Operations Research

In This Issue

To cite this article:
(2026) In This Issue. Operations Research 74(4):v-ix. https://doi.org/10.1287/opre.2026.inthisissue.v74.n4

https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 74, No. 4, July–August 2026, pp. v–ix
ISSN 0030-364X (print), ISSN 1526-5463 (online)

In This Issue

https://doi.org/10.1287/opre.2026.inthisissue.v74.n4

Smarter Labeling to Detect Hidden Human Trafficking Risks

Human trafficking investigators face the immense challenge of sifting through vast amounts of online data to uncover illicit activities. In "Reducing Manual Labeling Effort in Imbalanced Data Sets: Active Learning for Detecting Illicit Massage Business Reviews," Tobey, Mayorga, Bosisto, and Özaltın present a novel framework that uses reinforcement learning–based active learning to reduce the burden of manual data labeling, improving detection of illicit massage business reviews on Yelp. By strategically selecting the most informative reviews for expert annotation, the approach achieves strong performance despite limited and imbalanced data sets, easing the emotional and time costs of reviewing disturbing content. The study demonstrates that their method outperforms benchmark active learning strategies, remains effective even with large query batches, and generalizes across regions. Beyond combating human trafficking, the framework offers a scalable solution for other domains with scarce, sensitive, or costly-to-label data.

Properties of Nash Flows over Time

In "Continuity, Uniqueness, and Long-Term Behavior of Nash Flows over Time," Olver, Sering, and Vargas Koch study a dynamic model of traffic based on the Vickrey bottleneck model. In this model, a continuous flow of users travels over time through a network from an origin to a destination, aiming to minimize their travel time. Equilibrium flows in this model, in which no user has a better route available to the user considering delays because of the other users, have recently received significant attention. The paper answers several basic questions that have remained open in the single-commodity setting. It shows that equilibrium travel times are unique. Moreover, it proves that equilibria are continuous, meaning that small changes in the network parameters or starting conditions do not lead to drastically different outcomes. Surprisingly, the understanding of the long-term behavior of equilibria turns out to be crucial.

Stabilizing the Unstable: How Long-History PCA Sharpens Portfolio Risk Forecasts

Investors focus on market volatility, but a more subtle challenge is "second-order risk"—the discrepancy between a portfolio's predicted risk and its actual performance. Standard practice relies on short snapshots of data (typically one year) to estimate the covariance structure under the assumption that older data are irrelevant in fast-changing markets. However, this "short-memory" approach often mistakes random noise for real market signals, leading to optimized portfolios that are riskier than they appear. In "Long-History Principal Component Analysis in a Dynamic Factor Model with Weak Loadings," Anderson, Kim, and Ryu challenge this norm. They introduce Long-History Principal Component Analysis (LH-PCA), demonstrating that extending the data history to six years acts as a crucial stabilizer. Theoretically, the authors prove PCA remains a consistent estimator even in dynamic markets with "weak" loadings, provided the historical timeline is sufficiently large. Both simulation and empirical tests on U.S. and European markets confirm that looking back six years significantly reduces estimation bias.

Fair Fares for Vehicle Sharing Systems

Vehicle sharing systems—such as bike, scooter, and car sharing—play a key role in urban transportation, yet algorithmic pricing can lead to inequalities across locations. In "Fair Fares for Vehicle Sharing Systems," Elmachtoub and Kim study how platforms can incorporate fairness into pricing decisions and the resulting outcomes. The authors propose two notions of fairness: price fairness, which limits price differences across locations, and access fairness, which equalizes the proportion of demand at each location that has access to the system, determined by affordability and availability. Using a stylized two-location model, the authors analyze how fairness constraints affect platform revenue, consumer surplus, and social welfare. Although price fairness can sometimes increase consumer surplus at both locations, access fairness always reduces consumer surplus at both locations. A convex relaxation approach is developed for larger networks, and a case study using New York City data illustrates the tradeoffs between fairness and efficiency in practice.

