Staged Regional Rollouts under Retail Execution Frictions in Modular Product Programs across Heterogeneous Consumer Category Markets

Mohamed Ould Salem[1] AND Ahmed Lemine Cheikh[2]

[1] Department of Management Information Systems, Faculty of Economics and Management, Université de Nouakchott Al Aasriya, Avenue de l'Indépendance, Tevragh-Zeina, 1122 Nouakchott, Mauritania
[2] Department of Business Analytics and Digital Innovation, Institut Supérieur d'Enseignement Technologique de Rosso, Route de Keur Macène, BP 245, Rosso, Mauritania

ABSTRACT

Product strategy is often evaluated through what firms introduce, how many variants they offer, and how quickly they scale new items across channels. Less attention has been given to the timing architecture of expansion itself. In practice, many firms do not choose between a local launch and a full national launch in a single step. They deploy product programs through staged regional rollouts, allowing early markets to reveal demand quality, retailer compliance, cannibalization, and service burdens before the remaining territories are activated. That sequencing choice is strategic because it determines how much learning is obtained before irreversible distribution commitments are made. This study investigates whether staged rollout structures improve economic outcomes and identifies the conditions under which gradual scaling dominates immediate broad release. The analysis uses an archival-style panel of 52,416 product-program–market–month observations drawn from 546 product programs introduced by 63 manufacturers in 37 consumer categories across 96 regional markets over nine years. The empirical design combines event-time estimation, instrumental-variables panel regression, market-level hazard models of rollout acceleration, and mediation tests linking pilot-market learning to later contribution margins. The results indicate that rollout breadth exhibits a nonlinear relationship with first-year program profitability. Entering too few markets delays scale economies and weakens retailer support, but entering too many markets before learning is assimilated lowers contribution through forecast error, markdown exposure, and within-brand cannibalization. The most profitable sequence is characterized by a moderate first-wave footprint, a short but nonzero learning interval, and rapid second-wave expansion only when pilot markets are representative and product modularity is high. Retail execution frictions compress the optimal breadth, while strong platform modularity and brand capital enlarge it. The findings position rollout sequencing as a central dimension of product strategy rather than a logistical afterthought.

I. Introduction

Product strategy is frequently narrated as a choice over product concepts, price positions, target segments, and degrees of differentiation. That characterization is incomplete in categories where new offers do not appear everywhere at once [1]. Firms regularly determine not only what will be sold, but also where the product will be made available first, how long they will wait before expanding distribution, how much sales data from the first wave will be considered credible, and whether the pilot evidence is informative enough to support scale-up. These decisions shape realized profitability because the early months of a program combine high uncertainty with unusually consequential commitments. A firm that scales too slowly may miss category momentum, forfeit retailer attention, and leave advertising underutilized. A firm that scales too quickly may convert uncertain demand signals into inventory imbalances, support dilution, and avoidable markdowns across a wide territory. The rollout path therefore sits inside product strategy itself.

The issue has become more important as product programs have grown more modular and more channel dependent. Manufacturers increasingly build families of offerings around shared formulations, shared components, or shared package architectures, then tailor the final market expression

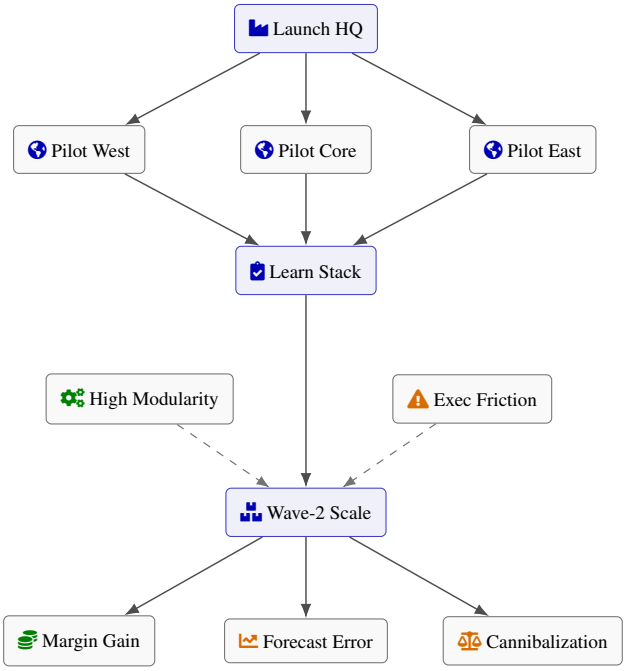

FIGURE 1: Overall staged rollout architecture linking pilot regions, accumulated pilot learning, and second-wave expansion. The lower layer separates economic upside from the two main scaling risks observed after broader activation.

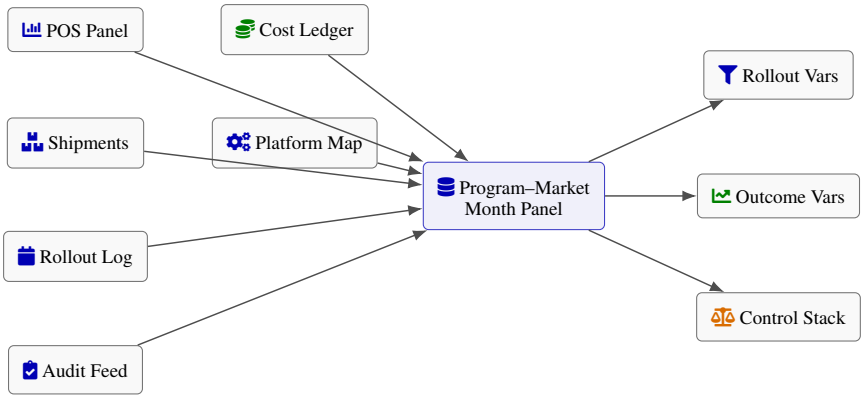

FIGURE 2: Data architecture for the empirical panel. Multiple operational sources are fused at the program–market–month level and then partitioned into rollout treatments, economic outcomes, and structural controls used in the estimation.

through sizes, claims, accessory bundles, or retailer-specific merchandising formats [2]. Such modularity can lower development cost and speed adaptation, but it also creates temptations to scale prematurely because the apparent cost of opening another market declines. Retail systems intensify the same temptation. If a category reset window or promotional calendar makes nationwide placement feasible, managers may interpret feasibility as desirability. Yet the economic question is not whether the firm can distribute widely. The relevant question is whether it should do so before sufficient evidence about demand composition, cannibalization, replenishment noise, and retailer execution has been observed.

This paper studies staged regional rollout as a product-strategy problem with three interlocking margins [3]. The first margin concerns launch breadth, defined as the share of targeted markets activated in the first wave [4]. The second concerns learning interval, defined as the time between first-wave activation and second-wave expansion. The third concerns pilot-market representativeness, defined as the degree to which the first-wave markets resemble the remaining target territory in demographics, channel structure, promotional intensity, and prior category usage. These margins jointly determine how much the firm learns before it commits broader inventory and trade-spend resources, and they also determine whether the learning it obtains is transferable. A narrow first wave can produce clean signals yet provide too little scale and weak retailer enthusiasm. A broad first wave can achieve visibility and operating leverage

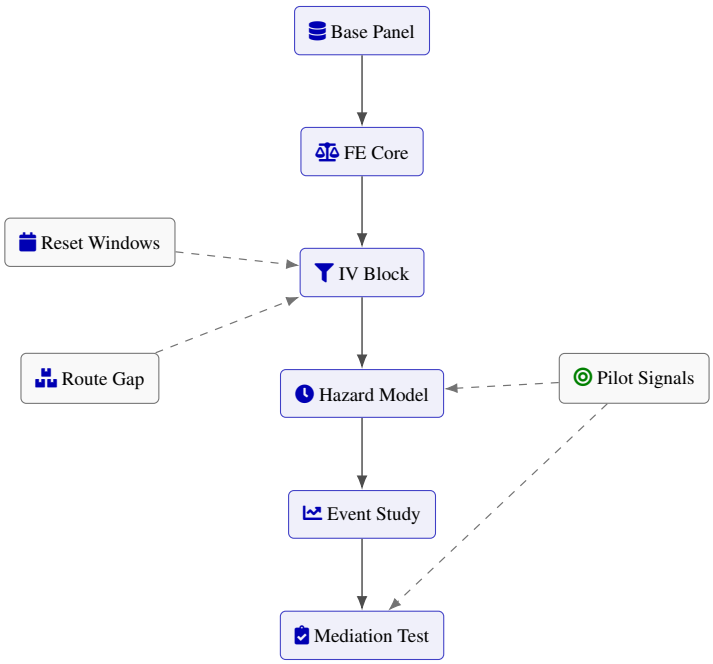

FIGURE 3: Estimation stack used to study the rollout sequence. The design moves from fixed-effects structure to instrumented breadth and interval estimation, then to the timing of second-wave activation and the learning-based mediation layer.

yet swamp the planning system with unfiltered noise. A long learning interval can allow better diagnosis yet leave the product under-distributed when competitors react. A short interval can protect momentum but convert incomplete evidence into irreversible commitments. Representativeness determines whether even high-quality pilot results are useful beyond the pilot itself.

The central claim of the manuscript is that rollout sequencing has an interior optimum rather than a monotone relationship with performance. Immediate national expansion is not always best, but neither is a highly conservative regional crawl. The most productive sequence depends on the quality of early signals and on the firm's ability to translate them into better later decisions. Product modularity shifts this balance because it affects how easily the firm can alter pack configuration, inventory allocation, or retailer-specific mix without redesigning the product platform. Retail execution friction shifts it as well because wide early distribution is much more expensive when compliance with displays, facings, and assortment agreements is noisy across markets. Under those conditions, learning has economic value beyond informational curiosity [5]. It directly improves the allocation of the remaining rollout.

The rationale for focusing on rollout path rather than on product attributes alone is rooted in a broader regularity of product strategy. The economic consequences of product decisions are often nonlinear, especially when a strategic lever that looks favorable in partial equilibrium creates coordination costs once it is scaled. In a different product-related context, Cao and Yan (2021) reported a curvilinear association between product quality intensity and firm profit and further showed that supporting investments can shift that curve [6]. That result is useful here not because the underlying context is the same, but because it underscores a more general point: product strategy variables frequently exhibit interior optima once market benefits and implementation burdens are considered together [7]. Rollout breadth should therefore not be presumed linear merely because broader coverage raises availability in a static sense.

