальної інтерпретації [8]. Для менеджменту це оформлюється у «паспорт ефекту», що містить DAG, передумови, протоколи (pre-

analysis), діагностики (pre-trends, placebo, чутливість), межі узагальнення та правила масштабування.

Методологія та дизайн дослідження проведемо, перш за все, через дерево вибору ідентифікації (рис. 2).

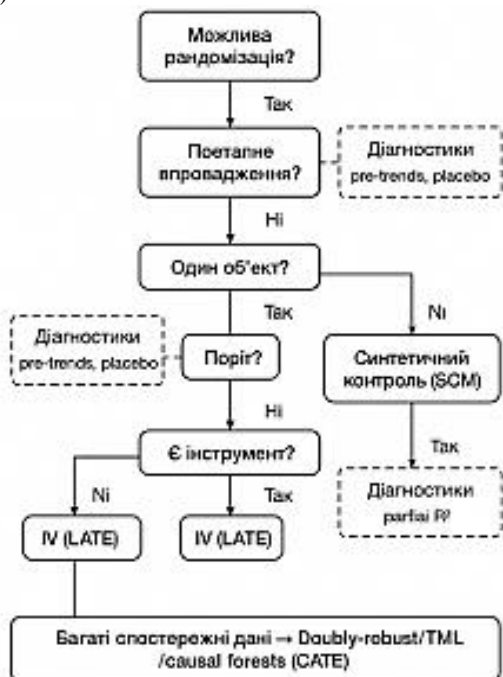


Рис. 2. Дерево вибору ідентифікації

Таким чином логіка вибору дизайну стартує з формулювання контрфактичного питання та побудови DAG, що було вищезазначено. Далі – принцип «спочатку дизайн, потім оцінювач»:

1. Якщо рандомізація прийнятна – RCT/польові А/В з попередньо зареєстрованим планом аналізу та правилами зупинки.

2. Якщо політика впроваджується поетапно – DiD для багатоперіодного таймінгу з перевірками паралельних трендів і вагами, що уникають зміщень TWFE [7, 13, 14].

3. Якщо втручання одиничне (один завод/регіон) – синтетичний контроль (SCM); для стагерованого прийняття політик – поєднання SCM/DiD з діагностикими передумов [5].

4. Якщо рішення керується порогами – RDD з тестом МакКрери та робастністю до вибору вікна [7].

5. Якщо існує інструмент, що впливає на дію, але не впливає напряму на результат – IV із перевіркою релевантності (partial RI, F) та коректною LATE-інтерпретацією під гетерогенністю [10, 15].

6. Якщо є багаті обсерваційні дані без «шоку» – doubly-robust / TMLE / causal forests для CATE та таргетованого призначення [8, 12].

Важливим у процесі прийняття рішень є робота з даними, вибіркою і якістю вимірювань, при цьому ми маємо орієнтуватись, перш за все на такі джерела: фінанси (P&L, валова/операційна маржа), операції (випуск, COPQ, простой), маркетинг (канал, промо, когорти), HR (абсентеїзм, плинність, зміни), формуючи єдину модель даних з чіткою часовою прив'язкою (Unit4Period), яка мінімізує похибки вирівнювання. Також не можна лишати поза увагою мета-дані експериментів/квазіекспериментів, а саме мета, DAG, правило призначення, первинні/вторинні KPI, вікна спостереження, ризику/етика, власники даних [6]. І з рештою ми маємо провести контроль якості: повнота, стабільність визначень KPI, політики аномалій та audit-trail перетворень. Таким чином *стандартизований шаблон документування інтервенції для відтворюваності має таку структуру (табл. 1):*

Таблиця 1. Паспорт управлінського експерименту (мінімальний набір метаданих)

Поле	Що фіксуємо	Приклад
Ціль втручання	Бізнес-мета, основний KPI	Підвищити маржу без падіння обороту
Каузальне питання	Формулювання в термінах контрфакту	Який вплив -5% ціни на маржу SKU?
DAG	Перелік confounders/медіаторів	Seasonality, Competitor price, Inventory
Дизайн	RCT / DiD / SCM / RDD / IV / DR	DiD (staggered adoption)
Правило призначення	Як відбираємо/рандомізуємо	Хвилі впровадження по регіонах
Вибірка/період	Одиниця аналізу Ч час	Магазин-тиждень, 18 міс.
Первинні KPI	1–2 ключові	Валова маржа, GMV
Вторинні KPI	Контрольні	Частка промо, NPS
Pre-analysis plan	Гіпотези, моделі, зупинки	Зареєстровано, α -корекція
Перевірки	Pre-trends, placebo, чутливість	Event-study, донорський пул
Приватність/етика	Обмеження, анонімізація	Role-based доступ
Ризики	Операційні/репутаційні	Короткий дефіцит складу
Власники/версії	Хто відповідальний	ExO, v1.2 (дата)

Протоколи валідації та діагностики даної таблиці:

- Pre-trend/Placebo для DiD/SCM; чутливість до донорського пулу та часових вікон [5,14].

- Hidden linearity bias у TWFE за коваріат-залежних трендів: діагностика через переписування TWFE як вагового оцінювача й застосування альтернативних DiD-оцінювачів [13].

- RDD: безперервність коваріат довкола порогу; робастність до bandwidth/polynomial [7].

- IV: partial RI (Shea), F-статистика, обґрунтованість виключення, локальність LATE [15].

- Doubly-robust/ML: honest splitting, крос-валідація для CATE, чутливість до пропущених коваріат [12].

- Редакційна “рамка 6 запитань” для каузальної інтерпретації в обсерваційних роботах [8].

