

ARTICLE

Evaluating operational AI-enabled supply-chain innovation: A disclosure-aware boundary framework for inventory and working capital in Chinese manufacturing

Sen Ye*

Liaoning Finance & Trade College, Huludao, Liaoning, China

Abstract

Artificial intelligence (AI) is embedded in demand forecasting, replenishment, and supply-chain planning, yet evidence on its operating consequences remains scarce. Public disclosures can verify a system as operating during a fiscal year without revealing first internal use, production acceptance, or the exact go-live quarter; coding the report date as adoption would turn disclosure into a fictitious intervention. Therefore, this study investigates a public operating-use milestone, not a deployment event. Four hand-verified operating-use disclosures by Chinese A-share manufacturers, covering operating years 2019–2024, were linked to an audited quarterly CSMAR panel. For each firm, synthetic controls were estimated under first-quarter and fourth-quarter assignments to the disclosed operating year. These are maximal quarterly codings of an annual marker: they test whether a conclusion survives the greatest within-year reclassification that the public record permits, but do not locate deployment or numerically bound effects at interior quarters. The study distinguishes public operating-use evidence from separate accounting-gap evidence. Across the four inventory/deviation cases and three cash-cycle cases, no pooled diagnostic was below 0.05 under both assignments: inventory days ($p = 0.670/0.904$); industry-median deviation ($p = 0.363/0.386$); and cash conversion cycle ($p = 0.072/0.340$). The study does not show that operational AI improves inventory efficiency or working-capital performance. Its principal contribution is methodological: it demonstrates how disclosed operating years should be treated as coarsened time information and subjected to boundary sensitivity rather than converted into precise deployment dates. The results provide descriptive comparative patterns, including a sign-based distinction between inventory scale and industry-relative alignment, without causal or general performance claims. For systematic-innovation research, the contribution is an ex-post, disclosure-aware evaluation protocol for a realised intelligent operating innovation: it links a named operational application, transparent time coding, accounting outcomes, and explicit inference boundaries. It is not a method for generating inventions, a TRIZ extension, or an evaluation of the proprietary AI models firms use.

Keywords: Artificial intelligence; Supply-chain forecasting; Inventory; Working capital; Synthetic control; China

***Corresponding author:**
Sen Ye
(yesen007@gmail.com)

Citation: Ye, S. (2026). Evaluating operational AI-enabled supply-chain innovation: A disclosure-aware boundary framework for inventory and working capital in Chinese manufacturing. *Int J Systematic Innovation*. 10(5): 026300225.
[https://doi.org/10.6977/IJoSI.202610_10\(5\).0014](https://doi.org/10.6977/IJoSI.202610_10(5).0014)

Received: July 23, 2026

Revised: August 29, 2026

Accepted: September 10, 2026

Published online: October 27, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Artificial intelligence (AI) has moved from the perimeter of the firm into the routines by which materials are ordered, inventories are positioned, and production plans are revised. In contemporary supply chains, prediction is not merely an informational activity. A forecast changes a replenishment quantity; a replenishment quantity changes working capital; and changes in working capital can either protect the firm from disruption or leave cash immobilised in the wrong stock. This chain of consequences makes AI-enabled forecasting an unusually demanding test of whether digital capability creates operating value.

The promise is intuitive. Machine-learning models can detect nonlinear demand patterns, combine high-frequency operational signals, and revise decisions faster than static rules. Prior research links AI and big-data analytics to operational performance, resilience, and firm growth (Babina *et al.*, 2024; Baryannis, Dani, & Antoniou, 2019; Baryannis, Validi, *et al.*, 2019; Belhadi *et al.*, 2024; Dubey *et al.*, 2020). Within the Asia Pacific literature, evidence from Chinese organisations indicates that big-data use contributes to supply-chain effectiveness through agility rather than through a simple direct pathway, while work on Chinese manufacturers places organisational capabilities between digital transformation and competitive advantage (Wang & Ali, 2023; Li *et al.*, 2025). The organisational pathway from a model to a financial-statement outcome is therefore neither direct nor frictionless. Data must be reliable, forecasts must be translated into planning parameters, and procurement, production, logistics, and sales teams must act on the output. The productivity benefits of general-purpose technologies depend on complementary organisational investment and may arrive only after an adjustment period (Bresnahan *et al.*, 2002; Brynjolfsson & Hitt, 2000; Brynjolfsson *et al.*, 2021).

This tension is especially consequential in China. Chinese manufacturers operate across dense domestic networks and internationally exposed supply chains while demand volatility, rapid digitalisation, and trade-related disruptions increase the value of reconfiguration and organisational learning (Jiang *et al.*, 2026). The setting is therefore fertile for studying operational AI, but it is also challenging. Public annual reports often state that an AI-enabled forecasting or replenishment application was in use during a reporting year. They rarely disclose the first production go-live date, organisational coverage, or the quarter in which the application began influencing decisions. An annual report filed in the following spring is evidence about operations during the prior year; it is not a deployment timestamp. Coding the filing date as adoption would misclassify disclosure as treatment and, in many

instances, place the marker after the system had already been operating.

This paper examines a deliberately narrower and more defensible question: what operating patterns are observed around the first publicly verifiable report that describes operational use of an AI-enabled supply-chain forecasting, replenishment, or planning application among Chinese A-share manufacturers, once uncertainty about quarterly alignment is retained? This study does not claim that the report marks the firm's first internal use, production acceptance, pilot start, or full rollout. The estimand is tied to a public operating-use milestone, not to an unknowable technological birth date or to a causal deployment effect.

This study builds a curated evidence registry from corporate annual reports and applies inclusion rules fixed before examining outcome estimates. The disclosure must identify an AI or machine-learning capability; connect it to forecasting, replenishment, inventory, or supply-chain planning; and describe operational use rather than aspiration, research, or procurement. Four manufacturing firms meet these requirements within the 2018–2025 accounting window: Shangpin Home Collection, Midea Group, Winner Medical, and Bear Electric. A fifth fully adjudicated candidate, Guangdong SACA Precision Manufacturing, was excluded because China Stock Market and Accounting Research (CSMAR) classifies it as retail in the disclosed operating year. The registry is a verified case series, not an exhaustive census of every unreported or differently worded AI application. The four cases cover different manufacturing industries and disclosed operating years, precluding a conventional staggered-adoption design with a large treated population but allowing transparent comparative case analysis.

The financial panel was constructed from licensed CSMAR quarterly balance-sheet, income-statement, and cash-flow data. Cumulative Chinese interim flows were converted into stand-alone quarters before operating ratios were calculated. Then, a separate synthetic control was estimated for each firm and outcome, matching within the same two-digit manufacturing industry on up to eight pre-assignment quarters and pre-assignment covariates. To expose annual timing sensitivity, every case was estimated twice. The early model assigned the disclosed operating-year marker to its first quarter; the late model assigned the same marker to its fourth quarter. These endpoints were selected because an annual disclosure supplies no within-year ordering: Q1 and Q4 impose the maximum possible contrast in quarterly pre/post classification while preserving the same disclosed year and donor rule. They are not a confidence interval, a deployment interval, or a numerical bound on results under unobserved Q2/Q3

or gradual-rollout specifications. Case-level placebo tests and pooled randomisation inference assessed whether the observed comparative gaps are unusual relative to comparable pseudo-treated firms.

Three outcomes separate inventory scale from working-capital consequences. Inventory day captures the scale of stock relative to quarterly cost of goods sold. The cash conversion cycle adds receivables and payables to inventory, linking supply-chain operations to working-capital liquidity. An industry-median inventory deviation measures the absolute distance between a firm's log inventory days and the median for its industry and quarter. The median is a descriptive reference distribution, not a structural optimum, service-level target, or performance frontier. Accordingly, a fall in inventory days, cash conversion cycle, or deviation was not labelled a supply-chain improvement unless corroborated by unobserved service, availability, disruption, or profitability evidence.

The results do not support a general efficiency or working-capital claim. Midea's inventory-days gap was approximately -37.6% under Q1 (case placebo $p = 0.032$) and -16.0% under Q4 ($p = 0.355$), illustrating assignment sensitivity rather than a robust firm-level inference. Shangpin and Bear Electric showed positive inventory-day gaps under both assignments, whereas Winner Medical showed negative gaps. Industry-median deviations reported different signs across firms. Pooled diagnostics failed the stated two-assignment 0.05 rule for every outcome: inventory days ($p = 0.670/0.904$), deviation ($p = 0.363/0.386$), and cash conversion cycle ($p = 0.072/0.340$). The three-case cash-cycle evidence is particularly limited and incorporates receivables and payables beyond the named supply-chain application. The response labels therefore describe comparative accounting patterns, not performance rankings. The exploratory demand-volatility specification and safeguards are described in Section S6 and are not part of the study's main empirical contribution.

The study makes four contributions with different evidentiary strengths. First, its substantive contribution is to focus on operational AI in explicitly named supply-chain processes—forecasting, replenishment, and planning—rather than on generic corporate AI rhetoric. Second, its measurement contribution is to distinguish this task-specific operating-use construct from broad adoption proxies such as textual AI intensity, patents, or workforce measures. Third, and most centrally, its methodological contribution concerns the temporal interpretation of corporate disclosures: a report that describes operations during a year provides coarse operating-year evidence, not a precise deployment event, and should be analysed through disclosed-year boundary sensitivity. Fourth, its

result contribution is descriptive: it separates inventory scale relative to a synthetic comparison from industry-relative inventory alignment, thereby preventing a one-dimensional low-stock interpretation. None of these contributions claims that the four cases establish an average AI effect on inventory or working-capital performance. The complementarity argument is used as an interpretive boundary framework; its mechanisms were not measured or tested here.

The paper also contributes to systematic innovation at the evaluation stage. A systematic innovation process requires a traceable way to assess whether a realised intelligent solution has entered operating routines and whether publicly observable evidence supports any claimed operational consequence. The disclosure-aware protocol developed here specifies four linked elements: a task-specific operational-AI registry, a distinction between operating-use evidence and outcome evidence, extremal disclosed-year sensitivity assignments, and comparative diagnostics with stated non-causal boundaries. This is an evaluation framework for a realised innovation, not a procedure for generating inventions or training an AI model.

The remainder of the paper develops the theoretical argument (Section 2), describes the evidence and boundary-assignment synthetic-control design (Section 3), reports case and pooled results (Section 4), and discusses their implications for digital transformation in the Asia-Pacific business environment (Section 5). Section 6 concludes the study.

