

Applications of Operations Research in Retail Supply Chains: A Literature-Based Analytical Study of Inventory Replenishment and Store Distribution

Author Details:

Arnav Singh Rawat, Ahaan Shah, Arpit Mishra, Aradhaya Chadha, Ananya Bhatia

¹ B.Sc. Finance / Anil Surendra Modi School of Commerce (ASMSOC) / NMIMS, Mumbai, India



<https://doi.org/10.55041/ijssmt.v2i10.033>

Cite this Article: Rawat, A. S., Shah, A., Mishra, A., Chadha, A. & Bhatia, A. (2026). Applications of Operations Research in Retail Supply Chains: A Literature-Based Analytical Study of Inventory Replenishment and Store Distribution. International Journal of Science, Strategic Management and Technology, 02(10), 1-9. <https://doi.org/10.55041/ijssmt.v2i10.033>



License: This article is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting use, distribution, and reproduction in any medium, provided the original author(s) and source are properly credited.

ABSTRACT

Retailers repeatedly decide how much stock to order, how to split it across stores and how to ship it, under uncertain demand and with limited stock, labour and vehicle capacity. This study examines how Operations Research (OR) has been applied to these inventory replenishment and distribution decisions. Five published studies are reviewed, two in full and three as supporting papers, covering four stages of the retail supply chain: ordering, allocation, shipping and the store shelf record. For each study, the decision variables, objective function, constraints, assumptions, solution method, software and validation are set out and assessed, and the integer programme of one study is re-solved with its own published data. In four of the five studies, the model outperformed the benchmark its authors chose, from a

I. INTRODUCTION

A. Operations Research

Operations Research (OR) is the use of mathematical models to choose the best course of action when resources are limited. Every OR model has three parts: the **decision variables**, which are the quantities the decision-maker controls; the **objective function**, which

profit gain of 2.23% to 4.92% of sales at a catalogue retailer to an estimated 3% to 4% more sales at Zara, using existing stock, staff and vehicles. None of the problems was solved exactly in daily use; each relied on a tested simplification, usually the newsvendor trade-off or a linear approximation. The re-solved model reproduces its reported optimum of 342 pallets, but shows that the paper's claimed gain over past dispatch rests on the wrong figure, and that relaxing the full-load rule alone would raise the optimum to 379 pallets. The quality of validation is what separates the strongest studies from the weakest.

Keywords— Operations research; inventory replenishment; retail supply chain; integer programming; newsvendor model; stock allocation

is the measure to be maximised or minimised; and the **constraints**, which are the limits the decision must respect. A solution that satisfies every constraint is feasible, and the feasible solution with the best objective value is optimal [1, Ch. 1].

An OR study moves through five phases: (1) definition of the problem, (2) construction of the model, (3)

solution of the model, (4) validation of the model and (5) implementation of the solution [1, Ch. 1].

B. The Real-Life Problem and Its Industry

The five papers reviewed in this study all belong to the **retail supply chain**. They share one question: how much of each product should be ordered, sent to each store and kept on each shelf, and when? Sending too little means lost sales and unhappy customers. Sending too much ties up cash, fills trailers and shelves with slow stock and ends in markdowns.

The numbers are large. ILOS (2014), cited in [2], puts transport, inventory and storage costs for Brazilian companies at 8.7% of net revenue. Zara offers about 11,000 references in a season, against 2,000 to 4,000 for its key competitors (Ghemawat and Nueno, 2003, cited in [3]). The data behind these decisions are also often wrong: an audit of nearly 370,000 inventory records across 37 stores of one retailer found 65% of them to be inaccurate (DeHoratius and Raman, 2004, cited in [4]).

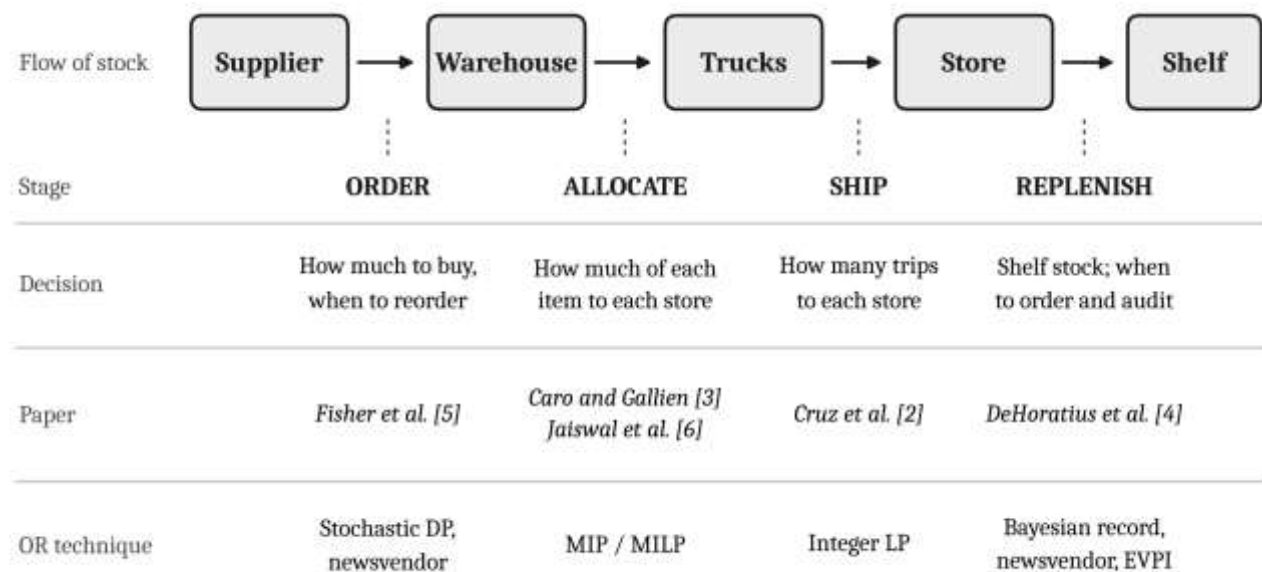


Figure 1: The Retail Supply Chain and the Main Stage Studied by Each Paper (prepared for this study from the five papers; primary papers: Cruz et al. [2] and Caro and Gallien [3])