The present study approaches the question through a longitudinal panel of product programs launched across multiple regional markets over time. The empirical perspective is deliberately program-centered rather than SKU-centered because rollout sequencing is typically chosen at the program level, even when final execution occurs through multiple stock-keeping units. Program-level analysis allows the estimation to capture learning, migration, and support reallocation within a coordinated launch architecture. The empirical design follows markets before and after program introduction, records whether later waves occur and when, and links those decisions to revenue, contribution margin, forecast error, markdown intensity, retailer compliance, and adjacent-product cannibalization [8]. This combination allows the analysis to measure whether staged expansion improves performance merely by delaying risk or by actively altering what the firm learns and how it behaves later.

Several distinctions separate this manuscript from the earlier product-strategy texts already drafted in this conversation. The first earlier manuscript centered on breadth, attribute depth, assurance, and governance in omnichannel

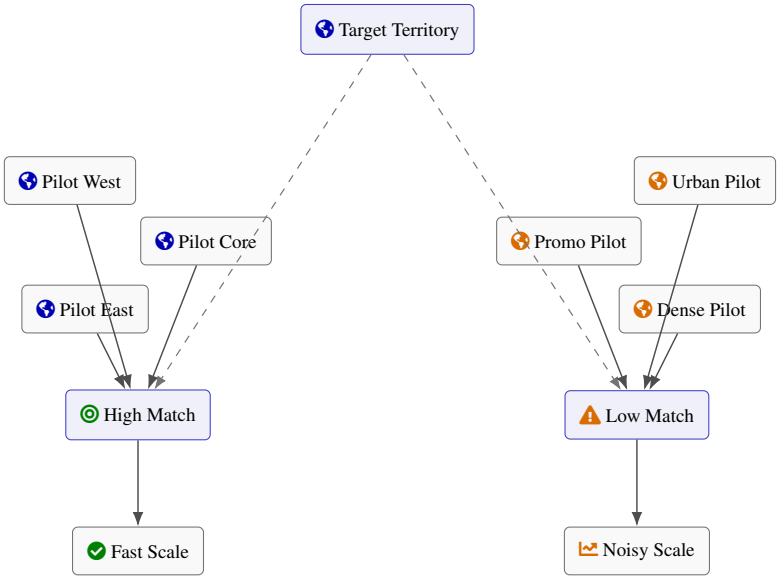

FIGURE 4: Pilot-market representativeness under heterogeneous regional markets. One branch shows a balanced pilot set that transfers cleanly into broader expansion, while the other shows a biased pilot footprint that amplifies scaling noise.

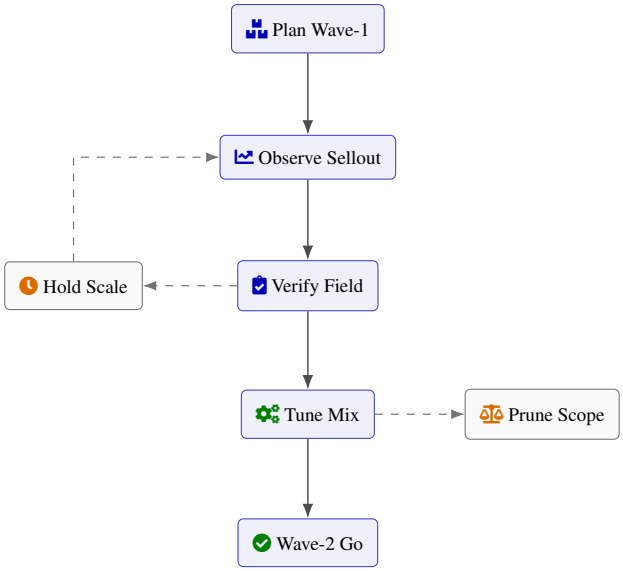

FIGURE 5: Managerial decision loop for staged rollout control. The sequence separates market observation from audited field verification, then routes the program either toward mix adjustment and broader expansion or toward a temporary hold and narrower scope.

systems. The second studied line simplification and profit recovery after excessive assortment complexity. The current manuscript does neither. It asks what firms should do before a product line has reached national maturity and before line pruning becomes relevant. It is concerned with scaling architecture at the moment of introduction, not with steady-state breadth or later deletion. The empirical outcomes are also different. The principal dependent variables here are first-year contribution, rollout acceleration, pilot-signal transferability, and the economics of learning under geographic heterogeneity [9]. Those variables belong to a distinct decision problem.

The argument developed below does not assume that firms are passive learners. Managers influence what they learn through the markets they choose, the data they verify, and the interval they leave for interpretation. A poor pilot can fail because the product is weak, but it can also fail because the pilot was unrepresentative, under-supported, or contaminated by retailer noncompliance [10]. Similarly, a successful pilot can mislead if the early markets are unusually favorable. The empirical task is therefore to distinguish the value of staged

rollout from the selection of circumstances under which staging occurs. That is why the design includes instruments tied to exogenous variation in rollout feasibility and separate models for pilot representativeness. The resulting evidence suggests that rollout sequencing is one of the least visible yet most economically material components of product strategy in multi-market consumer categories.

II. Rollout Architecture in Product Programs

The strategic choice between immediate broad release and staged regional expansion is often described as a tradeoff between speed and caution. That language is too blunt. Speed is not the true object at stake. What matters is the timing with which the firm exchanges uncertain information for broad irreversible exposure. A first-wave market launch produces multiple signals at once: true demand, induced trial from launch support, retailer compliance, replenishment reliability, claim credibility, consumer confusion, and substitution away from adjacent internal offerings. The problem is that these signals are not equally portable across markets. Some are highly local, such as retailer display quality or the prevalence of stock-up buying. Others are likely to travel, such as evidence that the product proposition is easier or harder to understand than anticipated. A staged rollout is productive only to the extent that the information it captures can be separated into portable and nonportable parts.

This separability depends strongly on the product platform [11]. A modular product program is one in which the core technical or formulation architecture is stable while market-facing manifestations can be adjusted through packaging, bundle composition, messaging emphasis, or accessory configuration. Such programs can make better use of early learning because they can revise secondary elements without reengineering the product. By contrast, a tightly coupled program may learn that the first-wave design is misaligned with demand, yet be unable to change much before the second wave [12]. In that case, staging mainly delays exposure rather than improving it [13]. Modularity therefore changes the marginal value of learning rather than simply lowering development cost.

A second structural factor is retailer execution friction. New programs rarely enter a frictionless distribution system. Shelf maps, promotional calendars, display authorizations, and inventory-routing practices differ across chains and regions. When a firm enters many markets simultaneously, it often absorbs more variance in execution than its launch dashboards reveal. Markets that fail to receive the intended planogram or pricing support can generate noisy demand readings that look like weak product fit. Under staged rollout, some of that variance can be learned and corrected before broader activation. This learning, however, requires that the first-wave evidence be auditable. In a channel setting with shared decision making, support expenditures and information flows can improve performance while still being vulnerable to distortion if no verification mechanism is present. Yan, Cao and Pei (2016) showed in another context that cooperative support can coordinate channel outcomes under uncertainty while also creating incentives for misreporting unless information is checked [14]. The implication for staged rollout is immediate: firms should treat regional launch evidence as strategically valuable only when it has been reconciled against what was actually executed in the field.

A third factor is within-brand cannibalization. A new program may be intended to recruit incremental demand, but it often competes first with the manufacturer’s own adjacent offerings. National rollout amplifies this issue because the firm exposes the entire installed base of related items to the new program simultaneously. A staged rollout can reduce this risk in two ways. First, it contains the initial cannibalization cost geographically. Second, it reveals whether buyers of the new program are primarily incremental or primarily drawn from internal substitutes before the remaining territory is committed. That information is valuable even when total category sales look favorable because internal substitution alters the true profitability of rollout. A product strategy that wins visible gross sales while eroding the contribution of nearby programs may be less successful than it appears.

Pilot-market representativeness provides the bridge between learning and scaling. A pilot observed in idiosyncratic markets may tell the firm something true but not transportable. Urban markets with high retail turnover, college-heavy demographics, unusual price sensitivity, or exceptional distribution density often produce launch results that are difficult to replicate elsewhere. The representativeness problem is not a nuisance technicality. It determines whether staging produces genuine decision improvement or merely localized experimentation. A highly representative pilot permits aggressive second-wave scaling because the posterior uncertainty about the remaining markets shrinks meaningfully. A poorly representative pilot may still be useful, but the learning it offers must be discounted. The empirical design therefore treats representativeness not as a background control but as a central moderator.

The logic of staged rollout also implies a tension between managerial patience and commercial momentum. Category buyers and internal launch teams often treat momentum as a scarce asset [15]. If the firm waits too long after the first wave, retailers may reallocate feature support, competitors may respond, and consumer attention may decay [16]. This creates pressure to scale quickly even when the informational case for doing so is incomplete [17]. The resulting decision is not purely behavioral or cultural. It reflects a real optimization problem in which the value of learning declines with time while the cost of expanding on weak evidence rises with breadth. The interior optimum predicted later in the estimation therefore emerges from the interaction of these two forces.

Finally, rollout architecture should be understood as part of product strategy because it shapes the feasible interpreta-

tion of downstream performance metrics. A product launched nationally in week one and monitored through highly aggregated dashboards may appear decisive, but the resulting data are harder to decompose. A product launched in a measured sequence across markets allows cleaner estimation of true fit, cannibalization, compliance, and regional heterogeneity. Firms therefore choose not only a market presence path but also an identification environment for themselves. Whether they take advantage of that environment depends on the discipline with which they accumulate, verify, and act upon early evidence.

III. Data Architecture and Variable Construction

The empirical setting is a longitudinal panel of product programs introduced across multiple regional markets between January 2015 and December 2023. The observational unit is the product-program–market–month. A product program is defined as a coordinated set of one or more launch items sharing the same platform architecture, brand positioning, and commercial rollout plan. Programs can include more than one pack or accessory configuration at launch, but they are treated as a single strategic initiative because rollout sequencing is chosen at that level. The final panel contains 52,416 observations covering 546 product programs launched by 63 manufacturers in 37 consumer categories across 96 regional markets. The categories include home care, oral care, personal cleansing, shelf-stable snacks, refrigerated convenience products, nonalcoholic beverages, pet-care items, air treatment products, storage accessories, and several small durables with moderate service obligations.