Smarter Branching Speeds Up Leading Exact Vehicle Routing Solvers

Researchers have introduced a learning-based branching strategy that substantially accelerates exact algorithms for vehicle routing problems (VRPs), a core challenge in logistics and transportation systems. In "Two-Stage Learning to Branch in Branch-Price-and-Cut Algorithms for Solving Vehicle Routing Problems Exactly," You, Yang, Wang, and Yin propose the first learning-to-branch framework tailored for branch-price-and-cut methods, where dynamic variables and dense constraints make traditional branching decisions computationally expensive. The novel two-stage learning-based branching (2LBB) approach effectively filters promising candidates using inexpensive features and then applies selective, partial testing to reduce costly evaluations. A theoretical model further guides dynamic adjustment of branching effort, balancing decision time with solution quality. Extensive experiments show runtime reductions of 45%–50% on standard capacitated VRP (CVRP) and VRP with time windows (VRPTW) benchmarks and a 47% speedup over the state-of-the-art VRPSolver when integrated into the open-source RouteOpt. These results highlight the growing potential of disciplined machine learning to enhance exact optimization algorithms.

Pricing Based on Matching Outcomes Improves Shared Rides Operations

Shared rides aspire to mitigate congestion and promote more sustainable urban transportation, but many platforms struggle to maintain a healthy and profitable shared rides product. Why have shared rides struggled, and how can platforms design a successful product? In "Pricing Shared Rides," Yan, Yan, and Shen discuss the drawbacks of platforms' existing pricing policies and show that a common approach to pricing suppresses demand and may even drive the platform out of the market. They then propose adjusting the price according to the extent she is matched with another rider—which proves to benefit both platform and riders alike in increased profitability and reduced payments. These benefits are especially pronounced in regimes with high cost and low demand, expanding shared rides access to areas where they have historically been most challenging to sustain.

Coordinating Dual-Sourcing for Multicomponent Assembly Systems

Assembly manufacturers are increasingly relying on dual sourcing—pairing a low-cost regular supplier with a faster but more expensive express option—to enhance supply chain resilience. Yet although dual sourcing is well understood for single items, managing it across multiple components in an assembly system has remained largely unresolved. In "Dual Sourcing in an Assembly System: A Simple and Effective Critical-Set Base-Surge Policy," Deng, Shao, Song, and Yu introduce a novel Critical-Set Base-Surge (CSBS) policy, which combines a constant order policy for regular sources to meet base demand with a Critical-Set policy for express sources to handle surge demand. By revealing a weaker form of additive separability and proving exponential state convergence under CSBS, they provide a performance bound and establish asymptotic optimality as regular lead times grow. Numerical experiments show that, compared to benchmark policies that apply the Tailored Base-Surge policy, the CSBS policy achieves substantial optimality gap reductions. Furthermore, we propose a CSBS policy with easy-to-compute parameters and demonstrate that it delivers performance comparable with the CSBS policy.

Context-Dependent Choice and Retail Decisions

Traditional assortment models assume that consumers evaluate products independently of the alternatives available (i.e., the "context"). In "Assortment and Price Optimization Under a Multiattribute (Contextual) Choice Model," Najafi, Jasin, Uichanco, and Zhao challenge this assumption by analyzing assortment and pricing decisions under a context-dependent choice framework known as the contextual concavity (CC) model. The CC model incorporates reference dependence across multiple attributes, such as price and quality, and captures well-documented context effects, including compromise and decoy effects. The study makes several contributions. It characterizes the structure of optimal assortments under multiattribute loss aversion, develops a polynomial-size mixed-integer linear programming formulation for solving the general problem, and analyzes the joint assortment and pricing decision. Numerical experiments show that ignoring context effects, by relying on standard context-independent models such as the multinomial logit, can lead to substantial profit losses, with gaps ranging from 3% to 63%. These findings highlight the strategic importance of incorporating contextual effects into retail decisions.