Figure 1 places the five papers on the chain, and Section II-B compares them.

C. Research Gap and Objectives

Each of the five studies examines a single decision, such as an ordering rule, an allocation model or a dispatch plan, and tests it against a benchmark of its authors' own choosing. Read separately, they do not show how the models connect along the supply chain, how strongly each was validated, or whether the reported figures hold when checked against the papers' own data. This study addresses that gap for retail inventory replenishment and distribution. Its objectives are to: (i) describe the real-life problem at each stage of the chain and explain why OR suits it; (ii) set out each study's mathematical model, with its decision variables, objective function, constraints and

assumptions; (iii) compare the solution methods, software and validation used; (iv) assess the reported results and their practical implications; and (v) evaluate the strengths and limitations of each study, including a re-solution of one model with its published data. Section II reviews the standard OR models and the five studies, Section III sets out the study design and the models, Section IV presents and discusses the results, and Section V concludes.

II. LITERATURE REVIEW

A. Standard OR Models

Table I sets out the three standard deterministic models that the rest of this study builds on [1, Chs. 2 and 5].

Table I: Standard OR Models

Model	Decision variables	Objective	Constraints	Usual solution method
Linear programming problem (LPP)	x_j = level of activity j	Max or Min $Z = \sum_j c_j x_j$	For each resource i : $\sum_j a_{ij} x_j \leq b_i$ (or $=$ or $\geq$); $x_j \geq 0$	Graphical method (two variables); simplex method
Transportation problem	x_{ij} = units sent from source i to destination j	Min $Z = \sum_i \sum_j c_{ij} x_{ij}$	Supply of source i : $\sum_j x_{ij} = a_i$; demand of destination j : $\sum_i x_{ij} = b_j$; $x_{ij} \geq 0$	Starting solution by north-west corner, least-cost or Vogel's method (VAM); optimality test by MODI (method of multipliers)
Assignment problem	$x_{ij} = 1$ if worker i does job j , and 0 otherwise	Min $Z = \sum_i \sum_j c_{ij} x_{ij}$	Each worker gets one job: $\sum_j x_{ij} = 1$; each job gets one worker: $\sum_i x_{ij} = 1$	Hungarian method

The three are related. The transportation problem is an LPP with a special structure, written above in its balanced form (total supply equals total demand), and the assignment problem is a transportation problem in which every supply and demand equals 1. Two extensions matter for this study. When some variables must be whole numbers, the model becomes an integer or mixed-integer programme (MIP or MILP). When demand is uncertain, the model becomes stochastic; the basic case

Table II: The Five Papers at a Glance

Paper	Retail setting	Type of problem	OR technique	Why OR fits
Fisher et al. [5] (supporting)	Catalogue retailer; 120 styles/colours of women's dresses	Stochastic, two-stage ordering	Stochastic dynamic programming (DP); newsvendor heuristic	The first order is fixed before demand is seen and only one reorder follows, so the costs of buying too much and too little must be balanced
Caro and Gallien [3] (primary)	Zara (fast fashion); more than 1,000 stores, two warehouses	Stochastic allocation of scarce stock	Stochastic sales model inside a MIP	Fixed stock must be split each week across 1,000+ stores and several sizes: too many linked choices to set by hand
Jaiswal et al. [6] (supporting)	Global retailer; one warehouse, about 9,000 items, 158 stores	Deterministic allocation	MILP; greedy heuristic; optimal transport	Millions of item-store-day quantities share limited labour and trailer space
Cruz et al. [2] (primary)	Mid-size chain, São Paulo; one distribution centre, 20 stores, 12 vehicles	Deterministic distribution (transport)	Integer LP	20 stores compete for limited vehicle minutes; loads were chosen by intuition
DeHoratius et al. [4] (supporting)	"Gamma" (name disguised); more than 1,000 stores	Stochastic, partly unobserved inventory	Bayesian updating; newsvendor; expected value of perfect information (EVPI)	Shelf stock cannot be observed, so orders and audits must be decided under uncertainty

C. Why OR Techniques Are Suitable

Three features make this an OR problem. The decisions are too large to balance by hand: Zara's warehouse team set shipments of several hundred references, in several sizes, to more than a thousand stores every week [3], and the model of Jaiswal et al. [6] has millions of variables. The resources are fixed and shared: stock, labour hours,

is the single-period newsvendor model, which weighs the cost of ordering too much against the cost of ordering too little.

B. The Five Studies

Table II gives the setting, the type of problem and the reason OR fits for each of the five papers, and marks the two primary papers.

trailer space and driver minutes, which is the situation LP and integer programming are designed for. And demand is uncertain, so every decision trades lost sales against leftover stock, which is the newsvendor trade-off.