2. Theory and analytical expectations

2.1. Operational AI as an information-processing capability and the evaluation stage of systematic innovation

Supply-chain decisions are made under incomplete and changing information. Demand is only partially observed, lead times vary, and local decisions are connected through inventories and production schedules. AI can expand information-processing capacity by extracting signals from large, heterogeneous, and rapidly updating data. Machine-learning demand forecasting has long shown promise in supply-chain settings (Carbonneau *et al.*, 2008). In risk management, machine-learning models can identify patterns associated with delays or disruptions, although greater predictive performance may reduce interpretability (Baryannis, Dani, & Antoniou, 2019). In inventory management, learning algorithms can select replenishment policies as conditions change (Priore *et al.*, 2019). Applications reported in practice span forecasting, scheduling, logistics, supplier selection, and exception management (Helo & Hao, 2022).

Systematic innovation involves more than the generation of novel concepts. Once an intelligent solution enters a business process, a further methodological problem is how to evaluate that realised innovation without converting a public narrative into a fictitious deployment event or an unverified performance claim. This study addresses that ex-post evaluation problem. It treats operational AI as the object being evaluated and the disclosure-aware boundary design as the evaluative method; it does not claim to design the firms' algorithms, invent a new replenishment policy, or extend Theory of Inventive Problem Solving (TRIZ)-based ideation.

These capabilities create value only when embedded in a decision system. A more accurate forecast has no operating consequence unless planners trust it, master data are sufficiently clean, system interfaces transmit it, and responsible employees alter orders or schedules; poor data quality can undermine analytics throughout the supply chain (Hazen *et al.*, 2014). The distinction parallels the wider literature on information technology and organisational complements. Technology investments have been most productive when combined with workplace reorganisation, skilled labour, and new decision processes (Brynjolfsson & Hitt, 2000; Bresnahan *et al.*, 2002). Data-driven decision-making diffused rapidly across manufacturing plants, but adoption was correlated with complementary information technology (IT), human capital, and organisational learning (Brynjolfsson & McElheran, 2016). AI can therefore be understood as a general-purpose predictive input whose realised value is conditional on complementary intangible capital (Brynjolfsson *et al.*, 2021).

This perspective cautions against treating operational disclosure as homogeneous. The four cases studied here differ in task, organisational history, and stage of digital development. Shangpin described an upgrade to an existing intelligent supply-chain system rather than the first appearance of a digital system. Midea placed an intelligent replenishment algorithm within a broader architecture of algorithms, material-requirements planning, scheduling, and channel-inventory integration. Winner Medical linked machine learning to inventory, margin, and replenishment optimisation and reported factory deployment of its Win+ platform. Bear Electric described AI-supported demand forecasting used to formulate production and logistics plans. Their common element is operational predictive support for supply-chain decisions; their complements and initial conditions are not common.

The theoretical contribution is deliberately bounded. Complementarity provides a reason not to infer a common performance implication from a named AI application: operating outcomes plausibly depend on data quality,

system integration, decision rights, planner trust, coverage, and rollout stage. None of these complements is observed in the present design. The framework therefore guides construct definition and interpretation of heterogeneous comparative gaps; it is not a tested mediation model and cannot explain why any focal firm differs from its synthetic comparison.

2.2. Sources of heterogeneity in inventory outcomes

Inventory is both a cost and an option. Excess stock consumes cash, incurs holding and obsolescence costs, and can signal weak planning. Too little stock, however, increases stockouts, service disruptions, and risk. Classical inventory theory predicts that demand uncertainty, lead time, and gross margin shape appropriate inventory levels, and firm-level evidence shows that these relationships survive aggregation beyond individual products (Rumyantsev & Netessine, 2007). Empirical inventory-performance metrics also vary with sales surprises, capital intensity, and margin, making raw turnover an imperfect universal benchmark (Gaur *et al.*, 2005).

An AI forecasting or replenishment application is theorised to be associated with inventory through at least three channels. A precision channel could coincide with lower forecast error and less safety stock. A coordination channel could coincide with more synchronised procurement, production, warehouses, and channels. A learning-and-transition channel could coincide with larger buffers while models are calibrated, data coverage expands, or old and new routines operate in parallel. Managers may also respond to improved risk signals by holding more, rather than less, inventory where the expected cost of disruption is higher. These channels organise interpretation; the present accounting data do not measure them or adjudicate among them.

Consequently, lower inventory days cannot be equated automatically with better supply-chain management. Inventory theory treats stock as both a carrying cost and a service-and-resilience buffer; the appropriate level depends on demand uncertainty, lead times, margins, product mix, and the cost of stockouts. The industry-median inventory deviation therefore has a limited theoretical role: it records how far a firm sits from the contemporaneous distribution of comparable firms, which can flag an unusually high or low relative position without asserting that the median is optimal. In the comparative design, a negative observed-minus-synthetic gap means that the focal firm's post-assignment deviation is smaller than its synthetic comparison's; it does not establish a movement towards a firm-specific optimum, better service, lower disruption risk, or superior performance.

Taken jointly, inventory scale and industry-relative deviation define four descriptive response modes. For each firm and boundary assignment, the exact unrounded mean post-assignment log-inventory-days gap determines scale: a value below zero is labelled leaner, and a value above zero is buffering. The exact unrounded mean post-assignment deviation gap determines alignment: a value below zero is convergence, and a value above zero is divergence. A zero value is neutral and is not assigned a quadrant. A firm receives a named case-level mode only when the two signs are the same under both Q1 and Q4 assignments; otherwise, it is labelled boundary-sensitive or unclassified. These are classification rules, not significance tests, performance rankings, or evidence of managerial intent. They prevent a negative inventory gap from being treated automatically as optimisation and make cross-case heterogeneity interpretable without imposing a single performance ranking.

The preceding mechanisms yield a directional expectation but also a strong heterogeneity qualification:

- Analytical expectation 1 (non-causal). Around a disclosed operating-use year, inventory-days and industry-median-deviation gaps may display different signs across cases because organisational complements, initial inventory positions, and the function of buffer stock are unobserved. This is an interpretive expectation, not a test of a causal mechanism or a performance hypothesis.

2.3. Working-capital transmission

Supply-chain operations extend beyond physical stock. The cash conversion cycle combines days inventory outstanding, days sales outstanding, and days payables outstanding. It captures the time between cash committed to operating inputs and cash recovered from customers. Working-capital research links shorter collection and inventory periods to profitability, while emphasising that receivables, payables, financing constraints, and growth opportunities shape the value of operating working capital (Deloof, 2003; Kieschnick *et al.*, 2013).

Artificial intelligence-supported forecasting may coincide with a shorter cash cycle when inventory and order fulfilment patterns change, but integration across commercial and procurement processes may also coincide with countervailing movements. Better demand prediction can accompany customer credit or service commitments that increase receivables; strategic supplier relationships can change payment terms; and a firm building resilience may accept a longer cash cycle in exchange for continuity. Because the cash conversion cycle contains financial-policy components not directly observed in the named

operating-use disclosure, it is a particularly indirect and noisy accounting outcome in this study.

The same total cash-cycle gap can therefore conceal different operating pathways. A reduction may be inventory-led, collection-led, or financing-led through longer payment periods, while movements in one component may offset movements in another. An additive accounting decomposition using fixed synthetic-control weights can reveal these channels exactly because the un-winsorised identity is days inventory outstanding (DIO) plus days sales outstanding (DSO) minus days payables outstanding (DPO) (i.e., $DIO + DSO - DPO$). Such a decomposition is diagnostic rather than causal mediation: it identifies where the accounting gap sits, not why that component changed.

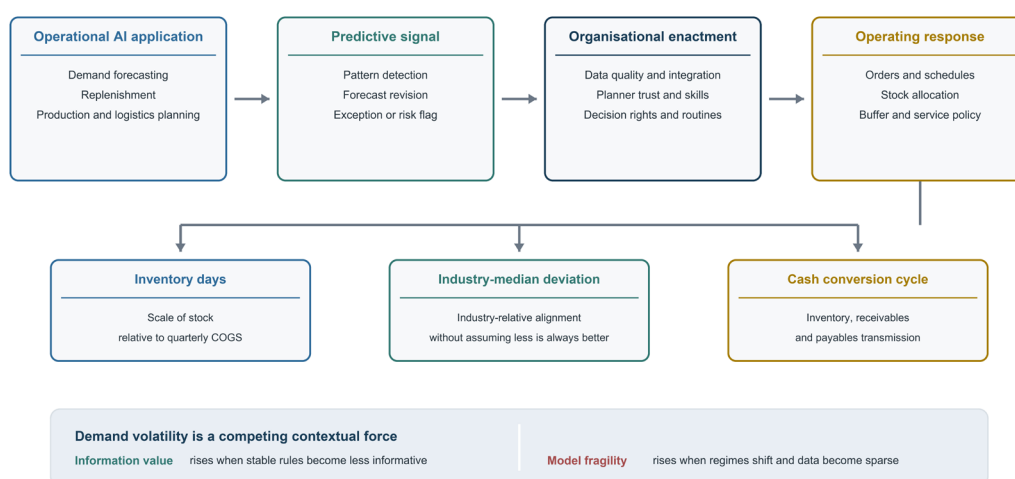
- Analytical expectation 2 (non-causal). Cash-conversion-cycle gaps may be less uniform than inventory-specific gaps because receivables, payables, and financing policies can move independently of the named supply-chain function. The study treats this as a reason for caution, especially because only three cases have complete cash-cycle inputs.

2.4. Demand volatility: exploratory appendix analysis

Demand volatility is retained only as an exploratory contextual description. It is not a theory test, a moderation analysis, or a contribution of the main paper. With four cases, a rolling firm-quarter sales-growth proxy, repeated observations, and cells with as few as one contributing case, the data cannot support an interaction estimator or an inference about when operational AI is more valuable. The exploratory proxy definition, design, and inferential safeguards are described in Section S6 for transparency.

The broader theoretical framework recognises that unstable demand can increase both the value of information and the difficulty of model calibration. The study does not empirically test either mechanism. Figure 1 is therefore an interpretive map of possible organisational pathways, not an empirical model and not evidence of demand-volatility moderation.

Figure 1 integrates the analytical argument. A named operating-use disclosure identifies a predictive technology and a supply-chain decision context, but accounting outcomes are observed only after unmeasured organisational enactment, including data quality, system integration, planner trust, and decision rights. The figure is used to delimit interpretation: the four-case design does not separately observe these mechanisms, their sequence, or their performance consequences.



Note: arrows represent theoretical pathways. The four-case design estimates comparative outcome gaps, not the mechanisms separately.

