Integrating qualitative data into transit service design: a stochastic estimate-then-optimize approach

Alexandre Jacquillat
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02142
alexjacq@mit.edu

Shriya Karam
Operations Research Center
Massachusetts Institute of Technology
Cambridge, MA 02142
karam809@mit.edu

Abstract

Transportation planning models are typically calibrated using coarse *numerical* data. However, these data *alone* may fail to capture travel demand patterns at a granular spatiotemporal-level, and hence, may lead to a mismatch between service levels and passenger demand. At the same time, the availability of *qualitative* data offers opportunities to better align service with true travel demand patterns. Yet, unstructured qualitative data is rarely incorporated into demand estimation and decision-making, and doing so remains an open question with no readily-available solution. To address this gap, we develop a stochastic estimate-then-optimize approach that leverages unstructured qualitative data to derive operational value. Our approach combines natural language processing for transit-specific topic modeling, a novel approach to estimate origin–destination demand based on qualitative and quantitative data, and stochastic optimization to optimize transit frequencies. Our results demonstrate that our demand estimation phase outperforms numerical data-only benchmarks and that our optimization phase improves total passenger waiting time by 7%.

1 Introduction

Transportation planning models are typically calibrated using coarse numerical data, which rely on low-granularity spatial and temporal information [1, 2]. However, current operational strategies based on coarse data *alone* may fail to capture travel demand patterns at a granular spatiotemporal-level. By building these low-dimensional representations of complex travel patterns, demand estimation methods have led to incomplete understandings of true demand, and hence, a mismatch between service levels and passenger demand. As transit agencies near the brink of a "fiscal cliff" – declining transit ridership and revenue and rising operational expenses amid the expiration of pandemic-era funding – there is an ever-pressing need to better align services to meet travel demand patterns [3].

At the same time, technology-driven solutions have emerged as a way to evaluate and respond to new urban mobility interventions. In particular, the availability of *qualitative* data offers opportunities to better align service offerings with true travel demand patterns [4]. Yet, unstructured qualitative feedback is rarely incorporated into demand estimation or decision-making models. Currently, systematically integrating these data into demand estimation or decision-making models remains an open question with no readily-available solution.

To address this methodological gap, we develop a stochastic estimate-then-optimize approach that leverages unstructured qualitative data to derive operational value. We first use natural language processing techniques for transit-specific topic modeling on tweets and customer comment cards from Washington Metropolitan Area Transit Authority (WMATA). We then develop a novel demand estimation approach that combines qualitative feedback instances, based on our topic model outputs, with coarse numerical data to estimate origin–destination demand. Leveraging the topic probabilities

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MLxOR: Mathematical Foundations and Operational Integration of Machine Learning for Uncertainty-Aware Decision-Making.

from the topic model, we characterize uncertainty in demand estimates that are subsequently used in a two-stage stochastic optimization problem to optimize rail frequencies. Our results demonstrate that our demand estimation phase outperforms numerical data-only benchmarks and that our optimization phase improves total passenger waiting time by 7%.

2 Stochastic estimate-then-optimize

2.1 Natural language processing

Our stochastic estimate-then-optimize approach uses both coarse numerical data and instances of reported overcrowding to estimate demand and subsequently optimize rail frequencies. Qualitative feedback instances, collected from Twitter and WMATA's customer resource management system [5], are used as inputs in the estimation phase to align demand with instances of reported overcrowding. To discern the qualitative feedback instances that correspond to overcrowding, we implement a topic model, described in Section A, where we classify instances into 20 transit-specific categories. We specifically use BERTopic, a natural language processing model that clusters text embeddings to identify similar feedback instances. For our demand estimation model, we consider feedback instances whose topic corresponds to "Rush Hour" in order to represent instances where demand exceeds capacity.

2.2 Demand estimation

Transit agencies oftentimes have access to aggregated ridership statistics, which can provide visibility into the total passengers entering and exiting a transit station over a time period, for instance. However, these aggregated statistics do not provide visibility origin-destination demand, which materializes at a more granular spatiotemporal level. In our approach, we therefore seek to combine coarse numerical data with qualitative data to estimate origin-destination demand at a granular level. That is, we use average daily station entrances and exits [6] along with reported instances of overcrowding to estimate demand D_p, where p represents a passenger with an origin station, destination station, and departure time. Inspired by Bertsimas and Yan [7], we use a bi-objective optimization approach to estimate D_p, where we jointly ensure consistency with aggregated station entrances and exits along with consistency in overcrowding, directly estimated from qualitative data inputs.