The papers also show what happens without a model: dispatchers in Cruz et al. [2] chose loads by intuition, and

the rule-based planners described by Jaiswal et al. [6] produced plans the warehouse could not always carry out. The last column of Table II gives the specific reason OR fits each paper.

III. METHODOLOGY

A. Study Design

The problem type chosen from the suggested list is **retail inventory replenishment and distribution**, which the brief places under transportation problems. None of the five papers models it as a textbook transportation problem: in practice the problem has one origin feeding many stores, trips or trailers that must be whole numbers, and uncertain demand. Section III-B shows how the closest paper departs from the transportation model.

The brief asks for one or two papers. Two are analysed in full against every item of the brief: Cruz et al. [2], the paper closest to the models of Table I, and Caro and Gallien [3], the only paper whose model was tested live in a working company. Three more, Fisher et al. [5], Jaiswal et al. [6] and DeHoratius et al. [4], are used as supporting papers, because replenishment runs through four stages of the retail supply chain (Figure 1) and the two primary papers cover only allocation and shipping. For the two stochastic supporting papers, which go beyond linear programming, this study explains the logic of each model and keeps only its central equations. Section IV-D re-solves the Cruz et al. [2] model with the paper's own data.

Each paper was analysed against the same set of items: the real-life problem and why OR suits it; the mathematical model, with its decision variables, parameters, objective function, constraints and assumptions; the solution method and software; the validation design; and the reported results. Figures were taken from the versions of the papers read and were recalculated wherever a paper gives enough data to do so; Section IV-C reports where they do not hold. The resolution in Section IV-D was carried out in Python, using the MILP function of SciPy, which calls the HiGHS solver.

B. Mathematical Models

The two primary papers come first, followed by the three supporting papers from the simplest model to the most complex. Each is set out with its decision variables, parameters, objective function, constraints and assumptions, and a plain-language reading. Symbols are

defined separately for each model and are local to its subsection.

1) Shipping from one distribution centre: integer LP (Cruz et al. [2])

A distribution centre serves 20 stores with 10 trucks (14 pallets each) and 2 van trucks (12 pallets each). Dispatchers had no rule for which store to serve first. The paper reports an average of 415 pallets waiting each morning, but its store-level demand table sums to 427; the model, and this study, use 427. The model plans the day's trips. It is restated here in cleaner notation, because equations (3) to (5) of the paper use the letters j and l for both a vehicle type and its capacity.

Decision variables. x_i = truck trips to store i ; y_i = van-truck trips to store i , for $i = 1, \dots, 20$.

Parameters. TC_i = cycle time for store i in minutes (loading + outward journey + unloading + return), ranging from 118 to 228; DP_i = pallets demanded by store i (427 in total). Available time is vehicles $\times$ (shift – lunch): $10 \times (500 - 60) = 4,400$ truck minutes and $2 \times (500 - 60) = 880$ van-truck minutes.

Objective function. Maximise the number of pallets shipped in the day:

$$\max Z = \sum_{i=1}^{20} (14 x_i + 12 y_i)$$

Table III lists the constraints.

Table III: Constraints of the Shipping Model

Constraint	What it means
$\sum_i TC_i x_i \leq 4400$	All truck trips must fit into the trucks' working minutes
$\sum_i TC_i y_i \leq 880$	All van-truck trips must fit into the van trucks' working minutes
$14 x_i + 12 y_i \leq DP_i$ for every store i	No store is sent more pallets than it demanded
$x_i, y_i \in \{0, 1, 2, \dots\}$	Trips are whole numbers; a fraction of a trip means nothing

Link with Table I. The model looks like a transportation problem, with one source (the distribution centre) and 20 destinations, but it is better read as an integer LPP of the resource-allocation type. Its capacity constraints, $\sum_i TC_i x_i \leq 4400$ and $\sum_i TC_i y_i \leq 880$, have the form $\sum_j a_{ij} x_j \leq b_i$ of the LPP row in Table I. A transportation

problem with a single source is trivial, and its constraints carry coefficients of 1, not cycle times. The objective also maximises pallets moved instead of minimising cost, and trips must be whole numbers, so the model is an integer programme, which the authors solve in Excel Solver and not by VAM and MODI.

Assumptions. Demand and cycle times are fixed at their three-month averages. Every trip serves one store and returns. A trip is always counted at full vehicle capacity, so part-loads are never sent. Section IV-C shows that this rule, not vehicle time, is what limits the plan. Time is pooled across all vehicles of a type instead of being tracked vehicle by vehicle.

2) Allocating scarce fashion stock: a sales model inside a MIP (Caro and Gallien [3])

Zara ships each reference to its stores once a week. Two features make the problem unusual. Warehouse stock of a reference is limited and is normally not replenished. And a store takes a reference off display as soon as one of its **major sizes** (for example S, M or L) sells out, so a store left with only odd sizes sells nothing. The authors tested this display rule on sales data and found that stores followed it 89% of the time on average.