Figure 1. Disclosure-aware evaluation framework for the realised operating stage of an AI-enabled supply-chain innovation. Named applications are publicly observable inputs; organisational enactment is unmeasured; and accounting outcomes are limited comparative indicators. The pathways are theoretical and are not identified, mediated, or tested by the four-case design. Predictive capability affects accounting outcomes only when information is translated into operating action.

Abbreviation: COGS: Cost of goods sold.

In this evaluation framework, the named AI application is the documented innovation input, organisational enactment is an unobserved implementation layer, and the accounting comparisons are limited outcome indicators. The framework is deliberately asymmetric: it can make the public evidence and the resulting uncertainty auditable, but it cannot recover proprietary model quality, internal decision rules, or an innovation mechanism from annual-report text.

2.5. Disclosed-operating-year timing as a disclosure-aware innovation-evaluation problem

Research on digital adoption frequently extracts event dates from public text. This approach becomes problematic when a report confirms operation during a year but omits both the quarter and whether use began earlier. Filing dates occur after the fiscal period and are determined by disclosure rules, not production rollout. Coding them as treatment introduces a measurement error that is systematic rather than random.

Partial-identification research cautions against replacing interval-valued or coarsened information with an arbitrary point (Manski & Tamer, 2002). This study's evidence is weaker than a known deployment interval because first internal use may precede the disclosed operating year. Therefore, it does not claim a formally identified set. The annual record leaves the within-year coding wholly unspecified. Q1 and Q4 are used as extremal calendar assignments: they maximise the number of fiscal-

year quarters placed on opposite sides of the marker while holding the same report, operating year, donor screen, covariates, and horizon rule fixed. If a substantive interpretation fails to survive this maximal annual-alignment contrast, it is rejected as timing-sensitive. This logic justifies the endpoints as a stringent sensitivity test for annual coding; it does not make them sufficient to recover a plausible deployment window, establish an interior-quarter effect, or resolve earlier use. Prior-report checks and an explicit left-censoring limitation address, but cannot remove, possible earlier use.

This timing discipline complements a case-specific synthetic-control design (Abadie, 2021; Abadie *et al.*, 2010, 2015). Synthetic controls are useful here not because four cases can estimate a population average, but because each focal firm can be paired transparently with a non-extrapolating, same-industry weighted comparison based on pre-assignment outcomes and covariates. The approach makes donor composition, pre-assignment fit, and placebo ranks inspectable, which is preferable to pooling four disclosure cases in a conventional regression. It does not establish that donors are untreated, remove time-varying confounding, create independent replications, or transform disclosure timing into a causal intervention. The resulting quantities are comparative case gaps, not average treatment effects.

Taken together, the registry, evidence distinction, extremal timing assignments, and transparent case comparisons constitute a disclosure-aware evaluation

protocol. Its role is to make the evaluation stage of an operational innovation reproducible at the level of public evidence and research choices. It cannot validate the underlying proprietary AI technology, reconstruct a firm's innovation-development process, or establish that the named system caused an accounting change.

3. Data and research design

3.1. Financial panel and sample construction

The accounting data were acquired from the CSMAR database. The source files cover A-share company identifiers and industry classifications, listing-status changes, quarterly balance sheets, income statements, direct-method cash-flow statements, accounts receivable, and supplier/customer concentration for 2017–2025. The analysis began in 2018Q1 because a beginning-of-quarter balance is required for ratios that use average balance-sheet stocks.

Chinese interim income statements are cumulative within a fiscal year. Therefore, this study recovered stand-alone quarterly flows before constructing outcomes: second-quarter flow equals the half-year cumulative value minus the first quarter; third-quarter flow equals the nine-month cumulative value minus the half-year value; and fourth-quarter flow equals the annual cumulative value minus the nine-month value. Balance-sheet items remained quarter-end stocks. Automated checks verified firm-quarter uniqueness, accounting identities, and the differencing logic. The audited construction produced 93,380 manufacturing firm-quarter observations for 3,732 firms from 2018Q1 through 2025Q4. After excluding special-treatment observations and balance-identity errors, the estimable panel contained 90,529 observations for 3,709 firms.

Manufacturing was defined using CSMAR industry codes beginning with C. Classification was assessed in the disclosed operating year rather than inferred from a firm's historical name. Observations flagged as special treatment (ST or *ST) were removed because distress-related accounting and operating changes can dominate the outcomes. No focal firm was selected because of its subsequent performance.

3.2. Curated event registry and disclosed-operating-year marker

The registry was assembled through targeted Chinese- and English-language searches of listed-company annual reports and public filing indexes using combinations of terms for AI, machine learning, intelligent replenishment, demand forecasting, inventory, supply chain, and planning. Screening was frozen on July 21, 2026. To

enter the registry, a disclosure had to (i) identify the listed entity; (ii) name AI, machine learning, or a clearly identified learning algorithm; (iii) connect that capability to forecasting, replenishment, inventory, or planning; and (iv) describe operational use during the fiscal year. Plans, tenders, generic digitalisation language, patents, and vendor product announcements did not qualify. Where available, the prior annual report was checked to reduce repeated coding of the same named application. Table 1 records the issuer, source report, public report date, stated operating year, named technology, affected supply-chain function, and the specific operational-use rationale for every included case. Because wording and disclosure incentives vary and non-disclosure is unobservable, this protocol produced a curated, verified case series rather than an exhaustive population of adopters.

The source event is the first public operational-use disclosure located within the observation window, not first internal adoption. This distinction is essential. The registry provides evidence that a firm publicly described a named technology as being used in a named supply-chain function during an operating year. It does not provide performance evidence, a production-acceptance date, system coverage, an implementation start date, or a causal attribution. Performance evidence is instead the separately constructed, post-assignment accounting comparison with a synthetic control. Firms can use AI before describing it; disclosure practices vary; and a system can evolve continuously. The analytical marker aligns quarterly outcomes with the operating year described by that disclosure. It does not estimate the effect of the first algorithm ever used by the firm.

Table 1 provides the transparent registry. It separates what the source directly supports (the public document, date, operating year, technology, and named operating function) from the coding judgement (why the language qualifies as operational use). Shangpin's 2019 report described an AI analytical-engine upgrade supporting dynamic sales prediction and automatic replenishment. Midea's 2020 report identified an intelligent supply-chain replenishment algorithm within integrated planning and channel-inventory applications. Winner Medical's 2022 report described machine learning used to optimise inventory and automate replenishment forecasts. Bear Electric's 2024 report described the use of big data and AI to improve demand forecasting and formulate production and logistics plans. Guangdong SACA's 2022 gradient-boosted replenishment system satisfied the functional test but was excluded because its contemporaneous CSMAR industry is retail (F52).

Table 1. Transparent registry of public operational-use evidence and disclosed-operating-year boundary assignments

Firm (ticker)	CSMAR industry	Disclosure source and public date	Disclosed operating year	AI application and affected supply-chain function	Technology nature and basis for coding operational use	Quarterly sensitivity assignments
Shangpin Home Collection (300616)	C21	2019 Annual Report; April 18, 2020 (CNINFO issuer filing)	2019	AI analytics-engine upgrade; regional sales forecasting, procurement direction, delivery coordination, and automatic replenishment.	AI analytical engine using dynamic data, monitoring, simulation, and optimisation. Coded as operational because the report describes the upgraded supply-chain system directing procurement, delivery, and replenishment, rather than a plan or an R&D project.	2019Q1; 2019Q4
Midea Group (000333)	C38	2020 Annual Report; April 30, 2021 (CNINFO issuer filing)	2020	Intelligent supply-chain replenishment algorithm; order fulfilment, replenishment, integrated planning, and channel-inventory management.	Named intelligent replenishment algorithm embedded in operating supply-chain work. Coded as operational because the report attributes order-fulfilment and customer-service effects to its use; it is not a procurement announcement or generic digital strategy.	2020Q1; 2020Q4
Winner Medical (300888)	C17	2022 Annual Report; April 25, 2023 (CNINFO issuer filing)	2022	Machine-learning inventory optimisation and automatic replenishment forecasting; inventory positioning, capacity balancing, and planning.	Algorithm engine and machine-learning AI. Coded as operational because the report states that these tools were used in full-life-cycle merchandise operations and that the Win+ operating platform had been deployed in factories; no go-live quarter is claimed.	2022Q1; 2022Q4
Bear Electric (002959)	C38	2024 Annual Report; April 9, 2025 (CNINFO issuer filing)	2024	Big-data and AI-supported demand forecasting; production and logistics planning.	Big-data and AI decision support. Coded as operational because the report describes use of the technology to improve demand forecasts and formulate more flexible production and logistics plans; it does not establish pilot, production-acceptance, or full-rollout dates.	2024Q1; 2024Q4

Notes: Source dates are the issuer's public report dates. The table records evidence of named operating use during the stated year, not a first-use, pilot, production-acceptance, or full-rollout date. Q1 and Q4 are extremal sensitivity assignments around the disclosed operating year; they test maximal within-year calendar recoding and do not bracket first internal use. The reported technology/function fields are disclosure evidence, whereas the synthetic-control outcomes reported later are performance evidence.

Abbreviations: AI: Artificial intelligence; CNINFO: China Information; CSMAR: China Stock Market and Accounting Research; R&D: research and development.

Regarding classification and system scope, the four disclosures described predictive machine-learning, analytics, or optimisation decision support for forecasting, replenishment, or planning. They were not coded as generative-AI or large-language-model applications. Public reports do not disclose model architecture, training data, hyperparameters, prompt templates, forecast-error benchmarks, or production-monitoring records; the registry therefore does not claim AI-system reproducibility or technical validation.

3.3. Outcomes and matching covariates

Inventory days were calculated using average beginning- and end-of-quarter inventory and stand-alone quarterly

cost of goods sold, scaled by a constant 91.25 days (365/4). The variable was winsorised within quarter at the 1st and 99th percentiles and transformed as $\log(1 + \text{inventory days})$. For interpretation in figures, log gaps were converted to percentage differences using $100[\exp(\text{gap}) - 1]$.

The cash conversion cycle equals inventory days plus days sales outstanding minus days payables outstanding. Receivables and payables used average beginning- and end-of-quarter balances and stand-alone quarterly revenue or cost of goods sold. The cycle was winsorised within quarter at the 1st and 99th percentiles. Complete quarterly inputs were required; as a result, cash-cycle models could not be estimated for Shangpin under either boundary assignment, leaving three cases for that outcome.

The industry-median inventory deviation is the absolute difference between the firm's log inventory days and the inclusive median log inventory days in its two-digit CSMAR industry and quarter. A smaller value indicates a position closer to the contemporaneous industry distribution, not better supply-chain performance. Its theoretical purpose is to avoid a one-dimensional low-inventory norm when buffer stock can protect service and resilience. The measure is not an operational target, an estimate of optimal inventory, or evidence of customer-service, availability, or profitability gains.