В результаті роботи з таблицею 1, ми отримуємо оцінені причинні ефекти з 95% довірчими інтервалами: загальний ATE і сегментні CATE (рис. 3).

Етичні та організаційні аспекти в такому разі набувають досить специфічних форм. Для експериментів – інформована згода/де-ідентифікація, мінімізація ризиків, equipoise; після завершення – справедливий доступ до ефективних політик. Для обсерваційних дизайнів – прозорість припущень (публічні DAG, pre-analysis plan), мінімізація доступів і суворий контроль приватності. Управління процесом – комітет з експериментів (ExO + юристи + функціональні керівники) [8].

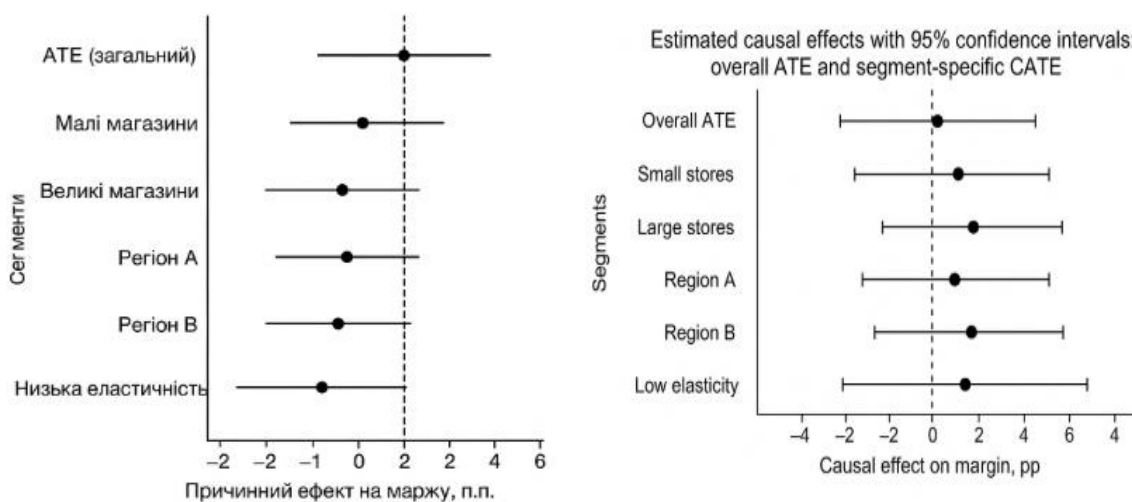


Рис. 3. Візуалізація ефектів та невизначеності (ATE/CATE)

Менеджерська інтеграція результатів спирається на очікувану прирідну цінність (EPV) та вартість додаткової інформації (VOI). EPV: $EPV = p_{win} \cdot \Delta\pi - C_{impl}$, де p_{win} – ймовірність позитивного ефекту (з урахуванням невизначеності оцінки), $\Delta\pi$ – причинний приріст прибутку/маржі, C_{impl} – витрати впровадження.

VOI: чи виправданий ще один раунд даних/експериментів перед масштабуванням (EVSI, EVPPI). Актуальні методичні огляди пропонують покрокові

інструкції та інструменти, придатні поза медичною – для бізнес-рішень із невизначеністю та витратами затримки [16].

Порогові правила: масштабувати, якщо $EPV > 0$ і $EVSI$ наступного раунду $<$ вартість затримки (втрачені вигоди за період); інакше – добирати інформацію, коли її очікувана цінність перевищує витрати і ризики.

Отже, провідник для вибору дизайну та контролю якості оцінок матиме такий вигляд (табл. 2):

Таблиця 2. «Дизайн → Припущення → Діагностики → Типові КРІ»

Дизайн	Ключові припущення	Діагностики/робастність	Типові управлінські КРІ
RCT/A-B	Валідна рандомізація, відсутність спіловуру	Баланс коваріат, attrition, alpha-spending	Маржа, конверсія, LTV
DiD (multi-period)	Паралельні тренди (умовні), стабільність складу	Event-study, placebo-у-часі/просторі, ваги	Маржа/GMV, продуктивність
SCM	Лінійна комбінація донорів відтворює pre-trend	Placebo-у-часі/просторі, чутливість до пулу	Продуктивність цеху/заводу
RDD	Безперервність за порогом, відсутність маніпуляції	McCrary, bandwidth/polynomial робастність	Бонуси/штрафи, якість
IV	Релевантність, екзогенність, LATE	Partial RI, F-стат., оверідентифікація	Запаси, ціни, випуск
DR/TMLE/CF	Правильність хоча б однієї моделі; стабільна CATE	Honest splitting, CV, чутливість	Таргетинг промо/цін

Кожну інтервенцію супроводжує «паспорт ефекту»: питання → DAG → дизайн → припущення → pre-analysis plan → діагностики → основні/second-order результати → гетерогенність (CATE) → EPV/VOI-рішення → межі узагальнення → (за можливості) дані/код у репозиторії. Така структура підвищує відтворюваність і відповідає сучасним редакційним практикам [8].

Таким чином перехід до каузального менеджменту в компанії вимагає чіткої архітектури ролей. Пропонуємо створити Експериментальний офіс (ЕхО) як міжфункціональну одиницю, що: (i) допомагає підрозділам формулювати контрфактичні питання та будувати DAG; (ii) обирає ідентифікаційні дизайни (RCT/DiD/SCM/RDD/IV/DR) і реєструє pre-analysis plan; (iii) забезпечує незалежну валідацію результатів і їх відтворюваність; (iv) веде реєстр «паспортів ефектів» і портфель EPV/VOI [6, 8].