We denote $\mathcal{S}$ as the set of transit stations, $\mathcal{L}$ as the set of transit lines, $\mathcal{Q}_\ell$ as the set of vehicles for line ℓ, and $\mathcal{N}_{q\ell}$ as the set of transit nodes for vehicle q on line ℓ, where a transit node includes a transit station and boarding/alighting time. We describe the objective function herein for one realization of reported instances of overcrowding, with the remaining details in Section C.1.

$$\min \gamma \cdot \underbrace{\sum_{s \in \mathcal{S}} (\epsilon_s^n + \epsilon_s^x)}_{\text{Station entrances/exits}} + \cdot \underbrace{\sum_{\ell \in \mathcal{L}} \sum_{q \in \mathcal{Q}_\ell} \sum_{n \in \mathcal{N}_{q\ell}} R_{nq\ell} \cdot \left| \left((W_{nq\ell} - K)^+ - c_{nq\ell} \right) \right|}_{\text{Reported overcrowding}}$$

The objective function consists of two terms: in the first term, we minimize the total deviation between simulated and observed station entrances (ϵ_s^n) and exits (ϵ_s^x) weighted by γ. The second term minimizes the deviation between passengers exceeding capacity and reported overcrowding, where $W_{nq\ell}$ is the load at transit node n and vehicle $q \in \mathcal{Q}_\ell$ for line ℓ, and K is the vehicle capacity. The amount of reported overcrowding $c_{nq\ell}$ is directly calibrated from text instances categorized as "Rush Hour" from our BERTopic model, described in Section E.1. We further weight $c_{nq\ell}$ by $R_{nq\ell}$, which is the recurrence of reported overcrowding at node n for vehicle q.

2.3 Stochastic optimization

A key limitation of using only one realization of reported instances of overcrowding is that the topic model yields a probability distribution over topics for each instance, rather than a deterministic topic. Consequently, demand estimates derived from topic classifications may be unstable if there is significant variation in topic probabilities. This misclassification error could further propagate to our downstream optimization phase, resulting in poor quality decisions in a deterministic predict-then-optimize framework [8, 9].

In response, we leverage the topic model probabilities of each feedback instance to generate demand scenarios. To generate scenarios Ω, we consider the normalized probability of each feedback instance being categorized as a "Rush Hour" topic. To calibrate the parameters $c^\omega_{nq\ell}$ and $R^\omega_{nq\ell}$, the amount of overcrowding and recurrence of overcrowding for scenario ω, we sample from the normalized topic distribution for each feedback instance Ω times. Given $c^{\omega}_{nq\ell}$ and $R^{\omega}_{nq\ell}$ as inputs, we can estimate D^{ω}_{p} for each scenario $\omega \in \Omega$ independently.

Using estimated demand D^{ω}_{p}, we formulate a two-stage stochastic optimization problem to optimize rail frequencies, described in Section C.2. The first-stage variable selects transit frequencies for each line from a discrete frequency set, given a total fleet size constraint. The second-stage variables allocate passengers to transit lines by minimizing passenger waiting and in-vehicle travel time.

3 Results

We implement our method for a representative weekday morning peak period (5:00 - 9:30 A.M.) for the WMATA rail system. The system features six transit lines (Red, Orange, Yellow, Green, Blue, and Silver), which all run at frequencies between 5 to 11 minutes. For Sections 3.1 and 3.2, we provide results for our deterministic model, where we use feedback instances whose highest probability topic is "Rush Hour,", while Section 3.3 quantifies the benefits of incorporating the topic probability distribution. We discuss the remaining experiment details in Section E.2.

3.1 Demand estimation benchmarks

We first illustrate the impact of our demand estimation method by comparing our method to numerical data-only demand estimation methods. We consider two classical transportation demand estimation methods: the first is the maximum entropy model, detailed in Section D.1, which estimates origin-destination demand by finding the most likely demand matrix while respecting total station entrances and exits. The second is the doubly-constrained gravity model, detailed in Section D.2, which integrates travel time disutility in the maximum entropy model. We evaluate these models using three metrics: average deviation in observed entrances, exits, and overcrowding, detailed in Section D.3.

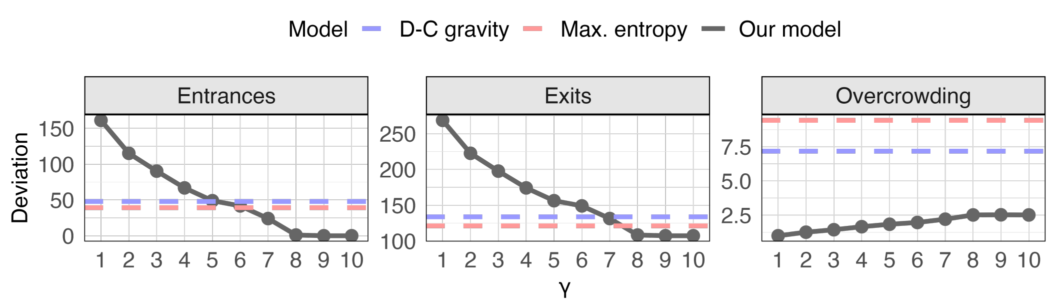

Figure 1: Our demand estimation method (gray) compared to the maximum entropy (red) and doubly-constrained gravity (blue) models for average deviation in entrances, exits, and overcrowding metrics.