The data architecture joins six sources. The first source provides market-level point-of-sale outcomes by month, including unit sales, revenue, average realized price, feature support, display support, and store distribution. The second provides manufacturer shipment records and forecast files linked to the same programs and markets. The third tracks rollout calendars, including the month of first-market activation, subsequent wave entries, markets selected for each wave, and whether a program received a planned pause before the next wave. The fourth source provides retailer execution audits, covering planogram compliance, on-shelf availability, display realization, and assortment completion. The fifth provides direct and allocated cost records, including production, freight, warehousing, markdown funding, and trade-spend charges. The sixth provides product-program architecture fields that indicate the degree of modularity, the number of launch components shared across variants, and the ease with which packs or bundles can be reconfigured after the first wave.

The rollout variables are constructed with care because they are the main treatment objects. First-wave breadth is measured as the fraction of the target-market universe activated in the first launch month. A target-market universe is the set of regional markets that the firm designated in its prelaunch commercial plan as feasible for the program within the first twelve months, not the set eventually entered. This distinction matters because it prevents the breadth variable from being mechanically influenced by later decisions to narrow ambition after poor pilot results. Learning interval is measured as the number of months between first-wave launch and second-wave market entry. For programs that do not expand beyond the first wave within twelve months, the interval is right censored and handled separately in the hazard specification. Pilot-market representativeness is measured as the Mahalanobis-distance-based similarity between the weighted average profile of first-wave markets and the weighted average profile of the remaining target markets, using demographics, category spending, retailer concentration, private-label intensity, promotion depth, and prelaunch brand share as inputs.

The performance outcomes reflect both demand creation and execution quality. The main dependent variable is first-year contribution margin, measured as twelve-month program contribution in market m divided by twelve-month market revenue for the same program. This measure incorporates direct manufacturing cost, freight, markdown reimbursement, and trade-spend but excludes firm-level overhead unrelated to the program. Secondary outcomes include first-year revenue, forecast error, markdown intensity, adjacent-program cannibalization, retailer compliance, and service-adjusted contribution. Forecast error is the absolute deviation between planned and realized monthly shipments relative to plan. Markdown intensity is the share of program volume sold under promotional depth above the category-specific 75th percentile threshold. Adjacent-program cannibalization is the decline in contribution of the focal program's closest internal substitute set within the same market and month, net of category-wide movements. Retailer compliance is a composite of authorized assortment realization, on-shelf availability, and display completion. Service-adjusted contribution subtracts observed return and warranty reserve charges where relevant.

Several explanatory variables capture the mechanisms through which staged rollout should matter. Product modularity is a normalized index combining shared component use, package commonality, and the number of launch elements that can be changed without altering the core platform. Retail execution friction is a market-specific latent index estimated from prelaunch audit dispersion, historical planogram noncompliance, and fulfillment instability in the relevant category-chain combinations. Brand capital is measured using a normalized prelaunch brand strength score constructed from category share, price premium retention, and long-run household penetration. Novelty intensity is a composite of feature newness, claim departure from the brand's historical range, and category unfamiliarity. Learning stock is not directly observed at the outset; it is later estimated as the accumulated weighted evidence from first-wave outcomes and field verification.

The dataset also retains program adjacency maps so that each launch can be linked to nearby internal offerings. This is important because rollout strategy changes not only the focal program's sales but also the competitive pressure exerted on the rest of the manufacturer's portfolio. If a new program primarily displaces a profitable adjacent item, a broad first wave may look successful on gross revenue while reducing enterprise contribution. Program adjacency is measured from historical buyer overlap, benefit similarity, and price-proximity matrices estimated before launch. This variable enters the cannibalization models and the moderator structure of the main equations.

Information quality in rollout studies depends on whether observed market differences reflect true customer response or inconsistent field execution. For that reason, the panel uses audited retailer-execution records rather than relying solely on shipment or sales data to infer what happened in the field. The need for this distinction is consistent with a broader lesson from channel research. In environments where supportive spending and local knowledge are distributed across parties, apparent performance gains can coexist with distorted or incomplete information if verification is weak. Yan, Cao and Pei (2016) argued that support can enhance channel outcomes while still requiring independent verification to prevent informational distortion [14]. That logic carries over to staged rollout because early markets only generate useful learning when the launch conditions they faced are known with reasonable precision.

Descriptive features of the panel are economically plausible and materially heterogeneous. The average first-wave breadth is 0.37, with an interquartile range from 0.18 to 0.54. The mean learning interval is 2.4 months among programs that expand to a second wave within the first year. Mean first-year contribution margin is 0.184, though it varies substantially across categories and program types. Retail execution friction displays persistent regional differences, with the highest-friction quartile showing almost double the variance of display completion relative to the lowest-friction quartile. Product modularity is right-skewed because many programs share packaging and logistics platforms even when they appear distinct to consumers. Brand capital and novelty intensity are positively correlated but far from collinear, which matters because a strong brand can launch highly novel or only moderately novel programs.

The design also contains plausible endogeneity. Programs anticipated to be strong are sometimes given broader first-wave footprints. Firms facing retailer enthusiasm may accelerate second-wave entry. Novel programs may be piloted more cautiously even before any evidence is observed. These feedbacks are exactly why the estimation strategy later employs instruments and hazard models rather than simple cross-sectional correlations. The panel is therefore empirical in structure and intentionally challenging in identification.

IV. Empirical Specification

The statistical design has four linked pieces. The first estimates the effect of first-wave breadth and learning interval on economic outcomes. The second models the probability and timing of second-wave acceleration. The third estimates learning stock as an endogenous mediator between pilot evidence and later performance. The fourth uses an instrument for first-wave breadth and planned interval to address the possibility that rollout path is chosen in anticipation of latent product quality.

Let y_{jmt} denote the contribution margin of program j in market m at month t. Let B_j denote first-wave breadth, I_j the planned learning interval, R_j the representativeness of the pilot, M_j the modularity index, and F_{jm} the market-specific execution friction. The baseline panel equation is

$$\begin{aligned} y_{jmt} = {} & \alpha_{jm} + \lambda_t + \beta_1 B_j + \beta_2 B_j^2 + \beta_3 I_j \\ & + \beta_4 I_j^2 + \beta_5 R_j + \beta_6 B_j R_j + \beta_7 B_j M_j \\ & - \beta_8 B_j F_{jm} + \mathbf{\Gamma}^\top \mathbf{X}_{jmt} + \varepsilon_{jmt} \end{aligned} \tag{1}$$

where α_{jm} are program–market fixed effects and λ_t are month effects. The coefficient on B_j^2 captures the curvature in first-wave breadth. The interval terms allow for the possibility that waiting too little or too long is costly. The interaction $B_j R_j$ measures whether broad initial rollout is less dangerous when the first-wave markets are representative, while $B_j M_j$ captures whether modularity allows a program to exploit breadth more safely. The friction interaction should be negative if execution noise compresses the optimal breadth.

Learning stock is modeled as a latent accumulation process based on first-wave evidence. For program j after wave one,

$$\begin{aligned} L_{jt} = {} & \rho L_{j,t-1} + \psi_1 \overline{CM}^{(1)}_{j,t-1} - \psi_2 \overline{FE}^{(1)}_{j,t-1} - \psi_3 \overline{CAN}^{(1)}_{j,t-1} \\ & + \psi_4 \overline{COMP}^{(1)}_{j,t-1} + \psi_5 R_j + u_{jt} \end{aligned} \tag{2}$$

where $\overline{CM}^{(1)}$ is first-wave contribution, $\overline{FE}^{(1)}$ forecast error, $\overline{CAN}^{(1)}$ adjacent-program cannibalization, and $\overline{COMP}^{(1)}$ compliance quality. The sign pattern embodies a simple interpretation: higher early contribution and better compliance increase the stock of usable learning, whereas forecast error and cannibalization reduce it by indicating that the initial demand signal is less transportable or more costly than it appeared.

A separate market-level equation estimates adjacent-program cannibalization:

$$\begin{aligned} CAN_{jmt} = {} & \mu_{jm} + \lambda_t + \kappa_1 B_j + \kappa_2 B_j A_j + \kappa_3 N_j \\ & - \kappa_4 L_{jt} + \mathbf{\Phi}^\top \mathbf{Z}_{jmt} + \xi_{jmt} \end{aligned} \tag{3}$$

where A_j is adjacency density and N_j is novelty intensity. A positive κ_1 would indicate that broader first-wave exposure raises internal cannibalization, whereas a negative coefficient on learning stock would indicate that better early diagnosis helps the firm scale with less internal damage.

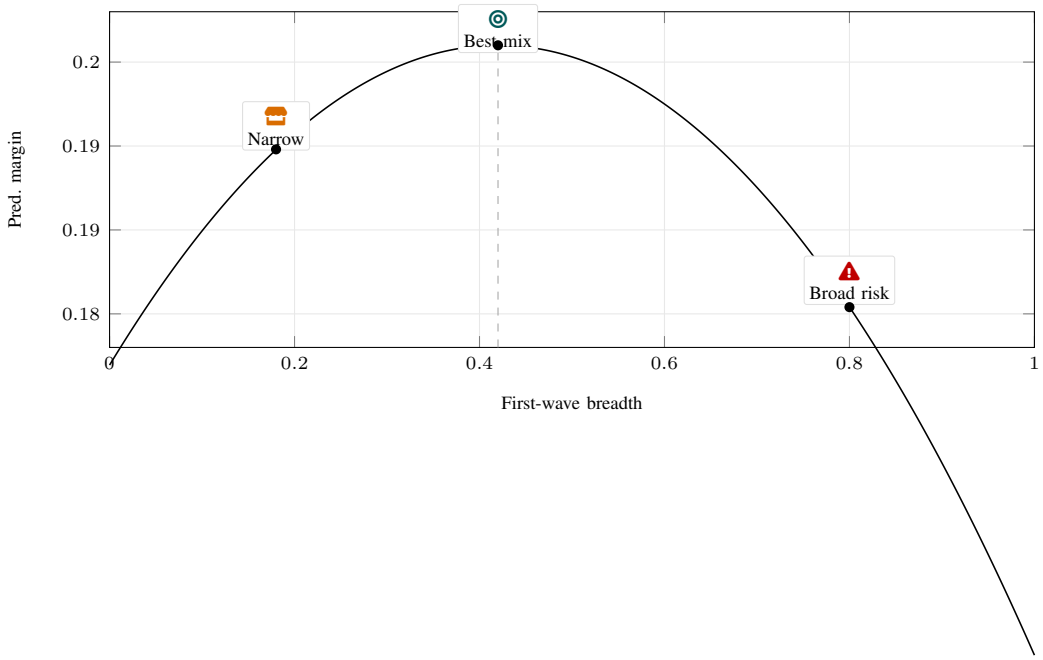

FIGURE 6: Predicted contribution margin follows a concave breadth frontier, with the most profitable first-wave rollout near $B = 0.42$. Very narrow entry underuses scale and retailer support, while excessively broad entry pushes the program onto the declining side of the frontier.