Власники політик (marketing, operations, HR, finance) відповідають за реалізацію втручань і дані;

Комітет з експериментів (ЕхО + юридичний/етичний офіс + C-level) ухвалює рішення Go/No-go та слідкує за ризиками, приватністю і відповідністю стандартам прозорості [8]. Для прозорості вводимо правило: жодна ключова зміна не масштабується без «паспорта ефекту» з валідованою ідентифікацією [6]. Додаткову дисципліну забезпечують політики відкритості даних і коду згідно з сучасними TOP-стандартами [17], а отже 12-тижневий план (спринти 1–4) мають озподілятися приблизно так:

Спринт 1 (тижні 1–3). Фокус і підготовка.

– затвердити 3–5 бізнес-ефектів (дохід, маржа, продуктивність, ризики, задоволеність клієнтів) і baseline-KPI;

– провести «літкод-курс» для керівників: контрфакти, DAG, відмінність між прогнозом та причинністю [7];

– затвердити політику відкритості: реєстрація гіпотез, шаблони звітності, зберігання даних/коду [8, 17].

Спринт 2 (тижні 4–6). Дизайн і протоколи.

– відібрати 2–3 швидких втручання з максимальним управлінським значенням (ціни/промо; мікронавчання змін; технологічний апгрейд);

– для кожного втручання – DAG, «дерево вибору» дизайну pre-analysis plan, ризик-матриця та етичні обмеження [6, 8];

– погодити схему даних (Unit/Period), контроль якості, журнал трансформацій (audit-trail). Методичні орієнтири: DiD-підходи для стагерованого таймінгу [1, 2, 11], синтетичний контроль для одиничних втручань [5, 9], IV за наявності релевантного інструмента [10].

Спринт 3 (тижні 7–9). Збір даних та оцінювання.

– запустити RCT або квазіексперименти (DiD/SCM/RDD/IV/DR) залежно від контексту;

– виконати діагностики: pre-trends і placebo для DiD/SCM [1], щільність і bandwidth для RDD [7], partial RI, F-статистику та перевірку припущення виключення для IV [10];

– знайти та задокументувати гетерогенність ефектів (CATE) із застосуванням сучасних каузальних лісів/DR-оцінювачів [8,12].

Спринт 4 (тижні 10–12). Рішення, масштабування, навчання організації.

– розрахувати EPV та VOI для кожного втручання: масштабувати, якщо $EPV > 0$ і очікувана цінність додаткових даних (EVSI) менша за вартість затримки; інакше – провести ще один цикл збору доказів [16];

– оприлюднити внутрішній white-paper з «паспортом ефекту», додати кейс у каталог; зафіксувати уроки (what worked/what failed);

– інтегрувати результати у OKR/бюджетний цикл; оновити політики найму/навчання персоналу для закріплення нових практик [6].

Кожне втручання отримує уніфікований запис: мета → каузальне питання → DAG → дизайн → припущення → план аналізу → діагностики → результати (ATE/CATE) → рішення EPV/VOI → межі узагальнення → дата-версія → відповідальні. Каталог виконує роль корпоративної пам'яті і «захисту від помилкових перемог», оскільки повторні ініціативи перевіряються на відтворюваність у нових умовах [8, 17]. Для підсилення зовнішньої валідності запускаємо реплікаційний цикл: щокварталу 1–2 ключові політики перевіряються на інших майданчиках (заводи/ринки) [7].

Результати каузального оцінювання безпосередньо переписують OKR на наступний квартал: цілі формулюються через валідовані інтервенції (напр., «+1,0 п.п. маржі на SKU-кластері X завдяки перевірній політиці ціноутворення DiD з CATE-таргетуванням»). Для бюджетування застосовується портфельний підхід: кожна ініціатива має EPV/VOI-картку; фронтір будують у просторі «очікувана цінність – ризик – ресурс». Пріоритет: спершу масштабуються інтервенції з високим EPV і низькою VOI [16]. Захист від KPI-маніпуляцій забезпечують стандарти відкритості й відтворюваності (версійність метрик, незалежна валідація, аудит коду/даних) відповідно до TOP-рекомендацій [17].

Матриця зрілості та бенчмаркінг прогресу

Запропоновано 5 рівнів зрілості: 0 – Дашборди; 1 – Експерименти ad-hoc; 2 – Керована практика (ExO, каталоги, pre-analysis); 3 – Портфель EPV/VOI з реплікаціями й CATE-таргетуванням; 4 – Каузальна культура (паспорт ефекту для всіх ключових політик, щорічний зовнішній аудит, відкриті white-papers).

Перехід між рівнями оцінюємо чек-листами прозорості, повнотою метаданих та часткою масштабованих ініціатив із позитивним EPV [6, 8].

Емпіричні віньетки (кейси) з різних функцій

1. Ціноутворення в роздрібі (B2C): короткий A/B + CATE-таргетування

Запит. Чи дає –5% корекція ціни в SKU-кластері X приріст маржі без втрати обороту?

Дизайн. Польовий A/B-експеримент на рівні «магазин-тиждень»: 40/60 рандомізація; primary KPI – валова маржа, secondary – GMV. Pre-analysis plan зафіксовано.

Діагностики. Перевірено баланс коваріат; контроль spillover; α -spending для послідовного моніторингу [14].

Оцінювання. ATE з robust-SE; гетерогенність через causal forests (CATE) з honest-splitting [12].