Figure 1 shows the values of each metric as a function of γ for each model. As expected, for low values of γ, the benchmarks match the coarse numerical data better than our approach by ensuring greater consistency with total station entrances and exits. However, they also lead to a large discrepancy between passenger demand and the signal from qualitative data. Therefore, for higher values of γ ($\gamma \geq 8$), our method achieves greater consistency with observed station demand, while maintaining consistency with reported overcrowding.

3.2 Practical impact

We next illustrate the practical impact of re-optimizing transit frequencies using qualitative data. In doing so, we compare our optimized frequencies, which rely on demand information estimated from qualitative data, with the current ("historical") WMATA transit frequencies.

The impact is significant: our model using qualitative data results in a 7% total waiting time improvement, primarily driven by re-balancing supply with demand. To illustrate why, we consider the total waiting times for each line (Figure 2). Across transit lines, the improvements (orange, blue, and green) and reductions (silver and yellow) are driven by the waiting times under the current system. That is, lines with higher waiting times for the historical solution are prioritized for service improvements, reflecting a greater need to balance demand and service levels. These effects are particularly pronounced for the orange and blue lines, where increased frequencies not only lead to higher ridership, but also lower total waiting time.

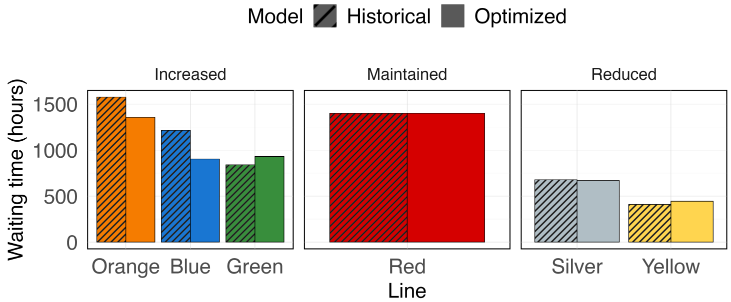

Figure 2: Total waiting time (hours) for each line for the historical and optimized frequencies, separated by whether the optimized model increased, maintained, or reduced frequency.

3.3 Benefits of uncertainty

To evaluate the benefits of incorporating topic probabilities into optimizing transit frequencies, we compare the stochastic estimate-then-optimize method against the deterministic method that selects feedback instances whose highest probability topic is "Rush Hour." The stochastic method achieves a 1% improvement in total waiting time, primarily driven by the underestimation of overcrowding on the yellow line in the deterministic model. As shown in Figure 3, since the yellow line has more instances with lower "Rush Hour" topic probabilities, the stochastic model can capture this variability. In doing so, there is an incentive to increase frequency for the yellow line in the stochastic model. This requires reducing frequencies on other lines; in this example, the orange line decreases its frequency due to its relatively low total waiting times. While the impacts are modest (1%), incorporating topic probabilities still offers practical benefits, totaling 1260 hours per day on average.

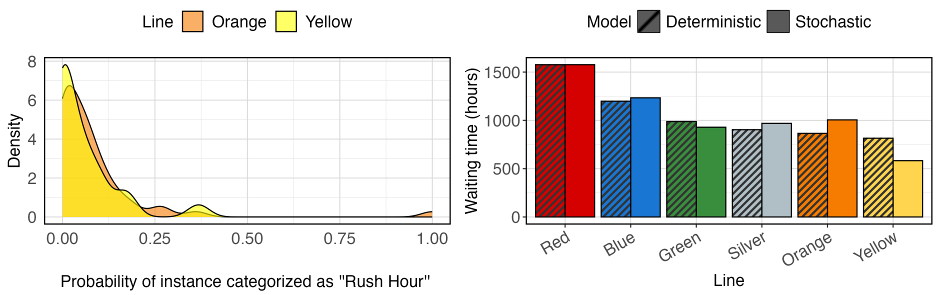

Figure 3: "Rush Hour" topic probability for instances associated with the orange and yellow line (left). Total waiting time (hours) for the deterministic estimate-then-optimize and stochastic estimate-then-optimize models (averaged across scenarios) by line (right).

4 Conclusion, limitations, and future work

This work provided a stochastic estimate-then-optimize approach for integrating qualitative data into decision-making for transit service design. Our estimation phase used qualitative instances of reported overcrowding, along with coarse numerical data, to estimate origin-destination demand. To ensure our downstream decisions were robust against uncertainty in topic classification, we leveraged the topic probabilities to generate demand scenarios, which were used as inputs in a two-stage stochastic optimization. By leveraging qualitative data, we found a 7% total waiting time improvement.