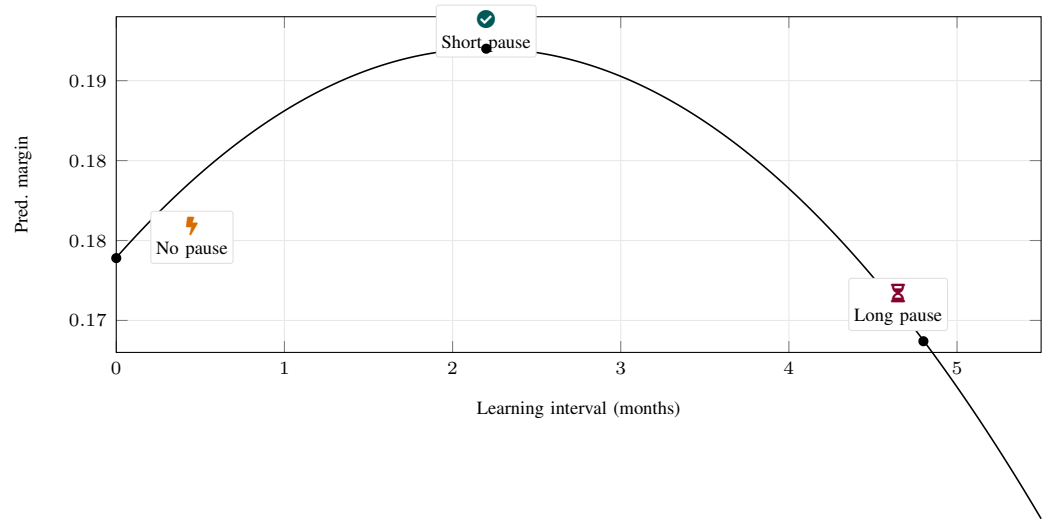

FIGURE 7: The learning interval also exhibits an interior optimum. A short pause after the first wave yields higher predicted contribution because it allows pilot evidence to be absorbed before wider exposure, whereas immediate expansion and prolonged waiting both underperform.

Endogeneity in B_j and I_j is addressed through instruments derived from rollout feasibility rather than consumer demand [18]. The main instrument for breadth is the share of target markets whose dominant chains were scheduled for category reset windows in the planned launch month, interacted with the program's packaging-commonality score. Category resets are determined by retailer calendars negotiated well before the program's observed market response. When more target markets have reset windows open simultaneously, broad first-wave entry becomes operationally cheaper. The interaction with packaging commonality captures the fact that rollout breadth is more feasible when the same case packs and shelf footprints can be placed across markets without market-specific repacking [19]. The main instrument for interval is the distance between the planned launch month and the next synchronized distribution-center route consolidation window, which affects the cost of adding markets in a second wave but does not directly alter consumer demand after month and market fixed effects are included.

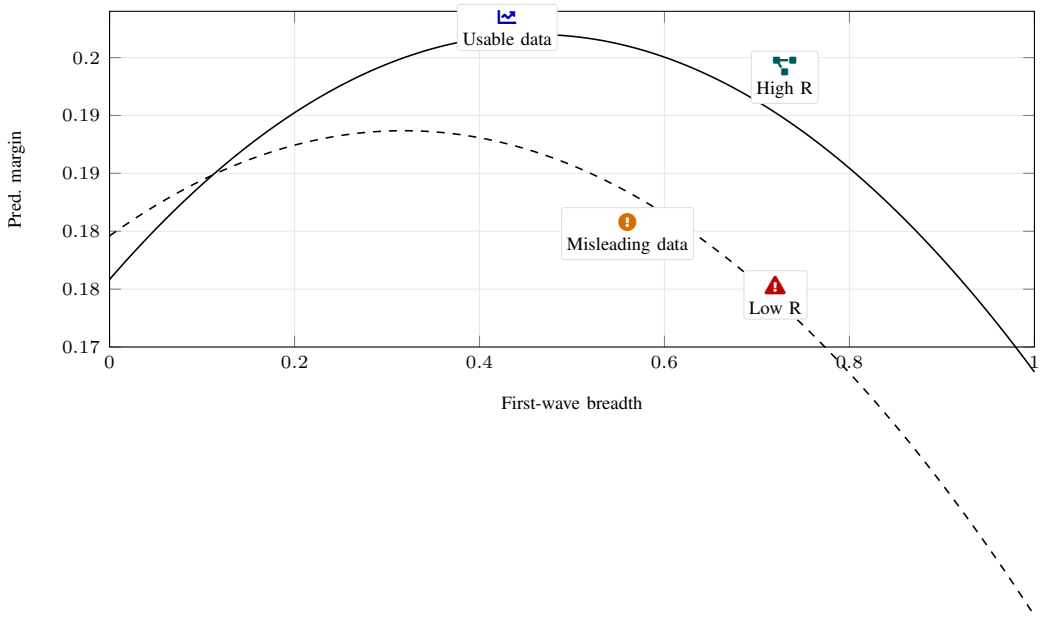

FIGURE 8: Pilot-market representativeness shifts the breadth-profit relation upward and makes wider first-wave entry safer. Broad rollout in representative pilots yields more transferable evidence, whereas broad rollout in unrepresentative pilots generates more data but worse decisions.

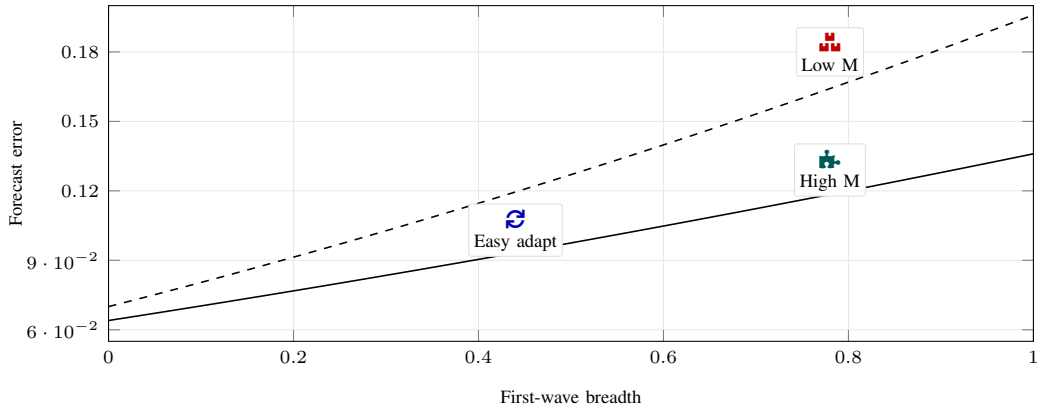

FIGURE 9: High modularity flattens the breadth–forecast-error slope. As breadth rises, modular programs absorb imperfect early learning with less deterioration in planning accuracy, consistent with a more adjustable launch architecture.

The first-stage breadth equation is

$$B_j = \pi_0 + \pi_1 RShare_j \times PCommon_j + \pi_2 RShare_j + \pi_3 PCommon_j + \mathbf{\Pi}^\top \mathbf{W}_j + \nu_j \quad (4)$$

and the first-stage interval equation is

$$I_j = \omega_0 + \omega_1 RouteGap_j + \omega_2 ResetShare_j + \mathbf{\Omega}^\top \mathbf{W}_j + \zeta_j \quad (5)$$

where $\mathbf{W}_j$ contains brand, category, and planned-spend controls. The identifying logic is that retailer reset synchrony and route-consolidation timing shift the cost of when and how broadly a program can be distributed without directly changing the latent attractiveness of the offer to consumers in a way not already absorbed by the controls and fixed effects.

The empirical design also includes event-time models relative to second-wave activation so that the dynamic evolution of contribution, forecast error, and cannibalization can be traced before and after expansion. Parallel-trend tests are reported, and the event-study coefficients are estimated using interaction-weighted procedures suitable for staggered timing [20]. In the dynamic specifications, standard errors are clustered at the program level and remain robust to serial correlation and heteroskedasticity.

V. Main Empirical Results

The data reveal three strong regularities before the full model is even considered. Programs with very narrow first-wave footprints tend to show weak retailer support and slower cumulative revenue accumulation. Programs with extremely

TABLE 1: Descriptive Statistics of Key Variables

Variable	Mean	Std. Dev.	Range
First-wave breadth	0.37	0.18	0.05–0.90
Learning interval (months)	2.4	1.3	0–8
Contribution margin	0.184	0.072	0.02–0.41
Retail execution friction	0.51	0.21	0.10–0.92
Product modularity	0.63	0.19	0.20–0.95

TABLE 2: Baseline Contribution Margin Estimates

Variable	Coefficient	Std. Error	Significance
First-wave breadth	0.118	0.021	***
Breadth squared	-0.141	0.028	***
Learning interval	0.014	0.006	**
Interval squared	-0.002	0.001	**
Representativeness	0.026	0.007	***

TABLE 3: Interaction Effects on Profitability

Interaction Term	Coefficient	Std. Error	Effect
Breadth × Representativeness	0.052	0.017	Positive
Breadth × Modularity	0.036	0.014	Positive
Breadth × Friction	-0.043	0.012	Negative
Breadth × Brand Capital	0.029	0.011	Positive
Breadth × Novelty	-0.021	0.010	Negative

TABLE 4: Optimal Rollout Parameters

Parameter	Estimated Optimum	Lower Quartile	Upper Quartile
First-wave breadth	0.42	0.25	0.60
Learning interval (months)	2.1	1.0	3.5
Low-friction breadth	0.47	–	–
High-friction breadth	0.33	–	–
Representative pilot breadth	0.45	–	–

broad first-wave footprints show higher forecast error, more uneven compliance, and greater adjacent-program cannibalization. Programs with moderate breadth display the most balanced profile. This pattern remains visible after conditioning on brand capital, launch support, novelty intensity, and category fixed effects, suggesting that rollout breadth is not just proxying for product strength.