The model is built in two steps. First comes a sales model for one store: expected weekly sales equal the demand rate multiplied by the expected time the reference stays
Table IV: Constraints of the Fashion Allocation Model

Constraint	What it means
$\sum_j x_{sj} \leq W_s$ for every size s	Shipments of a size cannot exceed warehouse stock
$z_j \leq (\sum_{s \in S^+} \lambda_{sj}) y_j + \sum_{s \in S^-} \lambda_{sj} v_{sj}$ for every store j	Expected sales = demand rate $\times$ display time, for major and minor sizes
$y_j \leq a_i(\lambda_{sj})(I_{sj} + x_{sj} - i) + b_i(\lambda_{sj})$ for every major size s and line i	Display time is capped by the scarcest major size
$v_{sj} \leq a_i(\lambda_{sj})(I_{sj} + x_{sj} - i) + b_i(\lambda_{sj})$ and $v_{sj} \leq y_j$ for every minor size s	A minor size sells only while it is in stock and the reference is on display
$z_j, y_j, v_{sj} \geq 0$ and $x_{sj} \in \{0, 1, 2, \dots\}$	Shipments are whole units

A high K makes shipments cautious and a low K makes them aggressive. Because y_j is capped by the scarcest major size, the model prefers to send complete size sets to fewer stores instead of scattering broken sets over many.

Assumptions. Demand for each size follows a Poisson process and is independent across sizes and references, so customers do not substitute. Unmet demand is lost. Forecasts are taken as given. The model looks one week

on display, and that time ends when the first major size runs out. Second, this sales function is approximated by a few straight lines, each of the form slope $\times$ (stock $- i$) + intercept, so that it can be placed inside a linear model for the whole network.

Decision variables. x_{sj} = units of size s shipped to store j (the main decision); z_j = approximate expected sales in store j ; y_j and v_{sj} = helper variables for the expected display time of the major sizes and of minor size s .

Parameters. P_j = selling price in store j ; W_s = warehouse stock of size s ; I_{sj} = stock already in the store; λ_{sj} = forecast demand rate; a_i, b_i = slope and intercept of straight line i ; K = the value given to a unit kept back in the warehouse; S^+ and S^- = the sets of major and minor sizes.

Objective function. Maximise this week's expected sales revenue plus the value of the stock held back for later weeks:

$$\max \sum_j P_j z_j + K \sum_s \left(W_s - \sum_j x_{sj} \right)$$

Table IV lists the constraints.

ahead, with K standing in for the future. Revenue, not profit, is maximised.

3) Allocating warehouse stock with a look ahead: MILP (Jaiswal et al. [6])

INVALS plans, for one warehouse and all its stores, how many warehouse-packs of each item go to each store and how many trailers carry them. Its central idea is **pull-forward**: on a quiet day the warehouse also ships part of what the stores will need over the next few days, so that labour and trailers are not left idle.

Decision variables. d_{ij}^t = quantity of item i sent to store j against the need of day t ($t = 0$ is the current period; $t \geq 1$ are pull-forward days); x_j = number of trailers sent to store j (a whole number); b_j = shortfall of the last trailer below its minimum load. For short, $s_{ij} = \sum_t d_{ij}^t$ is the total of item i sent to store j , $s_j = \sum_i s_{ij}$ is the total sent to store j , and $y_j \in \{0,1\}$ shows whether store j gets any trailer, linked to x_j by $y_j \leq x_j \leq R_j y_j$. The paper writes this link as $y_j = \min(1, x_j)$, which is not linear; the two inequalities are its standard linear form.

Parameters. D_{ij}^t = store need; S_i = warehouse stock; H_l = labour capacity for handling channel l ; I_l = set of items handled by channel l ; M, m = maximum and minimum trailer load; C_{ij} = shelf capacity; R_j = maximum trailers for store j ; p_j = store priority; q_{ij}^t = item-store priority, higher for faster-selling items; $\alpha[t]$ = a weight that falls as t rises; β, γ = large positive constants, with $\gamma > \beta$.

Objective function.

$$\max \sum_{i,j,t} \alpha[t] q_{ij}^t d_{ij}^t + \beta \sum_j p_j x_j - \gamma \sum_j b_j$$

The first term rewards every unit shipped, and rewards it more when the unit is needed now and sells quickly. The second term makes sure priority stores get their trailers. The third term penalises trailers that leave below the minimum load.

Table V lists the constraints.

Table V: Constraints of the INVALS Model

Constraint	What it means
$\sum_j s_{ij} \leq S_i$ for every item i	Cannot ship more than the warehouse holds
$\sum_{i \in I_l} \sum_j s_{ij} \leq H_l$ for every channel l	Cannot ship more than warehouse labour can handle
$x_j \leq R_j$, with $x_j \in \{0,1,2, \dots\}$	Trailers per store are limited and are whole numbers
$s_j \leq M x_j$	Total load cannot exceed trailer capacity
$s_j \geq M (x_j - y_j) + m y_j - b_j$, with $0 \leq b_j \leq m$	Earlier trailers are full; the last one carries at least m , or pays the penalty on b_j
$s_{ij} \leq C_{ij}$	Cannot send more than the store shelf can take
$0 \leq d_{ij}^t \leq D_{ij}^t$	Cannot send more than the store needs for that day

Assumptions. Store needs come from demand forecasts and are treated as known. The objective is a utility score, not a money value, and its weights are set by the business. The quantities d_{ij}^t are allowed to be fractional to keep the problem solvable; the authors report that 99.99% of them come out as whole numbers anyway.

4) Ordering a short-season product: stochastic DP and a newsvendor heuristic (Fisher et al. [5])

A catalogue retailer sells a dress for one season of 22 weeks. The supplier's lead time is 12 weeks, and the paper allows only one reorder. The retailer places an initial order Q_1 , watches early sales x , and then places a reorder Q_2 . Before the reorder arrives, unmet demand is backordered; after it arrives, unmet demand is lost.

Decision variables. Q_1 and Q_2 .