In our ongoing future work, we plan to expand on both the demand estimation and stochastic optimization phases. In our demand estimation phase, our ongoing work develops a theoretical framework to consider the impact of selection bias, inherent in qualitative data inputs, on demand estimation quality. In doing so, we use observed demand data to make distributional assumptions and predict overcrowding, thus generalizing to cities without observed demand data. In our optimization phase, one limitation is that we only incorporate uncertainty from topic probabilities, so our future work will incorporate more sources of uncertainty and alternative methods to model uncertainty.

References

[1] Erwin Van der Laan, Jan van Dalen, Michael Rohrmoser, and Rob Simpson. Demand forecasting and order planning for humanitarian logistics: An empirical assessment. *Journal of operations management*, 45:114–122, 2016.

[2] Dimitris Bertsimas, Yee Sian Ng, and Julia Yan. Data-driven transit network design at scale. *Operations Research*, 69(4):1118–1133, 2021.

[3] McKinsey & Company. Finding a route to fiscal stability for u.s. transit agencies. `https://www.mckinsey.com/industries/infrastructure/our-insights/finding-a-route-to-fiscal-stability-for-us-transit-agencies`, 2025. Accessed: September 5, 2025.

[4] Metro Magazine. Competitive mass transit: Using technology and data to revolutionize city commute. `https://www.metro-magazine.com/10237472/competitive-mass-transit-using-technology-and-data-to-revolutionize-city-commute`, 2025. Accessed: September 5, 2025.

[5] Michael Leong, Awad Abdelhalim, Jude Ha, Diane Patterson, Gabriel L Pincus, Anthony B Harris, Michael Eichler, and Jinhua Zhao. Metroberta: Leveraging traditional customer relationship management data to develop a transit-topic-aware language model. *Transportation Research Record*, 2678(9):215–229, 2024.

[6] Washington Metropolitan Area Transit Authority. Metrorail ridership summary. `https://www.wmata.com/initiatives/ridership-portal/Metrorail-Ridership-Summary.cfm`, 2025. Accessed: 2025-09-05.

[7] Dimitris Bertsimas and Julia Yan. From physical properties of transportation flows to demand estimation: An optimization approach. *Transportation Science*, 52(4):1002–1011, 2018.

[8] Adam N Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1):9–26, 2022.

[9] Priya Donti, Brandon Amos, and J Zico Kolter. Task-based end-to-end model learning in stochastic optimization. *Advances in neural information processing systems*, 30, 2017.

[10] Washington Metropolitan Area Transit Authority (via Transit.land). Gtfs feed: f-dqc-wmata~rail (metrorail gtfs static). `https://www.transit.land/feeds/f-dqc-wmata~rail`, 2025. Last fetched by Transit.land: 2025-09-05; feed accessed static GTFS URL; language: en.

A Topic modeling

We use tweets and WMATA customer resource management data from Leong et al. [5], which include feedback instances from October 2023 – March 2024. Each feedback instance is associated with a transit station, a transit line, and an hour at which the feedback instance was submitted. While the dataset already contains a large number of manually labeled topics (61 distinct topics), there were too many topics to use as is. As a result, we implemented a topic model to categorize each instance into broader transit-specific categories.

In particular, we performed topic modeling using BERTopic. BERTopic transforms feedback instances into numerical representations using embeddings, which incorporates semantic meaning, and then clusters embeddings to identify similar feedback instances. Before implementing BERTopic, we removed niche categories from the original manually labeled topics and removed stopwords, punctuation, symbols, links, person names, and WMATA-specific terms. Furthermore, we performed stemming and lemmatization and removed feedback instances where there were fewer than three words. We fine-tuned the parameters associated with the minimum cluster size, the number of outliers, the number of topics, and the embedding model. Finally, we performed hierarchical clustering to group similar topics and manually label the final topic clusters. Altogether, this procedure results in 20 topics.

For our demand estimation model, we consider feedback instances whose topic corresponds to "Rush Hour" in order to represent instances where demand exceeds capacity. Further, we only consider text feedback instances that had a valid station and line pair and that were submitted between 5 – 9 A.M. on weekdays. Note that Leong et al. [5] performed sentiment analysis (positive, negative, or neutral) on these data using a RoBERTa sentiment model; we use these sentiment predictions to remove instances corresponding to positive sentiment.