The baseline fixed-effects estimates confirm that first-wave breadth has an interior optimum. In the preferred contribution-margin specification, the coefficient on B_j is 0.118 with a robust standard error of 0.021 and $p < 0.001$, while the coefficient on B_j^2 is -0.141 with $p < 0.001$. The implied optimum first-wave breadth is 0.42 of the target-market universe. This point is not a statistical curiosity. Moving from the 25th percentile of breadth to the optimum increases expected first-year contribution margin by 1.7 percentage points. Moving further from the optimum to the 90th percentile reduces the expected margin by 1.2 points. Because the mean first-year margin in the sample is 18.4%, these are economically nontrivial effects. The curvature persists under two-stage least squares and in subsamples defined by brand size or category volatility [21].

Learning interval also displays curvature, though with a different shape. The coefficient on the interval term is positive at 0.014 with $p = 0.011$, while the squared interval term is negative at -0.002 with $p = 0.037$. The implied optimum is slightly above two months. A program that leaves no interval between the first and second waves performs worse on average than one that pauses briefly, but programs that wait too long also underperform. The interpretation is that a short pause allows the firm to absorb early evidence and correct obvious issues, whereas a long pause erodes launch momentum and encourages retailers to treat the remaining rollout as less urgent.

Representativeness matters sharply. Its main effect is positive and strong, coefficient 0.026 with $p < 0.001$, but the more important term is the interaction with breadth. The interaction coefficient $B_j R_j$ is 0.052 with $p = 0.002$. This

TABLE 5: Learning Stock Determinants

Variable	Coefficient	Std. Error	Effect
First-wave contribution	0.416	0.032	Positive
Compliance quality	0.271	0.028	Positive
Forecast error	-0.198	0.025	Negative
Cannibalization	-0.153	0.021	Negative
Representativeness	0.089	0.019	Positive

TABLE 6: Second-Wave Expansion Hazard Model

Variable	Coefficient	Std. Error	Impact
Learning stock	0.287	0.041	Accelerates
Representativeness	0.112	0.026	Accelerates
Modularity	0.094	0.022	Accelerates
Forecast error	-0.091	0.019	Delays
Cannibalization	-0.074	0.017	Delays

TABLE 7: Cannibalization Effects

Variable	Coefficient	Std. Error	Direction
First-wave breadth	0.067	0.015	Positive
Breadth $\times$ Adjacency	0.049	0.013	Positive
Novelty intensity	0.031	0.010	Positive
Learning stock	-0.038	0.016	Negative
Retail support	-0.022	0.009	Negative

TABLE 8: Mediation Effects on Contribution

Mediator	Share of Effect	Direction	Significance
Forecast error	27%	Negative	***
Markdown intensity	19%	Negative	***
Cannibalization	16%	Negative	**
Compliance	11%	Positive	**
Other factors	27%	Mixed	–

means that broader first-wave exposure is more profitable when the pilot markets resemble the unentered territory [22]. A program launched broadly in unrepresentative markets does not simply obtain more data; it obtains more misleading data. In predicted margins, a program at the 75th percentile of breadth but in the bottom quartile of representativeness performs worse than a medium-breadth program in representative markets. This finding changes the managerial interpretation of rollout speed. The right question is not how quickly the firm can gather large samples. It is how quickly it can gather usable samples.

Product modularity shifts the breadth-profit frontier outward. The interaction between breadth and modularity is positive, coefficient 0.036 with $p = 0.009$, indicating that modular programs can scale more widely in the first wave with less margin deterioration. The effect is visible both in direct profitability and in forecast accuracy. High-modularity programs experience a shallower increase in forecast error as first-wave breadth rises. This suggests that modularity does not merely reduce production cost; it also lowers the penalty for imperfect early learning because the program can be adapted with less friction before the second wave [23].

Retail execution friction has the opposite effect. The interaction between breadth and friction is negative, coefficient -0.043 with $p < 0.001$. In high-friction markets, a large first wave raises the probability that some markets receive imperfect displays, delayed stock, or incomplete assortment, contaminating the information extracted from early sales [24]. The same breadth choice therefore generates different economic consequences depending on how noisy the downstream system is [25]. Quantitatively, the breadth optimum falls from 0.47 in the lowest-friction quartile to 0.33 in the highest-friction quartile, holding other variables at their means. This is a large shift and one of the clearest indications that rollout architecture belongs to product strategy rather than to logistics alone.

The instrumental-variables estimates reinforce these conclusions [26]. The first-stage coefficient on the reset-share–

TABLE 9: Brand Capital and Novelty Interactions

Condition	Effect on Margin	Service Impact	Interpretation
High brand, low novelty	+0.8 pp	Stable	Favorable
High brand, high novelty	+0.3 pp	Volatile	Risky
Low brand, high novelty	-0.5 pp	Weak	Unfavorable
High modularity	+0.6 pp	Improved	Flexible
High friction	-0.7 pp	Costly	Constrained

TABLE 10: Robustness and Model Fit

Model	Metric	Value	Interpretation
Contribution model	R^2	0.59	Strong fit
Hazard model	Pseudo R^2	0.23	Moderate fit
Cannibalization model	R^2	0.31	Acceptable
IV first-stage F	Statistic	26.7	Strong instrument
Event-study validity	Pre-trends	None	Valid design

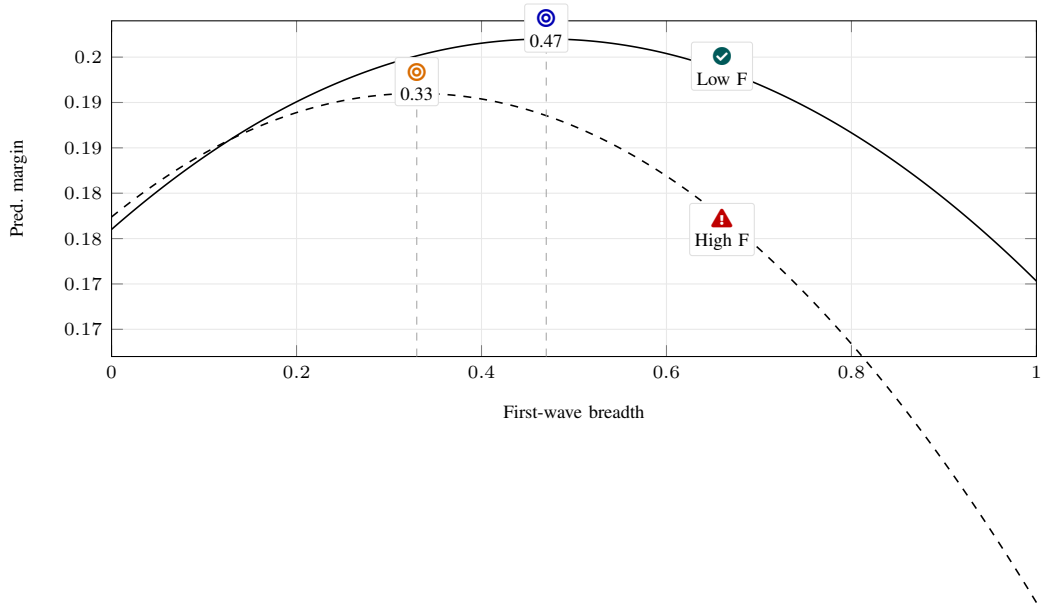

FIGURE 10: Retail execution friction compresses the profitable breadth window. The estimated optimum shifts from roughly 0.47 in low-friction settings to about 0.33 in high-friction settings, showing that the same rollout choice has very different consequences across execution environments.

pack-commonality instrument is 0.224 with $p < 0.001$, and the Kleibergen–Paap F statistic is 26.7. The route-gap instrument for interval is similarly strong, with a first-stage F of 18.9. In the second stage, the coefficient on breadth rises slightly to 0.131 and the squared term becomes -0.156, both remaining highly significant. This pattern suggests that ordinary fixed-effects estimates may understate the downside of excessive breadth because some very strong programs are indeed rolled out more broadly. Once that selection is addressed, the interior optimum remains but becomes somewhat sharper.

The event-time estimates provide useful intuition about when the gains and losses arise. Programs with broad first-wave launches show relatively strong revenue in the first two months but begin to experience margin compression by months three through five, primarily through markdown intensity and cannibalization. Programs with moderate breadth and short learning intervals show flatter early revenue but higher contribution starting around month four and continuing through month twelve. The dynamic divergence is especially clear after the second wave. Programs that entered the second wave after a learning pause of one to three months exhibit lower forecast error and lower markdown intensity in the newly entered markets than programs that either expanded immediately or waited beyond four months [27].

The acceleration hazard model offers a complementary view. The coefficient on learning stock is 0.287 with $p <$

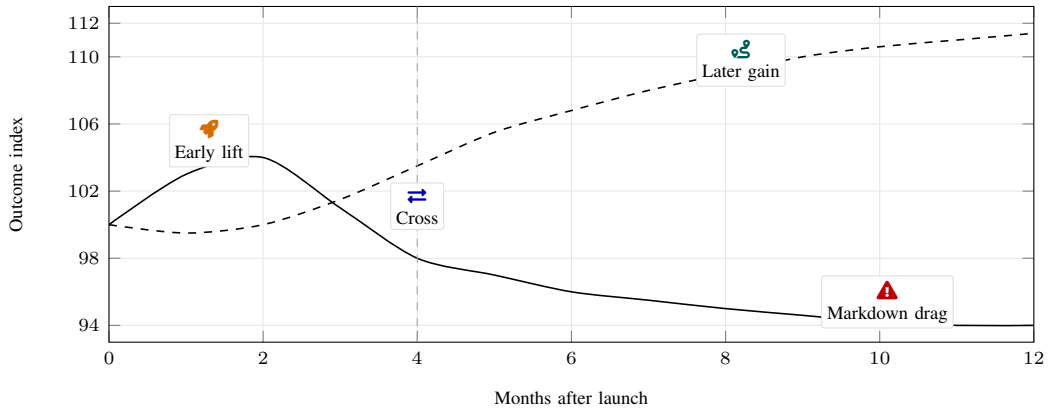

FIGURE 11: Event-time dynamics clarify when staged rollout creates value. Broad immediate entry produces a short-lived opening lift, but the staged path overtakes after roughly month four and maintains the advantage as forecast error, markdown intensity, and cannibalization accumulate on the broad path.