Результат. ATE \approx +0.6 п.п. маржі; найбільший CATE у «малих магазинах» (+1.2 п.п.), нульовий ефект у «великих».

Рішення. Масштабувати частково: таргетувати «малі магазини» ($EPV > 0$); у «великих» – VOI виправдує ще один цикл даних [16].

Примітка. Стандарт експериментального менеджменту та протоколів – [3, 14].

Візуалізація: див. Рис. 3 (ATE/CATE).

2. Капітальні інвестиції у виробництві: робот-зварник на одному заводі

Запит. Чи підвищує новий робот-зварник випуск і знижує COPQ у цеху №2?

Дизайн. Одиничне втручання → синтетичний контроль (SCM): донорський пул – інші 8 цехів без роботи; часовий зріз – 24 місяці «до», 12 «після» [5, 9].

Діагностики. Placebo-у-часі/просторі; чутливість до складу пулу.

Оцінювання. Ефект на продуктивність: +7–9% (стабільний у 3 з 4 placebo-перевірок). COPQ – –0.8 п.п.

Рішення. Roll-out на ще два цехи з поетапним «staggered adoption» і подальшим DiD-моніторингом [1].

3. HR-втручання на зміні: мікро-навчання бригад

Запит. Чи зменшить 12-тижнева мікро-програма навчання абсентеїзм і підвищить випуск?

Дизайн. DiD із багатоперіодним таймінгом: 6 змін впроваджують втручання у 3 хвили; контроль – 6 змін без програми [11].

Діагностики. Event-study претрендів; placebo-тести; робастність до альтернативних ваг [1, р. 1].

Оцінювання. АТТ: –0.6 дня абсентеїзму/працівника/міс; +2.1% випуску.

Рішення. Масштабувати; у чутливих підрозділах додатково перевірити CATE (новачки vs досвідчені) [7, pp. 148–151].

Етика/прозорість. Інформована згода, відкриті протоколи – згідно з редакційною рамкою для неекспериментальних доказів [8, pp. 1846–1850].

4. Ланцюги постачання: чи виправданий більший safety-stock?

Запит. Який причинний ефект збільшення safety-stock на відмови поставок і маржу з урахуванням випадкових затворів у логістиці?

Дизайн. IV-підхід: інструмент – випадкові простоти на прикордонному хабі, що впливають на рішення про підвищення safety-stock, але не впливають безпосередньо на маржу; дані – «завод-тиждень», 30 місяців [10].

Діагностики. Partial RI і F-статистика релевантності; обґрунтування виключення; аналіз локальності LATE [15].

Оцінювання. LATE: -1.3 п.п. відмов поставок; чистий ефект на маржу близький до нуля (витрати на утримання складу «з'їдають» виграш).

Рішення. Таргетоване підвищення safety-stock у «ризикових» ланцюгах; для інших – VOI: дешево перевірити альтернативні політики (перерозподіл запасів, контрактні штрафи) [16, pp. 1–4].

5. Розкриття інформації та попит (маркетинг/ESG)

Запит. Чи змінює розкриття ESG-інформації поведінку клієнтів?

Дизайн. Рандомізований польовий експеримент у CRM-каналах: клієнтські когорти випадково отримують різні комунікації (ESG-факти, соціальні сигнали, контроль); KPI – open/click/purchase.

Діагностики. Баланс коваріат, відсутність перетину каналів, α -корекції [14].

Оцінювання. Короткострокове підвищення purchase-rate на 3–5% у когорті «ESG-факти», ефект згасає після 6–8 тижнів – узгоджується з польовими доказами [4].

Рішення. Пульсова стратегія: періодичні ESG-кампанії + перевірка довших «нарративних» форматів; EPV>0 лише за низької собівартості контакту.

6. Узагальнення уроків для С-рівня

1. Спочатку дизайн, потім оцінювач. У всіх кейсах саме вибір дизайну (RCT/DiD/SCM/IV) забезпечував валідність, а не «красива» регресія [6, 7].

2. Гетерогенність важить. CATE-підхід змінив рішення у 5.1 і 5.3: масштабування стало таргетованим [12].

3. EPV/VOI дисциплінують масштабування. У 5.2 та 5.4 інвестиції приймалися/відкладалися з огляду на вартість інформації та ризик затримки [16].

4. Прозорість і «паспорт ефекту». Документування припущень, протоколів і діагностик зменшує KPI-маніпуляції та полегшує реплікації [8, 17].

Інтеграція результатів у процес прийняття рішень проведемо через механізм «гіпотеза → ідентифікація → оцінка → EPV/VOI → рішення». Таким чином інтеграція каузальних результатів у менеджмент починається з контрфактичного формулювання («що буде, якщо...»), фіксації припущень у DAG і вибору дизайну (RCT/DiD/SCM/RDD/IV/DR) [6, 7]. Після оцінювання ATE/CATE результати перекладаються у гроші через очікувану прирістну цінність (EPV) та вартість додаткової інформації (VOI): масштабувати, якщо EPV>0 і очікувана цінність ще одного раунду даних (EVSI) менша за вартість затримки; продовжувати збір доказів – якщо навпаки [16]. Такий «воронокподібний» механізм робить рішення відтворюваними й захищає від масштабу «хибних перемог» [8].

Ми пропонуємо три стандартизовані шаблони:

- Go-Scale. (а) ATE/CATE стабільні у діагностиках; (б) EPV>0; (в) EVSI < вартість затримки → масштабувати; зафіксувати «паспорт ефекту» й оновити політики [8, 16].

- Go-Pilot. ATE позитивний, але невизначеність велика або ефект неоднорідний → пілот у сегментах з найвищим CATE [7, pp. 148–151; 12, pp. 1–3].