B Notation

We first define the following notation used in our demand estimation and stochastic optimization phases. We have the following sets:

$\mathcal{P}$: passengers
$\mathcal{L}$: transit lines
$\mathcal{F}$: rail service frequencies
F_p: rail frequencies that can serve passenger p
$\mathcal{Q}_\ell$: vehicles required to operate line ℓ at the current rail frequency
$\mathcal{N}_{q\ell}$: transit nodes for vehicle $q \in \mathcal{Q}_\ell$
$\mathcal{S}$: set of transit stations
$\mathcal{H}$: set of hours within the time period
Ω: set of demand scenarios

Each passenger $p \in \mathcal{P}$ corresponds to a unique departure time as well as an origin and destination station. We define the following parameters:

E^n_s: average entrances for station s for the time period
E^x_s: average exits for station s for the time period
K: vehicle capacity
$c^\omega_{nq\ell}$: overcrowding at node n for vehicle $q \in \mathcal{Q}_\ell$ for scenario ω
$R^\omega_{nq\ell}$: recurrence of overcrowding at node n for vehicle $q \in \mathcal{Q}_\ell$ over the time period for scenario ω
γ: objective function coefficient
$\rho_{\ell f}$: number of vehicles needed to operate line ℓ at frequency f
V: total vehicles available
$w_{\ell f p}$: waiting time for passenger p for line ℓ and frequency f
$v_{\ell f p}$: in-vehicle time for passenger p for line ℓ and frequency f

E^n_s and E^x_s are calibrated from historical rail ridership data, using publicly available data from Washington Metropolitan Area Transit Authority [6]. The parameters $c^\omega_{nq\ell}$ and $R^\omega_{nq\ell}$ are calibrated from feedback instances, described in Section E.1.

C Stochastic estimate-then-optimize formulation

C.1 Demand estimation

We seek to estimate origin-destination-departure time demand based on information from tweets and customer resource management feedback data. Given $c^\omega_{nq\ell}$ and $R^\omega_{nq\ell}$ as inputs, we estimate D^ω_p for each scenario ω. The decision variables are the following:

D^ω_p: demand for passenger $p \in \mathcal{P}$ in scenario ω
$W^\omega_{nq\ell}$: load for node n for vehicle $q \in \mathcal{Q}_\ell$ in scenario ω
$\epsilon^{n,\omega}_s$: slack variable for entrances for station s in scenario ω
$\epsilon^{x,\omega}_s$: slack variable for exits for station s in scenario ω

The first set of constraints respects the load at each node for each vehicle across all scenarios.

$$W^\omega_{i(q),q,\ell} = \sum_{\substack{p \in \mathcal{P}: \\ \mathrm{sn}(p)=i(q)}} D^\omega_p, \quad \forall q \in \mathcal{Q}_\ell, \forall \ell \in \mathcal{L}, \forall \omega \in \Omega \tag{1}$$

$$W^\omega_{n,q,\ell} = W^\omega_{g(n),q,\ell} + \sum_{\substack{p \in \mathcal{P}: \\ \mathrm{sn}(p)=n}} D^\omega_p - \sum_{\substack{p \in \mathcal{P}: \\ \mathrm{en}(p)=n}} D^\omega_p, \quad \forall n \in \mathcal{N}_{q\ell}, \forall q \in \mathcal{Q}_\ell, \forall \ell \in \mathcal{L}, \forall \omega \in \Omega \tag{2}$$

In (1), we define $i(q)$ as the initial node for vehicle $q \in \mathcal{Q}_\ell$. (2) ensures that the load at node n is equivalent to the load at the preceding node in the vehicle sequence $g(n)$, plus any boardings at node n, minus any alightings at node n. We denote $\text{sn}(p)$ and $\text{en}(p)$ as the start and end transit nodes for passenger p, which is calculated by considering the lowest cost (i.e., lowest combined waiting time and travel time) transit route for passenger p. If passenger p can travel on more than one transit line from the origin to destination station, we assign passenger p to their lowest cost transit route.

The next set of constraints ensures consistency with total entrances and exits per station.

$$\left| \sum_{\substack{p \in \mathcal{P}: \\ u(p)=s}} D_p^\omega - E_s^n \right| \leq \epsilon_s^{n,\omega}, \quad \forall s \in \mathcal{S}, \forall \omega \in \Omega \tag{3}$$

$$\left| \sum_{\substack{p \in \mathcal{P}: \\ v(p)=s}} D_p^\omega - E_s^x \right| \leq \epsilon_s^{x,\omega}, \quad \forall s \in \mathcal{S}, \forall \omega \in \Omega \tag{4}$$

In the above equations, we define $u(p)$ and $v(p)$ as the stations corresponding to the origin and destination stations for passenger p. We ensure that the total boardings (alightings) at a station should be aligned with the number of entrances (exits) at that station within a tolerance.

The objective function consists of two terms: in the first term, we seek to minimize the total slack for entrances and exits. The second term minimizes the deviation in passengers exceeding vehicle capacity and the crowding information given from text data.

$$C_w = \min \gamma \cdot \sum_{s \in \mathcal{S}} (\epsilon_s^{n,\omega} + \epsilon_s^{x,\omega}) + \cdot \sum_{\ell \in \mathcal{L}} \sum_{q \in \mathcal{Q}_\ell} \sum_{n \in \mathcal{N}_{q\ell}} R_{nq\ell}^\omega \cdot \left| \left((W_{nq\ell}^\omega - K)^+ - c_{nq\ell}^\omega \right) \right|$$

Because there are no first-stage decisions, each scenario ω can be treated independently.