The matching feature set combined the full available pre-assignment outcome path with pre-assignment means of log assets, leverage, annualised return on assets, year-on-year sales growth, the eight-quarter rolling demand-volatility proxy, and supply-chain concentration. Continuous controls were winsorised by quarter. Missing pre-assignment covariates were imputed with the matching industry-period median; a covariate entirely unavailable in an early window was omitted rather than filled with later information. The pre-assignment outcome path was given greater weight than the six mean covariates collectively because close outcome fit is central to comparative interrupted-time-series identification. A list of variable definitions is presented in Table 2.

3.4. Case-specific synthetic controls

For focal firm i , accounting outcome k , and candidate boundary quarter G , the synthetic comparison is a convex combination of eligible firms in the same two-digit

manufacturing industry. These firms were not identified as operational-AI cases by the fixed screen, but they were also not verified to be untreated. Donor weights are non-negative and sum to one, thereby avoiding extrapolation outside the observed donor support. Potential donors must have complete outcome data throughout the required pre- and post-assignment window. The algorithm first retains the 30 closest donors in the standardised feature space, then estimates weights by minimising squared pre-assignment feature distance with a small ridge penalty. This produces a transparent comparison path, not a claim of experimental control.

Up to eight observable pre-assignment quarters were used, with a minimum of four. The early specification sets G to Q1 of the operating year and uses eight post-assignment quarters; the late specification sets G to Q4 and uses five. All 22 fitted models shared these assignment-specific horizons. Bear Electric therefore shared the same five-quarter late horizon as the other cases but lacks evidence about adjustment beyond that common window.

To reduce donor contamination, the donor pool excluded all four focal firms and a fixed list of firms with public operational-AI leads that did not meet the strict registry or fell outside the estimable window. These exclusions were made before outcome estimation. Across outcomes and boundary assignments, eligible pools ranged from 178 to 188 firms for Midea and from 228 to 242 for Bear Electric; Winner Medical had 32, and Shangpin's C21 industry had 18. All non-zero weights and pre-assignment diagnostics are reported in the supplement.

Table 2. Variable definitions

Construct	Operational measure	Interpretation
Inventory days	Log[1 + quarterly winsorised days inventory outstanding]	Lower values indicate fewer inventory days, conditional on cost of goods sold
Cash conversion cycle	Winsorised (DIO + DSO – DPO), in days	Lower values indicate faster recovery of operating cash
Industry-median inventory deviation	Absolute deviation of log inventory days from the inclusive industry-quarter median	Lower values indicate closer industry-relative alignment; not an optimum or firm target
Demand volatility	Rolling standard deviation of year-on-year quarterly sales growth over eight quarters (minimum four), compared with the industry-quarter median	Time-varying contextual proxy; not a causal moderator in this four-case design
Size	Log total assets	Pre-assignment matching covariate
Leverage	Total liabilities/Total assets	Pre-assignment matching covariate
Profitability	Annualised quarterly return on assets	Pre-assignment matching covariate
Growth	Year-on-year quarterly sales growth	Pre-assignment matching covariate
Supply-chain concentration	Arithmetic mean of valid customer and supplier concentration percentages	Pre-assignment matching covariate

Abbreviations: DIO: Days inventory outstanding; DPO: Days payables outstanding; DSO: Days sales outstanding.

For each fitted case, the reported estimate is the mean post-assignment comparative gap between the observed outcome and its synthetic counterpart:

$$\tau_{ik}(G) = \text{Mean}_i \geq G \left[Y_{ikt} - \sum_j w_{ijk}(G) Y_{ijk} \right] \quad (1)$$

The early- and late-assignment estimates, $\tau_{ik}(Q1)$ and $\tau_{ik}(Q4)$, are reported together. Their difference reveals how conclusions depend on quarterly alignment with the disclosed year. This study does not interpret the line joining them as a sampling confidence interval or a bracket on first internal use.

3.5. Placebo and pooled inference

Case-level inference used in-space placebos. The closest eligible donor firms were re-estimated as pseudo-focal units at the same boundary quarter, using the same industry restriction and horizon. Placebos with pre-assignment root mean squared prediction error (RMSPE) greater than five times that of the focal case were discarded. The two-sided placebo p -value is the rank of the absolute focal-firm gap within the absolute placebo-gap distribution, with a finite-sample correction.

Pooled summaries averaged the available case gaps only as a descriptive check on whether the four (or, for cash conversion cycle, three) case patterns point in one direction. Randomisation inference drew one placebo gap from each case-specific placebo distribution 20,000

times and compared the absolute observed mean with the simulated distribution of mean placebo gaps. These diagnostics neither create independent observations nor justify a causal population claim. The two-assignment reporting rule is intentionally stringent: a pooled pattern is described as robust only when $p < 0.05$ under both Q1 and Q4. No outcome satisfied that rule—inventories ($p = 0.670/0.904$), industry-median deviation ($p = 0.363/0.386$), and cash conversion cycle ($p = 0.072/0.340$)—and the study treats this as absence of robust pooled evidence, not as a marginal finding.

The exploratory demand-volatility specification and inferential safeguards are described in Section S6. They are retained only to document an exploratory descriptive exercise; repeated firm-quarter observations, a partly post-assignment proxy, and very small cells make them unsuitable for the main empirical design.

Figure 2 summarises how the two evidence streams enter the design. The verified public disclosures determined the four focal operating-year milestones, while the licensed quarterly panel provided outcomes, covariates, and same-industry donor firms. Each case was then estimated under both Q1 and Q4 assignments before case placebos and pooled randomisation diagnostics were calculated. The figure is a workflow, not a claim that the filters recover every operational AI user or that the two assignments locate first deployment.

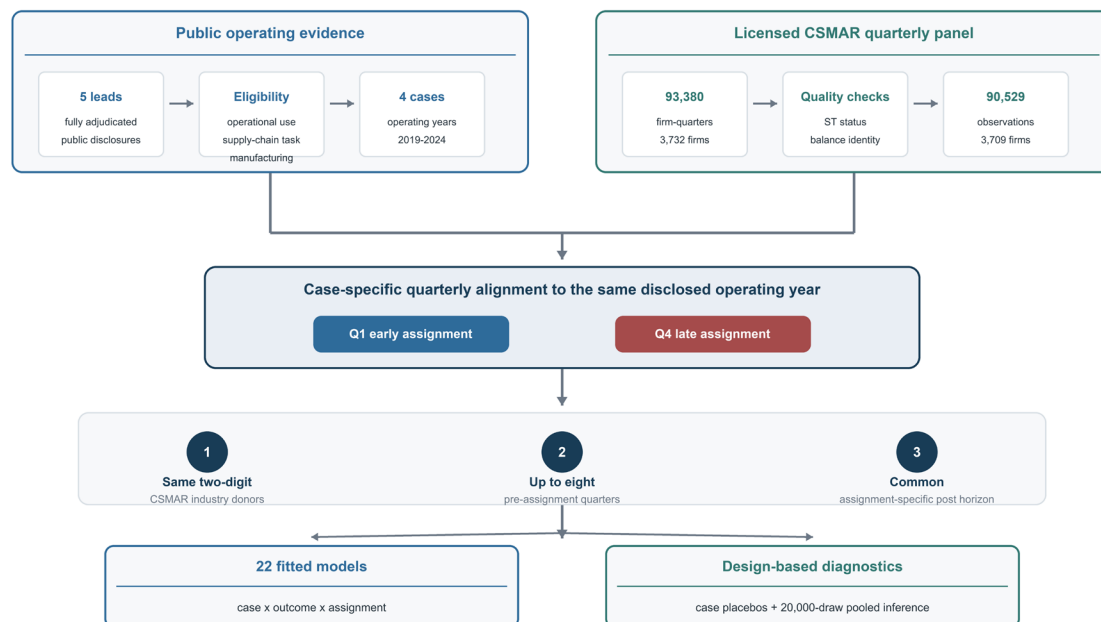


Figure 2. Evidence construction and disclosed-operating-year boundary design. Two evidence streams converge in case-specific synthetic-control comparisons under Q1 and Q4 alignment. Five fully adjudicated public leads yield four eligible manufacturing cases; the quarterly panel yields 90,529 non-special-treatment observations for 3,709 firms after the reported basic quality checks. Q1 and Q4 are sensitivity assignments around the same disclosed operating year, not deployment dates.

3.6. Evidence-strengthening diagnostics

Three pre-specified diagnostic operations were applied to the fitted results. First, the exact unrounded signs of the mean post-assignment inventory-days and industry-median-deviation gaps were combined into four descriptive response modes. Let $d_i(G)$ be the mean log inventory-days gap and $a_i(G)$ the mean industry-median-deviation gap. $d_i(G) < 0$ denotes leaner, $d_i(G) > 0$ denotes buffering, $a_i(G) < 0$ denotes convergence, and $a_i(G) > 0$ denotes divergence. A zero sign is neutral and receives no quadrant. A named case-level mode is reported only if both signs reproduce under Q1 and Q4; a sign change or zero under either assignment is reported as boundary-sensitive or unclassified. This deterministic rule was applied before narrative interpretation. It is not a statistical-significance rule, performance ranking, or measure of managerial intent.

Second, each fitted cash-cycle model was decomposed using its original donor weights. For every post-assignment quarter, the un-winsorised treated-minus-synthetic gaps in DIO, DSO, and DPO were calculated and combined as DIO gap plus DSO gap minus DPO gap. The component sum was checked against the direct un-winsorised cash-cycle gap. Because the main model winsorised the total cycle by quarter, a small difference could arise between the component sum and the reported model effect. The decomposition holds weights fixed and is an accounting attribution, not a mediation analysis.

Third, donor concentration and largest-donor sensitivity were reported. The effective number of donors equals $1/\Sigma w^2$. For each model, the donor with the largest original weight was removed from the selected donor set, the non-negative simplex weights were re-estimated on the same pre-assignment features, and the post-assignment gap was recalculated. This operation tests dependence on one comparison firm; it does not address all specification choices and does not generate a new confirmatory p-value.

3.7. Research-pipeline reproducibility and integrity checks

All data transformations, estimators, and evidence-strengthening diagnostics were scripted. The fixed random seed was 2607009. Integrity checks confirmed 4 included disclosures, 22 fitted case–outcome–assignment comparisons, all 8 required inventory-days, all 8 industry-median-deviation comparisons, and 2 missing cash-cycle comparisons attributable to the complete-quarter requirement. Reconstructed comparative gaps matched the saved results to within 8.33×10^{-17} , and the cash-cycle component identity held to within 5.69×10^{-14} days. No duplicate path keys occurred; all fitted comparisons

showed positive pre-assignment RMSPE and at least one valid placebo. Licensed CSMAR source data cannot be redistributed. Reproducibility, integrity checks, and data-availability arrangements are summarized in Section S7.