No-Regret Bayesian Recommendation to Homogeneous Users

In "No-Regret Bayesian Recommendation to Homogeneous Users," Feng, Tang, and Xu introduce and study the online Bayesian recommendation problem for a recommender system platform. The platform has the privilege to privately observe a utility-relevant state of a product at each round and uses this information to make online recommendations to a stream of myopic users. This paradigm is common in a wide range of scenarios in the current internet economy. The platform commits to an online recommendation policy that utilizes its information advantage on the product state to persuade self-interested users to follow the recommendation. Because the platform does not know users' preferences or beliefs in advance, the authors study the platform's online learning problem of designing an adaptive recommendation policy to persuade users while gradually learning users' preferences and beliefs en route. Specifically, the authors aim to design online learning policies with no Stackelberg regret for the platform, that is, against the optimal benchmark policy in hindsight under the assumption that users will correspondingly adapt their responses to the benchmark policy. Their first result is an online policy that achieves double logarithmic regret dependence on the number of rounds. They also present an information-theoretic lower bound showing that no adaptive online policy can achieve regret with better dependency on the number of rounds. Finally, by formulating the platform's problem as optimizing a linear program with membership oracle access, the authors present their second online recommendation policy that achieves regret with polynomial dependence on the number of states but logarithmic dependence on the number of rounds.

Learning to Optimize Assortment and Inventory Decisions with Unknown Demand

How should a retailer coordinate product assortment and inventory when customer preferences need to be learned on the fly and stockouts dynamically reshape demand? In "Online Joint Assortment-Inventory Optimization Under MNL Choices," Liang, Mao, and Wang study an online learning setting where customers follow a multinomial logit choice model with unknown parameters. The retailer repeatedly selects assortments and inventory levels across sales cycles, using observed choice data to update decisions, aiming to maximize the cumulative profit. Because stochastic stockouts induce complex substitution patterns, preference learning and decision making are intrinsically difficult. The paper introduces a new exploration–exploitation algorithm that combines a novel estimator for customer preferences, an adaptive mechanism that promotes informative experimentation, and an optimization oracle for inventory and assortment planning. The authors establish nearly optimal regret bounds and show that the algorithm remains effective when practical approximate optimization methods are used. They also extend the framework to incorporate inventory carryover and unknown arrival processes, with numerical studies confirming the strong performance of the proposed approach.

Breaking the $\mathcal{O}(\sqrt{T})$ Barrier in Online Linear Programming

How can fast first-order methods for online linear programming (OLP) match the performance of more computationally intensive LP-based approaches? In "Beyond $\mathcal{O}(\sqrt{T})$ Regret: Decoupling Learning and Decision Making in Online Linear Programming," Gao, Ge, Sun, Xue, and Ye develop a new framework that improves the regret guarantees of scalable first-order algorithms for online linear programming. They show that when the dual problem satisfies a mild error bound condition, first-order methods can achieve $o(\sqrt{T})$ regret, including $\mathcal{O}(T^{1/3})$ in continuous support settings and $\mathcal{O}(\log T)$ in finite support settings. The key insight is to decouple learning and decision making. The algorithm first learns an approximate dual solution and then localizes decisions to a neighborhood around that solution using an adaptive procedure. This decoupling narrows the gap between efficient gradient-based methods and LP-based benchmarks, providing strong theoretical guarantees while maintaining scalability for applications such as revenue management, online advertising, and cloud computing.

When Thousands of Items Arrive Every Hour: A New Approach to Online Experimentation

Researchers have reported a new approach to address a key challenge for online platforms. In "Short-Lived High-Volume Bandits," Jia, Li, Ravi, Oli, Duff, and Anderson introduce the Short-lived High-volume Bandits (SLHVB) framework. This models modern platforms where thousands of items—such as ads, stories, or interface designs—arrive each hour but remain available only briefly. The study develops a near-optimal online learning policy that balances exploration and exploitation. Theoretical analysis shows the algorithm achieves nearly the minimal possible loss as the number of user impressions grows. The team tested the policy in a large-scale field experiment with Glance, a leading lock-screen content platform. The policy increased viewing duration by 4.32% and click-through rates by 7.48% compared to the platform's existing deep-learning-based recommender system.