C.2 Frequency optimization

Given D_p^ω, we optimize frequencies via a two-stage stochastic optimization model where the first-stage variables determine the transit frequencies, and the second-stage variables allocate passengers to transit routes to minimize total passenger waiting time and in-vehicle travel time. We define the following decision variables:

$$x_{\ell f}\colon \begin{cases} 1 & \text{line } \ell \in \mathcal{L} \text{ operates at frequency } f \in \mathcal{F} \\ 0 & \text{otherwise} \end{cases}$$

$z_{\ell f p}^\omega$: number of passengers p served on line ℓ with frequency f for scenario ω

The first stage problem is written as the following, where (5a) ensures that we can select at most one frequency per line. (5b) ensures that we cannot use more than the total number of vehicles available, while (5c) enforces binary values for the frequency variables.

$$\min \quad \mathbb{E}[Y(\boldsymbol{x})]$$

$$\sum_{f \in \mathcal{F}} x_{\ell f} \leq 1, \quad \forall \ell \in \mathcal{L} \tag{5a}$$

$$\sum_{\ell \in \mathcal{L}} \sum_{f \in \mathcal{F}} \rho_{\ell f} x_{\ell f} \leq V \tag{5b}$$

$$x_{\ell f} \in \{0, 1\} \quad \forall \ell \in \mathcal{L}, \forall f \in \mathcal{F} \tag{5c}$$

The recourse function is the following, where $\boldsymbol{x}$ is the first-stage variable. For the second-stage variables $\boldsymbol{z}$, we ensure that the constraints hold for all scenarios and minimize total passenger waiting time and in-vehicle travel time.

$$\mathbb{E}[Y(\boldsymbol{x})] = \min_{z^{\omega}_{\ell f p}} \sum_{\omega\in\Omega} \sum_{p\in\mathcal{P}^{\omega}} \sum_{\ell\in\mathcal{L}} \sum_{f\in\mathcal{F}_p} \zeta^{\omega} \cdot z^{\omega}_{\ell f p} \cdot \left(w^{\omega}_{\ell f p} + v^{\omega}_{\ell f p}\right) \tag{6a}$$

$$z^{\omega}_{\ell f p} \leq D^{\omega}_{p} x_{\ell f}, \quad \forall \ell\in\mathcal{L}, f\in\mathcal{F}_p, p\in\mathcal{P}^{\omega}, \omega\in\Omega \tag{6b}$$

$$\sum_{\ell\in\mathcal{L}} \sum_{f\in\mathcal{F}_p} z^{\omega}_{\ell f p} = D^{\omega}_{p}, \quad \forall p\in\mathcal{P}^{\omega}, \omega\in\Omega \tag{6c}$$

$$z^{\omega}_{\ell f p} \in \{0,1\} \quad \forall \ell\in\mathcal{L}, f\in\mathcal{F}_p, p\in\mathcal{P}^{\omega}, \omega\in\Omega \tag{6d}$$

In (6a), we minimize the waiting and in-vehicle travel time for all passengers where ζ^{ω} denotes the probability of realizing scenario ω. In (6b), we ensure that we only serve passengers on line ℓ with frequency f if line ℓ and frequency f are selected. (6c) ensures all passengers for each scenario are served. We impose binary values for the second stage variable $\boldsymbol{z}$ in (6d).

D Demand estimation benchmarks

We consider two benchmarks to evaluate our demand estimation method against: the maximum entropy model and doubly-constrained gravity model, classical methods in transportation demand modeling. This section details both of these model formulations and the evaluation metrics used to compare the benchmarks with the original demand estimation model.

D.1 Maximum entropy

The maximum entropy demand estimation model seeks the "most likely" demand distribution. We define the decision variable D^{ME}_{od} as the demand between stations o and d. Note that we do not estimate demand at the departure time-level (this is determined post-processing). To ensure consistency with total entrances and exits per station, we define SE^{n}_{s} and SE^{x}_{s} as the *scaled* entrances and exits, such that the total entrances and total exits are equal. This allows us to model the problem without slack variables and enforce hard constraints.

We adapt the maximum entropy model to our setting as a constrained nonlinear optimization model. The objective function maximizes entropy, and the constraints ensure that the total entrances and exits are consistent with observed data.

$$\max \sum_{o,d} -D^{\text{ME}}_{od} \log(D^{\text{ME}}_{od})$$

$$\sum_{d\in\mathcal{S}} D^{\text{ME}}_{od} = SE^{n}_{o}, \quad \forall o\in\mathcal{S}$$

$$\sum_{o\in\mathcal{S}} D^{\text{ME}}_{od} = SE^{x}_{d}, \quad \forall d\in\mathcal{S}$$

After obtaining D^{ME}_{od}, we determine passenger departure times by equally allocating passengers to departure hours $h\in\mathcal{H}$, thus obtaining demand D^{ME}_{p} at the passenger p-level.