Parameters. C_u = cost of a lost sale (price – cost); C_o = cost of a leftover unit (cost – salvage value); C_b = cost of a backorder. Demand before the reorder, during the lead time and after the reorder arrives are random variables X, Y and W .

Objective function. Minimise the expected cost of backorders, lost sales and leftover stock over the season:

$$Z = \min_{Q_1 \geq 0} E[C_b \max(X - Q_1, 0) + C_2(Q_1, X)]$$

The first term is the cost of backorders before the reorder is placed. $C_2(Q_1, x)$ is the lowest expected cost that can still be reached once early sales x are known, by choosing the reorder well: backorders during the lead time, then lost sales and leftovers at the end of the season. This is a two-stage stochastic dynamic programme, because one decision is taken before demand is known and one after part of it has been seen.

Constraints. Orders cannot be negative: $Q_1 \geq 0$ and $Q_2 \geq 0$.

Solution rule. The authors show that C_2 is neither convex nor concave, so standard methods are not guaranteed to find the best answer. They propose the **two-period newsvendor heuristic**. Let F_1 be the distribution function of demand up to the arrival of the reorder ($X + Y$) and F_2 that of demand over the whole season ($X + Y + W$). The first order is the Q_1 that satisfies

$$F_1(Q_1) + \frac{C_o}{C_b} F_2(Q_1) = 1$$

Once x is seen, the reorder raises stock to the newsvendor quantity, which is the level that the remaining demand stays below with probability equal to the critical ratio

$$\frac{C_u - C_b}{C_u - C_b + C_o}$$

In words, the first order balances the risk of early backorders against the risk of leftovers at season end, and the reorder is an ordinary newsvendor quantity based on the improved forecast.

Assumptions. There is one reorder, at a time fixed inside the model; the authors choose that time outside the model by simulation (Table VI). Price and lead time are fixed, and leftovers are sold below cost. Season demand is normal, and early and late demand are jointly normal. A known fraction of sales (35% in the application) is returned and can be resold at once.

5) Replenishing when the stock record is wrong: Bayesian record, newsvendor and EVPI (DeHoratius et al. [4])

Store systems keep a recorded inventory: yesterday's record plus replenishment minus sales. The physical stock on the shelf is different, because of theft, damage, misplaced items and scanning errors. The authors bundle all of these into one unobserved "invisible demand" and write the physical stock at the end of day t as

$$I_t = I_{t-1} + R_t - S_t - U_t$$

where R_t is the replenishment received, S_t the sales recorded and U_t the units removed (or added) by invisible demand. Since U_t is never observed, the retailer cannot know I_t . Instead of a single number it keeps a **Bayesian inventory record**: a probability for each possible stock level, updated every day by Bayes' rule from the sales it sees. The intuition is that a day with no sales raises the chance that the shelf is empty, while a day with sales proves that it was not.

Decision variable 1: the order quantity R . With holding cost h per unit per period and penalty p per unit of lost sale, the objective is to minimise the one-period cost

$$C_1(R) = h E[\max(R + A - D, 0)] + p E[\max(D - A - R, 0)]$$

where A is the uncertain stock on hand before the order arrives and D is demand in the period when it arrives. The best order is the smallest whole number R for which the probability of covering demand, that is of $D - A \leq R$, is at least $p/(h + p)$. This is the newsvendor critical ratio again.

Decision variable 2: when to audit. A physical count costs K and removes the uncertainty. The benefit of removing it is measured by the expected value of perfect information, $EVPI(t)$. The rule is to audit a product when

$$t \cdot EVPI(t) > \frac{\gamma + 1}{\gamma} K$$

where t is the number of days since the last audit and γ , between 0 and 1, describes how quickly EVPI grows.

Constraints. Orders are whole and non-negative, $R \in \{0, 1, 2, \dots\}$; physical stock cannot be negative, $I_t \geq 0$; and the audit decision each day is yes or no.

Assumptions. The model covers one product reviewed daily. Unmet demand is lost and is not observed. All errors come from a single process that arrives after the day's sales. Demand and errors are independent and stationary. There is no fixed ordering cost, audits are perfect, and decisions look one period ahead.

C. Solution Methodology and Software

The five papers reach their answers in three ways (Table VI). Cruz et al. [2] solve their small integer programme exactly in Excel Solver. Caro and Gallien [3] could not solve the full dynamic programme for more than a thousand stores in the few seconds Zara allowed, so they simplified it to the one-week MIP of Section III-B and solved that exactly. The other three rely on heuristics built on the structure of the problem: the newsvendor critical ratio in Fisher et al. [5] and DeHoratius et al. [4], and in Jaiswal et al. [6] a greedy rule that adds trailers one at a time to the store that gains most, followed by one LP that fixes the item quantities. Jaiswal et al. also solve the full MILP exactly with CPLEX, but only as a benchmark.

Two of the validation designs in Table VI deserve a closer look. Caro and Gallien [3] used the new process only at the Arteixo warehouse and kept the old one at

Zaragoza, so the Zaragoza figures measure the error of their own method. Jaiswal et al. [6] switched the parts of

their objective on and off one at a time, which shows what each term contributes.