Mutually Fair Redistricting Even When Parties Disagree

Congressional redistricting is the process of partitioning a state into districts, each of which elects a representative to Congress. Several recent high-profile redistricting efforts aim to increase the political power of a party. This raises the question of whether "fair" redistricting plans exist. In "You Can Have Your Cake and Redistrict It Too," Benadè, Procaccia, and Tucker-Foltz propose a new theoretical model for redistricting inspired by classical cake-cutting models. In this model, it shown that is always possible to find redistricting plans that satisfy a particular notion of fairness, called the geometric target, simultaneously for both parties, even when the parties disagree about voter preferences. On real-world data, they find that this fairness constraint can be satisfied in all instances evaluated; moreover, requiring fairness comes at little cost in terms of traditional redistricting objectives. This suggests it is possible and practical to guarantee mutual fairness even in a climate of extreme partisanship.

On the Complexity of Finding Locally Optimal Solutions in Bilevel Linear Optimization

In "On the Complexity of Finding Locally Optimal Solutions in Bilevel Linear Optimization," Prokopyev and Ralphs study the computational complexity of finding *locally* optimal solutions to bilevel linear optimization problems (BLPs), from the leader's perspective. While BLPs are known to be *NP*-hard to solve globally, it was open whether a locally optimal solution could be found efficiently. The authors show this is not the case: for any constant $c > 0$, finding a leader's solution within Euclidean distance c/n of any locally optimal solution is *NP*-hard, where n is the number of variables. The proof adapts techniques from an analogous result for quadratic programs [Ahmadi AA, Zhang J (2022) On the complexity of finding a local minimizer of a quadratic function over a polytope. *Math. Programming*, 195(1–2):783–792], reducing the maximum independent set problem to a specially constructed BLP. Along the way, the authors derive a BLP reformulation of the classical Motzkin-Straus quadratic formulation of the maximum clique problem, highlighting a deep structural connection between combinatorial, quadratic, and bilevel optimization. The results show that finding a locally optimal solution is essentially as hard as finding global optima in BLPs.

A Risk-Based Lens on Online Allocation Algorithms

In "Risk Minimization as a Framework for Online Allocation in Display Advertising," Shamsi, Luenberger, and Ye revisit a foundational approach to real-time resource allocation. The study draws a conceptual bridge between earlier risk minimization models and more recent dual mirror descent methods, two influential paradigms in online optimization. Although both frameworks generate similar exponential price update rules, this research shows they arise from distinct modeling perspectives: one grounded in convex risk minimization and the other in Bregman divergence and mirror descent. By framing dual updates through a risk-aware lens, the authors recover widely used allocation strategies—including greedy, linear, and multiplicative weights—as special cases. The paper broadens the theoretical toolkit for online resource allocation and highlights how incorporating uncertainty directly into the optimization framework can yield flexible, interpretable, and robust algorithms across stochastic and adversarial settings.

A New Structural Insight into Stockout-Based Substitution

Stockout-based substitution creates complex stochastic dynamics in inventory systems even in highly symmetric settings. In "Optimal Inventory Allocation for Indifferent Goods Under Dynamic Substitution," Zhou, Wang, and Zhang examine a canonical setting in which a firm allocates a fixed inventory across multiple perfectly substitutable product types (e.g., colors or designs) and customers purchase uniformly from available options. Despite the model's symmetry, the induced stochastic dynamics are subtle and resist classical convexity or induction-based arguments. The authors introduce a continuous-time embedding of the discrete sales process, converting inventory depletion times into independent Erlang random variables. This probabilistic reformulation enables an exact characterization of marginal values and yields a clean structural result: the optimal allocation is balanced. Beyond solving the symmetric case, the method extends naturally to finite-horizon multinomial logit models, offering new analytical insight into fluid relaxations and dynamic substitution.

Smarter Matchmaking for Ride-Hailing Platforms

Ride-hailing platforms such as Uber, Lyft, and DiDi must assign drivers to riders every minute without knowing who will request a ride next. Most systems optimize each batch myopically, and this can leave some future riders waiting much longer than necessary. In "Matching Drivers to Riders: A Two-Stage Robust Approach," El Housni, Goyal, Hanguir, and Stein propose a two-stage robust matching model that incorporates uncertainty about future demand. The first stage matches current riders, reserving enough nearby drivers for plausible future scenarios; the second stage then tests these decisions against the most adverse demand patterns to guarantee good performance even in the worst case. The authors prove that finding the best such policy is computationally hard, but they develop constant factor approximation algorithms for several practically relevant cases and test them on large-scale taxi data from Shenzhen, China. Their methods substantially reduce maximum rider waiting times with little or no sacrifice in average travel distances.