- No-Go/Redesign. Порушені припущення ідентифікації або EPV≤0 → зупинка/перепроєктування (оновити DAG і протокол) [6].

Для CAPEX/операцій корисне «правило двох дверей»: масштабування лише після незалежної валідації (інший майданчик/часовий зріз) і оприлюд-

нення пакета діагностик (pre-trend/placebo, чутливість, LATE-інтерпретація тощо) [1, 5, 10].

Результати мають бути зрозумілі без статистичної підготовки:

- для кожної політики – ATE і CATE з 95% ДІ та «світлофором рішень» (Go-Scale/Go-Pilot/No-Go);

- причинні лісплоти (forest plot) для сегментів, лінія нульового ефекту (див. рис. 3);

- окремий блок припущення → діагностики → вердикт (pre-trends, placebo, RDD-щільність, IV-релевантність), щоб не змішувати прогностичну та причинну інформацію [8].

Такі дашборди дисциплінують інтерпретацію й унеможливають «прикрасу» результатів (p-hacking), оскільки фокусують увагу на припущеннях і стійкості ефектів [6].

Кожне Go-Scale тягне за собою переписування OKR: ціль формулюється через каузально валідоване втручання (напр., «підвищити валову маржу на +1.0 п.п. у кластері X через перевірену політику ціноутворення DiD; охоплення – сегменти з CATE≥0.8 п.п.») [7]. Бюджет переходить від «проектів» до портфеля втручань, ранжованих за EPV/ризиком/ресурсом; портфель переглядається щокварталу разом із реплікаційним планом (мінімум одна зовнішня валідація на квартал) [16].

Розподіл відповідальності: власники політик відповідають за виконання та якість даних; ExO – за дизайн, аналіз і «паспорт ефекту»; Комітет з експертів – за етику, приватність і остаточне «Go/No-Go» [8]. Запроваджується правило «ніяких масштабувань без паспорта» і TOP-вимоги (версійність коду/даних, репозитарій артефактів) [17].

Щоб уникнути «підтасування KPI», застосовуються:

1. Розділення ролей: підрозділ виконує політику, ExO аналізує, окрема команда валідує [8, pp. 1846–1850].

2. Незмінність визначень KPI на період експерименту; зміни – лише з наступним циклом OKR.

3. Попередня реєстрація (pre-analysis plan) і аудит трансформації даних [6, pp. 1–4; 17, pp. 2–6].

4. «Крива вартість-час»: будь-яке відтермінування масштабування повинно мати позитивну VOI-оцінку [16].

Порівняння з альтернативними підходами (ML-прогнози без ідентифікації)

Результати статті підтверджують, що прогностичні моделі без ідентифікації часто плутають кореляцію з причинністю: вони добре описують «що буде, якщо нічого не змінювати», але не дають надійної відповіді на «що станеться, якщо ми втрутимось». Саме тому в управлінських задачах критичними є DAG-мислення та відповідний дизайн (RCT/DiD/SCM/RDD/IV), а не складність алгоритму [6, 7]. Практика провідних медичних і соціально-економічних журналів також вимагає від авторів явного формулювання каузального питання, передумов і діагностик – інакше висновки лишаються описовими [8]. Таким чином, каузальна валідність має пріоритет над «точністю прогнозу» у відриві від контрфактичної логіки.

Інтеграція EPV/VOI переводить каузальні оцінки у мову бюджету: масштабування виправдане там, де очікувана прирістна цінність позитивна, а додаткові дані вже мало що змінять у рішенні [16]. Разом із тим, зовнішня валідність потребує реплікацій – на інших майданчиках/ринках та у часі. Для одиничних інтервенцій (на кшталт CAPEX) синтетичний контроль пропонує прозорі критерії правдоподібності та чутливості до пулу донорів [5, 9], а для стагерованого

впровадження політик сучасні варіації DiD допомагають уникати зміщень TWFE та надають діагностики паралельних трендів [1]. Підсумково, поєднання EPV/VOI з реплікаціями створює портфельну логіку масштабування у менеджменті.

Обмеження та загрози ідентифікації також мають місце. По-перше, невимірні confounders можуть спотворити оцінки у спостережних даних; це зменшується через проєктування (front-door/back-door), інструментальні зміни та чутливісні аналізи [6, 10]. По-друге, динамічна гетерогенність ефектів загрожує валідності класичних панельних оцінювачів; сучасні підходи DiD адресують це через подієві дослідження й альтернативні ваги [1]. По-третє, ризик р-hacking і «сад стежок» мінімізується попередньою реєстрацією, чіткими протоколами та відкритістю даних/коду за TOP-принципами [6, 17]. Нарешті, обмеження ресурсів та етика можуть звужити простір експериментів; тоді ключовими стають квазіекспериментальні дизайни з прозорими припущеннями та діагностиками [8].