Table VI: Solution Method, Tools and Validation

Paper	Solution method	Software and tools named	How the solution was validated
Fisher et al. [5]	Two-period newsvendor heuristic; Monte Carlo simulation to choose the reorder week	Not named (runs timed on a Dell Pentium II PC)	Back-test on actual sales of 120 styles/colours against the retailer's rule; compared with simulation-based optimisation
Caro and Gallien [3]	Piecewise-linear approximation; MIP	AMPL; CPLEX 10.0	Live pilot at Zara, July to November 2006; difference-in-differences against matched control references
Jaiswal et al. [6]	Greedy trailer assignment; optimal transport (DRM); final LP	CPLEX; GLOP (OR-Tools); Brent's root-finding method	19 real replenishment cycles; seven model variants; compared with the exact MILP optimum
Cruz et al. [2]	Exact integer LP	Excel Solver	Compared with past dispatch decisions; data cross-checked through six site visits, interviews and documents
DeHoratius et al. [4]	Bayesian update; critical-ratio order rule; EVPI audit rule	Not named	Simulation of 500 sample paths over 90 days (360 for audits), calibrated to real audit and sales data; sensitivity runs

IV. RESULTS AND DISCUSSION

A. Results and Findings

Table VII summarises the main result of each paper against the benchmark its authors chose.

Table VII: Key Results Reported in the Five Papers

Paper	Compared against	Main result
Fisher et al. [5]	The retailer's existing ordering rule	Profit higher by 2.23% to 4.92% of current sales (for a backorder cost of \$15 down to \$5), where pre-tax profit is about 3% of sales; heuristic runs in under a minute against about 40 hours for simulation-based optimisation
Caro and Gallien [3]	Old manual process, through matched control references	Sales higher by an estimated 3 to 4% (4.1% measured at Arteixo against a 0.7% error measured at Zaragoza); fewer transshipments; more time on display
Jaiswal et al. [6]	A plan covering the immediate need only (the authors' own baseline)	Labour utilisation from 61.73% to 96.45%; trailer utilisation from 42.61% to 79.69%; same allocation as the exact MILP in 17 of 19 cycles
Cruz et al. [2]	Past dispatch by intuition	342 of 427 pallets shipped in 21 truck trips and 4 van-truck trips; 86.18% of truck minutes and 88.30% of van-truck minutes used; 85 pallets left over; the paper claims 26 pallets more than past dispatch (p. 639), which this study finds doubtful (Section IV-C)
DeHoratius et al. [4]	A "naive" policy that trusts the record; zero-balance-walk audits	20% lower holding cost for a 93% fill rate; no "freezing"; Bayesian ordering with EVPI audits gives the lowest total cost

Four points lie behind these figures. In Fisher et al. [5], the model and the retailer bought almost the same total quantity: at a backorder cost of \$10, 21,946 units against 21,866. The model placed a smaller first order (18,015 against 19,050) and a larger reorder (3,931 against 2,816), so the gain comes from moving purchases to the point where better information is available.

In Jaiswal et al. [6], pull-forward alone is not enough. Without the minimum-load penalty, trailer utilisation

was only 48.11% and the plan needed 1.92 times as many trailers as the exact MILP solution.

In Caro and Gallien [3], the benefit differed by product type. Basic references gained mainly in demand cover (10.1%), fashion references in shipment success (6.4%). After the pilot, Zara adopted the model for all its stores and references.

In DeHoratius et al. [4], “freezing” is the case in which the record shows stock while the shelf is empty, so the system never reorders. The benchmark audit policy, the zero-balance walk, checks the shelf only when the record shows zero. The Bayesian record avoids freezing because days without sales raise the probability that the shelf is empty (Section III-B).

B. Practical Implications

More efficient use of existing resources. None of the improvements required additional physical assets. Zara’s additional revenue came from allocating the same stock better, and INVALS loaded the same trailers and used the same warehouse labour more fully. Cruz et al. [2] conclude that the vehicle fleet is appropriately sized for this problem, but their own results do not confirm this: shipping all 427 waiting pallets with single-store trips needs at least 7,514 vehicle-minutes, more than the 5,280 available (Section IV-D). This is possible only if far fewer pallets arrive each day than are waiting each morning, a figure the paper does not report.

Modelling provides advice; people make the decisions. At Zara, the model’s shipments were presented to warehouse staff as recommendations that they could override. Once the initial iterations had passed, overrides became rare, and the staff’s work shifted from data entry to handling exceptions.

The models place a monetary value on decisions made by operations managers. Fisher et al. [5] use their model to value a shorter lead time, which shows whether a closer but more expensive supplier is worth choosing over a more distant, cheaper one. **Some models need no specialist software.** The model of Cruz et al. [2] runs in Excel Solver.

C. Critical Analysis

1) Why the OR methods work here

The problem has the structure OR needs: a decision repeated every day or week, a goal that can be measured, limits that can be written as inequalities, and data to feed the model. The more interesting finding is what the papers do when the exact model is too hard or too slow for daily use, as it is in four of the five: a cost function that is neither convex nor concave (Fisher), dynamic programmes with far too many states (Caro and Gallien; DeHoratius) and an NP-hard MILP (Jaiswal) that is solved exactly only as an offline benchmark. In each case the authors replace the exact model with a simpler one.

Two of them measure what this costs, and the loss is small: within 1% for Fisher’s heuristic in the case the authors highlight, where early and late demand are strongly correlated, and the same allocation as the exact MILP in 17 of 19 cycles for INVALS. Caro and Gallien [3] and DeHoratius et al. [4] test their simplified models against existing practice or simple policies instead, so how far they fall short of the true optimum is not known. The practical value of OR here lies less in exact optimality than in a good approximation that has been properly tested.