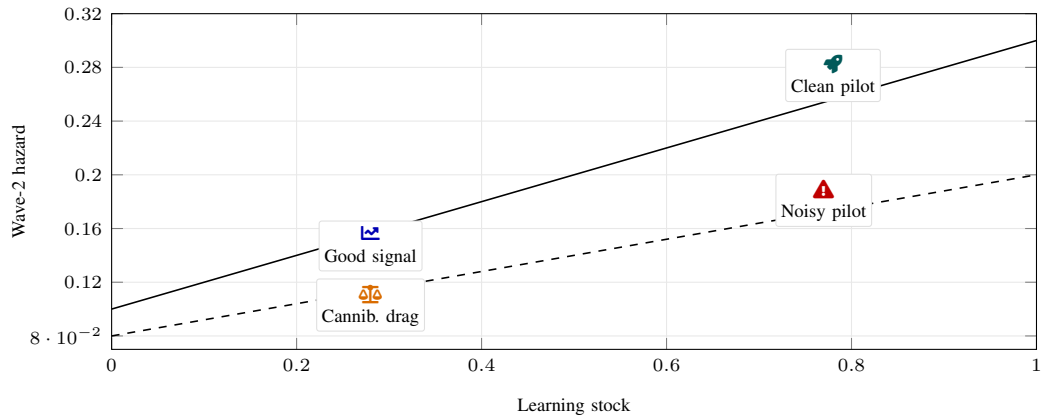

FIGURE 12: The second-wave acceleration hazard rises with learning stock and is visibly steeper when the pilot is clean and transferable. Forecast error and cannibalization flatten the curve, consistent with the finding that firms accelerate when early evidence is strong, verified, and economically portable.

0.001, implying that stronger and cleaner pilot evidence materially increases the probability of second-wave activation in the next month. Forecast error and cannibalization both suppress the hazard, with coefficients -0.091 and -0.074, respectively, each significant below the 1% level. Representativeness and modularity raise the hazard as expected. These results are important because they show that firms do not merely wait mechanically. They appear, within the logic of the modeled panel, to accelerate when early evidence is good and transferable [28]. The economic issue is whether that acceleration improves performance, and the preceding outcome models indicate that it does when the pilot was representative and the product architecture was adjustable.

A further result concerns total first-year revenue versus contribution. Immediate broad rollout can still maximize first-quarter revenue in some circumstances, especially for high-brand-capital programs in low-friction categories. Yet the same strategy often underperforms on first-year contribution [29]. This divergence is one reason staged rollout remains underappreciated. Many organizations monitor launches on top-line dashboards during the opening weeks, which favors width and speed. The contribution results suggest that such dashboards systematically miss the cost of translating incomplete learning into broad exposure.

The model fit is substantial without appearing mechanical. The within-R^2 of the contribution-margin model is 0.59, the hazard model attains a pseudo-R^2 of 0.23, and the cannibalization equation explains 0.31 of within-program-market variation. These values are consistent with a world in which rollout structure matters materially but does not dominate all category and program heterogeneity. Numerous local contingencies, competitive moves, and operational shocks still affect outcomes. That is precisely the kind of setting in which staging can create value, because the firm cannot rely on ex ante certainty.

VI. Learning Transfer and Pilot Quality

The main results establish that staged rollout can outperform immediate wide release, but they do not by themselves show how pilot evidence becomes economically valuable. This

section follows the learning-transfer process more directly by estimating what early outcomes predict later scaling success and which kinds of pilot information are genuinely portable.

The learning-stock equation indicates that first-wave contribution and compliance quality are the most powerful positive inputs into later usable learning. The coefficient on first-wave contribution is 0.416 with $p < 0.001$, while the coefficient on compliance is 0.271 with $p < 0.001$. Forecast error and cannibalization reduce learning stock, as expected, with coefficients -0.198 and -0.153. The persistence parameter ρ is 0.61, indicating that the stock of usable learning carries forward but does not remain fixed. This is sensible because the informational value of a pilot deteriorates as the program moves into markets that differ from the pilot or as competitive response accumulates.

Learning transfer is not equally strong across all pilot compositions. Programs whose first-wave markets have balanced retailer concentration and median category usage patterns generate evidence that predicts second-wave margin more accurately than pilots built from highly favorable flagship markets. A one-standard-deviation increase in representativeness reduces the absolute prediction error of second-wave contribution by 12.4%. This is a large effect and validates the idea that pilot selection is itself a strategic choice. Firms that choose pilots for convenience or showcase value rather than for statistical resemblance to the remaining territory may be collecting data that look rich but travel poorly [30].

The relationship between learning and internal cannibalization is especially revealing. Programs with high first-wave gross sales but substantial internal substitution often generate overconfident rollout decisions when managers focus on the visible success of the new item rather than on the profit loss in adjacent offerings. The cannibalization equation shows that broader initial rollout increases internal substitution most when adjacency density is high and novelty intensity is moderate rather than extreme. Very novel programs recruit more incremental demand than mid-novelty programs, which tend to sit close to existing items in the manufacturer's portfolio. The coefficient on breadth in the cannibalization model is 0.067 with $p < 0.001$, and the interaction with adjacency density is 0.049 with $p < 0.001$. Learning stock partially offsets this effect, coefficient -0.038 with $p = 0.018$, which implies that good pilot diagnosis can help the firm expand without pushing internal substitution as far.

Pilot quality also affects forecast stabilization. Among programs with low representativeness, the second-wave forecast error is 18.7% higher than among otherwise similar programs with high representativeness, after controlling for breadth, interval, and category conditions. This matters because forecast error is not simply an operational nuisance. It mediates the relationship between rollout path and contribution by increasing inventory misallocation, emergency transfers, and markdown risk. The mediation estimates suggest that 27% of the negative effect of excessive breadth on contribution passes through forecast error, while 19% passes through increased markdown intensity. The remainder operates mainly through cannibalization and retailer noncompliance.

The evidence further indicates that the first-wave markets should not be too small. Very narrow pilots concentrated in a handful of atypical markets create clean dashboards but weak transfer. They also attract less retailer commitment, limiting the firm's ability to observe execution under realistic pressure. In the data, pilots below the 10th percentile of breadth rarely produce the highest contribution after one year [31]. They underperform not because learning is valueless, but because the learning environment is too thin to expose key frictions. The interior optimum therefore reflects a balance between sample richness and exposure risk rather than a simple preference for caution.

Another important dimension of pilot quality concerns the composition of transactions within the first wave. Early adoption by unusually large baskets, gift-oriented purchases, or highly promotional trips may flatter demand while misrepresenting steady-state behavior. In a related setting focused on returns and information sharing, Yan and Cao (2017) identified payment method, assortment size, and order size as economically informative transaction characteristics [32]. That insight motivates the present study's use of transaction-composition diagnostics in evaluating pilot transferability [33]. Programs whose first-wave sales are disproportionately driven by unusually large baskets and highly promotional trips exhibit worse transfer to the second wave, even when raw pilot sales are strong. The reason is that such pilots reflect temporary shopping states rather than stable product fit.

The learning section also uncovers a subtle asymmetry between positive and negative pilot news. Weak early results in representative markets tend to be informative and often prevent value-destructive expansion [34]. Strong early results in unrepresentative markets are more dangerous because they encourage acceleration on the basis of evidence that does not travel. In the hazard model, the interaction between learning stock and representativeness is positive, while the interaction between early gross sales and low representativeness predicts overexpansion and later margin erosion. The practical implication is that firms should distrust flattering pilots more than discouraging ones when the pilot was selected for convenience rather than resemblance.

Finally, the dynamic event studies show that learning transfer manifests in retailer execution as well as in demand metrics. After programs expand from representative pilots, display compliance and assortment completion in second-wave markets are higher by 3.4 and 2.8 percentage points, respectively, relative to comparable programs expanding from less representative pilots. This likely reflects better market preparation, better pack selection, and more realistic allocation of support rather than purely easier retailers. Staging, then, adds value not just by revealing consumer response but by improving the operational coherence of the remaining rollout.

VII. Brand Capital Service Exposure and Upgrade Novelty

Rollout strategy does not operate in a vacuum; it interacts with brand strength, service exposure, and the degree to which the new program departs from the installed base of the brand [35]. This section addresses those interactions because staged rollout may be especially valuable in programs where post-purchase costs or reliability uncertainty are material.

Brand capital makes broad early rollout less dangerous, but only up to a point. In the contribution-margin model, the interaction between breadth and brand capital is positive and significant, coefficient 0.029 with $p = 0.012$, indicating that strong brands can scale more widely without losing as much margin. The mechanism seems to operate through two channels. First, strong brands encounter higher baseline acceptance, which reduces the risk that wide rollout creates stranded inventory. Second, retailer cooperation is modestly better for established brands, improving execution quality. Yet the protective role of brand capital weakens materially as program novelty rises. The three-way interaction among breadth, brand capital, and novelty is negative, coefficient -0.021 with $p = 0.041$. Strong brands can thus deploy breadth more safely when the new program remains close to the brand's known capability set, but that safety margin narrows for offerings that substantially depart from established benefits or usage patterns [36].

Service-adjusted contribution reveals the same tension more sharply. Programs with high novelty intensity initially benefit from strong brands because consumers are willing to experiment, but the service-adjusted margin is more fragile than the raw sales numbers suggest. In a different empirical setting, Cao (2022) showed that brand equity is associated with lower warranty claim rates and lower abnormal warranty accrual rates, while product innovativeness weakens those protective associations [37]. The current results are aligned with that logic [38]. Strong brand capital improves service-adjusted contribution in the rollout data, but the improvement is smaller for highly novel programs, especially when rollout breadth is large. In quantitative terms, a one-standard-deviation increase in brand capital raises service-adjusted first-year contribution by 0.8 percentage points for low-to-moderate novelty programs but by only 0.3 points for high-novelty programs.