Stream SGD: Fast Learning and Valid Inference from Dependent Data

Many online optimization problems in operations research rely on data generated by Markovian systems whose dynamics depend on the decision parameters, creating both statistical dependence and biased gradient information. In "Convergence and Inference of Stream Stochastic Gradient Descent, with Applications to Queueing Systems and Inventory Control," Li, Liang, Chen, and Zhang develop a unified theory for stream stochastic gradient descent (SGD), a sample-efficient method that uses just one observation per iteration. Using Poisson-equation techniques, the authors quantify and control gradient bias and dependence, proving an optimal $\mathcal{O}(T^{-1})$ convergence rate and a state-of-the-art $\mathcal{O}(\log T)$ regret bound. Beyond optimization performance, the paper introduces an online inference framework for uncertainty quantification and establishes a functional central limit theorem that underpins valid asymptotic inference. A new Wasserstein-type divergence yields verifiable conditions via coupling arguments tailored to operations research models. Applications to queueing and inventory problems demonstrate how the theory translates into practical, scalable algorithms.

End-to-End Optimization for Queues Beyond PTO

Most queueing optimization methods rely on a predict-then-optimize (PTO) paradigm, which first estimates key model primitives (e.g., demand rates and service-time distributions) and then optimizes system decisions using these estimates as if they were exact. In practice, however, queueing performance formulas are often highly sensitive to estimation errors, especially under congestion, making such approaches fragile and potentially misleading. In "Online Learning and Optimization for Queues with Unknown Arrival Rate and Service Distribution," Chen, Hong, and Liu develop a new end-to-end online learning framework for jointly optimizing pricing and capacity decisions in service systems, called LiQUAR (learning in queue with unknown arrival rate). LiQUAR is specialized to queueing systems by explicitly leveraging workload dynamics and transient behavior, allowing it to learn directly from arrival and service data without requiring prior knowledge of the demand function or the service-time distribution. Designing online learning algorithms for queues poses unique challenges: Data are temporally correlated, system dynamics are disrupted by policy updates, and performance is highly sensitive to congestion. LiQUAR addresses these challenges through queue-aware algorithm design and analysis, establishing regret bounds that capture the effect of traffic intensity and demonstrating superior performance to PTO and gradient-based reinforcement learning methods, particularly in heavily loaded systems.

Sharper Approximation for Multistage Decisions Under Uncertainty

In "A Note on Piecewise Affine Decision Rules for Robust, Stochastic, and Data-Driven Optimization," Thomä, Schiffer, and Wiesemann revisit piecewise affine decision rules, a widely used approximation for multistage stochastic programs. Building on the framework of Georghiou et al. [Georghiou A, Wiesemann W, Kuhn D (2015) Generalized decision rule approximations for stochastic programming via liftings. *Math. Programming* 152(1–2): 301–338], the authors propose an algorithmic refinement that yields strictly better policies in stochastic settings while retaining tractability. Beyond stochastic programming, the framework naturally extends to multistage robust optimization and to modern data-driven models based on Wasserstein ambiguity sets. The paper shows how the resulting policies can be computed efficiently and provides numerical evidence demonstrating consistent performance improvements.

Novel Insights on Robust Markov Decision Processes with Average Reward and Blackwell Optimality Criteria

Robust Markov decision processes (RMDPs) have been studied extensively when the objective is the discounted return, but little is known for average optimality and Blackwell optimality. In "Beyond Discounted Returns: Robust Markov Decision Processes with Average and Blackwell Optimality," Grand-Clément, Petrik, and Vieille show that average optimal policies can be chosen stationary and deterministic for sa-rectangular RMDPs, but perhaps surprisingly, they show that for s-rectangular RMDPs average optimal policies may not exist, and if they do exist, they may not be stationary. The authors also study Blackwell optimality for sa-rectangular RMDPs, where they show that approximately Blackwell optimal policies always exist, although exact Blackwell optimal policies may not exist. They provide a general sufficient condition for their existence. They then discuss the connection between average and Blackwell optimality, and they describe several algorithms to compute the optimal average return. Interestingly, the approach leverages the connections between RMDPs and stochastic games. Overall, the authors emphasize the superior practical properties of distance-based sa-rectangular models over s-rectangular models for average and Blackwell optimality.