The papers also share building blocks. The newsvendor critical ratio appears as $(C_u - C_b)/(C_u - C_b + C_o)$ in Fisher et al. [5] and as $p/(h + p)$ in DeHoratius et al. [4]. Linear constraints on stock and capacity are shared by the Cruz, Jaiswal and Caro models.

2) Strengths

The strongest research design is in Caro and Gallien [3]: a controlled field experiment in a working company, with a second warehouse acting as a check on the measurement. The 4.1% gain measured at Arteixo, less the 0.7% error measured at Zaragoza, leaves 3.4 points, consistent with the estimate of 3 to 4%. The design is strong even though the sample is small (see below). Their model also captures real store behaviour, the major-size display rule, which a standard allocation model would miss. Fisher et al. [5] handle demand that has no sales history by using the spread of four buyers’ forecasts as a measure of uncertainty. DeHoratius et al. [4] show that their parameters can be estimated from audit data a retailer already holds, and a retailer with perfect information still does better in their tests, so the case for preventing errors remains. INVALS is built for scale. The model of Cruz et al. [2] is the easiest to understand and to copy.

3) Limitations and assumptions

The evidence is uneven. Only Zara’s model was tested live. Fisher et al. [5] and Jaiswal et al. [6] replay past data, Cruz et al. [2] plan a single average day, and DeHoratius et al. [4] rely on simulation of one product. The sources also differ in status. Two were read as journal articles. Two were read as working papers that were later published in *Operations Research* (2010) and *Manufacturing & Service Operations Management* (2008). One, Jaiswal et al. [6], is a preprint. All figures in this study come from the versions read, and the published versions may differ in detail.

Each model also leans on assumptions that may not hold elsewhere:

Cruz et al. [2]. Demand is fixed and trips go out only as full loads. The plan therefore leaves 85 pallets (19.91%) behind while about 14% of truck time and 12% of van-truck time is unused. The full-load rule, not vehicle time, causes this: re-solving the model with the paper's data, full loads alone can never ship more than 348 of the 427 pallets, even with unlimited vehicle time. Allowing part-loads with the same fleet and minutes raises the optimum to 379 pallets, 37 more (+10.82%; Section IV-D). Routes with several stops are not considered, a limit the authors acknowledge.

Jaiswal et al. [6]. Demand is treated as certain. The objective is a score whose weights are confidential, so the result cannot be reproduced or converted into money. Submodularity, on which the greedy step rests, is supported by 7,600 test cases but not proved.

Caro and Gallien [3]. The estimate rests on ten references and has no confidence interval. The model maximises revenue, not profit, and gives no rule for choosing K .

Fisher et al. [5]. Results come from one department in one season. The backorder cost C_b had to be estimated within a range of \$5 to \$15, and the profit gain more than halves across that range. With pre-tax profit at about 3% of sales, the gains imply profit multiples of $1.74\times$ at \$15, $2.17\times$ at \$10 and $2.64\times$ at \$5, so the abstract's claim of an improvement "by enough to double profits" [5, p. 230] holds at \$10 and \$5 but not at \$15.

DeHoratius et al. [4]. All sources of error are merged into one process, and the model covers one product with no substitution.

A check of the published numbers. Working through the tables gave three points. First, Table 4 of Cruz et al. [2, p. 639] adds up to 342 pallets shipped, which agrees with the 342 stated on p. 638, yet the text below the table says the model "sent 427 pallets to the stores, 26 pallets more than the operation performed on the DC in question" (p. 639). The 427 is the total demand, not the shipment. Re-solving the model as stated, in a check carried out for this study (Python, SciPy with the HiGHS solver; Section IV-D), gives the same 342, so the optimum itself is correct, but the claimed gain rests on the wrong figure. Read literally, past dispatch was about 401 pallets ($427 - 26$), so the model's actual 342 would

be 59 fewer than the dispatchers achieved. If the 26 was measured from 342, past dispatch was 316. The first reading is not even possible inside the model, whose LP relaxation ships at most 399.36 pallets, so the dispatchers must have used part-loads, multi-stop trips or extra hours that the model rules out. Either way, the paper does not show that the model beats current practice. Second, the profit-increase row in Table 1 of Fisher et al. [5, p. 239] can be reproduced if the current rule's profit is re-costed at each value of C_b ; for C_b of \$10, $[445,384 - (431,696 - 6,534 \times 5)] \div 1,317,889 = 3.52\%$. The paper does not state this step, so it is an inference. Third, the gains of 34.70% and 37.08% quoted in the abstract of Jaiswal et al. [6] are percentage-point differences between the utilisation figures on p. 6, not relative increases: $96.45 - 61.73 = 34.72$, about the 34.70 quoted, and $79.69 - 42.61 = 37.08$. In relative terms, the increases are 56.24% and 87.02%.

4) Improvements, extensions and future applications

The clearest extension comes from reading the papers together. The allocation models of Caro and Gallien [3] and Jaiswal et al. [6] both take store inventory as a known input. Evidence cited by DeHoratius et al. [4] shows how wrong this input can be: 65% of records at one retailer were inaccurate (DeHoratius and Raman, 2004, cited in [4]). Feeding the Bayesian inventory record into an allocation model, in place of the system figure, would join the two lines of work.

Other extensions follow from the limits above. INVALS could take demand uncertainty into account, which its authors list as future work together with a multi-echelon version. The Cruz model could allow part-loads, which alone raise the optimum from 342 to 379 pallets (Section IV-D), and multi-stop routes, which would turn it into a vehicle routing problem. With a shorter lead time, which Section IV-B notes the model can price, Fisher's model could allow a second reorder. Zara's model could maximise profit and add the multi-colour version that the authors formulated but had not yet implemented.