Напрями подальших досліджень: 1) каузальний ML і CATE-таргетування – потрібні протоколи поєднання causal forests/DR-оцінювачів з рішеннями реального часу – з формальними тестами стабільності CATE у продакшені [12]; 2) каузальні панелі для CAPEX – розвиток гібридів SCMЧDiD для багатосайтових модернізацій і довгих горизонтів оцінювання [5]; 3) стандарти реплікацій у менеджменті – оновлені редакційні вимоги (пре-реєстрація, публічні DAG, мінімальний набір діагностик) варто гармонізувати між галузями, зокрема із вимогами TOP-ініціатив [8,17]; 4) вартість інформації поза охороною здоров'я – адаптація VOI/EVSI до промислових і комерційних рішень із явним обліком «вартості затримки» [16, pp. 1–4]; 5) управлінська освіта – розробка коротких «каузальних мінімумів» для С-рівня; емпірична оцінка впливу такої освіти на якість стратегічних рішень [3]

Висновки:

1. Від описовості до керованої причинності. Запропонована дорожня карта демонструє, що трансформація від дашбордної аналітики до каузального менеджменту є передусім інституційним завданням. Розгортання практик DAG-мислення, прозорих ідентифікаційних дизайнів (RCT/DiD/SCM/RDD/IV/DR) та стандартизованих діагностик переводить аналітичні висновки у площину відтворюваних управлінських рішень.

2. Єдина «операційна мова» рішень. Формалізація ефектів через ATE/CATE, «паспорт ефекту» й мінімальний набір перевірок (pre-trends/placebo, щільність у RDD, релевантність інструмента в IV, робастність CATE) створює спільну мову для менеджерів, аналітиків і аудиторів. Це знижує ризик «хибних перемог» і розв'язує хронічне непорозуміння між статистикою та бізнес-процесами.

3. Рішення в грошовому вимірі. Інтеграція EPV/VOI дисциплінує масштабування: експерименти і квазіексперименти оцінюються не лише за статистичною значущістю, а за очікуваною прирістною цінністю з урахуванням вартості затримки. Це переводить каузальні оцінки в портфельний підхід до ресурсів (капітал, час, увага керівників).

ЛІТЕРАТУРА

1. Baker A., Larochelle N., Li S., Roth J. Difference-in-Differences Designs: A Practitioner's Guide. *arXiv preprint arXiv:2503.13323*. 2025. URL: <https://arxiv.org/abs/2503.13323>

2. Caetano C., & Callaway B. Difference-in-Differences when Parallel Trends Holds Conditional on Covariates. *Working paper*. 2024. URL: https://bcallaway11.github.io/files/DID-Covariates/Caetano_Callaway_2024.pdf

4. Таргетоване масштабування замість «середніх» рішень. Систематичний облік гетерогенності (CATE) показує, що «середній» ефект часто приховує управлінську можливість: масштабування доцільне саме в тих сегментах, де очікуваний причинний вплив є найбільшим. Отже, політики ціноутворення, навчання персоналу чи управління запасами мають впроваджуватись селективно, а не рівномірно.

5. Організаційна архітектура каузальності. Запропонований Експериментальний офіс (ExO), комітет з експериментів, каталог «паспортів ефектів» і матриця зрілості переводять каузальні практики зі статусу «ентузіазму окремих команд» у постійну управлінську функцію. Це забезпечує сталість, спадкоємність знань і незалежну валідацію.

6. Етика й прозорість як частина керованості ризику. Пре-реєстрація гіпотез, відкриті припущення та контроль доступу до даних не є «науковим формалізмом», а працюють як механізми управління репутаційним і регуляторним ризиком. Етична доцільність і пропорційність втручань закладаються в дизайн на рівні з технічною ідентифікацією.

7. Зовнішня валідність через реплікації. Поєднання локальних експериментів/квазіекспериментів із запланованими реплікаціями на інших майданчиках і в інші періоди забезпечує переносимість рішень. Для одиничних втручань доцільно застосовувати SCM із системними placebo-перевірками; для стадерованих політик – сучасні варіації DiD із подієвими дослідженнями і коректними вагами.

8. Практичний ефект для С-рівня. Запропонована рамка дозволяє швидко переходити від гіпотези до керованого рішення: «Go-Scale» (масштабування), «Go-Pilot» (таргетовані пілоти) або «No-Go/Redesign» (зупинка чи перепроєктування). Прийняття рішень стає прозорим, аргументованим і зв'язаним з бюджетним циклом та OKR.

9. Обмеження та кордони застосовності. Емпіричні стратегії залишаються чутливими до невимірних конфаундерів, маніпуляцій за порогоми та локальності інструментів. Тому результати слід інтерпретувати як умовно-каузальні відносно конкретно сформульованих припущень і підтримувати їх регулярними чутливими аналізами та незалежними повтореннями.

10. Подальша траєкторія розвитку. Найперспективнішими є: інтеграція каузального ML у реальні операційні цикли з контролем стабільності CATE; гібриди SCMЧDiD для довгих CAPEX-горизонтів; уніфікація редакційних і внутрішньокорпоративних стандартів прозорості; поширення VOI-методів поза охороною здоров'я на виробничі та комерційні рішення.

Підсумовуючи, робота формує цілісний методично-організаційний каркас каузального менеджменту – від постановки контрфактичних питань і вибору дизайну до портфельних рішень і інституціоналізації процесів. Це створює відчутну управлінську перевагу: зменшення вартості помилок, швидше «відсіювання» неефективних ініціатив, таргетоване масштабування ефективних політик і краща керованість ризиком у мінливому середовищі.