3.8. Scope of operational-AI system reproducibility

This study distinguishes two reproducibility objects. The first is the researchers' evaluation pipeline: event-screening rules, source registry, stand-alone-quarter transformations, donor selection, ridge-regularised synthetic-control estimation, placebo calculations, random seed, and integrity checks were documented and provided to the extent permitted by the CSMAR licence. The second is the firms' operational AI systems. Those systems are not reproducible from public annual reports because their architecture, version, training data, feature engineering, optimisation constraints, deployment coverage, human overrides, and monitoring results are proprietary and largely undisclosed.

The absence of firm-level model specifications is a substantive limitation, not a gap that can be filled by reporting prompts. The named applications are predictive and optimisation-oriented decision systems rather than generative-AI or large-language model (LLM) systems. Accordingly, prompt sensitivity, factual hallucination, and stochastic text-output variability are not analytical constructs for the operating technologies evaluated here. Appropriate technical validation would require internal forecast-error comparisons, inventory-policy baselines, service-level records, stockout data, and deployment logs, none of which are present in the public disclosure record.

4. Results

4.1. Pre-assignment fit and case support

The design used four comparative case studies rather than a pooled treatment model. This small number is a central limitation, not a condition remedied by synthetic control. Pre-assignment history ranged from four quarters for Shangpin's Q1 comparisons to eight quarters for most later cases. Positive-weight donor counts ranged from three to nine, even when the eligible pool was considerably larger, reflecting sparse convex solutions. Pre-assignment RMSPE for log inventory days ranged from 0.038 to 0.207 under Q1 and from 0.084 to 0.194 under Q4. The path plots showed a close fit in several, but not all, pre-assignment quarters. In particular, Midea and the later cases retained visible pre-period noise. The resulting case gaps should be read as transparent comparative descriptions alongside RMSPE ratios, placebo ranks, and donor sensitivity, not as precise effect estimates or independent replications.

Figure 3 makes the disclosure-time problem explicit. Each shaded annual band shows the operating year named in the first located public disclosure. Nothing in the source record justifies choosing the annual-report filing date, establishes a common rollout quarter across firms, or rules out earlier internal use. The Q1 and Q4 models therefore asked whether the substantive conclusion survives alternative quarterly alignment with the same disclosed year.

4.2. Inventory days and industry-median alignment

Table 3 reports case-specific mean post-assignment gaps. The inventory-days results are markedly heterogeneous. Shangpin recorded a positive log gap under both assignments (0.211 for Q1 and 0.236 for Q4), equivalent to 23.5% and 26.7% more inventory days, respectively, than

its synthetic control. Neither estimate was unusual in the placebo distribution ($p = 0.429$ and 0.294 , respectively). Because the disclosure concerned an upgrade to an existing intelligent supply-chain platform, the positive gap could reflect expansion, transition, greater buffer protection, or forces unrelated to the upgrade. The design cannot distinguish among these mechanisms.

Midea recorded the largest negative comparative gap. Under the Q1 assignment, its log inventory-days gap was -0.471 , or -37.6% , and the case placebo p -value was 0.032 . Under Q4, the gap was -0.174 , or -16.0% , with $p = 0.355$. The direction survived, but the magnitude and unusualness did not. A substantial decline was observed before the Q4 marker. Aligning the disclosed operating year with Q1 places the decline in the comparison period, whereas aligning it with Q4 places the decline in the matching

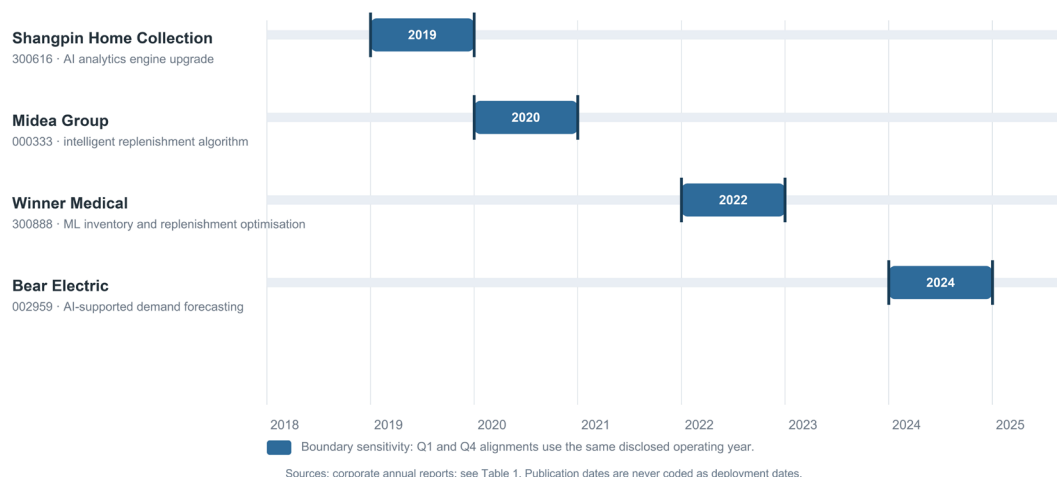


Figure 3. Disclosed-operating-year boundary assignments for the four verified operational AI cases. The shaded Q1–Q4 bands identify the operating years described in the disclosures; they do not bracket first internal use. Sources: Corporate annual reports; see Table 1.

Table 3. Case-specific synthetic-control gaps under disclosed-operating-year boundary assignments

Firm	Assignment	Inventory-days log gap	Inventory-days difference (%)	Industry-median deviation gap	Cash-cycle gap (days)
Shangpin	Q1 early	0.211 (0.429)	23.5	-0.177 (0.235)	–
	Q4 late	0.236 (0.294)	26.7	-0.104 (0.556)	–
Midea	Q1 early	-0.471 (0.032)	-37.6	0.376 (0.032)	-38.14 (0.059)
	Q4 late	-0.174 (0.355)	-16.0	0.055 (0.677)	-34.56 (0.129)
Winner Medical	Q1 early	-0.026 (0.935)	-2.6	-0.272 (0.161)	-35.44 (0.310)
	Q4 late	-0.139 (0.581)	-13.0	-0.057 (0.871)	-34.92 (0.258)
Bear Electric	Q1 early	0.112 (0.548)	11.8	-0.198 (0.097)	-7.37 (0.710)
	Q4 late	0.125 (0.258)	13.3	-0.189 (0.097)	24.19 (0.290)

Notes: Entries are mean post-assignment observed-minus-synthetic gaps. Parentheses contain two-sided case placebo p -values. Inventory percentages are $100[\exp(\log \text{ gap}) - 1]$. Negative industry-median deviation gaps indicate a smaller post-assignment absolute deviation than the synthetic comparison, not necessarily within-firm movement towards an optimum. Q1 and Q4 are boundary assignments, not confidence limits or first-use dates. Cash-cycle models are unavailable for Shangpin because complete quarterly inputs are insufficient.

history. Neither assignment proves when the algorithm first influenced operations, but a single-point code would conceal this sensitivity.

Winner Medical showed modestly lower inventory days under the Q1 assignment (-0.026 , -2.6%) and a larger decline under Q4 (-0.139 , -13.0%). Neither estimate was significant in the case placebos. Bear Electric showed positive gaps under both assignments (0.112 and 0.125 , 11.8% and 13.3% , respectively). Its latest model used the same five-quarter horizon as the other Q4 cases; however, the 2024 disclosure prevented assessment of adjustment beyond that common window.

The industry-median deviation changed the interpretation. The comparative deviation gap was negative for Shangpin under both assignments (-0.177 and -0.104), as it was for Winner Medical (-0.272 and -0.057) and Bear Electric (-0.198 and -0.189). Midea, by contrast, had a positive gap under Q1 (0.376 , $p = 0.032$) and a smaller positive gap, close to neutral, under Q4 (0.055). Thus, Midea's negative inventory-days gap is not evidence of unambiguously superior industry-relative alignment. The pattern could be consistent with deliberate operating leanness, a shifted product mix, or an inventory contraction below the contemporaneous industry norm, but the descriptive benchmark cannot distinguish among these interpretations. It prevents the lower-stock result from being labelled automatically as optimisation.

Figure 4 places the Q1 and Q4 estimates side by side, making both cross-case heterogeneity and assignment sensitivity visible. Figure 5 shows the corresponding

observed and synthetic inventory paths under the Q4 assignment.

Combining the two inventory measures produced a more informative cross-case pattern under the stated sign rule (Figure 6). Both Midea assignments fell in the leaner-divergence quadrant: the exact mean inventory-days gap was below zero while the exact mean deviation gap was above zero. Both Winner Medical assignments showed leaner convergence. Shangpin and Bear Electric showed buffering convergence under both assignments: their inventory-days gaps were above zero while their deviation gaps were below zero. No case fell in buffering divergence. The eight boundary-assignment points therefore formed three reproducible descriptive response modes rather than one efficiency gradient. The labels do not imply that convergence is optimal, that lower inventory improved performance, or that higher inventory was deliberately chosen for resilience.

4.3. Cash conversion cycle

The cash-cycle evidence was more uniformly negative among Midea and Winner Medical, but not robust in aggregate (Figure 7). Midea's mean cycle was 38.1 days below its synthetic control under the Q1 assignment and 34.6 days below under Q4. Winner Medical's corresponding gaps were -35.4 and -34.9 days. The case placebo p -values ranged from 0.059 to 0.310, reflecting both small placebo sets and volatile quarterly working-capital components.