The same structure, a central stock shared among many outlets under uncertain demand, also appears in vaccine distribution, food grain distribution under the PDS and hospital resource allocation, so the ideas reviewed here could be adapted to them.

D. Re-solving the Model of Cruz et al. [2]

The model of Section III-B was re-solved for this study with the store data in Tables 1 and 2 of Cruz et al. [2], using Python (the milp function of SciPy, which calls the HiGHS solver). The optimum matches the paper: 342 pallets in 21 truck trips and 4 van-truck trips, using 3,792

of 4,400 truck minutes (86.18%) and 777 of 880 van-truck minutes (88.30%). Table VIII gives one optimal plan; other plans can reach the same total.

Table IX changes one assumption at a time to show what limits the plan.

Table VIII: One Optimal Dispatch Plan (Data from Cruz et al. [2, Tables 1 and 2])

Store	Cycle time (min)	Demand (pallets)	Truck trips	Van-truck trips	Pallets shipped	Pallets left
1	125	16	1	0	14	2
2	197	23	1	0	14	9
3	196	39	1	2	38	1
4	169	26	1	1	26	0
5	118	28	2	0	28	0
6	199	49	3	0	42	7
7	166	19	1	0	14	5
8	176	29	2	0	28	1
9	197	21	1	0	14	7
10	196	23	1	0	14	9
11	202	31	2	0	28	3
12	157	11	0	0	0	11
13	153	15	1	0	14	1
14	208	18	1	0	14	4
15	179	15	1	0	14	1
16	189	15	1	0	14	1
17	185	8	0	0	0	8
18	216	13	0	1	12	1
19	228	18	1	0	14	4
20	188	10	0	0	0	10
Total	–	427	21	4	342	85

Table IX: What Limits the Plan

Case	Pallets shipped	What it shows
Model as stated: full loads only; 4,400 truck and 880 van-truck minutes	342	85 pallets (19.91%) left behind
Full loads only; unlimited vehicle time	348	79 pallets can never go as full loads, so time explains only 6 of the 85
Part-loads allowed; same minutes	379	37 more pallets (+10.82%) with the same fleet
Fractional trips (LP relaxation); same minutes	399.36	Upper bound for any plan with single-store trips
Ship all 427 pallets with single-store trips	–	Needs at least 7,514 vehicle-minutes, 42.31% more than the 5,280 available

Sources of the data: store cycle times and pallet demand from Cruz et al. [2, Tables 1 and 2]; calculations by the authors of this study.

V. CONCLUSION

The five studies follow stock from the supplier's order to the store shelf, each applying a different OR technique to its own stage: stochastic DP for ordering, mixed-integer programming for allocation, integer LP for shipping, and Bayesian updating for the shelf record. In four of the five papers the model beats the benchmark its authors chose,

from profit gains of 2.23% to 4.92% of sales at the catalogue retailer to an estimated 3 to 4% more sales at Zara. Only Zara's benchmark was existing practice measured live, and Cruz et al. [2] do not show a clear gain over past dispatch.

Two lessons stand out. The gains came from using existing stock, staff and vehicles better, not from new physical assets. And none of the original problems was solved exactly in daily use: each relied on a well-chosen simplification, usually the newsvendor trade-off or a



linear approximation, that was then tested against data. The quality of that testing is what separates the strongest papers from the weakest. Between the two primary papers the contrast is sharp: Caro and Gallien [3] tested their model live against a control, while Cruz et al. [2] give the clearest model but no reliable evidence that it beats past dispatch.

ACKNOWLEDGMENT

The authors thank Ms. Tejaswini Angre for guidance on this study.

REFERENCES

- [1] H. A. Taha, “Operations Research: An Introduction,” 10th ed.: Pearson, 2017.
- [2] R. S. Cruz, E. R. Viana, A. M. da Silva, and A. das N. Gomes, “Optimization model for materials expedition: An application [sic] in a retail chain store,” *Independent Journal of Management & Production*, vol. 7, no. 5, pp. 627–642, 2016, doi: 10.14807/ijmp.v7i5.454.
- [3] F. Caro and J. Gallien, “Inventory management of a fast-fashion retail network,” MIT Sloan School of Management, Cambridge, MA, USA, Working Paper 4656-07, 2007. [Online]. Available: <http://ssrn.com/abstract=1007075>. Later published in *Operations Research*, vol. 58, no. 2, pp. 257–273, 2010, doi: 10.1287/opre.1090.0698.
- [4] N. DeHoratius, A. J. Mersereau, and L. Schrage, “Retail inventory management when records are inaccurate,” unpublished manuscript, Graduate School of Business, The University of Chicago, Chicago, IL, USA, Nov. 10, 2005. Later published in *Manufacturing & Service Operations Management*, vol. 10, no. 2, pp. 257–277, 2008, doi: 10.1287/msom.1070.0203.
- [5] M. Fisher, K. Rajaram, and A. Raman, “Optimizing inventory replenishment of retail fashion products,” *Manufacturing & Service Operations Management*, vol. 3, no. 3, pp. 230–241, 2001, doi: 10.1287/msom.3.3.230.9889.
- [6] S. K. Jaiswal, K. S. Gurumoorthy, E. Agarwal, and S. Manchenahally, “INVALS: A forward looking inventory allocation system,” *arXiv:2310.04305v2*, 2025.