3. Camuffo A., Gambardella A., Messinese D., Novelli E., Paolucci E., Spina, C. A scientific approach to entrepreneurial decision-making: Large-scale replication and extension. *Strategic Management Journal*. 2024. Vol. 45, Iss. 6. P. 1209–1237. <https://doi.org/10.1002/smj.3580>
4. Leonelli S., Muhn M., Rauter T., Sran G. How Do Consumers Use Firm Disclosure? Evidence from a Randomized Field Experiment (BFI Working Paper No. 2024-04). Becker Friedman Institute, 2024. <https://doi.org/10.2139/ssrn.4687694>
5. Ben-Michael E., Feller A., Rothstein J. Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society: Series B*. 2022. Vol. 84, Iss. 2. P. 351–381. <https://doi.org/10.1111/rssb.12448>
6. Hernan M. A., Robins J. M. Causal Inference: What If. Harvard T.H. Chan School of Public Health, 2024, January 2. URL: https://content.sph.harvard.edu/wwwhsph/sites/1268/2024/01/hernanrobins_WhatIf_2jan24.pdf
7. Imbens G. W. Causal inference in the social sciences. *Annual Review of Statistics and Its Application*. 2024. Vol. 11. P. 123–152. <https://doi.org/10.1146/annurev-statistics-033121-114601>
8. Dahabreh I. J., Bibbins-Domingo K. Causal inference about the effects of interventions from observational studies in medical journals. *JAMA*. 2024. Vol. 331, Iss. 21. P. 1845–1853. <https://doi.org/10.1001/jama.2024.7741>
9. Abadie A. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*. 2021. Vol. 59, Iss. 2. P. 391–425. <https://doi.org/10.1257/jel.20191450>
10. Mogstad M., Torgovitsky A. Instrumental variables with unobserved heterogeneity in treatment effects. *Handbook of Labor Economics*. 2024. Vol. 5. P. 1–114. Elsevier. <https://doi.org/10.3386/w32927>
11. Callaway B., Sant’Anna P. H. C. Difference-in-Differences with multiple time periods (vignette). 2024. September 10. URL: <https://bcallaway11.github.io/did/articles/did-basics.html>
12. Rehill P., Gronsbell J., van der Laan M. J. How do applied researchers use the causal forest? A methodological review. *International Statistical Review*. 2025. Vol. 93, Iss. 2. P. 288–316. <https://doi.org/10.1111/insr.12610>
13. Caetano C., Callaway B. Difference-in-Differences when Parallel Trends Holds Conditional on Covariates. *arXiv:2406.15288*. 2024. URL: <https://arxiv.org/abs/2406.15288>
14. Callaway B. Introduction to DiD with Multiple Time Periods (did vignette). 2024, September 10. URL: <https://bcallaway11.github.io/did/articles/multi-period-did.html>
15. Mogstad M., Torgovitsky A. Instrumental Variables with Unobserved Heterogeneity in Treatment Effects (NBER Working Paper No. 32927). 2024. <https://doi.org/10.3386/w32927>
16. Value of Information for Healthcare Decision-Making / Eds. by Heath A., Kunst N., Jackson C. Taylor & Francis, 2024. <https://doi.org/10.1201/9781003156109>
17. Transparency and Openness Promotion (TOP) Guidelines, Version 2.0. Center for Open Science. Charlottesville, VA, 2024. URL: <https://www.cos.io/initiatives/top-guidelines>
18. Thomke S. Experimentation Works: The Surprising Power of Business Experiments. Harvard Business Review Press, 2020. <https://doi.org/10.1080/08956308.2020.1762443>
19. Kohavi R., Tang D., Xu Y. Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing. Cambridge University Press, 2020. <https://doi.org/10.1017/9781108653985>

Надійшла до редакції 25.09.2025 р.
Прийнята до друку 23.10.2025 р.

REFERENCES

1. Baker, A., et al. (2025). *Difference-in-Differences Designs: A Practitioner’s Guide*. arXiv preprint arXiv:2503.13323. <https://arxiv.org/abs/2503.13323>
2. Caetano, C., & Callaway, B. (2024). *Difference-in-Differences when Parallel Trends Holds Conditional on Covariates*. Working paper. https://bcallaway11.github.io/files/DID-Covariates/Caetano_Callaway_2024.pdf
3. Camuffo, A., Gambardella, A., Messinese, D., Novelli, E., Paolucci, E., & Spina, C. (2024). A scientific approach to entrepreneurial decision-making: Large-scale replication and extension. *Strategic Management Journal*, 45(6), 1209–1237. <https://doi.org/10.1002/smj.3580>
4. Leonelli, S., Muhn, M., Rauter, T., & Sran, G. (2024). *How Do Consumers Use Firm Disclosure? Evidence from a Randomized Field Experiment* (BFI Working Paper No. 2024-04). Becker Friedman Institute. <https://doi.org/10.2139/ssrn.4687694>
5. Ben-Michael, E., Feller, A., & Rothstein, J. (2022). Synthetic controls with staggered adoption. *Journal of the Royal Statistical Society: Series B*, 84(2), 351–381. <https://doi.org/10.1111/rssb.12448>
6. Hernan, M. A., & Robins, J. M. (2024, January 2). *Causal Inference: What If*. Harvard T.H. Chan School of Public Health. https://content.sph.harvard.edu/wwwhsph/sites/1268/2024/01/hernanrobins_WhatIf_2jan24.pdf
7. Imbens, G. W. (2024). Causal inference in the social sciences. *Annual Review of Statistics and Its Application*, 11, 123–152. <https://doi.org/10.1146/annurev-statistics-033121-114601>
8. Dahabreh, I. J., & Bibbins-Domingo, K. (2024). Causal inference about the effects of interventions from observational studies in medical journals. *JAMA*, 331(21), 1845–1853. <https://doi.org/10.1001/jama.2024.7741>
9. Abadie, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2), 391–425. <https://doi.org/10.1257/jel.20191450>
10. Mogstad, M., & Torgovitsky, A. (2024). Instrumental variables with unobserved heterogeneity in treatment effects. *Handbook of Labor Economics*, 5, 1–114. Elsevier. <https://doi.org/10.3386/w32927>
11. Callaway, B., & Sant’Anna, P. H. C. (2024). *Difference-in-Differences with multiple time periods* (vignette, updated September 10, 2024). <https://bcallaway11.github.io/did/articles/did-basics.html>