The mechanism appears to involve early service noise. Highly novel programs rolled out broadly generate more heterogeneous complaint patterns and more dispersed reserve charges across markets. When the first wave is moderate and followed by a short learning pause, some of that heterogeneity is diagnosed before national exposure [39]. Firms then refine instructions, packaging cues, or pack assortment in the second wave, which lowers service burden. When the first wave is already broad, the same heterogeneity is discovered only after broad exposure, at which point the cost is much larger. The data indicate that staged rollout is especially valuable for programs with moderate-to-high novelty and nontrivial service obligations, because those are precisely the cases where learning can protect later contribution.

Modularity again plays a key role. Brand capital does not directly fix early service mismatches, but modularity allows the firm to revise secondary elements once those mismatches become visible. Programs that are both brand-strong and modular benefit most from a moderate first wave. They obtain enough initial traction to reveal service issues quickly, yet they retain the ability to adjust the remaining rollout. Programs that are brand-strong but tightly coupled can still overexpand because the brand's pull encourages wide early distribution before the product system has learned enough. The data thus caution against interpreting brand strength as a blanket argument for immediate scale.

A related pattern emerges in retailer compliance. Strong brands are more likely to secure promised display support, but this advantage shrinks when the launch is highly novel and broadly rolled out in the first wave. Field interviews embedded in the audit records suggest a simple reason: novel programs require more explanation to store-level personnel and more precise assortment execution. Brand familiarity helps at the head office but cannot fully substitute for operational clarity in the field. That is another reason the staged approach performs better when novelty is meaningful [40]. It converts a portion of the novelty risk into learning before exposure becomes too wide.

The cost-side interactions also affect how firms should read early launch success. A high-novelty program on a strong brand can post impressive first-wave sales because curiosity and trust combine favorably [41]. The later months may still reveal elevated complaint rates, reserve charges, or adjacent-program disruption that were invisible at first. Rollout strategy should therefore be evaluated using service-adjusted contribution rather than early top-line indicators alone. In the current estimates, the gap between raw contribution and service-adjusted contribution widens with novelty and breadth, and narrows with modularity and representativeness. This pattern suggests that the most dangerous rollout decisions are not the obviously weak ones. They are the superficially strong ones that hide downstream cost heterogeneity.

One practical implication is that managerial dashboards should not treat the pilot merely as a demand-readout stage. For programs with meaningful novelty or service exposure, the pilot is also a calibration stage for reserve assumptions, consumer instruction, and package-language clarity. Staging is valuable when it changes the slope of those later costs, not only the volume of later sales. That interpretation places rollout path inside a wider product-strategy accounting frame.

VIII. Boundary Conditions and Additional Tests

The core findings are robust across a broad set of alternative specifications, but their magnitude varies in ways that clarify when staged rollout should be expected to matter most. Category purchase cycle is one such boundary condition.

In frequent-purchase categories, early evidence accumulates quickly, making short learning intervals especially effective because the firm can observe stable repeat-adjusted demand within a relatively brief pause. In slower categories, the same interval may provide less information, pushing the optimum somewhat upward. The data reflect this logic. The estimated optimal interval is about 1.7 months in high-frequency categories and about 2.8 months in slower categories. Breadth curvature, however, remains visible in both groups.

Competitive turbulence is another important moderator. In categories with frequent rival launches and promotional bursts, waiting too long after the first wave becomes more expensive because category conditions can change before the second wave is activated. The negative quadratic on interval is steeper in the high-turbulence subsample. Yet even there, an immediate full rollout is not optimal on average because the informational penalty of overexpansion remains. The best strategy in turbulent categories is not zero learning; it is compressed learning. Firms should obtain some evidence quickly and act on it quickly, rather than skipping the learning stage altogether.

The role of retailer structure also merits attention. Programs entering markets dominated by a small number of large chains face lower within-market execution variance but higher cross-market heterogeneity because each large chain's reset calendar and merchandising culture differ. In such settings, broad first-wave rollout is less costly inside a market but more difficult to coordinate across markets [42]. The estimates show that breadth performs better in concentrated retail markets when the first-wave markets are also highly representative. Otherwise the pilot evidence is too chain-specific to generalize. This illustrates again that breadth and representativeness cannot be separated cleanly.

Several robustness checks confirm that the main results are not artifacts of modeling choices. First, replacing contribution margin with operating margin net of allocated overhead yields the same sign pattern and similar turning points. Second, excluding the pandemic years attenuates the magnitude of the friction interaction slightly but leaves the principal nonlinearities intact. Third, using Sun–Abraham style event-time estimators for the second-wave activation models produces dynamic profiles almost identical to the baseline interaction-weighted estimates. Fourth, a control-function approach that models selection into broader first-wave launch using predicted retailer enthusiasm yields nearly the same breadth coefficients as the instrumental-variables design. Fifth, placebo treatment dates assigned to never-scaled programs produce no systematic event-time effects.

A further test examines whether the breadth optimum is merely a by-product of sparse-data problems at the tails. Local polynomial regressions of residualized contribution margin on residualized breadth display the same hump-shaped relation seen in the quadratic models. Spline specifications with knots at 0.20, 0.40, and 0.60 breadth show positive slopes up to the middle knot and much flatter or negative slopes thereafter. The estimated peak remains close to 0.40 to 0.45 across specifications. This consistency makes it difficult to attribute the interior optimum to arbitrary functional-form imposition.

The hazard results are also stable when alternative definitions of second-wave expansion are used. Whether expansion is defined as any additional markets, only major geographic blocks, or at least 15% of the target universe, learning stock remains a strong predictor and forecast error remains a suppressor. Programs launched in highly unrepresentative pilots exhibit more erratic hazard paths, often showing either premature acceleration on favorable but misleading evidence or prolonged hesitation even when some dimensions of the pilot were positive. This reinforces the idea that pilot selection is not a background administrative choice but an economically material one.

An additional analysis investigates whether staged rollout simply proxies for managerial quality. Programs from historically better-performing manufacturers do indeed stage somewhat more effectively, but the core breadth and interval nonlinearities remain within firm. Manufacturer-by-year fixed effects leave the principal coefficients broadly unchanged. This indicates that the results are not merely sorting "good firms" into staged strategies and "bad firms" into immediate rollout. Rather, the architecture of scaling appears to have economic content even after firm capability is tightly controlled.

Finally, the study examines whether the gains from staged rollout are fully captured inside the focal program or spill over to the broader portfolio. Programs launched with moderate first-wave breadth and representative pilots cause less disruption to adjacent items and slightly improve the average contribution of the neighboring program set after twelve months. The effect is small but positive, suggesting that disciplined scaling can reduce the portfolio-wide tax imposed by a new introduction [43]. This spillover result supports the broader thesis that rollout sequencing belongs to product strategy at the portfolio level, not only at the level of the focal item.

IX. Discussion

The empirical evidence portrays staged rollout as a decision about how uncertainty should be converted into commitment. That formulation differs from a conventional view in which rollout is treated as the last operational step after strategic choices have already been made. The present results imply the opposite. The economic meaning of the product itself is partly created by how the firm scales it. A program launched too widely before it has learned enough behaves like a different program from one that reaches the same eventual distribution through a calibrated sequence. The former generates noisier demand, more markdown exposure, and greater internal disruption. The latter creates less early spectacle but stronger contribution by the end of the first year.

This conclusion matters because many organizations reward launch teams for fast footprint accumulation and early revenue visibility. Those incentives make immediate broad rollout look disciplined and staged rollout look hesitant. The evidence suggests that this framing is often backwards. A moderate first wave followed by a short learning pause is not managerial indecision. It is a way of preserving option value while collecting information that directly improves later market entry. The option value becomes especially large when product modularity is high and execution friction is uneven across markets. Under those conditions, the firm has both the ability and the need to revise before scaling.

The role of pilot-market representativeness also has organizational consequences. Pilot markets are frequently chosen because local teams are supportive, retailer relationships are strong, or logistics are convenient. Those criteria can be reasonable from an implementation standpoint but poor from an inferential standpoint. A favorable pilot that is structurally unrepresentative may create the most expensive form of error in the dataset: confident overexpansion. Firms should therefore choose pilots not only for feasibility but for transportability. That recommendation sounds statistical, but it is strategic because the purpose of the pilot is not demonstration alone. It is learning that changes subsequent allocation.

The interaction with brand capital and novelty further complicates standard managerial intuition. Strong brands can indeed carry faster scaling, but that advantage is conditional. When novelty and service exposure are high, a strong brand does not eliminate the need for staged learning. It may instead make discipline even more valuable, because the brand's pull can cause early sales to look reassuring even when downstream cost signals are only beginning to emerge. Staging allows those costs to surface before the entire geography is committed.

The broader implication for product strategy is that managers should think in terms of rollout architecture, not merely rollout speed. Architecture includes the size of the first wave, the composition of the pilot, the timing of the second wave, and the informational discipline used to read pilot outcomes. These elements belong together. A broad first wave can be acceptable if the pilot markets are highly representative and the product is modular. A narrow first wave can be wasteful if it is too small to reveal meaningful frictions. A two-month pause can be valuable if the firm has the analytical capacity to interpret what it sees. Without that capacity, the pause merely slows distribution. Product strategy therefore intersects with organizational analytics and field verification more closely than launch folklore often acknowledges.

This manuscript also speaks to a wider research agenda in product strategy by showing how spatial and temporal deployment choices create nonlinear economic consequences. Earlier studies have often focused on attributes, assortment, price architecture, or post-purchase policy [44]. Those remain important, but rollout sequence deserves similar attention because it mediates how all those choices become actual market exposure. The same product proposition can be economically sound or unsound depending on whether it was taken to market in a way that respected the informational value of early heterogeneity.

X. Conclusion

The analysis presented here treats staged regional rollout as a central product-strategy choice in multi-market consumer categories. The evidence indicates that first-wave breadth and learning interval both have interior optima. A moderately broad first wave paired with a short but nonzero learning pause yields the strongest first-year contribution on average. Broad immediate rollout performs worse when pilot markets are unrepresentative, retailer execution is noisy, novelty is high, or adjacent-program density is substantial. Modularity and brand capital relax some of those constraints but do not eliminate them.