Bear Electric changed sign across the assignments: -7.4 days under Q1 and 24.2 days under Q4. This is precisely

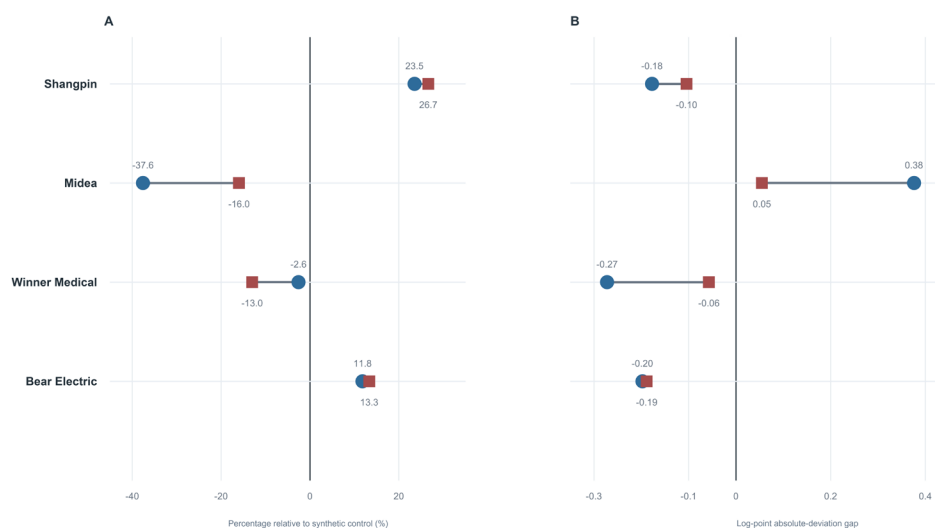


Figure 4. Case-specific comparative gaps under Q1 and Q4 disclosed-operating-year assignments. (A) Inventory days. (B) Industry-median deviation. The gaps are heterogeneous and timing-sensitive. Connected points visualise timing sensitivity and are not confidence intervals. Negative industry-median deviation gaps indicate a smaller absolute deviation than the synthetic comparison. For mean post-assignment gaps, negative values indicate fewer inventory days or smaller median deviation than the synthetic comparison.

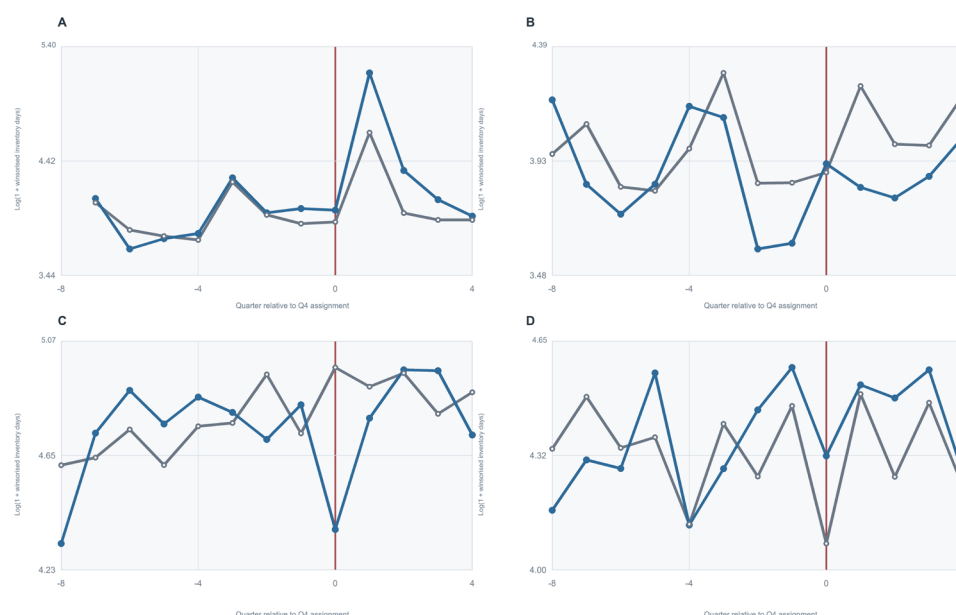


Figure 5. Observed and synthetic log inventory-days paths under the Q4 boundary assignment. (A) Shangpin (300616). (B) Midea (000333). (C) Winner Medical (300888). (D) Bear Electric (002959). The vertical line marks Q4 of the disclosed operating year; it does not identify the first deployment. Pre-assignment fit uses up to eight observable quarters.

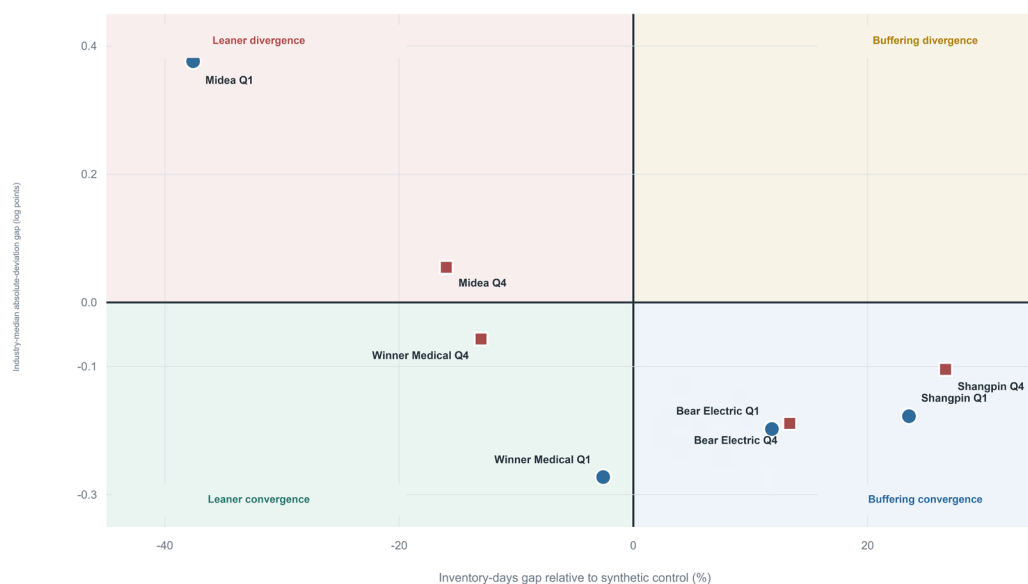


Figure 6. Joint inventory-response typology. Operational artificial intelligence disclosures map to three inventory-response modes. Horizontal position is the percentage inventory-days gap relative to the synthetic comparison; vertical position is the industry-median absolute-deviation gap. Quadrants describe comparative response modes and are not causal mechanism estimates or performance rankings.

the kind of result for which a single imputed quarter would be misleading. The Q1 alignment placed more of 2024 in the comparison period, suggesting a modestly shorter cycle; Q4 placed those quarters in the matching history, suggesting a longer cycle over the common five-quarter follow-up. Neither estimate was significant, and neither identified actual go-live timing.

Across the three cash-cycle cases, the Q1 mean was -26.99 days with a randomisation p -value of 0.072 and the Q4 mean was -15.10 days with $p = 0.340$. Neither value satisfies the study's two-assignment 0.05 rule, and the Q1 value is not interpreted as a marginal performance finding. The three-case average was largely shaped by Midea and Winner Medical, while the outcome also incorporated

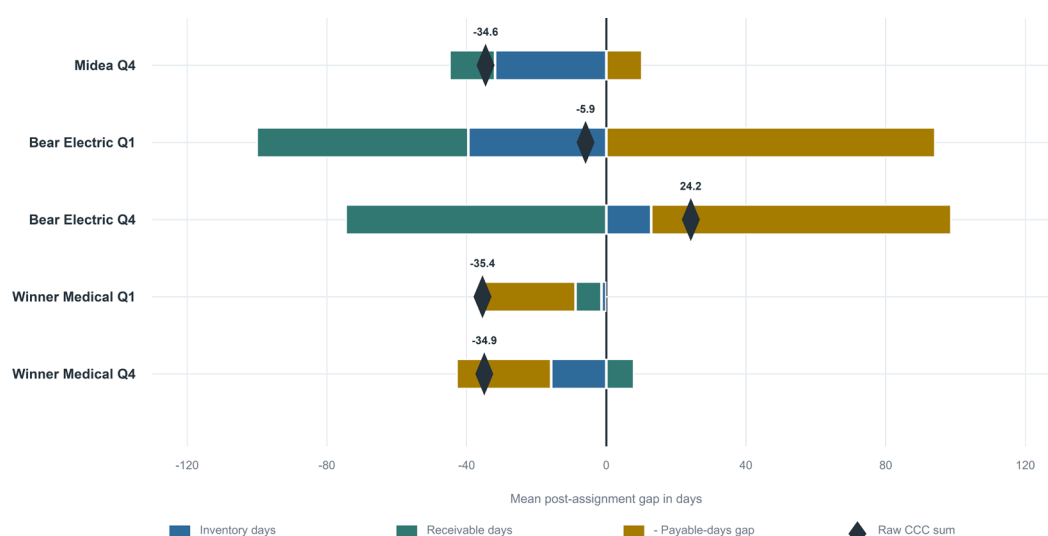


Figure 7. Additive cash-conversion-cycle (CCC) decomposition using each original CCC model's donor weights and un-winsorised accounting components (inventory, receivable, and payable days). Positive signed components lengthen the cycle, and negative components shorten it. The payable component is plotted with its CCC sign. The decomposition attributes the accounting gap; it does not identify mediation.

receivables and payables beyond the named supply-chain function. It provides no robust pooled evidence that operational AI improves working-capital performance.

The component decomposition shows that similar cash-cycle totals can contain different accounting components. Midea's raw Q1 gap of -38.1 days combined -40.9 inventory days, -23.2 receivable days, and a 26.0 -day payable component; its Q4 gap of -34.6 days remained primarily inventory-led. Winner Medical's Q1 gap of -35.4 days was dominated by the payable component (-26.6 days), whereas its Q4 gap combined -15.8 inventory days, 8.0 receivable days, and -27.1 payable-component days. Bear Electric contained the strongest offsets: under Q1, -39.5 inventory and -60.6 receivable days were almost offset by a 94.2 -day payable component, leaving a raw cycle gap of -5.9 days; under Q4, 12.9 inventory and 86.0 payable-component days outweighed -74.6 receivable days, leaving 24.2 days. The original model's Q1 total comparative gap was -7.4 days because the reported total cycle was winsorised within the quarter. This

decomposition attributes accounting components; it does not establish a mechanism or a performance gain.

4.4. Pooled estimates and donor sensitivity

Table 4 summarises the pooled descriptive diagnostics and provides the clearest limit on inference. The mean log inventory gap was -0.044 under Q1 ($p = 0.670$) and 0.012 under Q4 ($p = 0.904$); the sign changed, and neither result was unusual relative to the placebo distribution. The industry-median deviation gap was -0.068 under Q1 ($p = 0.363$) and -0.074 under Q4 ($p = 0.386$), again not unusual. The cash-cycle mean was -26.99 days under Q1 ($p = 0.072$) and -15.10 days under Q4 ($p = 0.340$). Thus, no pooled result was below 0.05 under both boundary assignments, and the study does not claim that operational AI improves inventory efficiency or working-capital performance.