D.2 Doubly-constrained gravity model

The second benchmark model integrates the idea of the maximum entropy model (i.e., ensuring consistency with total entrances and exits to obtain the most likely demand distribution) with travel time impedance. This model builds upon the gravity model, where demand between two locations is assumed to be proportional to the attractiveness of the two locations and the amount of physical separation. Our version of the gravity model is considered doubly-constrained, because we constrain both origin and destination totals to equal observed entrances and exits data.

We define the following notation:

D^{G}_{od}: demand between origin o and destination d

A_o: production potential at station o

B_d: attraction potential at station d

F_{od}: willingness to travel between o and d

β: deterrence parameter for the impedance function

The traditional gravity model assumes that demand is proportional to the production and attraction potential at the origin and destination stations, as well as the willingness to travel, given by the following equation:

$$D^{\mathrm{G}}_{od} = A_o B_d F_{od} \tag{8}$$

The doubly-constrained model adds the following constraints:

$$\sum_d D^{\mathrm{G}}_{od} = SE^n_o, \quad \forall o \in \mathcal{S}$$
$$\sum_o D^{\mathrm{G}}_{od} = SE^x_d, \quad \forall d \in \mathcal{S}$$

Starting with the total entrances, we can write:

$$\sum_d D^{\mathrm{G}}_{od} = \sum_d A_o B_d F_{od} = SE^n_o, \quad \forall o \in \mathcal{S}$$

Rearranging the last summation, this implies that:

$$A_o = \frac{SE^n_o}{\sum_d B_d F_{od}} \quad \forall o \in \mathcal{S} \tag{9}$$

For the total exits, we can write:

$$\sum_o D^{\mathrm{G}}_{od} = \sum_o A_o B_d F_{od} = SE^x_d, \quad \forall d \in \mathcal{S}$$

Rearranging the last summation, this implies that:

$$B_s = \frac{SE^x_d}{\sum_o A_o F_{od}} \quad \forall d \in \mathcal{S} \tag{10}$$

Combining (9) and (10) with (8), we have that:

$$D^{\mathrm{G}}_{od} = \frac{SE^n_o}{\sum_{d'} B_{d'} F_{od'}} \cdot \frac{SE^x_d}{\sum_{o'} A_{o'} F_{o'd}} \cdot F_{od} \tag{11}$$

To determine the parameters A_o and B_d, we run the following algorithm to iteratively update each parameter until convergence:

1. Initialize A_o and B_d with values 1.
2. Update the parameters:
$$A_o = \frac{1}{\sum_d B_d SE^x_d F_{od}} \quad \forall o \in \mathcal{S}$$
$$B_d = \frac{1}{\sum_o A_o SE^n_o F_{od}} \quad \forall d \in \mathcal{S}$$
3. Repeat until D^{G}_{od} according to (11) converges to stable values (i.e., a maximum difference of 1 passenger for o and d).

We model F_{od}, the willingness to travel between origin o and destination d, using a negative exponential function. That is, we let $F_{od} = \exp(-\beta c_{od})$, where β is a deterrence parameter and c_{od} is the travel time between origin station o and destination d. This ensures that the willingness to travel decreases as travel time increases according to a negative exponential function.

Similar to the maximum entropy model, after obtaining D^{G}_{od}, we determine passenger departure times by equally allocating passengers to departure hours $h \in \mathcal{H}$, thus obtaining demand D^G_p at the passenger p-level.

D.3 Evaluation metrics

We consider the following evaluation metrics on which to compare our demand estimation method with the benchmark models (maximum entropy and doubly-constrained gravity). We illustrate these metrics for one scenario ω, as reported in Sections 3.1 and 3.2.

- **Average deviation in station entrances:** We first measure the average deviation in passengers originating at a transit station and the observed entrances, which is equivalent to the average value over all ϵ_s^n. We calculate this metric as follows:

$$\hat{E}^n = \frac{\sum_{s \in \mathcal{S}} \left| \sum_{p:u(p)=s} D_p - E_s^n \right|}{|\mathcal{S}|}$$

- **Average deviation in station exits:** Similarly for exits, we calculate the following metric:

$$\hat{E}^x = \frac{\sum_{s \in \mathcal{S}} \left| \sum_{p:v(p)=s} D_p - E_s^x \right|}{|\mathcal{S}|}$$

  This is equivalent to the average value over all ϵ_s^x.