Stochastic Gradient Descent with Adaptive Data

Stochastic gradient descent (SGD) is a central tool in modern optimization, but its classical theory relies on the assumption that data are independent of the decisions being optimized. In many operations research settings, this assumption fails: policies influence system dynamics, and the resulting data feed back into subsequent updates. In "Stochastic Gradient Descent with Adaptive Data," Che, Dong, and Tong address this challenge by developing a general framework for analyzing SGD when data are generated adaptively by policy-dependent Markov processes. Their analysis shows that fully adaptive SGD can still attain convergence rates comparable to the classical i.i.d. setting, provided the underlying system satisfies mild ergodicity and continuity conditions. The theory is illustrated through canonical applications in operations research and reinforcement learning. Overall, the paper provides rigorous and reassuring theoretical foundations for deploying learning algorithms in dynamic environments where decisions and data are fundamentally intertwined.

A New Perspective on Dynamic Programming Operators

Dynamic programming is a central tool in operations research and economics, but many classical results rely on strong assumptions such as contraction that are not satisfied in important applications. In "Dynamic Programming in Ordered Vector Space," Peng and Stachurski develop a new framework for dynamic programming in ordered vector spaces, allowing for a broader class of operators and preference structures. By exploiting order-theoretic properties rather than metric contraction, the authors establish existence, uniqueness, and convergence results for value functions. The framework accommodates applications involving risk-sensitive preferences, nonstandard discounting, and other settings where traditional approaches fail. These results provide a unifying perspective on several recent extensions of dynamic programming and open the door to new applications in economics and operations research.

Reducing Long Waits with Optimal Nudging

Many systems traditionally use a basic first-come, first-served scheduling policy. It is simple to implement and perceived as fair to arriving jobs. In recent years, Nudge policies were introduced, showing that large delays occur less frequently by allowing small changes in the order in which jobs are served—often improving all delay quantiles. In "Tail Optimality and Performance Analysis of the Nudge*(M) Scheduling Algorithm," Charlet and Van Houdt introduce the Nudge*(M) policy and show that this policy is optimal among a broad class of Nudge-like policies. A key feature of these policies is that they do not exploit arrival time information, only the arrival order and job labels. The authors provide an explicit formula for the asymptotic tail improvement ratio and show how to numerically find the waiting and response time distributions.

The Pandora's Box Problem with Sequential Inspections

In "The Pandora's Box Problem with Sequential Inspections," Aouad, Ji, and Shaposhnik study the fundamental problem of choosing among competing uncertain alternatives when the decision maker can invest resources to reduce uncertainty. The novelty of this work lies in examining settings in which uncertainty can be reduced either through a small investment or eliminated altogether through a more substantial one. The central question is how to navigate the selection process efficiently in order to achieve a good outcome. This type of problem arises, for example, in job search, when firms must choose among a large number of candidates whose fit for a position is uncertain and when uncertainty can be reduced through a small effort (e.g., an online or artificial intelligence–assisted interview) or a more substantial one (e.g., an in-person visit). The authors show that, whereas the problem is computationally challenging in general, there are two types of indices that can be computed efficiently for each alternative and that capture the attractiveness of a small or a large investment. These indices can guide the selection process, yielding optimal results in some cases and near-optimal results in all others.

GraphL0: Sparse Gaussian Graphical Models with Discrete Optimization

Recovering sparse dependency graphs in undirected Gaussian graphical models is a well-known problem in statistical machine learning. Given samples from a p-dimensional Gaussian distribution, the task amounts to estimating the $p \times p$ precision (inverse covariance) matrix under the assumption that only a small fraction of its entries are nonzero. In "Sparse Gaussian Graphical Models with Discrete Optimization: Computational and Statistical Perspectives," Behdin, Chen, and Mazumder introduce GraphL0. GraphL0 is an estimator based on an ℓ_0-penalized pseudo-likelihood, departing from the more common ℓ_1 relaxation. The resulting formulation is a convex mixed-integer program, which becomes challenging for standard commercial solvers at moderate-to-large p. To make the approach practical, the authors develop a custom nonlinear branch-and-bound algorithm, alongside scalable approximate solvers. The authors also provide new statistical guarantees for estimation accuracy and support recovery, and experiments on synthetic and real data sets show substantial computational gains over off-the-shelf solvers and competitive runtime and accuracy versus leading alternatives.