Leave-one-case-out results reinforce the absence of a common pattern. Removing Midea changed the Q1 mean inventory-days gap from negative to positive. Removing

Table 4. Pooled comparative gaps and randomisation inference

Outcome	Cases	Q1 mean gap	Two-sided p	Q4 mean gap	Two-sided p
Log inventory days	4	-0.044	0.670	0.012	0.904
Industry-median inventory deviation	4	-0.068	0.363	-0.074	0.386
Cash conversion cycle (days)	3	-26.99	0.072	-15.10	0.340

Notes: Pooled p -values compare the observed mean with 20,000 means drawn from the case-specific placebo distributions. They are diagnostic randomisation-inference measures, not evidence of random assignment or a causal population effect.

Bear Electric made the cash-cycle average more negative. The Q1 industry-median-deviation average among the two later-disclosure cases was unusual in its pooled placebo distribution, but its Q4 counterpart was smaller and not unusual. Given the two-case subset, assignment sensitivity and post hoc nature of this contrast, it is not presented as confirmatory evidence. No pooled outcome met the stated robustness criterion of $p < 0.05$ under both boundary assignments.

The largest-donor diagnostic addresses a different fragility. Convex weights were concentrated: the median effective donor count was 4.39 for inventory days, 3.16 for industry-median deviation, and 3.08 for the cash cycle, while the median largest weight was 0.332, 0.496, and 0.427, respectively. After removing that largest-weight donor and re-estimating weights, the direction was preserved in all eight inventory models, all eight industry-deviation models, and five of six cash-cycle models. The only reversal was Bear Electric's Q1 cash-cycle estimate, which changed from -7.37 days to 0.45 days. Median pre-RMSPE ratios after removal were 0.94, 1.03, and 1.14 across the three outcomes. Thus, the inventory response types are not artefacts of a single dominant donor, but one already weak cash-cycle result is donor-sensitive. Case-level results are reported in Section S3, and the largest-donor sensitivity summary is reported in Section S4.

4.5. Scope statement on exploratory demand-volatility descriptives

The exploratory demand-volatility specification and inferential safeguards are described in Section S6. It used repeated focal-firm quarters, a rolling proxy that could include post-assignment observations, and cells with as few as one contributing cash-cycle case. These features prevent an interaction estimate, independent inference, or meaningful comparison across volatility groups. Therefore, they were not used to support a volatility claim, an AI-performance claim, or a contribution beyond transparent reporting of the exploratory analysis.

5. Discussion

5.1. Interpretation of the heterogeneous results

The central result is not that AI failed or succeeded. Four public operating-use disclosures did not yield a common comparative accounting trajectory, and the pooled diagnostics provided no robust evidence of an inventory-efficiency or working-capital improvement. The descriptive response labels indicate that Midea is leaner and farther from the industry benchmark, Winner Medical is leaner and closer, and Shangpin and Bear Electric carry more inventory while remaining closer to the benchmark

than their synthetic comparisons. These patterns are compatible with many unmeasured organisational and market processes. Because data quality, integration, decision rights, planner trust, coverage, and rollout stage were not measured, the cases cannot determine whether organisational complements account for the observed gaps.

The Midea case illustrates the danger of one-dimensional efficiency claims. Its large inventory-days reduction is the strongest case-level finding, yet the Q1 industry-median deviation widens. If lower inventory is the only criterion, this appears unequivocally beneficial. If industry-relative alignment is also considered, the interpretation becomes ambiguous. An efficient supply chain minimises neither inventory nor cash cycle in isolation; it balances capital, service, continuity, and strategic flexibility.

Winner Medical and Bear Electric provide a different signal. Both have negative industry-median comparative deviation gaps under both boundary assignments, even though their inventory-days gaps differ. The consistency of direction is managerially interesting but statistically weak. It may be consistent with forecasting used to adjust a firm's position relative to sector conditions rather than to pursue universal leanness, but the design does not identify that mechanism. Such an interpretation accords with machine learning as a decision aid: the model can support a context-dependent policy rather than a fixed low-inventory rule (Priore *et al.*, 2019).

The cash-cycle decomposition adds a second qualification. Midea's negative gap is largely inventory-led, whereas Winner Medical's is more strongly associated with the signed payable component; Bear Electric shows substantial offsets across all three components. A shorter cash-cycle accounting gap should therefore not be attributed automatically to forecasting accuracy, even when an inventory application is named in the disclosure. Consolidated statements combine operating, commercial, and financing choices. The 21-of-22 sign-preservation result provides a limited donor-sensitivity check, while the Bear Electric Q1 reversal also shows why the cash-cycle descriptions remain conditional.

5.2. Contribution hierarchy: operational-AI measurement and descriptive outcomes

The substantive and measurement contributions concern what is counted as AI. Much firm-level research uses workforce profiles, patents, textual intensity, or broad capability surveys. Those approaches illuminate diffusion and average associations, but they need not identify an application with a direct operating connection to a supply-chain decision. This study instead requires public evidence

that AI or machine learning is used in a named forecasting, replenishment, inventory, or planning function. This task specificity improves construct clarity but limits the number of verifiable cases. It is not a substitute for large-sample adoption studies and does not by itself deliver a performance estimate.

For systematic innovation, the additional contribution is methodological rather than algorithmic. The framework operationalises the evaluation stage of a realised AI-enabled process innovation: it asks what the public evidence establishes about the operational application, separates that evidence from accounting outcomes, tests the consequence of coarsened timing, and states the limits of the resulting comparison. It does not claim to be a generative-assistant method, a classification-model contribution, a TRIZ procedure, or a technical benchmark of the firms' proprietary AI systems.

The result contribution is correspondingly descriptive. A replenishment algorithm has a more direct conceptual connection to inventory outcomes than a generic AI score, yet lower inventory is not a complete definition of value. The joint response labels separate inventory scale from industry-relative alignment, while the cash-cycle decomposition separates inventory, collection, and supplier-financing components. These distinctions make heterogeneous accounting patterns visible; they do not establish an efficiency ranking, causal transmission, or average effect. The four-case registry is therefore best read as a high-construct-validity comparative case series with limited generalisability.

The theory contribution is an interpretive boundary, not a mechanism test. The complementarity account explains why a technology-only narrative would be implausible: organisational history, system integration, data quality, managerial response, and buffer strategy could all condition how a predictive tool is used. The present data do not observe these constructs, and the dispersed case gaps cannot test whether any of them explains a particular pattern. The productivity J-curve likewise motivates caution about transition periods but is not evaluated here. The common eight-quarter Q1 and five-quarter Q4 horizons may capture transition rather than steady-state accounting patterns, and Bear Electric cannot reveal adjustment beyond that window.

5.3. Contribution to disclosure-based identification

The more general methodological contribution concerns time. Corporate reports are attractive sources because they are public, firm-specific, and relatively standardised. Their temporal semantics are easily misread. A report published in April 2023 that stated a machine-learning replenishment

system operated in 2022 identifies an operating year, not an April 2023 intervention. Using publication as deployment can misclassify before-and-after periods and create artificial lags.

The disclosed-operating-year boundary design offers a transparent annual-alignment sensitivity test. It does not solve left censoring or selective disclosure and does not claim that Q1 and Q4 bracket first internal use. Its narrower contribution is to test whether an inference rests on one arbitrary quarterly coding when the source provides only an operating year. Because Q1 and Q4 are the maximal calendar recodings compatible with that annual evidence, agreement across them supports only robustness to this specific annual-alignment choice; disagreement is reported as timing sensitivity. Midea's inventory magnitude and Bear Electric's cash-cycle sign demonstrate that this is not a cosmetic adjustment. The right response to disagreement is not to select the more favourable result *ex post*, but to limit the claim and seek direct deployment records.

This lesson is relevant beyond China and beyond AI. Annual environmental, social, and governance reports; cybersecurity disclosures; and digital-transformation narratives often identify a period of activity without a production date. Researchers should distinguish announcement, procurement, pilot, acceptance, operational use, and full rollout. Where exact dates are absent, boundary sensitivity, qualitative event histories, or additional primary evidence are preferable to false precision.

5.4. Asia-Pacific business implications

The Chinese setting sharpens the managerial implications. Firms in the region combine rapid technology investment with heterogeneous organisational systems, supplier structures, and international exposure. Asia-Pacific evidence shows that supply-chain analytics operates through agility and information sharing, while disruption orientation builds resilience through learning and resource reconfiguration (Jiang *et al.*, 2026; Wang & Ali, 2023). The value of AI forecasting should therefore be assessed through operating architecture rather than procurement alone. Managers should specify which decisions the model changes, who owns those decisions, how overrides are recorded, and whether performance is evaluated against service and resilience objectives as well as inventory reduction.

Three actions follow. First, firms should maintain auditable rollout records: pilot date, production acceptance, plants and product categories covered, model version, and downstream decision rule. Such records improve internal governance and make external claims

credible. Second, dashboards should separate inventory scale from industry-relative and firm-specific benchmarks. A lower stock position may be efficient in stable conditions and hazardous when lead times or disruption risk rise. Third, AI investment should be accompanied by master-data governance, interface integration, and training. The evidence does not imply that organisational complements guarantee gains; it implies that technology without those complements is an incomplete intervention.

For policymakers and investors, disclosure quality is part of the digital business ecosystem. Current annual narratives often combine aspirations, infrastructure, and operational outcomes. More structured voluntary disclosure of production deployment and coverage would allow capital markets to distinguish substantive operational capability from generic AI rhetoric. Standardisation need not reveal proprietary model details. Dates, process scope, human oversight, and validated outcome categories would materially improve comparability.

5.5. Limitations and future research

The paper's central limitation is its size: it contains four verified operating-use disclosures and only three cash-conversion-cycle cases. These observations cannot identify an average treatment effect, estimate population heterogeneity, or support a general claim that operational AI improves inventory efficiency or working-capital performance. The pooled diagnostics make this limit concrete: inventory days ($p=0.670/0.904$), industry-median deviation ($p=0.363/0.386$), and cash conversion cycle ($p=0.072/0.340$) under Q1/Q4. Focal firms were selected based on public verifiability, not random assignment, and firms that disclose named applications may differ systematically in scale, governance, or communication. The targeted search may also miss applications worded differently or not disclosed.

A further boundary concerns the type of innovation contribution. The study does not observe or develop an innovation-generation method, TRIZ workflow, opportunity-identification algorithm, LLM prompt process, or proprietary predictive model. Its systematic-innovation relevance is limited to a transparent ex-post evaluation protocol for a realised operational innovation. Claims about algorithmic novelty, forecasting superiority, hallucination control, or prompt robustness would require technical artefacts and internal validation data that are outside this design.