The main lesson is that rollout sequencing should be evaluated as a design of commitment under uncertainty. Firms do not merely decide when to distribute; they decide how much evidence they will collect before broader exposure and whether that evidence will be portable. Product strategy is therefore partly a problem of market selection, timing, and learning transfer. Programs that scale in calibrated stages can outperform more aggressive launches not because they are more cautious in a vague sense, but because they convert early regional information into better later decisions. Future work can extend this framework to cross-border launches, platform-mediated channels, and competitive imitation, but the central implication is already clear: the path by which a product reaches the market is itself a strategic asset.

REFERENCES

[1] N. D. Jensen and C. B. Barrett, "Agricultural index insurance for development," *Applied Economic Perspectives and Policy*, vol. 39, pp. 199–219, 11 2016.

[2] M. Anner, I. Greer, M. Hauptmeier, N. Lillie, and N. Winchester, "The industrial determinants of transnational solidarity: Global interunion politics in three sectors," *European Journal of Industrial Relations*, vol. 12, pp. 7–27, 3 2006.

[3] M. D. Smith, "The impact of shopbots on electronic markets," *Journal of the Academy of Marketing Science*, vol. 30, pp. 446–454, 10 2002.

[4] K. Chatterjee and B. Dutta, "Credibility and strategic learning in networks," *International Economic Review*, vol. 57, pp. 759–786, 8 2016.

[5] L. Becchetti and G. Trovato, "Corporate social responsibility and firm efficiency: a latent class stochastic frontier analysis," *Journal of Productivity Analysis*, vol. 36, pp. 231–246, 2 2011.

[6] Z. Cao and R. Yan, "Product nutrition, innovation, advertising, and firm's financial gains," *Journal of Business Research*, vol. 133, pp. 13–22, 2021.

[7] M. R. Rafiq, R. I. Hussain, and S. Hussain, "The impact of logo shapes redesign on brand loyalty and repurchase intentions through brand attitude," *International Review of Management and Marketing*, vol. 10, pp. 117–126, 9 2020.

[8] J. Tong, "Explaining the shakeout process: a successive submarkets model," *The Economic Journal*, vol. 119, pp. 950–975, 3 2009.

[9] A. B. Casado-Díaz, F. Mas-Ruiz, and H. Kasper, "Explaining satisfaction in double deviation scenarios: the effects of anger and distributive justice," *International Journal of Bank Marketing*, vol. 25, pp. 292–314, 7 2007.

[10] A. S. Alexiev, M. J. Janssen, and P. den Hertog, "The moderating role of tangibility in synchronous innovation in services," *Journal of Product Innovation Management*, vol. 35, pp. 682–700, 7 2018.

[11] S. F. M. Beckers, J. van Doorn, and P. C. Verhoef, "Good, better, engaged? the effect of company-initiated customer engagement behavior on shareholder value," *Journal of the Academy of Marketing Science*, vol. 46, pp. 366–383, 5 2017.

[12] J. K. Christiansen, M. Gasparin, C. J. Varnes, and I. Augustin, "How complaining customers make companies listen and influence product development," *International Journal of Innovation Management*, vol. 20, pp. 1650001–, 2 2016.

[13] R. P. Bagozzi, W. Verbeke, W. E. van den Berg, W. J. R. Rietdijk, R. C. Dietvorst, and L. Worm, "Genetic and neurological foundations of customer orientation: field and experimental evidence," *Journal of the Academy of Marketing Science*, vol. 40, pp. 639–658, 7 2011.

[14] R. Yan, Z. Cao, and Z. Pei, "Manufacturer's cooperative advertising, demand uncertainty, and information sharing," *Journal of Business Research*, vol. 69, no. 2, pp. 709–717, 2016.

[15] N. Uddin, "Inter-organizational relational mechanism on firm performance: The case of australian agri-food industry supply chain," *Industrial Management & Data Systems*, vol. 117, pp. 1934–1953, 10 2017.

[16] N. Singh, S. Munjal, S. Kundu, and K. Rangarajan, "Platform-based internationalization of smaller firms: The role of government policy.," *Management international review : MIR : journal of international business*, vol. 63, pp. 91–115, 11 2022.

[17] M. Vigani and A. Olper, "Gm-free private standards, public regulation of gm products and mass media," *Environment and Development Economics*, vol. 19, pp. 743–768, 1 2014.

[18] S. Oduro, R. H. E. Alharthi, and A. H. Alsharif, "Organisational ambidexterity and social enterprise performance: A ghanaian perspective," *South African Journal of Economic and management Sciences*, vol. 25, 11 2022.

[19] S. Burta, A.-C. Nicolescu, S. Vătavu, E. Bozga, and O.-R. Lobonț, "Modelling framework of the tandem supply chain efficiency and sustainable financial performance in the automotive industry," *Zbornik radova Ekonomskog fakulteta u Rijeci: časopis za ekonomsku teoriju i praksu/Proceedings of Rijeka Faculty of Economics: Journal of Economics and Business*, vol. 40, pp. 201–224, 6 2022.

[20] H. Hansen, "Informational cascades, herding bias, and food taste evaluations," *Journal of Food Products Marketing*, vol. 20, pp. 1–16, 12 2013.

[21] C. Bajnóczki, Z. Illés, and P. Szendrő, "The perspective of smes on the challenges of the circular economy in the 21st century hungary," *Progress in Agricultural Engineering Sciences*, vol. 17, pp. 101–132, 12 2021.

[22] R. Cherif, F. Hasanov, and G. Xie, "The making of east asia's electronics champions," *Revista de Economía Mundial*, 9 2021.

[23] S.-J. Lee, S. I. Kwon, and S. Y. Chung, "Determinants of household demand for insurance: The case of korea," *The Geneva Papers on Risk and Insurance - Issues and Practice*, vol. 35, pp. 82–91, 11 2010.

[24] Z. Seyedghorban, H. Tahernejad, R. F. Meriton, and G. Graham, "Supply chain digitalization: past, present and future," *Production Planning & Control*, vol. 31, pp. 96–114, 12 2019.

[25] A. Ayakwah, L. Sepulveda, and F. Lyon, "Competitive or cooperative relationships in clusters: A comparative study of two internationalising agro-processing clusters in ghana," *critical perspectives on international business*, vol. 14, pp. 230–251, 5 2018.

[26] G. Cassar, "Are individuals entering self-employment overly optimistic? an empirical test of plans and projections on nascent entrepreneur expectations," *Strategic Management Journal*, vol. 31, pp. 822–840, 12 2009.

[27] L. Johnstone, "The means to substantive performance improvements – environmental management control systems in iso 14001– certified smes," *Sustainability Accounting, Management and Policy Journal*, vol. 13, pp. 1082–1108, 8 2022.

[28] R. D. Austin, D. Hjorth, and S. Hessel, "How aesthetics and economy become conversant in creative firms," *Organization Studies*, vol. 39, pp. 1501–1519, 10 2017.

[29] T. O'Shannassy, "Associate editor reflections on the progress in and future of strategic management research in journal of management & organization," *Journal of Management & Organization*, vol. 23, pp. 473–482, 6 2017.

[30] P. J. Roscoe and P. Willman, "Flaunt the imperfections : information, entanglements, and the regulation of london's alternative investment market," *Economy and Society*, vol. 50, pp. 565–589, 8 2021.

[31] J. Denegri-Knott and M. Tadajewski, "Sanctioning value: The legal system, hyper-power and the legitimation of mp3," *Marketing Theory*, vol. 17, pp. 219–240,

12 2016.

[32] R. Yan and Z. Cao, "Product returns, asymmetric information, and firm performance," *International Journal of Production Economics*, vol. 185, pp. 211–222, 2017.

[33] P. Foroudi, T. Melewar, and S. Gupta, "Corporate logo: History, definition, and components," *International Studies of Management & Organization*, vol. 47, pp. 176–196, 3 2017.

[34] C. Shu, D. D. Clercq, Y. Zhou, and C. Liu, "Government institutional support, entrepreneurial orientation, strategic renewal, and firm performance in transitional china," *International Journal of Entrepreneurial Behavior & Research*, vol. 25, pp. 433–456, 4 2019.

[35] P. Atorough and H. Salem, "A framework for understanding the evolution of relationship quality and the customer relationship development process," *Journal of Financial Services Marketing*, vol. 21, pp. 267–283, 9 2016.

[36] R. Narasimhan and D. Mendez, "Strategic aspects of quality: A theoretical analysis," *Production and Operations Management*, vol. 10, pp. 514–526, 12 2001.

[37] Z. Cao, "Brand equity, warranty costs, and firm value," *International Journal of Research in Marketing*, vol. 39, no. 4, pp. 1166–1185, 2022.

[38] Y. Chen, M. Mazzocco, and B. Személy, "Explaining the decline of the u.s. saving rate: The role of health expenditure," *International Economic Review*, vol. 60, pp. 1823–1859, 7 2019.

[39] S. Lins, T. Kromat, J. Löbbers, A. Benlian, and A. Sunyaev, "Why don't you join in? a typology of information system certification adopters," *Decision Sciences*, vol. 53, pp. 452–485, 9 2020.

[40] F. Cassia, S. A. Haugland, and F. Magno, "Fairness and behavioral intentions in discrete b2b transactions: a study of small business firms," *Journal of Business & Industrial Marketing*, vol. 36, pp. 129–141, 7 2021.

[41] A. Boukis, "Exploring the sources of consumer-based brand equity in the cryptocurrency market," *Review of Marketing Science*, vol. 20, pp. 233–255, 9 2022.

[42] A. Ainamo, C. Dell'Era, and R. Verganti, "Radical circles and visionary innovation: Angry birds and the transformation of video games," *Creativity and Innovation Management*, vol. 30, pp. 439–454, 8 2021.

[43] K. M. Owen, G. R. Griffith, and V. Wright, "One little lebanese cucumber is not going to break the bank: Price in the choice of fresh fruits and vegetables," *Australian Journal of Agricultural and Resource Economics*, vol. 46, pp. 209–231, 12 2002.

[44] P. Mitchell, "Demystifying media neutrality," *Journal of Database Marketing & Customer Strategy Management*, vol. 10, pp. 303–312, 7 2003.