A Distribution-Free Playbook for Dual Sourcing

In "Dual Sourcing Made Easy: Distributionally Robust Optimization of Inventory Systems Under Independent Demand," Jiang, Li, Bi, Teo, and Huang study a common dual-sourcing dilemma: a low-cost supplier with a long lead time versus a faster, more expensive backup. Rather than committing to a parametric demand model, the authors propose a distributionally robust formulation that relies only on the demand mean and variance, while leveraging independence across periods. This yields a simple, closed-form way to set a tailored base–surge (TBS) policy: regular orders cover the predictable component of demand, and emergency orders buffer the remaining uncertainty, with inventory targets reflecting lead-time differences. The resulting policy is lightweight to calibrate, easy to deploy, and straightforward to stress test. In a field application using data from a multinational food manufacturer, the robust TBS settings improve service performance and reduce stockouts relative to the firm's current rule and standard benchmarks, without sacrificing cost efficiency under volatility.

Product-Form Steady-State Approximation of Generalized Jackson Networks

In "Asymptotic Product-Form Steady State for Generalized Jackson Networks in Multiscale Heavy Traffic," Dai, Glynn, and Xu revisit a long-standing question in queueing network theory. For open Jackson networks in which interarrival and service distributions are exponential, steady-state queue lengths are independent and admit a product-form distribution, making performance analysis tractable and scalable. For generalized Jackson networks with general primitive distributions, the product-form structure no longer holds. The authors show that under a multiscale heavy traffic condition—where different stations operate at widely separated levels of congestion—the stationary distribution of suitably scaled queue lengths admits an asymptotic product-form limit. In this limit, each component is exponentially distributed, with a rate that depends only on the first two moments of the primitive distributions. This result, along with its recent generalizations, creates the potential for a scalable performance analysis tool for generalized Jackson networks and beyond.

Making Robust Contextual Decisions with Causal Transport

Modern decision systems—from supply chains to financial planning—often rely on *side information*, such as customer attributes or environmental conditions, to guide better choices under uncertainty. Yet real-world data are imperfect, and naive models can fail when the underlying distribution changes. In "Decision Making with Side Information: A Causal Transport Robust Approach," Yang, Zhang, Chen, Gao, and Hu develop a new framework that integrates side information into *distributionally robust optimization* while preserving the causal structure between covariates and uncertain outcomes. The approach uses a *causal transport distance* to construct uncertainty sets that respect the conditional relationships learned from data. The authors show that the resulting worst-case distributions maintain this information structure and derive a tractable dual formulation for evaluating worst-case performance. For affine policies, the resulting optimization problem can be solved via convex programming, whereas more general settings reveal a new class of robust decision rules under convex costs.

What Makes Service Different from Computation?

Often, service systems are modeled through the same queueing theoretic frameworks that would be used in computational settings: a queue can be both a line of customers for an agent and a schedule of jobs for a machine. However, at the within-service level, queueing models are quite coarse: the duration of service is simply represented as some random variable. In "Asymmetries of Service: Interdependence and Synchronicity," Daw and Yom-Tov analyze a model of service that is more micro than macro. This history-driven interaction model illuminates the (a)symmetries in the speed and volume of work that the customer and agent each contribute to the service. Moreover, this model offers new insight into how real-world services can differ from queueing theoretic models. For instance, multitasked services have widely exhibited "inverted U-shaped" throughput as a function of the number of simultaneous customers, yet the corresponding processor-sharing queues only take on such shapes via explicit assumption. By contrast, the interaction-level model establishes conditions under which an inverted U arises; in fact, it also shows that the optimized throughput will be U-shaped as a function of the customer's relative volume of contributions. In both cases, these insights are consequences of asymmetry between the customer and agent, demonstrating how micro-level details matter for services.