Second, the marker is a disclosure-linked operating year, not a known deployment interval. First internal use may precede the located report, and an application can move gradually from pilot to partial and then full rollout. Prior-

report checks reduce but do not eliminate left censoring. Q1 and Q4 assignments provide a deliberately stringent test of the maximum within-year reclassification implied by one annual marker; they do not bound unknown multi-year implementation histories, infer Q2/Q3 results, or recover a plausible deployment window. Direct acceptance records, enterprise-system logs, or firm confirmation would permit a more precise design.

Third, consolidated financial statements are coarse measures of process change. Inventory days and the cash conversion cycle are influenced by product mix, acquisitions, pricing, supply shocks, and financial policy. The inclusive industry-quarter median is only a descriptive benchmark and includes the focal firm; it is not a firm-specific optimum. Future research should use monthly or weekly SKU data, forecast errors, stockouts, supplier on-time-in-full performance, shortage downtime, emergency freight, and alert-response logs. Such data would reveal whether an algorithm improves sensing, decision speed, or realised continuity.

Fourth, the analysis deliberately uses common assignment-specific horizons: eight quarters after Q1 and five after Q4 for every fitted case. This improves comparability, but sacrifices longer follow-up available for early disclosures, and Bear Electric cannot reveal adjustment beyond the common five-quarter Q4 window. A longer panel would separate transition from steady-state patterns and allow common longer-horizon robustness. It would also permit explicit analysis of shocks after verified deployment, provided route, customer, or input exposure is measured rather than assumed.

Fifth, the exploratory demand-volatility specification and inferential safeguards are described in **Section S6**. Its time-varying sales-growth proxy can include post-assignment observations; the focal firms contribute repeated quarters; and some cells contain only one case. It is therefore not used for a moderation claim, a main-text result, or theoretical confirmation. Larger verified cohorts and genuinely pre-treatment environmental measures would be required for risk-set designs or formal interactions.

Finally, synthetic-control choices matter. Same-industry restriction improves comparability but can reduce support, especially for Shangpin. The convex-hull constraint prevents extrapolation but may leave imperfect fit. Removing the largest-weight donor preserves 21 of 22 directions, but this is not a complete specification curve: it does not vary industry definitions, ridge penalties, matching covariates, windows, or estimator class. The accounting decomposition likewise attributes cash-cycle gaps to statement components without identifying mediation.

Future work could compare augmented synthetic controls, matrix completion and Bayesian structural time-series models under a pre-registered specification.

6. Conclusion

Algorithms operate in supply chains through organisational doors. They are trained on firm data, inserted into planning systems, and acted upon by people whose objectives include cost, service, continuity, and growth. It should therefore be unsurprising that the same broad technology does not yield a single operating outcome.

Across four publicly verified Chinese manufacturing cases, named AI-enabled forecasting and replenishment disclosures coincide with heterogeneous comparative accounting gaps. Midea falls in the leaner-divergence label, Winner Medical in leaner convergence, and Shangpin and Bear Electric in buffering convergence under both boundary assignments. The cash-cycle decomposition shows that similar aggregate gaps can have different inventory, receivable, and payable components, and the largest-donor check preserves direction in 21 of 22 comparisons. These descriptive features do not overturn the main statistical limitation: no pooled diagnostic is below 0.05 under both disclosed-operating-year assignments (inventory days [$p = 0.670/0.904$]; industry-median deviation [$p = 0.363/0.386$]; cash conversion cycle [$p = 0.072/0.340$]). The study therefore does not show that operational AI improves inventory efficiency or working-capital performance.

The principal contribution is methodological rather than a claimed performance effect. When public evidence identifies an operating year but not a go-live quarter—or whether use began earlier—researchers should treat time as coarsened information, not as a precise intervention date. Boundary sensitivity makes the inferential cost of arbitrary calendar coding visible. The substantive and measurement contributions are to focus on named operational AI tasks and to distinguish inventory scale from industry-relative alignment. The theoretical complementarity account sets an interpretive boundary, but its mechanisms remain unmeasured. This produces a deliberately modest comparative case study rather than a causal or general evaluation of AI performance.

Viewed through a systematic-innovation lens, the contribution is a disciplined evaluation method for the realised operating stage of an AI-enabled supply-chain innovation. It combines task-specific evidence, disclosure-aware temporal coding, transparent comparative diagnostics, and explicit boundaries between operational-use evidence, accounting evidence, and proprietary AI-system evidence. It does not claim to generate

inventions, extend TRIZ, reproduce the firms' algorithms, or validate generative-AI behaviour. These limits are essential to the method: they prevent public corporate narratives from being mistaken for either technical model documentation or causal proof of innovation value.

Acknowledgments

None.

Funding

This research received no external funding.

Conflict of interest

The author declares no competing interests.

Author contributions

This is a single-authored article.

Ethics approval and consent to participate

Not applicable. The study used publicly available corporate disclosures and licensed secondary financial data and involved no human participants.

Consent for publication

Not applicable.

Availability of data

The licensed CSMAR data used in this study cannot be redistributed. Reproducibility, integrity checks, and data-availability arrangements are summarized in Section S7.

Further disclosure

During the preparation of this manuscript, the author(s) used AI and AI-assisted technologies solely to improve the readability and language of the manuscript, and not to complete key researcher work such as producing scientific insights, analysing or interpreting data, or drawing scientific conclusions. The author has reviewed and edited the output and takes full responsibility for the content of this publication.

References

- 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>
- Abadie, A., Diamond, A., & 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.

- <https://doi.org/10.1198/jasa.2009.ap08746>
- Abadie, A., Diamond, A., & Hainmueller, J. (2015). Comparative politics and the synthetic control method. *American Journal of Political Science*, 59(2), 495-510.
- <https://doi.org/10.1111/ajps.12116>
- Babina, T., Fedyk, A., He, A., & Hodson, J. (2024). Artificial intelligence, firm growth, and product innovation. *Journal of Financial Economics*, 151, Article 103745.
- <https://doi.org/10.1016/j.jfineco.2023.103745>
- Baryannis, G., Dani, S., & Antoniou, G. (2019). Predicting supply chain risks using machine learning: The trade-off between performance and interpretability. *Future Generation Computer Systems*, 101, 993-1004.
- <https://doi.org/10.1016/j.future.2019.07.059>
- Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply chain risk management and artificial intelligence: State of the art and future research directions. *International Journal of Production Research*, 57(7), 2179-2202.
- <https://doi.org/10.1080/00207543.2018.1530476>
- Belhadi, A., Mani, V., Kamble, S. S., Khan, S. A. R., & Verma, S. (2024). Artificial intelligence-driven innovation for enhancing supply chain resilience and performance under the effect of supply chain dynamism: An empirical investigation. *Annals of Operations Research*, 333(2-3), 627-652.
- <https://doi.org/10.1007/s10479-021-03956-x>
- Bresnahan, T. F., Brynjolfsson, E., & Hitt, L. M. (2002). Information technology, workplace organization, and the demand for skilled labor: Firm-level evidence. *The Quarterly Journal of Economics*, 117(1), 339-376.
- <https://doi.org/10.1162/003355302753399526>
- Brynjolfsson, E., & Hitt, L. M. (2000). Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic Perspectives*, 14(4), 23-48.
- <https://doi.org/10.1257/jep.14.4.23>
- Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5), 133-139.
- <https://doi.org/10.1257/aer.p20161016>
- Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The productivity J-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1), 333-372.
- <https://doi.org/10.1257/mac.20180386>
- Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, 184(3), 1140-1154.
- <https://doi.org/10.1016/j.ejor.2006.12.004>
- Deloof, M. (2003). Does working capital management affect profitability of Belgian firms? *Journal of Business Finance & Accounting*, 30(3-4), 573-588.
- <https://doi.org/10.1111/1468-5957.00008>
- Dubey, R., Gunasekaran, A., Childe, S. J., Bryde, D. J., Giannakis, M., Foropon, C., Roubaud, D., & Hazen, B. T. (2020). Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations. *International Journal of Production Economics*, 226, Article 107599.
- <https://doi.org/10.1016/j.ijpe.2019.107599>
- Gaur, V., Fisher, M. L., & Raman, A. (2005). An econometric analysis of inventory turnover performance in retail services. *Management Science*, 51(2), 181-194.
- <https://doi.org/10.1287/mnsc.1040.0298>
- Hazen, B. T., Boone, C. A., Ezell, J. D., & Jones-Farmer, L. A. (2014). Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications. *International Journal of Production Economics*, 154, 72-80.
- <https://doi.org/10.1016/j.ijpe.2014.04.018>
- Helo, P., & Hao, Y. (2022). Artificial intelligence in operations management and supply chain management: An exploratory case study. *Production Planning & Control*, 33(16), 1573-1590.
- <https://doi.org/10.1080/09537287.2021.1882690>
- Jiang, Q., Wang, N., Jiang, B., & Bu, F. (2026). Relationship between supply chain disruption orientation and firm's resilience in China: The dynamic capabilities perspective. *Asia Pacific Business Review*, 32(1), 62-89.
- <https://doi.org/10.1080/13602381.2024.2352602>
- Kieschnick, R., Laplante, M., & Moussawi, R. (2013). Working capital management and shareholders' wealth. *Review of Finance*, 17(5), 1827-1852.
- <https://doi.org/10.1093/rof/rfs043>
- Li, W., Zhu, X., & Feng, Z. (2025). How digital transformation can reshape competitive advantages in Chinese manufacturing firms: The mediating effects of organizational capabilities. *Asia Pacific Business Review*. Advance online publication.
- <https://doi.org/10.1080/13602381.2025.2516841>
- Manski, C. F., & Tamer, E. (2002). Inference on regressions with interval data on a regressor or outcome. *Econometrica*, 70(2), 519-546.
- <https://doi.org/10.1111/1468-0262.00294>
- Priore, P., Ponte, B., Rosillo, R., & de la Fuente, D. (2019).

Applying machine learning to the dynamic selection of replenishment policies in fast-changing supply chain environments. *International Journal of Production Research*, 57(11), 3663-3677.

<https://doi.org/10.1080/00207543.2018.1552369>

Rumyantsev, S., & Netessine, S. (2007). What can be learned from classical inventory models? A cross-industry exploratory investigation. *Manufacturing & Service Operations*

Management, 9(4), 409-429.

<https://doi.org/10.1287/msom.1070.0166>

Wang, Y., & Ali, Z. (2023). Exploring big data use to predict supply chain effectiveness in Chinese organizations: A moderated mediated model link. *Asia Pacific Business Review*, 29(3), 632-653.

<https://doi.org/10.1080/13602381.2021.1920704>