12. Rehill, P., Grönsbell, J., & van der Laan, M. J. (2025). How do applied researchers use the causal forest? A methodological review. *International Statistical Review*, 93 (2), 288-316. <https://doi.org/10.1111/insr.12610>
13. Caetano, C., & Callaway, B. (2024). *Difference-in-Differences when Parallel Trends Holds Conditional on Covariates*. arXiv:2406.15288. <https://arxiv.org/abs/2406.15288>
14. Callaway, B. (2024, September 10). *Introduction to DiD with Multiple Time Periods* (did vignette). <https://bcallaway11.github.io/did/articles/multi-period-did.html>
15. Mogstad, M., & Torgovitsky, A. (2024). *Instrumental Variables with Unobserved Heterogeneity in Treatment Effects*. NBER Working Paper No. 32927. <https://doi.org/10.3386/w32927>
16. Heath, A., Kunst, N., & Jackson, C. (Eds.). (2024). *Value of Information for Healthcare Decision-Making*. Taylor & Francis. <https://doi.org/10.1201/9781003156109>
17. Center for Open Science. (2024). *Transparency and Openness Promotion (TOP) Guidelines, Version 2.0*. Charlottesville, VA. <https://www.cos.io/initiatives/top-guidelines>
18. Thomke, S. (2020). *Experimentation Works: The Surprising Power of Business Experiments*. Harvard Business Review Press. <https://doi.org/10.1080/08956308.2020.1762443>
19. Kohavi, R., Tang, D., & Xu, Y. (2020). *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press. <https://doi.org/10.1017/9781108653985>

Received: 25.09.2025

Accepted: 23.10.2025

Ткаченко А. М., Чорний Р. С. Дорожня карта для топ-менеджера від описової статистики до каузального менеджменту

За останнє десятиліття корпоративні системи аналітики швидко еволюціонували від ручної звітності до масштабованих платформ бізнес-інтелекту із розвинутими дашбордами, а також прогностичні і рекомендаційні моделі. Незважаючи на це, значна частина управлінських рішень і далі спирається на кореляційні залежності або евристики, які рідко відповідають на ключове контрфактичне питання: що станеться, якщо ми втрутимося і модифікуємо політику ціноутворення, рівень запасів, технологію, навчальні програми, маркетингові комунікації чи інституційні правила? Відсутність операційного мосту між «описом/прогнозом» і «причинністю» породжує системні ризики: помилкові атрибуції, масштабування «хибних перемог», надмірне довір'я до короткострокових кореляцій і хронічне недоінвестування в інтервенції з високою очікуваною цінністю, але слабким історичним слідом у даних. На цьому тлі зростає потреба в дорожній карті, яка переводить фірму з режиму «дивимосся і здогадуємося» у режим керування доречності. У статті наголошено на правилах back-door/front-door, ідентифікації через подієві дослідження та ваги у DiD для стагерованого впровадження, особливостях синтетичного контролю для одиничних втручань, суворих вимогах до інструментальних змінних (релевантність, екзогенність, LATE-інтерпретація), а також на перевагах doubly-robust і targeted learning оцінювачів, що поєднують модель результату та модель схильності до лікування. Окремо висвітлено каузальний ML (uplift-моделі, causal forests) як інструмент виявлення стабільної гетерогенності ефектів для управлінського таргетування.

Ключові слова: каузальний менеджмент, DAG-діаграми, різниця-в-різницях, синтетичний контроль, інструментальні змінні, регресійна розривність, умовний причинний ефект, очікувана прирістна цінність, вартість інформації, експериментальний офіс.

Tkachenko A. M., Chornyi R. S. Roadmap for top managers from descriptive statistics to causal management

Over the past decade, corporate analytics systems have rapidly evolved from manual reporting to scalable business intelligence platforms with advanced dashboards, as well as predictive and recommendation models. As a result, a significant portion of management decisions continue to rely on correlations that rarely answer the key question: what will happen if we intervene and modify pricing policy, inventory levels, technology, training programmes, marketing communications, or institutional rules? The lack of an operational bridge between 'description/forecast' and "causality" creates systemic risks: false attributions, scaling of 'false victories,' overconfidence in short-term correlations, and chronic underinvestment in interventions with high expected value but with a weak historical track record in the data. Therefore, there is a growing need for a roadmap that moves the firm from a 'watch and guess' mode to a mode of managed causality.

Keywords: causal management, causal DAGs, DiD (difference-in-differences), SCM (synthetic control), IV (instrumental variables), RDD (regression discontinuity), CATE (conditional average treatment effect), EPV (expected profit value), VOI (value of information), ExO (experimental office).

Формат цитування:

Ткаченко А. М., Чорний Р. С. Дорожня карта для топ-менеджера від описової статистики до каузального менеджменту. *Вісник економічної науки України*. 2025. № 2 (49). С. 180-190. [https://doi.org/10.37405/1729-7206.2025.2\(49\).180-190](https://doi.org/10.37405/1729-7206.2025.2(49).180-190)

Tkachenko, A. M., & Chornyi, R. S. (2025). Roadmap for top managers from descriptive statistics to causal management. *Visnyk ekonomichnoi nauky Ukrainy*, 2(49), 180-190. [https://doi.org/10.37405/1729-7206.2025.2\(49\).180-190](https://doi.org/10.37405/1729-7206.2025.2(49).180-190)