- **Average deviation in overcrowding:** We also measure the deviation from reported overcrowding by first calculating the load, as shown in (1) and (2). We then calculate $y_{nq\ell}$ as the following:

$$y_{nq\ell} = \left| \left((W_{nq\ell} - K)^+ - c_{nq\ell} \right) \right|$$

  The final metric is calculated as:

$$\hat{\kappa} = \frac{\sum_{\ell \in \mathcal{L}} \sum_{q \in \mathcal{Q}_\ell} \sum_{n \in \mathcal{N}_{q\ell}} y_{nq\ell}}{\sum_{\ell \in \mathcal{L}} \sum_{q \in \mathcal{Q}_\ell} |\mathcal{N}_{q\ell}|}$$

E Experiment details

E.1 Estimating overcrowding from feedback instances

For each $\omega \in \Omega$, we use reported instances of overcrowding to calibrate the parameters $c^\omega_{nq\ell}$ and $R^\omega_{nq\ell}$ in our demand estimation model, which correspond to the overcrowding at transit node n for vehicle $q \in \mathcal{Q}_\ell$ for line ℓ. Recall that each instance is associated with an hour of the day h, transit line ℓ, and station s, and each transit node corresponds to a transit station and boarding/alighting time, which is at the minute-level. Therefore, we require a procedure to map each feedback instance to transit nodes $n \in \mathcal{N}_{q\ell}$.

The mapping of feedback instances to transit nodes consists of both temporal and spatial extrapolation. For the temporal component, we map each feedback instance to all transit nodes at station s and line ℓ, where the boarding/alighting time of the transit node occurs in hour h. For the spatial component, we extrapolate overcrowding instances at station s on line ℓ to nearby stations on the same transit line. We propose two methods for doing this: a Gaussian and Uniform kernel, each associated with a particular bandwidth, scaled by a fixed overcrowding amount. We present additional results with the Uniform kernel in Section F to demonstrate the robustness of our method to the choice of kernel smoothing. For the parameter $R^\omega_{nq\ell}$, we count the number of times node n is assigned a non-zero overcrowding amount.

E.2 Estimate-then-optimize parameters

For demand estimation, we use the Gaussian kernel with a bandwidth of 2 and a fixed overcrowding amount of 100 passengers. We let K (capacity of each train) be 600 passengers. For Sections 3.1 and 3.2, we use feedback instances whose highest topic probability was "Rush Hour" to estimate the overcrowding and recurrence parameters in our demand estimation model. We also set a time limit of 1 hour when solving each instance.

For frequency optimization, we consider the six transit lines in Washington D.C.: $\mathcal{L} = \{\text{Red, Orange, Yellow, Green, Blue, Silver}\}$. We also consider a discrete transit frequency set

$\mathcal{F} = \{3, 5, 7, 9, 11, 13, 15\}$ for each transit line, denoting the headways in minutes (e.g., 3, 5, etc.) for each transit line. We calculate transit boarding and alighting times, as well as in-vehicle travel times, using the travel times from the WMATA General Transit Feed Specification [10]. We let $V = 96$, which is the total fleet size required to operate the current WMATA frequencies for 5:00 – 9:30 A.M. For Sections 3.2 and 3.3, we use $\gamma = 9$.

F Additional results

We provide another comparison between our demand estimation methods and numerical data-only methods (i.e., the maximum entropy and doubly-constrained gravity models) using the Uniform kernel (with a bandwidth of 2), as discussed in Section E.1. Figure 4 shows how our demand estimation method outperforms both demand estimation benchmarks across all three metrics for values of $\gamma \geq 5$. Similar to Figure 1, our method achieves greater consistency with observed station demand, while maintaining consistency with reported overcrowding. This further shows the robustness of how our demand estimation method achieves superior performance as compared to benchmarks.

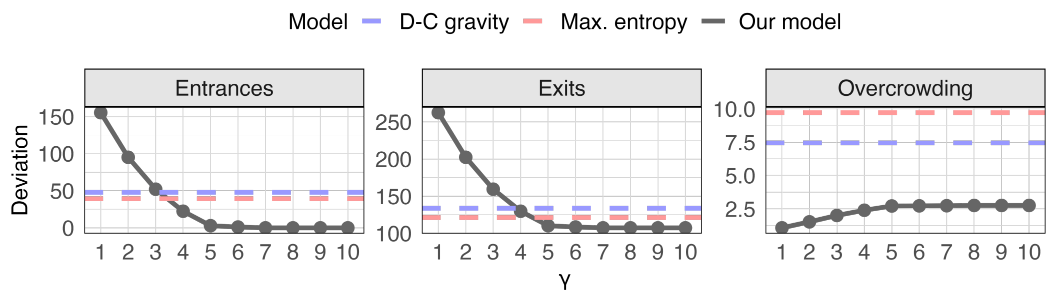

Figure 4: Our demand estimation method (gray) using the Uniform kernel compared to the maximum entropy (red) and doubly-constrained gravity (blue) models for average deviation in entrances, exits, and overcrowding metrics as a function of γ. Note that the vertical axis scale varies across plots.